jcls. Journal of Computational Literary Studies

conference version

A Stylometric Analysis of Seneca's disputed plays

Authorship Verification of *Octavia* and *Hercules Oetaeus*

Paschalis Agapitos[1]
Andreas van Cranenburgh[2]

1. P. M. de Lardizabal 4, Donostia International Physics Center, 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: *CCLS2024 Conference Preprints* 3 (1). 10.26083/tuprints-00027394

Date published 2024-05-28
Date accepted 2024-04-04
Date received 2024-01-22

Keywords
Seneca, stylometry, authorship verification, Latin, Stylo

Note
This paper has been submitted to the conference track of JCLS. It has been peer reviewed and accepted for presentation and discussion at the 3rd Annual Conference of Computational Literary Studies at Vienna, Austria, in June 2024.

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 Impostors 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 give set of candidates (Koppel et al. 2007, 1261). Authorship verification concerns the question of whether an author wrote a disputed document (Koppel et al. 2007, 1261;Juola 2015, 1106). The verification task is more challenging than the attribution task, because the verification 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 E. 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, 114). 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 impostor. 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 computation stylometric studies can be condensed into four points: richer dataset because of their high frequency, closeness of the set since function words are limited and fixed, content-independent, and, as mentioned above, 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 aim to contribute to the debate on Seneca's disputed texts by applying a variety of computational stylistic methods and testing several different scenarios. We do this using 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 impostor 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 that has a Roman historical subject) that survived until today from the corpus of Latin dramas (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 O. 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 for 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

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

[2] Manuscript tradition E saves the Senecan plays with the following order: *Hercules* (*Furens*), *Troades*, *Phoenissae*, *Medea*, *Phaedra*, *Oedipus*, *Agamemnon*, *Thyestes*, *Hercules* (*Oetaeus*); *Octavia* is omitted in tradition E. Manuscript tradition A gives the Senecan plays with 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).

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.

Concerning the stylistic aspect of O, 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 low levels of creativity when it comes to the provision of conclusions among the scenes.

HO also raises some concerns about the attribution of its authorship. As Marshall (2014, 40) points out, referring to Nisbet, the play follows a different approach of playwriting. 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 been also argued that *Oct.* and *H.O.* indeed carry the authorial fingerprint of Seneca. Concerning O, in lines 619–621, Agrippina lists some traditional punishments in an effort to predict the tyrant's (i.e., Nero's) imminent death (Seneca, *Oct.* 619–621). In this passage, the demise of Nero appears to be foretold what seems to rule out Seneca as an author. However, some scholars argue that the description of the punishments is not even close to what actually happened to Nero (i.e., suicide) 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 and mythological punishments (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 in order to be published after the death of Nero. This argument – merely a speculation since no additional evidence exists – can explain the inconsistencies in the text which scholars used to argue that *Oct.* is not a Senecan play. If we follow 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 O.

With respect to *H.O.*, the argument of the late composition is also used in support of the *H.O.* as a genuine Senecan play (Rozelaar, 1985; Nisbet 1995, p. 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 the anomalies, which 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 something new (e.g., Kestemont et al. 2016; Stover et al. 2016; Stover and Kestemont 2016). However, the number of such papers that consider Senecan texts is much smaller, and more so those that actually consider the authenticity of the two disputed Senecan plays, *Oct.* and *H.O.* per se.

Brofos et al. 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). However, their model also makes, as expected, many misclassifications; 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 stylistics techniques. Nolden (2019) 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 a 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, *Phoenissae* is considered Senecan, so this casts doubt on whether these methods are reliable. 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 of Seneca's 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 in 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 hastily design the experiments by not taking into account multiple variables connected to the texts per se or by considering these works as non-Senecan and focusing on the evaluation of authorship attribution/verification methods and software. Trying to fill this research gap, this paper takes into account as many variables as possible, validates the computational methods

[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 environents.

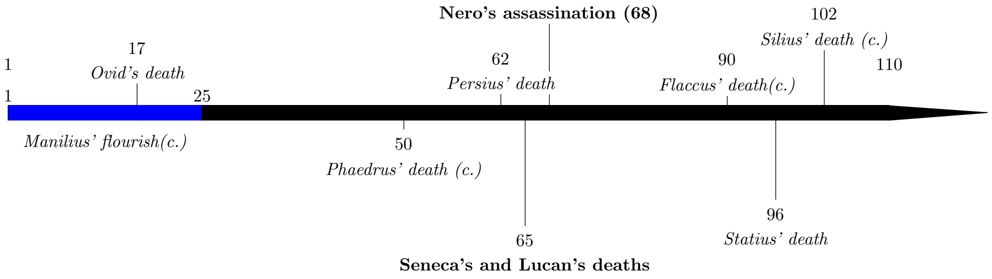


Figure 1: A timeline of the authors used in the dataset, centered around Nero's assassination, Seneca's suicide and Lucan's death. The two extremes in our corpus are Ovid and Silius Italicus.

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 (c. 4 BCE–65 CE). In the context of computational stylometric approaches, a distractor author, or "impostor", 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, Phaedrus, Persius, Lucan, Valerius Flaccus, Statius, and Silius Italicus (see Table 1). These authors, broadly associated with the literature of the early empire, that wrote within the first century of the Common Era (refer to Figure 1).[4]

In Scenario 5 presented in Table 4 we augment the dataset used by Kestemont et al. Kestemont et al. (2016) with our main corpus (see Table 1, therefore we consider of importance explaining what are the authors and the texts that populate this dataset, as well its main genre. Kestemont's dataset contains 1850 non-overlapping slices of 1000 tokens (for our analysis we split further these texts into non-overlapping slices of 500 tokens). The authors and the text present in the dataset are the following: *Res Gestae A Fine Corneli Taciti* by Ammianus Marcellinus (4th century AD), *Orationum Ciceronis Quinque Enarratio* by Quintus Asconius Pedianus (c. 9 B.C.E. - c. 76 C.E.), *Noctes Atticae* by Aulus Gellius (c. 125 C.E. - after 180), *Declamationes* by Calpurnius Flaccus (2nd century 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* by M. Tullius Cicero (106 B.C.E. - 43 B.C.E.),

[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.

Historiarum Alexandri Magni Libri Qui Supersunt by Quintus Curtius Rufus (1st century C.E.), *Breviarium Historiae Romanae* by Eutropius (4th century C.E.), *Festi Breviarium Rerum Gestarum Populi Romani* by Rufius Festus (c. 370 C.E.), *Epitome De T. Livio Bellorum Omnium Annorum DCC Libri Duo* by Florus (2nd century C.E.), *Historia Apollonii Regis Tyri* by unknown, *Fabulae* by G. Julius Hyginus (c. 64 B.C.E. - 17 C.E.), *Ab Urbe Condita Libri* by Titus Livius (59 B.C.E. - 17 C.E.), *Liber Memorialis* by Lucius Ampelius (c. 2nd century C.E.), *Commentarii in Somnium Scipionis* by Macrobius (flourished 400 C.E.), *Octavius* by M. Minucius Felix (c. 250 C.E.), *Panegyricus Constantino Augusto Dictus* by Nazarius (c. 4th century C.E.), *Epistularum Libri Decem*, and *Panegyricus* by Pliny the Younger (61-2 C.E. - c. 113 C.E.), *De Chorographia* by Pomponius Mela (flourished c. 43 C.E.), *Commentariolum Petitionis* by Quintus Tullius Cicero (102 B.C.E. - 43 B.C.E.), *Declamationes Maiores*, and *Institutiones* by Quintilian (35 C.E. - after 96 C.E.), *Bellum Catilinae*, *Epistola ad Caesarem I & II*, *Bellum Iugurthinum* by Sallustius (c. 86 B.C.E. - 35/4 B.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*, and *De Tranquilitate Animi* by Seneca the Younger (c. 4 B.C.E. - 65 C.E.), *Controversiae* by Seneca the Elder (c. 55 B.C.E. - 39 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 Caesarum-Tiberius*, *De Vitis-Caesaris-Titus*, *De Vitis Caesarum-Divus Vespasianus*, *De Vitis Caesarum-Vitellius* by Suetonius (c. 69 C.E. - after 122 C.E.), *Agricola*, *Annales*, *Historiae*, *Dialogus De Oratoribus* by Tacitus (56 C.E. - c. 120 C.E.), *Factorum Et Dictorum Memorabilium Libri Novem* by Valerius Maximus (flourished 30 C.E.), *De Lingua Latina*, *Rerum Rusticarum De Agri Cultura* by Varro (116 B.C.E. - 27 B.C.E.), *Historiae Romanae* by Velleius Paterculus (c. 19 B.C.E. - after 30 C.E.). Their dataset has mostly historiographical texts since in their paper they compare their corpus with Caesar's writings and it 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. Following Grieve (2007, 255), texts, disputed or not, are inherently tied to their historical era. Consequently, the dataset is designed to narrow the temporal scope, ensuring 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 temporarilly are located relatively close to Seneca's the Younger era and of 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 frequent character sequences rather than whole words (see subsection 4.1).

[5] We do test one scenario where we add 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 Kestemont's dataset (Kestemont et al. 2016)).

In addition to that, prior work on cross-genre and cross-topic stylometry has shown empirically that character-based authorship attribution is robust to such variation (e.g., P. D. Stamatatos et al. 2013, 343). 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 1 provides a complete list of authors and texts included in the dataset variations used for each experiment. All works, with the exceptioin of Manilius's *Astronomica*, were obtained from the Perseus Digital Library (*Perseus Digital Library* 2024) [6] because the latter was unavailable from the primary source. Thus, *Astronomica* was sourced from The Latin Library (*The Latin Library* 2024) [7].

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

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. 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. n.d., 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 3.

The extraction of relevant features involves character 4-grams in our study, a choice proven effective in cross-genre and cross-topic authorship attribution (Koppel et al. 2009, 12–13; E. Stamatatos 2009, 541–542; Eder 2011, 110; P. D. Stamatatos et al. 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 for word lemmatization or other normalization, as these features 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

[6] Available at: `https://github.com/cltk/lat_text_perseus`
[7] Available at: `https://github.com/cltk/lat_text_latin_library`
[8] For a very simple and informative definition of n-grams see Hagiwara (2021, 53–54).

Author	**Text**	**Filename**
Lucan	*Pharsalia*	luc_phars_{1-10}
Manilius	*Astronomica*	manil_astro_{1-5}
Ovid	*Amores*	ovid_am
	Medicamine Faciei Femineae	ovid_medicam
	Ars Amatoria	ovid_ars
	Remedia Amoris	ovid_remed
	Metamorphoses	ovid_meta
	Fasti	ovid_fasti
	Ibis	ovid_ibis
	Tristia	ovid_tristia
	Epistulae ex Ponto	ovid_ponto
	Epistulae or Heroides	ovid_epist
Persius	*Saturae*	persius_sati_{1-6}
Phaedrus	*Fabulae*	phaed_fables_{1-6}
Seneca the Younger	*Agamemnon*	sen_ag
	Hercules Furens	sen_her_f
	Hercules Oetaeus (disputed)	sen_her_o
	Medea	sen_med
	Octavia (disputed)	sen_oct
	Oedipus	sen_oed
	Phaedra	sen_phaed
	Phoenissae	sen_phoen
	Thyestes	sen_thy
	Troades	sen_tro
Silius Italicus	*Punica*	sil.ita_pun_{1-17}
Statius	*Thebaid*	stat_theb_{1-12}
	Silvae	stat_silv_{1-5}
	Achilleid	stat_achil
Valerius Flaccus	*Argonautica*	valflac_argon_{1-8}

Table 1: 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 90 texts (including the disputed Senecan plays) and 8 authors to compare against Seneca the Younger.

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 (E. Stamatatos 2009, 541–542; P. D. Stamatatos et al. 2013, 432–433; Eder 2011, 112). In the rest of this paper, we will use the the frequencies of the Most Frequent Character (MFC) n-grams. For example, 2000 MFC refers to the frequencies of the 2000 most common character n-grams.

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, based on its effectiveness in various test conditions and particular effectiveness for inflected languages (Jannidis et al. 2015, 6–8; Evert et al. 2017, ii9–

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 (wherever there are less than four characters displayed, the white-spaces are being counted as characters and are displayed using an underscore).

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	vobis
vos					

Table 3: A list of the 98 inflectional forms of 13 pronouns that are removed from every text of the corpus as provided by the software *Stylo* (Eder et al. 2016).

ii10; Eder 2022). Both the validation and main analysis phases utilize the 2000 most frequent character 4-grams (MFCs), a selection supported by studies indicating that 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 leads to better performance since the features capture more stylistic variation; however, beyond the 2000 MFCs, the character n-grams become more rare 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 2) illustrates this diminishing informativeness beyond the 2000th 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 Impostors (GI) method, each briefly outlined in the subsequent section.

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

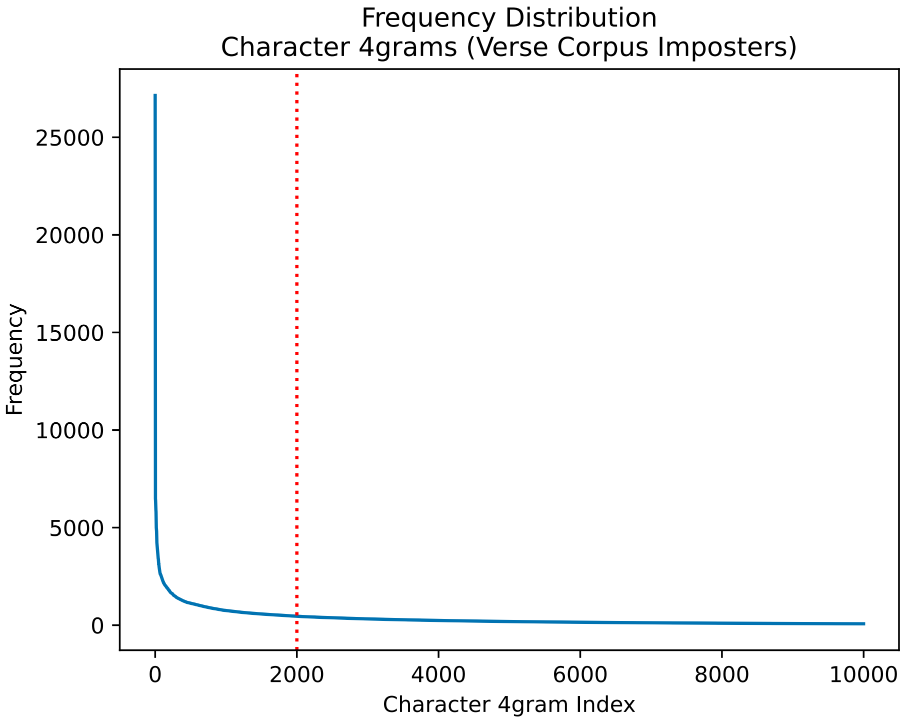


Figure 2: Frequency distribution of the character 4-grams in the whole corpus (i.e., 90 texts including the disputed plays). The vertical line is set to 2000 to show that characters 4-grams after this threshold start to become quite infrequent. The result is what we expect to see since the distribution 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*).

or variables initially present in the dataset (in our case the features that are generated 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 (E. 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 100-2000 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 replacement. 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, an assumption of this method is 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). Moreover, 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 2000 with a step of 100, and a consensus strength set at 0.5.

4.2.3 General Impostors Method

The GI method, initially introduced by Koppel and Winter (2014), has won for two consecutive years (i.e., 2013 and 2014) the first places in the PAN competitions for shared tasks in authorship verification (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, and (Kestemont et al. 2016; Stover and Kestemont 2016; Savoy 2020).

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 impostors. The process is akin to a witness identifying a suspect from a police lineup. Multiple iterations using different subsets of the 2000 most frequent character n-grams enhance the robustness of the results (Eder and Rybicki 2013). In each iteration, 50% of each impostor's text and feautures 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 with respect to the 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 2000 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 (refer to Table 1). 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.

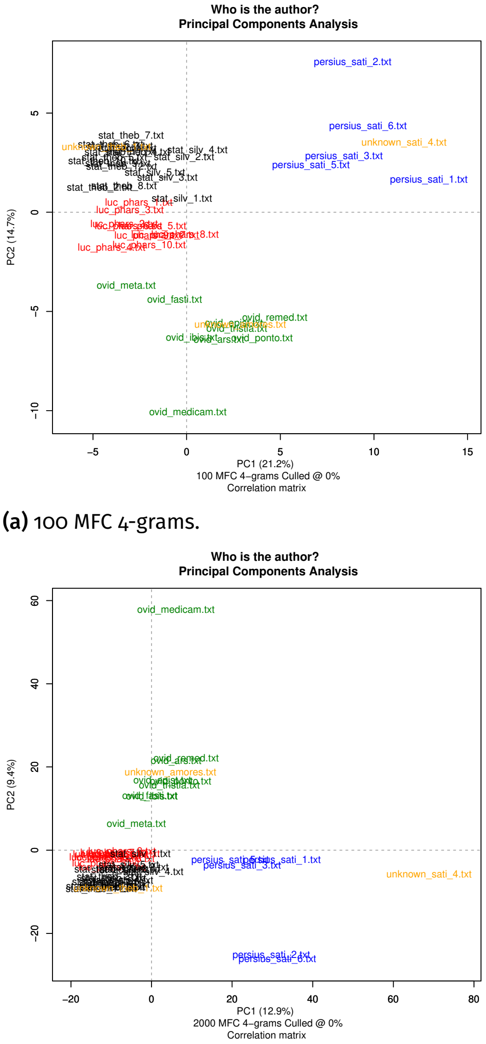
Figure 3 presents PCA results using the correlation matrix, showcasing the impact of different frequency bands (100 MFC 4-grams in Figure 3a and 2000 MFC 4-grams in Figure 3b). Observation reveals a consistent attribution in both cases, with larger frequency bands showing less distinct clusters. Notably, in Figure 3b, 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).

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 above subsubsection 4.2.2).

In this validation, the corpus is slightly changed, and file names were altered again to demonstrate the independence of the final result (unrooted tree and branches) from 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 to `unknown_medicam`.

[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) 2000 MFC 4-grams.

Figure 3: PCA using the correlation matrix to visualize the results. Figure 3a demonstrates how the attribution works given a small frequency band (i.e., 100 MFCs 4-grams). On the other hand, Figure 3b (on the right) demonstrates the authorship attribution given a larger frequency band (i.e., 2000 MFCs 4-grams).

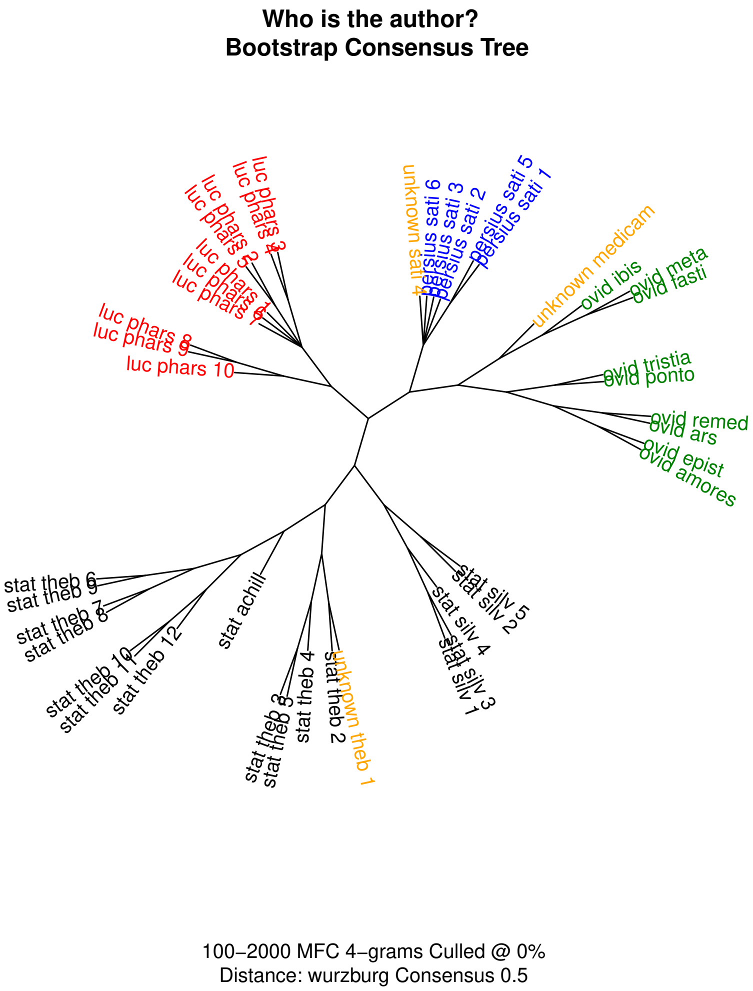


Figure 4: A Bootstrap Consensus Tree that was generated using the top 100-2000-100 (start-end-step) MFC 4-grams and Cosine Delta as distance metric (no culling set); pronoun culling was applied and a consensus strength of 0.5 was used.

The rest of the "unknown" texts remain consistent as in the previous validation test (see above subsection 5.1).

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

5.3 GI Method (Validation)

The GI method was validated using all known texts in our corpus, excluding the two disputed Senecan plays (O and *H.O.*), resulting in a total of 88 texts for validation. The Cosine Delta served as the distance metric, and frequency bands ranged from the top 100 to 2000 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 subsec-

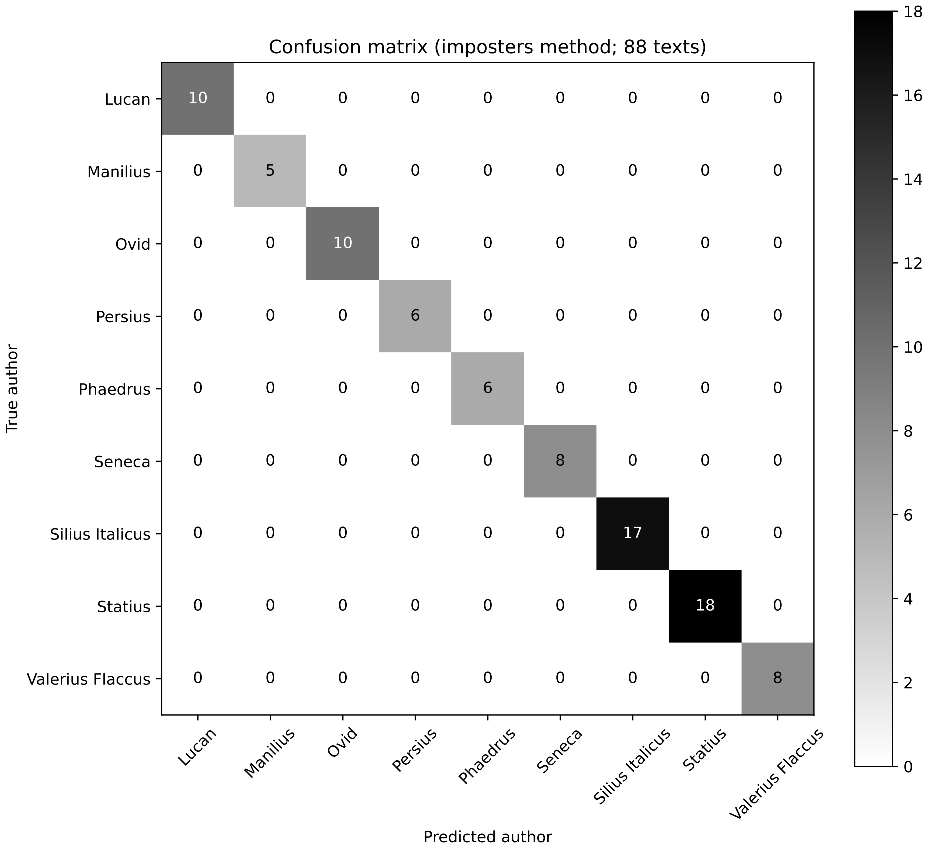


Figure 5: Confusion matrix that shows the results of the GI method on the validation dataset. P1 value = 0.35 and P2 value = 0.64. The result is based on the author that returned the highest score for a given text. The two disputed plays, *Oct.* and *H.O.*, by Seneca the Younger are excluded from the validation set.

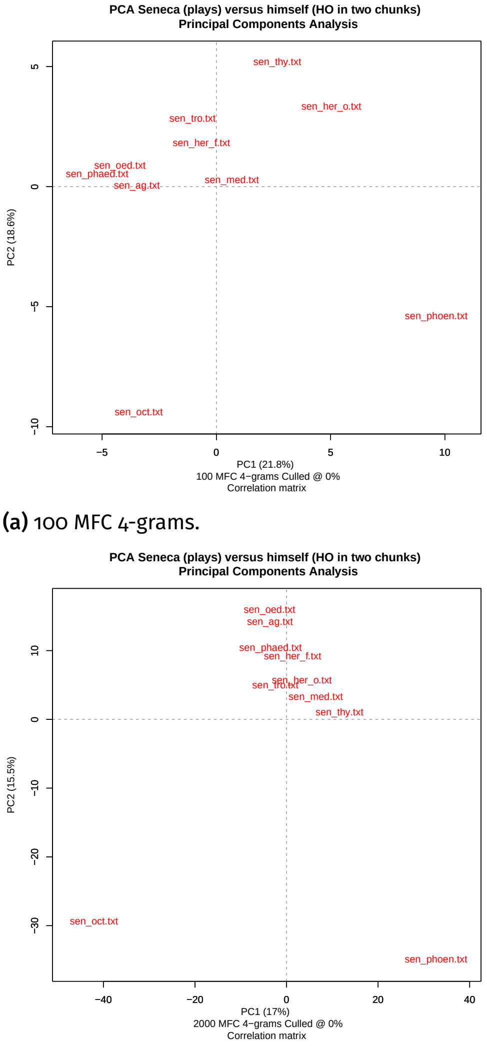
tion 4.1), tokenization and lower-casing, along with pronoun-culling. Subsequently, the GI method was applied to each text in the validation corpus.

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 3). 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 4). Similarly, the GI method reports a perfect accuracy for attributing the 88 texts (see Figure 5). These findings suggest that the selected frequency band (top 100 to 2000 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

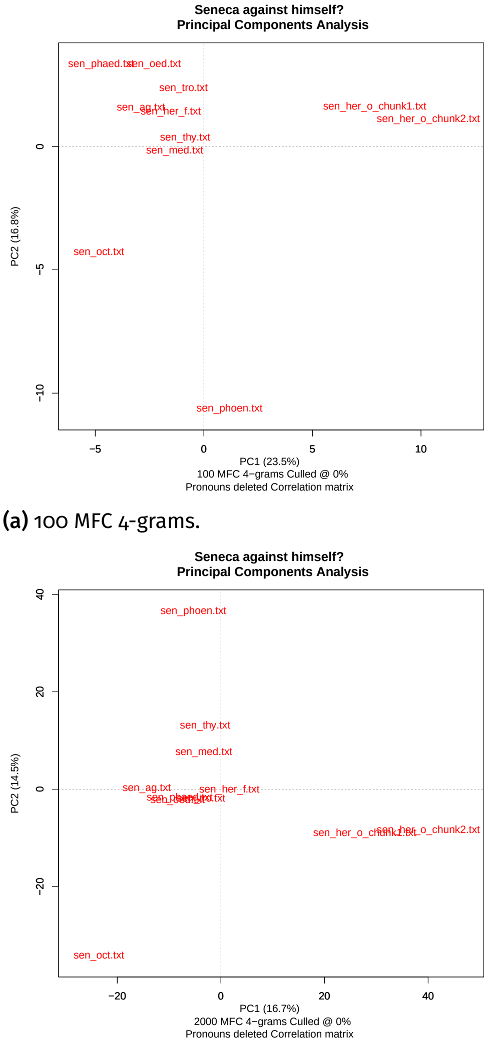
We first explore the stylometric properties of the Senecan plays using PCA, to see how they relate to each other. When treating the plays as a whole, it can be observed that



(a) 100 MFC 4-grams.

(b) 2000 MFC 4-grams.

Figure 6: PCA correlation matrix of the Senecan corpus of plays (disputed or 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’s 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) 2000 MFC 4-grams.

Figure 7: PCA correlation matrix of the Senecan corpus of plays (disputed or not), this time with *H.O.* split in half. *H.O.* starts to behave as outlier and *Oct.* remains among the outliers. It’s 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

from the two disputed texts, only *Oct.* behaves as outlier within the Senecan corpus of plays (see Figure 6). However, *H.O.* consists of 11.1147 tokens which, compared to the average size of a Senecan play (excluding *Oct.*) in terms of tokens, is almost double the size (average size of a Senecan play is 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 (refer to Figure 7). Meanwhile, *Oct.* consistently remains 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 8 shows a Bootstrap Consensus Tree (BCT) for the Senecan plays alongside two selected authors from the literature of the early empire, Lucan and Statius. Statius is included to test the hypothesis of Ferri (2003, 17–27), suggesting a temporal connection between the composition of *Oct.* and Statius. The BCT exhibits distinct branches for each author, placing both the disputed plays in proximity to the Senecan works, but *Oct.* is slightly gravitating towards the center of the unrooted tree. This again highlights the special nature of this specific text. On the other hand, *H.O.* remains among the Senecan cluster of plays.

Regarding the GI method, we test 5 different scenarios. However, since GI returns a confidence score as the final output we need to pick thresholds in order 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, 2, and 3 (see Table 4), this gives thresholds of 0.25 and 0.74 (i.e., under 0.25, Seneca is definitely not the author; above 0.74, Seneca is verified as the author; when the score is in between, no determination can be made). Unfortunately, the cross-validation method is too expensive to run with the larger datasets we use in the rest of our experiments (see scenarios 4 and 5 in Table 4) due to the nested loops and the bootstrapping that takes place which results to an increase of the time complexity of the algorithm. Therefore we will use a conservative threshold of 0.9 for all our experiments.

With the GI method, Scenario 1 and 2 confidently attribute Seneca the Younger as the author of the disputed plays (see Table 4). Next, in Scenario 3, we consider the cento-argument by Ferri (2014, 48).[10] We do this by identifying and removing lines 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

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

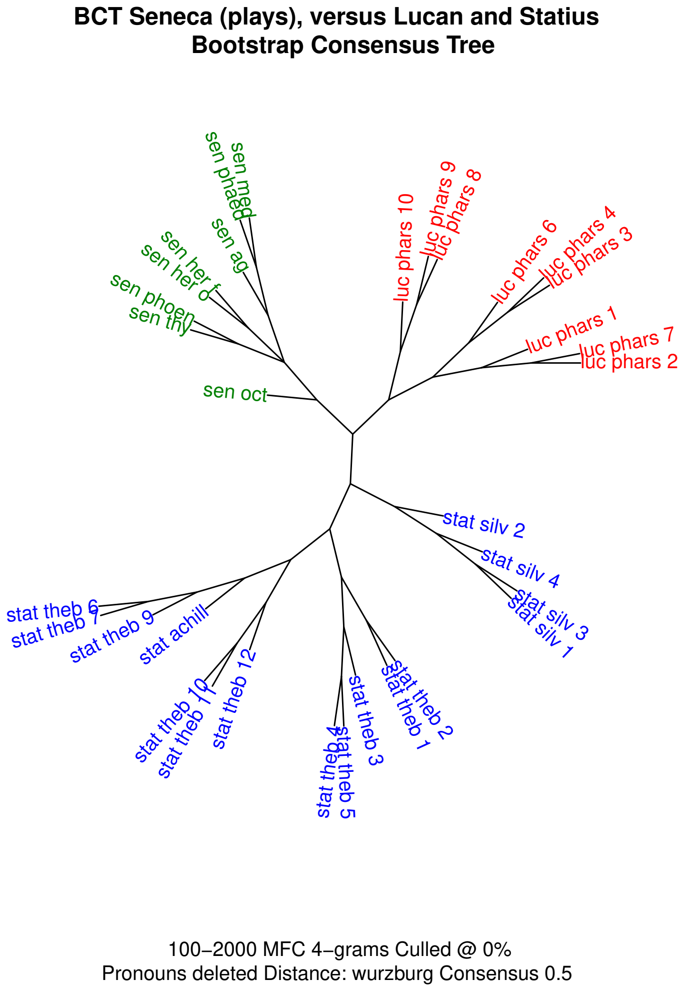


Figure 8: BCT of texts from Statius (*Achilleid*, *Thebaid*, *Silvae*), Lucan (*Pharsalia*), and Seneca (plays). The texts ‘`seneca_oct`‘ and *‘seneca_her_o‘* correspond to *Oct.* and *H.O.* respectively.

Scenario	Dataset	Results
Scenario 1: The GI method used against the disputed texts (no changes were applied to the texts per se)	90 text samples in verse written by authors that lived slightly before and after Seneca the Younger (see Figure 1 and 1).	*Octavia*: 1.0 *Hercules Oetaeus*: 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.	*Hercules Oetaeus* chunk 1: 1.0 *Hercules Oetaeus* 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.	*Octavia*: 1.0 *Hercules Oetaeus*: 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 1257 text samples.	For the scores for each chunk, see Figure 9 and 11
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 impostors dataset). The total size of this dataset including the disputed plays is 3061 text samples.	A larger dataset of mostly historiographical texts written in prose (a small number are in verse), augmented with the 500 token chunks of our main impostors dataset, resulting in 3051 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 10 and 12

Table 4: All the scenarios tested using the GI method, a brief description of the results, and the P1 & 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 'grey area' where no definitive conclusions should be drawn.

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

Table 5: 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).

exceeding a threshold of 0.6. The cosine similarity metric measures directional similarity between vectors, irrespective of magnitude or scale (Singhal et al. 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 5, wherein similarities are identified not only among various declensions of identical terms but also amid permutations in word order. For *Oct.* from a total 422 sentences, we identified and thus removed 2 (i.e., 0.46%) 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. Note that, Rolling Stylometry works by analyzing text in sequential segments to track stylistic patterns and changes over time within a document or corpus. The results for Scenario 4 (Figure 9 and Figure 11) 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 9 and 11.

For *Oct.* we observe a declining pattern in some text segments, especially for chunks 1, 3, 6, and 8 (Figure 9 and Table 6). However, excluding chunk 6 and 8 (score of 0.77), the rest of the scores are very close to 0.9 and thus the most prudent inference is that they remain of Senecan origin. Concerning chunk 6 (467-553) and chunk 8 (lines 634-733) the playwriter condenses the time in a way that seems unnatural for Seneca the Younger in

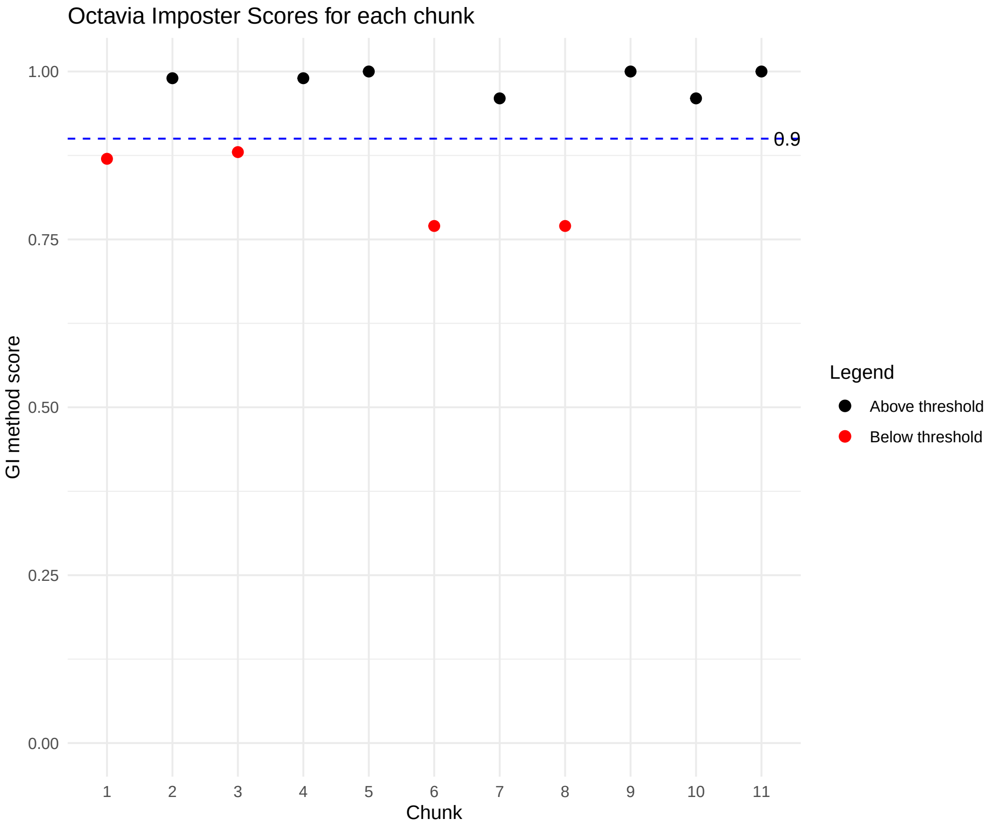


Figure 9: Results of the GI method for O's chunks (Scenario 4).

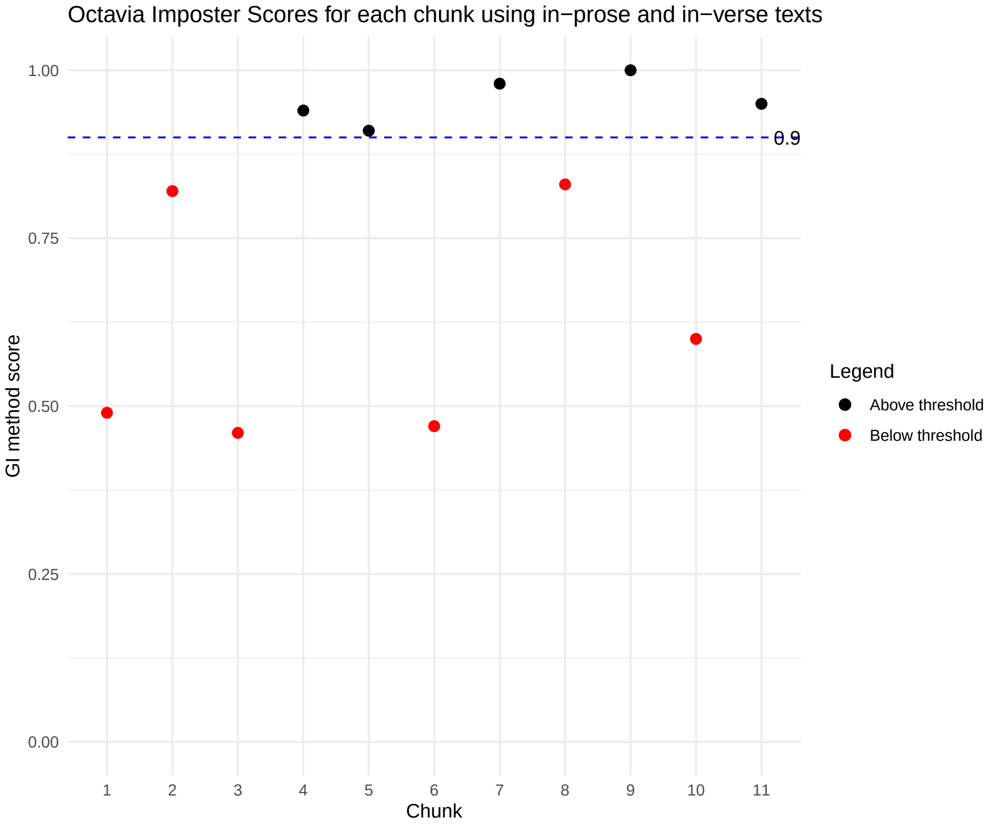


Figure 10: Results of the GI method for O's chunks using the dataset of Kestemont et al. (2016) (Scenario 5).

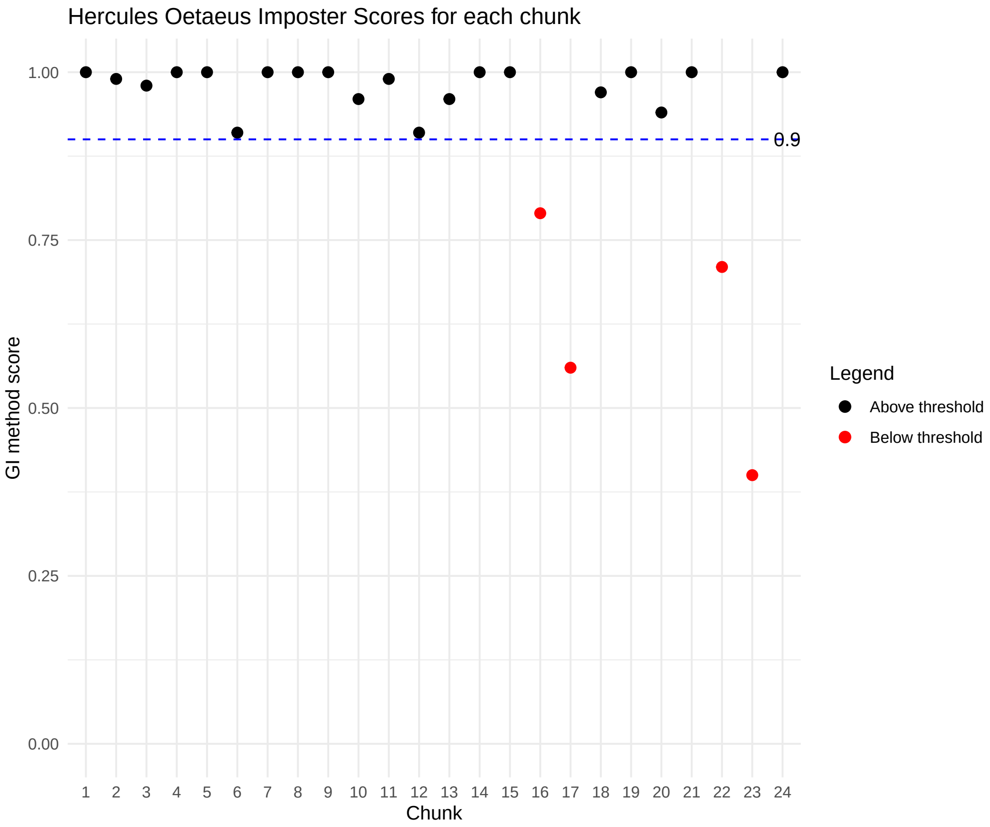


Figure 11: Results of the GI method for *H.O.*'s chunks (Scenario 4).

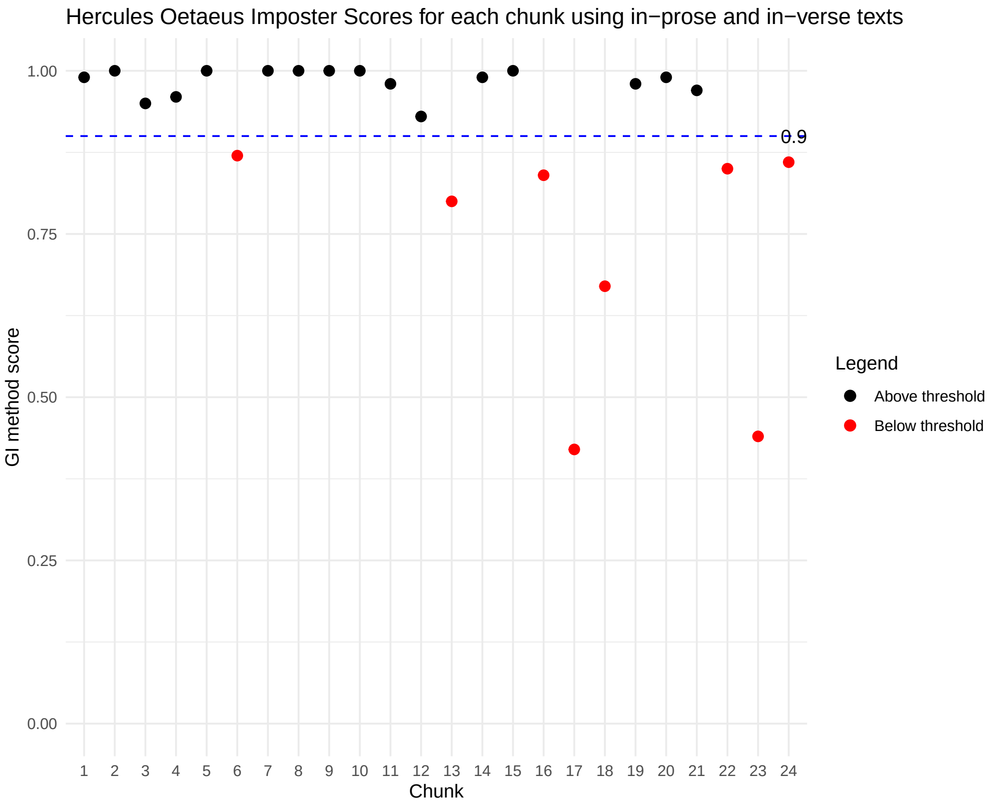


Figure 12: Results of the GI method for *H.O.*'s chunks using the dataset of Kestemont et al. Kestemont et al. (2016) (Scenario 5).

order to present a large number of events in a small amount of time (Ferri 2014, 307–309). Moreover, in both of the chunks the direct critique to Nero's reign in this passage can be considered as a task that is difficult to perform by someone (i.e., Seneca) who is working as the advisor of the emperor.

Furthermore, building upon the earlier discoveries, Figure 11 illustrates a noteworthy pattern within the *H.O.* text (see Table 7). Beyond chunk 16 (i.e., line 1297 and onwards), there is a small number of chunks with scores below the specified threshold of 0.9, indicating that they might have not been written by Seneca. This observation to some extent aligns with the hypothesis positing that the first half of the text originates from Seneca, while the remainder was finished by someone else (Tarrant 2017, 97). However, according to our results, most of the chunks in the second half were written by Seneca, which suggests that the second half is a case of mixed authorship, rather than having been completely written by someone else.

Lastly, in Scenario 5 we consider the dataset used by Kestemont et al. (2016) which mainly consists of historiographical texts that span from the 4th century B.C.E. until the 4th centure C.E.. We augment their corpus with our current corpus of impostors resulting in 3015 text samples and a mix of texts in prose and verse. Notably, the corpus also contains additional texts by Seneca (in prose). In this scenario, the texts are more dissimilar in terms of genre and chronology. On the other hand, the number of impostor authors is larger (in total 35 authors), should make it more difficult to pick out the right author and increase the reliability of the result (similar to picking out a subject from a larger police lineup). The results for *Oct.* (Figure 10) are highly similar to the results of Scenario 4 (Figure 9), where the dataset contains only texts in verse but the chunks that indicate mixed authorship grow in number (chunks 1, 2, 3, 6, 8, 10 (see Table 8)). Concerning *H.O.* (Figure 12 and Table 9), when compared against Kestemont's dataset, the signal for mixed authorship is becoming stronger too, especially after chunk 13 (lines 1027ff.). However, it should be noted again that chunks 6, 16, 22, and 24 still fall very close to the threshold of 0.9, thefefore most likely remain of Senecan origin.

Chunk no.	Lines	Score
Chunk 1	l. 1-102	0.87
Chunk 3	l. 184-276	0.88
Chunk 6	l. 467-553	0.77
Chunk 8	l. 634-733	0.77

Table 6: Chunks of *Oct.* that return a score below 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	Score
Chunk 16	l. 1319-1398	0.79
Chunk 17	l. 1398-1480	0.56
Chunk 22	l. 1819-1917	0.71
Chunk 23	l. 1918-1996	0.40

Table 7: Chunks of *H.O.* that return a score below 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	Score
Chunk 1	l. 1-102	0.49
Chunk 2	l. 102-185	0.79
Chunk 3	l. 185-276	0.46
Chunk 6	l. 467-553	0.47
Chunk 8	l. 634-733	0.62
Chunk 10	l. 825-914	0.59

Table 8: Chunks of *Oct.* that return a score below the threshold of 0.9 using Kestemont's corpus. The lines correspond to their online version in the Perseus Digital Library.

Chunk no.	Lines	Score
Chunk 6	l. 430-508	0.89
Chunk 13	l. 1027-1149	0.78
Chunk 16	l. 1319-1398	0.83
Chunk 17	l. 1398-1480	0.42
Chunk 18	l. 1480-1573	0.69
Chunk 22	l. 1819-1917	0.88
Chunk 23	l. 1918-1996	0.45
Chunk 24	l. 1970-end	0.88

Table 9: Chunks of *H.O.* that return a score below the threshold of 0.9 using Kestemont's corpus. The lines correspond to their online version in the Perseus Digital Library.

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 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, thus to mitigate their impact on the final results, we employ preprocessing techniques.

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 instances, they do not. For instance, when using only the Senecan plays, the genre and thus the meter seems to win over the authorial fingerprint and variables like the size of the plays (see the cases of *Phoenissae* and *H.O.* in Figure 6).

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 from both disputed plays lines that are similar to lines 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 with the works of Seneca cannot be explained by borrowed phrases. Nevertheless, the fourth scenario highlights segments in *Oct.* and *H.O.* that are likely not attributable to Seneca, implying the involvement of a distinct author or editor. By concentrating on the fourth GI scenario for *H.O.* (refer to Figure 9 and 11) and observing a diminishing trend in confidence after the 13th chunk, though remaining proximate to the average scores for each chunk, we posit that an editor of the text may have edited or added certain portions to 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, it is impossible to construct a perfect distractor corpus, due to limitations of extant texts. Therefore, while our empirical findings cannot positively confirm Seneca as the author of the disputed plays, our main contribution is that, perhaps contrary to expectation 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), in collaboration 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 that wrote in different meters would make it possible to quantify its 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 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, which could also detect paraphrases by taking into account semantic similarity.

9. Data Availability

Data and code: https://github.com/PaschalisAg/seneca_stylometry

10. Software Availability

Data and code: https://github.com/PaschalisAg/seneca_stylometry

11. Acknowledgements

We extend our sincere gratitude to all the anonymous reviewers for their invaluable feedback. Their insightful comments illuminated aspects of the paper that might otherwise have escaped our notice. Special thanks are due to Vasileios Dimoglidis, a PhD student 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: Conceptualization, Writing – original draft

Andreas van Cranenburgh: Formal Analysis, Writing – review & editing

References

Beullens, Pieter, Wouter Haverals, and Ben Nagy (Apr. 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, A. 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.

Daelemans, Walter (2013). “Explanation in Computational Stylometry”. In: *Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing - Volume 2*. 14th International Conference on Computational Linguistics and Intelligent Text Processing. Vol. 2. CICLing’13. Springer, 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; Issue 1*. ISSN: 1732-8160. https://www.ejournals.eu/SPL/2011/SPL-vol-6-2011/art/1171/.

— (2012). "Computational stylistics and Biblical translation: How reliable can a dendrogram be?" In: *The Translator and the Computer*. The Translator and the Computer. Ed. by Tadeusz Piotrowski and Łukasz Grabowski. Wrocław: Wyższa Szkoła Filologiczna we Wrocławiu.

— (Nov. 2013). "Does size matter? Authorship attribution, small samples, big problem". In: *Digital Scholarship in the Humanities* 30.2. _eprint: https://academic.oup.com/dsh/article-pdf/30/2/167/21517531/fqt066.pdf, 167–182. ISSN: 2055-7671. 10.1093/llc/fqt066. https://doi.org/10.1093/llc/fqt066.

— (Sept. 1, 2016). "Rolling stylometry". In: *Digital Scholarship in the Humanities* 31.3, 457–469. 10.1093/llc/fqv010.

— (2018). *Authorship verification with the package stylo*. Computational Stylistics. https://computationalstylistics.github.io/docs/imposters.

— (2022). "Boosting word frequencies in authorship attribution". In: *arXiv e-prints*. 10.48550/arXiv.2211.01289.

Eder, Maciej and Jan Rybicki (June 1, 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 (Dec. 1, 2017). "Understanding and explaining Delta measures for authorship attribution". In: *Digital Scholarship in the Humanities* 32 (suppl_2), ii4–ii16. 10.1093/llc/fqx023.

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

— (Jan. 1, 2014). "Octavia". In: Brill, 521–527. 10.1163/9789004217089_043.

Frank, M. (July 17, 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.

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, M. (2021). *Real-World Natural Language Processing: Practical Applications with Deep Learning*. Manning.

Herington, C. J. (1961). "Octavia Praetexta: A Survey". In: *The Classical Quarterly* 11.1, 18–30. 10.1017/S0009838800008351.

Hoover, David L. (Nov. 1, 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*. ADHO. UWS. http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empi/JANNIDIS_Fotis_Improving_Burrows__Delta___An_empirical_.html.

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*. Association for Computational Linguistics, 20–29. 10.18653/v1/2021.acl-demo.3.

Jolliffe, Ian T. and Jorge Cadima (Apr. 13, 2016). “Principal component analysis: a review and recent developments”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065, 20150202. 10.1098/rsta.2015.0202.

Juola, Patrick (Dec. 1, 2015). “The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions”. In: *Digital Scholarship in the Humanities* 30 (suppl_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*. Vol. 35. Trends in Classics. De Gruyter. 335 pp. 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)*. 3rd Workshop on Computational Linguistics for Literature (CLFL). Association for Computational Linguistics, 59–66. 10.3115/v1/W14-0908.

Kestemont, Mike, Sara Moens, and Jeroen Deploige (June 1, 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 (Nov. 30, 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: *CLEF (Working Notes)*, 977–983. http://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-KonijEt2014.pdf.

Koppel, Moshe, Jonathan Schler, and Shlomo Argamon (Jan. 1, 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 (Jan. 1, 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*. 5th ed. Springer, 600.

Luyckx, Kim and Walter Daelemans (Apr. 1, 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, C.W. (2014). "The Works of Seneca the Younger and Their Dates". In: 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*. 1st ed. London: Routledge, 130–141.

Mikros K., George (2018). "Blended Authorship Attribution: Unmasking Elena Ferrante Combining Different Author Profiling Methods". In: *Drawing Elena Ferrante's profile*. Padova University Press, 85–96.

Newman, Matthew L., Carla J. Groom, Lori D. Handelman, and James W. Pennebaker (n.d.). "Gender Differences in Language Use: An Analysis of 14,000 Text Samples". In: *Discourse Processes* 45.3 (), 211–236. 10.1080/01638530802073712.

Nolden, Luuk (July 19, 2019). "Finding Seneca in Seneca: using Text Mining techniques of Hercules Oetaeus and Octavia". Bachelor Thesis. Leiden, The Netherlands: Leiden Institute of Advanced Computer Science (LIACS). https://theses.liacs.nl/pdf/2018-2019-NoldenLSJ.pdf.

Päpcke, Simon, Thomas Weitin, Katharina Herget, Anastasia Glawion, and Ulrik Brandes (Aug. 9, 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. Publisher: The Classical Association of the Middle West and South, 388–403. http://www.jstor.org/stable/3288405.

Perseus Digital Library (2024). Ed. Gregory R. Crane. Tufts University. https://www.perseus.tufts.edu/hopper/ (visited on 05/14/2024).

Philp, R. H. (1968). "The Manuscript Tradition of Seneca's Tragedies". In: *The Classical Quarterly* 18.1. Publisher: [Classical Association, Cambridge University Press], 150–179. http://www.jstor.org/stable/637696.

Poe, Joe Park (1989). "Octavia Praetexta and Its Senecan Model". In: *The American Journal of Philology* 110.3, 434–459. 10.2307/295219.

Potha, Nektaria and Efstathios Stamatatos (2017). "An improved impostors method for authorship verification". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings 8*. Springer, 138–144.

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 (Dec. 1, 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*. Springer International Publishing, 191–210. 10.1007/978-3-030-53360-1_8.

Seidman, Shachar (2013). "Authorship verification using the impostors method". In: *CLEF 2013 Evaluation labs and workshop–Working notes papers*, 23–26.

Seneca (Apr. 17, 2008). *Octavia: Attributed to Seneca*. Place: Oxford Publisher: Oxford University Press. http://oxfordscholarlyeditions.com/view/10.1093/actrade/9780199287840.book.1/actrade-9780199287840-book-1.

Singhal, Amit et al. (2001). "Modern information retrieval: A brief overview". In: *IEEE Data Eng. Bull.* 24.4, 35–43.

Stamatatos, Efstathios (Mar. 1, 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.

Stamatatos, Ph D et al. (2013). "On the robustness of authorship attribution based on character n-gram features". In: *Journal of Law and Policy* 21.2, 7.

Star, Christopher (Jan. 1, 2015). "Roman Tragedy and Philosophy". In: 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. Edition: 2017/01/30 Publisher: Cambridge University Press, 645–672. 10.1017/S0009838816000768.

Stover, Justin, Yaron Winter, Moshe Koppel, and Mike Kestemont (Jan. 1, 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. ISSN: 2330-1635. 10.1002/asi.23460.

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

The Latin Library (2024). http://www.thelatinlibrary.com/ (visited on 05/13/2024).

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