

# De précieux manuscrits

Jean-Charles Basson <sup>1</sup>, Dominique Labbé <sup>2</sup>

<sup>1</sup> CreSco ; IFERISS – Université de Toulouse 3 – Jean-charles.basson@univ-tlse3.fr

<sup>2</sup> PACTE – Université Grenoble Alpes – dominique.labbe@umrpacte.fr

## Abstract

The Archives départementales of the Haute-Garonne in Toulouse hold the manuscripts of five plays of the 17th century: Aétius, César, Juba, Virginie and Tachmas. Two of these plays were performed at the Comédie française - and one was published - under the name of Jean-Galbert Campistron. This Toulousian dominated the dramatic stage at the end of the 17th century, after the retirement of Jean Racine (1677), who was still alive when Campistron's success was achieved. It has often been said of the latter that he was Racine's successor because his plays are very similar to those of his illustrious predecessor. However, a contemporaneous witness stated that Racine was the Campistron's ghostwriter. Until now, no one had paid attention to this assertion because it is impossible to recognize the author of a text simply by reading it. Today, computer algorithms make it possible. They show that all of Campistron's plays, as well as four of the five manuscripts in the Toulouse Archives, are by the same author who wrote the tragedies presented by Racine between 1667 (Andromaque) and 1677 (Phèdre). These manuscripts also provide some insights regarding the process of creation of these plays. Statistics applied to the texts thus shed new light on the history of French theatre. Above all, these findings invite to take a closer look at these manuscripts and to rediscover the plays presented by Campistron:

**Keywords:** Authorship attribution; France XVIIth century theater; Racine; Campistron.

## Résumé

Les Archives départementales de la Haute-Garonne à Toulouse détiennent les manuscrits de cinq pièces de théâtre de la fin du XVIIe siècle : Aétius, César, Juba, Virginie et Tachmas. Deux de ces pièces ont été jouées à la Comédie française et une a été publiée, toutes sous le nom de Jean-Galbert Campistron. Ce Toulousain a dominé la scène dramatique à la fin du XVIIe, après la retraite de Jean Racine (1677) qui était encore en vie lors des succès de Campistron. On a souvent dit de ce dernier qu'il était le continuateur de Racine tant ses pièces ressemblent à celles de son illustre prédécesseur. Or, à cette époque, un témoin a affirmé que Campistron était le prête-nom de Racine. Jusqu'à maintenant personne n'avait prêté attention à cette affirmation car il est impossible de reconnaître l'auteur d'un texte à la simple lecture. Aujourd'hui, des procédures informatiques le permettent. Elles montrent que toutes les pièces de Campistron, ainsi que quatre des cinq manuscrits des Archives de Toulouse, sont de la même plume que les tragédies présentées par J. Racine entre 1667 (Andromaque) et 1677 (Phèdre). Ces manuscrits éclairent également le processus de création d'une pièce de théâtre. La statistique appliquée aux textes jette ainsi une lumière nouvelle sur l'histoire du théâtre français. Elle invite surtout à examiner de plus près ces manuscrits et à redécouvrir les pièces présentées par Campistron..

**Mots clés :** Attribution d'auteur ; théâtre français du XVIIe siècle ; Racine ; Campistron.

## 1. Introduction

La carrière de Jean Racine (1639-1699) se déroule en deux temps. De 1664 à 1677, il présente dix pièces de théâtre dont certaines furent de grands succès (Annexe 1). En 1677, il est nommé historiographe du roi et cesse de produire pour le théâtre. Officiellement, il ne fera que deux exceptions pour Mme de Maintenon (*Esther* 1689 et *Athalie* 1691).

Après 1677, la scène tragique est dominée notamment par Jean-Galbert Campistron (1656-1723). Les pièces de Campistron sont présentées en annexe 3 (biographie dans : Jones, 1972 ; Groperrin, 2004 ; Campistron, 2002 ; Basson & Labbé, 2015). Les arguments des pièces et

leurs audiences sont résumés dans Basson & Labbé, 2015. Comme Racine, Campistron interrompt sa carrière théâtrale quand il entre au service du roi (vers 1693-1695, il devient secrétaire des galères royales).

Plusieurs contemporains ont fait le parallèle entre Racine et Campistron. Ce fut aussi le cas de Voltaire et de Hugo. Les spécialistes contemporains du théâtre du XVIIe sont d'accord sur cette proximité entre les deux œuvres, par exemple : Hepp, 1991 ; Truchet, 1992 ; Mazouer, 1993 ; Groperrin, 2004. Habituellement, on fait de Campistron un imitateur mais il y a un cas où Campistron aurait influencé Racine (Norman, 2004).

Des liens entre Racine et Campistron sont attestés par plusieurs témoins de l'époque. En particulier le père Dominique Colonia, lui-même dramaturge, écrit en 1695 que Racine prépare une tragédie sur Juba et affirme que Campistron "tient dignement la place" de Racine depuis longtemps (préface de Colonia, 1695).

Dans les années 1950, les descendants de Campistron ont remis, aux Archives départementales de la Haute-Garonne à Toulouse, les papiers de la famille parmi lesquels figurent les manuscrits de cinq pièces en alexandrins : la première et la dernière des tragédies présentées par Campistron : *Virginie*, *Aétius* (qui n'a jamais été publiée). Trois autres n'ont jamais été représentées et sont également inédites : *Juba*, *César*, et une ébauche (*Tachmas*).

Dès lors, au moins deux questions sont posées :

- *Juba* est-elle la pièce que, d'après Colonia, Racine préparait en 1695 et, dans l'affirmative, comment a-t-elle fini dans les papiers de Campistron ?
- la formule selon laquelle Campistron "tient dignement la place de Racine" signifie-t-elle que le premier occupe, sur la scène théâtrale, la place du second ou qu'il en est le prête-nom ?

Jusqu'à maintenant, il était impossible de répondre à ces questions car il n'existait aucun moyen d'identifier l'auteur d'un texte anonyme ou d'origine douteuse, du moins à la lecture de ce texte (Love, 2002). Il n'en est plus de même aujourd'hui grâce à de nouveaux outils (section 2) dont on présentera les résultats sur des cas indiscutables (section 3). Ils seront ensuite appliqués à Racine et Campistron (section 4). Cette application révélera également l'influence du temps sur la création littéraire (section 5). La conclusion discutera la portée de ces attributions et l'intérêt des manuscrits de Toulouse.

## 2. L'attribution d'auteur assistée par ordinateur

Comment reconnaître la plume qui a composé un texte ? L'ordinateur a donné une nouvelle actualité à cette question.

### 2.1. Etat de la question

On trouvera un tableau des principales méthodes dans : Argamon et al., 2009 ; Koppel et al., 2009 ; Stamatatos, 2009 ; Korcher & Savoy, 2019.

Toutes les méthodes d'attribution par ordinateur reposent sur un postulat plus ou moins explicite : chaque écrivain se singularise par certaines habitudes d'écriture. On peut donc établir un "profil" des écrivains connus puis celui du texte anonyme ou douteux et l'attribuer à l'auteur dont le profil sera le plus proche.

La plupart des expériences publiées ont révélé certaines limites. Quelques-unes sont particulièrement évidentes.

- Absence d'une norme de standardisation des graphies qui rend difficiles les comparaisons

entre chaque expérience.

- Pas d'accord sur les procédures de validation des méthodes. Par exemple, la plupart du temps, ces méthodes sont appliquées directement à des textes problématiques (les papiers fédéralistes américains, les œuvres du mormon J. Smith ou celles de Shakespeare...) sans avoir été au préalable testées sur des cas incontestables.

- La plupart sélectionnent certains mots - par exemple les mots outils - ou certaines classes de fréquence, ouvrant ainsi la possibilité de manipuler à l'infini les seuils et les listes de mots jusqu'à ce que la "bonne" solution soit atteinte...

Enfin, la marge d'erreur n'est pas indiquée.

De ces défauts découlent quelques principes simples, inspirés des sciences de l'ingénieur (Labbé & Labbé, 2001) : standardiser soigneusement les graphies ; considérer l'ensemble des textes (aucune sélection parmi les mots) ; donner à chaque mot un poids dans le calcul égal au nombre de fois qu'il apparaît dans chacun des textes ; minimiser les risques d'erreur, quitte à laisser des cas non-résolus ; organiser des expériences préalables visant à prendre la méthode en défaut ; établir des étalons afin de comparer les résultats obtenus sur des corpus différents et pour prendre des décisions en minimisant le risque d'erreur.

## 2.2. La distance intertextuelle

Un exposé est paru en ligne dans *Images des mathématiques* (Labbé & Labbé, 2011).

Soit deux textes A et B : combien ont-ils de mots en commun et de mots différents ? Les effectifs absolus sont convertis en nombres relatifs pour obtenir un indice (D) variant entre 0 et 1.

Une distance de 0,50 signifie que, entre les deux textes, un mot sur deux est commun, l'autre étant différent ; une distance de 0,25 que les trois quarts des mots sont communs, etc. Les principaux facteurs déterminant cette distance sont, par importance décroissante :

- une distance incompressible chez un même écrivain à la même époque, sur des thèmes voisins (pour le XVII<sup>e</sup> siècle, elle varie entre 0.15 chez Pierre Corneille (1606-1684) et 0.19 chez Racine) ;
- le genre : oral et écrit, prose, vers, comédie et tragédie, etc.
- l'écrivain ;
- l'époque pendant laquelle le texte a été rédigé (le lexique évolue avec le temps) ;
- le thème (vocabulaire propre à ce thème, noms des personnages, lieux, principaux motifs).

L'équation (1) ci-dessous résume ces facteurs.

$$D(A,B) = f (D_{\min}(A,B), (\text{GenreA,GenreB}), (\text{AuteurA,AuteurB}), (\text{EpoqueA,EpoqueB}), (\text{ThèmeA,ThèmeB})) \quad (1)$$

Pour déterminer la plume qui a composé un texte d'origine douteuse ou inconnue, ce texte est confronté à d'autres – dont l'origine n'est pas douteuse - écrits dans un même genre et à la même époque. Autrement dit : le théâtre est comparé au théâtre et les tragédies entre elles, les comédies entre elles, etc.

## 2.3. Validation de la méthode

Cette méthode a passé un grand nombre de tests (par exemple, Monière & Labbé, 2006 ; Labbé, 2007) ; dernièrement, l'identification de la plume dans l'ombre d'E. Ferrante (Savoy, 2018). Dans ce dernier cas, la conclusion a été validée par les résultats de sept équipes de recherche à travers le monde (Ratinaud, 2018 ; Tuzzi, 2018 ; Tuzzi & Cortelazo, 2018).

Cette méthode est utilisée pour repérer certaines fraudes dans les publications scientifiques (Labbé & Labbé, 2012) avec des résultats qui ont été signalés à trois reprises par *Nature* (Van-Norden, 2014 ; Phillips, 2017 ; Byrne, 2019). Ces outils sont intégrés dans le processus de publication du deuxième groupe mondial d'édition, Springer-Mac-Millan (Minh Tien, 2018). Ces algorithmes sont utilisés par les principaux éditeurs scientifiques. Récemment, ils ont permis de détecter une fraude importante dans la recherche sur le cancer (Byrne & Labbé, 2016), ce qui a valu à J. Byrne le titre de "Scientist of the Year 2017" (*Nature*).

Cette méthode a été appliquée à *Aétius*, *César*, *Juba* et *Tachmas* comparées à 236 autres pièces de théâtre de la seconde moitié du XVIIe, parues sous 32 noms différents (liste en annexe de Labbé, 2014). Cette expérience montre d'abord comment se présentent les distances entre textes écrits par une plume unique (distances "intra-auteur") et celles entre textes par deux auteurs différents (distances "inter-auteurs").

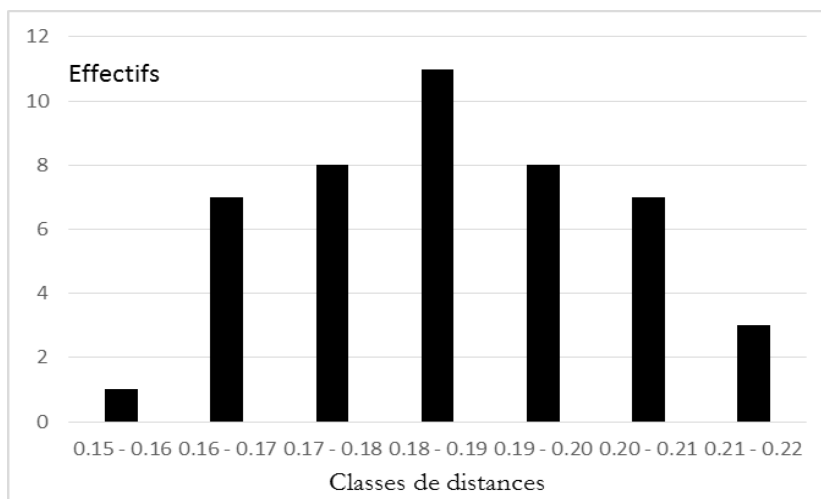
### 3. Distribution des distances observées sur un auteur unique ("intra")

Pour vérifier l'aptitude de la méthode à identifier la plume ayant composé une tragédie, il faut donc l'appliquer d'abord à des auteurs connus, par exemple : P. Corneille ou Racine.

#### 3.1. P. Corneille

Considérons les dix tragédies de P. Corneille entre 1659 (*Edipe*) et 1674 (*Suréna*), soit 45 distances différentes ("intra") entre ces pièces considérées deux à deux (matrice des distances dans Labbé, 2014). Elles sont rangées par ordre croissant et groupées dans des classes d'intervalles égaux (correspondant au seuil de précision de la mesure : 0.01) (Figure 1).

Figure 1. Histogramme des distances entre les tragédies de P. Corneille (1659-1674)



Trois valeurs centrales caractérisent cette série :

- la distance médiane (Me) qui partage les distances classées en parts égales (la moitié en dessous, la moitié en dessus) : 0,178, soit 0,18 ;
- la distance modale (Mo) (classe de distances dont l'effectif est le plus important) : [0.18 – 0,19[. Conventionnellement, le mode est le milieu de l'intervalle de cette classe : 0,185 ;
- la distance moyenne ( $\bar{D}$ ) : 0,181, soit 0,18.

La distribution en forme de cloche (avec symétrie autour de la classe centrale qui est aussi la plus peuplée) et l'égalité des trois valeurs centrales indiquent que les distances sont distribuées de façon normale (ou "gaussienne") autour de la moyenne. Il s'agit donc d'une

population homogène. Dès lors, on peut calculer un écart type ( $\sigma$ ), mesure standard de la dispersion des distances autour de la moyenne (l'écart-type est la moyenne quadratique des écarts entre les valeurs observées et la moyenne arithmétique de celles-ci) : 0,0154.

En cas de distribution normale, 95% des distances, entre les textes composant cette population homogène, sont comprises dans l'intervalle  $\bar{D} \pm 1,96\sigma$ , soit [0,15 et 0,21] et 99% entre  $\bar{D} \pm 2,58\sigma$  [0,14-0,22].

Dans l'équation (1) ci-dessus, le minimum est constant (0,15) et les variables "genre" et "auteur" sont neutralisées. Il ne reste que les thèmes et le temps. On remarque que les distances les plus élevées séparent les pièces les plus éloignées dans le temps : *Œdipe* (1659) avec *Agésilas* (1666). A contrario, la plus petite distance sépare *Tite et Bérénice* (1670) de *Pulchérie* (1672). Autrement dit, les tragédies présentées par P. Corneille durant la dernière partie de sa vie (sur une quinzaine d'années) sont sœurs (très proches) mais leur père aurait légèrement changé ses habitudes d'écriture entre 1659 et 1674, ce léger changement ne remettant pas en cause la distribution normale des distances.

### 3.2. Racine

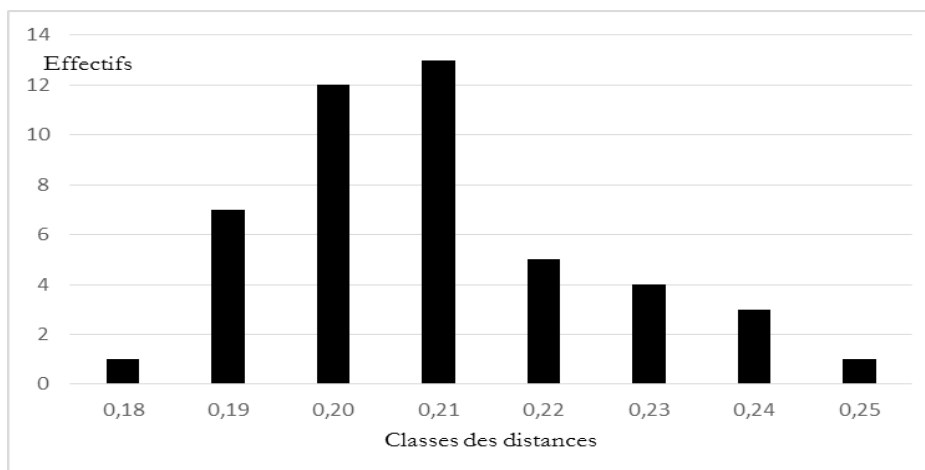
Considérons les tragédies présentées par Racine entre 1667 (*Andromaque*) et 1677 (*Phèdre*) : Me : 0,220 ;  $\bar{D}$  : 0,220 ; Mo : 0,225. L'égalité de ces trois valeurs centrales indique à nouveau une distribution normale. L'écart-type est pratiquement égal à celui constaté chez Corneille (0.0156). Il détermine un intervalle de fluctuation normale à 95% [0,190 - 0.245] et un à 99% [0.185 - 0.255]. Les valeurs centrales plus élevées que chez P. Corneille semblent tenir à la fois à une distance incompressible plus forte (0,19 au lieu de 0,15) et à une plus grande sensibilité au temps (ou versatilité), puisque les distances les plus fortes sont constatées entre les pièces dont la création est la plus éloignée dans le temps.

La distribution des fréquences (à gauche de la Figure 4) présente aussi un profil gaussien.

### 3.3. Campistron

Considérons les sept tragédies présentées par J.-G. Campistron auxquelles sont ajoutées d'une part *Pompéia* (jamais représentée du vivant de Campistron et publiée de manière posthume en 1750) et les 3 inédites supposées de la même plume : *Aétius*, *Juba* et *Tachmas*. Pour *César*, la moyenne de ses distances avec les autres sort de l'intervalle de fluctuation normale. Elle ne figure donc pas dans la suite de l'expérience (Figure 2).

Figure 2. Histogramme des distances entre les pièces présentées par J.-G. Campistron



La médiane est égale à 0.208, le mode à 0.215, la moyenne 0.210, l'écart-type à 0,0168. 95% des valeurs sont inscrites dans un intervalle de variation de [0,177 à 0.243]. Comme les autres séries "intra", la distribution est approximativement "gaussienne".

L'asymétrie de la distribution peut être expliquée par le caractère inachevé de *Juba* et *Tachmas* et, surtout, par les douze années séparant les premières des dernières pièces (entre 1683 et 1695), *Juba* étant la plus décalée. Cette distribution permet de conclure à un auteur unique. *Pompéïa*, *Aétius*, *Juba* et *Tachmas* sont bien de la même plume que les autres tragédies présentées par Campistron.

Conclusion de ces trois expériences : pour un même auteur, des textes produits dans un même genre à des dates proches, sont séparés par des distances faibles (ici : les distances moyennes sont comprises entre 0,18 et 0,22) et distribuées de manière gaussienne, ce qui permet de les doter d'un intervalle normal de fluctuation (la variable est "centrée réduite").

Selon l'équation (1) présentée ci-dessus, les textes produits par des auteurs différents, à la même époque et dans un genre unique (distances "inter-auteurs"), doivent tomber en dehors de ces intervalles, puisque le poids de la variable "auteur" l'emporte sur celui du thème et celui du temps.

#### 4. Les distances entre auteurs différents ("inter")

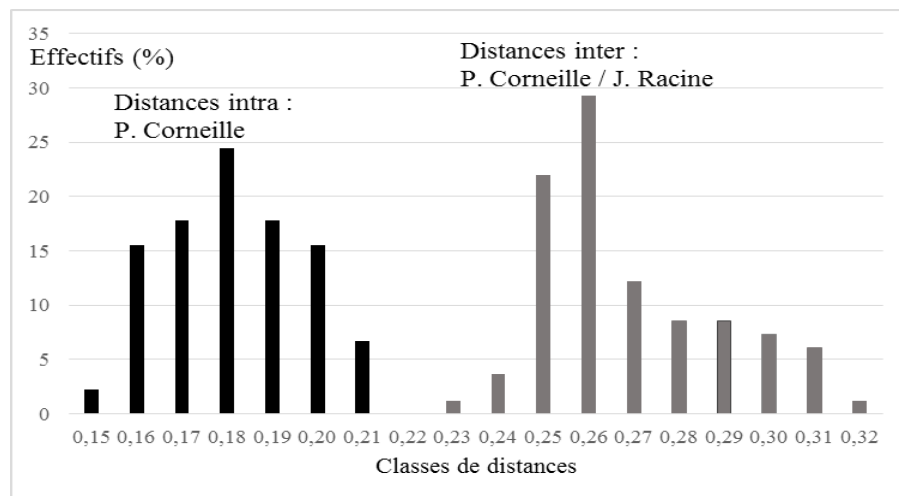
Si la distance observée entre deux textes tombe dans les intervalles définis ci-dessus, on ne peut écarter l'hypothèse d'un auteur unique (voir en conclusion les conditions d'acceptation de cette hypothèse). Si la distance excède la borne supérieure de l'intervalle, on écarte cette hypothèse avec un risque d'erreur de 5 ou de 1%.

Vérifions d'abord le raisonnement sur des cas indiscutables.

##### 4.1. Corneille et Racine

On calcule les 45 distances entre les sept tragédies de Racine et les dix (contemporaines) de P. Corneille, on leur applique le même traitement que ci-dessus, puis on rapproche cette série des deux séries de distances intra présentées au paragraphe précédent (Figures 3 et 4).

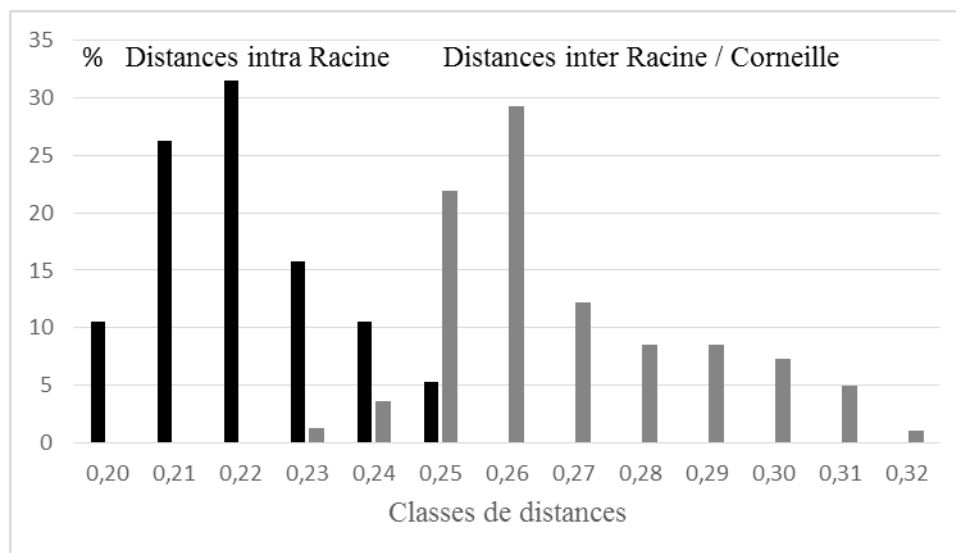
Figure 3 Histogramme des distances intra-P. Corneille et inter P. Corneille/J. Racine



Lecture : 24% des distances "intra" (tragédies de P. Corneille) sont comprises dans l'intervalle [0,18-0,19[ ; 29% des distances entre les tragédies de P. Corneille et de Racine sont comprises dans l'intervalle [0,26-0,27[, etc.

Pour les distances inter (en gris), la médiane 0,260 et le mode 0,265 sont proches de la moyenne (0,266), ce qui signale une population proche de la normale avec une légère asymétrie à droite. La variation autour de cette moyenne (écart type : 0,017) donne les intervalles suivants : 95% [0,235 – 0,300] et 99% [0,225 et 0,310].

Figure 4 Histogramme des distance intra Racine et inter Racine/Corneille.



Dans la figure 3, la classe 0,22 est vide : aucun recouvrement entre la série intra Corneille et la série inter Corneille-Racine. Dans la Figure 4 (intra Racine et inter Corneille-Racine), le recouvrement des deux séries concerne 5% du total des distances inter. Pour chacune des pièces, la moyenne des distances inter – avec toutes les pièces de l'autre auteur - tombe en dehors des deux intervalles "intra" à 1%. Les deux populations (tragédies par Corneille et par Racine) sont donc disjointes et indépendantes. L'hypothèse d'un seul écrivain est rejetée et l'existence de deux écrivains différents est retenue, avec des risques négligeables d'erreurs de première et de seconde espèces.

Le résultat peut sembler logique. Cependant, la tragédie classique en alexandrins était enserrée dans des règles strictes. Surtout, ces pièces ont été écrites en guettant celles du rival, en pensant aux mêmes actrices et acteurs et à un public restreint (le roi, sa famille, sa cour et les riches parisiens). Il est même arrivé, en 1670, que les deux auteurs traitent - en "aveugle" – le même thème (les amours de l'empereur Titus et de la reine Bérénice). Pourtant, même dans ces conditions limites, les auteurs sont identifiés avec un risque d'erreur négligeable.

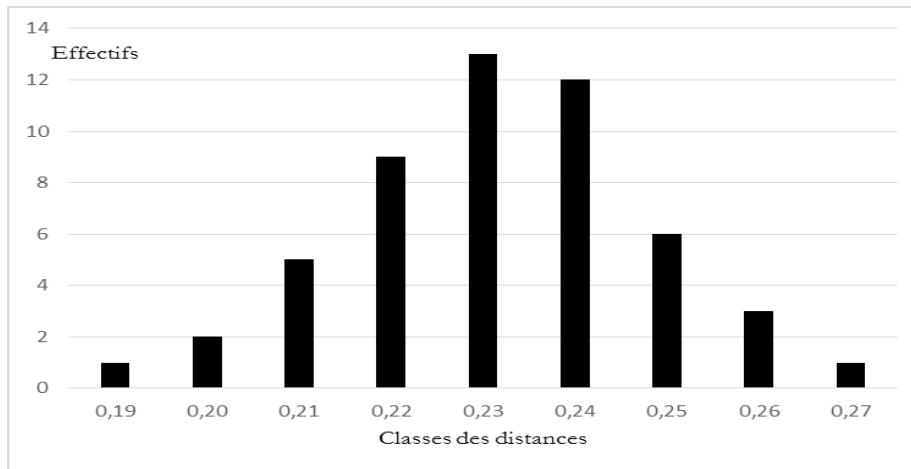
#### 4.2. Racine, Campistron et La Chapelle

Le même dépouillement a été effectué sur 239 pièces de théâtre de la seconde moitié du XVIIe. Une anomalie est apparue à propos des tragédies présentées par Racine, Campistron mais aussi avec celles de Jean de La Chapelle (annexe 2).

Les distances "inter" entre les trois corpus (Racine-La Chapelle ; Racine-Campistron et La Chapelle-Campistron) ont été classées de la même manière que précédemment (Figure 5). NB : s'il s'agissait de trois auteurs différents, on devrait obtenir des profils semblables à ceux des figures 3 et 4.

Les trois valeurs centrales sont égales (moyenne = 0,232 ; médiane = 0,229 ; mode = 0,235). La distribution suit une loi normale. Il s'agit donc d'une population unique, homogène (ou encore "stationnaire"). De plus, toutes les moyennes des distances "inter" tombent à l'intérieur des deux intervalles "intra" définis ci-dessus. On peut donc conclure à une seule plume.

Figure 5. Histogramme des distances inter-tragédies présentées par J. de La Chapelle, par J. Racine et J.-G. Campistron



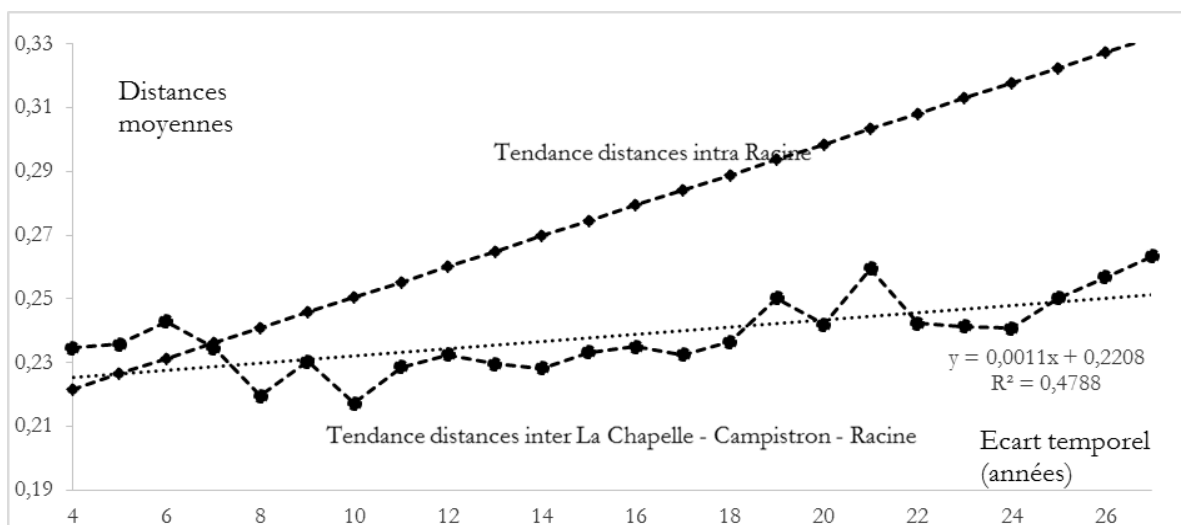
Cependant, les valeurs centrales sont légèrement plus élevées que dans les cas précédents. En suivant l'équation (1), la distance incompressible demeure approximativement constante (ici 0,18 pour Campistron, 0,19 pour Racine), de même que la variation entre les thèmes. Donc, seul le facteur temps pourrait expliquer ces distances un peu plus élevées ?

## 5. Le facteur temps

Les intervalles temporels séparant la création des pièces considérées deux à deux sont mesurés en années. Par exemple, il s'écoule dix ans entre *Andromaque* (Racine 1667) et *Phèdre* (Racine 1677). La dernière pièce de Racine (*Phèdre*, début 1677) apparaît quatre ans avant *Zaïde* (La Chapelle 1681) et six ans avant *Virginie* (Campistron 1683). L'intervalle le plus grand sépare *Andromaque* (1667) d'*Aétius* (1693) ou de *Juba* : 26 et 28 ans.

Pour mesurer l'influence de ces décalages temporels, on associe les intervalles, mesurés en années, aux moyennes des distances correspondantes (points gras sur la figure 6) puis on réalise un ajustement linéaire de ces données sur les 28 ans (pointillé maigre en bas). La tendance propre au corpus Racine (1667-1677) est prolongée jusqu'à 1695 (pointillé gras).

Figure 6. Accroissement de la distance en fonction du temps (distances entre les œuvres de J. Racine et celles présentées sous les noms de La Chapelle et Campistron).





La relation linéaire entre les intervalles temporels et les distances est observée dans tous les corpus d'auteur unique et homogènes quant au genre (par exemple, la correspondance de V. Hugo dans Labbé & Labbé, 2011).

D'après le trait gras pointillé, si le rythme d'accroissement des distances en fonction du temps, observé entre 1667 et 1677, s'était poursuivi jusqu'au bout, la distance entre *Aétius* et *Andromaque* aurait été de 0.33 alors qu'elle est de 0.26. En dehors des distances entre les sept tragédies de Racine et les trois premières pièces parues sous le nom de J. La Chapelle, toutes les autres sont "trop" proches par rapport à ce que les écarts temporels laisseraient attendre.

Après 1683, la tendance à l'accroissement des distances en fonction du temps est donc quatre fois plus lente qu'auparavant. La pente de la droite d'ajustement est d'ailleurs très faible (0.001). Mais la liaison demeure. Avec 25 degrés de liberté, le coefficient de corrélation (0.69) est même très significatif : on a moins d'une chance sur 1000 de se tromper en acceptant une augmentation faible mais continue des distances entre les pièces présentées sous le nom de J. Racine et celles présentées par J. de La Chapelle puis par J. G. Campistron, mais le temps n'explique que la moitié des fluctuations de la variable. Pour l'autre moitié, plusieurs explications sont possibles (et cumulables) :

- pour certaines pièces présentées par La Chapelle et Campistron, la composition serait antérieure à leur parution. Ceci vaut principalement pour les dernières (*Tiridate*, *Pompéia*, *Aétius*, *Juba*) qui semblent nettement antérieures ;

- les pièces qui ont assuré les plus grands succès avant 1677 (*Iphigénie*, *Andromaque*...) ont servi de modèles pour celles présentées sous les noms de La Chapelle puis de Campistron. En tous cas, cette unité du corpus à travers le temps renforce l'attribution à une même plume de toutes les tragédies présentées, entre 1666 et 1693, sous les noms de Racine, de La Chapelle puis de Campistron puisque les distances constatées sont inférieures aux valeurs attendues pour le seul Racine, étant donné les intervalles de temps qui séparent leurs dates de création.

## 5. Conclusions

Cette attribution est contrôlable et reproductible : les corpus et les logiciels sont à la disposition des chercheurs, les matrices de distance sont publiées (Labbé, 2014).

Quelle est la marge d'erreur ? Puisque, la distance moyenne "inter" de chacune des pièces présentées par l'un des auteurs tombe dans l'intervalle de confiance "intra" à 5% définis pour l'autre (Racine et Campistron), ou dans ces deux intervalles pour La Chapelle, il y a moins de 5% de chance de se tromper en rejetant l'hypothèse de trois auteurs différents au profit de celle d'une plume unique pour chacune des pièces. Puisque ces épreuves ne sont pas indépendantes, l'incertitude pesant sur le résultat final serait de 0,05<sup>20</sup> (il y a 21 pièces attribuées).

Cette attribution est confortée par diverses classifications automatiques et par d'autres indices stylistiques, comme les longueurs des phrases (Labbé, 2014). Les indices historiques (comme la présence des manuscrits dans le fonds Campistron ou l'allusion du P. de Colonia) renforcent encore cette attribution. C'est la convergence de tous ces indices qui permet de conclure qu'une même plume a composé toutes ces pièces. S'il n'y avait que quelques distances anormalement faibles, il serait impossible de choisir entre deux hypothèses : un seul écrivain ou un mimétisme entre deux auteurs différents, voire une collaboration plus ou moins ponctuelle.

En effet, il faudrait considérer une troisième hypothèse : une collaboration au long cours, selon des modalités à peu près stables, comme on en rencontre, par exemple, dans la recherche scientifique (Labbé & Labbé 2014). Cette collaboration semble rare en littérature, sans être impossible (Lafon & Peters 2006). Il faudrait donc calculer plusieurs risques de seconde espèce. En effet, dans le cas de deux hypothèses alternatives, le risque de seconde espèce (rejeter une hypothèse alors qu'elle est vraie) est complémentaire de celui de première espèce (accepter l'autre hypothèse alors qu'elle est fausse) ; en revanche, dans le cas d'hypothèses "composites", il y a lieu d'estimer le risque d'erreur de seconde espèce pour chacune des hypothèses rejetées. Cependant, dans le cas présent, cette troisième hypothèse doit être écartée car, du fait de leurs dates de naissance, La Chapelle et Campistron étaient trop jeunes pour avoir collaboré aux premières pièces parues sous le nom de Racine. De plus, pour Campistron, il existe, dans le fonds déposé aux Archives départementales de Toulouse, plusieurs lettres de sa main qui attestent une maîtrise limitée de la langue française.

Enfin, et surtout, ces manuscrits de Toulouse sont précieux pour la compréhension du théâtre du XVII<sup>e</sup>. En effet, on ne dispose d'aucun manuscrit des pièces des frères Corneille, de Molière, de Quinault, de Racine, etc. Or, les documents conservés à Toulouse offrent un panorama à peu près complet du processus de création des pièces de l'époque, depuis l'esquisse (*Tachmas*), les brouillons corrigés à divers stades (*Virginie*, *Juba*), la pièce jouée mais non publiée (*Aétius*) et enfin la pièce publiée (*Virginie*). Et l'intérêt redouble quand on constate que certains manuscrits comportent plusieurs écritures différentes...

## Remerciements

Mme G. Douillard, directrice des Archives départementales de la Haute-Garonne, et M. B. de Saint-Martin nous ont permis de photographier les manuscrits du fonds Maniban-Campistron. G. Bensimon, C. Labbé et J. Savoy ont relu une première version de cette communication et nous ont fait de très utiles remarques.

E. Arnold, G. Bensimon, J.-G. Bergeron, M. Brugidou, P. Hubert, N. & J. Leselbaum, C. Labbé, X. Luong, T. Merriam, D. Monière, J. Picard, A. Pibarot, M. Ruhlman, J. Savoy ont collaboré à la mise au point de la méthode d'attribution d'auteur.

Toutes nos recherches sont réalisées sans aide publique ni mécénat.

## References

- Argamon S., Koppel M., Pennebaker J. W. & Schler J. (2009). Automatically profiling the author of an anonymous text. *Communications of the ACM*, 52(2), p. 119–23.
- Basson J.-C. & Labbé D. (2015), *Jean Racine, Aétius, Juba, Tachmas. Tragédies inédites transcrites et présentées par J.-C. Basson et D. Labbé*. Montréal : Monière-Wollank Éditeurs et Bibliothèque Nationale du Québec, Paris : Bibliothèque Nationale de France.
- Byrne J. (2019). We need to talk about systematic fraud, *Nature*, 06 February 2019.
- Byrne J. & Labbé C. (2016). Striking similarities between publications from China describing single gene knockdown experiments in human cancer cell lines. *Scientometrics*, 28 December 2016.
- Campistron J.-G. (2002). *Tragédies (1684-1685). Arminius Andronic Alcibiade*. Edition établie, présentée et annotée par J.-P. Groperrin et J.-N. Pascal. Toulouse : Société de Littératures Classiques.
- Colonia D. de (1695). *Juba*. Lyon : Guerrier.
- Groperrin J.-P. (Ed 2004). *Campistron et consorts : tragédie et opéra en France (1680-1733)*. Littératures classiques. 52, automne 2004.

- Hepp N. (1991). De la littérature à la rhétorique : Campistron imitateur de Racine. In Wentzlaff-Eggebert C. (ed.). *Le langage littéraire au XVIIe siècle*. Tübingen : Gunter Narr, p. 295-303.
- Jones D. F. (1972). *Jean de Campistron: a study of his life and work*. University Mississippi Romance monographs.
- Koppel M., Schler J. & Argamon S. (2009). Computational Methods in Authorship Attribution. *Journal of the American Society for Information Science and Technology*. 2009, 60-1, p. 9-26.
- Kosher M. & Savoy J. (2019). Evaluation of text representation schemes and distance measures for authorship linking. *Digital Scholarship in the Humanities*. 34, p. 189-207.
- Labbé C. & Labbé D. (2001). Inter-Textual Distance and Authorship Attribution Corneille and Molière. *Journal of Quantitative Linguistics*. 8-3, december 2001, p. 213-231.
- Labbé C. & Labbé D. (2011). La classification des textes. Comment trouver le meilleur classement possible au sein d'une collection de textes ? *Images des mathématiques. La recherche mathématique en mots et en images*. (<http://images.math.cnrs.fr/La-classification-des-textes.html>). 28 mars 2011.
- Labbé C. & Labbé D. (2012). Duplicate and fake publications in the scientific literature: how many SCIdgen papers in computer science? *Scientometrics*, 22 June 2012.
- Labbé C. & Labbé D. (2014). *Who wrote this scientific text?* Grenoble : Laboratoire d'Informatique de Grenoble (LIG).
- Labbé D. (2007). Experiments on Authorship Attribution by Intertextual Distance in English, *Journal of Quantitative Linguistics*, 14(1), p. 33-80, April 2007
- Labbé D. (2014). *Qui a écrit Aétius, Juba et Tachmas ? Une attribution d'auteur par ordinateur*. Grenoble : Pacte (document consultable en ligne).
- Lafon M. & Peeters B. (2006). *Nous est un autre*. Paris : Flammarion.
- Love H. (2002). *Attributing Authorship: an Introduction*. Cambridge. Cambridge University Press.
- Mazouer C. (1993). La mémoire des classiques chez Regnard et Campistron. Les lieux de mémoire et la fabrique de l'œuvre. *Biblio 17 (Papers on French Seventeenth Century Literature)*, 80, p. 277-293.
- Minh Tien N. (2018). *Detection of automatically generated texts*. Ph-D Thesis. Grenoble-Alpes University. 3-04-2018.
- Monière D. & Labbé D. (2006). L'influence des plumes de l'ombre sur les discours des politiciens. In Condé C. et Viprey J.-M. *Actes des 8e Journées internationales d'Analyse des données textuelles*. Besançon : 19-21 avril 2006, II, p. 687-696.
- Norman B. (2004). L'univers nouveau de la tragédie en musique : d'Armide et Renaud à Achille et Polyxène. In Groperrin J.-P.(Ed). *Campistron et consorts : tragédie et opéra en France (1680-1733)*. Littératures classiques. 52, automne 2004, p. 231-241.
- Philipps N. (2017). Online software spots genetic errors in cancer papers. *Nature*. 20 November 2017.
- Racine J. (2015). *Aétius, Juba, Tachmas. Tragédies inédites transcrites et présentées par J.-C. Basson et D. Labbé*. Montréal : Monière-Wollank Editeurs (Consultable en ligne sur ResearchGate).
- Ratinaud P. (2018). The Brilliant Friend(s) of Elena Ferrante: A Lexicometrical Comparison between Elena Ferrante's Books and 39 Contemporary Italian Writers. In Tuzzi A. & Cortelazzo M. (dir.), *Drawing Elena Ferrante's Profile*. Padova : Padova University Press, p. 97-110.
- Savoy J. (2018). Is Starnone really the author behind Ferrante? *Digital Scholarship in the Humanities*, 33(4), p. 902-918.
- Stamatatos E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*. 60-3, p. 538-556.

Truchet J. (1992). Notice de Tiridate. *Théâtre du XVIIe siècle*. Gallimard (La pléiade), tome III, p. 1279-1284.

Tuzzi A. (2018). *It Takes Many Hands to Draw Elena Ferrante's Profile*. Padova University Press.

Tuzzi A. & Cortelazzo M. (2018). What is Elena Ferrante? A comparative analysis of a secretive bestselling Italian writer. *Digital Scholarship in the Humanities*, 33, 3, p. 685–7028.

Van Noorden R. (2014). Publishers withdraw more than 120 gibberish papers. *Nature*, 24 February 2014.

#### Annexe 1. Les pièces de Jean Racine

Titre	Genre	Création	Longueurs (Mots)	Vocabulaire
La Thébàide ou les frères ennemis	Tragédie	20 juin 1664	13 813	2 256
Alexandre le Grand	Tragédie	4 décembre 1665	13 864	2 383
Andromaque	Tragédie	17 novembre 1667	15 076	2 441
Les Plaideurs	comédie	novembre 1668	8 041	1 839
Britannicus	Tragédie	13 décembre 1669	15 387	2 748
Bérénice	Tragédie	21 novembre 1670	13 242	2 254
Bajazet	Tragédie	1er janvier 1672	15 297	2 598
Mithridate	Tragédie	23 décembre 1672	15 091	2 650
Iphigénie	Tragédie	18 août 1674	15 782	2 753
Phèdre	Tragédie	1er janvier 1677	14 394	2 918
Esther	Tragédie	26 janvier 1689	11 147	2 518
Athalie	Tragédie	5 janvier 1691	15 492	2 989
			166 626	10 145

#### Annexe 2. Les pièces de Jean de la Chapelle

Titre	Genre	Création	Longueurs (Mots)	Vocabulaire
Zaïde	Tragédie	26 janvier 1681	13 048	2 314
Cléopâtre	Tragédie	12 décembre 1681	12 158	1 342
Téléphonte	Tragédie	26 décembre 1682	12 425	1 342
			37 631	2 138

#### Annexe 3. Les pièces de Jean-Galbert Campistron

Titre	Genre	Création	Longueurs (Mots)	Vocabulaire
Virginie	Tragédie	12 février 1683	13 430	1 349
Arminius	Tragédie	19 février 1684	13 478	1 455
Amante amant (I')	Comédie	2 août 1684	16 331	1 621
Andronic	Tragédie	8 février 1685	12 457	1 400
Alcibiade	Tragédie	28 décembre 1685	13 182	1 389
Acis et Galatée	Pastorale	septembre 1686	4 369	799
Achille et Polixène	Tragédie musique	7 novembre 1688	5 990	924
Phocion	Tragédie	16 décembre 1688	10 713	1 315
Adrien	Tragédie	11 janvier 1690	12 835	1 499
Tiridate	Tragédie	12 décembre 1691	11 488	1 330
Aétius	Tragédie	28 janvier 1693	10 368	1 290
Alcide	Opéra	3 février 1693	5 563	926
Pompeia	Tragédie	1693	11 881	1 469
Juba	Tragédie inédite	1695 ?	9 757	1 367
Le jaloux désabusé	Comédie	1709	11 882	1 548
César	Tragédie inédite	?	15 322	1 716
Tachmas	Tragédie inédite	?	4 039	756
Total			183 085	4 233