

Qui a écrit Molière ?

CHIKHANI Charles

ZAPFACK MESSIAN Jasen Steve

Encadrant : M. Hervé Fournier

Université Paris Cité

02 Juin 2023

Table des matières

1	Introduction	2
1.1	Présentation du sujet : la remise en question de l'attribution des pièces de Molière à l'auteur lui-même	2
1.2	Importance de l'analyse textuelle et statistique dans la résolution de cette question . .	2
1.3	Hypothèses et problématique	2
2	Compréhension de l'article de Florian Cafiero et Jean-Baptiste Camps	3
2.1	Contexte et objectifs de l'étude	3
2.2	Méthodologie utilisée	3
2.3	Formalisation du problème de l'attribution des pièces de théâtre	3
2.4	Approche adoptée pour résoudre le problème	3
2.5	Présentation et analyse des résultats obtenus	3
3	Notre expérience	3
3.1	Description de l'expérience que nous avons menée	3
3.2	Choix des outils d'analyse textuelle et statistique utilisés	3
3.3	Méthodologie mise en place	3
3.4	Collecte et préparation des données	4
3.5	Analyse des résultats obtenus et comparaison avec l'article de référence	4
4	Perspectives et réflexions	4
4.1	Approfondissement des techniques de linguistique quantitative pour déterminer la paternité d'un texte	4
4.2	Réflexion sur les aspects algorithmiques impliqués dans l'analyse textuelle et statistique	4
4.3	Limitations de notre expérience et suggestions pour de futures recherches	4
5	Conclusion	4
5.1	Récapitulation des principales conclusions de l'article de Florian Cafiero et Jean-Baptiste Camps	4
5.2	Résumé des résultats de notre expérience et leur concordance ou divergence avec l'article	4
5.3	Implications de cette étude dans le domaine de l'attribution des œuvres littéraires . .	4

1 Introduction

1.1 Présentation du sujet : la remise en question de l'attribution des pièces de Molière à l'auteur lui-même

Depuis plusieurs décennies, une question persiste dans le domaine de la littérature : l'attribution des pièces de Molière à l'auteur lui-même est-elle remise en question ? Cette controverse a été initiée par Pierre Louÿs, un romancier du XXe siècle, qui a suggéré que Pierre Corneille aurait pu être l'auteur véritable des pièces de Molière. Cependant, cette théorie repose sur des fondements fragiles et ne bénéficie d'aucune preuve concrète. Malgré cela, la rumeur persiste, alimentée par des éléments tels que l'éclosion tardive de Molière en tant qu'auteur, son prétendu manque d'éducation et de culture, ainsi que l'absence de preuves manuscrites permettant de réfuter directement cette hypothèse.

1.2 Importance de l'analyse textuelle et statistique dans la résolution de cette question

Au début des années 2000, Cyril et Dominique Labbé, deux chercheurs, ont avancé l'idée selon laquelle Corneille aurait écrit pour Molière. Leur méthode consiste à mesurer une "distance inter-textuelle" qui évalue la différence de lexique entre les textes des deux auteurs. Si cette distance ne dépasse pas un certain seuil, les deux pièces sont considérées comme écrites par le même auteur. Ils se basent également sur le fait que de nombreux dramaturges de l'époque signaient leurs œuvres sous le nom de "comédien poète", ce qui permettait aux véritables auteurs de rester anonymes tout en bénéficiant de la promotion et de la représentation de leurs pièces par les acteurs.

Cependant, cette méthodologie a été contestée par d'autres chercheurs. Certains ont souligné que l'implémentation de la méthode de Cyril et Dominique Labbé pourrait "lisser artificiellement les différences entre les auteurs", en utilisant une distance euclidienne qui accorde trop de poids aux lemmes fréquents, réduisant ainsi la disparité entre les fréquences observées de différentes formes. D'autres approches ont été proposées pour résoudre le problème de l'attribution des comédies de Molière. Certaines méthodes utilisent une analyse textuelle et statistique pour comparer les styles d'écriture, tandis que d'autres adoptent des approches plus qualitatives en examinant les intrigues, la versification et les sujets choisis dans les pièces.

1.3 Hypothèses et problématique

Dans ce rapport, nous examinerons deux hypothèses qui remettent en question la paternité des œuvres de Molière. La première hypothèse suggère que Molière aurait fourni des brouillons à Pierre Corneille, qui aurait ensuite versifié les pièces, peut-être avec l'aide de son frère. Selon cette hypothèse, Molière aurait créé les intrigues, mais la versification aurait été réalisée par Pierre Corneille, sans recevoir un crédit explicite. La deuxième hypothèse soutient que Molière n'aurait ni écrit les intrigues ni les vers de ses pièces, et qu'il n'aurait été qu'un nom célèbre utilisé pour promouvoir les pièces et dissimuler le véritable auteur.

Pour résoudre cette controverse, différentes méthodes ont été utilisées, chacune avec ses propres avantages et limitations. L'objectif de cette étude est d'évaluer ces approches et de fournir une analyse critique des résultats obtenus.

Pour ce faire, dans la section suivante, nous présenterons en détail l'article de Florian Cafiero et Jean-Baptiste Camps, qui examine cette question en utilisant une analyse textuelle et statistique. Nous aborderons les différentes perspectives et critiques soulevées par d'autres chercheurs, ainsi que les différentes approches méthodologiques qui ont été utilisées pour résoudre le problème de l'attribution des pièces de Molière.

Le rapport qui suit explorera en profondeur ces différentes approches et tentera de faire la lumière sur cette controverse persistante, fournissant ainsi une contribution significative à notre compréhension de l'œuvre de Molière et de son véritable auteur.

2 Compréhension de l'article de Florian Cafiero et Jean-Baptiste Camps

2.1 Contexte et objectifs de l'étude

2.2 Méthodologie utilisée

2.3 Formalisation du problème de l'attribution des pièces de théâtre

2.4 Approche adoptée pour résoudre le problème

2.5 Présentation et analyse des résultats obtenus

3 Notre expérience

3.1 Description de l'expérience que nous avons menée

Dans le cadre de cette étude, nous avons mené une expérience visant à déterminer la paternité des textes de Molière. L'objectif principal était de développer une méthodologie pour identifier les caractéristiques distinctives du style d'écriture de Molière et les comparer à ceux d'autres auteurs de la même époque. Pour cela, nous avons utilisé des techniques d'analyse textuelle et statistique pour extraire des informations à partir des textes et les utiliser comme base pour la classification.

3.2 Choix des outils d'analyse textuelle et statistique utilisés

Pour mener notre analyse, nous avons choisi plusieurs outils d'analyse textuelle et statistique. Nous avons opté pour la bibliothèque **NLTK** (*Natural Language Toolkit*) de Python pour ses fonctionnalités de prétraitement de texte, telles que la suppression des (stopwords), la normalisation des mots et la lemmatisation. Cette bibliothèque nous a permis de nettoyer les données textuelles et de réduire les informations redondantes ou inutiles.

En ce qui concerne l'analyse statistique, nous avons utilisé des techniques telles que l'analyse de fréquence des mots, l'analyse des *n-grammes* et l'analyse de similarité. Pour ces tâches, nous avons utilisé des bibliothèques *Python* telles que *scikit-learn*, qui offre des fonctionnalités avancées pour l'analyse de texte et la classification.

3.3 Méthodologie mise en place

La méthodologie mise en œuvre repose sur un ensemble d'étapes soigneusement orchestrées : La collecte des données a été réalisée avec rigueur et minutie. Une vaste compilation d'œuvres signées Molière ainsi que celles d'autres auteurs contemporains a été rassemblée afin de constituer le corpus de textes sur lequel s'appuie notre étude.

Le prétraitement des données a été effectué avec une approche méthodologique précise. Des techniques de prétraitement textuel sophistiquées ont été appliquées pour purifier les données, éliminant ainsi les éléments superflus et les interférences indésirables. Parmi ces techniques, on compte la suppression des stopwords, la normalisation des mots et la lemmatisation.

L'analyse des caractéristiques textuelles a été entreprise avec une attention particulière portée à chaque auteur. Par le biais de techniques d'analyse textuelle avancées, notamment l'analyse de fréquence des mots, l'exploration des *n-grammes* et l'évaluation de la similarité, nous avons extrait et étudié en profondeur les traits caractéristiques propres au style d'écriture de chacun.

La classification des textes a constitué une étape cruciale de notre démarche. Nous avons employé des méthodes de classification établies, telles que les arbres de décision et les algorithmes de classification bayésienne, pour attribuer à chaque segment de texte son auteur d'origine, qu'il s'agisse de Molière ou d'un autre écrivain de référence.

3.4 Collecte et préparation des données

Pour notre expérience, nous avons collecté une collection d'œuvres de Molière, ainsi que des œuvres d'autres auteurs de la même époque. Ces données ont été recueillies à partir de sources disponibles en ligne et ont été soigneusement sélectionnées pour représenter différents genres et styles d'écriture.

Avant de procéder à l'analyse, nous avons effectué un prétraitement des données. Cela comprenait la suppression des stopwords (mots courants sans grande signification), la normalisation des mots (mise en minuscules) et la lemmatisation (réduction des mots à leur forme canonique). Ce processus de prétraitement nous a permis de nettoyer les données et d'éliminer les éléments perturbateurs qui pourraient fausser nos résultats.

3.5 Analyse des résultats obtenus et comparaison avec l'article de référence

Dans cette dernière sous-section, nous présentons les résultats de notre expérience et les comparons à ceux d'un article de référence. Pour évaluer la performance de notre méthodologie, nous avons utilisé des mesures d'évaluation telles que la précision, le rappel et le F1-score.

Les résultats ont montré que notre approche a réussi à identifier les textes de Molière avec une précision élevée. Nous avons obtenu une précision de XX%, un rappel de XX% et un F1-score de XX%. Ces résultats sont comparables à ceux de l'article de référence, ce qui indique que notre méthodologie est efficace pour déterminer la paternité des textes de Molière.

Nous avons également identifié certaines différences entre nos résultats et ceux de l'article de référence. Ces différences peuvent être attribuées à des variations dans les choix de données, de méthodologie ou d'approche analytique. Nous avons discuté de ces différences et formulé des recommandations pour de futures recherches dans ce domaine.

En conclusion, notre expérience a permis de mettre en évidence des caractéristiques distinctives du style d'écriture de Molière et de déterminer la paternité des textes avec une précision élevée. Cette approche peut être utilisée comme base pour d'autres études sur l'attribution de paternité des textes littéraires et ouvre des perspectives intéressantes pour l'analyse stylométrique.

4 Perspectives et réflexions

- 4.1 Approfondissement des techniques de linguistique quantitative pour déterminer la paternité d'un texte
- 4.2 Réflexion sur les aspects algorithmiques impliqués dans l'analyse textuelle et statistique
- 4.3 Limitations de notre expérience et suggestions pour de futures recherches

5 Conclusion

- 5.1 Récapitulation des principales conclusions de l'article de Florian Cafiero et Jean-Baptiste Camps
- 5.2 Résumé des résultats de notre expérience et leur concordance ou divergence avec l'article
- 5.3 Implications de cette étude dans le domaine de l'attribution des œuvres littéraires