The Un-Platonic *Menexenus*:
A Stylometric Analysis with More Data

Thomas Koentges

THE *CORPUS PLATONICUM* is one of the most well-known and influential works of ancient literature, and yet it still presents unresolved challenges regarding the authorship of some works. Since the middle of the nineteenth century, stylometry, the counting and statistical analysis of features in the language of an author, has been applied to the *Corpus Platonicum* in order to date works by Plato, to prove or disprove the authenticity of his work, or to analyze how he used different styles (for instance, to characterize different individuals in his dialogues).[1] This article revisits the second of these applications with a special focus on the question of the authorship of the *Menexenus*.

The results of previous stylometric analyses of the *Corpus Platonicum* were inconclusive on the question of authorship.[2] One reason for this probably lies in the selection of features, which covers the frequency of certain words, the number of words in a sentence, the quantities of *clausulae*, or any other element of the text that can be observed and counted and that is thought to be part of the stylistic fingerprints of the author. Previously, analyses employed a limited number of features, but as Burrows has observed, "in this sort of work on language … a wealth of variables, many of which may be weak discriminators, almost

---

[1] A good example of the latter has been published in this journal: D. Sansone, "Stylistic Characterization in Plato: Nicias, Alcibiades, and Laches," *GRBS* 58 (2018) 156–176.

[2] A more complete survey than can be given here was undertaken by L. Brandwood, "Stylometry and Chronology," in R. Kraut (ed.), *Cambridge Companion to Plato* (Cambridge 1992) 90–120.

always offer more tenable results than a small number of strong ones."[3] In the first non-computational stylometric analyses of Plato the number of extracted features was very low: Campbell and Dittenberger each counted eight lexical features and Billig used fourteen metric features.[4] One of the most extensive non-computational stylometric analyses was performed by Ritter using forty-three linguistic features.[5] Unfortunately, their studies only partially help to evaluate the authenticity of the *Menexenus*: Campbell did not focus on that work because the aim of his study was the analysis of the *Sophista* and the *Politicus*; Billig did not include all of the *Corpus Platonicum* and one of the dialogues that was omitted was the *Menexenus*; Dittenberger did not include it because it was considered inauthentic by most scholars in Germany at the time; and while Ritter saw similarities between the *Menexenus* and the *Symposium*, his research only included works in the *Corpus Platonicum* and therefore lacked textual comparanda.[6] In addition, Ritter admitted that the number of features he observed in the *Menexenus* was low.[7]

In one of the first computational stylometric analyses of the *Corpus Platonicum*, G. R. Ledger used thirty-seven lexical features as well as comparison texts. He found unusual results for the

---

[3] J. Burrows, " 'Delta'—A Measure of Stylistic Difference and a Guide to Likely Authorship," *Literary and Linguistic Computing* 17 (2002) 268.

[4] L. Campbell, *The Sophistes and Politicus of Plato* (Oxford 1867) I–XLV; W. Dittenberger, "Sprachliche Kriterien zur Chronologie der platonischen Dialoge," *Hermes* 16 (1881) 321–345; L. Billig, "Clausulae and Platonic Chronology," *JP* 35 (1920) 225–256.

[5] C. Ritter, *Untersuchungen über Plato. Die Echtheit und Chronologie der platonischen Schriften* (Stuttgart 1888), and *Platon: Sein Leben, Seine Schriften, Seine Lehre* (Munich 1910), esp. 232–272.

[6] Dittenberger, *Hermes* 16 (1881) 322 n.1; Ritter, *Platon* 496.

[7] Ritter, *Untersuchungen* 97–98; cf. 125, where he assumes that the noticeable change in content in the *Menexenus* might be evidence that Plato grew happier again: "Auch kann ich mich nicht entschliessen den *Menexenus* als unecht zu verwerfen, der wegen seiner leichten Haltung in die zweite Hälfte meiner ersten Periode nicht hereinpassen will. Ich muss so annehmen, dass … eine heitere und freiere Gemütsstimmung sich bei Plato herstellte."

———

Menexenus compared with other Platonic works; however, he did not trust those results, for he saw authenticity confirmed by Aristotelian references that may point to the *Menexenus*.[8] In fact, he excluded the *Menexenus* from a specific authenticity test that would have explicitly treated the work as independent from the *Corpus Platonicum*, writing that this was "partly through oversight, partly because what I am attempting here is illustrative rather than definitive, and partly because I believe [it] to be genuine" (105). In addition, the computational analysis of Greek works was more difficult thirty years ago. Given the challenges of encoding and therefore also counting Greek words in the 1980s, Ledger defined features based on the first and last letters of a word, arguing that this would be the best way to observe lexical features of an inflected, morphologically complex language when limited by processing and memory power.

Empowered by technological progress, however, the last decade of Natural Language Processing research has developed promising automated analytical methods to process and classify vast amounts of texts. For instance, the R-Package *stylo* makes it easy to extract lexical features, such as the frequency of words or the frequency of character n-grams (that is, a sequence of characters, where a 4-gram, for instance, consists of four characters).[9] Both feature sets—frequency of words and character n-grams—are complementary to each other. While the frequency of words almost entirely focuses on function words (that is, words with little semantic meaning that still mark a stylistic decision by the author, such as καὶ, δὲ, μὲν, τε), measuring the frequency of sequences of four characters will also show, for instance, an author's preference for certain compounds (e.g. περὶ, περίπλεω), certain endings, or even sounds.

Furthermore, equally empowered by technological progress, a

[8] G. R. Ledger, *Recounting Plato: A Computational Analysis of Plato's Style* (Oxford 1989) 163.

[9] M. Eder, J. Rybicki, and M. Kestemont, "Stylometry with R: A Package for Computational Text Analysis," *R Journal* 8 (2016) 107–121.

lot of machine-actionable textual data has been produced by the Perseus Digital Library and the Open Greek and Latin Project (OGL), whose staff has digitized and curated not only most of the *Corpus Platonicum* and the Greek authors commonly read in schools and universities, but also, in collaboration with the Center for Hellenic Studies (CHS),[10] digitized extant works of the first one thousand years of Greek literature.[11] Now, for the first time in history, it is possible to use computational analysis for the majority of Ancient Greek literature.

Using more data also enables a better understanding of Platonic style. While it is difficult to compare Platonic works with themselves, it is possible to detect features with greater discriminatory power when we compare as many works as possible. This observation is not new: in 1989 Ledger not only used works included in the *Corpus Platonicum* but also other prose works of the fifth and fourth centuries by Aeschines, Isaeus, Isocrates, Lysias, Thucydides, and Xenophon. Yet he admitted that his experiment ignored the component of corruption in the process of textual transmission. While I agree that given the amount of data some individual corruptions would not greatly affect the overall result, by preselecting only texts from the fifth and fourth centuries, Ledger additionally ignores (without mentioning it) the possibility of a later forgery of individual works. As we now have more data, I ran stylometric experiments without preselecting a certain style of Ancient Greek. The machine now can easily cluster (i.e. put the works into groups that have statistically similar features) larger groups and we do not run the risk of influencing the results through a priori selection bias.

Another advantage of more data is that morphological normalization (that is, reducing each word to its dictionary entry) is

[10] T. Koentges, "Computational Analysis of the *Corpus Platonicum*," *CHS Research Bulletin* 6.1 (2018), at http://www.chs-fellows.org/2018/04/30/report-corpus-platonicum/.

[11] For this part of the corpus see http://opengreekandlatin.github.io/First1KGreek/. The Perseus text can be retrieved from https://github.com/PerseusDL/canonical-greekLit. Most of the texts are also accessible using the Scaife viewer: https://scaife.perseus.org.

less urgent. The problem with analyzing morphologically complex languages with statistical tools is essentially a feature-frequency problem; a feature will be all but ignored if it occurs only a statistically irrelevant number of times. In Ancient Greek the same lexical unit can have more than ten different morphological alternatives. For the computer all of those are distinct features: Πλάτωνα differs from Πλάτων as much as Εὔβουλος if we choose to select the frequencies of morphologically unnormalized words for our statistical stylometric analysis. But by including more data the most frequent words are less prone to become statistically irrelevant. In fact, even a word's absence becomes a statistically relevant feature if we can otherwise find it among the most frequent words in the corpus. Additionally, measurements of the frequency of 4-character sequences are also less affected by this kind of feature-frequency problem. Both kinds of features were independently used in my experiments.

In what follows I describe a thorough series of stylometric experiments on the machine-actionable corpus of Ancient Greek text produced by Perseus, OGL, and the CHS and will focus on one finding of potentially many more to come. This finding is that although the *Menexenus* was transmitted in the Platonic canon of the tetralogies, there is very little stylistic similarity between it and the rest of the *Corpus Platonicum*. While this in itself does not necessarily prove that the *Menexenus* was not written by the authorial entity identified as Plato, it strongly supports this view if its inclusion cannot be otherwise justified by philological or philosophical arguments.

Data and method

This experiment used the corpus generated by Perseus, OGL, and CHS as of December 2017.[12] At the time of research, the corpus contained almost thirty million words in 1044 works.

---

[12] This corpus includes CTS identifiers for all works and workgroups. For instance, tlg0059 stands for the *Corpus Platonicum* and tlg0059.tlg028 for the *Menexenus*. As an aid to the reader in most visualizations I have replaced those identifiers with abbreviations for the author and the work.

Because they are not all entirely in Ancient Greek (e.g., Perseus includes some translations in their repository), I filtered out all non-Greek works, which reduced the corpus to twenty-three million words, and I then reduced this further by including only works with more than 1500 words of Ancient Greek.[13] This resulted in 825 works with a total of 23,113,247 words. However, this also included some works with more than one version, either because Perseus digitized more than one version or the modern language translation still contained more than 1500 Greek words. In such cases I selected the version that contained more words. Ultimately, the data used for the stylistic experiments comprised 800 works and about 22,250,000 words, which makes this research the largest stylometric experiment for Ancient Greek at the time of writing.[14]

From this corpus I extracted twenty-five different feature sets. A feature set is the frequency of certain features in the corpus. Authorship can be deduced from the preferences of authors for certain words (or other features). The process of feature-set and measure selection is further described below in the technical appendix.

Results of the statistical stylometric analysis

In the experiments I pursued two independent approaches. One is a t-SNE dimensionality reduction approach that simplifies the 1000-dimensional space to a 2-dimensional space to

[13] At a talk at DH2017, M. Eder suggested that a minimum of 2000 words be used for English. I set the threshold to 1500 so as to not exclude Plato's *Clitophon*. I have since repeated the experiments by setting higher thresholds of 2000 and even 3000 words, but the results do not change regarding the *Corpus Platonicum*. See M. Eder, "Short Samples in Authorship Attribution: A New Approach," a paper given at the DH2017 conference in Montreal; abstract at https://dh2017.adho.org/abstracts/341/341.pdf.

[14] I have made the data and the R-programming script that wrangled and analyzed the data available under open licenses at GitHub and Zenodo (MIT for the script and CC-BY for tables and visualizations) to allow readers to reproduce the experiment or expand on it: Thomas Koentges (2019), ThomasK81/PlatoStylometricAnalysis: Post-Peer-Review (Version 1.0.0), Zenodo, http://doi.org/10.5281/zenodo.3450542.

enable us to use simple Euclidean distances to cluster the data. The other approach measures several types of distances between points directly in the high-dimensional space and then uses those distances to cluster the data. Further information on these approaches and the agglomerative clustering used for both (Ward's method) is included in the appendix.

When using the methods based on t-SNE, the *Menexenus* clearly lies outside the cluster of Platonic works in almost all visualizations.[15] In two feature sets (500 and 300WekaWords) it sits in the proximity of the Platonic cluster. In general, the *Menexenus* does not cluster with many works; rather, it sits somewhere close to the works of Lucian and sometimes Plutarch and even Appian. That said, global distances are not very well preserved in the t-SNE visualization and the main finding is that the *Menexenus* does not generally cluster with other Platonic works. Aside from the *Menexenus*, two other works often appear outside the main cluster of Platonic works: the *Timaeus* and *Critias*. These almost always cluster with each other and are sometimes very close to the main cluster. At times, the *Clitophon* is also outside the main cluster.

Similar observations can be made when we look at the results of the second method: the cluster analysis of all distances. The *Menexenus* rarely clusters with Plato; indeed, it does so only in approximately 10% of the feature-set and measure combinations. In four additional cases it clusters alone, but after further examination of the hierarchical tree it can be seen that the parent cluster includes works from the *Corpus Platonicum*. The *Menexenus* also rarely clusters with more than two Platonic works simultaneously, often doing so with just one other Platonic work, as shown in TABLE 1.

While this already suggests that attribution of the *Menexenus* to Plato is not well supported by this analysis, the result becomes even more decisive if we compare the scores shown in TABLE 1

---

[15] Given their size and resolution it is not practical to include these visualizations in a print publication; they can be accessed at Koentges, Zenodo repository and Github, e.g. https://raw.githubusercontent.com/ThomasK81 /PlatoStylometricAnalysis/master/plots/MF4Gram1000t-SNE.png.

TABLE 1: Number of works in the *Corpus Platonicum* that cluster with the *Menexenus* in each feature-set and measure combination. Four cases where the *Menexenus* clusters alone, but is part of a Platonic parent cluster, are not specifically marked in this table but were included in the analysis; those are t-SNE at 500WekaWord and 300WekaWord, Eder's Delta at 810MenexMFW, and Eder's Simple at 1,000WekaWord.

	Euclidean	Jaccard	BurrowsDelta	EdersSimple	Cosine	CosineDelta	EdersDelta	ArgamonsDelta	t.SNE
1000MFW	0	0	0	0	0	0	0	0	0
500MFW	0	0	0	0	0	0	0	0	0
300MFW	0	0	0	0	0	1	0	0	0
200MFW	0	0	0	0	0	1	0	0	0
100MFW	0	0	0	1	0	1	0	0	0
1000MF4Gram	0	0	0	0	0	0	0	0	0
500MF4Gram	0	0	0	0	0	0	0	0	0
300MF4Gram	0	0	0	0	0	0	0	0	0
200MF4Gram	0	0	0	0	0	0	0	0	0
100MF4Gram	0	0	0	0	0	0	0	0	0
1000WekaWord	0	0	0	0	0	4	0	0	0
500WekaWord	0	1	0	0	0	0	0	8	0
300WekaWord	0	0	0	1	0	0	0	0	0
200WekaWord	0	0	0	0	0	2	2	3	0
100WekaWord	0	1	2	1	0	4	2	0	0
859Weka4Gram	0	0	0	0	0	0	0	0	0
500Weka4Gram	0	0	0	0	0	0	0	0	0
300Weka4Gram	0	0	0	0	0	0	0	0	0
200Weka4Gram	0	0	0	0	0	0	0	0	0
100Weka4Gram	0	0	0	0	0	0	3	0	0
810MenexMFW	0	0	0	0	0	0	0	0	0
500MenexMFW	0	0	0	0	0	3	0	0	0
300MenexMFW	0	0	0	1	0	0	0	1	0
200MenexMFW	0	0	0	1	0	1	0	0	0
100MenexMFW	0	0	0	7	0	1	0	0	0

———

with the individual scores of the feature-set and measure combinations in TABLE 2. The instances where the *Menexenus* clusters with Plato often have lower k-scores and there is no instance of the *Menexenus* clustering with another Platonic work when any of the five most frequent 4-gram sets were used (the feature sets with the highest mean k-score).[16] Only three of the top twelve highest-scored feature-set and measure combinations show a cluster that includes the *Menexenus* and one other Platonic work

---

[16] The k-score is a cluster score between 0 and 100 generated by testing the stylometric analysis results against the library metadata of the dataset. The higher the score the better, although a score of 100 is only theoretically possible (because of mistakes in the metadata or workgroups that only include one work). See the technical appendix.

TABLE 2: K-scores by feature-set and measure combination. The scores where the *Menexenus* clusters with another work of the *Corpus Platonicum* have a grey background.

	Euclidean	Jaccard	BurrowsDelta	EdersSimple	Cosine	CosineDelta	EdersDelta	ArgamonsDelta	t.SNE
1000MFW	76	75.38	78.12	78.62	75.25	78	78.25	76.5	74.75
500MFW	77	75.88	77.88	78.38	75.12	78.38	77.75	77.5	74
300MFW	76.12	75.38	77	77.75	74.62	78.75	78.5	76.12	73.5
200MFW	75.75	75	77.62	79.38	75.25	77.12	77.88	78.38	73.75
100MFW	75	74.12	78.5	79.25	74.75	78.5	78.12	77.38	73
1000MF4Gram	77.25	77.75	77.25	78	77.5	78.25	78.38	77.38	75.75
500MF4Gram	77.75	77.75	78.62	79.12	77.88	77.12	78	77.62	75.25
300MF4Gram	76.75	76.88	77.75	78.75	76.88	78	77.88	77.12	74.88
200MF4Gram	75.75	76.38	77	77.88	76.38	77.62	78	77.12	74.5
100MF4Gram	73.88	73.12	76.38	77	73.88	74.88	75.38	75	71.75
1000WekaWord	73.62	73.88	73.88	77.12	74.75	76.88	74.25	71.88	72.75
500WekaWord	75.25	75.62	72.75	76.75	74.75	76.62	73.62	71.25	74.88
300WekaWord	73.75	74.75	73.38	77.25	72.88	76.5	74.5	70	74.12
200WekaWord	73.12	73.75	73.5	77.12	72.25	74.88	72.88	68.25	71.62
100WekaWord	70.12	70.5	72.62	76.25	71.25	71.5	69.25	66.5	71
859Weka4Gram	76.88	77.12	76.75	78.38	76.12	79.62	77	77.25	74.62
500Weka4Gram	76.38	76.38	77.38	77.5	76.62	78.75	77.38	77.62	75.12
300Weka4Gram	75.75	76.5	76.88	77.75	75.88	77.88	77.38	75.38	75.25
200Weka4Gram	76.25	76.12	77.75	78.38	75.38	78.62	76.75	77.62	76
100Weka4Gram	75.25	76.25	76.75	79.25	74.38	76	75.5	74.5	74.12
810MenexMFW	75.88	74.88	74.75	77.38	74.38	76.25	75.88	70.75	73.25
500MenexMFW	74.75	75.25	76.12	78.12	74.5	77	77.25	73.75	73.25
300MenexMFW	75.12	74.75	76.5	78.62	75	76.75	78	77	73
200MenexMFW	74.75	75.25	77.5	78.12	74.38	77.5	78.38	73.88	72.88
100MenexMFW	74.88	74.25	78.38	78.38	74.75	77.12	77.88	76.25	72

———————

(*Crito* twice and the *Letters* once). The only time a single cluster contains all works of the *Corpus Platonicum* including the *Menexenus* occurs with Argamon's Delta at 500 WekaWord. The Argamon's Delta measure, however, is the most unreliable measure used in the analysis and this particular combination ranks 215 out of 225 in reliability (measured with the k-score).

To undertake a more detailed analysis, I constructed a correlation matrix for the individual feature sets for all Platonic works. In the example given in *fig.* 1, which shows the correlation matrix for the most-frequent 4-grams, the exclusively pale peach row clearly shows that the *Menexenus* does not have a positive correlation with other Platonic works. The *Parmenides* is another dialogue that often has a weaker connection to the rest of the *Corpus Platonicum*. Additionally, we can see that there are Platonic works that do correlate. For instance, the *Sophista*, *Politicus*, *Philebus*, *Leges*, *Epinomis*, and *Letters* show strong correlation (dark blue on the matrix in *fig.* 2). These are also works that are tradi-
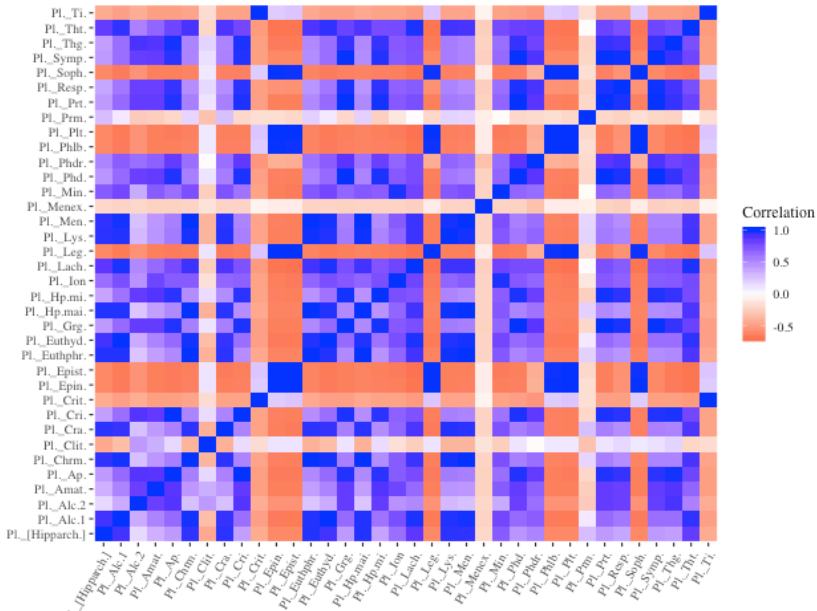
———————

Figure 1: Correlation matrix for the *Corpus Platonicum* analyzed with the MF4-gram feature sets, showing that the *Menexenus* does not have a positive correlation with the other works.

————

tionally considered to be late Platonic dialogues, three of which are characterized by an almost complete absence of dialogue. Similarly, the *Clitophon* shows only a weak correlation with other Platonic dialogues, although when using some other feature sets the correlation is stronger. The *Clitophon* shows the strongest correlation with the *Menexenus*, but it is still only very weak. The *Timaeus* and *Critias* always correlate strongly; here they also have a weak correlation with the late dialogues.

Furthermore, I employed the correlation with other Platonic texts to rank all works of the *Corpus Platonicum* according to their 'Platonic-ness' of style. *Fig.* 2 thus shows that the most 'Platonic' work seems to be the *Theaetetus*; however, when I used other feature sets, *Laches* and *Gorgias* also topped the group. While there was variance among the most 'Platonic,' every feature set ranked the *Menexenus* last, with very low correlation to other Pla-
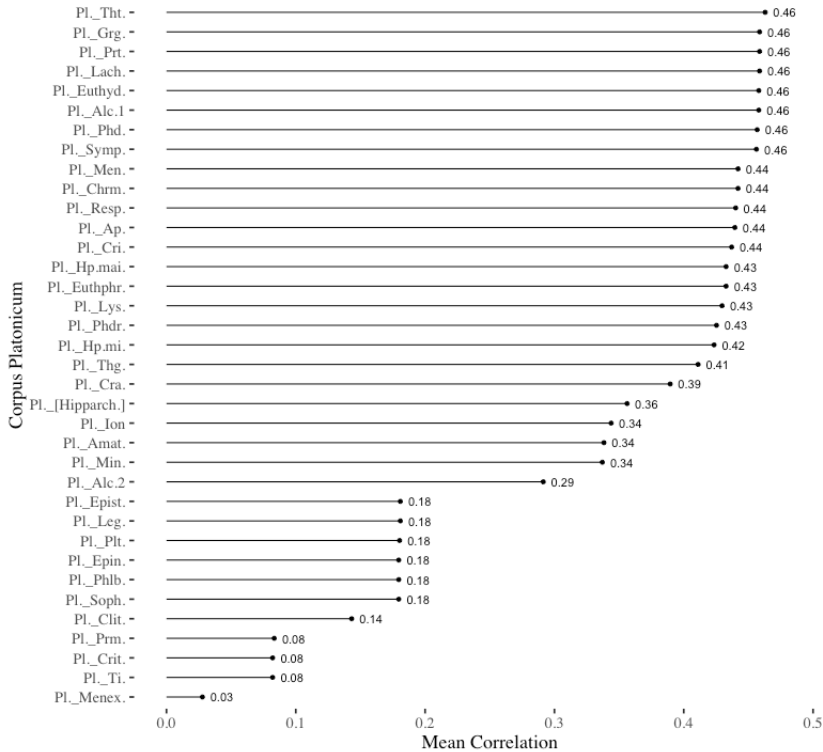
Figure 2: Works of the *Corpus Platonicum* analyzed with the MF4-gram feature sets and ranked by their correlation with other works.

————

tonic works. In fact, the *Menexenus* clusters better with Xenophon, albeit not convincingly (see *fig.* 3), or even Lucian, which is impossible because the *Menexenus* was written before Lucian was born.[17]

While those results are very conclusive,[18] to make the results more tenable for philologists working qualitatively I was also interested in whether I could find words where the *Menexenus* is a clear outlier of the *Corpus Platonicum*. Outliers in a data set are

[17] For more visualizations of the data supporting this see Koentges, Zenodo repository.

[18] Burrows, *Literary and Linguistic Computing* 17 (2002) 268.

————
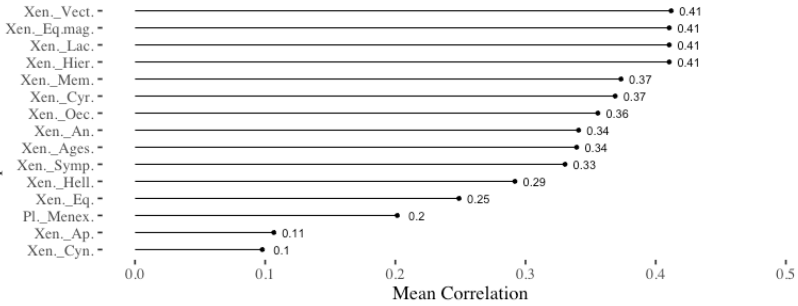
Figure 3: Works by Xenophon and the *Menexenus* analyzed with the MF4-gram feature sets and ranked by their correlation with other works.

———

values that differ significantly from the other observations. In statistics, every value that is 1.5 times the interquartile range below the lower or above the upper quartile can be considered an outlier.[19] In the table of the 5000 most frequent words generated by *stylo*, one can identify 294 clear outliers (345 for the most frequent 4-grams). Instead of looking at all of those outliers in detail, for reasons of space I chose to focus on the words that are extremely common in the *Corpus Platonicum*. Hence, for this article, I reduced the number of outliers to words that occur at least once every 500 words (see *fig.* 4). The full data is available on Zenodo.[20]

In *fig.* 4, the fact that καί is far more frequent and δὲ is far less frequent in the *Menexenus* than in the *Corpus Platonicum* as a whole shows a different preference for connecting parts of sentences that are on the same level. Furthermore, differentiating between just the speech in the *Menexenus* or only the preceding and concluding dialogue has almost no effect, for even in the dialogue the frequency of καί (6.42%) is at the upper limit of what has been observed throughout the *Corpus Platonicum*. Admittedly, the dialogues are not long enough to make the latter observation reliable, but the more general point still stands. Furthermore, if we

[19] Calculated with the R function *boxplot.stats*.

[20] Koentges, Zenodo repository.

———
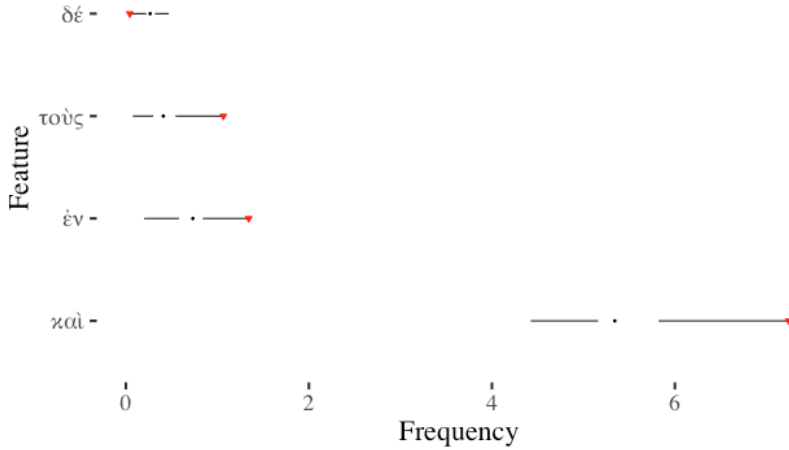
Figure 4: Most frequent words in the *Corpus Platonicum* that occur at least once every 500 words for which the *Menexenus* is an outlier. The frequency is given in occurrence per 100 words. The red triangle shows the value for the *Menexenus*.

———————

look only at the character Socrates' use of καὶ throughout the *Corpus Platonicum*, then the *Menexenus* is a clear outlier too (see *fig.* 5). This claim becomes even stronger if we follow Sansone's argument that Plato changes his usage of conjunctions like καὶ as specific linguistic markers to enhance the portrayal of his characters.[21] It must be emphasized, however, that this is only of limited value for the *Menexenus*, since the preponderance of the work is a speech in which Socrates is attempting to speak in the style of Aspasia.

Ledger's suggestion, however, that such differences might be explained by a genre signal is most likely wrong.[22] Although his suggestion sounds tempting, this hypothesis can be dismissed because: first, in the t-SNE visualizations orators like Lysias, Libanius, and Demosthenes have the tendency to cluster in proximity, but the *Menexenus* does not; and second, when I repeated the co-occurrence experiment by pairing the *Menexenus*

[21] Sansone, *GRBS* 58 (2018) 173.

[22] Ledger, *Recounting Plato* 163.

———————
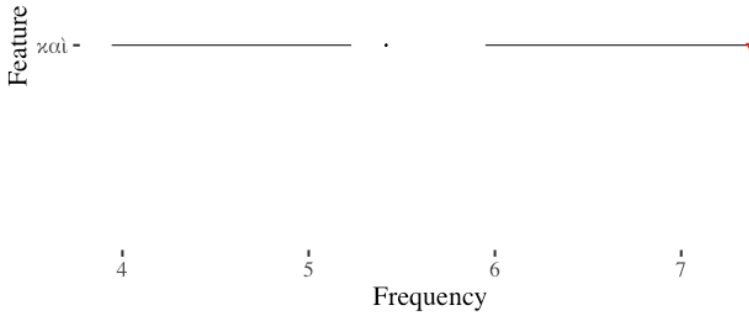
Figure 5: Use of καὶ by the character Socrates in the *Corpus Platonicum*. The frequency is given in occurrences per 100 words. The red triangle shows the value for the *Menexenus* (including the oration).

———————

with the orators, it did not show a positive co-occurrence, and when I repeated the outlier experiment for the orators and the *Menexenus*, the use of καὶ and ἐν is as unusual in that context as it is when compared with the works of the *Corpus Platonicum* (see *fig.* 6). On the basis of this large-scale analysis Ledger's hypothesis can be dismissed.
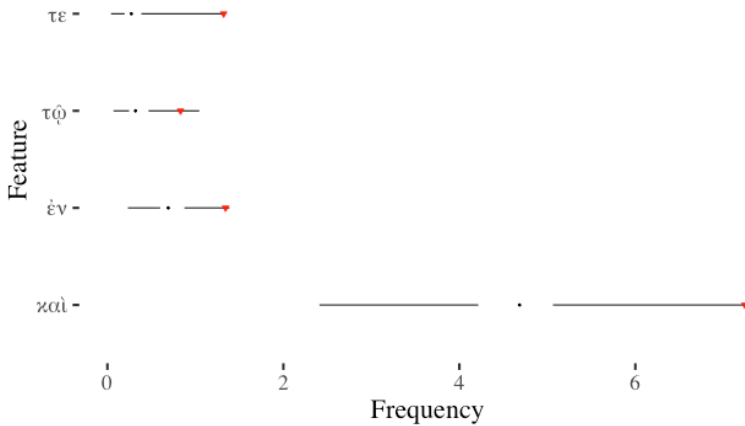
Figure 6: Most-frequent words used by the orators and the *Menexenus* that occur at least once every 500 words, where the latter is an outlier. The frequency is given in occurrences per 100 words. The red triangle shows the value for the *Menexenus*.

———————

Relying on these tests with more data, we can now offer significant new stylometric findings. Specifically, 201 out of 225 combinations of feature set and measure do not cluster the *Menexenus* with other works of the *Corpus Platonicum*. Furthermore, as mentioned above, nine of the twelve highest-scored feature set and measure combinations do not do so either. Instead, the *Menexenus* has a higher co-occurrence with Xenophon and Lucian than with other works in the *Corpus Platonicum*. Its use of καὶ, for instance, is unusual when viewed against other works in the *Corpus Platonicum* and the regular usage of Plato's Socrates. In combination, these findings seem to indicate strongly that the *Menexenus* was not written by Plato. Therefore, if the philological arguments for Platonic authorship of the *Menexenus* are not beyond reasonable doubt, we can safely conclude that the work was written by someone else.

The philological arguments for and against attributing
   the Menexenus *to Plato*

As mentioned above, for most of the nineteenth century the *Menexenus* was not thought to have been written by Plato. This changed when influential philologists like Ulrich von Wilamowitz-Moellendorff reversed their opinion and declared the work Platonic.[23] Since then, all philological defenses of the view that the *Menexenus* was written by Plato broadly follow the same pattern. One of the best summaries is given by Stavros Tsitsirides,[24] who divided the arguments into three categories: first, ancient works attribute the *Menexenus* to Plato; second, stylistic criteria support this attribution; and third, the *Menexenus* exhibits textual connections to other Platonic works.

The chief evidence for the first category can be reduced to two passages of Aristotle that may invoke the *Menexenus*. The rest date to at least three hundred years later. One of these, Cic. *Orat.* 151, clearly attributes a eulogy to Plato, which—according to Cicero, the only witness I could find for this statement—was

[23] E.g. Wilamowitz, *Platon: Beilagen und Textkritik* (Berlin 1920) 126 n.2.

[24] S. Tsitsirides, *Platons Menexenos: Einleitung, Text und Kommentar* (Stuttgart/Leipzig 1998) 34–41.

———————

restaged by the Athenians every year. Given that such an authoritative writer as Cicero attributes the *Menexenus* to Plato, it is not surprising that many have followed his opinion. That said, Cicero and his contemporaries mark a watershed for the widespread acceptance of the *Menexenus* as genuinely Platonic. From that period onward it was received as Platonic thought and influenced philosophical and rhetorical discussions. While certainly important for the reception of Platonic thought, this does not support the view that the *Menexenus* was written by Plato, only that it was perceived as having been written by Plato from as early as the first century BCE onward. In fact, there were probably several written *Menexenus* in circulation, one of which is attributed to a Glaucon, perhaps Plato's brother.[25] Late writers who attribute authorship are not necessarily reliable witnesses.

The only potentially valid witness for the *Menexenus* being written by Plato is Aristotle. Thus, the whole weight of the first category of arguments for Platonic authorship may be reduced to whether or not we believe that Aristotle is referring to Plato. Therefore, I will analyze the Aristotelian passages put forward as evidence by Tsitsirides after discussing the other two categories of arguments.

Tsitsirides' second category is the stylistic evidence, and it is

---

[25] Diogenes Laertius wrote in his entry on Glaucon (2.124): Γλαύκων Ἀθηναῖος· καὶ τούτου φέρονται ἐν ἑνὶ βιβλίῳ διάλογοι ἐννέα· … Μενέξενος. Thus, Glaucon, maybe Plato's brother, seems to have written nine works that were published in one book. In light of the *Menexenus*' stylometric similarity to Xenophon, analysis points to an author who lived at a time and in a space near Plato's. Therefore, his brother Glaucon would be a fitting alternative for authorship. This possibility had already been suggested by F. Ueberweg, *Untersuchungen über Echtheit und Zeitfolge platonischer Schriften und über die Hauptmomente aus Platons Leben* (Vienna 1861) 146. That Glaucon was interested in oratory and was training to become a politician (Xen. *Mem.* 3.6.1) makes the case even stronger. Glaucon's rhetorical proclivities make him a better candidate for authorship of a fictional eulogy than Plato, who was widely known as a critic of oratory. Unfortunately, Diogenes Laertius is late and does not quote from the *Menexenus* but only mentions the title. But we must ask: how likely is it that there was another work written by Plato's brother that was also called *Menexenus*? Or was Plato's *Menexenus* 'wrongly' attributed to Glaucon?

very much a weak point in his defense of Platonic authorship. While he suggests it is possible that a person could have written the dialogue shortly after Plato's death—maybe even a person belonging to Plato's school—and that the work entered the transmission lines of the *Corpus Platonicum* by accident, he dismisses evidence that supports this interpretation. Specifically, he focuses mainly on Ritter's nineteenth-century calculations,[26] although Ritter used only forty-three features, of which only ten occur in the *Menexenus* (most of them only once).[27] Tsitsirides also glosses over Ledger's computational analysis, because he seems to have misunderstood its result and misconstrued Ledger's remark that he "concede[s] a weakness in the stylometric evidence" to mean that he concedes a weakness in the stylometric *method*. Instead, Ledger meant to draw attention to the weakness of the stylometric evidence supporting the Platonic authorship of the *Menexenus*. He clearly stated that the "verdict is rather against" the *Menexenus'* being written by Plato, but he declared himself "inclined to think that that is more a reflection of its peculiar nature than a definitive declaration against Platonic authorship."[28] Thus, Ledger considers the *Menexenus* a Platonic work against the statistical evidence, which he dismisses with the suggestion that it might be skewed by a genre signal. Although Ledger concedes that the genre signal does not, as one might expect, seem to influence the *Letters*, including the seventh *Letter*, he was so convinced by the non-stylometric evidence provided by the Aristotelian references that he did not even include the *Menexenus* in his discriminant analysis authenticity study.[29]

Given that the *Menexenus* looks less Platonic in style the more data we include in our analysis, it becomes obvious that Tsitsirides' stylistic arguments involve cherry-picking, partially based

[26] Tsitsirides, *Platons Menexenos* 37–39.

[27] Ritter, *Untersuchungen* 99.

[28] Ledger, *Recounting Plato* 163. He also mentions (169) that although the stylistic evidence is against them, he thinks that both dialogues, *Menexenus* and *Clitophon*, seem "too outrageous to have been written by anyone other than Plato."

[29] Ledger, *Recounting Plato* 105, 163.

————————

on Ledger's doubting his own results, and partially simply ignoring evidence. Twentieth-century scholarship on the Platonic authorship of the *Menexenus* was largely carried out by researchers who really *wanted* it to be by Plato. An unbiased look at Tsitsirides' second category of arguments actually finds evidence against Platonic authorship.

The third category concerns the similarities in content of the *Menexenus* and other Platonic works. Tsitsirides admits that the lack of clear philosophical content in the *Menexenus* makes it difficult to compare it to Platonic thought in the other dialogues. Thus, he resorts to listing individual intertextual connections that in the aggregate clearly show a familiarity with Platonic works on the part of the author of the *Menexenus*. Apart from one passage, Tsitsirides does not find a direction for the intertextual connections and we do not know which work came first. The only pair of passages for which he sees a direction is *Menex.* 240A–B and *Leg.* 698C–E. Here Tsistirides follows Pohlenz's argument, which essentially states that the *Menexenus* must have been earlier because it is more detailed.[30] This is not persuasive, especially given that a similar story also features in Herodotus (Hdt. 6.31). At any rate, the third category of arguments certainly proves that the author of the *Menexenus* was familiar with Platonic works.

Since the stylistic evidence speaks against the *Menexenus* being written by Plato and the intertextual evidence at most proves familiarity with other Platonic works, and not dependence on them, and because later writers might be influenced by their reading of Aristotle and Cicero, the question whether the *Menexenus* was written by Plato can be reduced to whether Aristotle attributes the *Menexenus* to Plato. If he does, then the stylistic anomalies of the *Menexenus* might be a sign of Plato's wide stylistic range. If he does not, the stylistic evidence outweighs the testimony of late authors and of alleged intertextual connections to the *Corpus Platonicum*, and we must admit the possibility that the *Menexenus* was not written by Plato.

[30] Tsitsirides, *Platons Menexenos* 41; M. Pohlenz, *Aus Platos Werdezeit* (Berlin 1913) 278–281.

———

Aristotle and the Menexenus

In his *Rhetoric*, Aristotle does not name Plato as the author of the *Menexenus*, nor does he even name the *Menexenus* itself. He does, however, refer in two instances to one passage from the *Menexenus* (235D). The first instance is *Rh.* 1.9.30 (1367b):

> σκοπεῖν δὲ καὶ παρ᾽ οἷς ὁ ἔπαινος· ὥσπερ γὰρ ὁ Σωκράτης ἔλεγεν, οὐ χαλεπὸν Ἀθηναίους ἐν Ἀθηναίοις ἐπαινεῖν.

> one should also consider in whose presence the praise happens: for, as Socrates said, it is not difficult to praise Athenians among Athenians.

It has been argued and accepted that Aristotle's use of the imperfect ἔλεγεν suggests that he refers to the historic Socrates.[31] Given the almost proverbial nature of the passage, it is clear why this was accepted. *Menexenus* 235D, however, is much more detailed, and in fact, instead of stating what is not difficult, the Socrates in the *Menexenus* emphasizes how difficult it would be to praise someone in the presence of his enemies:

> εἰ μὲν γὰρ δέοι Ἀθηναίους ἐν Πελοποννησίοις εὖ λέγειν ἢ Πελοποννησίους ἐν Ἀθηναίοις, ἀγαθοῦ ἂν ῥήτορος δέοι τοῦ πείσοντος καὶ εὐδοκιμήσοντος· ὅταν δέ τις ἐν τούτοις ἀγωνίζηται οὕσπερ καὶ ἐπαινεῖ, οὐδὲν μέγα δοκεῖν εὖ λέγειν.

> For should it be necessary to speak well of Athenians among Peloponnesians or Peloponnesians among Athenians, then one would need a good orator to persuade and gain credit: but whenever someone competes among those he also praises, speaking well does not seem a big deal.

This first Aristotelian passage by itself proves neither the existence of the dialogue nor its authorship: it could simply be reporting a well-known anecdote about Socrates. The second instance, however, is not so easily dismissed (*Rh.* 3.14.11, 1415b):

> ὃ γὰρ λέγει Σωκράτης ἐν τῷ ἐπιταφίῳ, ἀληθές, ὅτι οὐ χαλεπὸν Ἀθηναίους ἐν Ἀθηναίοις ἐπαινεῖν, ἀλλ᾽ ἐν Λακεδαιμονίοις.

---

[31] Tsitsirides, *Platons Menexenos* 34; Ueberweg, *Untersuchungen* 143.

> For what Socrates says in his funeral oration is true, that it is not difficult to praise Athenians among Athenians, but among Lacedaemonians.

I will only mention in passing (because I think it has been sufficiently disproved) that critics have considered the third book of the *Rhetoric*, where Aristotle appears to refer to the *Menexenus*, inauthentic. After Diels's demonstration that the third book of the *Rhetoric* is Aristotelian, however, it was assumed that the passage was an interpolation.[32] I agree with Tsitsirides that this is unlikely: not least because the choice of the word ἐπιταφίῳ seems odd for someone who seeks to clarify.

Thus, for the sake of argument I will assume the more likely scenario that the reference is authentically Aristotelian: despite its brevity, reverse order, and Aristotle's replacing Peloponnesians with Lacedaemonians, this passage is close to the wording of the *Menexenus*. The present tense λέγει and the title ἐν τῷ ἐπιταφίῳ also seem to point rather to a written resource and not necessarily to the historic Socrates.[33] It is interesting, however, that Aristotle calls the work *Epitaphios* and not *Menexenus* and that he does not name Plato as its author. That may explain why Bonitz in his *Index Aristotelicus*, although he lists the passages under the works by Plato, does not classify them under Group A, which comprises references where either Plato or Socrates and a Platonic work are mentioned.[34] Instead, Bonitz refers to Ueberweg's theory that Glaucon is the author,[35] so that readers can judge for themselves whether this is a clear Platonic reference. Ueberweg points out that although the formula ὃ γὰρ λέγει Σωκράτης ἐν τῷ ἐπιταφίῳ fits with how Aristotle, at times, refers

---

[32] Tsitsirides, *Platons Menexenos* 35, and H. Diels, *Über das dritte Buch der aristotelischen Rhetorik* (Berlin 1886), esp. 20–23. Diels thinks that the *Menexenus* might not be Plato's masterpiece (in fact he thinks it is a "taube Nuß") but he considered it authentic, because he believed the third book of the *Rhetoric* to be authentic.

[33] Ueberweg, *Untersuchungen* 144.

[34] H. Bonitz, *Index Aristotelicus* (Berlin 1870) 598, s.v. Πλάτων.

[35] Ueberweg, *Untersuchungen* 146.

———

to Platonic works, it does not entail that the *epitaphios* in question be of Platonic authorship.[36] In order to judge whether Aristotle is referring to Plato, we must look at all the references to Plato or Socrates in Aristotle's *Rhetoric*.

There, Aristotle refers to Plato by name only three times. In the first instance (1.15, 1376a) it is not entirely clear whether he means the philosopher or the comic writer Plato. In either case, it is a historical person and not the work of an author. In the second instance (2.23.12, 1398b) Aristotle reports Aristippus' criticism of a confident declaration by Plato, saying that Socrates never spoke like that. Only in the third instance (3.4, 1406a) does Aristotle clearly refer to Platonic passages. A missing reference to Plato does not necessarily preclude Platonic authorship and one must look at the instances that mention Socrates and clearly refer to a Platonic work.

Apart from these passages Aristotle refers to a Socrates ten times. At 1.1, 1356b, the name Socrates appears twice, and then only as a stand-in for an individual or an individual case, while 1.2, 1357b, offers a statement about the historical Socrates to clarify a point. A third instance again refers to Socrates as an individual (2.4, 1382a). The fourth (2.15, 1390b) offers the historical Socrates or, rather, his descendants as an example. At 2.20, 1393b, Aristotle describes the effect of Socratic comparisons. If this points at all to a written account, then Xenophon's *Memorabilia* offers the closest parallel (1.2.9). Its context (Xenophon's reporting on one of Socrates' accusers) suggests that Aristotle has the historical Socrates in mind. Similarly, at 2.23.8, 1398a, Aristotle regards either Xenophon (*Ap.* 16–17) or the historical Socrates.[37] At 2.23.13, 1399a, he mentions Theodectes'

---

[36] Ueberweg, *Untersuchungen* 144.

[37] In their *Rhetoric of Aristotle* (Cambridge 1877) *ad. loc.*, E. M. Cope and J. E. Sandys argue that the reference is to Plato's *Apology*, because Socrates there draws an explicit inference (absent in Xenophon) from *to daimonion* to gods. I focus instead on the connection that according to Aristotle Socrates made between *hubris* and his inability to repay Archelaus. This recalls Socrates'

Conclusion

In deviating from the generally accepted hypothesis that the *Menexenus* was written by Plato, this paper might actually praise Lacedaemonians among Athenians. Yet, as Ledger stated and I now have shown more thoroughly here, there is very little stylistic evidence that the *Menexenus* was written by Plato. The stylometric measures by and large do not cluster the *Menexenus* with other works of the *Corpus Platonicum*. The *Menexenus* has a higher co-occurrence with Xenophon and Lucian than with Platonic works, and the frequency of καὶ in it is unusual not only when set against other works in the corpus but also when compared with the usage of Plato's Socrates. Furthermore, the genre signal hypothesis was disproved. Taken together, these facts strongly suggest that the *Menexenus* was not written by Plato.

As Tsitsirides has demonstrated, the philological evidence, on the other hand, shows that the author is very familiar with the *Corpus Platonicum*, and the stylistic evidence also exhibits similarities with other Athenian authors from the fifth and fourth centuries BCE. Aside from Aristotle's, ancient testimonies are too late to be useful, and if one looks at how he allegedly refers to the *Menexenus*, it is not clear that he has Plato's *Menexenus* in mind as opposed to anecdotes about the historical Socrates. It is tempting to attribute the *Menexenus* to Glaucon, as Ueberweg following Diogenes Laertius suggested. Ultimately, one must reckon with the fact that the *Menexenus* was included in the canon of Platonic tetralogies, so that the burden of proof lies on those who argue against Platonic authorship.

The answer to the question of the *Menexenus*' authorship depends on how much weight the reader puts on the stylistic evidence as opposed to the fact that the second Aristotelian passage may be referring to a Platonic work by just mentioning Socrates. That Aristotle frequently refers to various non-Platonic Socrates in the *Rhetoric* might tip the scales against Platonic authorship. I should also note that, if one insists that the *Menexenus* was written by Plato, there are no stylistic reasons to dismiss any of the other dialogues considered doubtful in the past and subjected to stylistic analysis: *Second Alcibiades, Epinomis, Amatores, Hipparchus,*

Minos, and the *Letters* are more Platonic in style than the *Menexenus*. Therefore, readers must seriously consider the possibility that the *Menexenus* was not written by Plato. The stylistic evidence now conclusively indicates that it was not.

APPENDIX: Technical Notes

The selection of the feature sets in the experiments

The feature sets in my experiments consisted either of words or 4-grams in the corpus. The twenty-five feature sets could themselves be divided into five different groups. The first group was solely based on word frequency: the 1000, 500, 300, 200, and 100 most frequent words were used as features. For the second group I made a similar selection of 4-grams (the 1000, 500, etc. most frequent ones). Both of those feature sets were extracted using the R library *stylo*.[39] In the third group the words were ranked according to their frequency in the corpus but were excluded if they did not occur in the *Menexenus*. This gave every feature that does not occur in the *Menexenus* the weight 0, which assumes that words not in the *Menexenus* are unimportant, or that their importance cannot be assessed and they should be excluded. It remains open whether that assumption is correct, but I nevertheless allowed the bias of the experiment slightly to favor detecting a weak Platonic signal in the *Menexenus*. This group and the fourth and fifth groups have been used to double-check the clear result that the *Menexenus* does not seem Platonic in style obtained solely with the first two groups. The fourth and fifth groups consist of two sets, one of words and one of 4-grams, each built from the sets generated by *stylo* in groups one and two.

To produce the fourth and fifth group, machine learning was used to prune the feature sets. Simply put, because the CTS workgroups often represent works by one author, we can use this bibliographic information to train algorithms using a program such as the Data Mining software WEKA.[40] This means that we divide the corpus into

---

[39] Eder et al., *Literary and Linguistic Computing* 26 (2011) 107–121.

[40] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "The WEKA Workbench," online appendix for M. Kaufmann (ed.), *Data Mining: Practical Machine Learning Tools and Techniques* (2016). Available at https://www.cs. waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf.

training and test set(s),[41] and then train the algorithm optimizing it for the task of classifying a work as belonging to a workgroup on the basis of the frequencies of the features. For instance, if we have Plutarch's *Life of Theseus*, we can train the algorithm so that when it looks at word frequencies in that text it determines that it belongs to the workgroup 'works by Plutarch'.

In WEKA we can train algorithms to classify the data or use out-of-the-box classifiers to see how well they perform on the desired classification task. For instance, the algorithm in WEKA called ZeroR always guesses the most frequent class in training set(s): no matter which work you feed into the algorithm, it always guesses Libanius, the author with the most works in our corpus. Of course, this is not a very smart algorithm, but it provides a baseline to test how effective our algorithms are: ZeroR achieved an accuracy of 12.13%. OneR is another simple WEKA algorithm. It attempts to find the most discriminating feature in the dataset (that is, the one word or 4-gram that results in the most accurate classification of the training set) and uses its value as the rule to determine the class. When I employed OneR on 4-grams the feature was α ὶ _ ἐ, which achieved 22.88% accuracy, and when I used it on words the feature was μὲν, with 20%. While these do not demonstrate great accuracy either and merely serve as a baseline for better algorithms, they show that, by simply measuring the mean frequency of the word μὲν and building a rule from this, the odds are one in five that we will correctly identify the author of an Ancient Greek work.

Two slightly more complex algorithms performed considerably better on the data: J48 and Naïve Bayes. J48 is a classifier that builds a decision tree based on the training data and then uses it to classify the test data.[42] Using the stylometric feature sets, J48 applied to 4-gram frequencies classifier correctly classified 50.75% of the instances, while with word frequencies its success rate was 50.5%. Naïve Bayes is a probabilistic classifier that, for each feature attribute, computes as many probabilities as there are classes and then chains them to make a classification.[43] Using the stylometric feature sets, Naïve

---

[41] Since I used 10-fold cross-validation when training the algorithms, plural "sets" is more correct, but the principle relies on one test set and one training set.

[42] Witten et al., in *Data Mining* 92.

[43] Witten et al., in *Data Mining* 34.

Bayes applied to 4-gram frequencies correctly classified 65.88% of the instances, whereas with word frequencies its success rate was 64.75%.

While WEKA can determine how well each trained algorithm performs and thus can generate a baseline that our stylometric-specific measures must beat, we can also employ our analysis of the classifiers for feature selection. In WEKA this is called 'Attribute Selection'. For instance, if we look at the results of J48 and Naïve Bayes for the *Corpus Platonicum*, we see that, while the overall accuracy of Naïve Bayes with word frequencies was higher than that of J48, its numbers for Plato were lower (the precision of Naïve Bayes was 0.4 while J48 attained 0.8). With WEKA's *Ranker* method, we can rank the attributes according to their usefulness. In the two sets built with WEKA, I ranked the feature sets that were extracted by *stylo* and removed the attributes that *Ranker* evaluated negatively. For word frequency, *Ranker* evaluated the attributes using J48; for 4-gram frequencies it used Naïve Bayes because of their greater precision when classifying Platonic works in each feature set.

I should stress that initially I had used only the most frequent words and most frequent 4-grams; however, I later decided to include more feature sets because the *Menexenus* did not seem to cluster with Plato when using the original sets. My aim was to calibrate the features so as to detect a potentially weak *Menexenus* signal of Platonic authorship. The feature sets three, four, and five were my attempt to defend this attribution. Not surprisingly, the WekaWord feature and measure combination best detected similarities between the *Menexenus* and the *Corpus Platonicum*.

Measuring the frequency of each feature resulted in a large table, where the rows are the unique identifiers of the works and the columns are the names of the measured features. Since the experiments observe up to 1000 features at once (either 1000 words or 1000 n-grams) and cover 800 ancient works, the largest matrix has 800,000 cells. Clearly this would be too much data for manual qualitative analysis. Nor can graphs be used to express the data, because they are essentially discrete sets: although some connections between frequencies of individual words exist (e.g., between μὲν and δὲ), the frequency of ἀνθρώπους, for instance, has no direct influence on the frequency of καὶ. Each data point (or work) is best thought of as a set of 1000 values. As the order of those values is always the same for each data point, we can think of each work as a point in a high-dimensional coordinate system, where each dimension is the frequency of a word or n-gram. Thus, when using the frequencies of the most-frequent 1000 words as features in

the stylometric analysis, I created a 1000-dimensional space in which each work is a single point in that space, represented by its set of 1000 values. The challenge is that high-dimensional spaces are difficult for the human brain to visualize, imagine, and interpret.

In order to make sense of the data, we must address this challenge. As noted in the main text, I pursued two independent approaches: one is a t-SNE dimensionality reduction approach that by mapping the 1000-dimensional space to a 2-dimensional space allows us to use simple Euclidean distances to cluster the data; the other measures several types of distances between points directly in the high-dimensional space and then uses those distances to cluster the data. The same clustering method of grouping the works according to their data similarity was applied. The clustering used Ward's method (the ward.D2-implementation of the programming language R),[44] which is an agglomerative clustering method; that is, it produces a tree by combining close clusters with each other until there is nothing to join. Because we know what step created each cluster we can easily cut any number of clusters out of that tree (the minimum number of clusters we can cut is one and the maximum is the number of data points, here 800). Because our data was grouped by humans into 153 workgroups representing something like an author, I cut the tree for automatic evaluation into 153 clusters. Additionally, I also looked at the trees as wholes.[45]

The abbreviation t-SNE stands for t-Distributed Stochastic Neighbor Embedding and is a popular algorithm for dimensionality reduction.[46] It attempts to visualize data points that have close proximity in the high-dimensional space by placing them in close proximity in a low-dimensional space. For example, to visualize a high-dimensional data set in two dimensions we can observe the t-SNE characteristic groups of data points: the data that is similar to each other appears to build an 'island' on a map. To visualize that same data set in three dimensions we would observe 3-dimensional

[44] F. Murtagh and P. Legendre, "Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm," arXiv:1111.6285v2 [stat.ML] (11 Dec 2011), https://arxiv.org/pdf/1111.6285v2.pdf.

[45] These can be accessed at T. Koentges, Zenodo repository.

[46] L. v. d. Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research* 9 (2008) 2579–2605, available at: http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.

bodies of data points vaguely resembling 'blobs'. Those islands or blobs are visual clusters of similar data. The fact that a data point is outside the cluster indicates its degree of difference in the higher-dimensional space from points in the cluster.

Besides t-SNE combined with two-dimensional Euclidean distances, I also calculated distances directly with the high-dimensional data. Because we do not know what kind of space the feature set creates, I used eight different distances to cluster the data: Euclidean distance (segment lengths joining points in ordinary physical space); Jaccard Distance, which measures the similarities between sets; Cosine Distance and Cosine Delta, two distances based on the cosine of the two vectors expressed through the values of each data point; and four other distances developed for authorship attribution: Burrows' Delta, Eder's Simple, Eder's Delta, and Argamon's Delta.

Although Eder's Delta, Burrows' Delta, and especially Cosine Delta have been considered very reliable for authorship attribution, re-searchers disagree on whether these measures are the best for all languages, because language-dependent factors may influence the result.[47] For instance, the best measure for German might not be the best measure for Greek when the most frequent words are used. For this reason I used all eight measures. To evaluate them, each feature-set and measure combination was tested and scored.

Testing the measures

In order to test the 225 feature-set and measure combinations, I once again used the fact that the CTS URN of each work indicates its corresponding author/workgroup. Digital archivists have ordered the 800 works into 153 workgroups. This was not done in a single meta-

---

[47] See J. Rybicki and M. Eder, "Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26 (2011) 315–321. For a ranking of the algorithms based on English, French, and German novels see F. Jannidis, S. Pielström, C. Schöch, and T. Vitt, "Improving Burrows' Delta – An Empirical Evaluation of Text Distance Measures," paper given at the DH2015 conference in Sydney; abstract accessible at https://www.researchgate.net/profile/ Steffen_Pielstroem/publication/280086768_Improving_Burrows'_Delta_- _An_empirical_evaluation_of_text_distance_measures/links/573ad8ae08ae 9f741b2d3d40.pdf. For a better understanding of Delta authorship attribu-tion measures see S. Evert and T. Proisl, "Understanding and Explaining Delta Measures for Authorship Attribution," *Digital Scholarship in the Human-ities* 32 (2017) ii4–ii16, https://doi.org/10.1093/llc/fqx023.

———————

data sprint; it happened over decades under traditional assumptions in order to facilitate the use of the collection. As a result, a workgroup sometimes contains more than one author (e.g., tlg0013 stands for the *Homeric Hymns*), and given the complexity of textual transmission there will sometimes be authorship attribution errors. But given the final outcome of 153 workgroups, we can compare these with what a machine will do if tasked with clustering the corpus into 153 groups using a selection of different measures. This comparison should help us to estimate how well the individual measures performed.

If works are clustered under a single author, we can consider them correctly classified, whereas if they are assigned to a mixed group, we consider them ascribed to the author that holds the majority in that group; if there is no majority author or if they are by a minority author in the group, we consider them misclassified. One may then measure the percentage of works that passed the test as the proportion of works correctly classified and thereby obtain an estimate of how well an individual feature-set and measure combination performs in an authorship attribution task applied to a real-world corpus. Since the score is based on membership in a cluster (k), and since it does not really measure the precision in a controlled environment where we are completely certain about the authorship of a work but only how well the algorithm performs compared to human classification, I call this the k-score. Although it is theoretically possible to reach a k-score of one, this is very unlikely for two reasons: first, because it is improbable that all works are correctly attributed when preparing the URNs; and second, because each feature-set and measure combination may be distracted by different signal noise. *Fig.* 7 shows the distribution of the k-scores of the individual measures. They look very promising, especially if we compare them to the classification scores in WEKA mentioned above.

One stylometric research result of this test is that, even though all delta measures other than Argamon's Delta performed better than all non-authorship attribution specific measures, Eder's Simple—not Eder's Delta—performed most precisely and reliably of all. This outcome is counter-intuitive, given that for English, German, and French Cosine Delta outperforms Eder's Simple, and it can only be explained by the fact that the Latinist and stylometry expert Eder built his authorship attribution measures for morphologically more complex

languages.[48] That said, it is worth noting that the Simple measure out-performs the Delta measure in Ancient Greek, which features a more complex morphology than Latin, and that in general all measures performed well.[49]
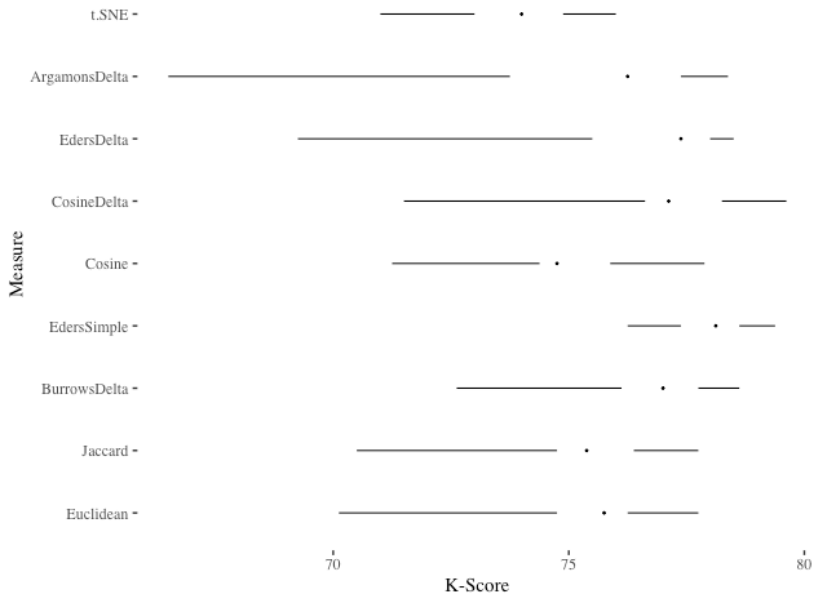

Figure 7: Tufte-style box-and-whiskers plot of the k-scores of the measures used in the experiment. The dot shows the median value of all k-scores for the 25 different feature sets, and the whiskers show the range of the value. Eder's Simple is the most consistent and has the highest median, while Cosine Delta measure has the highest individual k-score.

———————

[48] Cf. Jannidis et al. (preceding note), and S. Evert, T. Proisl, F. Jannidis, S. Pielström, C. Schöch, and T. Vitt, "Towards a Better Understanding of Burrows's Delta in Literary Authorship Attribution," *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature* (Denver 2015) 79–88.

[49] To confirm that Eder's Simple is the best measure currently available for Ancient Greek, this observation should be followed by a stylometric experiment based solely on works whose authorship is completely certain (that is, the analysis of a more controlled sample).

———————

Given that the combinations of feature-set and measure perform mostly between 0.75 and 0.8, we can consider any of the measures reliable, although Eder's Simple is the most reliable. However, given that each combination might be affected by different kinds of noise in the data-set (e.g. a strong genre signal) and given that all measures score comparably well, I compared the results of all 225 combinations in my analysis of the *Menexenus* and the *Corpus Platonicum*.[50]

February, 2019   Center for Hellenic Studies,
                                    Harvard University
                                Digital Humanities,
                                    Universität Leipzig
                                thomas.koentges@gmail.com