

7e Journées d'Analyse des Données Textuelles

Louvain-la-Neuve 11 mars 2004

Table ronde

CORNEILLE ET MOLIÈRE

Sous la présidence de M. Ludovic Lebart

Avec la participation de :

Mme Valérie Beaudouin

MM. Dominique Labbé, Jean-Marie Viprey et Hippolyte Wouters.

Dossier établi par :

Dominique Labbé

dominique.labbe@iep.upmf-grenoble.fr

25 mars 2004

Avec deux documents supplémentaires (2008)

Résumé

Compte-rendu des interventions de l'auteur lors de la table ronde, sur Corneille et Molière, réunie pendant des Journées d'Analyse statistique des Données Textuelles à Louvain-la-Neuve le 11 mars 2004. Un des participants ayant affirmé qu'il faut utiliser les "collocations" pour l'attribution d'auteur, il est démontré que les collocations des mots les plus fréquents prouvent la paternité de Corneille sur les pièces en vers de Molière, ainsi que sur Dom Juan et l'Avare. En réponse aux interventions, on rappelle les limites de validité de l'échelle de la distance intertextuelle et, dans ces limites, l'indépendance de cette distance par rapport aux différences de longueur entre les textes.

Mots-clefs : collocation ; distance intertextuelle ; attribution d'auteur ; statistique textuelle ; Corneille ; Molière

Sommaire

Avertissements (2008 et 2004)

I. Le sens des mots est le même chez Corneille et chez Molière	5
1. <i>L'amour dans le théâtre du XVIIe</i>	6
2. <i>Cœur et autres mots usuels du théâtre du XVIIe</i>	11
3. <i>Madame, aimer, monsieur...</i>	14
<i>Conclusions</i>	19
II. Distance intertextuelle et taille des textes	21
III Réponse à la communication de M. Viprey	24
IV Réponse à la communication de Mme Beaudouin	29

Annexes

I. Les vingt lignes introduisant le débat	30
II. Tableau des œuvres de Corneille	32
III. Par qui ont été écrites les pièces de "Molière" ?	33
IV. Compte-rendu de la table ronde par M. Benoît Peters	35
IV. Compte-rendu de l'intervention de M. J.-M. Viprey par lui-même	37

Avertissement

(mars 2008)

Le dossier ci-dessous a été mis en ligne juste après la table ronde (mars 2004).

En mars 2008, il est augmenté de deux documents (placés à la fin) :

- le compte rendu rédigé par l'une des personnes ayant assisté à cette table ronde (M. Benoît Peeters). Celui-ci l'avait immédiatement placé en ligne sur le site de sa maison d'édition – où il est toujours consultable - et il l'avait fait circuler dans plusieurs listes de discussion (notamment LITOR).

- M. J.-M. Viprey avait estimé que le document de M. B. Peeters ne rendait pas exactement compte de ses propos. En avril 2004, il avait rédigé son propre compte-rendu qu'il l'avait placé sur sa page personnelle et communiqué à plusieurs listes de discussion (notamment LITOR).

A l'automne 2007, ce dernier document a disparu de sa page personnelle.

Ce texte éclaire bien la discussion qui a eu lieu ce soir-là à Louvain. Nous le reproduisons en annexe à la fin de ce dossier, dans la version originale qui avait été mise en ligne et diffusée aux participants à la liste LITOR.

Enfin, dans mon avertissement de mars 2004 (ci-dessous), il est fait allusion à deux communications qui se voulaient des "réfutations" de notre propre travail. Comme toutes les communications au congrès, elles peuvent être consultées sur le site de la revue électronique Lexicometrica (<http://www.cavi.univ-paris3.fr/lexicometrica/>) :

http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_009.pdf

http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_115.pdf

Avertissement

(mars 2004)

En octobre 2003, les organisateurs des 7e journées de LLN ont pris contact avec H. Wouters et moi-même pour une table ronde autour de Corneille et Molière. J'ai simplement demandé que l'équilibre soit respecté et que chaque participant remette un document préparatoire qui serait soumis au comité de lecture, publié dans les actes du congrès et envoyé à l'avance à chacun des participants.

A la mi-février 2004, cette table ronde était confirmée, les noms de nos contradicteurs étaient annoncés (Mme V. Beaudouin et M. J.-M. Viprey), mais les documents préparatoires étaient réduits à "20 lignes" (reproduites en annexe à la fin de ce dossier).

A l'ouverture du congrès, le 10 mars au matin, les Actes ont été remis aux participants. J'ai eu la surprise de constater que mes deux contradicteurs avaient présenté des communications qui se voulaient des "réfutations" de mon propre travail. Les organisateurs avaient ces textes

en main quand ils ont refusé que soient échangés des textes préparatoires, comme il est d'usage dans un débat de ce genre... En 24 heures, j'ai élaboré une réponse écrite à chacun de ces deux textes. On pourra lire ces réponses à la fin de ce dossier, telles qu'elles ont été distribuées à l'assistance lors de la table ronde. A la fin de ces deux textes, un certain nombre de questions sont posées publiquement aux organisateurs de ces journées. Après le congrès, elles ont été réitérées par courrier. A ce jour, elles sont restées sans réponse. Naturellement, il n'est pas possible de reproduire les textes originaux de Mme Beaudouin et de M. Viprey. Le lecteur curieux devra se procurer les Actes ou demander copie aux auteurs. A moins que ceux-ci ne placent ces textes sur leurs pages personnelles ?

Malheureusement, les organisateurs n'ont pas "songé" à enregistrer la table ronde. Il reste donc les témoignages des présents et les notes des participants. Personne ne conteste les points suivants :

M. Viprey a avoué qu'il n'est pas statisticien (depuis un an, il s'est présenté comme "expert" dans de multiples journaux, séminaires et document électronique). Il a admis que la question de la paternité des œuvres de Molière est posée, ce qu'il niait jusqu'à maintenant.

On retiendra de l'intervention de Mme Beaudouin que l'étude de la prosodie ne permet pas de distinguer les deux Menteurs de Corneille des pièces en vers de Molière. Depuis un an et demi, nos critiques prétendaient l'inverse.

Enfin M. Benoît Peeters a tenté un résumé des différentes interventions et a placé ce document sur le site des Impressions Nouvelles (<http://www.lespierides.com/in/moliere.html>). Il faut espérer que d'autres spectateurs feront part de leurs souvenirs et de leurs impressions.

Lors de cette table ronde, j'ai d'abord présenté brièvement le calcul de la distance intertextuelle et la méthode d'attribution d'auteur. Puis, comme M. Viprey avait affirmé qu'il fallait utiliser pour cela les "collocations", j'ai montré qu'elles convergent totalement avec nos conclusions antérieures. Enfin, lors d'un deuxième tour de table, j'ai répondu à quelques objections. On trouvera ces réponses dans les pages suivantes.

Ce dossier a été établi le 25 mars 2004. Il sera augmenté si de nouveaux éléments venaient à être publiés à ce sujet.

I. Le sens des mots est le même dans les pièces de Corneille et dans celles qu'il a écrites pour Molière¹

Dans ses "vingt lignes" publiées à la fin des Actes de ce congrès (texte reproduit en annexe à ce dossier) M. J.-M. Viprey affirme que :

— notre méthode est "probabiliste". C'est une erreur : notre mesure est une distance qui enregistre une réalité physique, comme la distance en kilomètres séparant deux villes ;

— les interjections seraient caractéristiques de Molière. S'il a raison, certaines pièces de "Molière" qui ne contiennent pratiquement aucune interjection, comme Dom Garcie de Navarre ou le Misanthrope — il n'y a pas de "holà" dans le Misanthrope — ne sont donc pas de "Molière" ?

— "Molière" s'inspirerait de la Commedia dell'Arte et pas Corneille. L'auditoire pourra juger du sérieux de cette affirmation en se reportant, d'une part, aux sources qui inspirent les pièces de Molière telles que M. Bourqui les a reconstituées (Les sources de Molière, Paris, SEDES, 1999) et, d'autre part, à tous les grands critiques qui ont souligné la profondeur des influences italiennes chez Corneille, à commencer par Chateaubriand.

— notre méthode ne prendrait pas en compte les "collocations" et autres combinaisons entre les mots.

Cette dernière affirmation est également fautive. Nous avons écrit dès l'origine que les combinaisons de mots sont une manière possible de confirmer ou de mettre en doute une attribution d'auteur fondée sur la distance intertextuelle. Nous avons signalé que la similarité de sens entre les mots les plus fréquents chez Corneille et dans les pièces qu'il a écrites pour Molière ne se rencontre jamais à ce point chez deux auteurs différents, même quand ces deux auteurs sont contemporains, qu'ils écrivent sur les mêmes thèmes en utilisant un même genre.

Nous allons développer ce point car il pose des questions qui intéressent les participants à un congrès de statistique textuelle : comment mesurer le sens qu'un auteur donne à un mot ? Comment mesurer la similarité de sens dans deux œuvres différentes ?

Les vingt lignes de notre contradicteur suggèrent d'utiliser les "collocations". Eh bien ! examinons ces collocations dans le théâtre du XVIIe.

On peut définir cette notion de la manière suivante : *"le vocabulaire entourant le mot pivot dans un espace limité — généralement la phrase — sans prendre en considération l'ordre des mots"*.

L'analyse proposée repose sur deux postulats implicites. D'une part, le sens d'un mot lui vient du "contexte" étroit qui tisse une sorte de réseau de relations ; d'autre part, chaque auteur donne aux mots qu'il emploie un sens qui lui est propre et qui se différencie de celui des autres... Naturellement, ce second postulat doit être manié avec prudence : le sens d'un mot chez un même auteur pourra être différent suivant le genre (comédie, tragédie...) le thème ou l'époque.

¹ Cette intervention et celles qui suivent ont dû être raccourcies pour respecter les temps de parole très brefs accordés par les organisateurs et pour permettre à la salle de s'exprimer. On lira ci-dessous le texte que nous avons préparé quand les "vingt lignes" de M. Viprey nous ont été communiquées.

Comparons donc les collocations des principaux mots dans les trois corpus (pour le détail des pièces, se reporter aux annexes II et III) :

— "Corneille" : les 33 pièces sont conservées ensemble mais l'on dispose pour chacune d'une indication concernant son "genre", ce qui permet de découper des sous-corpus, notamment : tragédies et comédies ;

— "Molière" : ce corpus est subdivisé en trois sous ensembles : "Molière écrit par Corneille" ; "Molière non-écrit par Corneille" ; les pièces qui contiennent probablement une part importante de la main de Corneille mais qui ne sont pas attribuées pour l'instant et seront donc laissées de côté pour ces expériences...

— un corpus "témoin" : les pièces de Racine.

Dans chacun de ces corpus et sous-corpus, on recueille les phrases contenant le "mot pivot" recherché. On limite toutefois la taille des plus grandes phrases (l'empan maximum sera la taille moyenne de la phrase augmentée de deux écart types).

1. *L'amour dans le théâtre du XVIIe*

Nous avons choisi ce mot simplement parce qu'il est le substantif le plus employé dans les trois corpus principaux (Corneille, Racine, Molière).

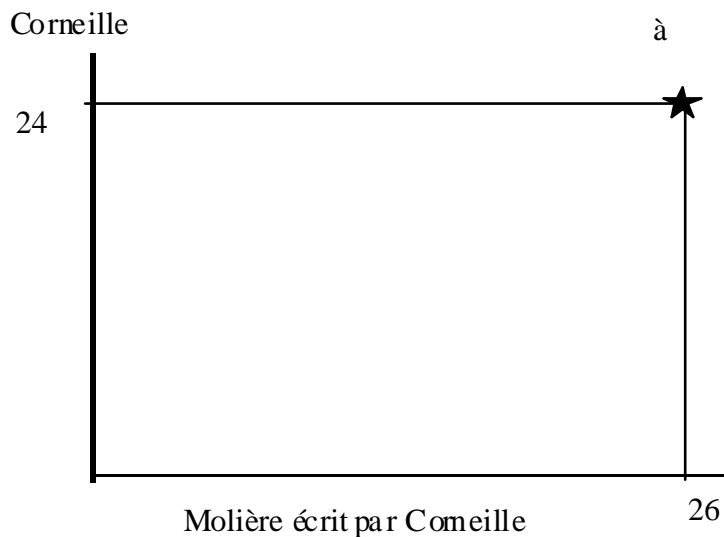
Le logiciel commence par isoler toutes les phrases contenant le vocable (en respectant la limite de taille indiquée ci-dessus). Il établit l'index alphabétique et calcule les fréquences de chacun des vocables présents dans ces phrases. Par exemple, on voit dans le tableau I que, dans les phrases contenant "amour", on rencontre le verbe "aimer" 226 fois chez Corneille et 59 fois dans les pièces qu'il a écrites pour Molière. Pour neutraliser les différences de taille entre ces corpus, on convertit les fréquences absolues en fréquences relatives.

Tableau I. Les quatre premières collocations de "amour" chez Corneille et dans les pièces de "Molière" écrites par Corneille.

Vocables	Corneille		Molière écrit par Corneille	
	Fréquence absolue	Fréquence relative	Fréquence absolue	Fréquence relative
à (pré)	1 506	23,8	405	26,3
aimer (v)	226	3,6	59	3,8
aller (v)	104	1,6	22	1,4
amant (n m)	73	1,2	15	1,0

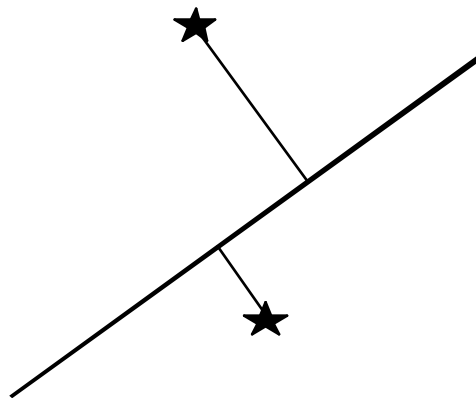
Comment comparer ces données ?

Tout d'abord un graphique s'impose : le "nuage de points". Sur le graphique, chaque vocable est représenté par un point ayant comme coordonnées sa fréquence dans les deux corpus comparés :



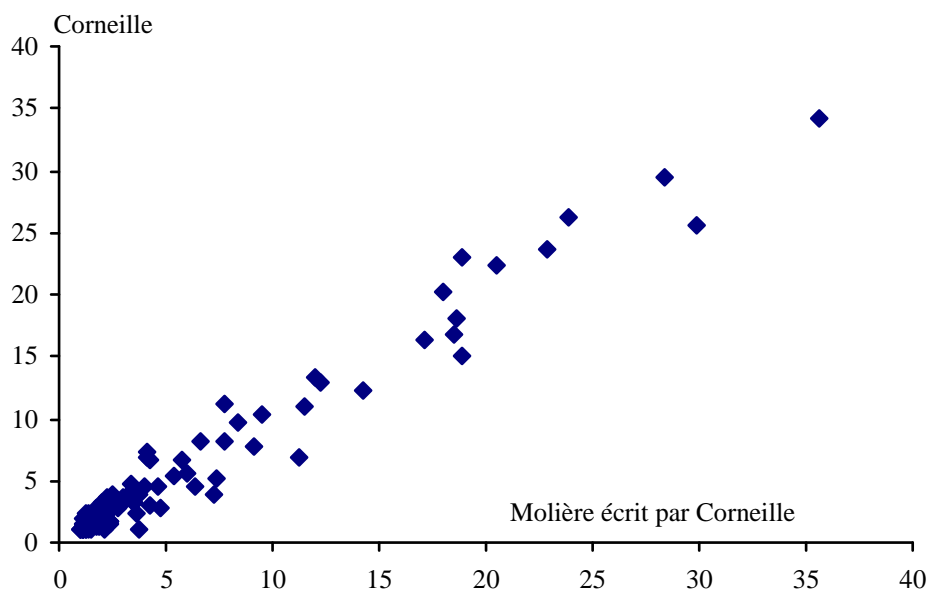
Si les fréquences sont égales dans les deux corpus, les points seront alignés sur la première bissectrice : les collocations (contextes) sont les mêmes dans les deux corpus. En fonction de notre second postulat, cela plaiderait en faveur d'un auteur unique. Plus on s'éloigne de la bissectrice plus les contextes sont différents, plus il y a de chance pour que les auteurs soient différents.

En plus de l'examen direct des nuages de points, on peut calculer l'écart plus ou moins grand par rapport à la situation théorique idéale : tous les points alignés. La qualité de cette "régression linéaire" est donnée par le coefficient de "corrélation". Nous expliquerons en conclusion que les fréquences d'un même mot dans deux corpus différents ne peuvent être considérées comme deux variables indépendantes. Au contraire, quel que soit l'auteur, elles sont fortement reliées. Le coefficient indique donc simplement la qualité plus ou moins grande de l'ajustement du nuage de points par la droite.



Voici le nuage des collocations de "amour" dans les corpus "Corneille" et "Molière écrit par Corneille" :

Tableau II. Nuage des collocations de « amour » chez Corneille et dans les pièces écrites pour Molière



96 vocables communs ($F > 1\%$) ; $R = .980$

L'examen visuel de ce graphe montre que le nuage de points est pratiquement aligné sur la première bissectrice : les densités d'emploi sont les mêmes dans les deux corpus.

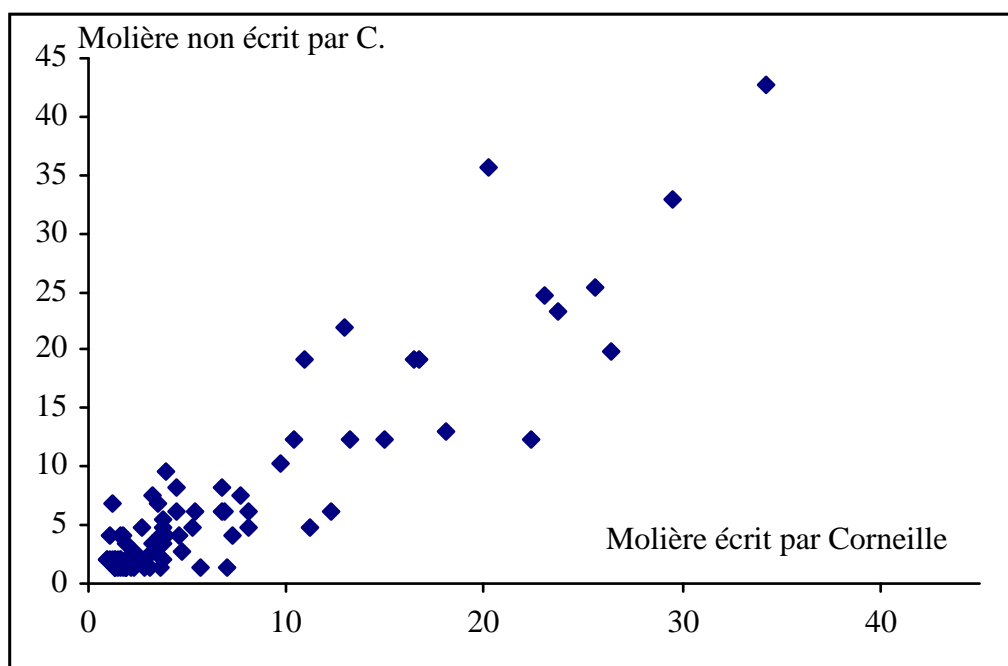
Deux informations supplémentaires peuvent être retenues :

— le nombre de vocables dont les fréquences relatives sont supérieures au seuil choisi (ici 1%) dans les deux corpus considérés. Plus ce nombre est important, plus les deux corpus partagent le même vocabulaire. Ainsi, les collocations de "amour" dans les corpus "Corneille" et "Molière écrit par Corneille" partagent 96 vocables dont la fréquence est supérieure à 1% ;

— la proximité des points par rapport à la bissectrice. Classiquement, on mesure la qualité de cet ajustement linéaire grâce au coefficient de "corrélation" du même nom. Plus ce coefficient est proche de 1, plus l'ajustement du nuage par la bissectrice est bon. Ici le coefficient est égal à 0.980. L'ajustement est excellent : les fréquences sont les mêmes dans les deux corpus.

Voyons ce même nuage pour les collocations de "amour" dans les deux corpus "Molière écrit par Corneille" et "Molière non-écrit par Corneille"

Tableau III. Nuage des collocations de "amour" dans "Molière écrit par Corneille" et dans "Molière non-écrit" par Corneille



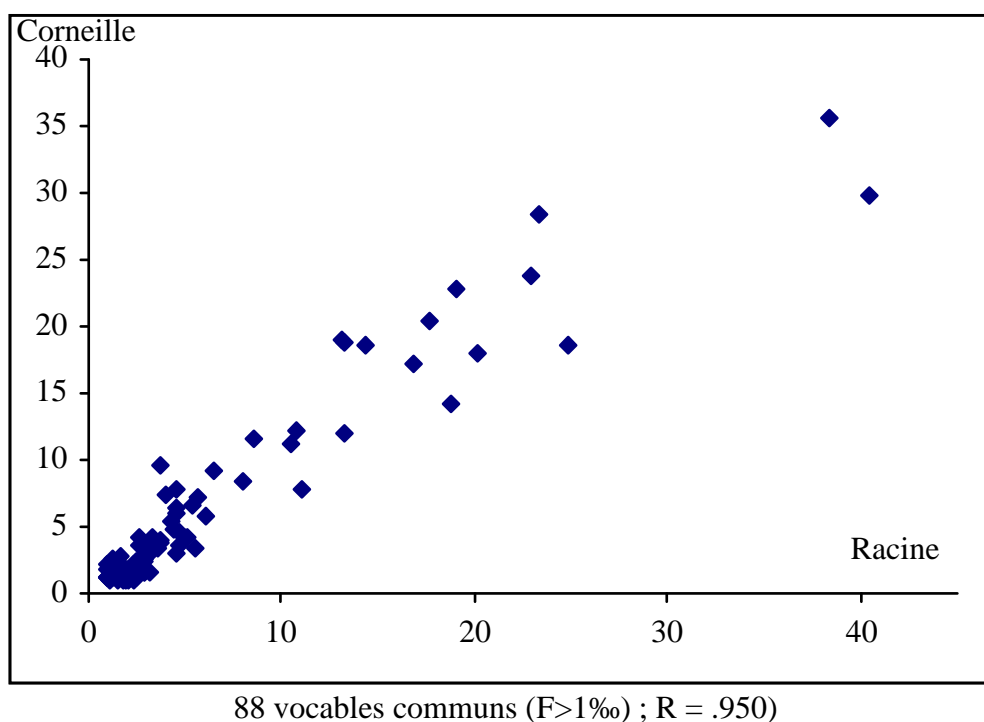
80 vocables communs ($F > 1\%$) ; $R = .922$

Non seulement ces mêmes collocations dans les deux corpus "Molière" (écrit par Corneille et non-écrit par Corneille) partagent moins de mots mais, de plus, les fréquences de ces vocables sont significativement différentes comme l'indique la dispersion du nuage et le coefficient de régression plus faible. Si l'auteur de ces deux corpus était le même, comme le pensent la plupart des littéraires, ce devrait être l'inverse d'autant plus que le corpus Corneille contient aussi des tragédies et s'étend sur plus de 40 ans alors que les deux corpus "Molière" ne comportent que des comédies et sont contemporains.

Toutefois, une limite mérite d'être signalée. Les collocations de "amour" dans le corpus "Molière non-écrit par Corneille" sont en nombre restreint (à peine plus de 1500 mots). Le bas du tableau porte donc sur des mots dont la fréquence absolue est très faible, ce qui peut expliquer en partie la dispersion du nuage au moins dans sa partie basse. On peut faire l'effort mental de couper le bas du tableau par exemple à 10% (soit une fréquence absolue au moins égale à 15 dans le corpus Molière non-écrit par C.) : le nuage reste très dispersé. Malgré tout, Cette remarque souligne la principale limite de la technique : elle ne peut être appliquée qu'à quelques mots très fréquents et elle doit l'être avec prudence...

Naturellement il n'est pas possible de conclure sans avoir auparavant tenté l'expérience sur d'autres auteurs. Le corpus "Racine" permet de réaliser cette expérience. Pour la comparaison avec Corneille, le test est durci : seules les tragédies sont retenues afin d'avoir une unité de genre et une moins grande dispersion chronologique : on mesurera donc précisément l'influence du facteur "auteur".

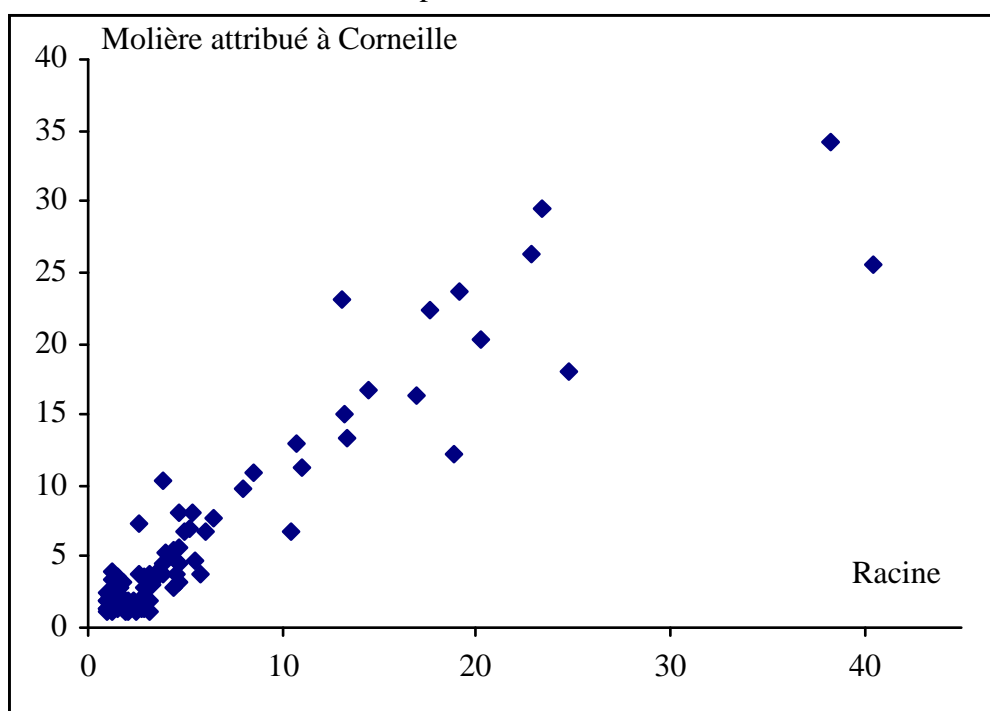
Tableau IV. Nuage des collocations de « amour » dans les tragédies de Corneille et de Racine



Si l'ajustement est meilleur qu'entre les deux corpus "Molière", il reste tout de même nettement moins bon qu'entre "Corneille" et "Molière écrit par Corneille" alors que l'unité de genre devrait normalement aboutir au résultat inverse.

Pour parfaire l'expérience, comparons l'amour chez Racine (en incluant les Plaideurs) et chez "Molière écrit par Corneille".

Tableau V. Les collocations de "amour" chez Racine et dans les pièces de "Molière" écrites par Corneille.



75 vocables communs (F>1%) ; R = .933)

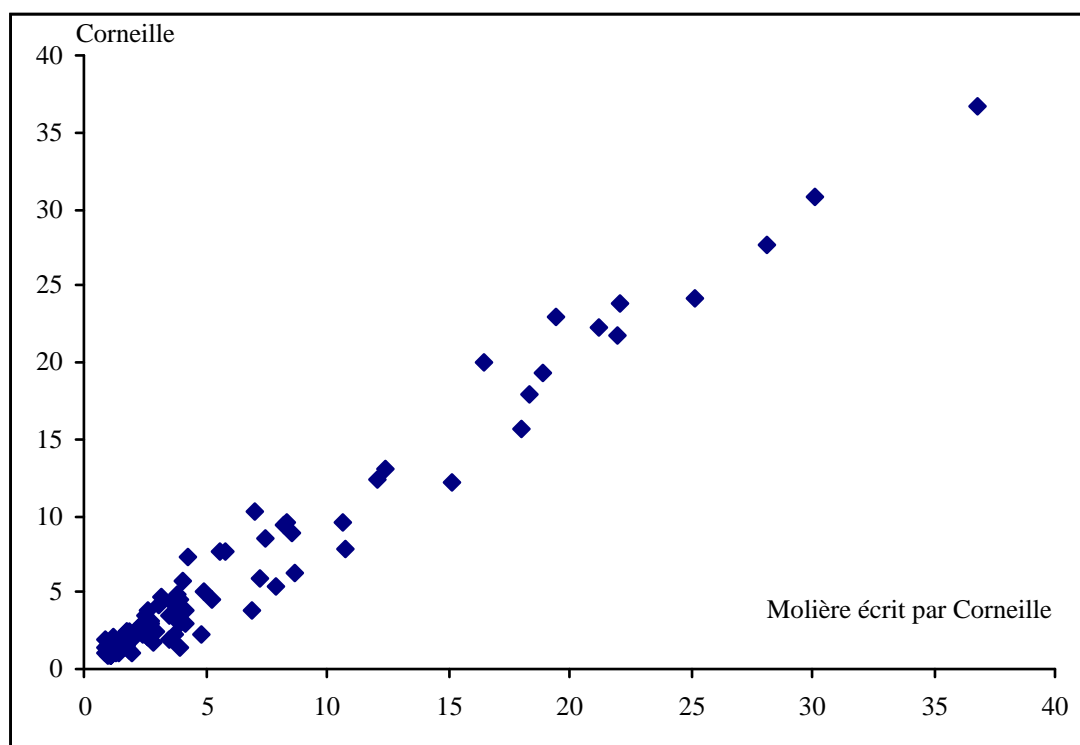
Autrement dit, les collocations de "amour" chez ces trois auteurs corroborent les conclusions tirées de la distance intertextuelle et des classifications automatiques (le statut de cette "preuve" est discuté en conclusion). Nous l'avons écrit clairement.

Certes, on pourra prétendre qu'un seul test est insuffisant. Voyons donc les autres mots usuels, du moins ceux qui sont présents en densité suffisante dans les différents corpus.

2. "Cœur" et autres mots usuels du théâtre du XVIIe

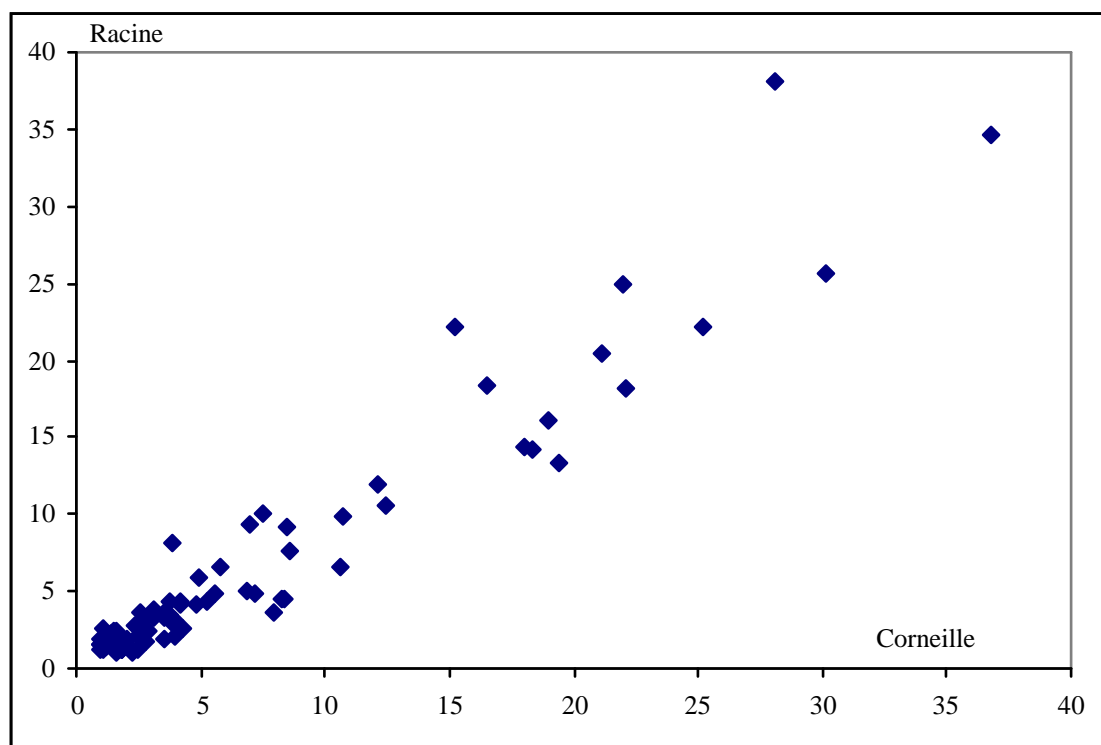
Les constats tirés de l'examen des collocations de "amour" sont confirmés par l'examen des mots suivants dans la hiérarchie des fréquences. Ci-dessous les trois graphiques obtenus sur les différents corpus avec le mot "cœur".

Tableau VI. Les collocations de "cœur" chez Corneille et dans les pièces de Molière écrites par Corneille



102 mots (F > 1%) R = 0.981

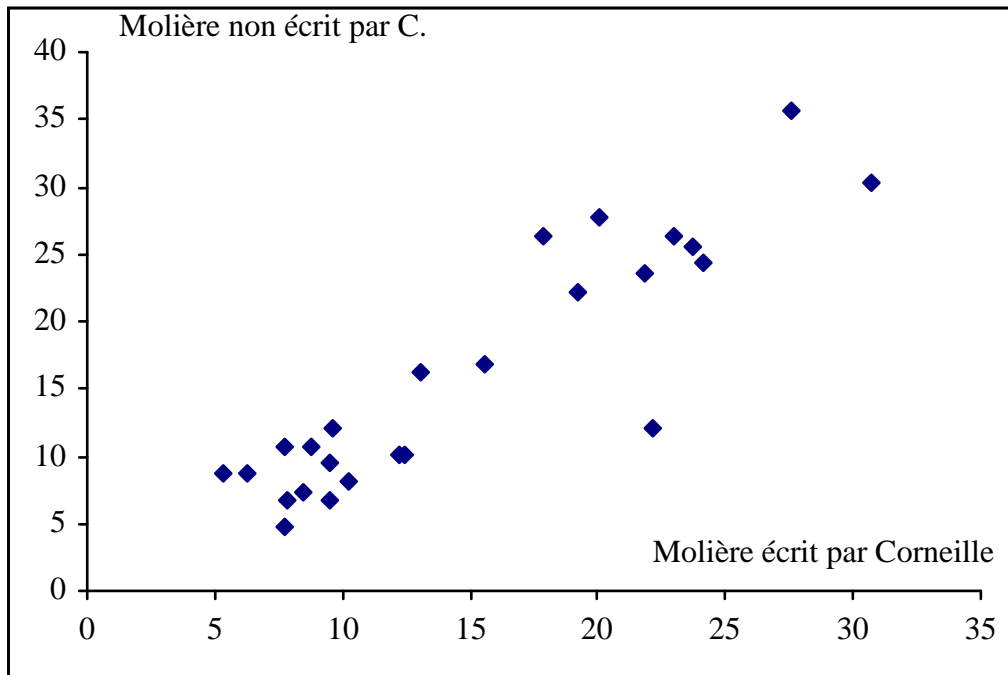
Tableau VII. Comparaison des collocations de "cœur" dans les tragédies de Racine et de Corneille



87 vocables (F > 1%) R = 0.954

La comparaison avec "Molière non écrit par Corneille" est plus difficile car, là encore, les collocations ont une surface trop faible (1460 mots). Les fréquences relatives sont donc calculées sur de petits effectifs. Dans une population aussi faible, les "accidents" pèsent lourd. Voici tout de même le graphe en interrompant le calcul au dessus de 5‰ (fréquence absolue égale à 8). Il ne reste pratiquement plus que de mots outils qui sont normalement ceux dont les différences entre auteurs sont les plus faibles.

Tableau VIII. Collocations de "cœur" dans les pièces de Molière écrites par Corneille et dans celles non écrites par Corneille



(R = 0.923)

Tous les substantifs suivants (notamment "œil", "chose", "père", "fils") ainsi que les principaux verbes (faire, dire, vouloir, voir) présentent des profils semblables. On observe juste que les nuages obtenus avec les verbes sont légèrement moins bons mais que l'écart entre les corpus reste le même : l'ajustement est toujours meilleur entre le corpus "Corneille" et "Molière écrit par Corneille" qu'entre toutes les autres combinaisons imaginables.

Il faut s'arrêter sur les quelques cas un peu moins favorables. Ceux-ci s'expliquent probablement par l'influence du genre. Le verbe "aimer" et les substantifs "madame", "monsieur" fournissent les exemples les plus défavorables à notre démonstration sans pour autant l'infirmier.

3. "Madame", "aimer", "monsieur"

Parmi tous les mots pour lesquels le calcul a un sens, les trois mots les plus défavorables à notre démonstration sont "madame", "aimer" et "monsieur". Voici les résultats obtenus :

Tableau IX. Collocations de "madame" dans les corpus Corneille et dans les pièces écrites par Corneille pour Molière

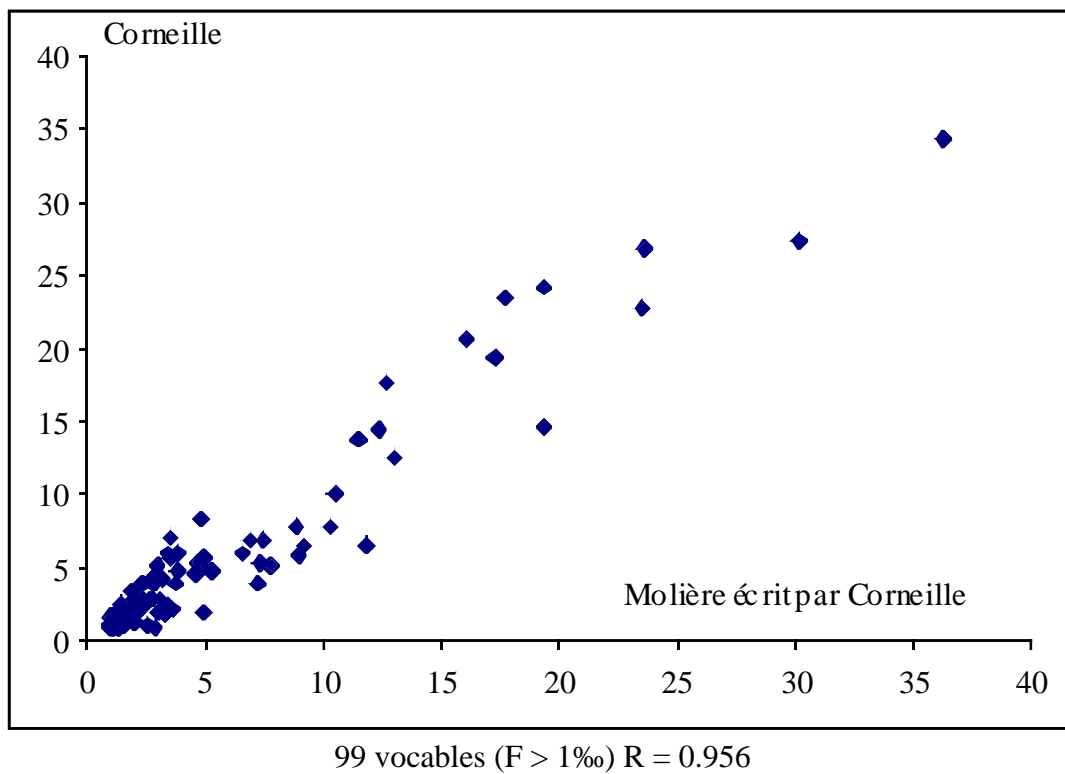
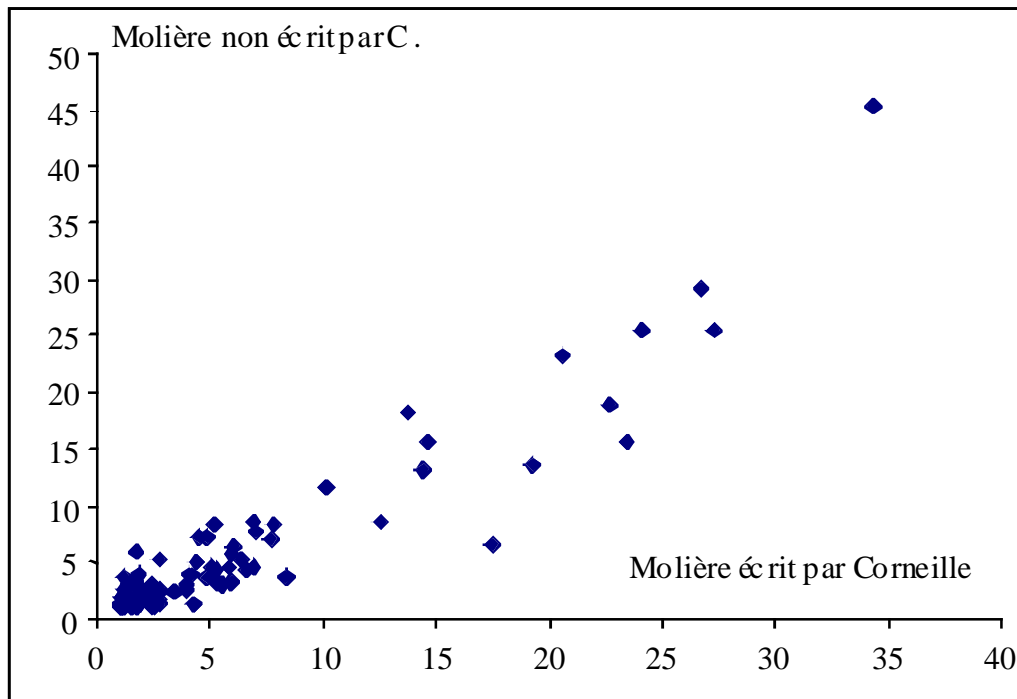
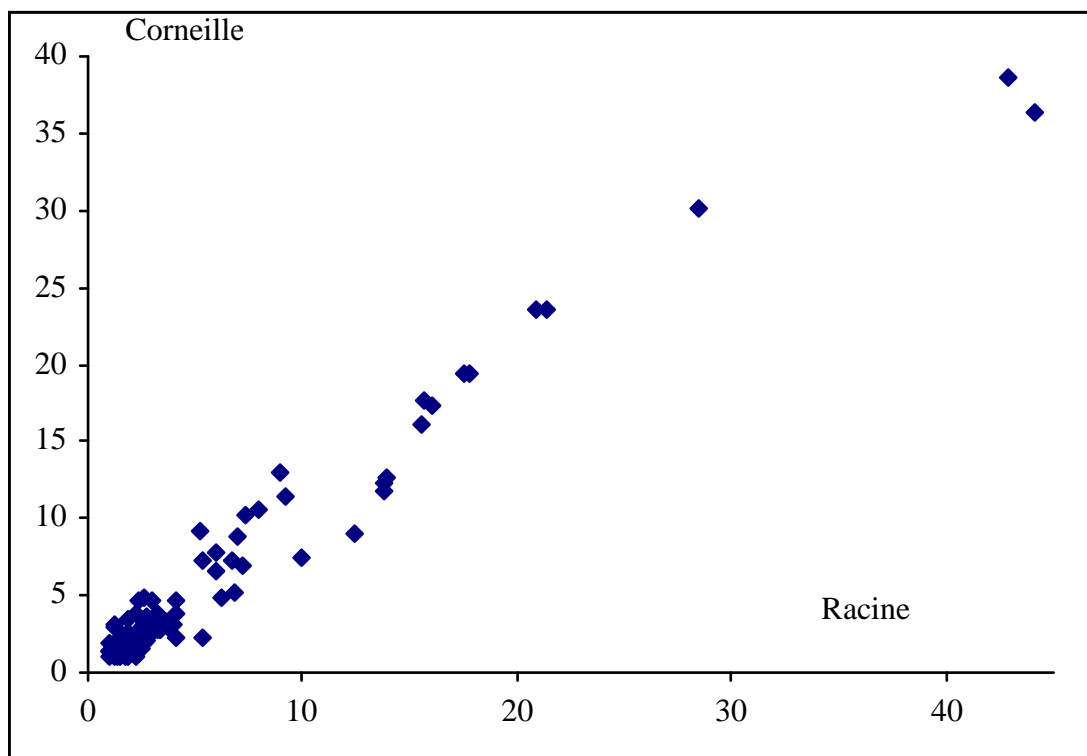


Tableau X. Collocations de "madame" dans les corpus Molière écrit par Corneille et Molière non écrit par Corneille



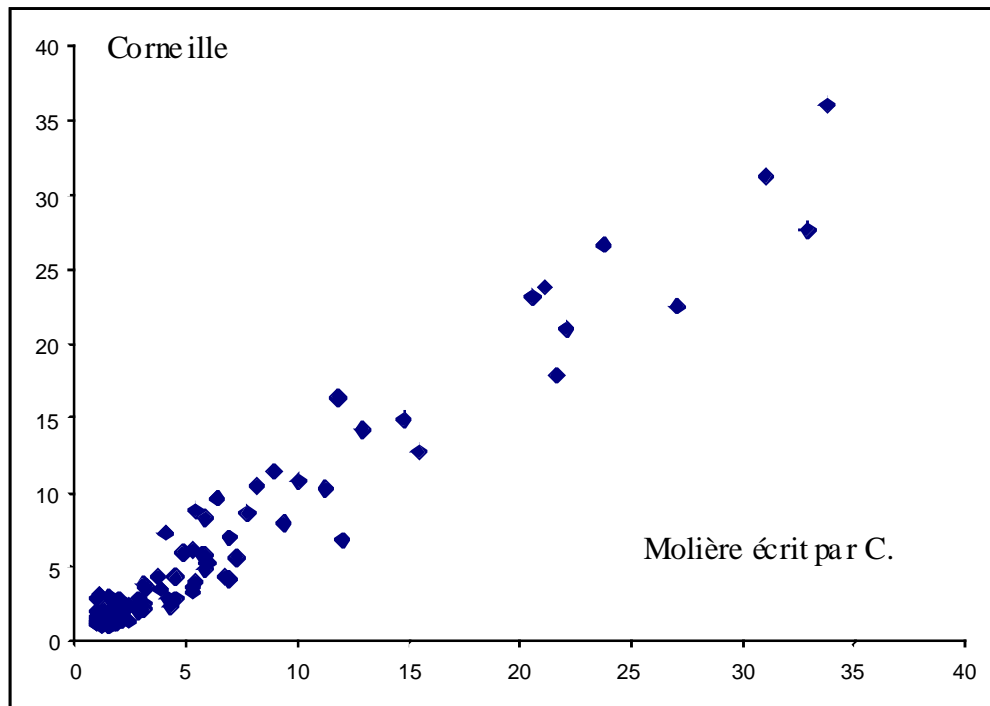
89 vocables (F > 1‰) R = 0.932

Tableau XI. Collocations de "Madame" dans les tragédies de Corneille et de Racine



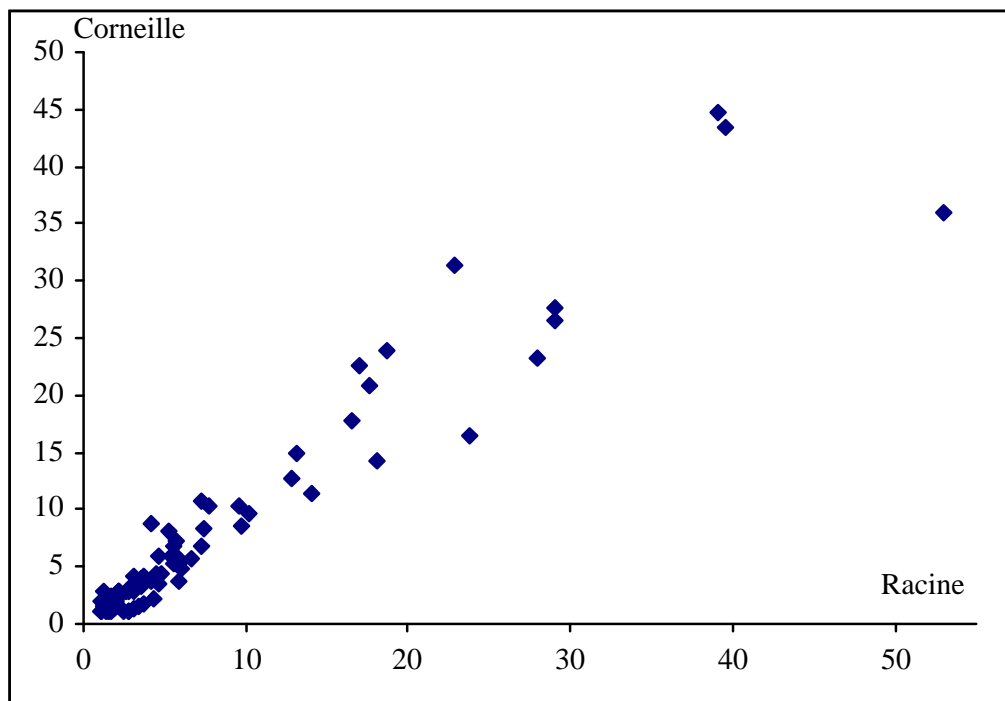
99 vocables (F > 1‰) R = 0.952

Tableau XII. Collocations du verbe "aimer" dans les corpus Corneille et dans les pièces écrites par Corneille pour Molière



92 vocables ($F > 1\text{‰}$) $R = 0.973$

Tableau XIII. Collocations du verbe "aimer" dans les tragédies de Corneille et de Racine



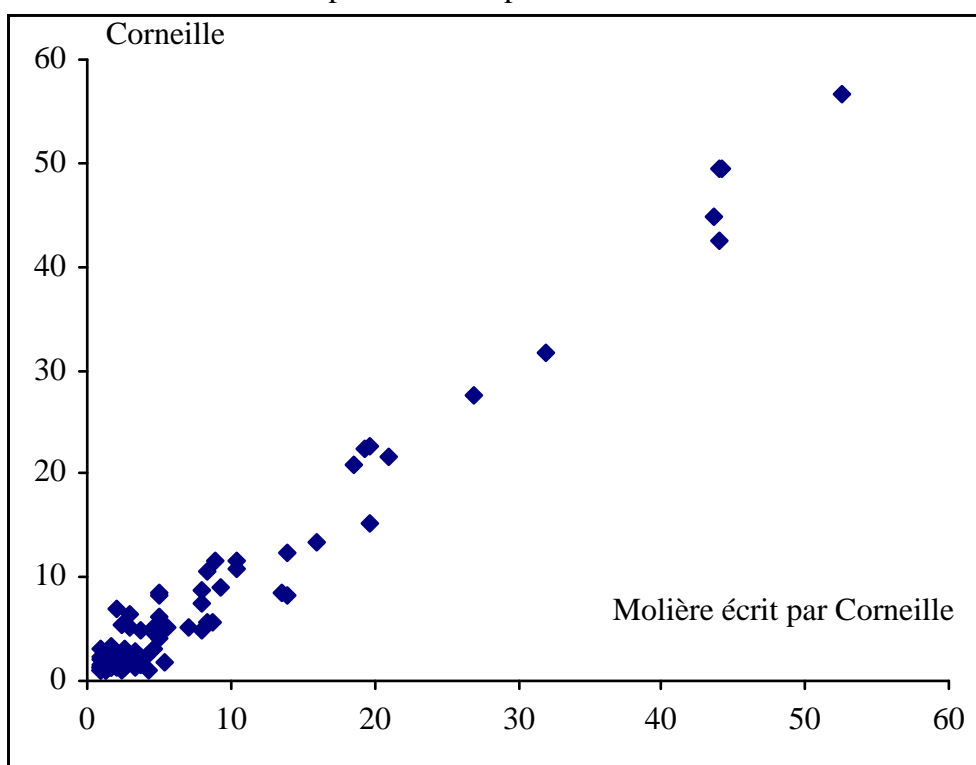
89 vocables communs ($F > 1\text{‰}$; $R = 0.929$)

Les collocations de ce verbe dans "Molière non écrit par Corneille" ont des effectifs trop faible (968 mots) pour se prêter au calcul.

Le vocable "monsieur" pose un problème un peu semblable mais ici, c'est le corpus "Corneille" qui est concerné. En effet, Corneille n'emploie ce vocable que dans les comédies et spécialement dans les dernières qu'il signe sous son nom. Aussi les collocations de "monsieur" ne comportent-elles que 3005 mots chez lui contre 10 358 dans les pièces qu'il a écrites pour Molière et 7 916 dans celles que Corneille n'a pas écrites. On retrouve ici la limitation qui a été signalée pour "amour" au début de cette note.

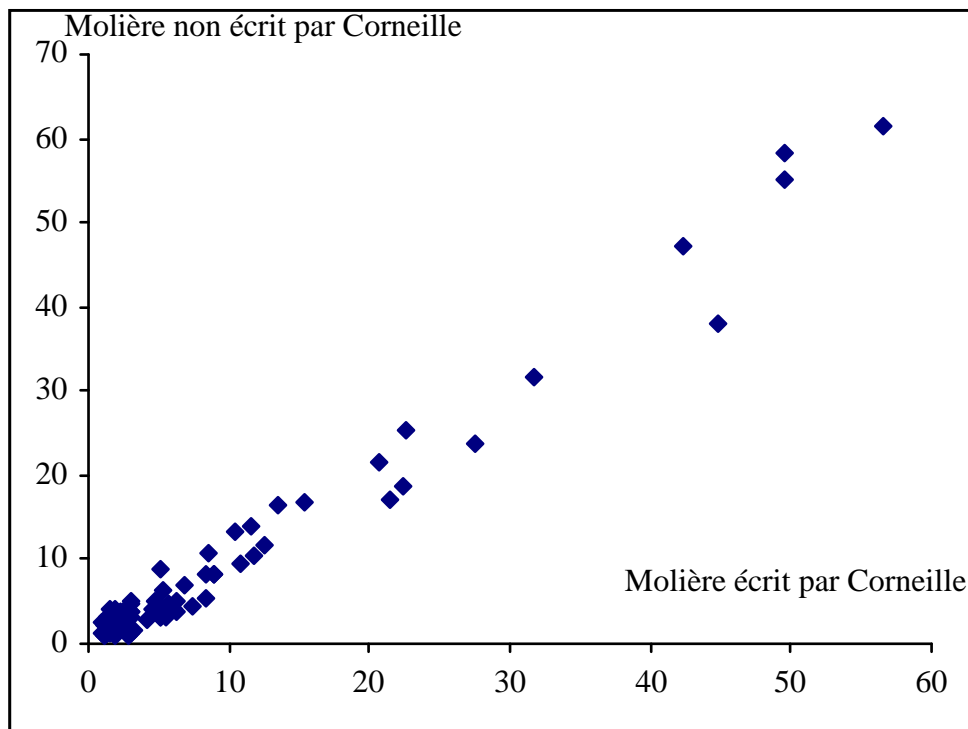
Voici les résultats (Tableau XIV) :

Tableau XIV. Collocations de "monsieur" dans les corpus Corneille et dans les pièces écrites par Corneille pour Molière



92 vocables communs (F>1‰ ; R = 0.970)

Tableau XV. Collocations de "monsieur" dans les corpus Molière écrit par Corneille et Molière non écrit par Corneille



88 vocables communs (F>1‰ ; R = 0.971)

Dans le premier tableau, les "perturbations" se produisent uniquement en bas à gauche, l'alignement est ensuite légèrement meilleur dans le tableau XIV que dans le tableau XV. De plus, le vocabulaire commun est légèrement plus fourni dans le premier que dans le second alors que ce devrait être l'inverse. Il faut en plus rappeler que les deux corpus (Corneille et Molière écrit par Corneille) sont ici séparés de plus de 15 ans alors que "Molière écrit par Corneille" et "Molière non écrit par C." sont contemporains. Si l'on ne peut pas écarter l'idée d'un auteur unique dans le tableau XV, a fortiori n'y en a-t-il qu'un dans la situation décrite par le tableau XIV.

Encore une fois : c'est le cas le plus défavorable.

Conclusions

Si nous avons dû infliger à l'auditoire cette série de chiffres et de tableaux, c'est que nos contradicteurs s'ingénient à contredire toutes nos affirmations, même quand ils savent bien qu'elles sont fondées. L'auditoire a ainsi en main les principaux éléments de la démonstration et cette démonstration est effectuée sur le terrain même qui a été choisi, dans ses "vingt lignes", par M. Viprey: les collocations. Au passage, on pourra juger de la solidité de ses prétendues "réfutations".

Pour toutes les collocations, les ajustements "Corneille vs Molière (écrit par Corneille)" sont nettement meilleurs que ceux obtenus sur "Molière vs Molière" (quand les effectifs permettent le calcul) et sur "Corneille vs Racine".

Si les auteurs étaient différents dans le premier cas ("Corneille vs Molière écrit par Corneille") et semblables dans le second ("Molière écrit par Corneille" vs "Molière non écrit par Corneille"), ce devrait être l'inverse. De plus dans le premier cas, les tragédies voisinent avec des comédies, la création de ces pièces s'étend sur plus de 40 ans ; dans l'autre un seul auteur, seulement des comédies, une période de création beaucoup plus courte... Mais la comparaison avec Racine est aussi essentielle. Elle donne une idée de la distance qui peut séparer les collocations d'un même mot chez deux auteurs différents travaillant sur des thèmes proches, dans un même genre. Ces valeurs sont toujours moins bonnes que celles obtenues entre les pièces que Corneille a publiées sous son nom et celles qu'il a écrites pour Molière.

Cette expérience sur les collocations soulève quelques questions que nous allons évoquer en conclusion :

1. Pourquoi les nuages sont-ils si allongés et les différences relativement faibles ?

Les "contextes" d'un mot sont influencés par trois facteurs, classés ci-dessous par ordre de puissance décroissante :

— le composant syntaxique de la langue. Par exemple, "de" et "le" seront toujours les mots les plus employés quel que soit le pivot examiné, les verbes "être" et "avoir" précéderont toujours "faire", etc. Tout simplement parce que le français est ainsi fait et qu'aucun auteur utilisant cette langue ne peut s'en affranchir ;

— le composant sémantique du français - c'est-à-dire le lexique dont les dictionnaires donnent une image imparfaite - délimite le sens des principaux mots - il ouvre un nombre limité de combinaisons possibles ;

— les choix de l'écrivain au sein de ces combinaisons possibles. Naturellement ces choix sont eux aussi limités par les conventions. Par exemple, dans les collocations de "amour", dans le théâtre du XVIIe, on est certain de retrouver "flamme" !

Naturellement, les deux premiers facteurs pèsent sur tous les locuteurs. Les fréquences d'un même mot usuel chez divers auteurs seront donc toujours fortement corrélées. Les deux variables ne sont pas indépendantes. Le calcul n'indique que la qualité de l'ajustement et non pas une "corrélation" au sens usuel de ce terme.

En définitive, l'auteur est de loin le facteur le plus marginal. Pour voir en pleine lumière ce troisième facteur, il faut neutraliser les deux premiers.

2. Les « collocations » ne peuvent répondre à une question essentielle : Quel sens spécifique chacun des auteurs donne-t-il au mot étudié ?

Pour répondre à cette question, il faut calculer :

— les mots que l'auteur sur-emploie quand il parle d'amour... Dans son esprit, ces vocables sont associés ensemble ;

— les mots qu'il sous-emploie ou qu'il "évite": dans son esprit, ils appartiennent à des univers différents.

Si ce calcul est effectué sur des textes lemmatisés, on peut neutraliser les deux premiers composants. Les choix de l'auteur apparaissent alors en toute clarté. L'ensemble de ces relations d'attraction ou d'exclusion forme "l'univers lexical" du mot.

Ce raisonnement a déjà été présenté en séance plénière aux JADT de Lausanne en 2000 avec l'exemple de... "amour" chez Corneille.

<http://www.cavi.univ-paris3.fr/lexicometrica/article/numero3.html>

Le même calcul a été opéré sur l'univers lexical de "amour" dans les pièces que Corneille a écrites pour Molière : les deux univers se superposent presque parfaitement, ou, plus précisément : le second est inclus dans le premier.

3. S'agit-il d'une "preuve" ?

Ces expériences ne sont qu'une manière de conforter ou de mettre en doute l'attribution d'auteur fondée essentiellement sur la distance intertextuelle et les classifications. Elles ne peuvent s'y substituer car il est impossible de descendre au niveau de chaque pièce et l'on ne peut examiner que quelques mots très fréquents.

Il n'est pas difficile de comprendre que cette expérience est un peu redondante par rapport à la distance intertextuelle. Si la distance entre deux textes est faible, c'est que les mêmes mots sont employés dans ces deux textes avec des fréquences assez proches. Les "collocations" ne font qu'enregistrer cette même similarité pour des segments déterminés de textes. Il est donc logique que l'on retrouve dans le détail, ce que l'on a mesuré globalement à l'étape précédente.

En définitive, il s'agit d'une simple confirmation.

Quand tous ces indices convergent dans la même direction, c'est qu'il y a un seul auteur. Pour infirmer cette conclusion, il faudrait trouver au moins un cas où ces indices, associés à des distances intertextuelles très faibles, se retrouvent chez deux auteurs dont on est certain qu'ils sont différents.

Depuis deux ans, ce défi est lancé à nos contradicteurs : nous attendons avec sérénité.

II. Distance intertextuelle et taille des textes

Dans sa première intervention, mon contradicteur a répété deux contre-vérités qu'il a proférées à plusieurs reprises dans des textes mis en ligne sur sa page personnelle :

— nous aurions affirmé que notre échelle des distances, utilisée pour l'attribution d'auteur, aurait une portée universelle quelle que soit la taille des textes. C'est faux. Dès l'origine, nous avons indiqué que cette échelle est étalonnée pour 10 000 mots ;

— notre calcul sur Corneille et Molière aurait été biaisé par l'influence de la taille sur la distance intertextuelle.

Arrêtons-nous sur cette dernière affirmation. En théorie, les différences de taille entre les textes sont totalement neutralisées puisque, aux arrondis près, notre calcul ramène le plus grand des deux textes à la taille du plus petit. Cependant, dès notre première publication, nous avons signalé que, lorsque les différences de taille sont très importantes, les plus longs textes ont un certain avantage par rapport aux autres, avantage qui tient à l'étendue de leur vocabulaire, à la "spécialisation lexicale" — phénomène que Pierre Hubert et moi-même avons mis en lumière, il y a vingt ans — et à la déformation de la distribution des fréquences. Dans une collection de textes de tailles très différentes, ce sont pratiquement toujours les textes les plus longs qui se trouvent au centre de la distribution et les plus courts qui sont les plus décalés. C'est pourquoi nous avons recommandé de ne pas appliquer le calcul sur des textes trop courts (en tous cas, pas moins de 1.000 mots) et de s'en tenir à une échelle inférieure à 1:10 dans les différences de taille.

Dans cette première publication, nous avons indiqué que ces limitations ne concernaient pas l'expérience sur Corneille et Molière. D'une part, le plus petit texte de Molière écrit par Corneille (*Mélicerte*) dépasse 5 000 mots (à ce niveau, les perturbations dues aux basses fréquences sont négligeables) et la plus longue pièce (*L'Avare*) en compte 21 000. D'autre part, nous avons signalé que plusieurs tests avaient permis de contrôler cette absence d'influence des différences de taille sur les résultats de l'expérience.

Revenons donc sur ce dernier point.

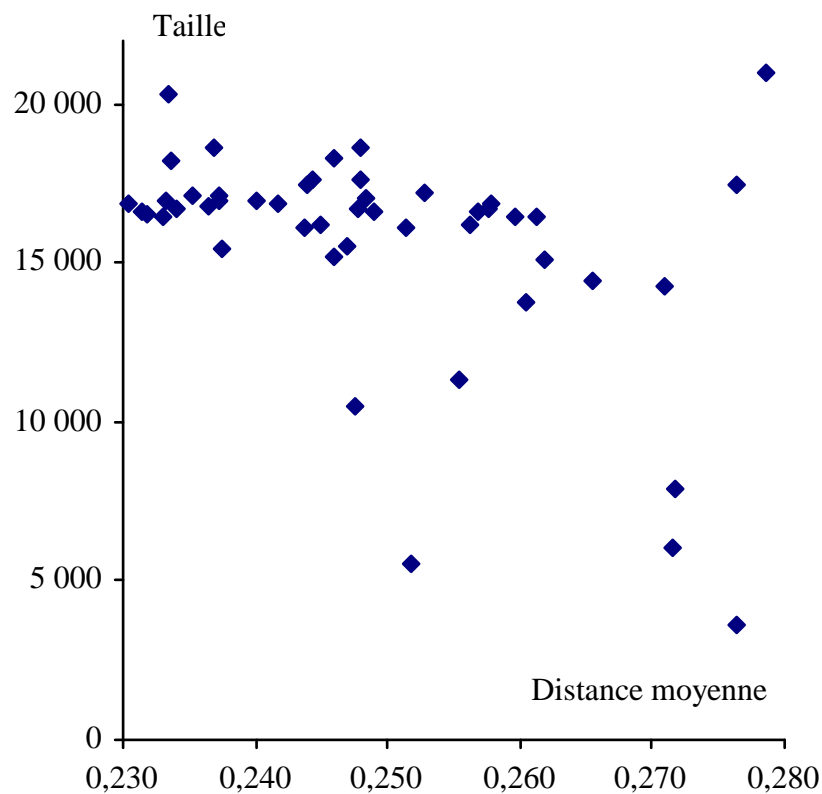
On peut vérifier cette absence de biais de mille manières. La plus simple est la suivante. Considérons les 49 pièces écrites par Corneille sous son nom et sous celui de Molière. Chacune de ces pièces est séparée de toutes les autres par 48 distances : calculons la moyenne de ces 48 valeurs et comparons la avec la taille du texte considéré. Si mon contradicteur a raison, distances moyennes et tailles doivent être corrélées.

Le tableau ci-dessous récapitule l'expérience (les pièces sont rangées par distance moyenne croissante).

Il n'y a aucune liaison entre les deux colonnes. On pourra le vérifier visuellement sur le graphique ci-dessous. Chaque pièce est matérialisée par un point dont l'abscisse est sa distance moyenne à toutes les autres et l'ordonnée sa taille en mots. La forme du nuage montre que les deux variables sont indépendantes. Tout juste notera-t-on à gauche, un noyau plus dense formé des tragédies écrites par Corneille dans sa maturité auxquelles se trouve

mêlées un certain nombre de pièces écrites à la même époque pour Molière : le Dépit amoureux, le Tartuffe, l'Ecole des maris, Dom Garcie...

Pièce	Distance moyenne	Taille (mots)
Sertorius	0,230	17 675
Sophonisbe	0,230	16 858
Pulchérie	0,231	16 630
Suréna	0,232	16 545
Polyeucte	0,233	16 472
Othon	0,233	16 971
Toison d'or	0,234	20 343
Agésilas	0,234	18 227
Tite et Bérénice	0,234	16 697
Théodore	0,235	17 121
Attila	0,237	16 789
Oedipe	0,237	18 618
DonSanche	0,237	16 947
Illusion comique	0,237	15 428
Pertharite	0,237	17 121
Nicomède	0,240	16 923
Rodogune	0,242	16 842
Galerie	0,244	16 140
Héraclius	0,244	17 433
Suite du menteur	0,244	17 675
<i>Dépit amoureux</i>	0,245	16 243
Suivante	0,246	15 160
<i>Tartuffe</i>	0,246	18 272
Andromède	0,247	15 514
<i>Ecole des Maris</i>	0,248	10 536
Mélite	0,248	16 690
<i>Etourdi</i>	0,248	18 674
Veuve	0,248	17 661
<i>Dom Garcie</i>	0,248	17 049
Menteur	0,249	16 653
Cinna	0,251	16 126
<i>Mélicerte</i>	0,252	5 540
<i>Misanthrope</i>	0,253	17 182
<i>Elide</i>	0,255	11 333
<i>Psyché</i>	0,256	16 182
<i>Ecole des Femmes</i>	0,257	16 625
Cid	0,258	16 677
<i>Femmes Savantes</i>	0,258	16 865
Horace	0,260	16 482
Place Royale	0,260	13 801
Pompée	0,261	16 492
<i>Amphitryon</i>	0,262	15 117
Clitandre	0,266	14 402
Médée	0,271	14 269
<i>Sganarelle</i>	0,272	6 042
<i>Fâcheux</i>	0,272	7 922
Comédie des Tuileries	0,276	3 627
<i>Dom Juan</i>	0,276	17 454
<i>Avare</i>	0,279	21 033



Notre calcul sur Corneille et Molière n'est pas biaisé par les différences de taille entre les textes.

Puisque mon contradicteur a affirmé le contraire publiquement, mettons les choses au point :

— ce genre de détails ne peut trouver place dans les articles de revue où l'on privilégie légitimement les conclusions. A propos de tests standards comme ceux-ci, où le risque d'erreur est faible, on fait plutôt confiance aux auteurs... ;

— il n'est pas du tout choquant de vouloir vérifier ce genre de choses. Au contraire, les discussions sur les tests, les procédures ou les unités de mesure font souvent progresser la connaissance plus que les débats théoriques fumeux, à condition toutefois que ces discussions soient menées de bonne foi. Il suffit de demander à l'auteur les explications nécessaires, de les discuter si elles ne paraissent pas satisfaisantes, voire de publier les désaccords s'il en persiste à l'issue de la confrontation ;

— évidemment, mon contradicteur n'a pas du tout procédé ainsi. On pourra le vérifier en lisant, sur son site internet, les textes qu'il a mis en circulation contre nous : les coups de force y côtoient les artefacts et les raisonnements obscurs...

III. Réponse à J.-M. Viprey

"Analyse séquencée de la micro-distribution lexicale"

(in G. Purnelle, C. Fairon, A. Dister (eds), Le poids des mots, Vol 2, p 1165-1176)

Cette communication contient :

1. Deux erreurs et une manipulation

Sous le titre 3 (p. 1171), on nous prêle une méthode dépassée depuis longtemps : les "collocations", c'est-à-dire le relevé des mots entourant le pivot étudié (ici "amour" qui est le substantif le plus fréquent dans le théâtre du XVIIe). Il manque le calcul nécessaire pour répondre à une question simple : quand il parle d'amour, quels mots l'auteur préfère-t-il ? Et quels sont ceux qu'il évite ? Notre méthode répond à cette question et améliore considérablement les "collocations". Cette méthode a été appliquée, en 1990, sur le vocabulaire de Mitterrand ; P. Hubert et moi-même l'avons présentée aux JADT de Rome en 1995 ; elle a été illustrée à Lausanne en 2000 — avec l'exemple du substantif "amour" chez Corneille — et à Saint-Malo en 2002 (ces communications sont en ligne sur le site de la revue Lexicometrica). Par quel étrange aveuglement aurions-nous abandonné cet outil performant ?

Dans le paragraphe suivant, on prétend examiner «les pivots lexicaux mentionnés par D. Labbé (...) les vocables "amour" et "aimer" (cumulés)». Nous n'avons jamais amalgamé ces deux mots dont les distributions sont très différentes. Un peu plus loin, on répète l'opération avec "père"- "fils". Cela n'a rien à voir avec notre recherche.

NOTRE CLASSEMENT DES PIÈCES DE MOLIÈRE N'EST ABSOLUMENT PAS RESPECTÉ. Nous avons isolé TROIS groupes de pièces (tableau à la fin de cette note) :

- (1) celles qui sont attribuées avec certitude à Corneille ("Molière" écrit par Corneille) ;
- (2) celles qui ne sont certainement pas de lui ("Molière" non-écrit par Corneille) ;
- (3) les "non classées" dont certaines contiennent une proportion importante de passages écrits par Corneille. C'est le cas notamment du *Bourgeois gentilhomme* qui — en dehors des passages contenant du faux turc — est de la main de Corneille ou du *Malade imaginaire* (A ce stade de la recherche, nous avons refusé de découper les pièces).

On prétend que nous n'attribuons pas le troisième groupe à Corneille et on l'amalgame au second. C'est une MANIPULATION.

Pourquoi cette manipulation ? Fallait-il dissimuler que, dans les graphiques 6 et 7, à part les *Précieuses*, les pièces de la catégorie (2) se démarquent bien des autres (catégorie 1 et 3) ? Fallait-il aussi améliorer le calcul de "corrélation" ?

2. Un curieux calcul de corrélation

Il est impossible de savoir sur quoi porte le calcul de corrélation présenté p. 1173-1174. On prétend avoir rangé, par ordre décroissant de fréquence, le vocabulaire des phrases contenant le substantif "amour" (et le verbe aimer !). Mais le tableau n'est pas reproduit et une seule indication est donnée sur son contenu (p 1173) :

"Faire occupe toujours et partout le premier rang et *voir* ne descend jamais au-delà du 4e".

Dans un texte en français, les mots les plus fréquents sont : "de", "le", "à", "et", "que" (conjonction et pronom), suivis par quelques pronoms personnels, puis par les verbes "être" et "avoir" (même en excluant les auxiliaires)... le verbe "faire" se classe, au mieux, aux alentours de la vingtième position (et en 3e position des verbes). Nous avons donc voulu vérifier : en annexe de cette note, on lira le tableau des "collocations" de "amour" relevées dans les différents groupes de pièces. Les phrases contenant "amour" n'échappent pas à la

règle générale : "faire" est dans les profondeurs du classement ; quant à "voir", il figure beaucoup plus bas encore (il est d'ailleurs quasiment absent des collocations de "amour" dans les pièces que n'a pas écrites Corneille).

Un grand nombre de mots ont donc été éliminés sans que le lecteur en soit prévenu. Ils gênaient la "démonstration" ?

Aucune explication n'est fournie sur les raisons qui justifient l'utilisation du coefficient de "corrélation des rangs" ("de Spearman") et beaucoup de choses sont passées sous silence :

— quelle est la "norme" utilisée pour arrêter la liste : le corpus entier ou telle ou telle partie choisie pour mieux "démontrer" une thèse personnelle ?

— comment a-t-on classé les ex-aequo (en gras dans notre tableau en annexe). Cette question est cruciale pour les pièces non écrites par Corneille où il y a beaucoup d'ex-aequo.

— qu'a-t-on fait des "absents" et de ceux qui "sortent de la liste par le bas" (en italiques dans le tableau) ?

Pourquoi avoir passé sous silence ces difficultés caractéristiques du coefficient de Spearman et bien connues des statisticiens ?

Puisque ce coefficient est une approximation — le rang ne prend pas en compte la taille des intervalles séparant les individus étudiés — pourquoi le tableau 8 (au bas de la p. 1173) donne-t-il les résultats avec quatre décimales ? Les tables de Spearman n'en utilisent que deux et la prudence est recommandée quant à la seconde.

Enfin et surtout, on garde le silence sur les effectifs entrant dans le calcul (il est simplement dit dans le titre du tableau 8 : "les plus forts cooccurents").

UNE CORRÉLATION SANS INDICATION DU NOMBRE DE DEGRÉS DE LIBERTÉ...

Cette énormité est publiée dans les actes d'un congrès "scientifique"...

Si on cache cette information cruciale, n'est-ce pas parce qu'on est descendu plus ou moins bas dans le tableau pour "embellir" les résultats ?

Enfin, l'annexe 1 montre que, les mots étant rangés par ordre de fréquence décroissante, le dernier de la liste pèse vingt fois moins que le premier. Dans le coefficient de Spearman, chaque individu rangé compte pour un... Il n'était donc pas applicable au cas présent.

3/. Deux confirmations de nos conclusions

Dans un cas semblable, un graphique s'impose : le nuage de points (Tableau à la fin de cette note). Tout mot, dont les fréquences relatives sont supérieures au seuil (1‰) dans les deux corpus comparés, est représenté par un point dont les coordonnées sont ces deux fréquences. Outre que les collocations de "amour" dans "Corneille" et dans "Molière" écrit par Corneille partagent plus de mots, le nuage de points est bien aligné sur la première bissectrice. A l'inverse, ces mêmes collocations dans "Molière" écrit par Corneille et dans "Molière" non-écrit par Corneille partagent moins de mots et leurs fréquences sont significativement différentes comme l'indique la dispersion du nuage. Si mon contradicteur avait raison, ce devrait être l'inverse d'autant plus que le corpus Corneille contient aussi des tragédies et s'étend sur plus de 40 ans alors que les deux corpus "Molière" ne comportent que des comédies et sont contemporains :

Les collocations de "amour" confirment que Corneille est l'auteur des pièces de "Molière" que nous lui attribuons et que quelqu'un d'autre a écrit les 9 que nous ne lui attribuons pas.

Dans les actes, les graphiques 6 et 7 (p. 1172-1173) confirment cette conclusion :

— sur le premier axe factoriel (le plus significatif), les comédies de Corneille — notamment *l'Illusion comique* (1636), *Mélite* (1630) et les deux *Menteurs* (1642-1643) — figurent au milieu des premières pièces de Molière alors qu'elles leur sont antérieures de 20 à 30 ans ! Pour se tirer d'affaire, on a mis en gris des pièces que nous attribuons à Corneille et on affirme que la "diachronie" (la chronologie) est "essentielle" alors que la communication ne date aucune pièce ! Seule l'année 1662 est mentionnée comme une rupture dans l'œuvre de Molière. C'est l'année où Corneille s'installe à Paris : sa collaboration avec Molière va donner naissance à une suite de chefs-d'œuvre...

— *Dom Garcie de Navarre* (1661) — comédie en vers de "Molière" — se retrouve au milieu des tragédies de Corneille. On nous dit que c'est parce qu'il s'agit d'une "comédie héroïque". Sur le graphique 6, les pièces de Corneille qui entourent *Dom Garcie* sont : *Pompée* (1642), *Pulchérie* (1672), *Nicomède* (1651), *Othon* (1664) *Agésilas* (1666), etc. Ces tragédies n'étaient donc que des "comédies héroïques" déguisées ?

Remarquons enfin que les petits carrés, en bas à droite, indiquent que l'expérience est très pauvre : le premier plan factoriel rend compte de 15% de l'inertie totale du nuage...

4/ Une curieuse conception de la recherche

Les premières pages de cette communication n'apportent aucune lumière sur les calculs effectués — notamment le traitement des grandes matrices creuses où l'essentiel de l'information est concentrée dans quelques cases. La bibliographie renvoie seulement aux deux manuels de C. Muller (vieux de 30 et 36 ans), qui n'abordent pas l'analyse factorielle. Pourquoi devrions-nous faire confiance à une "boîte noire", au motif que les sorties sont en couleur et qu'il faut passer à la caisse ?

A l'inverse, nos données, calculs et algorithmes sont dans le domaine public. Nous n'avons rien à cacher et rien à vendre.

Nous avons manifestement deux conceptions opposées de la recherche. En veut-on une preuve supplémentaire ? Quand on utilise les données fournies par un autre chercheur, il est de tradition de lui communiquer les résultats avant publication. Notre étrange contradicteur utilise les fichiers que nous lui avons gracieusement remis. Pourquoi avons-nous dû attendre l'ouverture du congrès pour découvrir son mauvais pamphlet ?

En conclusion, il reste trois questions à poser :

— pour figurer dans les Actes du congrès, ce texte a été accepté par un comité "scientifique", après avoir été "expertisé" par deux de ses membres. Ces personnes auront-elles l'amabilité d'exposer les raisons qui ont justifié la sélection de ce texte ?

— dans l'édition scientifique, quand un chercheur est mis en cause par un texte, il est de tradition de le lui communiquer avant publication pour lui donner la possibilité de présenter ses objections. Pourquoi les organisateurs du congrès n'ont pas procédé ainsi à mon égard ?

— en octobre dernier, quand ces organisateurs ont pris contact, pour l'organisation de cette table ronde, il leur a été indiqué qu'un débat scientifique ne se conçoit pas sans textes préparatoires. Ils ont décliné cette proposition alors qu'ils avaient en main les communications de mes trois contradicteurs. Résultat ? 30 pages et 3 fois 25 minutes pour attaquer un travail présenté en 20 lignes et dix minutes... Enfin, je n'ai disposé que de quelques heures pour déchiffrer cette prose obscure et rédiger cette réponse qui, elle, ne figurera pas dans les Actes du congrès, alors que les erreurs, les manipulations et les anachronismes s'y étalent avec la complicité des organisateurs.

Tableau I Les 50 premières collocations du substantif "amour" dans les trois sous-corpus*

Rang	Corpus total Vocabulaire	Fréquence (‰)	Corneille entier Rang F(‰)	Molière écrit par C. Rang F(‰)	M. non- écrit par C Rang F(‰)
1	le (det)	(60,5)	1 (59,2)	1 (62,6)	1 (70,8)
2	de (pré)	(56,0)	2 (53,9)	2 (60,0)	2 (70,1)
3	je (pro)	(35,8)	3 (35,6)	3 (34,2)	3 (42,6)
4	et (cj)	(28,6)	4 (28,3)	4 (29,5)	5 (33,0)
5	à (pré)	(24,1)	5 (23,8)	5 (26,3)	9 (19,9)
6	que (cj)	(22,7)	6 (22,9)	6 (23,7)	7 (23,4)
7	un (det)	(20,4)	7 (20,5)	8 (22,3)	14 (12,4)
8	être (v)	(19,8)	8 (18,9)	7 (23,0)	6 (24,7)
9	vous (pro)	(19,0)	12 (18,1)	9 (20,2)	4 (35,7)
10	mon (det)	(18,1)	10 (18,6)	10 (18,0)	13 (13,1)
11	il (pro)	(18,1)	9 (18,9)	13 (15,0)	14 (12,4)
12	ne (adv)	(18,1)	11 (18,5)	11 (16,7)	10 (19,2)
13	avoir (v)	(17,3)	13 (17,1)	12 (16,4)	10 (19,2)
14	son (det)	(13,5)	14 (14,3)	16 (12,2)	26 (6,2)
15	que (pro)	(12,9)	15 (12,2)	15 (12,9)	8 (22,0)
16	ce (det)	(12,0)	16 (12,0)	14 (13,3)	14 (12,4)
17	pour (pré)	(11,7)	17 (11,5)	18 (10,9)	12 (19,2)
18	le (pro)	(10,1)	18 (11,3)	26 (6,8)	26 (6,2)
19	faire (v)	(9,9)	19 (9,5)	19 (10,4)	14 (12,4)
20	en (pré)	(8,7)	20 (9,2)	23 (7,7)	22 (7,6)
21	qui (pro)	(8,7)	21 (8,4)	20 (9,7)	18 (10,3)
22	votre (det)	(8,4)	22 (7,7)	17 (11,2)	33 (4,8)
23	ce (pro)	(7,8)	22 (7,7)	21 (8,1)	26 (6,2)
24	se (pro)	(6,8)	26 (6,6)	21 (8,1)	33 (4,8)
25	en (pro)	(6,7)	24 (7,3)	31 (5,3)	33 (4,8)
26	plus (adv)	(6,4)	25 (7,3)	38 (3,8)	48 (3,4)
27	pouvoir (v)	(6,0)	29 (5,8)	27 (6,7)	26 (6,2)
28	voir (v)	(5,9)	28 (6,1)	29 (5,6)	78 (1,2)
29	si (cj)	(5,9)	27 (6,4)	32 (4,6)	38 (4,1)
30	mais (cj)	(5,4)	30 (5,3)	30 (5,4)	26 (6,2)
31	on (pro)	(5,0)	35 (4,1)	24 (7,3)	38 (4,1)
32	tout (det)	(5,0)	34 (4,2)	27 (6,7)	20 (8,2)
33	vouloir (v)	(4,7)	32 (4,6)	34 (4,5)	20 (8,2)
34	coeur (n m)	(4,6)	35 (4,1)	25 (7,0)	78 (1,2)
35	tu (pro)	(4,2)	31 (4,7)	54 (2,7)	—
36	lui (pro)	(4,0)	33 (4,2)	50 (3,1)	78 (1,2)
37	par (pré)	(4,0)	37 (4,0)	35 (4,4)	26 (6,2)
38	moi (pro)	(3,8)	39 (3,8)	38 (3,8)	32 (5,5)
39	dans (pré)	(3,7)	44 (3,4)	33 (4,6)	51 (2,7)
40	pas (adv)	(3,6)	43 (3,5)	46 (3,3)	22 (7,6)
41	nous (pro)	(3,6)	46 (3,3)	36 (4,0)	19 (9,6)
42	aimer (v)	(3,5)	40 (3,6)	38 (3,8)	102 (0,7)
43	savoir (v)	(3,5)	43 (3,5)	38 (3,8)	33 (4,8)
44	sans (pré)	(3,3)	42 (3,5)	48 (3,2)	102 (0,7)
45	si (adv)	(3,3)	40 (3,6)	36 (2,3)	78 (1,2)
46	ton (det)	(3,1)	38 (3,8)	95 (1,2)	—
47	jour (n m)	(3,1)	48 (3,0)	42 (3,7)	102 (0,7)
48	point (adv)	(3,0)	47 (3,8)	48 (3,2)	48 (3,4)
49	bien (adv)	(2,9)	53 (2,5)	60 (3,9)	38 (4,1)
50	tout (pro)	(2,8)	52 (2,5)	43 (3,6)	78 (1,2)

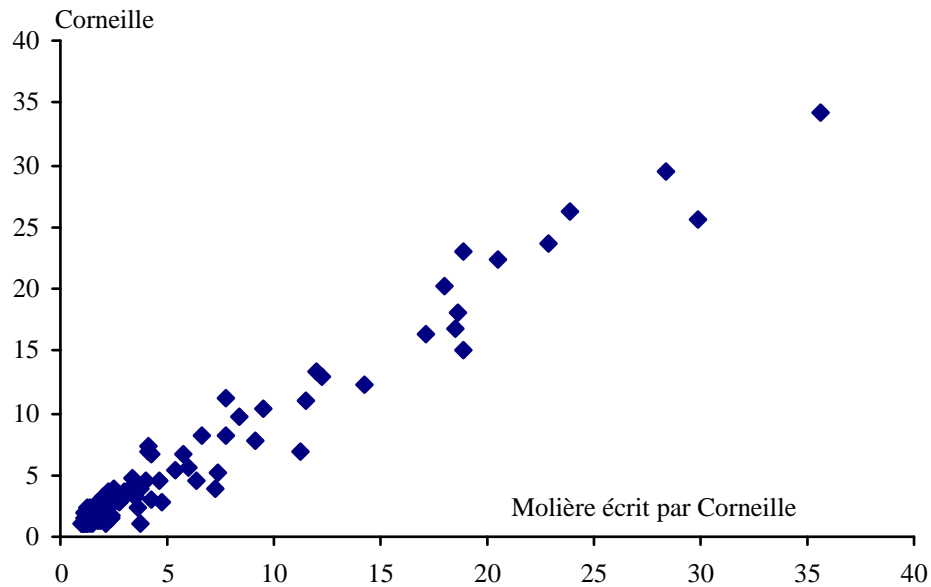
(en gras les ex-aequo (premier rang), en italiques : les rangs supérieurs à 50 ; — absent des collocations.)
 • "Molière" écrit par Corneille : L'étourdi (1658), Le dépit amoureux (1658), Sganarelle (1660), Dom Garcie de Navarre (1661), l'Ecole des maris (1661), Les Fâcheux (1661), l'Ecole des femmes (1662),

La Princesse d'Elide (1664), Le Tartuffe (1664), Dom Juan (1665), Le Misanthrope (1666), Méricerte (1666), Amphitryon (1668), l'Avare (1668), Psyché (1671), Les femmes savantes (1672).

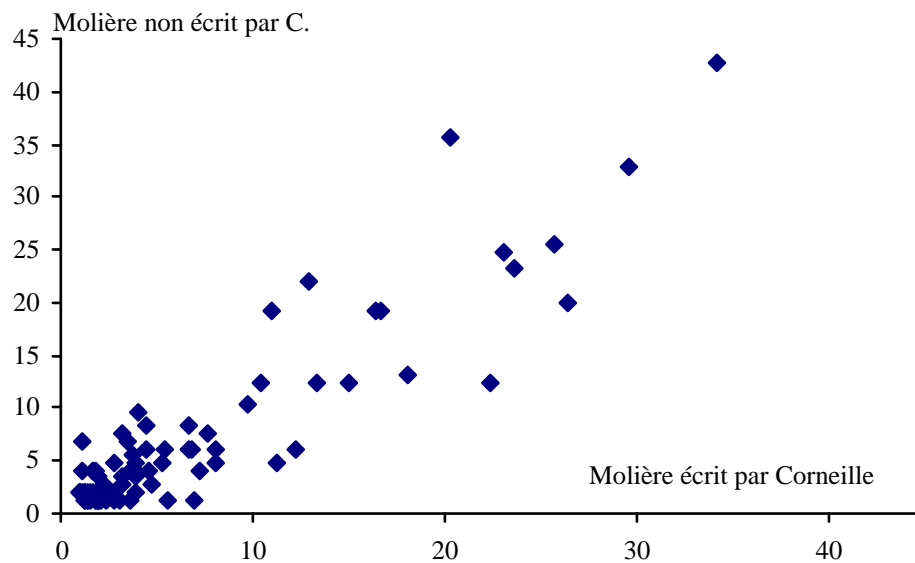
• "Molière" non-écrit par Corneille : La Jalousie du barbouillé (avant 1658), Le Médecin volant (avant 1658), les Précieuses ridicules (1659), Critique de l'Ecole des femmes (1663), l'Impromptu de Versailles (1663), le Mariage forcé (1664), l'Amour médecin (1665), le Médecin malgré lui (1666), la Comtesse d'Escarbagnas (1671).

Tableau II

Nuage des 96 vocables de fréquence supérieure à 1‰ dans les collocations de "amour" communs aux corpus "Corneille total" et "Molière" écrit Corneille" ($R = .980$).



Nuage des 80 vocables de fréquence supérieure à 1‰ communs aux collocations de "amour" dans les corpus "Molière" écrit par Corneille et "Molière" non écrit par Corneille ($R = .922$).



IV. Réponse à V. Beaudouin et F. Yvon "Contribution de la métrique à la stylométrie"

(in G. Purnelle, C. Fairon, A. Dister (eds), Le poids des mots, vol 1, p. 107-117)

Saluons l'ambition de ce travail : attribuer "avec une très grande précision" chaque vers du théâtre classique à son auteur et à son genre (p 115) ; nos travaux sont plus modestes et nous ne sommes jamais descendu en dessous de 1000 mots (soit plus d'une centaine de vers)...

Toutefois, le lecteur sera surpris : cette communication omet tout repère chronologique. La création des pièces étudiées s'étalant sur plus de 60 ans — Mélite (Corneille, 1630) ; Athalie (Racine, 1691) — une comparaison sérieuse ne doit-elle pas prendre en compte leur succession dans le temps ? Par exemple, la légende du tableau 4 indique que 10% des vers des comédies de Corneille — de Mélite (1630) au Menteur (1642) — sont attribués à Racine (né en 1639). Un peu plus loin, on apprend que le programme attribue à ce même Racine un nombre important de vers de Cinna, créé par Corneille en 1641 : Racine avait deux ans ! Dans la même page, on écrit qu'un "fort pourcentage des vers du Menteur" (Corneille, 1642) "sont attribués à Molière" (sur la figure 2, on voit que c'est 30%). Molière, né en 1622, n'était pas encore monté sur les planches et n'avait pas écrit une ligne...

Le bon sens aurait voulu que l'on interdise au "métromètre" de violer la chronologie. Pourquoi ne pas l'avoir fait ? Cela aurait révélé le poids prépondérant de Corneille dans les œuvres de "Molière". C'est d'ailleurs aussi pour cela que l'expérience décrite p. 115-116 ne porte sur aucune pièce de "Molière" (Psyché n'est pas de "Molière") : Dom Garcie, le Tartuffe ou le Misanthrope sont indiscutables ?

Revenons au cas de Cinna : 3 vers sur 10 seraient de Racine et, d'après la figure 2, un autre dixième serait de "Molière". Autrement dit, sur une pièce très caractéristique de Corneille, largement antérieure aux œuvres de Racine et de Molière, le métromètre se trompe 4 fois sur 10. Ce programme est peut-être intéressant pour détecter certains procédés utilisés dans une pièce, mais pourquoi prétendre, contre toute évidence, qu'il pourrait en identifier l'auteur ?

Il y a plus extraordinaire encore : selon cette communication, Molière aurait écrit 60 à 80% des Plaideurs (Racine, 1668). Concernant cette pièce, trois choses sont certaines : les alexandrins sont techniquement impeccables, ils sont de Racine et ils sont postérieurs à la plupart des comédies en vers publiées sous le nom de Molière. Cela montre simplement que Racine s'est massivement inspiré des comédies antérieures écrites par Corneille. Conclusion :

L'écrivain a un plein contrôle sur les techniques de versification, il peut en changer à sa guise, LA PROSODIE — TOUT COMME LA RIMAISON — N'ONT AUCUNE UTILITÉ POUR L'ATTRIBUTION D'AUTEUR.

Si Racine peut "faire du Molière" (sur le plan prosodique), pourquoi Corneille ne le pourrait-il pas ? D'autant plus que, 15 ans avant la première comédie écrite pour Molière, il avait déjà fait une bonne partie du chemin. En effet, la figure 2 confirme ce que nous avons écrit : les Menteurs (1642) sont plus proches des comédies en vers de "Molière" (largement postérieures) que des autres comédies de Corneille (immédiatement antérieures) ou de ses tragédies... Les deux Menteurs sont les sœurs aînées des grandes pièces en vers de "Molière", tout simplement parce que l'auteur est le même et qu'il s'agit de Corneille.

Ce n'est pas la première fois que des chercheurs rejettent la pépite à la rivière et gardent la gangue de boue. Dans un congrès, le comité "scientifique" est là pour empêcher qu'on imprime de pareilles aberrations. Si celui des JADT 2004 a failli à cette mission, c'est que certains en son sein voulaient alimenter la polémique contre nous. Ils n'ont pas compris que ce papier manque son but et qu'il compromet la réputation de l'analyse des données textuelles.

Annexes

I.

Les vingt lignes échangées par les participants à la table ronde de Louvain-la-Neuve

(ces textes figurent à la fin du deuxième tome des Actes du congrès de Louvain)

Les vingt lignes de M. J.-M. Viprey

Les statistiques lexicales sont-elles destinées à fournir des preuves? Ne servent-elles pas plutôt à renforcer ou affaiblir des hypothèses, et surtout à les réorienter ? Une méthode probabiliste ne saurait, à elle seule, prouver qu'un auteur peut être tenu pour responsable de telle ou telle oeuvre particulière.

Corneille et Molière sont intervenus dans un même champ littéraire, à des dates proches, et dans un genre très contraignant, la comédie (voire la comédie en vers). Corneille a mis la main à Psyché. Molière a écrit Dom Garcie de Navarre sous l'influence de la comédie héroïque cornélienne. A mon sens, les résultats obtenus jusqu'ici accèdent plutôt la thèse communément admise : Molière invente peu à peu, et de façon décisive à partir de 1662 (L'Ecole des femmes), une expression théâtrale radicalement nouvelle.

L'analyse lexicale massive, qui ne prend pas en compte les collocations ni a fortiori les configurations plus complexes, me paraît insuffisante pour caractériser une écriture. Par exemple, l'examen des interjections employées exclusivement dans les textes signés par Molière manifeste la référence spécifique de ce dernier à la Commedia dell'Arte (à l'exclusion de toutes les pièces de Corneille), ce qui affaiblit notablement la thèse présentée par D. Labbé.

Les vingt lignes de Dominique Labbé

Les travaux qui sont à l'origine de ce débat sont techniques, transparents et publiés. Nous n'avons pas une thèse, polémique ou pas, sur le sujet, mais une méthodologie dont un résumé de 20 lignes (demandé un peu tardivement) ne peut rendre compte. Nous pensons plus utile de donner quelque références.

Notre travail a été publié dans une revue internationale à comité de lecture :

Labbé Cyril, Labbé Dominique. "Inter-Textual Distance and Authorship Attribution Corneille and Molière". Journal of Quantitative Linguistics. 8-3. December 2001. 213-231.

La distance intertextuelle a été présentée en français (avec une discussion de ses propriétés) :

Labbé Cyril, Labbé Dominique. "La distance intertextuelle". Corpus. 2-2003. 95-118.

La méthode a été validée par plusieurs expériences techniquement irréprochables ; elle est utilisée par plusieurs chercheurs. Des publications sont en cours. Deux articles ont déjà paru :

Merriam Thomas. "Intertextual Distances between Shakespeare Plays, with Special Reference to Henry V (verse)". Journal of Quantitative Linguistics. 9-3. December 2002. P 260-273.

Merriam Thomas. "An Application of Authorship Attribution by Intertextual Distance in English". Corpus. 2. 2003. p 167-182.

Une synthèse concernant l'ensemble du dossier est accessible à :

<http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/LabbeCorneilleMoliere.pdf>

(On trouvera sur cette page une série d'autres documents sur ce thème).

L'Université d'Oxford a mis en ligne toutes les pièces de Corneille, Molière et Racine — normalisées et lemmatisées par nos soins — avec d'autres corpus traités de la même manière, des publications et la plupart de nos logiciels : <http://ota.ahds.ac.uk/2466>.

Nous n'avons rien à vendre et nous n'utilisons aucune boîte noire.

Annexe II.

Les pièces de Corneille.

(Toutes les pièces de Corneille sont en vers)

Corneille	Année de création	Genre	Taille 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	Atilia	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

Psyché	Année de création	Genre	Taille en mots	
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

Le corpus Corneille compte 34 pièces, dont 32 complètes. Il est long de 553 190 "occurrences". Son vocabulaire comporte : 15 535 formes graphiques normalisées 6 258 vocables.

Annexe III

Par qui ont été écrites les pièces de Molière ?

16 comédies sont attribuées à Corneille (par ordre chronologique de création)

Titre	Actes	Genre	Date	Taille (mots)
L'étourdi	5	Vers	1658 ?	18 674
Le Dépit amoureux	5	Vers	1658 ?	16 243
Sganarelle ou le cocu imaginaire	1	Vers	1660	6 042
Dom Garcie de Navarre	5	Vers	1661	17 049
L'Ecole des maris	3	Vers	1661	10 536
Les fâcheux	3	Vers	1661	7 922
L'Ecole des femmes	5	Vers	1662	16 625
La princesse d'Elide	5	Vers et prose	1664	11 333
Le Tartuffe	5	Vers	1664	18 272
Dom Juan	5	Prose	1665	17 454
Le Misanthrope	5	Vers	1666	17 182
Mélicerte	2	Vers	1666	5 540
Amphytrion	3	Vers libres	1668	15 117
L'Avare	5	Prose	1668	21 033
Psyché	5	Vers	1671	16 182
Les Femmes savantes	5	Vers	1672	16 865

9 comédies de Molière n'ont pas été écrites par Corneille
(par ordre chronologique de création)

Titre	Actes	Genre	Date	Taille (mots)
La jalousie du barbouillé	1	Prose	Avt 1658	3 501
Le médecin volant	1	Prose	Avt 1658	3 876
Les précieuses ridicules	1	Prose	1660	6 651
Critique de l'école des femmes	1	Prose	1663	8 613
Impromptu de Versailles	1	Prose	1663	7 170
Le mariage forcé	1	Prose	1664	6 059
L'amour médecin	3	Prose	1665	6 148
Le médecin malgré lui	3	Prose	1666	9 319
La comtesse d'Escarbagnas	1	Prose	1671	5 565

7 pièces sont douteuses (elles peuvent avoir été écrites par Corneille mais elles sont trop éloignées des *Menteurs* ou de *Psyché* pour que l'on puisse conclure avec une certitude raisonnable)

Titre	Actes	Genre	Date	Taille (mots)
Le sicilien ou l'amour peintre	1	Prose	1667	5 375
Georges Dandin	3	Prose	1668	11 009
Monsieur de Pourceaugnac	2	Prose	1669	11 803
Les amants magnifiques	5	Prose	1670	11 983
Le bourgeois gentilhomme	5	Prose	1670	17 136
Les fourberies de Scapin	3	Prose	1671	14 245
Le malade imaginaire	3	Prose	1673	19 920

NB : le *Bourgeois gentilhomme* et le *Malade imaginaire*, bien que plus éloignés des *Menteurs*, peuvent être rattachés à la première catégorie — celles des pièces écrites par Corneille car en enlevant les dernières scènes du *Bourgeois* ou les passages en latin de cuisine du *Malade* (ainsi que les intermèdes en italien), leur distance avec les deux *Menteurs* tombe en dessous de 0.25.

Annexe IV

Compte-rendu de la table-ronde par Benoît Peeters
(document placé en ligne sur le site des Impressions nouvelles)

Corneille et Molière : nouveaux éléments

L'Université de Louvain-la-Neuve (Belgique) a accueilli du 10 au 12 mars 2004, les 7e journées internationales d'analyse des données textuelles (JADT). Ce congrès réunit, une fois tous les deux ans, les meilleurs spécialistes européens de la statistique appliquée aux textes et à la langue. Quelque 200 personnes participaient aux journées de Louvain.

Le jeudi 11 mars en fin d'après midi, le congrès a assisté à une table ronde portant sur l'attribution d'auteur. La question posée aux participants de cette table ronde était la suivante : "La statistique peut-elle apporter une aide dans la recherche de l'auteur d'un texte anonyme ou d'origine douteuse ?"

Plus précisément, il s'agissait de savoir si Corneille pouvait ou non avoir écrit certaines comédies de Molière comme l'affirme Dominique Labbé (de l'Institut d'Etudes Politiques de Grenoble) dans son essai *Corneille dans l'ombre de Molière* (Paris-Bruxelles, les Impressions Nouvelles, 2003).

Ludovic Lebart — professeur à l'Ecole nationale supérieure des télécommunications de Paris et l'un des meilleurs spécialistes mondiaux de l'analyse des données textuelles — a ouvert cette table ronde en rendant hommage à Dominique Labbé à la fois pour ses travaux souvent pionniers, pour la rigueur et la transparence de ses méthodes.

Dominique Labbé a présenté une synthèse de la méthode d'attribution d'auteur développée depuis plusieurs années dans son laboratoire. Cette méthode combine le calcul de la distance entre textes avec plusieurs classifications automatiques. Elle a, jusqu'à maintenant, résisté aux tests les plus sévères. Elle est utilisée par plusieurs équipes à travers le monde et a déjà donné lieu à plusieurs publications scientifiques.

Elle révèle que P. Corneille a écrit l'essentiel des pièces en vers de Molière et d'au moins quatre pièces en prose : *l'Avare*, le *Bourgeois gentilhomme*, *Dom Juan*, le *Malade imaginaire*. Il a enfin été rappelé que ces conclusions sont confirmées par de nombreux indices lexicaux et stylistiques mais aussi historiques.

Jean-Marie Viprey de l'Université de Besançon a exposé ses réticences anciennes contre le travail de Dominique Labbé mais il en a reconnu le sérieux et la rigueur. Il n'a pas contesté les proximités lexicales existant entre les deux oeuvres et il a admis que celles-ci rendaient pour le moins "plausible" l'hypothèse d'une "collaboration" régulière et approfondie entre Corneille et Molière. Il a également reconnu que de nombreux indices littéraires ou historiques pouvaient conforter cette hypothèse. Le dossier est ouvert et le travail doit donc se poursuivre avec comme but principal la compréhension de l'oeuvre.

Valérie Beaudouin (Chercheuse à France Télécom R&D) a remercié D. Labbé pour son travail qui relance l'étude du théâtre classique. Ses conclusions ne sont pas remises en cause par l'analyse statistique des vers des deux auteurs qu'elle a elle-même réalisée. Au contraire, cette analyse montre une proximité remarquable de deux comédies de Corneille (*Le Menteur*

et la Suite du Menteur) avec les grandes pièces de Molière. Cependant, elle s'est demandée si l'on n'attache pas trop d'importance à la notion d'auteur et s'il ne vaudrait pas mieux abandonner cette vision "romantique" pour se centrer sur l'intertextualité.

Hippolyte Wouters (Avocat au barreau de Bruxelles) a résumé l'hypothèse présentée dans son essai Molière ou l'auteur imaginaire ? (Bruxelles, Complexe, 1990). Molière serait une sorte de raison sociale pour un groupe chargé de satisfaire le roi Louis XIV qui voulait offrir à sa cour des spectacles de bonne qualité. Ce groupe animé par Jean-Baptiste Poquelin comprenait une troupe d'acteurs, des musiciens, des ballets... et plusieurs auteurs dont le principal était Corneille. Les délais impartis par le roi étaient généralement très courts, de telle sorte que des vers ou des scènes de transition étaient parfois ajoutées au dernier moment par d'autres que Corneille.

Cette explication a semblé rallier l'auditoire et les participants ont paru tous d'accord pour estimer que les travaux de D. Labbé ouvrent de nouvelles perspectives pour l'analyse du théâtre du XVIIe siècle et pour l'histoire littéraire.

Benoît Peeters - mars 2004

Annexe V

Compte-rendu de l'intervention de M. J.-M. Viprey par lui-même
(Document communiqué à diverses listes de discussion et placé en ligne sur sa page
personnelle de 2004 à 2007)

From: "Jean-Marie Viprey" <jean-marie.viprey@univ-fcomte.fr>

Date sent: Sun, 14 Mar 2004 15:51:04 +0100

Voici la teneur précise de ce que j'ai prononcé lors de la table-ronde organisée sur le thème "Attribution d'oeuvre : le débat Corneille-Molière" avec Valérie Beaudouin, Dominique Labbé, Me Hippolyte Wouters, et Ludovic Lebart dans la fonction de modérateur.

Il ne s'agit pas d'un verbatim, celui-ci n'ayant pas été prévu ; pour ma part, j'ai parlé d'après des notes détaillées tout en essayant de réagir aux interventions.

Mes interventions :

[en écho à l'introduction de Ludovic Lebart] :

Première intervention

(1.1)

Il est tout-à-fait légitime d'interroger les certitudes communes. Bien qu'allant contre ces certitudes, les thèses développées dans le livre d'Hippolyte Wouters ["Molière ou l'auteur imaginaire ?", paru en 1990 aux Editions Complexe] méritent un authentique débat, qui n'a peut-être pas, en ce qui nous concerne, encore eu complètement lieu.

Nous [les divers partenaires des journées JADT, le monde de la statistique lexicale et de l'analyse de textes] avons, par ailleurs, eu tort de ne pas dialoguer plus tôt avec l'article de MM. Labbé [celui paru dans le Journal of Quantitative Linguistics, 2001, vol.3].

La communauté des experts littéraires doit accepter qu'il y ait matière à débat.

Nous [cf. supra] devons en échange être prudents et circonstanciés. Il ne faut donc pas, ce qu'a fait en substance Dominique Labbé, refermer le dossier avant même de l'avoir ouvert. Plus largement que la question du statut de la "preuve" dans notre domaine, c'est le mode d'exposition péremptoire "C'est Corneille et non Molière qui a écrit Dom Juan et un certain nombre d'autres pièces signées de Molière" qui pose problème.

(1.2)

Il faut donc discuter sur les principes.

La question est : "comment l'Analyse des Données Textuelles intervient-elle dans les divers champs disciplinaires ?". Je répondrais : avec prudence, dans le respect de l'expertise de ces champs, dans un esprit de dialogue.

Je ne suis pas statisticien. Je suis un utilisateur expert de la statistique linguistique et lexicale. De ma connaissance du domaine, du dialogue avec les statisticiens, je tire l'enseignement qu'il ne peut exister de "preuves" en statistiques probabilistes. Plutôt que de discuter de vocabulaire (parler de "quasi-preuves", mieux : de "présomptions" serait bienvenu), il faut en

discuter la philosophie : quelle que soit la qualité des "indices", une échelle de validation péremptoire ("au-dessous de 0.20 il est impossible que ce soient deux auteurs différents, au-dessus d'un certain seuil et dans certaines conditions il est impossible que ce soit un seul auteur"), une telle échelle s'expose à être démentie à n'importe quel moment.

OUI, Corneille et Molière sont exceptionnellement proches, mais il faut bien qu'existe le couple d'auteurs le plus « intime », sans que cela empêche en rien qu'ils soient bien deux auteurs distincts.

Deuxième intervention :

(2.1)

[Dans des termes que je ne peux reconstituer exactement pour avoir été dans le feu de l'action, j'approuve l'intervention de Valérie Beaudouin, je remercie Ludovic Lebart et André Salem d'avoir organisé le débat, je me réjouis de ce que la discussion prend une tournure dépassionnée]

Le dossier doit être pleinement ouvert (sinon réouvert) ; nous devrions envisager une publication collective.

(2.2)

Sur la technique proposée par Dominique Labbé (la distance intertextuelle), mes observations [disponibles sur le site <http://laseldi.univ-fcomte.fr>], faites à partir de l'application de l'indice de D.Labbé, me conduisent à supposer un biais dans cet indice : il croît tendanciellement à mesure que les textes analysés s'allongent. N'étant pas moi-même statisticien mais utilisateur expert, je soumetts cette question aux statisticiens tout en maintenant pour l'instant ma position. [Ludovic Lebart souligne à ce moment que, dans une publication plus récente, D.Labbé a précisé que l'échelle n'est valide que pour des textes de 10'000 mots].

(2.3)

Oui, toutes les techniques, y compris celle de Dominique Labbé malgré ce biais que je lui suppose, convergent, sur le vocabulaire massif, pour indiquer une grande parenté lexicale, exceptionnelle, mais aussi [ici, je crois avoir nettement marqué mon emphase] pour laisser ouverte l'hypothèse d'un Molière autonome, sans autre intervention de Corneille que celle unanimement connue [Psyché].

(2.4)

Venons-en au fond.

La question n'est pas de savoir si les indices montrent une grande proximité du vocabulaire massif : nous en sommes TOUS d'accord.

La question est de savoir quels sont les facteurs de cette proximité, avant d'en venir à la réattribution positive. On a évoqué, à juste titre (sans que je sache pour ma part en déterminer les parts respectives) le genre et le thème. Mais en cette occurrence, il faut faire sa part à l'influence, et en cette occurrence précise à celle de l' "incorporation". La considération des seules données textuelles tend à désincarner les textes et à les décontextualiser. Molière disait

chaque jour les vers de Corneille. Comment pourrait-on ne pas en retrouver la trace dans ses propres écrits ?

(2.5)

C'est ici que voudrais expliquer que la thèse d'Hippolyte Wouters n'est pas la même que celle de Dominique Labbé. H.Wouters suppose de toute évidence de Corneille a écrit des PARTIES des pièces de Molière, ou plutôt que Molière est venu "abîmer" les pièces que Corneille écrivait sous son nom (le nom de Molière). Je cite [op.cit] pp.99-100 :

"Mais alors pourquoi à nouveau dans cette pièce admirable qu'est Dom Juan ces « passages-farces » parfaitement inutiles ? Que vient faire Monsieur Dimanche dans tout cela ? Qu'ajoute-t-il au personnage de Dom Juan ?"

"Sans doute est-ce un tort de ne pas payer ses fournisseurs, mais cela paraît quand même assez dérisoire à côté de son mépris affiché pour la religion, ses parents et l'amour ! Ne serait-ce pas encore une fois le metteur en scène, soucieux de plaire au public qui plaque sur des scènes superbes des arlequinades dépourvues de sens et de style ?"

La thèse est donc bien celle d'un plaquage du mauvais sur du bon.

La question serait donc, du point de vue qui nous concerne, à savoir celle de l'expertise littéraire : que serait Dom Juan, et le théâtre de Molière, sans ce "plaquage", que pour ma part j'appellerai au moins "dissonance" ?

Je propose à l'assistance de se demander de qui sont ces vers, tirés de leur contexte :

"Ruissent de tes toits de neiges et de suies"

"Le tambour de ton coeur fait craquer son tombeau"

Puis de ces vers, tirés du même poème :

"[...] j'entends les sirènes de police

Je préfère tes sirènes, celles qui ont la peau lisse".

Ma manière de rendre hommage à Claude Nougaro, disparu il y a une semaine, puisqu'ils sont tirés tous du même texte, "Harlem" (recueil "Nougayork", 1987). Le calembour est très mauvais : chaque gamin le connaît. Mais que serait ce texte sans cette proximité ?

La dissonance, intentionnelle ou non (c'est-à-dire exactement : quel que soit le point qu'elle occupe entre ces deux pôles), n'est-elle pas une part de ce qui rend Molière moderne (c'est-à-dire aussi ce qui explique son audience à travers les siècles et les époques), comme Shakespeare, autre "sujet d'énigmes", autre homme de tréteaux, d'arlequinades mélangées au sublime, etc ? Tous deux ont subi et sublimé la contrainte de la commande d'un théâtre vivant, de leur propre troupe, voire de leur propre jeu. Je partage l'avis de Valérie Beaudouin quant à la conception romantique de l'auteur qui préside à la thèse selon laquelle Molière n'a pu écrire ses vers (ou ses phrases de prose) "sublimes".

En résumé [ici, je me tourne vers Hippolyte Wouters, et non vers Dominique Labbé], VOTRE thèse est certes PLAUSIBLE, mais en l'état actuel des connaissances, je n'y crois pas. Toute ma culture littéraire me pousse à conserver la thèse établie, tout en m'engageant à tout faire pour la questionner encore et encore.

Voilà pour ce que j'ai dit.

On comprendra peut-être pourquoi j'ai été gêné de trouver le compte rendu de Benoît Peeters. Je n'ai en effet ni "reconnu le sérieux et la rigueur" des travaux de Dominique Labbé SUR CETTE QUESTION. Je n'ai en effet "pas contesté les proximités lexicales existant entre les deux oeuvres", et je n'ai pas employé "pour le moins" à propos de la plausibilité de l'hypothèse de Me Wouters. Je n'ai "reconnu" AUCUN "indice[s] littéraire[s] ou historique[s] pouva[ie]nt conforter cette hypothèse".

L'autre raison de ma gêne est la méconnaissance totale des usages que manifeste la publication unilatérale d'un compte rendu dans une matière controversée, sans consultation des participants. Cette publication risquait de ruiner l'effort de maîtrise qui avait permis une discussion parfaitement courtoise et qui, me semble-t-il, avait offert aux auditeurs l'occasion (appréciée d'eux) d'une meilleure connaissance de divers aspects du dossier.

Jean-Marie Viprey