

# **Shakespeare Authorship problem**

Shanshan Liu S5336651

## **Introduction**

Mainstream Shakespeare scholars have no debate on his identity, the William Shakespeare of Stratford-upon-Avon, but since the 16th century there has been a voice that Shakespeare was a pseudonym. The Shakespeare authorship debate is the contention that works attributed to WShakespeare were written by someone other than Shakespeare. Whether there was “true author” behind this literary master has always been one of the focal points of debate in the literary world, and many people do not believe that an English squire could write many monumental works. As early as centuries ago, questions were raised because Shakespeare was a world-renowned playwright who had many works celebrated by future generations, but there are still many unknown aspects of his life, and besides, he did not personally leave such texts. Thus, there is very little material about Shakespeare's life, which makes Shakespeare a veil of mystery.

Stylometry has been used to address a variety of questions in the field of literary scholarship, including authorship attribution, genre classification, and the study of literary influence. By analyzing linguistic patterns in texts, stylometry can provide valuable insights into the authorship of anonymous or disputed works.

In the case of the Shakespeare authorship problem, stylometry has been used to compare linguistic features in the plays and poems attributed to Shakespeare with those of other writers of the time. Through this analysis, researchers have been able to identify patterns of language use that are characteristic of Shakespeare's work, and to distinguish these from patterns found in the works of other writers.

Overall, I have chosen to work with stylometry in my research on the Shakespeare authorship problem because of its potential to shed new light on this longstanding question in literary scholarship. By providing a quantitative analysis of linguistic patterns in the texts, stylometry can help to support or refute existing theories about Shakespeare's authorship, and to uncover new insights into the language and literary style of this iconic figure in English literature.

## **Literature review**

The article titled "Questioning Authorship: Attribution and Beyond" has inspired me to make a relevant research. It explores the challenges and complexities of determining authorship, specifically in the context of literary works. The article highlights

the ongoing debate surrounding the authorship of Shakespeare's plays and the controversies that have arisen due to the lack of conclusive evidence.

The article discusses various methods that have been employed to determine authorship, including stylometry, linguistic analysis, and historical context. It also explores the limitations and potential pitfalls of these methods and emphasizes the need for a multi-disciplinary approach that incorporates a range of techniques and perspectives. Furthermore, the article discusses the broader implications of the authorship debate, including the significance of authorship in the interpretation and reception of literary works, and the relationship between authorship, identity, and power.

Based on this article, I make the following research report.

## **Hypothesis**

There are many hypotheses of this Shakespeare authorship question: some support the true author as Francis Bacon, Christopher Marlowe, Edward de Vere, illegitimate son of Queen Victoria, even Queen Victoria herself (Leigh et al., 2019).

The Bacon hypothesis of Shakespeare authorship asserts that the plays attributed to William Shakespeare were written by Sir Francis Bacon, a philosopher, writer, and physicist. Various theories have been advanced for this supposed deception, the most popular being that Bacon's climb to high rank would have been hampered if it became known that he authored plays for the public theater. The theory was first put forth in the mid-nineteenth century, based on perceived correspondences between the philosophical ideas found in Bacon's writings and the works of Shakespeare. (Leigh et al., 2019).

The Marlowe theory: In 1819, an anonymous writer for *The Monthly Review*, or *Literary Journal*, argued that 'Christopher Marlowe' was a pseudonym used by Shakespeare for a while. In the late nineteenth century, scholars studying contemporaneous playwrights: Christopher Marlowe and Shakespeare. They found a subtle consistency in their diction, style, and point of view, and came up with the hypothesis that the two were the same person. (Leigh et al., 2019).

The De Vere theory: as earl of Oxford, many of Shakespeare's plays resemble the life of de Vere, who was well educated and social enough to write at length about the classics, the law, foreign countries, and languages, and who traveled frequently, spending most of 1575 in Italy, while 14 of Shakespeare's plays have Italian settings, and seem more convincing than an English country squire. (Looney, 2013d)

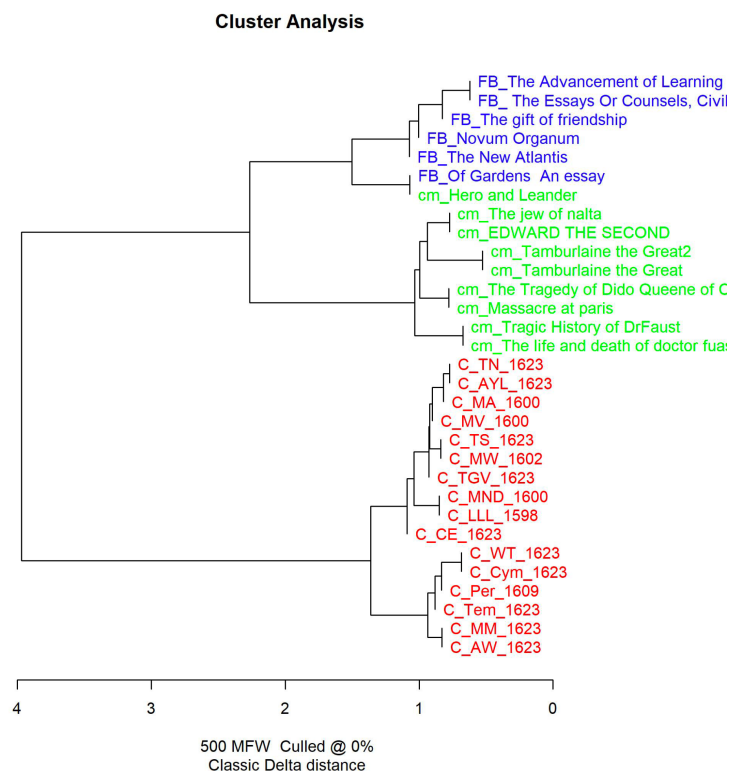
## **Method**

All the evidence suggests that the most suspicious person may be among these Three. In order to testify the hypothesis, I downloaded all the books contributed by the four writers (including Shakespeare) in the format of text from the Gutenberg project as corpus. However, it was difficult to find anything of Edward de Vere except the name, so

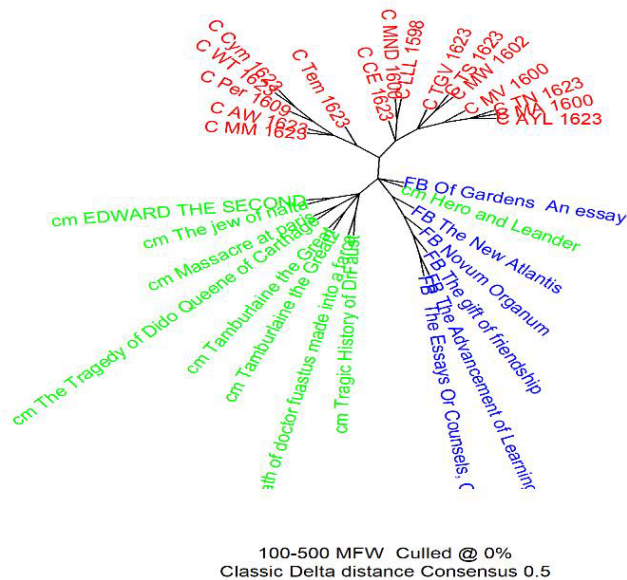
I changed it to compare the three authors by stylo with 31 Shakespeare works, 9 Marlowe works and 6 Bacon works.

The texts of Christopher Marlowe, Francis Bacon, and Shakespeare's works are compared and analyzed by using stylo, where the text beginning with 'cm\_' is Christopher Marlowe's work, the text beginning with 'FB\_' is Francis Bacon's work, and the text beginning with 'C\_' is Shakespeare's work. The text beginning with 'FB\_' is the work of Francis Bacon, and the text beginning with 'C\_' is the work of Shakespeare. All the text is read with the `load.corpus` function, and then the xml data is processed with the `delete.markup` function. Then use the `txt.to.words.ext` function to split the text sentence into words and convert them all to lowercase. Then the `delete.stop.words` function is used to remove deactivated words to avoid the impact of common words on the text analysis. The 3000 high-frequency words are extracted using the `make.frequency.list` function.

Based on the previous processing, the following graph was generated using cluster analysis, setting the minimum value of MFW to 300 and the maximum value to 1000. From the graph, it can be seen that Francis Bacon and Christopher Marlowe's works are close, and the words used in Christopher Marlowe's *Hero and Leander* and *The Jew of Malta* are similar to Francis Bacon's *Of Gardens*. *An Essay* has a higher degree of wording similarity.

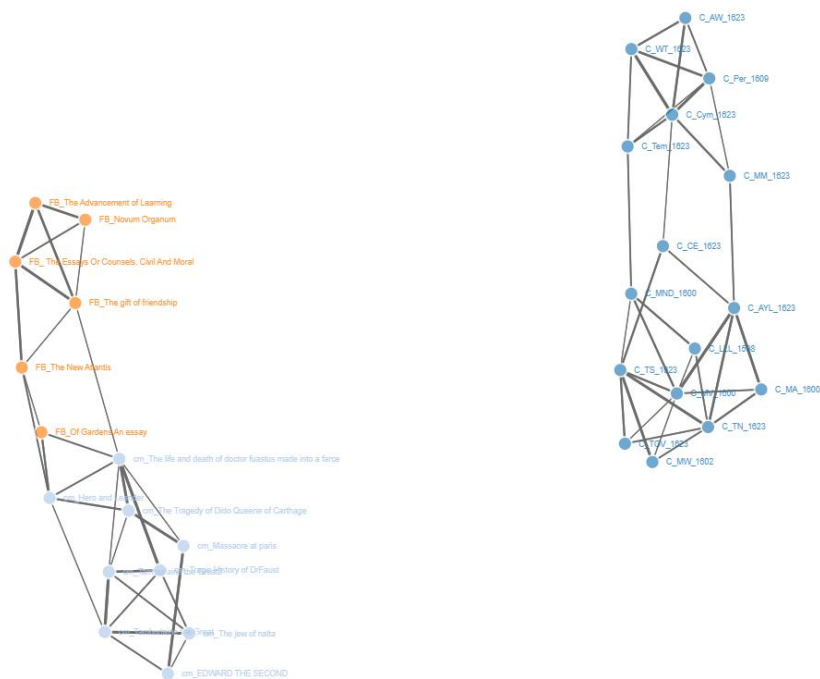


### Bootstrap Consensus Tree

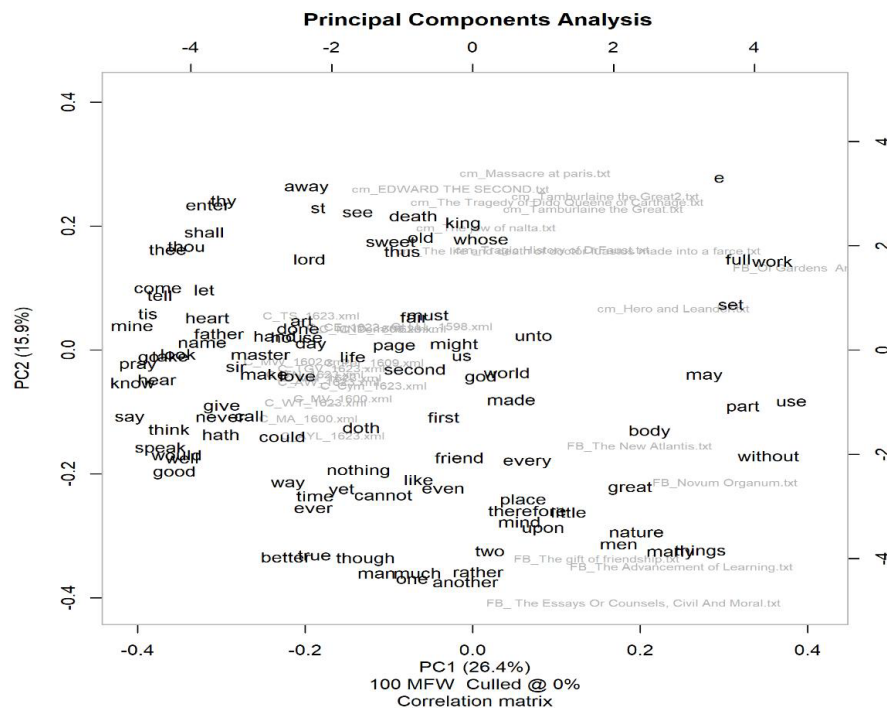


Changing the minimum value of MFW to 100 and the maximum value to 500, the tree diagram shown in the figure below was obtained. From the figure, we can see that the works of the three authors belong to three branches, Hero and Leander and Of Gardens An essay are closer, and their clustering effect is basically consistent with the previous clustering analysis results.

The degree of relevance of each text is represented as a network graph using the `stylo.network` function, setting both the minimum and maximum MFW values to 300, and obtaining the results shown below. Once again, it is confirmed that the works of Francis Bacon and Christopher Marlowe are highly correlated and distant from Shakespeare's works.



The high-frequency words commonly used by each author were analyzed by the method of Principal Components Analysis. The high frequency words used by Francis Bacon are body, great, nature, men, another, rather, etc. The high frequency words for Christopher Marlowe are full, work, set, king, sweet, etc. and for Shakespeare they are art, master, think, life, love, father, etc.



## Conclusion

The Stylo's genre analysis employed in my study was unable to establish with certainty that Francis Bacon or Christopher Marlowe could be the genuine Shakespearean author. It is worth noting that the Stylo tool has limitations, as it provides only a rough comparison among texts without a more robust method such as Burrows Delta. Moreover, it does not offer a detailed account of the specific similarities and differences between the texts. Nonetheless, the study revealed a remarkable resemblance in the writing styles of Marlowe and Bacon, which could explain why they are both prominent candidates in the Shakespeare authorship debate, and awaits further investigation.

## Reference

- Burrows, J. (2003). Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York. *Computers and the Humanities*, 37(1), 5–32. <http://www.jstor.org/stable/30204877>
- Leigh, R. J., Casson, J. J., & Ewald, D. (2019). A Scientific Approach to the Shakespeare Authorship Question. *SAGE Open*. <https://doi.org/10.1177/2158244018823465>
- Looney, J. T. (2013d). *Shakespeare Identified in Edward de Vere, the Seventeenth Earl of Oxford*. Theclassics.Us.