

10th International Conference
Statistical Analysis of Textual Data
SAPIENZA University of Rome
(11 june 2010)

Ce que disent les phrases de Corneille et Molière

Dominique LABBE
(PACTE – IEP Grenoble)
dominique.labbe@iep-grenoble.fr

Abstract

The sentence lengths of the plays of the French 17th century depend on the authors, except for Corneille and Molière who seem to have done the same stylistic choices. This atypical identity is confirmed by several statistical indices: intertextual distances, classifications, combinations of the most usual words, meanings of keywords. Testimonies of some contemporaneous confirm that Molière was not the author of the plays he presented and three of them pointed out that Corneille wrote *le Dépit amoureux*, *le Bourgeois gentilhomme* and *Psyché*. Corneille and Molière conformed to the standards: during the second half of the French 17th century, 9 out of 10 comedies were presented by some comedians – like Molière – and not by the writers who wrote them.

Résumé

Dans le théâtre français du XVII^e siècle, les longueurs de phrases singularisent chacun des auteurs, sauf Corneille et Molière. Cette proximité est confirmée par d'autres indices : distances entre textes, classifications, combinaisons des mots fréquents, sens des vocables usuels. On rappelle ensuite le témoignage de plusieurs contemporains qui indiquent clairement que Molière n'est pas l'auteur des pièces qu'il présentait et qui désignent P. Corneille comme étant sa plume de l'ombre dans trois cas (*le Dépit amoureux*, *le Bourgeois gentilhomme*, *Psyché*). Il s'agissait d'un système : à cette époque, neuf comédies sur dix n'ont pas été présentées par leur auteur mais par un comédien poète, comme Molière.

Key Words: French Theatre ; French 17th century ; authorship attribution ; sentence lengths ; Corneille ; Molière

Mots clefs : Théâtre ; XVII^e siècle français ; attribution d'auteur ; longueur des phrases ; Corneille ; Molière.

Introduction

Notre recherche - publiée sous le titre *Ce que disent leurs phrases* (Labbé & Labbé 2010) - examine la longueur des phrases dans le théâtre du XVII^e siècle. Elle aboutit aux conclusions suivantes.

Premièrement, ces pièces sont composées de quatre sortes de phrases, soit par longueur croissante :

- l'interpellation et l'interjection (phrases très brèves),
- la conversation courante (phrases courtes),
- le récit (longueur moyenne),
- le soliloque et le monologue (phrases longues).

Deuxièmement, chacun de ces types de phrases se caractérise par une densité particulière des catégories grammaticales et par une construction plus ou moins complexe.

Troisièmement, dans un genre donné (comédie ou tragédie) chaque auteur privilégié tel ou tel type de phrase. Dès lors la longueur des phrases différencie les auteurs et met en lumière un cas singulier : Corneille et Molière.

La première partie de notre communication examine ce point plus en détail ; les deux parties suivantes présentent les indices lexicaux et historiques qui viennent éclairer cette question.

I. Longueurs de phrases et attribution d'auteur

Les valeurs centrales - mode, médiane, moyenne et médiale, rappelées en annexe VI – montrent que les longueurs de phrases sont une caractéristique propre au style d'un auteur, du moins quand il écrit dans un genre donné.

Prenons d'abord l'exemple des tragédies en alexandrins de Corneille (annexe 1) et de Racine (annexe III). Ces pièces sont contemporaines et le genre tragique est soumis à des contraintes particulièrement fortes (les fameuses « règles » et la versification). Malgré ces contraintes, les différences de longueurs de phrases entre Corneille (A) et Racine (B) sont-elles significatives ?

La procédure est la suivante :

- Les longueurs de phrases sont rangées par ordre croissant ; on calcule les effectifs de chacune des i classes de longueurs ;

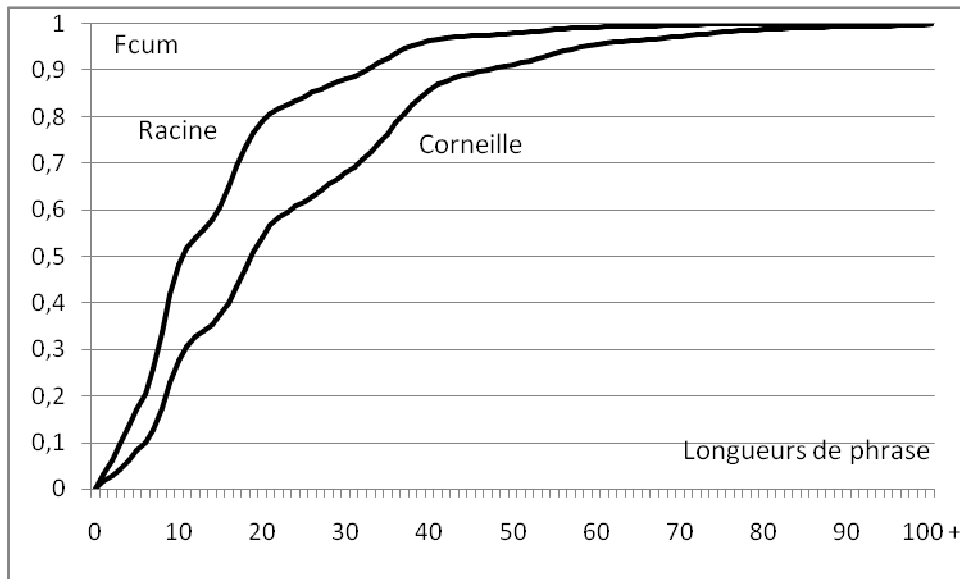
- On convertit les effectifs absolus de chacune des n_a et n_b classes en fréquences relatives (F_i) afin

$$\text{que : } \sum_{i=1}^{i=n_a} F_{i_a} = 1 ; \sum_{i=1}^{i=n_b} F_{i_b} = 1$$

- On cumule ces fréquences : $F_{cum_{i_a}}$; $F_{cum_{i_b}}$

Les deux fonctions de répartition de la variable « longueur de phrase » sont présentées dans le graphique 1. Les deux courbes divergent considérablement. En l'occurrence, Racine ne construit pas ses phrases de la même manière que Corneille. Même quand Racine et Corneille traitent, à la même époque, de sujets proches dans un même genre, les phrases du premier sont significativement plus courtes que celles du second. Comme l'indique notre communication écrite, cela signifie que les pièces de Racine contiennent plus d'interpellations et de conversations courantes que celles de Corneille et moins de récits, de soliloques et de monologues. Ces choix stylistiques sont répétés de pièce en pièce.

Graphique 1 Diagramme des fréquences cumulées. Longueurs de phrases (en mots) dans les tragédies de Corneille et Racine (pour les titres, voir annexes I et III)

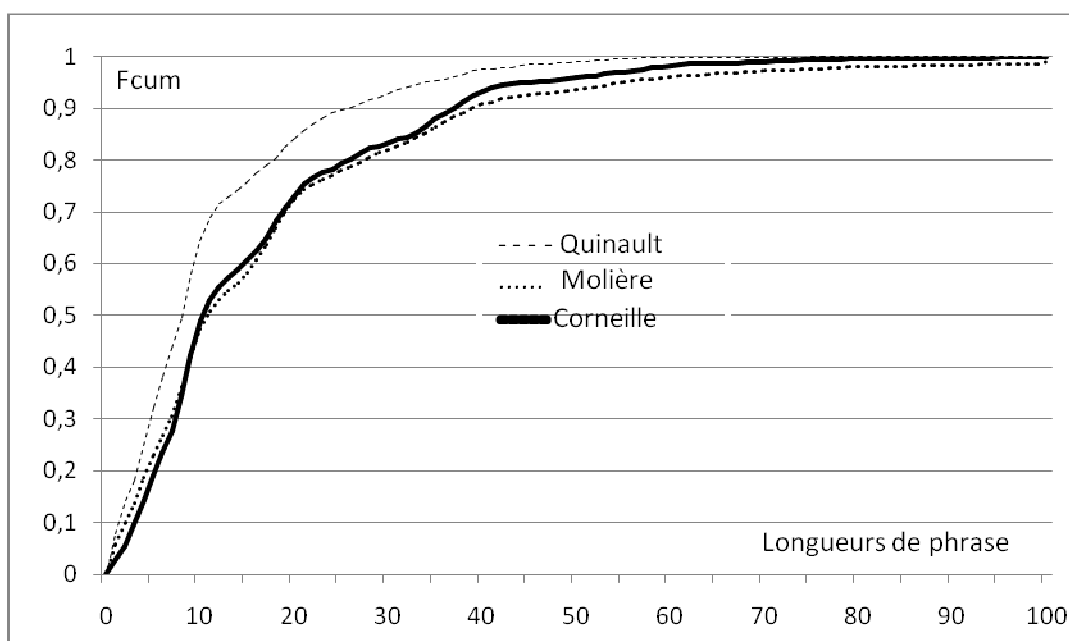


Cette expérience est confirmée par d'autres identiques effectuées sur les 4 siècles de littérature. Elle permet de conclure que, dans un genre donné, à une époque précise, chaque auteur se singularise par des longueurs de phrases différentes, c'est-à-dire par des choix différents en faveur de l'une ou l'autre des fonctions décrites dans la dernière partie de notre communication écrite.

Dans le théâtre du XVIIe, il existe deux exceptions à cette règle. Elles concernent Corneille-Molière :

Première exception : la distribution des longueurs de phrase dans le *Menteur* et la *Suite du Menteur* (P. Corneille) ne diverge pas de celle observée dans les onze comédies en alexandrins présentées par Molière. Cela est vrai pour chacun des 24 couples de pièces considérés séparément. Le graphique 2 ci-dessous en donne un résumé parlant.

Graphique 2. Diagramme des fréquences cumulées. Longueurs de phrases (en mots) dans le *Menteur* et la *Suite du menteur* de P. Corneille, la *Mère coquette* de P. Quinault et les comédies en alexandrins présentées par Molière (pour les titres, voir annexes I, II et V)



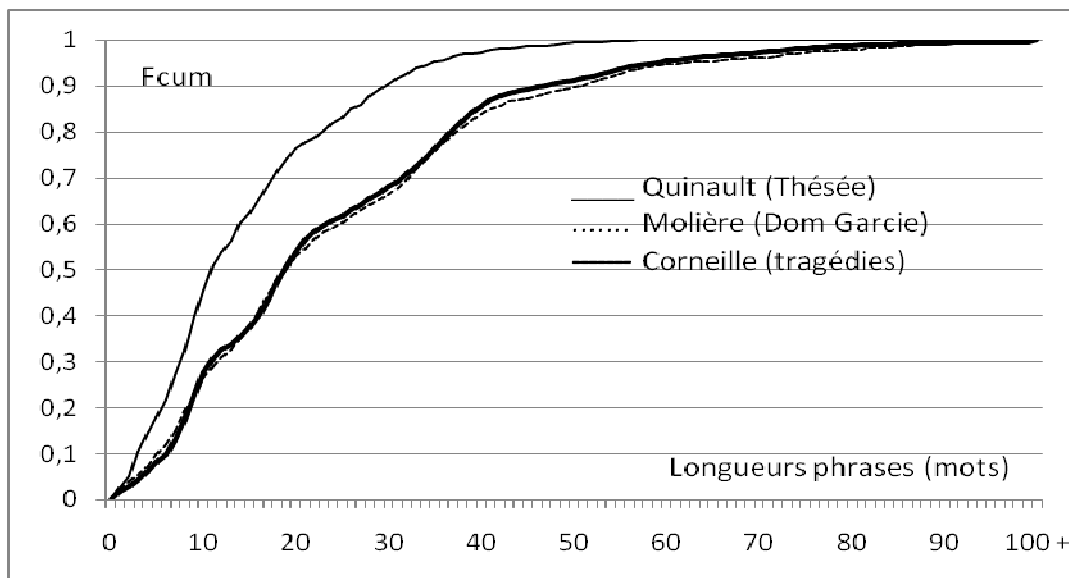
Les deux courbes sont quasiment confondues, y compris dans leurs "accidents" (les modes principaux et secondaires, communs aux deux corpus). Les comédies de Molière utilisent simplement un peu plus d'interjections et d'interpellations – que ne s'en permet Corneille dans son œuvre officielle –, ce qui se traduit par un léger excédent de phrases très courtes.

En revanche, Quinault se sépare nettement des choix communs à Corneille et «Molière». Pourtant la *Mère coquette* (1665) est contemporaine des comédies présentées par Molière et elle a été écrite spécialement pour la compagnie rivale (l'hôtel de Bourgogne) afin de faire pièce au succès des comédies présentées par Molière au Palais royal. A l'inverse, les deux *Menteurs* ont été créés par P. Corneille 15 ans avant la première pièce présentée par Molière...

En tous cas, la similitude de la distribution des longueurs de phrase chez Corneille et Molière ne peut s'expliquer par le genre.

Deuxième exception : la distribution des longueurs de phrases dans *Dom Garcie* (Molière) ne diverge pas significativement de celle observée dans les tragédies (et les tragi-comédies) en alexandrins de P. Corneille (graphique 3). La comparaison Racine/Corneille est déjà donnée dans le graphique 1. Dans le graphique 3, on a ajouté Quinault comme élément de comparaison.

Graphique 3. Diagramme des fréquences cumulées. Longueurs de phrases (en mots) dans les tragédies de P. Corneille, *Thésée* de P. Quinault et *Dom Garcie* présentée par Molière (voir annexes I, II et V)



Quinault utilise plus de phrases courtes que ne le font Corneille et Molière. De nouveau, Corneille et Molière utilisent des longueurs de phrases semblables. Ici ce sont 18 pièces de Corneille qui sont concernées (de *Rodogune* à *Suréna*). Enfin, on obtient exactement le même graphique avec *Psyché*, mais ce n'est pas une surprise puisque le premier éditeur a vendu la mèche : P. Corneille a écrit la grande majorité de cette pièce...

Quelle est la portée de ces expériences ?

En premier lieu, ces résultats ne peuvent être le fait du hasard. Il y a entre Corneille et Molière trop de « proximités » anormales – par rapport à ce que laissent attendre les effectifs concernés - et, de plus, ces proximités se concentrent sur quatre pièces : le *Menteur* et la *Suite du Menteur* d'une part, *Dom Garcie* et *Psyché* d'autre part, au lieu d'être dispersées au hasard sur l'ensemble des deux œuvres. Il s'agit donc d'un système et non de fluctuations aléatoires.

Deuxièmement, qu'en conclure du point de vue de l'histoire littéraire ?

Les dates de création des pièces (annexes I et II) ne laissent pas de doute quant à l'antériorité de P. Corneille par rapport à Molière. Le hasard devant être écarté, il ne reste que deux explications qui s'excluent mutuellement :

- 1 : P. Corneille et Molière sont deux auteurs distincts mais Molière a "imité" les choix stylistiques de Corneille pendant toute son activité créatrice,

- 2 : Corneille a écrit toutes les pièces en vers de Molière ainsi que tout ou partie de certaines pièces en prose, notamment *Dom Juan* et *l'Avare*...

Il existe un très grand nombre d'indices statistiques et historiques convergents - en faveur de la seconde explication - indices que nous allons rappeler maintenant.

II. Quatre indices statistiques convergents

1. La distance intertextuelle

La distance entre deux textes se mesure en superposant ces textes et en comptant le nombre de vocables différents que l'on divise par le nombre total de mots compris dans ces deux textes (le calcul et les propriétés de la distance intertextuelle sont présentés notamment dans : Labbé & Labbé 2001b ; Labbé et Labbé 2003 ; Labbé & Labbé 2006). La distance est nulle si les textes se recouvrent totalement et elle est égale à l'unité s'ils n'ont aucun vocable en commun. L'indice varie uniformément entre 0 (tous les vocables ont les mêmes effectifs dans les deux textes) et 1 (aucun vocable commun).

Un grand nombre d'expériences, menées en suivant les procédures usuelles dans les sciences de l'ingénieur, ont permis de conclure que la distance intertextuelle présente bien les propriétés d'une distance et qu'elle est influencée par quatre facteurs mesurables : le genre, l'époque, l'auteur et le thème (cités par ordre décroissant d'importance). Parmi ces expériences publiées : les entretiens sociologiques (Bergeron & Labbé 2000, Bergeron & Labbé 2004, Labbé & Labbé 2001a), les premiers ministres québécois et canadiens (Labbé & Monière 2000, Labbé & Monière 2003, Labbé & Monière 2008), les premiers ministres français (Labbé 1998), les romanciers et poètes français du XIXe siècle (Labbé & Labbé 2007a), la correspondance de Hugo, Flaubert et Maupassant (Labbé & Labbé 2009), les romanciers anglais du XIXe siècle (Labbé 2007), etc.

Un grand nombre d'expériences en aveugle, organisées dans les conditions les plus sévères, ont montré l'efficacité de la méthode pour l'attribution à un auteur connu de textes d'origine douteuse ou inconnue (par exemple : Labbé 2002, Labbé et Monière 2006, Labbé 2007). Dans un vaste ensemble de textes contemporains, appartenant à un même genre, et à condition que tous les auteurs aient au moins deux textes, les distances les plus faibles¹ concernent les textes de mêmes auteurs, et les distances les plus fortes ceux d'auteurs différents².

Appliquée au théâtre du XVIIe siècle, cette méthode met en lumière des distances « anormalement » faibles entre d'une part le *Menteur* et la *Suite du menteur* (Corneille) et les comédies en alexandrins parues sous le nom de Molière (annexe VII), et d'autre part *Dom Garcie* et *Psyché* (Molière) et toutes les tragédies de P. Corneille (annexe VIII). Là encore, ces distances sont trop nombreuses et systématiquement concentrées sur certaines pièces. On peut donc en conclure que ces textes-là sont trop proches par rapport à ce qui est attendu pour deux auteurs différents.

2. Les classifications

Trois classifications aboutissent aux mêmes conclusions : classification automatique (Labbé & Labbé 2001b), classification arborée (Luong 1988 & 1994 ; Ruhlman 2003 ; Labbé & Labbé 2001b, 2004,

¹ C'est-à-dire inférieures à la moyenne des distances diminuée de deux écarts-types. Etant donné la distribution gaussienne des distances autour de la moyenne (Labbé 2007), il y a moins de 5% de chances de se tromper en considérant que les distances « anormalement » faibles ne peuvent être le fait du hasard.

² C'est-à-dire supérieures à la moyenne des distances augmentée de deux écarts-types. Mêmes remarques que ci-dessus.

2008), classification automatique effectuée sur le premier plan factoriel (Labbé & Labbé, 2007b, p. 95-98). Cette dernière classification prend un relief particulier car elle a été imaginée par l'un de nos critiques selon une méthode préconisée par un autre de ceux-ci. Elle est reproduite en annexe IX.

Dans ces trois expériences, *le menteur et la suite du menteur* (Corneille) sont rattachées à toutes les comédies en alexandrins présentées par Molière ; *Dom Garcie* et les deux parties de *Psyché* – notamment les passages censés avoir été écrits par « Molière » - sont rattachées aux tragédies contemporaines de Corneille...

Deux autres indices statistiques conduisent aux mêmes conclusions.

3. Les combinaisons de mots préférées.

Les combinaisons de mots les plus fréquents employées par Corneille et Molière révèlent les mêmes proximités entre ces deux auteurs supposés (Annexe X).

Racine ne partage avec Corneille et Molière que trois combinaisons (soulignées dans le tableau en annexe X) : "pouvoir voir", "pouvoir faire" et "pouvoir être", mais avec un classement et des densités très différentes. En revanche, Corneille et Molière en ont cinq en commun (en gras) dont les trois premières dans le même ordre et avec des densités voisines (en italiques). Etant donné le nombre des combinaisons possibles, la probabilité pour qu'une telle "coïncidence" survienne au hasard est infinitésimale.

Il existe un seul cas comparable dans les 4 derniers siècles de littérature française : Gary et Ajar³. Depuis huit ans, personne n'a pu en trouver un autre concernant deux auteurs réellement différents...

4. Le sens des mots usuels.

Autre indice intéressant : le sens spécifique que chaque auteur donne aux principaux mots qu'il emploie. Grâce à l'étude des réseaux sémantiques, nous pouvons affirmer que, chez Corneille et Molière, les principaux vocables ont le même sens, ou plutôt, que ceux de Molière s'inscrivent comme un sous-ensemble dans ceux de Corneille. Le plus évocateur est le mot "amour" – substantif le plus employé par Corneille comme par Molière (Labbé & Labbé 2006). Là encore, ces significations sont propres à Corneille et ne se retrouvent pas chez ses contemporains. Cette démonstration a été refaite, en utilisant les « collocations » lors des Journées Internationales d'Analyse des Données Textuelles de Louvain-la-Neuve (Labbé 2004b). La démonstration est en ligne depuis 6 ans et elle n'a pas été contredite à ce jour.

Tous les indices statistiques convergent : on peut attribuer, sans risque d'erreur, à Corneille toutes les pièces en vers présentées par Molière ainsi que *l'Avare*, *Dom Juan* et la plupart du *Bourgeois gentilhomme* et du *Malade imaginaire*.

Avant nous, au moins trois personnes - P. Louÿs⁴, H. Poulaille et H. Wouters - avaient relevé de nombreuses ressemblances troublantes entre les œuvres présentées par Corneille et par Molière. Ils en avaient déduit que Corneille était bien l'auteur des pièces présentées par Molière.

Cette attribution d'auteur est d'autant plus assurée qu'elle est confirmée par de nombreux indices historiques.

III. De nombreux indices historiques

De nombreux indices historiques⁵ viennent confirmer les conclusions tirées des analyses statistiques.

1. La personnalité de Molière

Molière était un financier et un homme de cour. Dans ses deux maisons richement meublées, il y avait

³ Annexe XI. Voir également les combinaisons préférées d'une « plume de l'ombre » qui a travaillé pour deux Premiers ministres québécois (Molière et Labbé 2006).

⁴ Les articles de Pierre Louÿs sont reproduits en annexe de Boissier 2004.

⁵ Ce dossier historique est présenté en détail dans : Labbé 2009.

très peu de livres et aucune pièce, aucun meuble réservés à l'écriture. Molière ne laisse pas de manuscrit, aucune trace de correspondance. Son emploi du temps quotidien – connu grâce au « registre » de La Grange – ne lui permettait pas de composer en moyenne deux pièces de théâtre par an.

Molière ne s'est pas comporté en écrivain. Notamment, il a laissé subsister nombre de coquilles et d'incorrections dans la publication de « ses » pièces dont il n'a livré aucun examen critique.

2. Le témoignage des contemporains

Aucun des contemporains de Molière ne l'a traité en écrivain. Au contraire, de son vivant de nombreuses rumeurs ont couru sur son compte. Par exemple :

- Molière est l'« auteur *prétendu* des *Précieuses ridicules* » (Baudeau de Somaize, 1660), « ses enfants ont plus d'un père » (Donneau de Visé, 1663), « Molière n'est pas une source vive mais un bassin qui reçoit ses eaux d'ailleurs », il est comme ces « ânes seulement capables de porter de grands fardeaux » (Robinet, 1663). Personne ne conteste que ces trois personnes étaient bien informées des coulisses du théâtre de l'époque ;

- Boileau a affirmé, à propos de *l'École des Femmes*, qu'un *P. Cornelius* se cache derrière Molière comme il y avait un *P. Cornelius*⁶ derrière Térence (1663), que Molière ne sait pas faire des vers (1665), qu'il n'est pas l'auteur du *Misanthrope* (1674) et, qu'il est un « faux plaisant, à grossière équivoque, qui pour me divertir, n'a que la saleté » (1674),

- dans sa *Gazette rimée* du 22 novembre 1670, Robinet écrit que le *Bourgeois gentilhomme* est de P. Corneille,

- deux éditeurs indiquent, dans des notes placées au début de la première édition, que P. Corneille est l'auteur de tout ou partie de deux pièces qu'ils publient sous le nom de Molière : le *Dépit amoureux* (1662) et *Psyché* (1670),

- en 1683, l'éditeur de *Don Juan* affirme que Molière a « fait jouer » cette pièce de son vivant et qu'elle est toujours « représentée sous son nom »...

Enfin, d'Aubignac, en 1663, et Boileau en 1674, accusent les frères Corneille (Pierre et Thomas) d'être « saouls de gloire et affamés d'argent », d'être des « écrivains mercenaires » et d'être les auteurs de farces inavouables.

3. Le système du comédien poète

La clef essentielle est fournie par la production théâtrale de ce temps⁷. En effet, durant la seconde moitié du XVIIe siècle, 6 pièces sur 10 – et 9 comédies sur 10 - ne sont pas présentées par les écrivains qui les ont composées mais par des comédiens qui s'en prétendent "auteurs". Les principaux "comédiens poètes", confrères de Molière : Baron, Brécourt, Champmeslé, Dancourt, Desjardins, Dorimond, Hauteroche, La Thuilerie, Montfleury, Poisson, Raisin, Rosimond, Villiers... Desjardins passe pour l'inventeur du système (Nishida 2008). Il était l'un des comédiens de la troupe de l'illustre théâtre où Molière a fait ses débuts en scène (1643-1645).

Ce système du *Comédien poète* donne son titre à une comédie présentée, en 1673, par la troupe de Molière, sous le nom de Montfleury alors qu'elle est de... Thomas Corneille, le frère de Pierre. En effet, T. Corneille, comme les autres écrivains de ce temps, présentait des pièces sérieuses sous son propre nom et des comédies sous le nom de Montfleury puis de Hauteroche. Autres exemples : Boursault avec le comédien Poisson, La Fontaine avec Champmeslé, Le Clerc avec Beauregard...

Ce système s'explique par la situation particulière du théâtre français au XVIIe siècle :

⁶ Publius Cornelius dit Scipion Emilien. Boileau considérait qu'il était l'auteur des comédies de Térence. Le patronyme « Corneille » dérive du nom de famille romain Cornelius.

⁷ Ce recensement a été rendu possible, notamment grâce à un site en ligne anglais ("Cesar") qui présente toutes les pièces de théâtre connues des XVIIe et XVIIIe siècle français avec leur date de création et leur "auteur". La plupart des pièces publiées sont consultables sur le site Gallica de la Bibliothèque Nationale de France.

- les comédies satiriques, comme celles de Molière, plaisaient au public parisien et elles étaient nécessaires à l'équilibre économique des troupes, mais elles étaient condamnées par l'Eglise, par une partie de la Cour et par l'Académie. Les écrivains célèbres qui composaient ces comédies préféraient rester dans l'ombre,

- il n'y avait pas de propriété intellectuelle et les troupes n'avaient pas la personnalité juridique : pour garder le contrôle sur ces textes, un comédien achetait la pièce au nom de la troupe et assurait son exploitation,

- le comédien poète présentait le texte à l'assemblée de la troupe qui décidait si elle le jouerait ; il avançait une partie des sommes nécessaires pour l'achat, pour les décors, la musique, les ballets ; il surveillait la distribution des rôles, les costumes, la mise en scène ; il veillait à ce que la pièce ne soit pas retirée trop vite de l'affiche ; il assistait au partage de la recette qui avait lieu chaque soir après la représentation et touchait la part de l'"auteur" ; il négociait la publication avec les éditeurs et... il encaissait les moqueries et les critiques à la place de l'écrivain.

Conclusions

En ce qui concerne l'attribution d'auteur, la distance intertextuelle - associée à l'échelle normalisée des distances et à la classification arborée – fournissent des outils fiables. Les longueurs de phrases, les associations de mots fréquents, le sens des mots viennent compléter cette analyse. Naturellement, tout cela exige que les textes (y compris la ponctuation) ait été traités selon des normes rigoureuses (Labbé 1990 et 2002a).

En ce qui concerne Corneille et Molière, tous nos travaux sont consultables sur le site "archives en ligne" du CNRS (HAL-SHS) ou à partir de notre page personnelle. Nous avons suivi les procédures classiques en sciences de l'ingénieur. Nos programmes et nos données sont dans le domaine public. La plupart des documents historiques cités sont en ligne, les autres ont été publiés. Tout est donc vérifiable. Enfin, depuis neuf ans, nous proposons à nos contradicteurs de réaliser les expériences de leur choix...

Quelles conclusions tirer du faisceau d'indices précis, sérieux, concordants – et tous vérifiables – que l'on vient de résumer ?

Du volet statistique, on conclut qu'il existe entre certaines œuvres de Corneille et les principales pièces de Molière des proximités uniques dans l'histoire littéraire française, pour deux auteurs contemporains travaillant dans un même genre sur des thèmes proches.

Du volet historique, on conclut que cette proximité est logique puisque Corneille a fait comme les autres écrivains de son temps : il a présenté ses tragédies sous son nom et ses comédies sous le nom d'un comédien poète (Molière).

L'oubli de ce système du comédien poète a permis à la légende du Molière grand auteur de prendre corps, longtemps après sa mort. Cette légende aveugle beaucoup de littéraires et quelques "spécialistes" de "statistique textuelle", dont certains organisateurs de ce congrès. Nous profitons donc de leur présence à Rome pour réaffirmer la conclusion de la démonstration qui vient d'être résumée.

Corneille a écrit les principales pièces présentées sous le nom de Molière :

- Le *Menteur* et la *Suite du Menteur* (Corneille) sont les sœurs aînées de toutes les comédies en vers présentées par Molière (l'*Etourdi*, le *Dépit amoureux*, *Sganarelle*, l'*Ecole des maris*, les *Fâcheux*, l'*Ecole des femmes*, la *Princesse d'Elide*, le *Tartuffe*, le *Misanthrope*, *Mélicerte*, les *Amants magnifiques*, les *Femmes savantes*) ainsi que du *Dom Juan*, de l'*Avare* et d'au moins les quatre cinquièmes du *Bourgeois gentilhomme* et du *Malade imaginaire* ;

- Les tragédies de Corneille (de *Cinna* à *Suréna*) sont les sœurs aînées des deux tragi-comédies présentées par Molière : *Dom Garcie* et *Psyché*.

Références

Nos travaux sont consultables sur le site en ligne HAL-SHS du CNRS.
D'autres documents sont également en ligne sur le site : corneille-moliere.org

- Aubignac François Hédelin, Abbé d' (1663). *Quatrième dissertation sur le poème dramatique servant de réponse aux calomnies de M. Corneille*. Paris : J. du Breil.
- Baudeau de Sommaize (1660). "Préface". *Les véritable précieuses*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome I, p. 35-67.
- Bergeron Jean-Guy & Labbé Dominique (2000). "L'évaluation de la négociation raisonnée par les acteurs. Une analyse lexicométrique". *Communication au XVI^e Congrès international de l'Association internationale des sociologues de langue française*. Québec : juillet 2000, 12 p. Reproduit dans Bernier Colette et Al. *Formation, relations professionnelles à l'heure de la société-monde*. Paris-Québec : L'Harmattan - Les Presses de l'Université Laval, 2002, p. 239-252.
- Bergeron Jean-Guy & Labbé Dominique (2004). "Analyser les entretiens sociologiques". In Purnelle Gérard, Fairon Cédric et Dister Anne (Eds). *Le poids des mots. Actes des 7e journées internationales d'analyse des données textuelles*. Louvain-la-Neuve : Presses Universitaires de Louvain, 2004, p. 136-147.
- Boileau Nicolas (1663). "Stances à M. Molière" in *Les délices de la poésie galante*. Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966, p. 246.
- Boileau Nicolas (1665). "Satire II". Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966, p. 17.
- Boileau Nicolas (1674). "L'art poétique". Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966, p. 178-179.
- Boissier Denis (2004). *L'affaire Molière*. Paris : Jean-Cyrille Godefroy.
- Donneau de Vizé (1663). *Réponse à l'Impromptu de Versailles*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome II.
- Labbé Cyril & Labbé Dominique (2001a). "Discrimination et classement au sein d'un groupe d'entretiens. Le cas du confort électrique". *Communication aux journées d'études du CIDSP*. Grenoble : 9 mars 2001.
- Labbé Cyril & Labbé Dominique (2001b). "Inter-Textual Distance and Authorship Attribution Corneille and Molière". *Journal of Quantitative Linguistics*. 8-3, December 2001, p. 213-231.
- Labbé Cyril & Labbé Dominique (2003). "La distance intertextuelle". *Corpus*. 2, p. 95-118.
- Labbé Cyril & Labbé Dominique (2005). "How to Measure the Meanings of Words ? Amour in Corneille's Work". *Langage Resources Evaluation*. 39, p. 335-351.
- Labbé Cyril & Labbé Dominique (2006). "A Tool for Literary Studies. Intertextual Distance and Tree Classification". *Literary and Linguistic Computing*. 21-3, p. 311-326.
- Labbé Cyril & Labbé Dominique (2007a). "Baudelaire, Rimbaud et Verlaine". VIIIe journées de l'ERLA, *Aspects linguistiques du texte poétique*. Brest 16-17 novembre 2007.
- Labbé Cyril & Labbé Dominique (2007b). *Corneille a écrit 16 pièces représentées sous le nom de Molière. Réponses à : Viprey Jean-Marie et Ledoux Claude-Nicolas, 'About Labbé's "Inter-textual Distance"*. Grenoble : PACTE-IEP.
- Labbé Cyril & Labbé Dominique (2008). "Peut-on se fier aux arbres ?". In Heiden Serge et Pincemin Bénédicte (Eds). *9^e Journées internationales d'analyse statistique des données textuelles (Lyon, 12-14 mars 2008)*. Lyon : Presses universitaires de Lyon, 2008, volume 2, p. 635-645.
- Labbé Cyril & Labbé Dominique (2009). "Existe-t-il un genre épistolaire ? Hugo, Flaubert et Maupassant". *Xe journées de l'Erla*. Brest novembre 2009.
- Labbé Cyril & Labbé Dominique (2010). "Ce que disent leurs phrases". In Bolasco Sergio, Chiari Isabella, Giuliano Luca (Eds). *Proceedings of 10th International Conference Statistical Analysis of Textual Data*. Rome : Edizioni Universitarie di Lettere Economia Diritto, 2010, Vol 1, p. 297-307.
- Labbé Cyril, Labbé Dominique & Monière Denis (1998). " Les styles discursifs des premiers ministres québécois de Jean Lesage à Jean Charest". *Revue canadienne de science politique*. 41:1, mars 2008, p. 43-69.
- Labbé Dominique (1990). *Normes de saisie et de dépouillement des textes politiques*. Grenoble : Cahiers du CERAT.

- Labbé Dominique (1998). "Les déclarations gouvernementales sous la Ve République (1959-1997)". Communication devant le séminaire du DEA Etudes politiques (3 décembre 1998). Publié dans : Autin Jean-Louis et Weill Laurence (Eds). *Le Droit figure du politique. Etudes offertes au professeur Michel Miaille*. Montpellier : Université de Montpellier I, 2008, tome I, p. 843-865.
- Labbé Dominique (2002a). *Qui a écrit quoi ? L'attribution d'auteur et la distance intertextuelle*. Grenoble : CERAT.
- Labbé Dominique (2002b). "La lemmatisation des grandes bases de textes. Un exemple : Corneille, Molière et Racine". Communication au colloque *L'édition électronique en littérature et dictionnaire, évaluation et bilan*. Rouen : 17-21 juin 2002.
- Labbé Dominique (2002c). "La lexicométrie appliquée au discours politique. Le général de Gaulle". Séminaire ARCATI. Paris : décembre 2002.
- Labbé Dominique (2004a). *Romain Gary et Emile Ajar*. Grenoble : Cerat-IEP, mai 2004.
- Labbé Dominique (2004b). *Corneille et Molière. Table ronde 7e Journées d'Analyse des Données Textuelles*. Louvain-la-Neuve 11 mars 2004. Grenoble : CERAT-IEP.
- Labbé Dominique (2007). "Experiments on Authorship Attribution by Intertextual Distance in English". *Journal of Quantitative Linguistics*, April 2007, 14-1. p. 33-80.
- Labbé Dominique (2009). *Qui a écrit Tartuffe ?* Montréal : Monière et Wollank. Réédition : *Si deux et deux sont quatre Molière n'a pas écrit Don Juan*. Paris : Max Milo.
- Labbé Dominique & Monière Denis (2000). "La connexion intertextuelle. Application au discours gouvernemental québécois". In Rajman Martin et Chappelier Jean-Cédric (Eds). *Actes des 5^e journées internationales d'analyse des données textuelles*. Lausanne : Ecole polytechnique fédérale, 2000, vol 1, p 85-94.
- Labbé Dominique & Monière Denis (2003). *Le vocabulaire gouvernemental. Canada, Québec, France (1945-2000)*. Paris : Champion.
- Labbé Dominique & Monière Denis (2006). "L'influence des plumes de l'ombre sur les discours des politiciens". In Condé Claude et Viprey Jean-Marie. *Actes des 8e Journées internationales d'Analyse des données textuelles*. Besançon : 19-21 avril 2006, II, p. 687-696.
- Labbé Dominique & Monière Denis (2008). *Les mots qui nous gouvernent*. Montréal : Monière-Wollank Editeurs.
- Luong Xuan (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Paris : Université de Paris V.
- Luong Xuan (1994). "L'analyse arborée des données textuelles : mode d'emploi". *Travaux du cercle linguistique de Nice*. 16, p. 25-42.
- Nishida Shikiko (2008). "Une réflexion historique et sociologique sur l'activité du premier comédien poète ; Desfontaines". *Cahiers de littérature et langue françaises Université Waseda*. Vol. 27, 2008, p. 65-73.
- Robinet Charles (1663). *Panegyrique de l'Ecole des femmes*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome I.
- Robinet Charles (1670). *Gazette rimée*. Reproduit dans Brooks William. "Le théâtre et l'opéra vus par les gazetiers Robinet et Laurent". *Papers on French Seventeenth Century Literature*. Tome XVII, 1993, 76.
- Ruhlmann Mathieu (2003). *Analyse arborée. Représentation arborée par la méthode des groupements*. Rapport de stage sous la direction de Labbé Cyril et Labbé Dominique. Grenoble : Polytech'Grenoble et Cerat-IEP, août 2003.
- Wouters Hippolyte & Ville de Goyet Christine de (1990). *Molière ou l'auteur imaginaire ?* Bruxelles : Complexe.

Annexes

Annexe I. Les pièces de Pierre Corneille. (sources sous l'annexe II)

		Création	Genre	Longueur en mots
1	Mélite	1630 ?	Comédie	16 690
2	Clitandre	1631	Tragi-comédie	14 402
3	La Veuve	1631	Comédie	17 661
4	La Galerie du Palais	1632	Comédie	16 140
5	La Suivante	1633	Comédie	15 160
6	Comédie des Tuileries*	1634	Comédie	3 627
7	Médée	1635	Tragédie	14 269
8	La Place Royale	1634	Comédie	13 801
9	L'illusion comique	1636	Comédie	15 428
10	Le Cid	1636	Tragi-comédie	16 677
11	Cinna	1641	Tragédie	16 126
12	Horace	1640	Tragédie	16 482
13	Polyeucte	1641	Tragédie	16 472
14	Pompée	1642	Tragédie	16 492
15	Le menteur 1	1642	Comédie	16 653
16	Le menteur 2	1643	Comédie	17 675
17	Rodogune	1644	Tragédie	16 842
18	Théodore	1645	Tragédie	17 121
19	Héraclius	1647	Tragédie	17 433
20	Andromède	1650	Tragédie	15 514
21	Don Sanche	1650	Comédie héroïque	16 947
22	Nicomède	1651	Tragédie	16 923
23	Pertharite	1651	Tragédie	17 121
24	Œdipe	1659	Tragédie	18 618
25	Toison d'Or	1661	Tragédie	20 343
26	Sertorius	1662	Tragédie	17 675
27	Sophonisbe	1663	Tragédie	16 858
28	Othon	1664	Tragédie	16 971
29	Agésilas	1666	Tragédie	18 227
30	Atilla	1667	Tragédie	16 788
31	Tite et Bérénice	1670	Comédie héroïque	16 697
32	Pulchérie	1672	Tragédie	16 630
33	Suréna	1674	Tragédie	16 545
34	Psyché Corneille	1671	Comédie en vers	10 067
35	Psyché Molière	1671	Comédie en vers	4 816
36	Psyché Quinault*	1671	Comédie en vers	1 399

* Pièces retirées des expériences à cause du trop petit nombre de phrases

Annexe II.
L'œuvre théâtrale représentée sous le nom de Molière

		Création	Genre	Longueur mots
1	La jalousie	Avant 1659	Comédie prose	3 501
2	Médecin volant	Avant 1659	Comédie prose	3 876
3	L'étourdi*	1659	Comédie vers	18 671
4	Dépit amoureux*	1659	Comédie vers	16 242
5	Précieuses ridicules	1660	Comédie prose	6 648
6	Sganarelle*	1660	Comédie vers	6 042
7	Dom Garcie*	1661	Comédie héroïque vers	17 049
8	L'école des maris*	1661	Comédie vers	10 536
9	Les fâcheux*	1661	Comédie vers	7 922
10	L'école des femmes*	1662	Comédie vers	16 625
11	Critique de l'école	1663	Comédie prose	8 610
12	L'impromptu	1663	Comédie prose	7 168
13	Mariage forcé	1664	Comédie prose	6 058
14	Princesse d'Elide*	1664	Comédie vers et prose	11 333
15	Le Tartuffe*	1664	Comédie vers	18 271
16	Dom Juan*	1665	Comédie prose	17 452
17	L'amour médecin	1665	Comédie prose	6 147
18	Le Misanthrope*	1666	Comédie vers	17 180
19	Médecin malgré lui	1666	Comédie prose	9 317
20	Mélicerte*	1666	Comédie vers	5 540
21	Comédie pastorale**	1667	Comédie vers libres	732
22	Le sicilien	1667	Comédie prose	5 375
23	Amphytrion*	1668	Comédie vers libres	15 117
24	Georges Dandin	1668	Comédie prose	11 009
25	L'avare*	1668	Comédie prose	21 033
26	M. de Pourceaugnac	1669	Comédie prose	11 803
27	Amants magnifiques*	1670	Comédie vers & prose	11 983
28	Bourgeois gentilhom.*	1670	Comédie prose	17 132
29	Fourberies de Scapin	1671	Comédie prose	14 245
30	Escarbagnas	1671	Comédie prose	5 564
31	Femmes savantes*	1672	Comédie vers	16 863
32	Malade imaginaire*	1673	Comédie prose	19 919

* Pièce écrite, en tout ou partie, par P. Corneille

** Pièce retirée des expériences à cause de sa trop petite taille.

Sources :

Corneille : Charles Marty-Laveaux. *Œuvres complètes de P. Corneille*. Paris : Hachette 1862. Collection Les Grands écrivains de la France.

Molière : Eugène Despois. *Œuvres complètes de Molière*. Paris : Hachette, 1876. Collection Les Grands écrivains de la France.

Annexe III. Les pièces de Jean Racine.

	Titre	Création	Genre*	Longueur en mots
1	La Thébaïde	1664	Tragédie	13 813
2	Alexandre	1665	Tragédie	13 864
3	Andromaque	1667	Tragédie	15 076
4	Les Plaideurs	1668	Comédie	8 041
5	Britannicus	1669	Tragédie	15 387
6	Bérénice	1670	Tragédie	13 242
7	Bajazet	1672	Tragédie	15 297
8	Mithridate	1673	Tragédie	15 091
9	Iphigénie	1674	Tragédie	15 782
10	Phèdre	1677	Tragédie	14 394
11	Esther	1689	Tragédie	11 147
12	Athalie	1691	Tragédie	15 492

* Toutes les pièces de J. Racine sont en alexandrins

Source : M. Paul Mesnard. *Œuvres de J. Racine*. Paris : Hachette, 1885 (Les Grands écrivains de la France).

Annexe IV. Les pièces de J. Mairet

	Titre	Création	Genre	Longueur (mots)
1	Sylvie	1621	Tragi-comédie (alexandrins)	19 813
2	Sophonisbe	1634	Tragédie (alexandrins)	16 166

Sources : *La Sylvie*. Texte établi par Jules Masan. Paris : Société nouvelle d'édition, 1905. *Sophonisbe*. Texte établi par Charles Dédéyan. Paris : Nizet, 1969.

Annexe V. Les pièces de P. Quinault

	Titre	Création	Genre	Longueur (mots)
1	La mère coquette	1665	Comédie (vers libre)	16 130
2	Thésée	1675	Tragédie (vers libres)	8 284
3	Atis	1676	Opéra (vers libres)	7 959

Sources : *La mère coquette*. Édition d'Étienne Gros. Paris : Champion, 1926. *Livrets d'opéra*. Édition de Norman Buford. Paris : Champion, 1979.

Annexe VI. Récapitulatif des valeurs centrales et écart-types dans les corpus étudiés.

	Mode	Médiane	Moyenne	Ecart-type	Médiale
Corneille (<i>Menteurs</i> , comédies en vers)	9	10,3	16,8	15,2	26,8
Molière (comédies en vers)	9	10,6	16,9	17,1	28,1
Corneille (tragédies en vers)	9	18,7	24,2	18,2	34,7
Corneille (comédies en vers)	9	14,2	18,5	14,8	26,3
Corneille (Dissertations en prose)	27 & 41	40,4	44,2	24,8	52,8
Mairet (2 tragédies)	18	18,7	26,3	22,2	36,2
Molière (comédies en prose)	1	8,5	13,6	14,6	21,8
Quinault (2 tragédies en vers)	9	10,6	14,4	10,4	19,3
Quinault (1 comédie en vers)	9	8,1	11,6	11,3	18,2
Racine (tragédies en vers)	9	10,5	15,1	12,2	19,2
Racine (Plaideurs, comédie en vers)	1 & 4	5,4	8,2	8,3	10,9

Annexe VII

Distances séparant les deux *Menteurs* (Corneille) et les *Plaideurs* (Racine) de toutes les pièces de Molière (en gras les valeurs inférieures au seuil de rejet avec $\alpha = 5\%$)

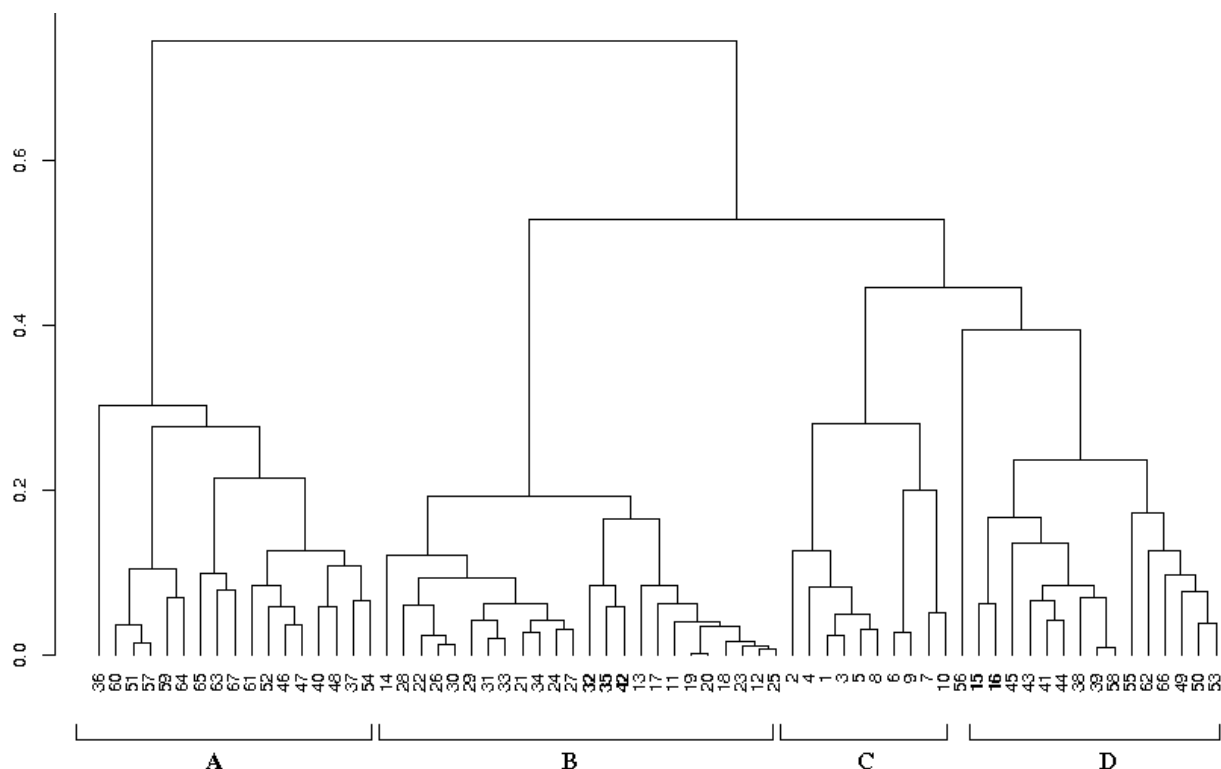
N°	Pièces	Genre	Le Menteur (Corneille 1642)	Suite du Menteur (Corneille 1643)	Les plaideurs (Racine : 1668)
15	Le Menteur (1642)	Vers	0,000	0,180	0,296
16	La suite du Menteur (1643)	Vers	0,180	0,000	0,293
34	Psyché Corneille (1671)	Vers	0,288	0,273	0,348
36	Psyché Molière (1671)	Vers	0,329	0,325	0,354
37	La jalousie du barbouillé (avant 1660)	Prose	0,341	0,331	0,327
38	Médecin volant (avant 1660)	Prose	0,310	0,293	0,302
39	L'étourdi (1658)	Vers	0,205	0,206	0,269
40	Dépit amoureux (1658)	Vers	0,215	0,212	0,270
41	Précieuses ridicules (1660)	Prose	0,315	0,314	0,314
42	Sganarelle ou le cocu imagin. (1660)	Vers	0,259	0,253	0,293
43	Dom Garcie de Navarre (1661)	Vers	0,280	0,273	0,359
44	L'école des maris (1661)	Vers	0,223	0,217	0,279
45	Les fâcheux (1661)	Vers	0,248	0,248	0,306
46	L'école des femmes (1662)	Vers	0,226	0,217	0,261
47	Critique de l'école des femmes (1663)	Prose	0,323	0,319	0,340
48	L'impromptu de Versailles (1663)	Prose	0,321	0,316	0,323
49	Mariage forcé (1664)	Prose	0,322	0,302	0,320
50	Princesse d'Elide (1664)	Vers Prose	0,251	0,243	0,314
51	Le Tartuffe (1664)	Vers	0,242	0,228	0,275
52	Dom Juan (1665)	Prose	0,259	0,248	0,281
53	L'amour médecin (1665)	Prose	0,292	0,289	0,287
54	Le Misanthrope (1666)	Vers	0,252	0,234	0,283
55	Médecin malgré lui (1666)	Prose	0,298	0,289	0,296
56	Mélicerte (1666)	Vers	0,257	0,250	0,322
57	Le sicilien ou l'amour peintre (1667)	Prose	0,277	0,260	0,301
58	Amphytrion (1668)	Vers libres	0,253	0,256	0,297
59	Georges Dandin (1668)	Prose	0,292	0,279	0,292
60	L'Avare (1668)	Prose	0,256	0,244	0,270
61	M. de Pourceaugnac (1669)	Prose	0,292	0,283	0,285
62	Amants magnifiques (1670)	Prose	0,282	0,279	0,329
63	Bourgeois gentilhomme (1670)	Prose	0,294	0,280	0,286
64	Fourberies de Scapin (1671)	Prose	0,269	0,263	0,281
65	Comtesse d'Escarbagnas (1671)	Prose	0,311	0,300	0,305
66	Femmes savantes (1672)	Vers	0,260	0,248	0,283
67	Malade imaginaire (1672)	Prose	0,282	0,270	0,278
	<i>Moyenne oeuvre de Molière</i>		0,275	0,266	0,299
	<i>Moyenne pièces en vers de Molière</i>		0,241	0,234	0,290
	<i>Moyenne oeuvre de Corneille</i>		0,252	0,249	0,347
	<i>Moyenne oeuvre de Racine</i>		0,314	0,311	0,376

Annexe VIII.

Distances séparant *Dom Garcie* (Molière) et *Psyché* (Corneille et Molière) des dernières pièces de Corneille.
En gras les valeurs inférieures au seuil de rejet (avec $\alpha = 5\%$)

Ultimes pièces de Corneille	<i>Dom Garcie</i> (Molière, 1661)	<i>Psyché</i> (Corneille 1671)
Rodogune (1644)	0,245	0,231
Théodore (1645)	0,234	0,245
Héraclius (1647)	0,248	0,273
Andromède (1650)	0,241	0,218
Don Sanche (1650)	0,224	0,251
Nicomède (1651)	0,244	0,264
Pertharite (1651)	0,235	0,263
Œdipe (1659)	0,223	0,226
Toison d'or (1661)	0,221	0,220
Sertorius (1662)	0,230	0,238
Sophonisbe (1663)	0,228	0,236
Othon (1664)	0,235	0,240
Agésilas (1666)	0,234	0,233
Attila (1667)	0,235	0,227
Tite et Bérénice (1670)	0,227	0,235
Psyché (1671)	0,230	—
Pulchérie (1672)	0,230	0,226
Suréna (1674)	0,216	0,224
<i>Moyenne Corneille</i>	0,243	0,244
<i>Moyenne Molière</i>	0,286	0,297

Annexe IX. Classification hiérarchique ascendante sur le premier plan factoriel (Corneille-Molière, 67 pièces, hapax exclus, Labbé & Labbé, 2007b, p 96-97)



Les "anomalies" sont en gras. De gauche à droite sur le dendrogramme :

Groupe A	Groupe B	Groupe C	Groupe D
Molière prose	Corneille tragédies	Corneille premières pièces	Molière Comédies en vers
36. Jalousie du B.	14. Pompée	2. Clitandre	15. Menteur (Corneille)
60. Avare	28. Othon	4. Galerie du Palais	16. Suite du Menteur
51. Dom Juan	22. Nicomède	1. Mélite	45. Ecole des femmes
57. Sicilien	26. Sertorius	3. Veuve	43. Ecole des maris
59. Georges Dandin	30. Atilla	5. Suivante	41. Sganarelle
64. Fourberies de Scapin	29. Agésilas	8. Place royale	44. Fâcheux
65. Escarbagnas	31. Tite et Bérénice	6. Comédie des T.	38. Dépit amoureux
63. Bourgeois gentilhomme	33. Pulchérie	9. Illusion comique	37. Etourdi
67. Malade imaginaire	21. Don Sanche	7. Médée	58. Amphytrion
61. Pourceaugnac	34. Suréna	10. Cid	55. Mélicerte
52. Amour médecin	24. Œdipe		62. Amants magnifiques
46. Critique de l'Ecole	27. Sophonisbe		66. Femmes savantes
47. Impromptu de V.	32. Psyché (Corneille)		49. Princesse d'Elide
40. Précieuses ridicules	35. Psyché (Molière)		50. Tartuffe
48. Mariage forcé	42. Dom Garcie (Molière)		53. Misanthrope
37. Médecin volant	13. Polyeucte		
54. Médecin malgré lui	17. Rodogune		
	11. Cinna		
	19. Héraclius		
	20. Andromède		
	18. Théodore		
	23. Pertharite		
	12. Horace		
	25. Toison d'Or		

Annexe X.

Les combinaisons "pseudo-auxiliaire + infinitif" chez Corneille, Molière et Racine
(fréquence pour 100.000 mots)

P. Corneille		Molière		Racine	
Syntagmes	F	Syntagmes	F	Syntagmes	F
<i>faire voir</i>	33,8	<i>faire voir</i>	31,5	aller voir	12,0
<u><i>pouvoir être</i></u>	18,8	<u><i>pouvoir être</i></u>	25,5	<u>pouvoir voir</u>	9,6
<u><i>pouvoir faire</i></u>	18,4	<u><i>pouvoir faire</i></u>	25,5	faire entendre	9,0
faire naître	13,9	vouloir dire	24,9	<u>pouvoir faire</u>	8,4
<u><i>pouvoir voir</i></u>	13,4	<i>vouloir faire</i>	19,5	aller chercher	7,8
devoir être	12,7	pouvoir dire	14,5	faire parler	7,8
pouvoir souffrir	10,8	pouvoir avoir	13,7	<u>pouvoir être</u>	7,8
<i>vouloir faire</i>	9,9	aller faire	13,2	venir chercher	7,2
faire connaître	9,6	avoir faire	13,2	faire éclater	6,6
devoir faire	8,7	<u><i>pouvoir voir</i></u>	12,3	falloir partir	6,6

Annexe XI. Les principaux groupes verbaux chez Ajar et Gary (fréquence pour 10.000 mots. Labbé 2004a)

Ajar	F	Gary	F
vouloir dire	7,20	vouloir dire	5,80
pouvoir être	4,90	pouvoir être	3,62
pouvoir faire	4,40	pouvoir faire	2,97
devoir être	3,10	devoir être	2,39
aller faire	2,90	aller faire	2,20
pouvoir vivre	2,70	devoir avoir	2,13
laisser tomber	2,30	pouvoir dire	2,13
devoir avoir	2,10	laisser tomber	1,87
aller voir	2,00	laisser aller	1,55
devoir faire	1,80	faire passer	1,49