jcls. Journal of Computational Literary Studies

OPEN ACCESS

Article

A Stylometric Analysis of Seneca's Disputed Plays

Authorship Verification of *Octavia* and *Hercules Oetaeus*

Paschalis Agapitos[1]
Andreas van Cranenburgh[2]

1. Donostia International Physics Center, P. M. de Lardizabal 4, Donostia/San Sebastian, Spain.
2. Computational Linguistics Department, University of Groningen, Groningen, The Netherlands.

Citation
Paschalis Agapitos and Andreas van Cranenburgh (2024). "A Stylometric Analysis of Seneca's Disputed Plays. Authorship Verification of Octavia and Hercules Oetaeus". In: *Journal of Computational Literary Studies* 3 (1). 10.48694/jcls.3919

Date published 2024-11-14
Date accepted 2024-09-29
Date received 2024-01-22

Keywords
Seneca, stylometry, authorship verification, Latin, stylo

Reviewers
Jacques Savoy, Christof Schöch (Extended Code Reviewer), Paul Vierthaler, Anonymous Reviewer

Note
This paper has passed through the conference track of JCLS. In addition to being peer reviewed, it was presented and discussed at the 3rd Annual Conference of Computational Literary Studies at Vienna, Austria, in June 2024. In addition to the regular peer review, this paper has successfully passed the extended data and code review.

Abstract. Seneca's authorship of *Octavia* and *Hercules Oetaeus* is disputed. This study employs established computational stylometry methods based on character n-gram frequencies to investigate this case. Based on a Principal Component Analysis (PCA) of stylistic similarities within the Senecan corpus, *Octavia* and *Phoenissae* emerge as outliers, while *Hercules Oetaeus* only stands out when the text is split in half. Subsequently, applying Bootstrap Consensus Trees (BCT) to a corpus of distractor texts, both disputed plays align with the Senecan cluster/branch. The General Imposters method confidently reports Seneca as the author of the disputed plays under various scenarios. However, upon closer examination of text segments, indications of mixed authorship arise. Based on computational stylometry, it appears that the disputed plays were in large part, but not wholly, written by Seneca.

1. Introduction

Computational stylometry is a quantitative text analysis method mostly concerned with authorship attribution and authorship verification problems. Authorship attribution involves identifying the most likely author of a disputed document from a given set of candidates (Koppel et al. 2007, 1261), whereas authorship verification concerns the question of whether an author wrote a disputed document (Koppel et al. 2007, 1261; Juola 2015, i106). The verification task is more challenging than the attribution task because the former task involves determining whether an observed similarity in style is sufficient to verify authorship, while the attribution task merely involves picking the most similar author from the given candidates (Potha and Stamatatos 2017, 138). It is important to also note that the authorship verification typically involves both close-set and open-set scenarios. In the close-set scenario, the suspected author is one of the candidates provided, whereas in the open-set scenario, the true author may not be among the known candidates.

The main assumption behind computational stylometry is that certain words are chosen unconsciously by the writer, which form a unique, individual fingerprint of an author (Evert et al. 2017, ii4). Since these words are predominantly function words that are used in a way that is hard for the author to control, imitating someone else's writing style is difficult for an imposter. In other words, there is an "immutable signal that authors emit involuntarily” (Päpcke et al. 2022, 1). The utility of function words in traditional and computational stylometric studies can be condensed into four points: 1) richer dataset because of their high frequency, 2) closeness of the set since function words are limited and fixed, 3) content-independent, and, as mentioned above, 4) unconscious use of them due to their high frequency (Kestemont 2014, 60; Beullens et al. 2024, 393–394).

The aim of this article is to examine whether Seneca the Younger wrote *Octavia* and/or *Hercules Oetaeus* (henceforward: *Oct.* and *H.O.*, respectively) since they are both tragedies of which a plethora of literary scholars have raised concerns about their attribution to Seneca. We contribute to the debate on Seneca’s disputed texts by applying a variety of computational stylistic methods and testing several scenarios. For this we use the *stylo* software, an R package created and developed by Eder et al. (2016).

The ensuing sections of this study are organized as follows. Initially, a concise literature review is provided addressing *Oct.* and *H.O.* (section 2). Subsequently, section 3 outlines the rationale for selecting a specific set of imposter texts and acknowledges potential limitations associated with the limited transmission of ancient texts and differences in genre and meter. Section 4 delves into the preprocessing steps and features employed in the study, while also offering a brief explanation of each method utilized in the primary analysis. Section 5 provides a validation of the methods on texts with known authorship. Section 6 presents the main results for the disputed texts and engages in a discussion of these findings. Finally, we present our conclusions concerning the findings and outline ideas for future research (section 7).

2. Literature Review

2.1 Non-quantitative Approaches

The disputed texts considered in this article, *Oct.* and *H.O.*, are Latin tragedies. *Oct.* is the only *fabula praetexta* (i.e., an ancient Roman tragedy with a Roman historical subject) that has survived from the corpus of Latin dramas until today (Ferri 2003, 1), whereas *H.O.* is a *fabula crepidata*, an ancient Roman tragedy with a Greek subject.[1]

A lot of arguments have been made over the years by literary scholars to support the idea that Seneca’s stylus could not have written *Oct.* According to Philp (1968, 151–153), the principal manuscript traditions for the Senecan tragedies are the traditions E and A as well as some excerpts and fragments. The A recension is the only one that transmits *Oct.* (Philp 1968, 151; Seneca 2008, 78). Based on the fact that the interest in Senecan tragedies increased at the beginning of the thirteenth century, there is the hypothesis that *Oct.* was included in the A recension at this time (Gahan 1985; Ferri 2014, 525). Moreover, in both recensions, the texts are given in a different order (Marti 1945, 220).[2] According to Ferri (2003, 31), the resemblance that *Oct.* bears with the other Senecan plays and the fact that Seneca “participates” as a persona in the play might have been the reason for classifying *Oct.* as a Senecan play.

[1] It should be noted that extant *fabulae crepidatae* are attributed to Seneca’s stylus.

[2] The manuscript tradition E saves the Senecan plays in the following order: *Hercules* (*Furens*), *Troades*, *Phoenissae*, *Medea*, *Phaedra*, *Oedipus*, *Agamemnon*, *Thyestes*, *Hercules* (*Oetaeus*); *Octavia* is omitted in tradition E. The manuscript tradition A gives the Senecan plays in the following order: *Hercules furens*, *Thyestes*, *Thebais*, *Hippolytus*, *Oedipus*, *Troades Medea*, *Agamemnon*, *Octavia*, *Hercules Oetaeus*. The order of the plays and their names follow Philp (1968, 151).

Concerning the stylistic aspect of *Oct.*, the same words are repeated a lot, and some poetic phrases seem artificial rather than the inspiration of the author; in other words, a weakening of the literary power is observed (Herington 1961, 24). Even though in the original Senecan plays the rhetorical style of Ovid was a major influence, the author of *Oct.* seems not to care about this aspect (Michalopoulos 2020). Moreover, Carbone (1977, 56) argues that it had been impossible for Seneca to know details about events that took place after his death with such great precision (e.g., the death of emperor Nero). Poe (1989, 435) suggests that *Oct.* is not Seneca's genuine work, but the product of an imitator with limited literary experience and a low level of creativity when it comes to the provision of conclusions among the scenes.

The text *H.O.* also raises some concerns about the attribution of its authorship. As Marshall (2014, 40) points out, referring to Nisbet (1995), the play follows a different approach to play-writing. For example, the length of this tragedy is twice as long as Seneca's other plays, which makes it the longest extant drama to survive from antiquity (Boyle 2009, 220; Star 2015, 255).

However, it has also been argued that *Oct.* and *H.O.* indeed carry the authorial fingerprint of Seneca. Concerning *Oct.*, in lines 619–621, Agrippina lists some traditional punishments in an effort to predict the tyrant's (i.e., Nero's) imminent death (Seneca [1921] 2007, *Oct.* 619–621). In this passage, the demise of Nero appears to be foretold, which seems to rule out Seneca as the author. Nevertheless, some scholars argue that the description of the punishments is not even close to what actually happened to Nero and that it should not be taken as a prophecy that requires knowledge of the historical event of the death of Nero, since the punishments described represent common punishments in mythology (Pease 1920, 390–391).

Furthermore, Pease (1920, 390) supports the idea that the public circulation of *Oct.* is a posthumous event, and that Seneca entrusted the manuscript of the play to friends for publication after the death of Nero. This argument – merely a speculation, since no additional evidence exists – can explain the inconsistencies in the text that scholars have used to argue that *Oct.* is not a Senecan play. Following the line of thought of this argument, someone could hypothesize that Seneca is the author of the play but an editor or a ghost author added or edited some segments of *Oct*.

With respect to *H.O.*, the argument of the late composition is also used to support *H.O.* as a genuine Senecan play (Rozelaar (1985); Nisbet (1995, 209–212) as cited in Marshall (2014, 40)). If *H.O.* was one of the last tragedies written by Seneca the Younger before his death, this could explain the haste and anomalies that might have caused the sheer length of the play in its current form.

2.2 Quantitative Approaches

There is a plethora of papers that apply computational stylistics to Latin texts, therefore the study of the authorial fingerprint of ancient Latin texts is not new (e.g., Kestemont et al. 2016; Stover et al. 2016; Stover and Kestemont 2016). However, the number of papers that consider Senecan texts is much smaller, and even smaller are those that actually consider the authenticity of the two disputed Senecan plays, *Oct.* and *H.O. per se*.

Brofos et al. (2014) use a machine learning model trained to recognize texts as Senecan or not, namely a "one-class SVM (i.e., Support Vector Machine) with functional n-gram probability features"[3]. The model predicts that *Oct.* and *H.O.* were not written by Seneca the Younger (Brofos et al. 2014, 8–9). Yet, as expected, their model also makes many misclassifications. I.e., it classifies some Senecan texts as non-Senecan, and when the model is augmented with prose texts in addition to tragedies, other authors are also classified as Senecan (Brofos et al. 2014, 9).

Nolden (2019) examines the authorship of *Oct.* and *H.O.* with a variety of computational stylistic techniques. He starts with the hypothesis that *Oct.* and *H.O.* were probably not written by Seneca and evaluates various methods in this light, including type-token ratio, compressibility, and dimensionality reduction. The results present a mixed picture: Some methods point to high similarity between all the ten plays attributed to Seneca (including the disputed ones), while other methods point to *H.O.*, but also *Phoenissae*, as outliers. However, since *Phoenissae* is considered Senecan, this casts doubt on the reliability of these methods. In the end, no strong conclusions can be drawn, as the differences are small and it is not certain whether the mixed results should be explained as unsuitability of particular methods or uncertainty about Seneca's authorship.

In addition, Gómez Caballero (2021), performing cluster analysis using a simple dendrogram, bootstrap consensus tree, and multidimensional scaling, found that *Oct.* and *H.O.* were actually products of the stylus of Seneca the Younger without, however, evaluating the methods and delving into the hypothesis of mixed authorship.

Lastly, it is worth mentioning the paper by Cantaluppi and Passarotti (2015). Even though the main aim of their paper is to cluster the works of Seneca and to show that certain statistical methods can be effective at detecting the genre of the text, their insights are useful for some of the limitations of the methods used in authorship attribution studies and the current study as well (e.g., Principal Component Analysis). For instance, they perform their analysis using the full size of the text, and as they show the Principal Component Analysis method can be affected by the topic and the genre of the text (see the clustering and the words that appear next to the filenames in Cantaluppi and Passarotti (2015)).

2.3 Literature Review Conclusion

In conclusion, "the language and style of these two tragedies [*Oct.* and *H.O.*], however, are identical to the language and style of the others; that is why the discussion of whether these two tragedies are genuine has not yet ceased" (Marshall 2014, 74). Moreover, both of the disputed plays can be considered tricky cases because of the small number of extant Roman tragedies and the fact that *Oct.* has no equivalent extant tragedy in its genre. Previous computational approaches seem to design the experiments hastily, not taking into account multiple variables connected to the texts *per se* or considering these works as non-Senecan and focusing on the evaluation of authorship attribution/verification methods and software. To fill this research gap, this paper takes into account as many

[3] An SVM is a supervised learning algorithm used for classification and regression tasks. It draws a line or a plane that maximizes the space between the data points, in our case the texts. It works both in linear (data points can be separated by a straight line) and non-linear (data points cannot be separated by a straight line) high-dimensional environments.

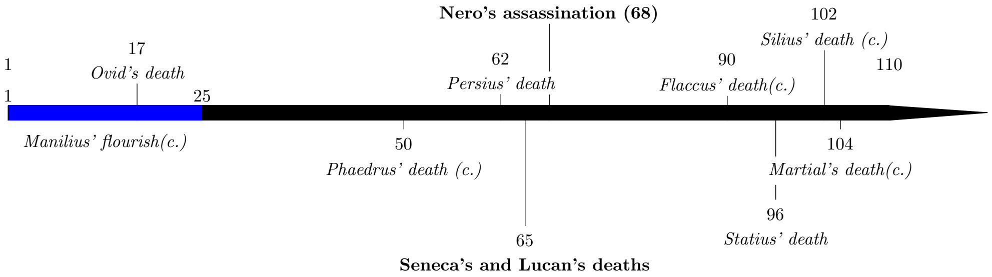


Figure 1: A timeline of the authors used in the main dataset. This dataset is used for the PCA, BCT, and the first four out of five scenarios of the GI method.

variables as possible, validates the computational methods before it applies them to texts, and uses the evaluated methods to contribute and shed new light on the arguments surrounding the authorship of the disputed tragedies. The main research question will be as follows: Were *Oct.* and *H.O.* written by Seneca the Younger or are they, at least in their present form, the product of an imitator or mixed authorship?

3. Dataset

The main dataset employed in this study comprises distractor authors and verse texts that slightly precede and follow the era of Seneca the Younger (ca. 4 BCE–65 CE). In the context of computational stylometric approaches, a distractor author, or 'imposter', is utilized for comparison with a disputed text. For clarity, consider a text X attributed to author A, with distractor authors B, C, and D, known not to be the author of X. The soundness of a stylometric method is affirmed by observing significantly higher similarity between X and other texts by A compared to B, C, and D, confirming A as the probable true author, or vice versa.

In our analysis of Seneca, the dataset includes authors such as Ovid, Manilius, Martial, Phaedrus, Persius, Lucan, Valerius Flaccus, Statius, and Silius Italicus (see Table 10). These authors, broadly associated with the literature of the Early Empire, wrote within the first century of the Common Era (see Figure 1).[4]

In Scenario 5 presented in Table 3 (section 6), we augment the dataset used by Kestemont et al. (2016) (see Table 9 in Appendix A) with our main corpus (see Appendix A for the main corpus and Figure 2 for a visual representation of the augmented dataset used in Scenario 5 of GI). Therefore, we consider it important to explain the authors and the texts that populate this dataset, as well as their main genre. Kestemont's dataset contains 1,850 non-overlapping text slices. The authors and the texts present in their dataset are listed in Table 9 in Appendix A. Since in their paper, they compare their corpus with Caesar's writings, their dataset contains mostly historiographical texts and covers a huge time span (from the 4th century B.C.E. up to the 4th century C.E.).

In authorship verification, the challenge of text and author selection inevitably involves some arbitrary or imperfect choices. This section aims to transparently justify our choices.

[4] Karakasis (2018) suggests Titus Calpurnius Siculus's connection to the reign of Nero, placing him within the Neronian literature. Due to the ongoing debate on Siculus's inclusion in this category, we exclude him from our dataset. In addition to that, we choose to omit the text that is called *Aratea* since there is an ongoing debate about its authorship (Baldwin 1981; Possanza 2003, 217–243).

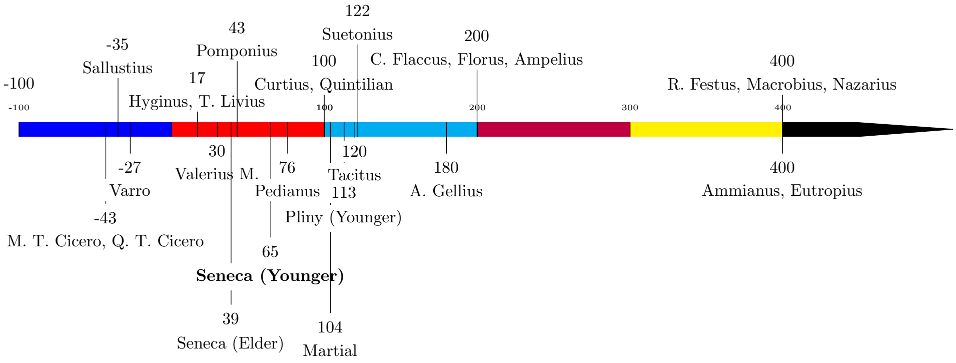


Figure 2: A timeline of the authors used in Kestemont et al. (2016)'s data augmented with our main dataset. This dataset is used in scenario 5 of the GI method.

According to Grieve (2007, 255), texts, whether disputed or not, are inherently tied to their historical era. Consequently, the dataset is designed to narrow the temporal scope to ensure a more focused linguistic comparison. However, we should highlight two important aspects that complicate the corpus selection.

First, besides the Senecan tragedies, there are no other extant Roman tragedies. Therefore, expanding the timeline is difficult in our case without at the same time increasing the linguistic variation and adding many different genres. Thus, our focus is to run most of the experiments using texts that temporarily are located relatively close to the era of Seneca the Younger and the same kind (in verse).[5] Second, there is the issue of the varying meter across the texts (e.g., iambic vs. hexametric), which constrains the vocabulary available to the author. For computational stylometry, different vocabulary means different features, and therefore dissimilarity between texts. While we cannot completely resolve this issue, we believe that we can limit its influence by considering patterns of very frequent character sequences rather than whole words (see subsection 4.1). In addition to that, prior work on cross-genre and cross-topic stylometry has empirically shown that character-based authorship attribution is robust to such variation (e.g., Stamatatos 2013, 343). On the one hand, it may be that this robustness also applies to the genre and meter variation in our case. On the other hand, it must be noted that since the disputed plays are compared to Senecan texts in the same genre and meter, while the imposter texts are in a different genre and meter, the likelihood of attributing the disputed plays to Seneca may be increased.

Table 10 in Appendix A provides a complete list of authors and texts included in the dataset variations used for each experiment. All works were obtained from the Perseus Digital Library (Crane 2024) except for Manilius' *Astronomica* due to unavailability in the source.[6] Thus, *Astronomica* was sourced from The Latin Library (Carey 2024).[7]

[5] We test one scenario adding historiographical texts in prose that span from the 4th century B.C.E. up to the 4th century of C.E. (see the description above about the dataset by Kestemont et al. (2016).
[6] Available at: `https://github.com/cltk/lat_text_perseus`.
[7] Available at: `https://github.com/cltk/lat_text_latin_library`.

ea	eae	eam	earum	eas	ego
ei	eis	eius	eo	eorum	eos
eum	id	illa	illae	illam	illarum
illas	ille	illi	illis	illius	illo
illorum	illos	illud	illum	is	me
mea	meae	meam	mearum	meas	mei
meis	meo	meos	meorum	meum	meus
mihi	nobis	nos	noster	nostra	nostrae
nostram	nostrarum	nostras	nostri	nostris	nostro
nostros	nostrorum	nostrum	sua	suae	suam
suarum	suas	sui	suis	suo	suos
suorum	suum	suus	te	tibi	tu
tua	tuae	tuam	tuarum	tuas	tui
tuis	tuo	tuos	tuorum	tuum	tuus
vester	vestra	vestrae	vestram	vestrarum	vestras
vestri	vestris	vestro	vestros	vestrorum	vestrum
vos	vobis				

Table 1: A list of the 98 inflectional forms of 13 pronouns removed from every text of the corpus as provided by the software *stylo* (Eder et al. 2016).

4. Feature Selection and Methods

The dataset was preprocessed and analyzed using the R package *stylo* (Eder et al. 2016) and *The Classical Language Toolkit* (CLTK) (Johnson et al. 2021).

4.1 Preprocessing and Feature Selection

The texts were initially tokenized with consideration for the non-differentiation of the letters ‘v’ and ‘u’ in certain text editions. To ensure orthographic consistency, ‘v’ was uniformly converted to ‘u’ where applicable. Pronoun-culling (i.e., eliminating personal pronouns from the text) was then applied to automatically remove frequency information primarily associated with personal pronouns (see Table 1 for the list of removed pronouns). This step aims to mitigate the impact of genre, topic, author’s gender, and narrative perspective on the analysis (Hoover 2004, 480; Newman et al. 2008, 233; Kestemont et al. 2015, 206). Given the varied meter of the texts, even within works by the same author, this approach reduces the ‘noise’ in texts due to the topic or the gender of the author. Both orthographic normalization and pronoun-culling followed the predefined steps of *stylo* (Eder et al. 2016, 110), with details on the pronoun-culling process outlined in Table 1. Lastly, to enhance the performance of every approach, we iterate over different feature sizes (from 100 to 2,000 MFCs).

The extraction of relevant features in our study involves character 4-grams, a choice proven effective in cross-genre and cross-topic authorship attribution (Koppel et al. 2009, 12–13; Stamatatos 2009, 541–542; Eder 2011, 110; Stamatatos 2013).[8] Despite appearing initially inconsequential, character n-grams, particularly of size 4, excel in capturing sub-word level information, including case endings and morphemes (Kestemont 2014, 62–64). In the context of Latin’s highly inflected nature, character n-grams preserve details from lower frequency words such as prepositions and determiners (Kestemont 2014, 60–61). Notably, the use of character n-grams eliminates the need

[8] For a very simple and informative definition of n-grams see Hagiwara (2021, 53–54).

for word lemmatization or other normalization, as these features (character n-grams) operate below the word level and are language-independent (Daelemans 2013, 4; Kestemont et al. 2015, 206). This approach, utilizing plain inflected surface tokens, has demonstrated increased stability compared to lemma/stem-based methods (Stover and Kestemont 2016). Slicing words into 4-character packages enhances observations, striking a balance between sparseness and information content (Daelemans 2013, 4–5). In general, character n-grams represent a widely adopted and reliable feature type in stylometry (Stamatatos 2009, 541–542; Stamatatos 2013, 432–433; Eder 2011, 112). In the rest of this paper, we will use the frequencies of the Most Frequent Character (MFC) n-grams. For example, '2,000 MFC' refers to the frequencies of the 2,000 most common character n-grams.

1) que_	2) _et_	3) ere_	4) _in_	5) _qua_
6) ibus_	7) sque_	8) _qu_	9) _bus_	10) usa_
11) _tus_	12) mque_	13) _tis_	14) _qui_	15) pro_
16) per_	17) sin_	18) quo_	19) con_	20) non_

Table 2: Most frequent character 4-grams of the entire corpus (underscores represent whitespaces).

4.2 Methods

All of the methods we employ estimate the stylistic similarity of texts as the distance between their features (i.e., character n-gram frequencies). For this, we pick the Cosine Delta distance metric because of its effectiveness in various test conditions and its particular effectiveness for inflected languages (Jannidis et al. 2015, 6–8; Evert et al. 2017, ii9–ii10; Eder 2022). Both the validation and main analysis phases utilize the 2,000 Most Frequent Character 4-grams (MFCs), a selection supported by studies indicating the performance of the Cosine Delta plateaus at this threshold for texts in Latin (Jannidis et al. 2015, 6–8; Evert et al. 2017, ii9–ii10).

In general, more MFCs lead to better performance since the features capture more stylistic variation. However, beyond the 2,000 MFCs, the character n-grams become rarer and are therefore not as informative. Therefore, we consider this point as adequate to capture the necessary amount of authorial fingerprint (Jannidis et al. 2015; Evert et al. 2017; Eder 2022). The frequency distribution plot (see Figure 3) illustrates this diminishing informativeness beyond the 2,000th character 4-gram.

The study employs two exploratory analysis methods and one authorship verification method, presented in ascending order of robustness. Firstly, Principal Component Analysis (PCA) is applied. Secondly, the Bootstrap Consensus Tree (BCT) is introduced, followed by the General Imposters (GI) method, each briefly outlined in the subsequent sections.

4.2.1 Principal Component Analysis

PCA, a widely used unsupervised algorithm in authorship attribution and verification studies, reduces dimensionality by identifying principal components (eigenvectors) that explain feature variation. In this context, dimensionality refers to the number of features or variables initially present in the dataset (in our case the features that are generated

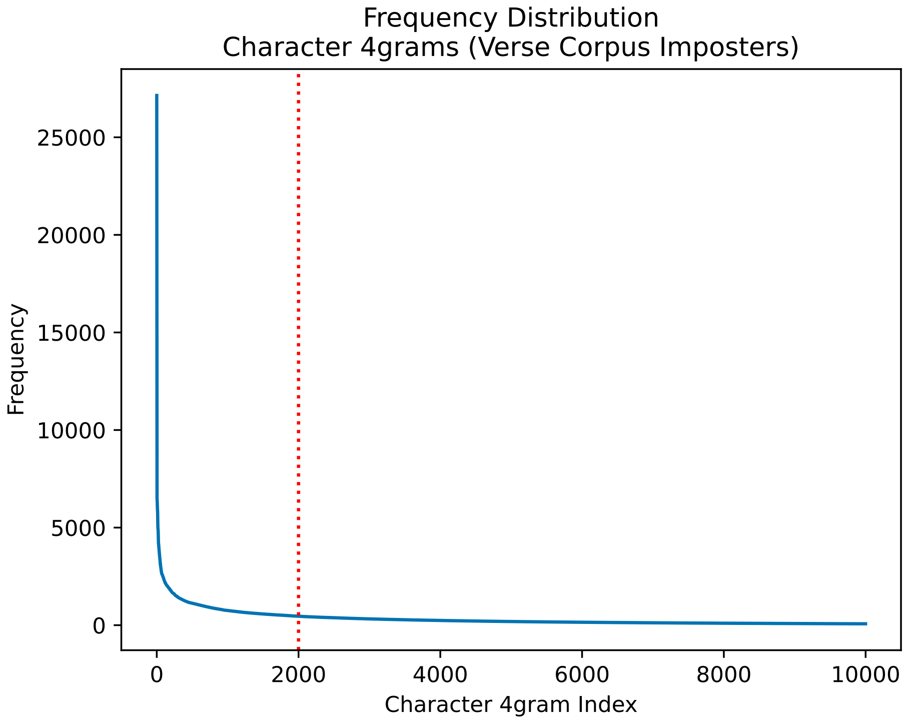


Figure 3: Frequency distribution of the character 4-grams in the whole corpus (i.e., 90 texts including the disputed plays). The vertical line is set to 2,000 to show that characters 4-grams after this threshold start to become quite infrequent. The result is what we expect to see since the distribution of the frequency of features in a given text follows Zipf’s law (the frequency *f* of a feature is inversely proportional to its rank *r*).

by character n-grams). PCA helps reduce this dimensionality by transforming the data into a new set of variables, where each successive variable captures less and less of the total variance in the data. To preserve maximal data variance, PCA zeroes out smaller principal components, employing only those capturing the highest variance (VanderPlas 2017, 436). These components position texts in a two-dimensional visualization, enhancing readability for human interpretation but at the same time losing some of the variation information (Stamatatos 2009, 545). Similarity in frequency distribution correlates with spatial proximity in the PCA plot, indicating text dissimilarity based on vector dissimilarity. Closeness may reflect temporal proximity, common genre, or shared authorship (Manousakis 2020, 171–172). Isolated data points suggest the opposite. Applied exclusively to the Senecan corpus, PCA results use a correlation matrix due to its invariance to linear changes in units of measurement, making it suitable for scaled variables like relative frequencies of character 4-grams (Jolliffe and Cadima 2016, 6). The correlation matrix accommodates the varied scale changes within the broad range of 100 to 2,000 Most Frequent Character 4-grams (MFCs).

4.2.2 Bootstrap Consensus Tree

While the Bootstrap Consensus Tree (BCT) originates from the field of phylogenetics, it was introduced as a method for computational stylometry by Eder (2012) and has since been increasingly used to identify authorial and translator fingerprints (Rybicki 2012; Rybicki and Heydel 2013). The fundamental idea behind bootstrapping is to randomly select a large number of samples with replacements. This process allows us to average the estimates of these samples, thereby enhancing the recurrence of patterns within a document (Jurafsky and Martin 2024, 75–77). Moreover, this method assumes that frequent patterns will reappear many times (robustness), but by increasing the number of iterations and using the consensus strength, we incorporate a larger and thus more diverse number of patterns within a single text (diversity). In other words, a higher number of samples guarantees a greater variety of patterns, making the results more representative of the population.

To clarify some of the concepts mentioned in the previous paragraph: Sampling with replacement involves sampling units returning to the data pool, allowing them to appear in multiple data 'snapshots.' This facilitates the identification of frequently occurring patterns, but also risks letting outliers excessively impact results. To balance the influence of outlier impact, a large number of iterations is usually preferred (Kuhn and Johnshon 2016, 72–73). In addition, another concept that is being implemented in our approach to further balance the impact of outliers is consensus strength. Consensus strength means that patterns present only in a certain percentage of iterations will be included in the final result. For instance, if we have a consensus strength of 0.5 (i.e., 50%), then only patterns that appeared in at least 50% of the iterations will be included. Unlike a simple dendrogram, a key advantage of BCT lies in its consensus strength, ensuring that more reliable relationships above a specified threshold will influence the final output. Parameters utilized include an MFC n-grams range from 100 to 2,000 with a step of 100, and a consensus strength set to 0.5.

4.2.3 General Imposters Method

The GI method, initially introduced by Koppel and Winter (2014), won first place in the PAN competitions for shared tasks in authorship verification for two consecutive years (i.e., 2013 and 2014) (Seidman 2013; Khonji and Iraqi 2014). Since then, it has proven effective in authenticating disputed writings attributed to Julius Caesar, attributing the text *Compendiosa expositio* to Apuleius, and identifying the author behind the pseudonym Elena Ferrante (Kestemont et al. 2016; Stover and Kestemont 2016; Savoy 2020; Tuzzi et al. 2024).

In the context of the GI method, authentication involves determining whether a text is consistently attributed to an author across many comparisons and quantifying the confidence in this determination. Unlike many other authorship attribution methods, the GI method handles open-set authorship verification problems, allowing for scenarios where the actual author may or may not be among the candidates.

The GI method verifies authorship based on the document's similarity to the purported author's writings and dissimilarity with imposters. The process is akin to a witness identifying a suspect from a police lineup. Multiple iterations using different subsets of the 2,000 Most Frequent Character n-grams enhance the robustness of the results (Eder and Rybicki 2013). In each iteration, 50% of each imposter's text and features are randomly selected for analysis, enabling consideration of numerous feature combinations and outlier detection, leading to more reliable outcomes (Eder et al. 2016). The method produces a score between 0 and 1 for each author in the lineup, indicating the proportion of times an author was identified. A higher score reflects greater confidence that the author wrote the disputed text (Eder 2018). This score not only gauges stylistic similarity, but also assesses how consistently an author is identified as opposed to imposters.

5. Validation

The methods described were assessed across multiple validation sub-corpora (detailed in respective subsections) to measure their efficacy for authorship attribution/verification tasks. Utilizing the Cosine Delta distance metric and a frequency band of the top 2,000 MFCs 4-grams, no culling parameter was applied to ensure an adequate feature set.[9]

5.1 PCA (Validation)

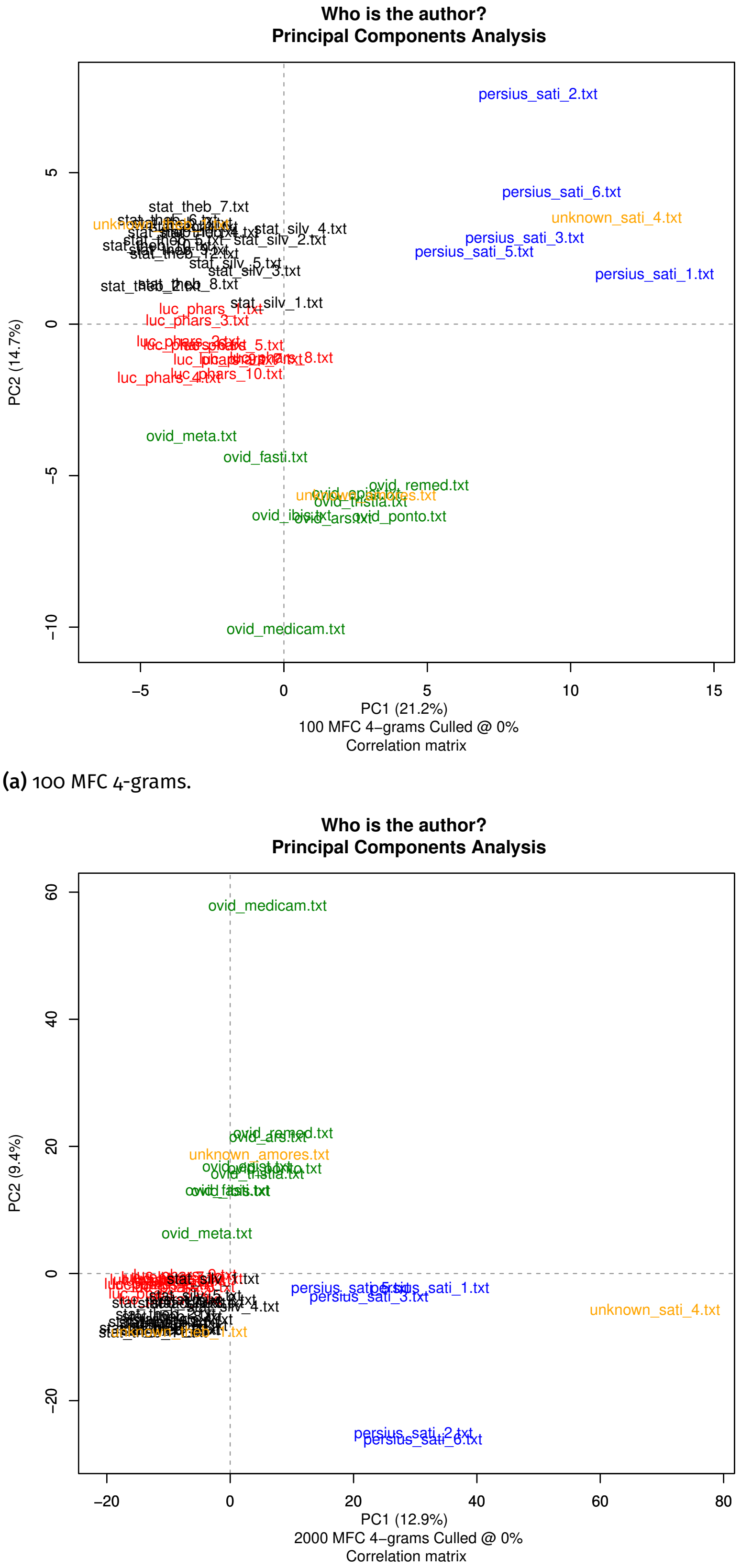
To validate PCA, a sub-corpus was created from the initial dataset, consisting of works by four authors: Ovid, Lucan, Persius, and Statius (see Table 10). These authors were chosen due to their temporal proximity to Seneca's work, despite differences in genre; while Lucan, Ovid, and Statius wrote epic poems, Persius focused on satires. Including Persius's works in this validation corpus was based on their relatively smaller size compared to the other works, posing a potential challenge for PCA analysis.

Demonstrating the method's emphasis on text variance over author names, three texts had their author names replaced with "unknown." The filenames were adjusted to `unknown_amores` for Ovid's *Amores*, `unknown_theb_1` for Statius' first book of *Thebaid*, and `unknown_sati_4` for Persius' fourth *Satura*. The first two texts were randomly chosen, while the last, due to its small size (392 tokens, including pronouns), posed a challenge for PCA.

Figure 4 presents the PCA results using a correlation matrix, showcasing the impact of different frequency bands (100 MFC 4-grams in Figure 4a and 2,000 MFC 4-grams in Figure 4b). Observation reveals a consistent attribution in both cases, with larger frequency bands showing less distinct clusters. Notably, in Figure 4b, Persius' fourth *Satura* and Ovid's text *Medicamina Faciei Femineae* exhibit some movement outside their relevant clusters. This deviation could be attributed to the small size of these texts relative to others in the corpus, as text size may influence authorship attribution or verification tasks (Luyckx and Daelemans 2011, 52; Eder 2013, 180).

Someone might ask at this point why the two PCAs, even though they use exactly the same corpus, differ in terms of variance they capture. The reason is that 100 MFC n-grams might capture more variance than when we include less frequent features. With 100 MFC n-grams, the dimensionality is lower, thus PCA more easily captures the variance explained by the features. With the addition of more features, more noise might be added and thus less variance might be captured (see Figure 4).

[9] Culling, with a ratio of 20, involves including only words occurring in at least 20% of documents in a corpus. While enhancing result comparability, especially with balanced corpora, it introduces a drawback. In unbalanced corpora like ours, with varying document lengths, culling may lead to insufficient features, resulting in an indistinguishable authorial fingerprint for some authors.



(a) 100 MFC 4-grams.

(b) 2,000 MFC 4-grams.

Figure 4: PCA using a correlation matrix to visualize the results. Figure 4a demonstrates how the attribution works given a small frequency band (i.e., 100 MFCs 4-grams). Figure 4b demonstrates the authorship attribution given a larger frequency band (i.e., 2,000 MFCs 4-grams).

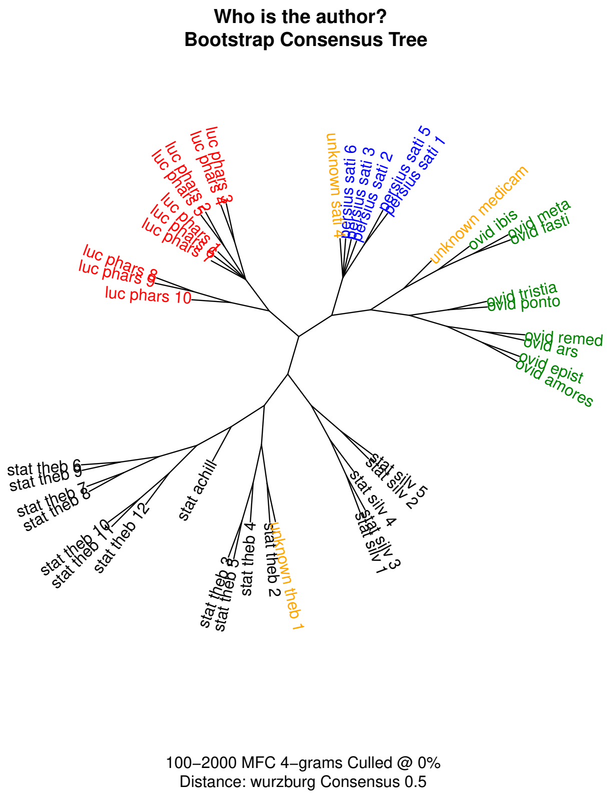


Figure 5: A BCT that was generated using the top 100-2,000-100 (start-end-step) MFC 4-grams and Cosine Delta as distance metric (no culling set); pronoun culling was applied and a consensus strength of 0.5 was used.

5.2 BCT (Validation)

At this point, it is crucial to note that the Bootstrap Consensus Tree (BCT) functions as a consensus, capturing more dimensions and information than PCA due to the robust patterns observed across different iterations (see subsubsection 4.2.2). Unlike PCA, which reduces the data to a few principal components by focusing on the largest sources of variance, BCT aggregates information from multiple bootstrap samples, thus integrating a wider range of variations and subtleties in the data. This allows BCT to provide a more comprehensive view of the data's structure and relationships, encompassing nuances that PCA might overlook.

In this validation, the corpus is slightly changed and the file names are altered again to demonstrate the independence of the final result (unrooted tree and branches) from the file names. Due to its very small size, this time instead of *Amores* we use *Medicamina Faciei Femineae* as part of the unknown texts by converting its filename to `unknown_medicam`. The rest of the "unknown" texts remain consistent as in the previous validation test (see subsection 5.1).

All texts in the test set were accurately attributed to their respective authors using BCT (see Figure 5). Notably, the texts renamed as "unknown," which presented challenges in PCA (i.e., Ovid's *Medicamina Faciei Femineae* and Persius' 4th Satura) were handled adeptly by BCT, emphasizing the robustness of BCT in authorship attribution tasks regardless of text size (see subsubsection 4.2.2 for further details).

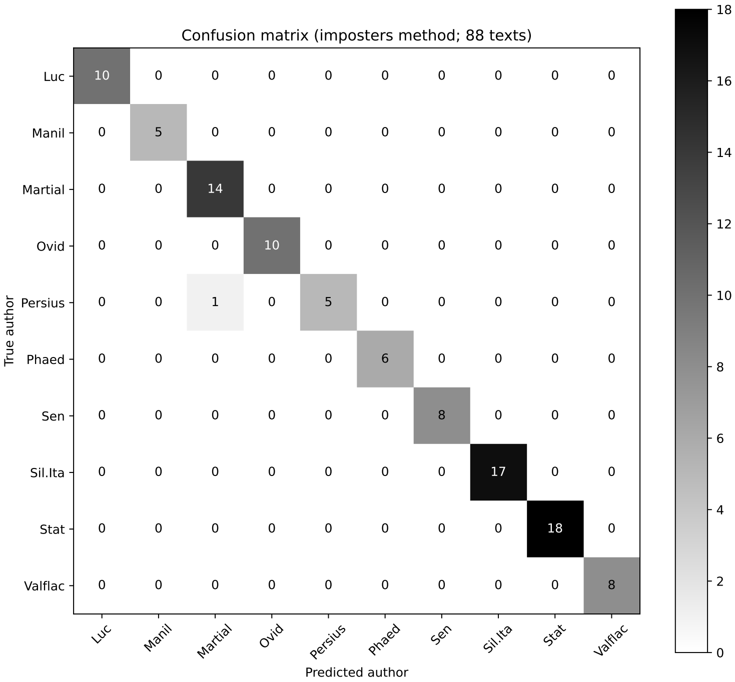


Figure 6: The confusion matrix shows the results of the GI method on the validation dataset. *P1* value = 0.31 and *p2* value = 0.67. The result is based on the author that returned the highest score for a given text. The two disputed plays, *Oct.* and *H.O.*, by Seneca the Younger are excluded from the validation set.

5.3 GI Method (Validation)

The GI method was validated using all known texts in our corpus, excluding the two disputed Senecan plays (*Oct.* and *H.O.*), resulting in a total of 88 texts for validation. The Cosine Delta served as the distance metric, and the frequency bands ranged from the top 100 to the 2,000 Most Frequent Character (MFC) 4-grams. The method is applied for 100 iterations per run to enhance performance. No culling parameter was set, and consistent preprocessing steps were applied, including orthographic normalization (see subsection 4.1), tokenization, and lower-casing, along with pronoun-culling. Subsequently, the GI method was applied to each text in the validation corpus.

5.4 Validation Findings

The validation indicates effective performance for all methods on the texts within the corpus, with PCA showing limitations for short texts (Figure 4). The BCT method demonstrates robust recognition of authorial fingerprints across varied text lengths, owing to their bootstrapping techniques, culminating in a consensus from multiple iterations (see Figure 5). Similarly, the GI method commits only one mistake (see Figure 6): Martial is being attributed as the author of the 1st *Satura* of Persius with a confidence score of 0.71 (+0.04 of *p1*). These findings suggest that the selected frequency band (top 100 to 2,000 Most Frequent Character 4-grams) is informative for capturing authorial fingerprints, yielding high success rates in each validation scenario. Consequently, the main analysis phase will replicate this process with a focus on the disputed texts.

6. Results and Discussion

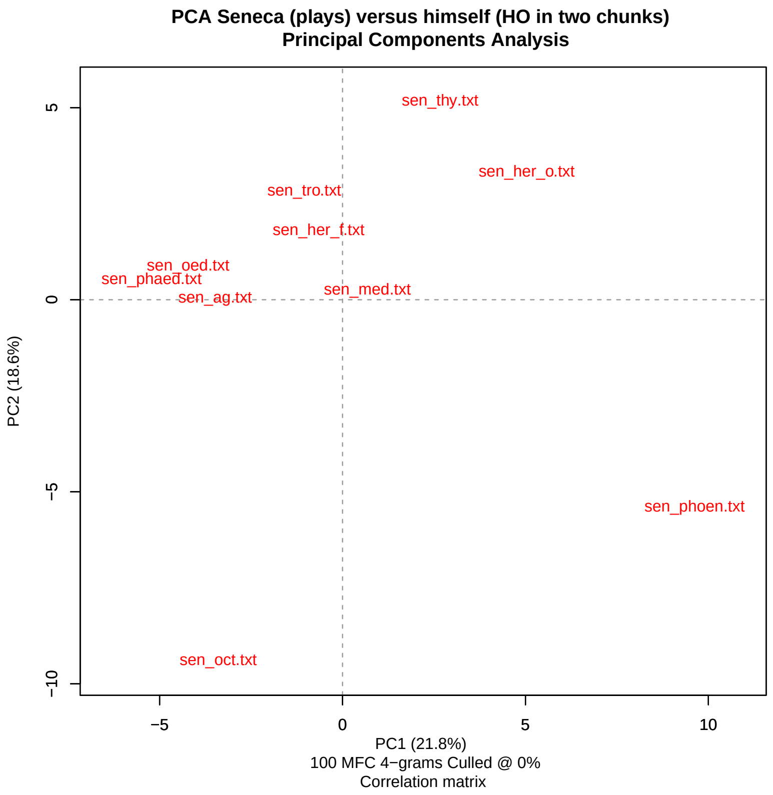
We first explore the stylometric properties of the Senecan plays using PCA to see how they relate to each other. When treating the plays as a whole, it can be observed that from the two disputed texts, only *Oct.* behaves as outlier within the Senecan corpus of plays (see Figure 7). However, *H.O.* consists of 11,1147 tokens, which is almost twice as many tokens as the average size of a Senecan play (excluding *Oct.*) ($\bar{x}$ = 6192.5 tokens). When *H.O.* is divided into two halves to align its size more closely with the average size of a Senecan play, it shifts away from the cluster of Senecan texts (see Figure 8). Meanwhile, *Oct.* remains consistently outside the cluster of Senecan plays. A possible explanation of why *Oct.* and *H.O.* behave as outliers is the fact that when considering the works of a single author using a PCA, the genre-related signal tends to become stronger than the author-related signal (Stover and Kestemont 2016, 659).

In addition to that, it should be stressed that in all of the PCA plots *Phoenissae* also behaves as an outlier within the Senecan corpus, while its authorship is not disputed. An explanation for this behavior could be that *Phoenissae* is an unfinished play and the shortest text in the Senecan corpus of plays. Furthermore, the aforementioned play has a lot of issues in terms of structure and unity. Based on the number of innovations that were attempted in the text, Frank (2018, 1–2) points out that this might be the reason why this text was abandoned by Seneca when he realized the difficulty of this venture.

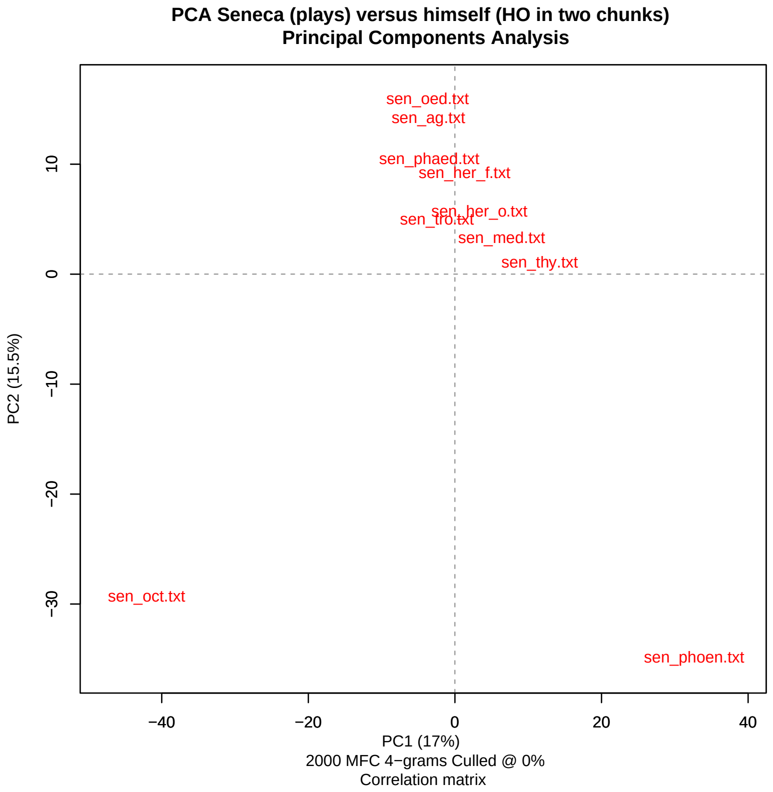
Figure 9 shows a Bootstrap Consensus Tree (BCT) for the Senecan plays alongside two selected authors from the literature of the Early Empire, namely Lucan and Statius. Statius is included to test the hypothesis of Ferri (2003, 17–27), which suggests a temporal connection between the composition of *Oct.* and Statius. On the one hand, the BCT exhibits distinct branches for each author, placing both disputed plays in proximity to the Senecan works, but *Oct.* gravitates slightly towards the center of the unrooted tree. This again highlights the special nature of this specific text. On the other hand, *H.O.* remains in the Senecan cluster of plays.

Regarding the GI method, we test five different scenarios. Since GI returns a confidence score as the final output, we need to pick thresholds for when to reject or accept the verification of an author. *Stylo* provides a method to automatically determine such thresholds using cross-validation (the `stylo.optimize()` method). For Scenario 1 it gives $p1 = 0.45$ and $p2 = 0.55$, for Scenario 2 the threshold is set to the values $p1 = 0.22$ and $p2 = 0.76$, whereas for Scenario 3 it returns $p1 = 0.00$ and $p2 = 0.98$ (for a brief description of the different GI scenarios see Table 3); below *p1* Seneca is not the author; above *p2* Seneca is identified as the author; when the score is in between, no determination can be made. Unfortunately, the cross-validation method is too expensive to run on the larger datasets we use in the rest of our experiments (see scenarios 4 and 5 in Table 3) due to the nested loops and the bootstrapping that takes place, which increases the time complexity of the algorithm. Therefore, we will use a conservative threshold of 0.9 for all our experiments.

Since there is no option to set a random seed in *stylo*'s environment, we applied a workaround to achieve more consistent and replicable results. Specifically, we employed the GI method by running the analysis 10 times, with each run consisting of 100 iterations. We report both the average score and the standard deviation across the 10 runs. This

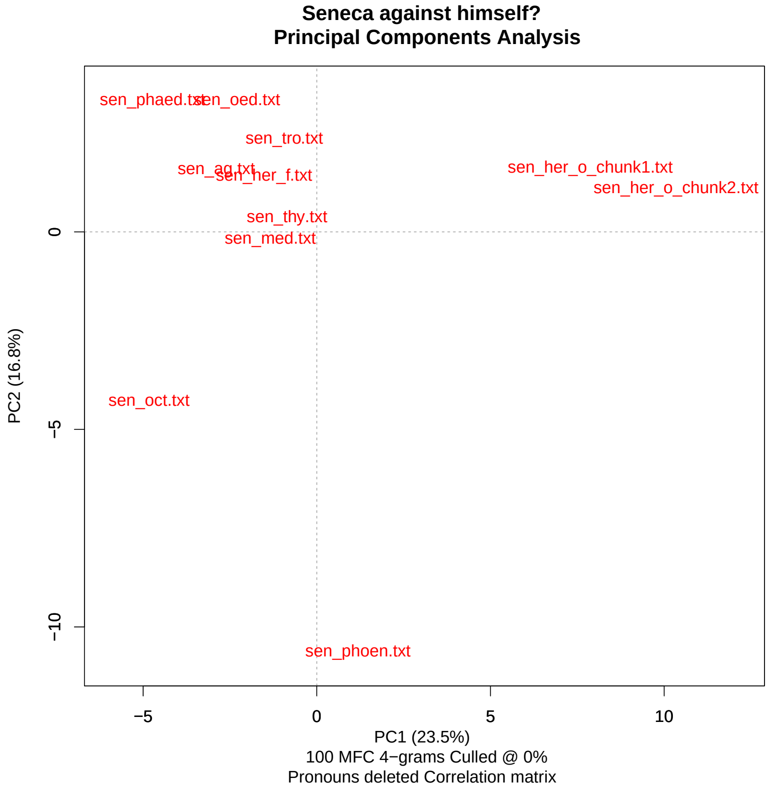


(a) 100 MFC 4-grams.

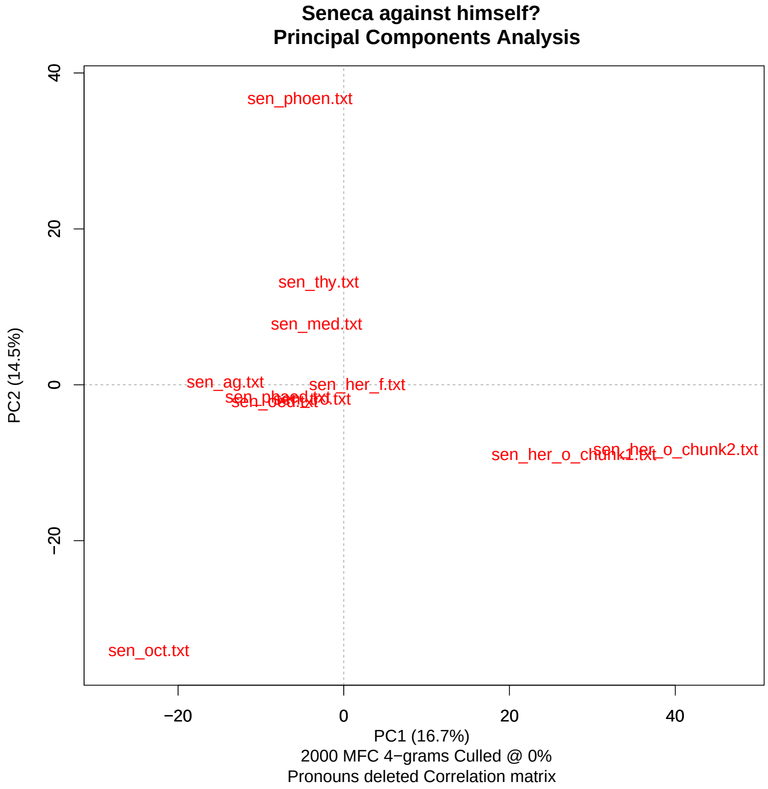


(b) 2,000 MFC 4-grams.

Figure 7: PCA correlation matrix of the Senecan corpus of plays (disputed and not). The texts `seneca_oct` and `seneca_her_o` correspond to *Oct.* and *H.O.*, respectively. In both cases, regardless of the size of the frequency band, *Oct.* and *Phoenissae* behave as outliers within the Senecan corpus, whereas *H.O.* is placed among the Senecan plays. It is important to highlight that the percentage shown in PC1 and PC2 varies in each plot because the principal components capture different amounts of variance each time.



(a) 100 MFC 4-grams.



(b) 2,000 MFC 4-grams.

Figure 8: PCA correlation matrix of the Senecan corpus of plays (disputed and not), this time with *H.O.* split in half. *H.O.* starts to behave as an outlier and *Oct.* remains among the outliers. It is important to highlight that the percentage shown in PC1 and PC2 varies in each plot because the principal components capture different amounts of variance each time.

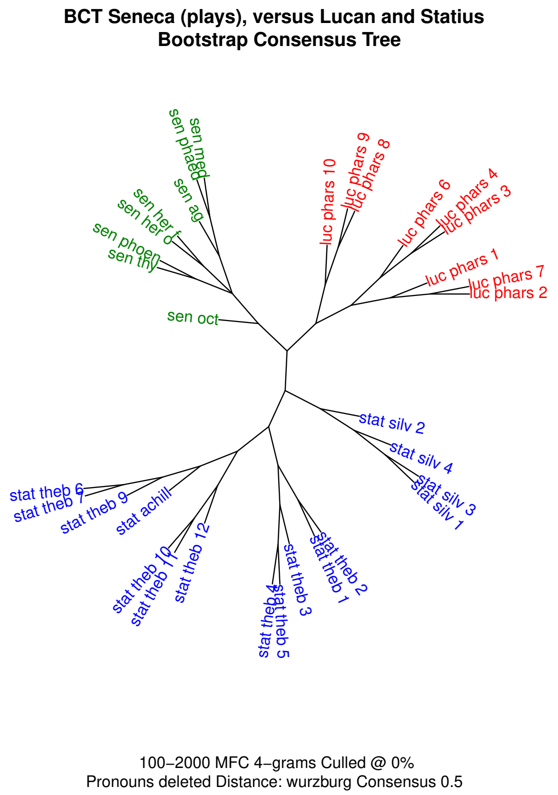


Figure 9: BCT of texts from Statius (*Achilleid*, *Thebaid*, *Silvae*), Lucan (*Pharsalia*), and Seneca (plays). The texts `seneca_oct` and `seneca_her_o` correspond to *Oct.* and *H.O.*, respectively.

approach allowed us to assess the stability of our results by observing variability and central tendencies over multiple independent runs, even without a fixed random seed. A drawback, however, is the increase in time complexity, as this method adds an external loop of 10 runs to the existing internal loop of 100 iterations, resulting in a total of $1,000$ iterations.

With the GI method, Scenario 1 and 2 confidently attribute Seneca the Younger as the author of the disputed plays (see Table 3). Next, in Scenario 3, we consider the cento-argument by Ferri (2014, 48).[10] We do this by identifying and removing sentences from the disputed texts resembling those in the Senecan corpus of plays. We operationalize sentence similarity using tf-idf (term frequency-inverse document frequency) vectors of the character 4-grams for each sentence, and cosine similarity as the metric for the similarity of pairs of sentences. We identify and exclude all sentences with a similarity exceeding a threshold of 0.6. The cosine similarity metric measures directional similarity between vectors, irrespective of magnitude or scale (Singhal 2001, 2–3). The presented methodology, when integrated with specific preprocessing procedures including the conversion to lowercase, elimination of punctuation marks (with the understanding that an editor may subsequently reintroduce punctuation marks), and the utilization of character 4-grams as distinctive features, exhibits the capability to discern similarities. This capability is exemplified in Table 4, wherein similarities are identified not only among various declensions of identical terms but also amid permutations in word order. For *Oct.* from a total of 422 sentences, we identified and thus removed 2 (i.e., 0.46%)

[10] A basic definition of a cento would describe it as a composition largely comprised of quotations from the works of other authors.

Scenario	Dataset	Results
Scenario 1: The GI method used against the disputed texts (no changes were applied to the texts *per se*)	104 texts in verse written by authors that lived slightly before and after Seneca the Younger (see Figure 1 and Table 10).	*Oct.*: 1.0 *H.O.*: 1.0
Scenario 2: The GI method is applied to *H.O.* split into two chunks.	Same as Scenario 1, but *H.O.* split into two chunks, thus in total 105 texts.	*H.O.* chunk 1: 1.0 chunk 2: 1.0
Scenario 3: The GI method is applied to the two disputed texts. *Oct.* and *H.O.* are cleaned by removing sentences that are above the similarity threshold (i.e., 0.6) in terms of cosine similarity.	Same as Scenario 1, but *Oct.* and *H.O.* are cleaned from similar lines with the rest of the Senecan corpus of plays.	*Oct.*: 1.0 *H.O.*: 1.0
Scenario 4: The GI method is applied to the two disputed texts (i.e., *Oct.* and *H.O.*). Each text in the corpus is split into non-overlapping chunks of 500 words if their length is above 500 tokens. This addresses a possible length bias due to shorter or longer texts. In addition, it enables checking for mixed authorship throughout the disputed texts.	The main corpus, but the texts are divided into chunks of 500 tokens, resulting in 1,344 text samples.	For the scores for each chunk, see Figure 10 and Figure 12.
Scenario 5: The GI method is applied to the chunks of the two disputed plays. This time, the texts are compared with texts in prose (the dataset is the one used by Kestemont et al. (2016) but augmented with the chunks of our imposters dataset).	A larger dataset of mostly historiographical texts written in prose (a small number are in verse), augmented with the 500 token chunks from our main imposters dataset, resulting in 3,090 text samples. This dataset includes texts written by Seneca the Younger in prose (e.g., *De Ira*, *De Providentia*, etc.).	For the score for each chunk, see Figure 11 and Figure 13.

Table 3: All the scenarios tested using the GI method, a brief description of the results, and the *p1* and *p2* values for each scenario. The interpretation of the *p1* and *p2* values is as follows: any score below *p1* suggests a negative answer to the question, "Can author A be confirmed as the author of disputed document X?" Conversely, any score above *p2* indicates a positive answer to the same question. Between *p1* and *p2* lies a 'gray area' where no definitive conclusions should be drawn.

Play	Line	Score
Phoenissae	scelus in propinquo est	
Oct.	nihil in propinquos temere constitui decet	0.40
Agamemnon	eheu quid hoc est	
H.O.	quid hoc	0.52
Phaedra	anime quid segnis stupes	
H.O.	quid stupes segnis furor	0.60
Medea	Profugere dubitas?	
Oct.	Parere dubias?	0.64
Thyestes	Viduam relinques?	
H.O.	Vitam relinques?	0.71
Phoenissae	Et hoc sat est	
Oct.	nec hoc sat est	0.74
Phaedra	quam bene excideram mihi	
H.O.	quam bene excideras dolor	0.77
Agamemnon	scelus occupandum est	
H.O.	scelus occupandum est	1

Table 4: Lines from Senecan and disputed plays with cosine similarity scores. The first two rows are examples of sentences that did not pass the threshold (< 0.6).

sentences above the similarity threshold (i.e., 0.6), whereas for *H.O.*, from a total of 1149 sentences we identified and removed 33 (i.e., 2.87%) sentences.

To address potential length bias and investigate possible mixed authorship throughout the disputed texts, in Scenario 4 each text exceeding 500 tokens is divided into non-overlapping chunks of 500 tokens. This approach, inspired by Rolling Stylometry (Eder 2016), simplifies the process by using non-overlapping segments instead of overlapping ones. Rolling Stylometry works by analyzing text in sequential segments to track stylistic patterns and changes throughout a document or corpus. The results for Scenario 4 (Figure 10 and Figure 12) reveal a nuanced internal composition, uncovering authorship diversity within the disputed plays. Although Seneca's authorship dominates, specific segments warrant attention, as highlighted in Figure 10 and Figure 12.

When we compare the chunks of *Oct.* with the chunks from the main corpus in Scenario 4, we see that chunks 1, 2, 3, 6, 8 are below the threshold of 0.90 (see Figure 10 and Table 5). However, for chunks 1, 2, 6, and 8, we obtain an average confidence score of 0.85, 0.83, 0.88, and 0.87, respectively, which indicates that Seneca authored these parts, but someone might have made minor changes. Focusing on chunk 3 (0.31), there are features that appear merely as a decoration (e.g., the change between iambics and anapaests) and parts of the text that are a critique of Nero (e.g., the rhetorical question of why Jupiter is killing the innocent and not those who deserve it, i.e., Nero; see Ferri (2014, 180–210)). Criticizing the emperor himself, especially by someone (Seneca) who was his advisor, was not an easy task.

In Scenario 5, the comparison reveals one more chunk of *Oct.* that might raise concerns (see Figure 11 and Table 7), chunk 6 (0.57). The playwriter here condenses the time in a way that seems unnatural for Seneca the Younger, in order to present a large number of events in a small amount of time (Ferri 2014, 307–309). The rest of the chunks (i.e.,

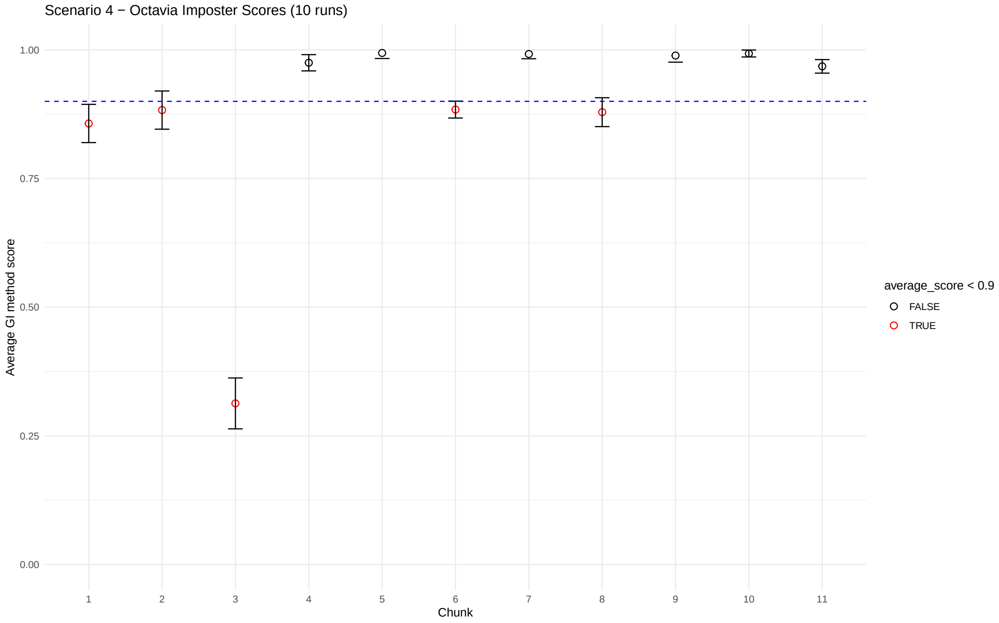


Figure 10: Results of the GI method for *Oct.*'s chunks using the corpus of texts in verse (Scenario 4). The dots represent the average score across ten runs, while the error bars indicate the standard deviation, reflecting the range of score variability observed throughout the entire process.

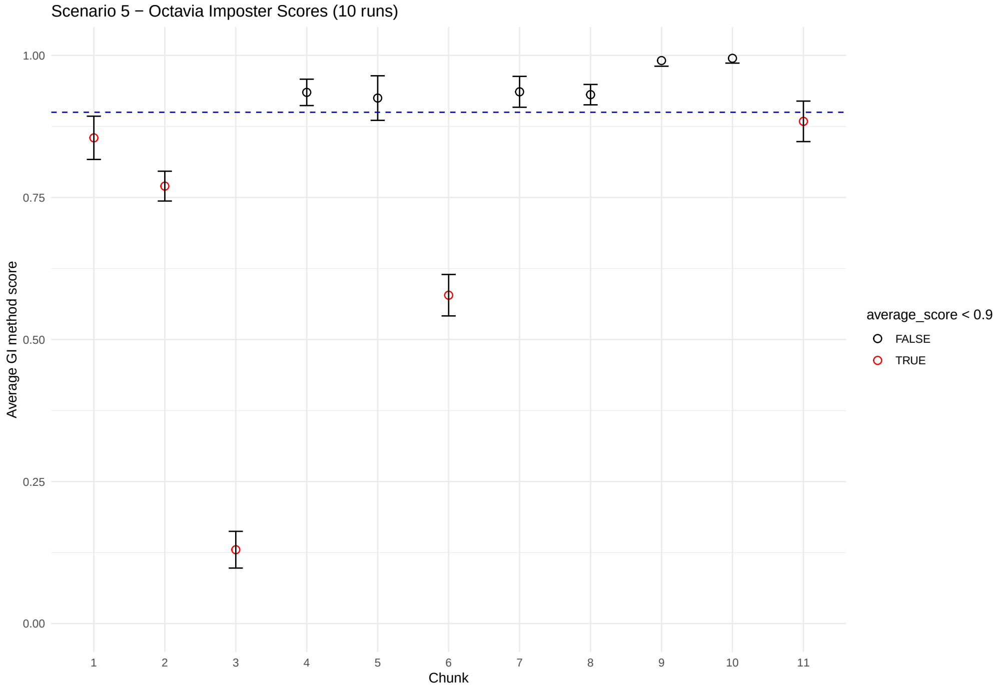


Figure 11: Results of the GI method for *Oct.*'s chunks using the dataset of Kestemont et al. (2016) (Scenario 5). The dots represent the average score across ten runs, while the error bars indicate the standard deviation, reflecting the range of score variability observed throughout the entire process.

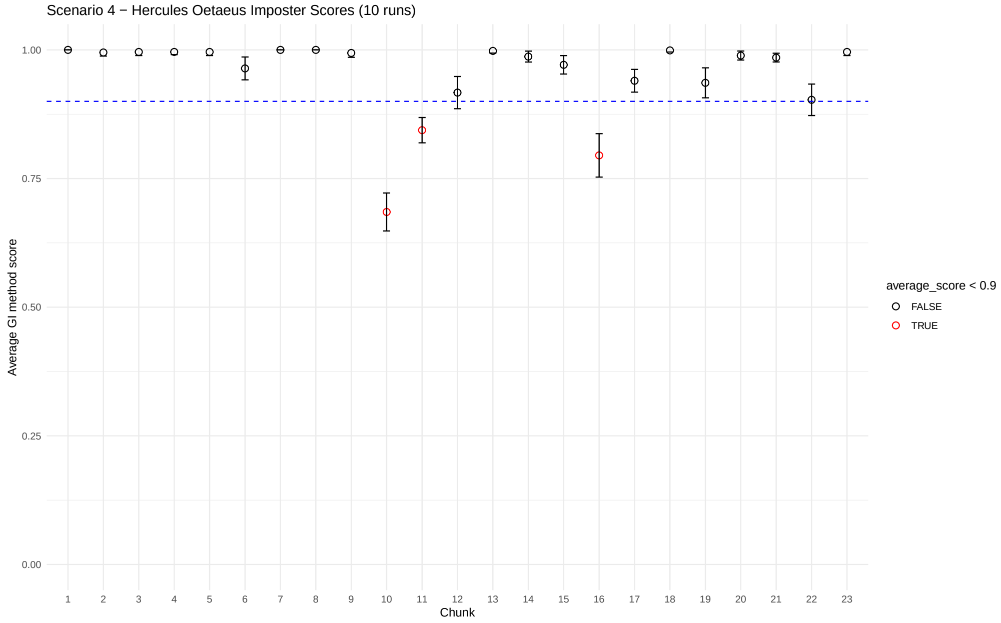


Figure 12: Results of the GI method for *H.O.*'s chunks using the corpus of texts in verse (Scenario 4). The dots represent the average score across ten runs, while the error bars indicate the standard deviation, reflecting the range of score variability observed throughout the entire process.

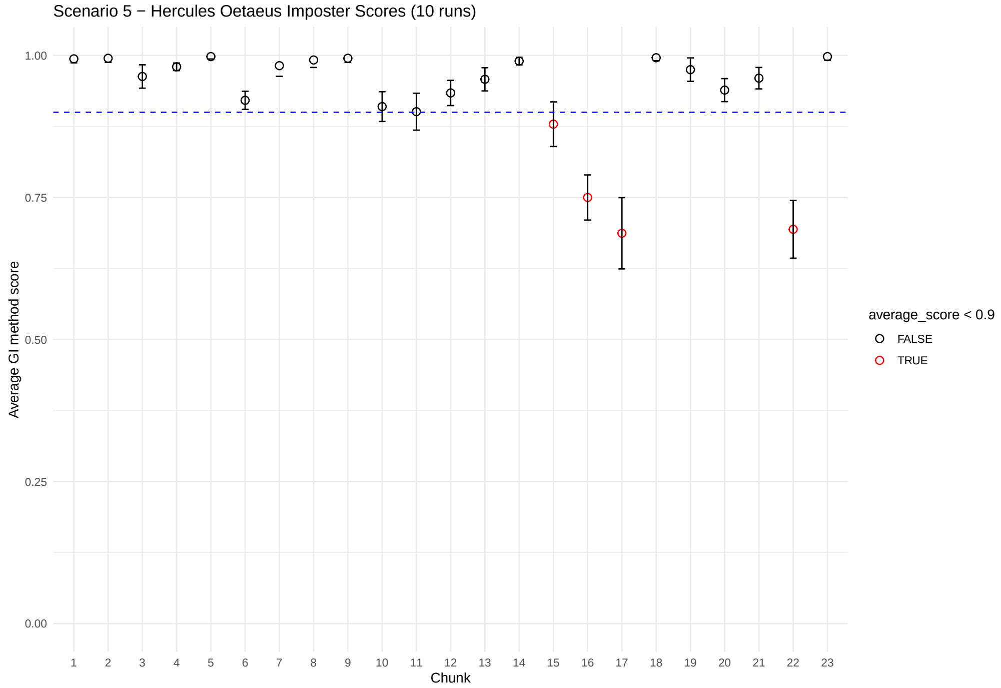


Figure 13: Results of the GI method for *H.O.*'s chunks using the dataset of Kestemont et al. (2016) (Scenario 5). The dots represent the average score across ten runs, while the error bars indicate the standard deviation, reflecting the range of score variability observed throughout the entire process.

Chunk no.	Lines	Average score	Standard deviation
Chunk 1	l. 1–107	0.85	0.037
Chunk 2	l. 107–193	0.88	0.037
Chunk 3	l. 193–291	0.31	0.049
Chunk 6	l. 467–553	0.88	0.016
Chunk 8	l. 668–762	0.87	0.028

Table 5: (Scenario 4) Chunks of *Oct.* that return a score below or equal to the threshold of 0.9 using the main corpus split into non-overlapping chunks of 500 tokens. The lines correspond to their online version in the Perseus Digital Library.

Chunk no.	Lines	Average score	Standard deviation
Chunk 10	l. 802–883	0.68	0.036
Chunk 11	l. 884–967	0.84	0.024
Chunk 16	l. 1,342–1,423	0.79	0.042

Table 6: (Scenario 4) Chunks of *H.O.* that return a score below or equal the threshold of 0.9 using the main corpus split into non-overlapping chunks of 500 tokens. The lines correspond to their online version in the Perseus Digital Library.

chunks 1, 2, 11) do not fall far from the threshold of 0.9, indicating that someone might have made minor adjustments.

Regarding *H.O.* both scenarios 4 (see Figure 12 and Table 6) and 5 (see Figure 13 and Table 8) reveal a pattern where the first half of the text is securely Senecan, and the second half shows chunks that do not meet the threshold. These results align to some extent with the hypothesis that the first half of the text originates from Seneca, while the remainder may have been finished by someone else (Tarrant 2017, 97). However, most of the chunks in the second half still have high scores, suggesting that the second half is a case of mixed authorship, rather than being written completely by someone else.

Lastly, note that for both Scenarios 4 and 5, the standard deviation of the text chunks that fall under the threshold of 0.9 tends to be larger than the standard deviation of the chunks above the threshold. This is another indicator that these chunks differ stylistically and warrant further study.

7. Conclusions

Our findings underscore the complexity of the authorship verification problem, particularly evident in the case of the disputed Senecan plays, *Oct.* and *H.O.*. Across

Chunk no.	Lines	Average score	Standard deviation
Chunk 1	l. 1–107	0.85	0.038
Chunk 2	l. 107–192	0.77	0.026
Chunk 3	l. 193–291	0.13	0.032
Chunk 6	l. 467–553	0.57	0.036
Chunk 11	l. 961–end	0.88	0.035

Table 7: (Scenario 5) Chunks of *Oct.* that return a score below or equal the threshold of 0.9 using Kestemont et al. (2016)'s corpus. The lines correspond to their online version in the Perseus Digital Library.

Chunk no.	Lines	Average score	Standard deviation
Chunk 15	l. 1,258–1,342	0.87	0.039
Chunk 16	l. 1,343–1,423	0.75	0.039
Chunk 17	l. 1,423–1,507	0.68	0.062
Chunk 22	l. 1,849–1,954	0.69	0.050

Table 8: (Scenario 5) Chunks of *H.O.* that return a score below or equal to the threshold of 0.9 using Kestemont et al. (2016)’s corpus. The lines correspond to their online version in the Perseus Digital Library.

experimental runs, varying results highlight the intricate nature of this challenge in computational stylometry.

Paraphrasing Stover and Kestemont (2016, 647), our aim is not to replace existing modes of analysis, but rather to illuminate longstanding issues by shedding new light through the application of innovative tools grounded in traditional methods. This analysis underscores the importance of considering genre and meter variations in our conclusions. As previously noted, these two factors can introduce complexities due to their influence on vocabulary. It is impossible to completely remove the influence of variation in meter and genre (Corbara et al. 2023), thus we employ preprocessing techniques to mitigate their impact on the final results.

Through the validation phase, we demonstrate the effectiveness of these techniques for our task. Consequently, we apply these techniques consistently to generate uniform features for each method. Notably, in the case of the two exploratory methods – PCA and BCT – *Oct.* and *H.O.* emerge as intriguing examples of texts concerning their authorship among the Senecan corpus of plays. In certain instances, they exhibit clustering with the broader set of Senecan plays, while in other cases they do not. For instance, when using only the Senecan plays, the genre and thus the meter and the size of the plays seem to win over the authorial fingerprint (see the cases of *Phoenissae* and *H.O.* in Figure 7).

The initial two scenarios of the GI method confidently verify Seneca as the author with a high degree of confidence (= 1.0). Moreover, after removing sentences from both disputed plays that are similar to sentences from other Senecan plays, the GI method still verifies Seneca as the author of the disputed plays. Therefore, the stylistic similarity of the disputed plays to the works of Seneca cannot be explained by borrowed phrases. Nevertheless, the fourth and the fifth scenarios highlight segments in *Oct.* and *H.O.* that are unlikely to be attributed to Seneca, implying the involvement of a distinct author or editor. Concentrating on *H.O.* (see Figure 12 and Figure 13), we posit that an editor of the text may have edited or added certain parts in the original play, even though it was primarily authored by Seneca. Lastly, the results hold up when the disputed plays are compared with a larger corpus of prose texts, suggesting that our findings are robust.

Against this algorithmic confidence, two objections can be made: First, we cannot rule out a highly skilled imitator; however, this seems implausible given the advanced nature of modern stylometry, of which an imitator could not have been aware. Second, the distractor texts differ in genre and meter from the Senecan texts. Unfortunately, due to the limitations of extant texts, it is impossible to construct a perfect distractor corpus. Therefore, while our empirical findings cannot positively confirm Seneca as the author of the disputed plays, our main contribution is that, perhaps contrary to expectations given the consensus against Seneca’s authorship, most of the text of the disputed plays is highly stylistically similar to Seneca’s writings. This means that Seneca cannot be ruled out as the author of the disputed plays based on stylometry. Moreover, our results provide evidence for mixed authorship in specific parts of the disputed plays.

8. Further Research

Deciphering the authorial fingerprint of the Senecan disputed plays requires further investigation and consideration of study limitations. Future work could take a closer look at the specific text chunks diverging from Seneca the Younger’s style. Employing Rolling Stylometry or using the General Imposters method with overlapping text segments (Eder 2016; Beullens et al. 2024), together with close reading approaches, could enable identification of authorship at the sentence level and enhance understanding of why these segments differ from Seneca’s style. Moreover, exploring the impact of prosody in ancient languages (e.g., Latin or Ancient Greek) on stylometric methods is another avenue for investigation. Controlled experiments using authors who wrote in different meters would make it possible to quantify their effect on the stylometric profile of texts. Furthermore, while the GI method has been shown to be robust and reliable in previous studies, including for Latin (Kestemont et al. 2016), it would be useful to examine and empirically test whether an imitator can successfully deceive the GI method. The Ferrante case shows that the pseudonym of an author who is highly motivated to hide his or her identity can be unmasked by pinpointing the gender, age, region, and city of the author profile (Mikros 2018). A potential improvement would be to use a large language model that could also detect paraphrases by taking into account semantic similarity.

9. Data Availability

Data used for the research can be found at: https://github.com/PaschalisAg/seneca_stylometry. It has been archived and is persistently available at: https://doi.org/10.5281/zenodo.14002368.

10. Software Availability

All code created and used in this research has been published at: https://github.com/PaschalisAg/seneca_stylometry. It has been archived and is persistently available at: https://doi.org/10.5281/zenodo.14002368.

11. Acknowledgements

We extend our sincere gratitude to all the reviewers for their invaluable feedback. Their insightful comments illuminated aspects of the paper that might otherwise have escaped our notice. We are grateful to Bram van der Velden for feedback on a draft. Special thanks are due to Vasileios Dimoglidis, a PhD candidate in Classics at the University of Cincinnati (UC), whose initial inspiration sparked this study. Additionally, we are grateful to Associate Professor Vasileios Pappas and Professor Helen Gasti from the

University of Ioannina (UOI) for generously providing an extensive bibliography to support our research into the non-quantitative approaches examined in this study.

12. Author Contributions

Paschalis Agapitos: Code, Data curation, Conceptualization, Investigation, Writing – original draft

Andreas van Cranenburgh: Conceptualization, Formal analysis, Methodology, Supervision, Writing – review & editing

References

Baldwin, Barry (1981). "The Authorship of the 'Aratus' Ascribed to Germanicus". In: *Quaderni Urbinati di Cultura Classica* 7, 163–172. 10.2307/20538670.

Beullens, Pieter, Wouter Haverals, and Ben Nagy (2024). "The Elementary Particles: A Computational Stylometric Inquiry into the Mediaeval Greek-Latin Aristotle". In: *Mediterranea. International Journal on the Transfer of Knowledge* 9, 385–408. https://journals.uco.es/mediterranea/article/view/16723.

Boyle, Anthony J. (2009). *Tragic Seneca: An Essay in the Theatrical Tradition*. Routledge.

Brofos, James, Ajay Kannan, and Rui Shu (2014). "Automated Attribution and Intertextual Analysis". In: *arXiv*. 10.48550/ARXIV.1405.0616.

Cantaluppi, Gabriele and Marco Passarotti (2015). "Clustering the Corpus of Seneca: A Lexical-Based Approach". In: *Advances in Latent Variables: Methods, Models and Applications*. Ed. by Maurizio Carpita, Eugenio Brentari, and El Mostafa Qannari. Springer International Publishing, 13–25. 10.1007/10104_2014_6.

Carbone, Martin E. (1977). "The 'Octavia': Structure, Date, and Authenticity". In: *Phoenix* 31 (1), 48–67. 10.2307/1087155.

Carey, William L. (2024). *The Latin Library*. http://www.thelatinlibrary.com/ (visited on 05/13/2024).

Corbara, Silvia, Alejandro Moreo, and Fabrizio Sebastiani (2023). "Syllabic Quantity Patterns as Rhythmic Features for Latin Authorship Attribution". In: *Journal of the Association for Information Science and Technology* 74 (1), 128–141. https://doi.org/10.1002/asi.24660.

Crane, Gregory R., ed. (2024). *Perseus Digital Library*. https://www.perseus.tufts.edu/hopper/ (visited on 05/14/2024).

Daelemans, Walter (2013). "Explanation in Computational Stylometry". In: *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing*. Vol. 2, 451–462. 10.1007/978-3-642-37256-8_37.

Eder, Maciej (2011). "Style-markers in Authorship Attribution A Cross-language Study of the Authorial Fingerprint". In: *Studies in Polish Linguistics* 1. https://www.ejournals.eu/SPL/2011/SPL-vol-6-2011/art/1171/ (visited on 10/16/2024).

— (2012). "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" In: *The Translator and the Computer*. Ed. by Tadeusz Piotrowski and Łukasz Grabowski. Wyższa Szkoła Filologiczna Press.

— (2013). "Does Size Matter? Authorship Attribution, Small Samples, Big Problem". In: *Digital Scholarship in the Humanities* 30 (2), 167–182. 10.1093/llc/fqt066.

Eder, Maciej (2016). "Rolling Stylometry". In: *Digital Scholarship in the Humanities* 31 (3), 457–469. 10.1093/llc/fqv010.

— (2018). "Authorship Verification with the Package stylo". In: *Computational Stylistics*. https://computationalstylistics.github.io/docs/imposters (visited on 10/16/2024).

— (2022). "Boosting Word Frequencies in Authorship Attribution". In: *Proceedings of Computational Humanities Research*. 10.48550/arXiv.2211.01289.

Eder, Maciej and Jan Rybicki (2013). "Do Birds of a Feather Really Flock Together, or How to Choose Training Samples for Authorship Attribution". In: *Literary and Linguistic Computing* 28 (2), 229–236. 10.1093/llc/fqs036.

Eder, Maciej, Jan Rybicki, and Mike Kestemont (2016). "Stylometry with R: A Package for Computational Text Analysis". In: *The R Journal* 8 (1), 107–121. 10.32614/RJ-2016-007.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt (2017). "Understanding and Explaining Delta Measures for Authorship Attribution". In: *Digital Scholarship in the Humanities* 32 (2), ii4–ii16. 10.1093/llc/fqx023.

Ferri, Rolando (2003). *Octavia: A Play Attributed to Seneca*. Cambridge Classical Texts and Commentaries. Cambridge University Press.

— (2014). "Octavia". In: *Brill's Companion to Seneca: Philosopher and Dramatist*. Ed. by Andreas Heil and Gregor Damschen. Brill, 521–527. 10.1163/9789004217089_043.

Frank, Marica (2018). *Seneca's Phoenissae: Introduction and Commentary*. Brill. 10.1163/9789004329430.

Gahan, John J. (1985). "Seneca, Ovid, and Exile". In: *The Classical World* 78 (3), 145–147. 10.2307/4349723.

Gómez Caballero, Iván (2021). "Estudio estilométrico del teatro latino: a vueltas con Octavia y Hercules Oetaeus de Séneca." In: *Cuadernos de Filología Clásica: Estudios Latinos* 41 (1).

Grieve, Jack (Sept. 1, 2007). "Quantitative Authorship Attribution: An Evaluation of Techniques". In: *Literary and Linguistic Computing* 22 (3), 251–270. 10.1093/llc/fqm020.

Hagiwara, Masato (2021). *Real-world Natural Language Processing: Practical Applications with Deep Learning*. Manning.

Herington, Cecil J. (1961). "Octavia Praetexta: A Survey". In: *The Classical Quarterly* 11 (1), 18–30. 10.1017/S0009838800008351.

Hoover, David L. (2004). "Delta Prime?" In: *Literary and Linguistic Computing* 19 (4), 477–495. 10.1093/llc/19.4.477.

Jannidis, Fotis, Steffen Pielström, Christof Schöch, and Thorsten Vitt (2015). "Improving Burrows' Delta – An Empirical Evaluation of Text Distance Measures". In: *Book of Abstracts of the Digital Humanities Conference 2015*. http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empi/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empirical_.html (visited on 10/16/2024).

Johnson, Kyle P., Patrick J. Burns, John Stewart, Todd Cook, Clément Besnier, and William J. B. Mattingly (2021). "The Classical Language Toolkit: An NLP Framework for Pre-modern Languages". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, 20–29. 10.18653/v1/2021.acl-demo.3.

Jolliffe, Ian T. and Jorge Cadima (2016). "Principal Component Analysis: A Review and Recent Developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065). 10.1098/rsta.2015.0202.

Juola, Patrick (2015). "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions". In: *Digital Scholarship in the Humanities* 30 (1), i100–i113. 10.1093/llc/fqv040.

Jurafsky, Dan and James H. Martin (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd ed. draft. https://web.stanford.edu/~jurafsky/slp3/ (visited on 05/03/2024).

Karakasis, Evangelos (2018). *T. Calpurnius Siculus: A Pastoral Poet in Neronian Rome*. Trends in Classics 35. De Gruyter. 10.33776/ec.v24i0.5007.

Kestemont, Mike (2014). "Function Words in Authorship Attribution. From Black Magic to Theory?" In: *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59–66. 10.3115/v1/W14-0908.

Kestemont, Mike, Sara Moens, and Jeroen Deploige (2015). "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux". In: *Digital Scholarship in the Humanities* 30 (2), 199–224. 10.1093/llc/fqt063.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans (2016). "Authenticating the Writings of Julius Caesar". In: *Expert Systems with Applications* 63, 86–96. 10.1016/j.eswa.2016.06.029.

Khonji, Mahmoud and Youssef Iraqi (2014). "A Slightly-modified GI-based Author-verifier with Lots of Features (ASGALF)". In: *Conference and Labs of the Evaluation Forum Working Notes*, 977–983. http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-KonijEt2014.pdf (visited on 10/16/2024).

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon (2009). "Computational Methods in Authorship Attribution". In: *Journal of the American Society for Information Science and Technology* 60 (1), 9–26. 10.1002/asi.20961.

Koppel, Moshe, Jonathan Schler, and Elisheva Bonchek-Dokow (2007). "Measuring Differentiability: Unmasking Pseudonymous Authors". In: *Journal of Machine Learning Research* 8 (6).

Koppel, Moshe and Yaron Winter (2014). "Determining if Two Documents Are Written by the Same Author". In: *Journal of the Association for Information Science and Technology* 65 (1), 178–187. 10.1002/asi.22954.

Kuhn, Max and Kjell Johnshon (2016). "Over-fitting and Model Tuning". In: *Applied Predictive Modelling*. Ed. by Max Kuhn and Kjell Johnshon. 5th ed. Springer, 61–92.

Luyckx, Kim and Walter Daelemans (2011). "The Effect of Author Set Size and Data Size in Authorship Attribution". In: *Literary and Linguistic Computing* 26 (1), 35–55. 10.1093/llc/fqq013.

Manousakis, Nikos (2020). *'Prometheus Bound' – A Separate Authorial Trace in the Aeschylean Corpus*. De Gruyter. 10.1515/9783110687675.

Marshall, Christopher W. (2014). "The Works of Seneca the Younger and Their Dates". In: *Brill's Companion to Seneca: Philosopher and Dramatist*. Brill, 33–44. 10.1163/9789004217089_003.

Marti, Berthe (1945). "Seneca's Tragedies. A New Interpretation". In: *Transactions and Proceedings of the American Philological Association* 76, 216–245. 10.2307/283337.

Michalopoulos, Andreas N. (2020). "Seneca Quoting Ovid in the Epistulae Morales". In: *Intertextuality in Seneca's Philosophical Writings*. Ed. by Myrto Garani, Andreas N. Michalopoulos, and Sophia Papaioannou. Routledge, 130–141.

Mikros, George K. (2018). "Blended Authorship Attribution: Unmasking Elena Ferrante Combining Different Author Profiling Methods". In: *Drawing Elena Ferrante's Profile*. Padova University Press, 85–96. `http://members.unine.ch/jacques.savoy/Articles/WorkshopFerrante.pdf` (visited on 10/16/2024).

Newman, Matthew L., Carla J. Groom, Lori D. Handelman, and James W. Pennebaker (2008). "Gender Differences in Language Use: An Analysis of 14,000 Text Samples". In: *Discourse Processes* 45 (3), 211–236. `10.1080/01638530802073712`.

Nisbet, Robert G. M. (1995). *Collected Papers on Latin Literature*. Oxford University Press.

Nolden, Luuk (2019). "Finding Seneca in Seneca: Using Text Mining Techniques of Hercules Oetaeus and Octavia". Bachelor Thesis. Leiden Institute of Advanced Computer Science (LIACS). `https://theses.liacs.nl/pdf/2018-2019-NoldenLSJ.pdf` (visited on 10/16/2024).

Päpcke, Simon, Thomas Weitin, Katharina Herget, Anastasia Glawion, and Ulrik Brandes (2022). "Stylometric Similarity in Literary Corpora: Non-authorship Clustering and Deutscher Novellenschatz". In: *Digital Scholarship in the Humanities*, fqac039. `10.1093/llc/fqac039`.

Pease, Arthur Stanley (1920). "Is the 'Octavia' a Play of Seneca?" In: *The Classical Journal* 15 (7), 388–403. `http://www.jstor.org/stable/3288405` (visited on 10/16/2024).

Philp, Robert H. (1968). "The Manuscript Tradition of Seneca's Tragedies". In: *The Classical Quarterly* 18 (1), 150–179. `http://www.jstor.org/stable/637696` (visited on 10/16/2024).

Poe, Joe Park (1989). "Octavia Praetexta and Its Senecan Model". In: *The American Journal of Philology* 110 (3), 434–459. `10.2307/295219`.

Possanza, Mark D. (2003). "Appendix A: Authorship and Date". In: *Translating the Heavens: Aratus, Germanicus, and the Poetics of Latin Translation*. Lang Classical Studies. Peter Lang, 217–243.

Potha, Nektaria and Efstathios Stamatatos (2017). "An Improved Impostors Method for Authorship Verification". In: *Proceedings of the 8th International Conference of the CLEF Association*, 138–144.

Rozelaar, Marc (1985). "Neue Studien zur Tragödie 'Hercules Oetaeus'". In: *Band 32/2. Teilband Sprache und Literatur (Literatur der julisch-claudischen und der flavischen Zeit [Forts.])*. Ed. by Wolfgang Haase. De Gruyter, 1348–1420. `10.1515/9783110861549-013`.

Rybicki, Jan (2012). "The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation". In: *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*. Ed. by Michael P. Oakes and Meng Ji. Studies in Corpus Linguistics. John Benjamins Publishing Company, 231–248. `10.1075/scl.51.09ryb`.

Rybicki, Jan and Magda Heydel (2013). "The Stylistics and Stylometry of Collaborative Translation: Woolf's Night and Day in Polish". In: *Literary and Linguistic Computing* 28 (4), 708–717. `10.1093/llc/fqt027`.

Savoy, Jacques (2020). "Elena Ferrante: A Case Study in Authorship Attribution". In: *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*.

Ed. by Jacques Savoy. Springer International Publishing, 191–210. 10.1007/978-3-030-53360-1_8.

Seidman, Shachar (2013). "Authorship Verification Using the Impostors Method". In: *Conference and Labs of the Evaluation Forum Working Notes*, 23–26. https://ceur-ws.org/Vol-1179/CLEF2013wn-PAN-Seidman2013.pdf (visited on 10/16/2024).

Seneca, L. Annaeus [1921] (2007). "Octavia". In: *Tragoediae*. Ed. by Rudolf Peiper and Gustav Richter. Perseus Digital Library. http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A2007.01.0006%3Acard%3D1 (visited on 10/29/2024).

— (2008). *Octavia: Attributed to Seneca*. Ed. by Anthony J. Boyle. Oxford University Press. 10.1093/actrade/9780199287840.book.1.

Singhal, Amit (2001). "Modern Information Retrieval: A Brief Overview". In: *IEEE Data Eng. Bull.* 24 (4), 35–43. http://singhal.info/ieee2001.pdf (visited on 10/16/2024).

Stamatatos, Efstathios (2009). "A Survey of Modern Authorship Attribution Methods". In: *Journal of the American Society for Information Science and Technology* 60, 538–556. 10.1002/asi.21001.

— (2013). "On the Robustness of Authorship Attribution Based on Character N-gram Features". In: *Journal of Law and Policy* 21 (2), 421–439. https://brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1048&context=jlp (visited on 10/16/2024).

Star, Christopher (2015). "Roman Tragedy and Philosophy". In: *Brill's Companion to Roman Tragedy*. Ed. by George W.M. Harrison. Brill, 238–259. 10.1163/9789004284784_013.

Stover, Justin and Mike Kestemont (2016). "Reassessing the Apuleian Corpus: A Computational Approach to Authenticity". In: *The Classical Quarterly* 66 (2), 645–672. 10.1017/S0009838816000768.

Stover, Justin, Yaron Winter, Moshe Koppel, and Mike Kestemont (2016). "Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author". In: *Journal of the Association for Information Science and Technology* 67 (1), 239–242. 10.1002/asi.23460.

Tarrant, Richard (2017). "Custode rerum Caesare: Horatian Civic Engagement and the Senecan Tragic Chorus". In: *Horace and Seneca. Interactions, Intertexts, Interpretations*. Ed. by Martin Stöckinger, Kathrin Winter, and Andreas T. Zanker. De Gruyter, 93–112. 10.1515/9783110528893-005.

Tuzzi, Arjuna, George Mikros, and Michele A. Cortelazzo (2024). "Applying General Impostors Method to the Ferrante Case". In: *Digital Stylistics in Romance Studies and Beyond*. Ed. by Robert Hesselbach, José Calvo Tello, Ulrike Henny-Krahmer, Christof Schöch, and Daniel Schlör. Heidelberg University Publishing, 299–313. 10.17885/heiup.1157.c19377.

VanderPlas, Jake (2017). "In Depth: Principal Component Analysis". In: *Python Data Science Handbook*. Ed. by Jake VanderPlas. O'Reilly Media, 433–445.

A. Datasets Overview

Author	Time	Text
Ammianus Marcellinus	4th century C.E.	*Res Gestae A Fine Corneli Taciti;*
Quintus Asconius Pedianus	ca. 9 B.C.E. – ca. 76 C.E.	*Orationum Ciceronis Quinque Enarratio;*
Aulus Gellius	ca. 125 C.E. – after 180	*Noctes Atticae;*
Calpurnius Flaccus	2nd century C.E.	*Declamationes;*
M. Tullius Cicero	106 B.C.E. – 43 B.C.E.	*Academica, Laelius de Amicitia, Pro Archia, Brutus, Pro Caecina, Pro Caelio, Cato Maior de Senectute, De Divinatione, De Fato, De Finibus, Pro Milone, De Natura Deorum, De Officiis, De Optimo Genere Oratorum, Orator, De Oratore, Paradoxa Stoicorum, In Pisonem, De Re Publica, Topica, Tusculanae Disputationes;*
Quintus Curtius Rufus	1st century C.E.	*Historiarum Alexandri Magni Libri Qui Supersunt;*
Eutropius	4th century C.E.	*Breviarium Historiae Romanae;*
Rufius Festus	ca. 370 C.E.	*Festi Breviarium Rerum Gestarum Populi Romani;*
Florus	2nd century C.E.	*Epitome De T. Livio Bellorum Omnium Annorum DCC Libri Duo;*
unknown	-;-	*Historia Apollonii Regis Tyri;*
G. Julius Hyginus	ca. 64 B.C.E. – 17 C.E.	*Fabulae;*
Titus Livius	59 B.C.E. – 17 C.E.	*Ab Urbe Condita Libri;*
Lucius Ampelius	ca. 2nd century C.E.	*Liber Memorialis;*
Macrobius	flourished 400 C.E.	*Commentarii in Somnium Scipionis;*
M. Minucius Felix	ca. 250 C.E.	*Octavius;*
Nazarius	ca. 4th century C.E.	*Panegyricus Constantino Augusto Dictus;*
Pliny the Younger	61–2 C.E. – ca. 113 C.E.	*Epistularum Libri Decem, Panegyricus;*
Pomponius Mela	flourished ca. 43 C.E.	*De Chorographia;*
Quintus Tullius Cicero	102 B.C.E. – 43 B.C.E.	*Commentariolum Petitionis;*
Quintilian	35 C.E. – after 96 C.E.	*Declamationes Maiores, Institutiones;*
Sallustius	ca. 86 B.C.E. – 35/4 B.C.E.	*Bellum Catilinae, Epistola ad Caesarem I & II, Bellum Iugurthinum;*
Seneca the Younger	ca. 4 B.C.E. – 65 C.E.	*De Beneficiis, De Brevitate Vitae, De Clementia, De Consolatione, Epistulae Morales Ad Lucilium, De Vita Beata, De Ira, Quaestiones Naturales, De Otio, De Providentia, De Tranquilitate Animi;*
Seneca the Elder	ca. 55 B.C.E. – 39 C.E.	*Controversiae;*
Suetonius	ca. 69 C.E. – after 122 C.E.	*De Vitis Caesarum-Augustus, De Vitis Caesarum-Gaius, De Vitis Caesarum-Divus Claudius, De Vitis Caesarum-Domotianus, De Vitis Caesarum-Galba, De Vitis Caesarum-Divus Iulius, De Vitis Caesarum-Nero, De Vitis Caesarum-Otho, De Vitis Caesarum-Tiberius, De Vitis-Caesaris-Titus, De Vitis Caesarum-Divus Vespasianus, De Vitis Caesarum-Vitellius;*
Tacitus	56 C.E. – ca. 120 C.E.	*Agricola, Annales, Historiae, Dialogus De Oratoribus;*
Valerius Maximus	flourished 30 C.E.	*Factorum Et Dictorum Memorabilium Libri Novem;*
Varro	116 B.C.E. – 27 B.C.E.	*De Lingua Latina, Rerum Rusticarum De Agri Cultura;*
Velleius Paterculus	ca. 19 B.C.E. – after 30 C.E.	*Historiae Romanae;*

Table 9: List of authors and texts present in the dataset used by Kestemont et al. (2016).

Author	Text	Filename
Lucan	*Pharsalia*	luc_phars_{1–10}
Martial	*Epigrammata*	martial_epigr_{1–14}
Manilius	*Astronomica*	manil_astro_{1–5}
Ovid	*Amores* *Medicamine Faciei Femineae* *Ars Amatoria* *Remedia Amoris* *Metamorphoses* *Fasti* *Ibis* *Tristia* *Epistulae ex Ponto* *Epistulae or Heroides*	ovid_am ovid_medicam ovid_ars ovid_remed ovid_meta ovid_fasti ovid_ibis ovid_tristia ovid_ponto ovid_epist
Persius	*Saturae*	persius_sati_{1–6}
Phaedrus	*Fabulae*	phaed_fables_{1–6}
Seneca the Younger	*Agamemnon* *Hercules Furens* *Hercules Oetaeus* (disputed) *Medea* *Octavia* (disputed) *Oedipus* *Phaedra* *Phoenissae* *Thyestes* *Troades*	sen_ag sen_her_f sen_her_o sen_med sen_oct sen_oed sen_phaed sen_phoen sen_thy sen_tro
Silius Italicus	*Punica*	sil.ita_pun_{1–17}

Statius	*Thebaid*
	Silvae
	Achilleid
		stat_theb_{1–12}
		stat_silv_{1–5}
		stat_achil

Valerius Flaccus	*Argonautica*	valflac_argon_{1–8}

Table 10: Authors and texts included in the dataset. All of the texts are written in verse, albeit the only plays are the Senecan tragedies. In total, our corpus comprises 104 texts (including the disputed Senecan plays) and nine authors to compare against Seneca the Younger.