

JURA TE **CH** RÉPUBLIQUE ET CANTON DU JURA

CEJEF
DIVISION LYCÉENNE
LYCÉE CANTONAL

2, place Blarer-de-Wartensee
Case postale 152
CH-2900 Porrentruy
t +41 32 420 36 80
f +41 32 420 36 81
lycee.cantonal@jura.ch



société jurassienne d'émulation

Avec le soutien :

- du Cercle de mathématiques et de physique
- du Cercle littéraire
- de la Section de Porrentruy

Mardi 2 novembre 2010



(Jean-Léon Gérôme *Une collaboration Corneille et Molière* –1874)

Qui a écrit Dom Juan ?

Molière est-il l'auteur des pièces parues sous son nom ?

Dominique Labbé

(Institut d'Etudes Politiques de Grenoble)

Résumé :

Cette conférence expose les raisons pour lesquelles Molière n'a pas composé les pièces présentées sous son nom. Après avoir présenté le dossier historique et la pratique consistant à présenter les comédies satiriques des grands auteurs sous le nom de « comédiens poètes », cette conférence expose les méthodes statistiques qui attribuent à Corneille les comédies les plus connues de Molière.

Abstract :

Some Molière's contemporaneous testified that he was not the author of the plays he presented and that Corneille wrote some of them. These testimonies are confirmed by several statistical indices: intertextual distances, classifications, combinations of the most usual words, meanings of keywords, lengths of the sentences. Corneille and Molière conformed to the standards: during the second half of the French 17th century, the comedies were presented by some comedians – like Molière – and not by the great authors who wrote them.

Pour remercier la Société Jurassienne de son invitation, j'ai rédigé le texte de cette conférence afin qu'elle puisse être consultée par ceux qui n'ont pu venir ce soir. Comme l'invitation venait à la fois du Cercle littéraire et du Cercle de mathématiques et de physique, cette conférence est partagée en deux parties, la première porte sur le dossier historique et littéraire, la seconde sur la recherche statistique, tout en veillant à ce que l'ensemble soit compréhensible par les uns et les autres.

Ce soir nous allons donc nous pencher sur une légende des Temps modernes.

Et à propos de légende, permettez-moi d'évoquer le film de John Ford, « L'homme qui a tué Liberty Valance » (1962). Un journaliste d'une ville de l'Ouest américain rencontre un vieux sénateur qui, dans sa jeunesse, s'est acquis la reconnaissance de ses concitoyens en débarrassant cette ville d'un dangereux bandit. Le vieux sénateur accepte de répondre aux questions et leur dévoile une histoire bien différente de la légende...

Je vous propose de mener l'enquête sur une légende du même genre, celle de Molière (1622-1673), de son vrai nom Jean-Baptiste Poquelin.

Le portrait qu'on en donne est à peu près celui-ci. Molière était le fils d'un bourgeois parisien, « tapissier » - c'est-à-dire marchand de meubles - et valet de chambre du roi. Molière aurait pu prendre la succession de son père mais, succombant aux charmes d'une actrice célèbre, il devint acteur, puis metteur en scène, auteur et directeur de la troupe du roi. Il a donné une trentaine de comédies (voir annexe II) dont certaines – comme *Dom Juan*, *Tartuffe*, *l'Avare* ou le *Bourgeois gentilhomme* - sont aujourd'hui encore les plus jouées (et les plus vendues). L'homme aurait été un peu bohème, et se serait moqué des travers des grands de ce monde (clergé et noblesse) ce qui lui aurait valu la censure de plusieurs de ses pièces, spécialement le *Tartuffe* et *Dom Juan*. Bref, Molière serait le premier « intellectuel » et la « première vedette » des Temps modernes, selon des formules de Roger Duchêne – dont la biographie de 1998¹ a été couronnée par l'Académie française.

Tout cela est-il exact ?

La réponse réside d'abord dans un examen dépassionné des documents de l'époque. Notre ouvrage *Si deux et deux sont quatre*, paru en 2009 chez Max Milo à Paris, présente en détail ce dossier historique que nous allons résumer dans la première partie de cette conférence. La réponse est surtout apportée par les méthodes statistiques d'attribution d'auteur et par les conclusions que l'on en tire quand elles sont appliquées à Molière. La seconde partie de cette conférence leur sera consacrée.

I. DOCUMENTS ET TEMOIGNAGES D'EPOQUE

Concernant Molière, il existe quelques documents de première main :

- une centaine d'actes officiels où figure sa signature. Ce sont les seuls « manuscrits » qui restent de Molière (Jurgens et Maxfield-Miller, 1963) ;
- un registre où sont consignées les séances de sa troupe depuis Pâques 1659, avec les recettes et quelques éléments sur la vie de la troupe (Young, 1977).

Enfin, une série de textes imprimés du vivant de Molière et la correspondance de ses contemporains.

Ces documents dressent le portrait de Molière.

¹ Bibliographie à la fin de ce texte.

1 . Riche comédien, financier et faux noble

Molière s'est présenté une centaine de fois devant un officier ministériel, un prêtre ou un fonctionnaire. La plupart du temps, il a déclaré être comédien d'un grand puis de Monsieur (frère du roi), enfin du roi lui-même. Mais il aussi déclaré être :

- « valet de chambre du roi » (28 fois),
- « sieur » (de) Molière (19 fois),
- « écuyer » (4 fois).

C'était autant d'usurpations.

- « valet de chambre du roi » : le père de Molière était effectivement « valet de chambre du roi » et Molière aurait pu lui succéder. En 1643, à 20 ans, il y a renoncé. La charge aurait été reprise par son frère, mort en 1660, puis par son père jusqu'à son décès en 1669 (Jurgens et Maxfield-Miller, 1963, p. 166-169). Cette charge immobilisait son titulaire, 3 mois par an, au service du roi (Da Vinha 2004). Pour la période postérieure à 1669, le registre de la troupe indique que Molière n'a pu l'exercer puisque la troupe jouait tout le temps. Pourtant, Molière s'est présenté toute sa vie comme « valet de chambre du roi »...

- « sieur de... » : seuls les nobles, propriétaires d'une seigneurie pouvaient utiliser le titre « sieur de... ». Molière n'était pas noble et n'était propriétaire d'aucune seigneurie.

- « écuyer » : c'était le titre de noblesse le plus bas avant chevalier. Les valets du roi pouvaient utiliser ce titre, du moins tant qu'ils exerçaient effectivement leur charge, ce que n'a pas fait Molière...

Ces usurpations de fonctions et de titres étaient sévèrement punies. Par exemple, La Fontaine, le célèbre fabuliste a été condamné en 1662 à une lourde amende pour s'être attribué, dans plusieurs actes notariés, le titre d'écuyer, exactement comme le faisait Molière. Cette sévérité s'explique aisément : sous l'ancien régime, les nobles ne payaient pas l'impôt. Il fallait donc identifier les nobles du reste de la population afin de percevoir le maximum d'impôts...

En tout cas, ces documents prouvent que Molière éprouvait de sérieuses démanagements de noblesse... exactement comme le *Bourgeois gentilhomme* !

Molière était un riche financier pas un « intellectuel ».

On a retrouvé plusieurs contrats de prêts qui attestent qu'il était financier et qu'il disposait de fonds importants. Il a même utilisé un prête-nom pour mettre la main sur la maison familiale peu avant la mort de son père.

L'inventaire de ses biens à son décès (Jurgens & Maxfield-Miller) montre qu'il avait un grand appartement à Paris et une maison de campagne à Auteuil, tous deux richement meublés – par exemple, le commissaire priseur a estimé le lit de Molière à une somme qui équivaut à 4 ans du salaire d'un ouvrier qualifié de l'époque. Sa garde robe était digne de celle du *Bourgeois gentilhomme*. En revanche, il avait très peu de livres (226) en majorité dans la garde-robe et dans la maison de campagne. Dans ces deux habitations, aucune pièce qui puisse servir de bureau et aucun meuble spécifiquement destiné à l'écriture... Enfin, Molière ne laisse absolument aucun manuscrit. Pas la moindre lettre, pas le moindre billet à un fournisseur. Il faut se rappeler qu'à l'époque, on correspondait beaucoup et que Mme de Sévigné n'était pas un cas particulier. Molière serait un cas unique dans l'histoire littéraire d'un écrivain qui n'entretenait aucune correspondance.

Molière n'était pas un metteur en scène ni un chef de troupe au sens moderne.

La troupe du Palais royal, comme les autres, ne fonctionnait pas comme les troupes modernes. Le choix des pièces était fait par l'assemblée générale des comédiens qui répartissait les rôles, discutait des décors et de la mise en scène. Tout juste était-il admis que l'auteur de la pièce puisse intervenir dans la répartition des rôles et dans certains aspects de la mise en scène. Enfin, chaque soir, après la représentation, les comédiens se partageaient la recette en parts égales.

Molière était si peu « administrateur » ou « chef » que ses camarades ne lui ont jamais confié la tenue du livre de comptes ni celle de la caisse... En plus d'être acteur et premier apporteur de pièces, la seule fonction de Molière était celle d'introducteur : au début du spectacle, il présentait la pièce de manière plaisante, pour obtenir l'attention du public et, après le spectacle, il annonçait le programme des prochaines séances.

Molière n'avait pas le temps de composer l'œuvre qu'on lui prête.

Le registre de La Grange montre que, entre 1659 et 1673, la troupe du Palais Royal a joué trois jours par semaine au minimum, avec une seule interruption pour les fêtes de Pâques. C'est un rythme épuisant qu'aucune troupe moderne ne tiendrait. D'autant plus que la majorité des séances comportaient deux pièces – une pièce longue et une petite comédie. Pour la période allant de 1659 à 1673 – période pendant laquelle Molière est censé avoir composé son œuvre et être valet de chambre du roi ! – cela représente plus d'une centaine de rôles à apprendre. Avec les répétitions, les assemblées de la troupe et les nombreux déplacements – sans compter ses apparitions à la cour et son activité de financier -, Molière n'avait pas le temps de composer les deux pièces (au minimum) présentées sous son nom, soit plus d'un millier de vers par an en moyenne, vers qui sont, pour la plupart, techniquement parfaits...

2. Parmi ses contemporains, personne n'a traité Molière comme un auteur

Pas un seul de ceux qui l'ont approché ne l'a traité en grand auteur.

Faisons d'abord appel aux témoins dont les noms vous sont familiers et dont les œuvres sont facilement accessibles, ce qui permettra de vérifier.

La Fontaine, Madame de Sévigné, Boileau

La Fontaine, connu surtout pour ses fables, s'intéressait beaucoup au théâtre. Il a écrit au moins deux pièces et, peut-être, plusieurs autres, comme on le verra plus loin. Il était l'ami de nombreux comédiens. En 1661, après avoir assisté à la première des *Fâcheux* et sans connaître Molière, il écrit une longue lettre enthousiaste à son ami Maucroix, lettre restée inédite jusqu'en 1729 (La Fontaine, 1968, p. 522-527). Puis il rencontre Molière – probablement en 1662 ou 1663. Il n'y a aucune trace d'une autre rencontre. La Fontaine ne mentionnera plus Molière dans sa correspondance, ni dans ses publications.

Depuis trois siècles, il y a là une énigme que les admirateurs de Molière ne peuvent expliquer. D'autant plus que, fréquentant lui-même des troupes et des acteurs, La Fontaine savait ce qui se passait dans les coulisses.

Les lettres de la Marquise de Sévigné sont un témoignage unique sur la vie mondaine et intellectuelle du temps. Après celles de Corneille, les pièces de Molière sont les plus citées dans cette correspondance. La marquise invitait à sa table toutes les personnalités marquantes. Et pourtant, elle semble n'avoir jamais rencontré Molière et, en tout cas, elle ne le mentionne pas dans ses lettres. Si celui-ci était l'auteur des pièces qu'elle aimait tant, cet ostracisme est inexplicable. En revanche, si l'acteur Molière n'était que financier et faux noble, cette attitude est fort logique...

Boileau est le témoin clef. Contrairement à ce qui est partout imprimé :

- Boileau n'a jamais considéré Molière comme un grand écrivain. Par exemple, l'œuvre majeure de Boileau, *l'Art poétique* (1674) se clôt, au chant IV, par un éloge des bons écrivains – P. Corneille, Racine, Benserade, Segrais – Molière n'y figure pas ;

- dans toute l'œuvre de Boileau, dans sa correspondance et dans les témoignages de ses familiers, il n'y a aucun indice, même indirect, d'une relation entre Boileau et Molière et, *a fortiori*, d'une amitié entre les deux hommes.

Depuis 3 siècles il y a là une énigme que les admirateurs de Molière sont incapables d'expliquer.... D'ailleurs, les quelques vers de Boileau, toujours cités en faveur de Molière, sont à double sens. Considérons par exemple, les premiers d'entre eux.

Stances à M. Molière Sur la comédie de l'Ecole des Femmes (Boileau 1663, p 246)

Que tu ris agréablement !
 Que tu badines savamment !
Celui qui sut vaincre Numance,
Qui mit Carthage sous sa loi,
Jadis sous le nom de Térence
 Sut-il mieux badiner que toi ?
 (...)

Boileau aurait donc vu dans Molière un nouveau Térence ? Térence (env. 190-159 av. JC) était un esclave romain affranchi, dont on ne sait rien sinon que 6 comédies ont été présentées sous son nom et ont connu un très grand succès. Beaucoup, comme Boileau, pensent que ces comédies ont été écrites par *Scipion Emilien* (185-129 av. JC), général romain vainqueur de Carthage et de Numance. Le nom de ce général était *P. Cornelius*. Le nom (et le prénom) français *Corneille* viennent du nom de cette famille romaine, comme *Jules* vient de la famille *Julia* dont était originaire César. Dès lors on peut lire ces vers de la manière suivante :

P. Cornelius se cache derrière Molière, comme il se cachait derrière Terence...

S'il n'y avait que cette allusion, on ne pourrait rien en tirer mais, dans le même ordre d'idée :

- à la fin de sa satire de 1665, Boileau dit clairement que Molière ne sait pas faire des vers (p 17) ;

- en 1674, Boileau suggère que Molière n'est pas l'auteur du *Misanthrope* (p 178) ;

- et dans le même passage, Boileau écrit que Molière est un « faux plaisant, à grossière équivoque, qui pour me divertir, n'a que la saleté » (p 178-179)..

Quant à Racine, il a très vite rompu avec Molière.

Comment expliquer cet ostracisme général alors même que le public faisait un grand succès aux comédies que présentait Molière ?

Trois critiques littéraires et trois éditeurs

Du vivant de Molière, de nombreuses rumeurs ont couru sur son compte. Prenons l'exemple de trois critiques littéraires qui étaient bien au courant des coulisses du théâtre.

En 1660, au début des triomphes de Molière, Baudeau de Somaize écrit que Molière est l'« auteur *prétendu* des *Précieuses ridicules* ».

En 1663, Donneau de Visé – co-fondateur avec Thomas Corneille (le frère cadet de Pierre) du *Mercure galant*, la première revue de critique littéraire - écrit, à propos de *l'Ecole des femmes*, que tout le monde sait que les « enfants de Molière ont plus d'un père ».

La même année, Robinet, l'autre critique littéraire du temps écrit, à propos de cette même pièce, que « Molière n'est pas une source vive mais un bassin qui reçoit ses eaux d'ailleurs », et qu'il est comme ces « ânes seulement capables de porter de grands fardeaux »...

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

Deux éditeurs indiquent, dans des notes placées au début des premières éditions, 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 *Dom Juan* – adapté et mis en vers par T. Corneille - affirme que Molière a « fait jouer » cette pièce de son vivant et qu'elle est toujours « représentée sous son nom ».

Molière ne serait donc pas l'auteur des *Précieuses ridicules*, du *Dépit amoureux*, de *l'Ecole des femmes*, du *Bourgeois gentilhomme* et de *Dom Juan*... ?

Une clef d'explication réside dans la manière dont fonctionnait le théâtre à cette époque².

3. les « comédiens poètes »

Pour la seconde moitié du XIXe siècle, on connaît à peu près toutes les pièces de théâtre et il existe un certain nombre d'archives incontestables, au premier rang desquels les livres de comptes de la troupe de Molière après sa mort, puis de la Comédie française à partir de 1680. Ces documents révèlent que la pratique du comédien poète prête-nom d'un grand auteur était un véritable système.

Un véritable système

L'examen de ces documents aboutit à quelques constats :

- 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" et que l'on appelait des « comédiens poètes » ;

- bien qu'ils connaissaient la vérité, les troupes, les critiques littéraires et les éditeurs étaient d'accord pour dissimuler le nom de l'auteur et présenter ces ouvrages sous le nom du comédien poète.

² L'un des deux chapitres de notre ouvrage *Si deux et deux sont quatre*, consacrés à ce fonctionnement est consultable en ligne sur le site corneille-moliere.org.

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... T. Corneille. 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.

Voici les principaux "comédiens poètes", confrères de Molière : Baron, Brécourt, Champmeslé, Dancourt, Desfontaines, Dorimond, Hauteroche, La Thuilerie, Montfleury, Poisson, Raisin, Rosimond, Villiers... Desfontaines 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).

On sait que derrière Poisson, il y avait un écrivain, Boursault, ami des Corneille ; que La Fontaine fournissait des comédies à Champmeslé ; que Le Clerc – un autre académicien - a utilisé au moins une fois Beauregard... Mais la plupart des associations – entre les écrivains et les comédiens poètes - gardent aujourd'hui leur secret.

Ce système s'explique par la situation particulière du théâtre français au XVII^e siècle :

- les livres de comptes montrent que les comédies satiriques, comme celles de Molière, plaisaient au public parisien et qu'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.

Les mobiles des Corneille

Nous connaissons donc les mobiles qui poussaient les troupes à utiliser ce système. On se doute bien des raisons pour lesquelles les écrivains désargentés fournissaient ces comédies inavouables.

Personne ne conteste que Thomas Corneille avait de gros besoins d'argent et qu'il a fait beaucoup de travaux alimentaires (Reynier 1892). Mais tout le monde semble oublier que Pierre et Thomas avaient épousé deux sœurs ; qu'ils ont toujours habité ensemble et qu'ils faisaient caisse commune. Les deux frères ont eu 10 enfants qu'ils ont établis avec de fortes sommes... d'origine inconnue. En effet, notre livre présente les principaux éléments connus concernant les ressources et les dépenses des frères Corneille. Le budget de la famille présente une impasse énorme. Il est impossible d'expliquer comment les deux frères ont fait face aux dépenses considérables qu'ils ont effectuées. Sauf à admettre qu'ils ont joué les « plumes mercenaires », ce dont Boileau les accuse justement dans son *Art poétique* (1674).

Mais je ne puis souffrir ces auteurs renommés,
 Qui dégoûtés de gloire et d'argent affamés,
 Mettent leur Apollon aux gages d'un libraire,
 Et font d'un art divin un métier mercenaire.

Tout le monde est d'accord : ces quatre vers visent les frères Corneille. Ils doivent être rapprochés des allusions de Boileau concernant Molière.

On peut aussi signaler que les Corneille pouvaient avoir d'autres mobiles pour écrire ces comédies satiriques. Par exemple, dans son essai, H. Wouters signale que P. Corneille aurait été humilié par les « précieuses » et que cela aurait été une motivation pour sa charge contre les *Femmes savantes* ou les *Précieuses ridicules*. Ce sont des pistes qu'il ne faut pas négliger.

Les conclusions de cette enquête historique peuvent se résumer facilement : Molière était un riche comédien et financier qui ne s'est jamais comporté en écrivain et qu'aucun de ses contemporains n'a traité comme tel. Il était un « comédien poète », comme une bonne douzaine de ses confrères qui servaient d'intermédiaires entre leurs troupes et les écrivains. Plusieurs témoignages désignent P. Corneille comme étant la plume de l'ombre de Molière. Comment en être certain ? De quelles comédies P. Corneille serait-il l'auteur ?

II. L'ATTRIBUTION D'AUTEUR PAR LES METHODES SCIENTIFIQUES.

De quels outils dispose-t-on pour l'attribution à un auteur connu de textes inconnus ou d'origine douteuse ? A quels résultats ces méthodes aboutissent-elles quand elles sont appliquées aux œuvres de Corneille et Molière ?

1. La distance intertextuelle

La question posée a une portée qui dépasse largement l'attribution d'auteur. Comment partitionner les vastes populations hétérogènes – ici des ensembles de textes – et y retrouver des sous-ensembles homogènes. L'extraordinaire expansion du web met aujourd'hui cette question au premier plan. Nous avons commencé à étudier cette question en 1997, avec Pierre Hubert, Cyril Labbé et Denis Monière et nous avons mis au point une mesure et une méthode à laquelle la seconde partie de cette conférence sera consacrée.

La mesure

Nous avons cherché une mesure de « distance » entre les textes qui présente les propriétés d'une distance « euclidienne ». On désigne ainsi la longueur du segment de droite joignant deux points, dans la géométrie qui repose sur le postulat bien connu (par un point il ne passe qu'une parallèle à une droite située hors de ce point). Les propriétés d'une distance euclidienne sont :

- l'identité (la distance d'un point à lui-même est nulle),
- la symétrie : le résultat est le même que l'on mesure AB ou BA (propriété très intéressante car elle permet de se passer des vecteurs),
- l'inégalité triangulaire (le chemin direct entre A et B est toujours plus court qu'en passant par un point C non situé sur le segment AB).

La mesure de la distance entre deux textes (baptisée pour cela « intertextuelle ») sera le nombre de mots différents. Pour pouvoir comparer des textes de longueur inégales, on transforme la distance absolue en une distance relative à la manière d'un pourcentage (présentation des calculs dans Labbé & Labbé 2003).

Les tests ont montré que la distance intertextuelle est influencée par quatre facteurs, soit par ordre décroissant d'importance : le genre, l'époque, l'auteur et le thème.

Le « genre » est la variable dominante. Exemples de « genres » : l'écrit et l'oral, le roman, le théâtre, la poésie, la correspondance... Ce point est important : pour déterminer l'auteur d'une pièce de théâtre, il faut comparer cette pièce à d'autres pièces de théâtre. Ou encore : si P. Corneille n'avait pas publié de pièces de théâtre sous son nom, cette méthode ne pourrait pas être utilisée pour retrouver celles qu'il aurait écrites pour Molière...

La méthode est maintenant très largement employée, également sur d'autres langues comme l'anglais (Arnold 2004) ou l'italien (Tuzzi 2010).

La mise au point de la méthode a été effectuée grâce à un grand nombre d'expériences semblables à celles utilisées dans l'industrie ou dans la recherche biomédicale. Plusieurs de ces expériences ont été publiées. Citons notamment : 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.

Voici l'une de ces expériences inédite.

Une expérience en aveugle

L'expérience s'est déroulée "en aveugle", c'est-à-dire que l'opérateur ne connaît pas les textes qu'il traite. Cependant, le choix n'a pas été arbitraire et a suivi des conventions précises :

- puisqu'il s'agit de mesurer l'importance de la variable auteur, il faut qu'il y ait plusieurs auteurs différents et que les autres facteurs soient neutralisés : tous les textes doivent appartenir au même genre et avoir été écrits à peu près à la même époque. Cela permettra également de s'assurer que l'auteur l'emporte sur le « facteur thème » ;

- chaque auteur a au moins deux textes pas forcément issus du même ouvrage. En fait, il y avait au moins deux extraits par ouvrage mais les effectifs impairs pesaient aussi lourds que les pairs, afin de pénaliser les algorithmes de classification qui fonctionnent en regroupant par couples les individus, ou les groupes d'individus...

Les 56 textes ont été anonymés, c'est-à-dire que les auteurs et les titres ont été remplacés par un numéro. En annexe IV, la liste des œuvres et des écrivains qui ont été révélés à la fin de l'épreuve. En fait, il s'agissait de 12 romanciers français du XIXe siècle.

L'algorithme lit successivement chacun des textes et mesure sa distance aux 55 autres, pris un par un. Cette opération engendre donc :

$$\frac{1}{2} 56*55 = 1540 \text{ distances différentes}$$

Le second terme de la multiplication – 55 et non pas 56 – est justifié par la propriété d'identité de la distance euclidienne ($d_{(aa)} = 0$) ;

La division par deux est justifiée par la propriété de symétrie de la distance euclidienne ($d_{(ab)} = d_{(ba)}$) ;

On calcule la distance moyenne (que l'on note \bar{x}) c'est-à-dire la somme des distances divisée par 1540 :

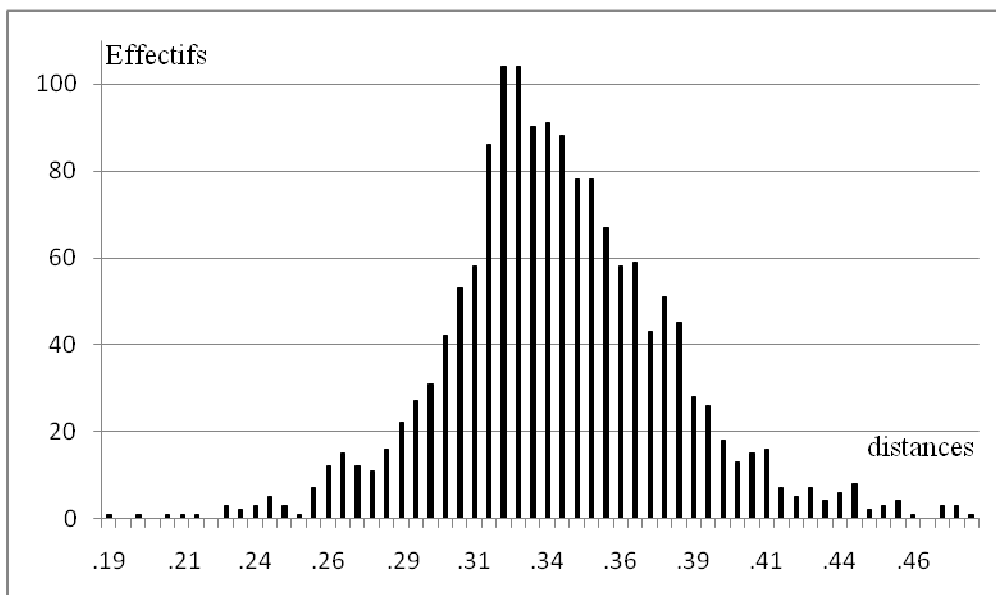
$$\bar{x} = .3358$$

On peut également lire cette valeur ainsi : "en moyenne dans les 56 textes comparés deux à deux, on observe 3 358 mots différents pour 10000 mots" ou plus simplement qu'en moyenne environ un tiers du vocabulaire est différent.

On associe à cette moyenne, une "déviatoin standard" ou écart type qui est la racine carrée de la variance (moyenne quadratique des écarts à la moyenne arithmétique). Elle est symbolisée par la lettre grecque sigma (σ). Comme son nom l'indique, elle donne une mesure de la dispersion des valeurs observées autour de la moyenne. Ici l'écart type est égal à 0,0388 (ou encore 388 mots différents pour 10 000).

Les distances sont rangées par ordre croissant et regroupées par classes d'intervalle égal – dans la figure ci-dessous l'intervalle est de .010.

Figure 1. Histogramme des distances classées par ordre croissant

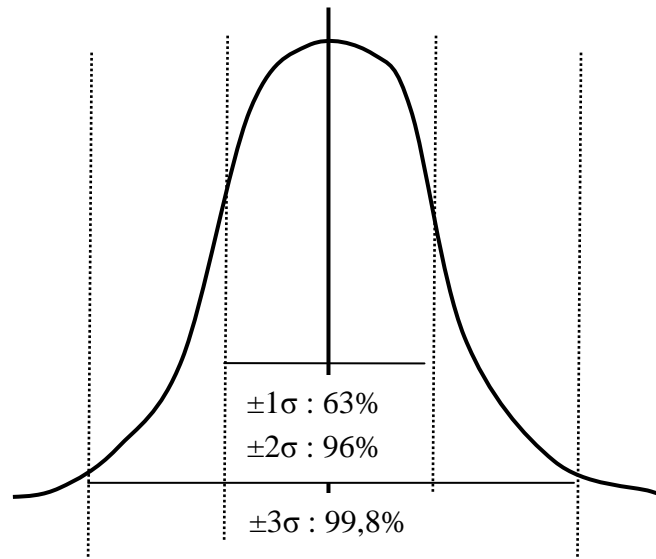


Ce profil est celui d'une distribution "gaussienne" - du nom du mathématicien F. Gauss (1777-1855). On l'appelle aussi "courbe en cloche", à cause de son profil caractéristique. Une distribution semblable est habituellement le résultat d' "épreuves aléatoires" dont le modèle est le tirage au hasard de boules de différentes couleurs dans une urne. Quand on rencontre une distribution de ce genre, on en déduit que la distribution du caractère étudié chez les individus composant cette population (ici les distances entre textes) est régie par le hasard.

La valeur la plus probable – dite pour cela "espérance mathématique" - coïncide avec la valeur la plus fréquemment observée (le bâton le plus élevé de l'histogramme) et avec la moyenne des observations (\bar{x}).

Une population de ce genre présente plusieurs propriétés intéressantes, résumées dans le schéma ci-dessous.

Figure 2 Les valeurs caractéristiques d'une distribution gaussienne



Les valeurs indiquées sur le schéma fournissent des repères commodes :

- environ les deux tiers des observations sont groupées autour de la moyenne dans un intervalle égal à plus ou moins un écart type ;
- moins de 5% de ces observations se situent en dehors de l'intervalle de deux écarts-types autour de la moyenne (2.5 en dessous et 2.5 au dessus) ;
- moins de 1% de ces observations se situent en dehors de l'intervalle de trois écarts-types (0.5 de chaque côté de l'intervalle).

Pour l'expérience, ces intervalles sont présentés dans le tableau ci-dessous. Les valeurs observées correspondent aux valeurs prédites par le schéma théorique ci-dessus. La distribution est bien gaussienne.

Tableau 1 Les valeurs attendues et observées dans l'expérience

Bornes	$D_{(ab)}$	Pourcentage attendu	Pourcentage observé
$\bar{x} - 3\sigma$.220	< 0.5	0.3
$\bar{x} - 2\sigma$.260	< 2.5	2.6
$\bar{x} + 2\sigma$.413	> 97.5	97.0
$\bar{x} + 3\sigma$.452	> 99.5	99.4

Considérons les individus "anormaux", c'est-à-dire les distances situées au-delà des bornes de l'intervalle de fluctuation standard autour de la moyenne. Notre théorie nous suggère que :

- les distances les plus courtes (inférieures à .260) s'observent entre textes d'un même auteur, lorsqu'ils portent sur des thèmes peu éloignés (annexe V). Par exemple, la distance la plus faible (0.1777) sépare les textes n° 15 et 26 ; la seconde (0.191) les textes 03 et 26. On en tire donc que ces trois textes sont d'un même auteur et extraits d'un même ouvrage). On répète l'opération pour les 39 autres distances inférieures à la moyenne moins deux écarts-types que l'on examine une par une en constituant des groupes comme celui donné en exemple ci-dessus.

- les textes "anormalement éloignés" (distances supérieures à .413) sont d'auteurs différents et portent sur des thèmes également éloignés (annexe VI). Par exemple, la deuxième plus longue distance (0.4737) sépare les n° 15 et 07. On en déduit que l'auteur du texte n° 07 n'est pas celui du groupe {03-15-26} que l'étape précédente a révélé.

Il y a 47 distances supérieures à la moyenne augmentée de deux écarts-types. On les examine une par une pour disjoindre les groupes constitués lors de la première étape.

Ces deux étapes permettent de classer 51 des 56 textes en 15 groupes différents (tableau 2).

Tableau 2 Détermination des textes et des auteurs grâce aux distances remarquables

N°	Textes	Auteurs
1	01 13	Balzac (Père Goriot)
2	03 15 26	Dumas fils (Dame aux camélias)
3	02 14 25 33 39	Dumas père (Monte Cristo, Trois Mousquetaires)
4	04 16	Flaubert (Bouvard et Pécuchet)
5	27 34 40 45 50 53	Flaubert (Bovary - Education sentimentale)
6	05 17	Gautier (Avatar)
7	06 18 28	Hugo (Misérables)
8	41 46	Hugo (Notre Dame de Paris)
9	08 20 29 36 42 47 51 54 55 56	Maupassant (Bel Ami, Fort comme la mort, Mont Oriol, Pierre et Jean, Une vie).
10	21 30	Stendhal (Chartreuse de Parme)
11	37 43	Stendhal (Rouge et Noir)
12	11 23 31	Vigny (Servitude de grandeur militaires)
13	12 24	Zola (la Bête humaine)
14	32 38	Zola (Germinal)
15	44 49 52	Zola (l'Argent)

Après l'expérience, les noms des auteurs ont été dévoilés (dernière colonne à droite) : il n'y a aucune erreur. Cependant, sept textes n'ont pas pu être mariés à un autre ni rattachés à un groupe existant : 07 et 19 (Lamartine), 09 et 48 (Stendhal), 10 et 22 (Vernes) et 35 (Hugo). De plus, si toutes les affectations sont exactes, certains auteurs se retrouvent dans plusieurs groupes (toujours pour des livres différents).

Cela ne doit pas être considéré comme un échec. En effet, notre méthode ne prétend pas classer tous les textes. En contrepartie, quand on conclut, c'est avec une certitude raisonnable. De ce point de vue, la réussite est totale...

Corneille et Molière

Appliquée au corpus Corneille-Molière (Labbé & Labbé 2001b), cette même méthode classe correctement la plupart des textes, mais elle apporte deux surprises :

- deux comédies en alexandrins de P. Corneille - *le Menteur* (1642) et *la suite du Menteur* (1643) – sont classées avec 14 comédies présentées par Molière entre 1659 et 1673 : douze comédies en alexandrins et deux comédies en prose (*Dom Juan* et *l'Avare*),

- deux comédies "sérieuses" de Molière, également en alexandrins, (*Dom Garcie et Psyché*) sont classées avec les tragédies et les comédies héroïques contemporaines de P. Corneille.

A ce stade, on peut formuler trois remarques :

Premièrement, la distance intertextuelle a été testée sur des milliers de textes, sans jamais rencontrer de tels croisements entre deux œuvres d'auteurs différents, sauf... quand les deux auteurs supposés ne font qu'un, comme dans le cas de Gary et Ajar (Labbé 2004a et Lafon et Peters 2006), ou quand la même plume de l'ombre a travaillé pour deux employeurs différents (Monière et Labbé, 2006).

Deuxièmement, en répétant les expériences, comme celle-ci, sur un grand nombre de textes de différents genres, nous avons pu étalonner empiriquement une échelle des distances pour le français. Si l'on exprime les valeurs en "mots différents pour 10 000 mots" :

- au-dessous de 2 000 (un mot sur 5) : les deux textes sont écrits dans un même genre, par un seul auteur, à la même époque et sur des thèmes proches,

- entre 2 000 et 2 500 : l'auteur et le genre sont identiques mais les dates de composition et (ou) les thèmes sont plus éloignés,

- entre 2 500 et 3 500, soit l'auteur est le même et alors plusieurs facteurs ont changé, soit ce sont deux auteurs différents, travaillant à la même époque dans un même genre sur des thèmes très proches,

- au-dessus de 3 500, si les deux textes sont dans un même genre, alors les auteurs sont différents. Si l'auteur est le même, alors ils ont été rédigés à des époques éloignées sur des thèmes très différents, plus probablement les genres sont différents.

- au dessus de 4 500, auteur et genre sont différents.

Nota : cette échelle est applicable à des textes dont les longueurs sont comprises entre 5000 et 30 000 mots, les valeurs sont des bornes sur un continuum, pas des seuils.

Appliquée au corpus Corneille-Molière, cette échelle confirme la proximité remarquable entre les deux menteurs de P. Corneille et les principales pièces présentées par Molière (annexe 3).

Troisième remarque : Racine a écrit une comédie en alexandrins (*les Plaideurs*, 1668) exactement contemporaine des comédies en alexandrins présentées par Molière (entre 1658 et 1672). Le facteur temps étant neutralisé, les distances séparant ces pièces devraient être plus petites qu'avec les deux *Menteurs* de Corneille qui sont antérieures de 17 à 30 ans par rapport aux pièces présentées par Molière. Or c'est l'inverse qui se produit (annexe 3).

Le XVIIe siècle fournit d'autres contre-épreuves intéressantes. Par exemple, Corneille et Racine ont écrit chacun une tragédie sur l'amour impossible entre un empereur romain (Titus) et une reine orientale (Bérénice). Ils l'ont fait au même moment, en "aveugle", en alexandrins et en respectant les fameuses "règles" qui enserraient la création théâtrale, spécialement la tragédie. Et tous deux avaient en tête la liaison entre Louis XIV et Henriette d'Angleterre, sa belle sœur. Le lieu de l'action et les personnages étaient donc identiques ainsi que les obstacles à cet amour partagé (d'où un vocabulaire commun important...). La distance entre ces deux pièces (0,256 ou 2 560 mots différents pour 10 000) est plus élevée que toutes celles constatées entre les pièces en vers de Molière et les deux *Menteurs* de Corneille ou entre *Dom Garcie* et les comédies héroïques de Corneille, alors que toutes ces pièces sont séparées par un laps de temps important, que les thèmes sont toujours différents et que le genre "comédie" en vers est nettement moins contraignant que celui de la grande tragédie en alexandrins.

Le XVII^e fournit d'autres exemples. Les *Sophonisbe* de Mairet et de Corneille se distinguent bien alors qu'ils ont travaillé dans le même genre, sur le même thème, les mêmes événements, les mêmes personnages, en suivant la même trame narrative. Quand Quinault, Corneille et Molière travaillent ensemble sur *Psyché*, Quinault se distingue des deux autres... qui sont impossibles à départager.

Conclusion de toutes ces expériences : les distances entre des textes appartenant à un même genre et écrits *par un même auteur* à une même époque, *sur des thèmes différents*, sont systématiquement et nettement inférieures à celles séparant deux auteurs différents, *même quand ils écrivent au même moment, sur un thème identique*.

Enfin, il est évident que les grands tableaux de chiffres sont lourds à manipuler manuellement (comme on vient de le faire) et que l'on peut passer à côté d'informations importantes. Des automates peuvent faire ce travail plus efficacement et offrir des représentations graphiques plus simples à consulter que les tableaux de chiffres.

2. Classification

L'objectif est double :

- rechercher les meilleurs groupements possibles sans intervention humaine. Deux critères sont utilisés : d'une part, les distances entre les individus composant un même groupe doivent être les plus courtes possibles ; d'autre part, les distances séparant les différents groupes ainsi constitués, doivent être les plus grandes ;

- offrir la meilleure représentation possible, en deux dimensions, de ce meilleur ordre possible alors que celui-ci comporte un grand nombre de dimensions.

Parmi les méthodes de classification automatique, la plus performante, dite "arborée", est très utilisée en biologie et en génétique. Cette méthode repose sur la propriété suivante : si toutes les distances séparant les individus étudiés sont euclidiennes, il existe un "arbre" qui représente exactement les positions respectives de ces individus les uns par rapport aux autres et les meilleurs groupements possibles (Luong1988 ; Ruhlmann 2003).

Appliquée à l'expérience en aveugle sur les romanciers du XIX^e, l'algorithme donne la figure 3 ci-dessous.

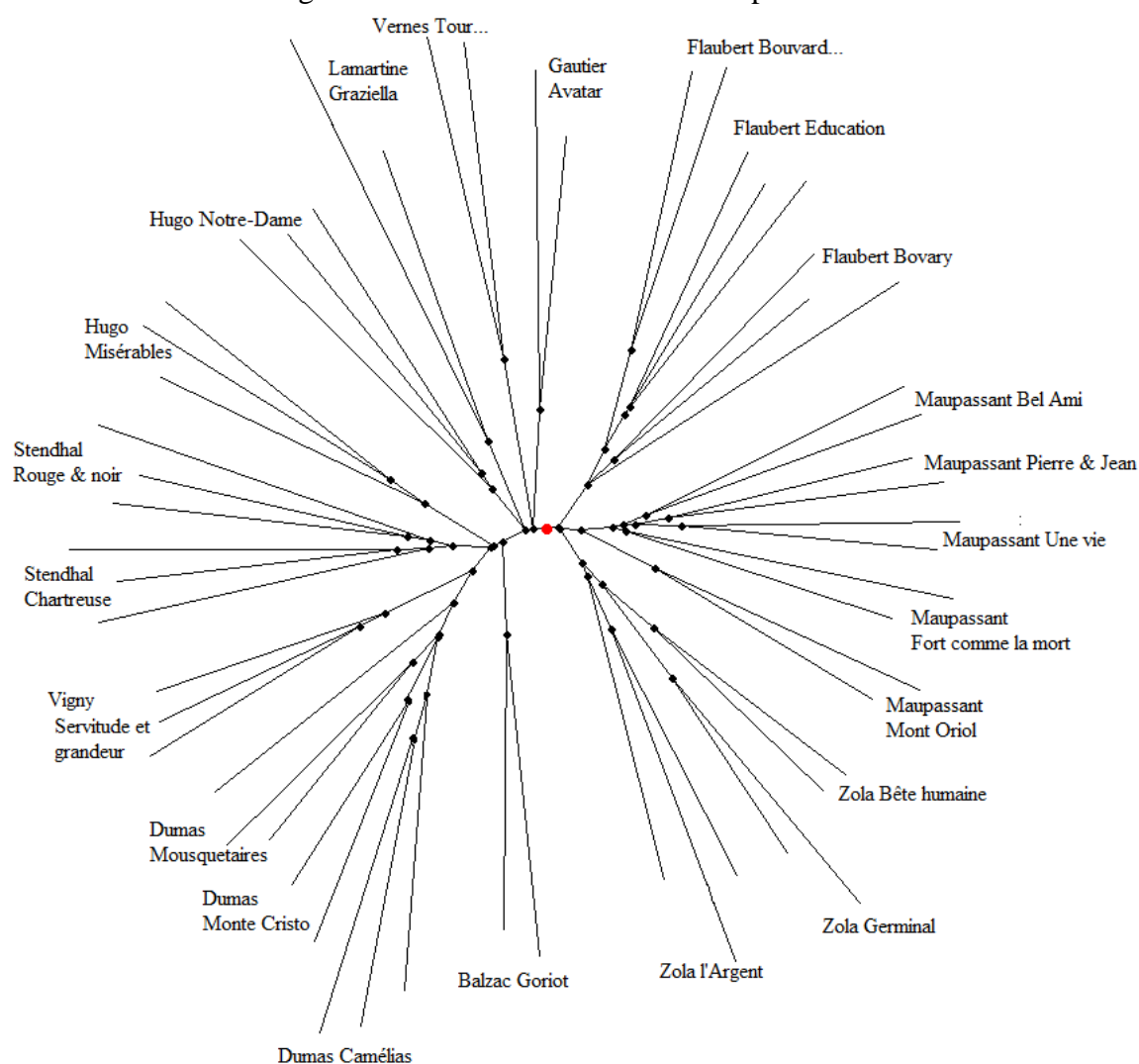
Dans cette figure, les **feuilles** terminales figurent chacune un texte ; les **nœuds** intermédiaires figurent les meilleurs groupements possibles, c'est-à-dire ceux pour lesquels les distances entre les éléments qui composent le groupe sont les plus faibles possibles et les distances les séparant des autres les plus grandes possibles. Les segments de droite, ou **arêtes**, sont des **branches** quand elles relient des feuilles à des nœuds et des **troncs** quand elles relient des nœuds entre eux. La distance entre deux points quelconques est figurée par le **chemin** unissant ces points et la longueur de ce chemin est proportionnelle à la distance originelle correspondante (Labbé et Labbé 2006 et 2008).

Tous les textes sont correctement attribués à leurs auteurs et aux œuvres respectives. Combien de combinaisons différentes peut-on réaliser avec 56 objets. En théorie, 56 ! (factorielle de 56), c'est-à-dire : $56 * 55 * 54... * 3 * 2 = 7,11 \text{ e}74$. Même en tenant compte des propriétés de symétrie et d'identité déjà mentionnées, on obtient encore un nombre de combinaisons comportant une trentaine de zéros... Il n'y a donc aucune chance pour que ce genre d'épreuve puisse être réussie par hasard.

Le graphique suggère bien d'autres conclusions. La plus intéressante concerne le partage du graphique en deux parties : une première période regroupe Balzac, Dumas,

Hugo, Lamartine, Stendhal, Vigny et la seconde : Flaubert, Maupassant, Zola. Les écoles littéraires et les deux parties du siècle sont donc clairement découpées.

Figure 3 Classification arborée du corpus



NB : les noms des auteurs et des œuvres ont été ajoutés après l'expérience.

La remarque la plus amusante est la proximité du fils Dumas (*la Dame aux camélias*) avec les œuvres que son père n'écrivait pas lui-même ! Et secondairement avec le roman autobiographique de Vigny (*Servitude et grandeur militaires*) qui, au fond, n'est pas très éloigné du sous-genre « cape et épée ».

Cet outil est évidemment plus puissant que l'examen manuel des distances. Il soulève deux questions. Quelle est la précision de la méthode ? Quel risque d'erreur ?

Dans l'arbre ci-dessus, les longueurs des branches sont égales aux distances correspondantes avec une incertitude moyenne de $\pm 3.9\%$ (voir Labbé et Labbé 2008). La représentation graphique est donc fiable. L'algorithme a cependant rencontré quelques légers problèmes notamment avec les textes des Dumas père et fils et de Vigny parce qu'ils sont d'un genre un peu différent des autres : dans l'annexe VI, la majorité des distances les plus longues séparent d'ailleurs les deux Dumas des auteurs contemporains...

Appliqué aux pièces de Corneille et de Molière, la classification arborée confirme les quatre anomalies signalée ci-dessus : les deux *Menteurs* de P. Corneille sont classés au milieu des pièces en vers présentées par Molière ; *Psyché* et *Dom Garcie* sont rattachées au reste de l'œuvre de P. Corneille (annexe VII).

3. Trois autres indices

Trois autres indices statistiques conduisent aux mêmes conclusions.

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 VIII).

Racine ne partage avec Corneille et Molière que trois combinaisons (soulignées dans le tableau en annexe) : "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 trouver un autre exemple concernant deux auteurs réellement différents...

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.

Longueur de phrase

Il y a mille manières de construire une phrase. Dans la conversation, on utilise généralement des phrases brèves et plus simplement construites que dans un récit ou un essai théorique. Dans un genre donné, certains auteurs font preuve d'une grande stabilité et leurs phrases offrent donc un outil pour l'attribution d'auteur (Labbé & Labbé, 2010). Par exemple, les tragédies de Corneille et de Racine. Même quand ils traitent, à la même époque, de sujets proches dans un même genre, les phrases de Racine sont significativement plus courtes que celles de Corneille. Cela signifie que les pièces de Racine contiennent plus d'interpellations et de conversations courantes que celles de

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

Corneille et moins de récits, de soliloques et de monologues. Ces choix stylistiques sont répétés de pièce en pièce.

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, la plupart des auteurs se singularisent par des longueurs de phrases différentes. *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.

Deuxième exception : la distribution des longueurs de phrases dans *Dom Garcie* (Molière) et dans *Psyché* (Corneille présenté sous le nom de Molière) ne diverge pas significativement de celle observée dans les tragédies (et les tragi-comédies) en alexandrins de P. Corneille. 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*).

Conclusions

Quelques remarques avant de conclure :

- Avant nous, au moins trois personnes - P. Louÿs⁴, H. Poulaille et H. Wouters - avaient relevé les ressemblances troublantes entre les œuvres présentées par Corneille et par Molière.

- La recherche qui vient de vous être présentée est collective. P. Hubert, C. Labbé et D. Monière ont collaboré à la mise au point de la distance intertextuelle ; C. Labbé, X. Luong et M. Ruhlman à celle de la classification arborée. Nous ignorons d'ailleurs le nom de certains de nos "collaborateurs" puisque les relecteurs de nos communications en congrès, de nos articles dans les revues scientifiques sont anonymes...

- Tous nos travaux sur cette question sont consultables sur le site "archives en ligne" du CNRS (HAL-SHS). Nos programmes et nos données sont dans le domaine public. La plupart des documents historiques que nous avons cités sont en ligne, les autres ont été publiés. Tout est vérifiable, tout est reproductible.

Enfin, depuis neuf ans, nous proposons à nos contradicteurs de réaliser les expériences de leur choix, comparables à celle qui vous a été présentée ce soir. Leur silence montre qu'ils reconnaissent le sérieux de notre travail...

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.

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

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).

- 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*.

L'oubli du système du comédien poète a permis à la légende du Molière grand auteur de prendre corps, longtemps après sa mort. Cette légende aveugle beaucoup d'universitaires, de journalistes et d'éditeurs français qui refusent de prendre en compte le dossier qui vient de vous être présenté.

Pour comprendre leur attitude, revenons aux journalistes du film *Qui a tué Liberty Valence*. A la fin du film, ces journalistes savent que l'homme qui a tué Liberty Valance n'est pas celui qu'ils croyaient et que l'histoire est moins brillante, plus compliquée... Mais ils décident : "Quand la légende dépasse la réalité, alors on publie la légende"...

Pour la même raison, la légende d'un Molière grand auteur, et premier "intellectuel" des temps modernes, continue à être imprimée, même si tout le monde se doute bien maintenant qu'elle ne correspond pas à la vérité.

Références

Tous nos travaux cités sont accessibles en ligne à partir de notre page personnelle. On consultera également le site corneille-moliere.org.

Arnold Edward (2005). "Le discours de Tony Blair (1997-2004)". *Corpus*, 4, p. 55-77.

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. 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.

Boileau Nicolas (1674). "L'art poétique". Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966.

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.
- Jurgens Madeleine & Elisabeth Maxfield-Miller (1963). *Cent ans de recherches sur Molière*, Paris, Imprimerie nationale.
- Duchêne Roger (1998), *Molière*, Paris, Fayard.
- La Fontaine Jean de (1968). *Œuvres diverses*. Paris : Gallimard, La Pléiade.
- 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". *Language 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 (2008). "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 (2010). "Ce que disent les phrases de Corneille et Molière". Communication devant les Xe Journées Internationales d'Analyse des Données Textuelles. Rome : 11 juin 2010.
- 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, II, p. 687-696.
- Labbé Dominique & Monière Denis (2008). *Les mots qui nous gouvernent*. Montréal : Monière-Wollank Editeurs.
- Lafon Michel & Peeters Benoît (2006). *Nous est un autre*. Paris, Flammarion.
- Luong Xuan (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Paris : Université de Paris V.
- 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.
- Reynier Gustave (1892). *Thomas Corneille, sa vie et son théâtre*. Paris : Hachette.
- 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.
- Tuzzi Arjuna, Popescu Ioan-Iovitz & Altmann Gabriel (2010). Quantitative Analysis of Italian Texts. *Studies in Quantitative Linguistics*, 6.
- Vinha da Mathieu (2004). *Les valets de chambre de Louis XIV*. Paris.
- Wouters Hippolyte & Ville de Goyet Christine de (1990). *Molière ou l'auteur imaginaire ?* Bruxelles : Complexe.
- Young Bert E. & Grace P. (1997). *Le registre de La Grange*, Genève, Slatkine.

Annexe I. Les pièces de Pierre Corneille.

		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	Oedipe	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

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

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

	Création	Genre	Longueur mots
37 La jalousie	Avant 1659	Comédie prose	3 501
38 Médecin volant	Avant 1659	Comédie prose	3 876
39 L'étourdi*	1659	Comédie vers	18 671
40 Dépit amoureux*	1659	Comédie vers	16 242
41 Précieuses ridicules	1660	Comédie prose	6 648
42 Sganarelle*	1660	Comédie vers	6 042
43 Dom Garcie*	1661	Comédie héroïque vers	17 049
44 L'école des maris*	1661	Comédie vers	10 536
45 Les fâcheux*	1661	Comédie vers	7 922
46 L'école des femmes*	1662	Comédie vers	16 625
47 Critique de l'école	1663	Comédie prose	8 610
48 L'impromptu	1663	Comédie prose	7 168
49 Mariage forcé	1664	Comédie prose	6 058
50 Princesse d'Elide*	1664	Comédie vers et prose	11 333
51 Le Tartuffe*	1664	Comédie vers	18 271
52 Dom Juan*	1665	Comédie prose	17 452
53 L'amour médecin	1665	Comédie prose	6 147
54 Le Misanthrope*	1666	Comédie vers	17 180
55 Médecin malgré lui	1666	Comédie prose	9 317
56 Mélicerte*	1666	Comédie vers	5 540
** Comédie pastorale	1667	Comédie vers libres	732
57 Le sicilien	1667	Comédie prose	5 375
58 Amphytrion*	1668	Comédie vers libres	15 117
59 Georges Dandin	1668	Comédie prose	11 009
60 L'avare*	1668	Comédie prose	21 033
61 M. de Pourceaugnac	1669	Comédie prose	11 803
62 Amants magnifiques*	1670	Comédie vers & prose	11 983
63 Bourgeois gentilhom.*	1670	Comédie prose	17 132
64 Fourberies de Scapin	1671	Comédie prose	14 245
65 Escarbagnas	1671	Comédie prose	5 564
66 Femmes savantes*	1672	Comédie vers	16 863
67 Malade imaginaire*	1673	Comédie prose	19 919

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

** Pièce retirée dans les expériences à cause de sa petite taille.

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

Annexe III
Distances séparant les deux *Menteurs* (Corneille) et les *Plaideurs* (Racine)
de toutes les pièces de Molière

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 IV. Romans sélectionnés pour l'expérience en aveugle.

N°	Auteur	Titre	N°	Auteur	Titre
1	Balzac	Père Goriot	29	Maupassant	Fort comme la mort
2	Dumas(père)	Monte Cristo	30	Stendhal	Chartreuse de Parme
3	Dumas(fils)	Dame aux camélias	31	Vigny	Grandeur et servitude
4	Flaubert	Bouvard & Pécuchet	32	Zola	Germinal
5	Gautier	Avatar	33	Dumas(père)	Trois mousquetaires
6	Hugo	Misérables	34	Flaubert	Bovary
7	Lamartine	Graziella	35	Hugo	Notre-Dame de Paris
8	Maupassant	Bel ami	36	Maupassant	Fort comme la mort
9	Stendhal	Chartreuse de Parme	37	Stendhal	Rouge et Noir
10	Vernes	Tour du monde	38	Zola	Germinal
11	Vigny	Servitude et grandeur	39	Dumas(père)	Trois mousquetaires
12	Zola	Bête humaine	40	Flaubert	Bovary
13	Balzac	Père Goriot	41	Hugo	Notre-Dame de Paris
14	Dumas(père)	Monte Cristo	42	Maupassant	Mont Oriol
15	Dumas(fils)	Dame aux camélias	43	Stendhal	Rouge et Noir
16	Flaubert	Bouvard & Pécuchet	44	Zola	l'Argent
17	Gautier	Avatar	45	Flaubert	Education sentimentale
18	Hugo	Misérables	46	Hugo	Notre-Dame de Paris
19	Lamartine	Graziella	47	Maupassant	Mont Oriol
20	Maupassant	Bel ami	48	Stendhal	Rouge et Noir
21	Stendhal	Chartreuse de Parme	49	Zola	l'Argent
22	Vernes	Tour du monde	50	Flaubert	Education sentimentale
23	Vigny	Servitude et grandeur	51	Maupassant	Pierre et Jean
24	Zola	Bête humaine	52	Zola	l'Argent
25	Dumas(père)	Trois mousquetaires	53	Flaubert	Education sentimentale
26	Dumas(fils)	Dame aux camélias	54	Maupassant	Pierre et Jean
27	Flaubert	Bovary	55	Maupassant	Une vie
28	Hugo	Misérables	56	Maupassant	Une vie

Annexe V Les distances remarquablement faibles
(inférieures à la moyenne diminuée de deux écarts-types)

Rang	Texte 1	Texte2	Distance	Auteur 1	Auteur 2
1	15	26	0,17773	Dumas fils	Dumas fils
2	03	26	0,19095	Dumas fils	Dumas fils
3	03	15	0,20389	Dumas fils	Dumas fils
4	11	31	0,20811	Vigny	Vigny
5	12	24	0,21089	Zola	Zola
6	32	38	0,22008	Zola	Zola
7	23	31	0,22010	Vigny	Vigny
8	11	23	0,22011	Vigny	Vigny
9	51	53	0,22824	Maupassant	Maupassant
10	02	14	0,22916	Dumas	Dumas
11	55	56	0,23012	Maupassant	Maupassant
12	45	50	0,23274	Flaubert	Flaubert
13	34	40	0,23465	Flaubert	Flaubert
14	25	39	0,23579	Dumas	Dumas
15	42	47	0,23639	Maupassant	Maupassant
16	27	34	0,23718	Flaubert	Flaubert
17	33	39	0,23756	Dumas	Dumas
18	50	53	0,23957	Flaubert	Flaubert
19	37	43	0,24243	Stendhal	Stendhal
20	54	55	0,24407	Maupassant	Maupassant
21	18	28	0,24412	Hugo	Hugo
22	08	20	0,24608	Maupassant	Maupassant
23	54	56	0,25002	Maupassant	Maupassant
24	45	53	0,25019	Flaubert	Flaubert
25	20	51	0,25056	Maupassant	Maupassant
26	14	33	0,25242	Dumas	Dumas
27	08	51	0,25356	Maupassant	Maupassant
28	06	28	0,25412	Hugo	Hugo
29	04	16	0,25500	Flaubert	Flaubert
30	21	30	0,25527	Stendhal	Stendhal
31	29	54	0,25535	Maupassant	Maupassant
32	20	54	0,25569	Maupassant	Maupassant
33	05	17	0,25661	Gautier	Gautier
34	29	36	0,25746	Maupassant	Maupassant
35	20	29	0,25850	Maupassant	Maupassant
36	36	54	0,25873	Maupassant	Maupassant
37	44	52	0,25891	Zola	Zola
38	27	40	0,25892	Flaubert	Flaubert
39	01	13	0,25897	Balzac	Balzac
40	41	46	0,25928	Hugo	Hugo
41	27	45	0,25950	Flaubert	Flaubert
42	20	56	0,26010	Maupassant	Maupassant

NB : les noms des auteurs et des œuvres ont été ajoutés après l'expérience.

Annexe VI. Les distances remarquablement fortes
(supérieures à la moyenne augmentée de deux écarts-types)

Rang	Texte 1	Texte2	Distance	Auteur 1	Auteur 2
1494	07	48	0,41725	Lamartine	Stendhal
1495	07	28	0,41751	Lamartine	Hugo
1496	07	10	0,41776	Lamartine	Vernes
1497	02	04	0,41838	Dumas(père)	Maupassant
1498	05	26	0,41962	Gautier	Dumas(fils)
1499	03	50	0,42015	Dumas(fils)	Flaubert
1500	03	10	0,42044	Dumas(fils)	Vernes
1501	10	15	0,42097	Vernes	Dumas(fils)
1502	07	18	0,42151	Lamartine	Hugo
1503	16	33	0,42195	Flaubert	Dumas(père)
1504	02	07	0,42303	Dumas(père)	Lamartine
1505	15	42	0,42419	Dumas(fils)	Maupassant
1506	15	27	0,42529	Dumas(fils)	Flaubert
1507	26	42	0,42575	Dumas(fils)	Maupassant
1508	26	27	0,42714	Dumas(fils)	Flaubert
1509	03	27	0,42920	Dumas(fils)	Flaubert
1510	07	33	0,43062	Lamartine	Dumas(père)
1511	03	22	0,43183	Dumas(fils)	Vernes
1512	10	26	0,43219	Vernes	Dumas(fils)
1513	15	49	0,43303	Dumas(fils)	Zola
1514	03	49	0,43364	Dumas(fils)	Zola
1515	03	05	0,43481	Dumas(fils)	Gautier
1516	26	50	0,43511	Dumas(fils)	Flaubert
1517	03	53	0,43527	Dumas(fils)	Flaubert
1518	26	49	0,43544	Dumas(fils)	Zola
1519	02	16	0,43565	Dumas(père)	Flaubert
1520	15	50	0,43584	Dumas(fils)	Flaubert
1521	15	22	0,43642	Dumas(fils)	Vernes
1522	03	45	0,43778	Dumas(fils)	Flaubert
1523	03	32	0,43797	Dumas(fils)	Zola
1524	22	26	0,44132	Vernes	Dumas(fils)
1525	15	53	0,44238	Dumas(fils)	Flaubert
1526	05	15	0,44579	Gautier	Dumas(fils)
1527	26	53	0,44583	Dumas(fils)	Flaubert
1528	15	45	0,44714	Dumas(fils)	Flaubert
1529	03	04	0,45118	Dumas(fils)	Flaubert
1530	04	15	0,45186	Flaubert	Dumas(fils)
1531	26	32	0,45260	Dumas(fils)	Zola
1532	15	32	0,45317	Dumas(fils)	Zola
1533	26	45	0,45502	Dumas(fils)	Flaubert
1534	04	26	0,46506	Flaubert	Dumas(fils)
1535	03	16	0,46711	Dumas(fils)	Flaubert
1536	07	26	0,46854	Lamartine	Dumas(fils)
1537	15	16	0,47232	Dumas(fils)	Flaubert
1538	03	07	0,47321	Dumas(fils)	Lamartine
1539	07	15	0,47369	Lamartine	Dumas(fils)
1540	16	26	0,47717	Flaubert	Dumas(fils)

Annexe VIII 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 IX. Les principaux groupes verbaux chez Ajar et Gary (fréquence pour 10.000 mots)

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