Fine-Tuning Pre-Trained Language Models for Authorship Attribution of the Pseudo-Dionysian Ars Rhetorica

Gleb Schmidt[1,*,†], Veronica Vybornaya[2,†] and Ivan P. Yamshchikov[3,†]

[1]Radboud University Nijmegen, Erasmusplein 1, 6525 HT, Nijmegen, The Netherlands
[2]Independent scholar, St. Petersburg, Russia
[3]CAIRO, THWS, Technische Hochschule Würzburg-Schweinfurt, Franz-Horn Str. 2, 97082 Würzburg, Germany

Abstract

This paper explores the use of pre-trained language models for Ancient Greek in the context of authorship attribution. The study adopts a two-step approach: first, the models are fine-tuned on a domain-specific corpus using a masked language modeling (MLM) objective; second, based on the fine-tuned model, a classifier is trained to address the authorship attribution task. The analysis centers on a corpus of texts on rhetorical theory from the Second Sophistic period, with particular focus on the Pseudo-Dionysian *Ars Rhetorica*. The results of the experiment suggest that this approach offers valuable insights into the authorship of ancient texts. Notably, the findings align with some traditional scholarly views on the *Ars Rhetorica* while also opening the door to reconsidering long-discarded hypotheses about the treatise's internal structure. This study highlights how the integration of natural language processing and classical philology can significantly advance discussions in ancient literary scholarship.

Keywords

pre-trained language models, authorship attribution, authorship analysis, historical languages, transfer learning, ancient greek (roman period), Ps.-Dionysius's Ars Rhetorica, BERT, RoBERTa

1. Introduction

Over the past several years, the application of transformer-based neural networks [1] has led to significant advancements in many NLP tasks related to historical languages [2, 3, 4]. However, unlike in the case of modern languages, where fine-tuning pre-trained transformers for linguistic forensics is very common [5, 6, 7, 8], the application of such models for authorship attribution tasks in historical languages remains relatively underexplored, although some excellent seminal studies and surveys have been recently published [9, 10, 11, 12]. The availability of state-of-the-art pre-trained language models [13, 3, 14, 15] excelling in multiple downstream tasks suggests that the situation with authorship analysis can be different as well.

Yamshchikov et al. [15] obtained a pre-trained model for Ancient Greek by fine-tuning a Modern Greek BERT model [16]. The resulting model subsequently served as the backbone for a classifier and proved effective for authorship attribution of the so-called Pseudo-Plutarch corpus. Interestingly, despite being fine-tuned on a limited amount of Ancient Greek data, the model obtained through transfer learning showed results comparable to those of the models trained from scratch on significantly larger corpora, as reported by Singh et al. [14] and Riemenschneider and Frank [3]. Drawing inspiration from Yamshchikov et al. [15], this study experiments with a similar approach focusing on the works of late Greek rhetoricians.

Greek prose on rhetorical theory from the period known as the Second Sophistic serves as a crucial

CHR 2024: Computational Humanities Research Conference, December 4–6, 2024, Aarhus, Denmark
*Corresponding author.
†These authors contributed equally.
gleb.schmidt@ru.nl (G. Schmidt); ivan.yamshchikov@thws.de (I. P. Yamshchikov)
https://github.com/glsch/ (G. Schmidt); https://www.yamshchikov.info (I. P. Yamshchikov)
0000-0001-6925-551X (G. Schmidt); 0000-0003-3784-0671 (I. P. Yamshchikov)


source, documenting the cultural and intellectual framework of Greek thought and literature in the first centuries AD [17, 18, 19, 20]. However, the study of this extensive corpus of texts, collectively referred to under the broad concept of *Rhetores Graeci* [21, 22], is significantly complicated by endless controversies surrounding authorship, dating, and contextual factors [23].

In this paper, we explore the potential of a transformer-based models, fine-tuned for sequence classification task, to provide further insights into the debate.

The focal point of our study is the text conventionally referred to as the *Ars Rhetorica* (Art of Rhetoric, hereafter *ars*). This work has long been attributed to, and frequently published under the name of, the rhetorician and historian Dionysius of Halicarnassus (*ca.* 60–7 BC). However, Sadée [24], followed by Usener [25], Usener and Radermacher [26], demonstrated that the *ars* most likely circulated anonymously, with its association to Dionysius emerging from a much later conjecture. This conjecture appears to have been based on an overinterpretation of a scholion (a marginal commentary) on chapter 10 of the text.

2. Ars Rhetorica

Several aspects of the *ars* must be discussed in the context of statistical modelling of its writing style.

2.1. Not one, but multiple works

The text has a complex structure. In *Parisinus Graecus* 1741 [27], the only manuscript that preserves all the material associated with the *ars* (ff. 1–37), the text is divided into 11 chapters. However, these chapters do not constitute a homogeneous work, as the text is generally understood to consist of two [28], three [25, 26, 29, 30], or even four [31] distinct parts.

The first part, covering ch. 1–7, provides concise instruction on ceremonial (epideictic) oratory, addressing seven epideictic genres. These chapters are connected by cross-references and recurring addresses to the author's former pupil, Echecrates, to whom the text is presented as a wedding gift. The remainder of the text, ch. 8–11, may be interpreted as a combination of two or three distinct works on separate topics. Ch. 8–9 explore the so-called "figured speeches", i.e., speeches intended to convey a hidden meaning that may conflict with the literal content and stated purpose of the speech, while ch. 10–11 focus on the criticism of declamation.

2.2. Authorship

Ch. 1–7 exhibit a consistent compositional pattern and a recognizable writing style, suggesting they were authored by the same rhetorician. However, whether these chapters form a coherent and complete treatise is a matter of debate. This portion of the *ars* has been interpreted as a collection of distinct letters or essays [31, 28], as remnants or excerpts from a much longer work [32, 25, 26], or as a unified treatise [33, 34, 35]. For ch. 8–11, the situation is even more ambiguous. Usener [25] speculated that ch. 8–9 were written by two different disciples attending separate lectures of the same teacher. Pendorf [36] and Schöpsdau [37] rejected the idea that ch. 8–9 had a single author, suggesting instead that these texts drew from various sources. Similarly, ch. 10–11 have been attributed either to the same author as ch. 8–9 (with Heath [28] tentatively identifying him as Sarapion Aelius, a 2nd-century Alexandrian rhetorician whose entire corpus is lost) or to two different authors unrelated to the rest of the *ars*. Table 1 summarizes the content and authorship hypotheses for the various sections of the *ars*.

2.3. *Ars Rhetorica*, Menanderian Corpus, and Pseudo-Hermogenes' *On Method*

Since the early days of scholarship on the *Ars Rhetorica*, it has been noted that the rhetorical instruction provided in ch. 1–7 and ch. 8–11 shows a clear methodological affinity with, respectively, the treatises

Table 1
Themes addressed in the *Ars Rhetorica* and alleged authorship of its different parts. Each Roman number stands for one author. II–III means that the section might have been written by two different persons.

ars	Theme	Author
1	Panegyrics	
2	Marriage speeches	
3	Birthday speeches	
4	Epithalamium	I
5	Addresses	
6	Funeral speeches	
7	Exhortations to athletes	
8 9	"Figured speeches"	I or II or II–III
10 11	Criticism of declamations	I or II or IV or IV–V

ascribed to Menander Rhetor (particularly the second one) and Pseudo-Hermogenes' *On Method*. The parallels with the second treatise attributed to Menander are especially noteworthy. In both works:

- the occasion — rather than the subject, as in the first treatise attributed to Menander — determines the genre;
- a very similar selection of genres is discussed (of the seven genres mentioned in the *ars*, only two are absent from Menander's purported work; see Table 7);
- the author addresses a former disciple throughout the text.

This affinity led Heath [28] to describe the *ars* as "comparable to, though less sophisticated than" Menander's work.

The numerous parallels between ch. 8–11 of the *ars* and Pseudo-Hermogenes' *On Method* [38, 28] have led scholars to hypothesize either a shared source [38] or a closer, albeit indirect, connection [29, 28].

2.4. Dates

The dates of the texts constituting the *ars* have been assessed differently. For ch. 1–7, a mention of the 2nd-century sophist Nicostratus (ch. 2, par. 9, p. 266, l. 14), along with the considerable focus on speeches addressing Roman magistrates, suggests a composition date no earlier than the High Empire [29, 35]. Race [39] posits that the first part of the *ars* is roughly contemporary with the corpus attributed to Menander Rhetor, which is datable to the late 3rd century AD. In contrast, ch. 8–11 may be a century earlier [28], i.e., 2nd century AD.

3. Aims

The hypotheses concerning the authorship of different parts of the *ars* have multiplied, as have suggestions regarding its potential relationship with other texts. However, the evidence presented in the scholarship so far is drawn almost exclusively from close reading and remains inconclusive. Additionally, unlike the case with ch. 8–11, no efforts have been made to identify the author responsible for ch. 1–7.

The aim of our investigation, therefore, is to apply modern natural language processing techniques to this rich textual material in order to gather new evidence about the structure of the *ars* and gain further insights into its authorship. The arguments formulated through language modeling could provide a novel and valuable contribution to the debate, particularly when considered alongside the accumulated philological evidence and existing codicological indications.

The main contributions of this work can be summarized as follows:

- We further fine-tune two pre-trained models for Ancient Greek and one model for Modern Greek on a corpus of Greek rhetoricians. We subsequently use the resulting models to train "open set" 10-class classifiers capable of attributing short fragments of text to different authors of the Second Sophistic period;
- Analyzing in more details the results provided by two best-performing models, we shed light on the history of the Pseudo-Dionysian *ars*, suggesting that:
  - Ch. 1–7 of the *ars* could have been authored by an individual from the same school as the author(s) of the Menandrian treatises;
  - Ch. 8–11 not only differ in authorship from ch. 1–7, but may have been written by two distinct individuals, one responsible for ch. 8–9 and another for ch. 10–11.

4. Corpus

The primary focus of our study is a corpus comprising at least 18 *rhetores Graeci* from the 1st–4th centuries AD and Dionysius of Halicarnassus. We only retained the authors whose teachings are relatively well-preserved, excluding those known only through fragmentary or indirect evidence. Importantly, we focus exclusively on rhetorical theory, i.e., works with a theoretical or pedagogical intent.

A significant limitation of this dataset is that many of the rhetorical corpora within it have notorious attribution problems of their own. In particular, there is a compelling case for the heterogeneity of the Hermogenean corpus (see Section 6). Similarly, the question of whether both treatises attributed to Menander were authored by the same person remains unresolved [40, 41, 39, 42]. Other corpora raise similar questions, too [23]. Being aware of this and currently working on a follow-up authorship verification study of these corpora (the importance of which was also insightfully emphasized by the reviewers of this work), for simplicity, we continue to group the studied texts by authorship as categorized in the *Thesaurus Linguae Graecae* (TLG), where our dataset stems from.[1] We deem this simplification legitimate. In most cases, these questionable attributions are rooted in long-standing traditions that date back to the early stages of textual transmission. For example, the Hermogenean corpus has been consistently attributed to Hermogenes of Tarsus since as early as the 5th century (for more details see Section 6). Therefore, with all necessary reservations, these conventional groupings can be considered to represent at least some kind of connection. Even if they do not link texts written by the same individual, they may still group works originating from the same school. After all, this is why such simplification is commonly used in scholarship.

4.1. <UNK> category

The literature on oratory theory was undoubtedly much richer than what has been preserved. To account for this, we created an "open set" scenario. For this purpose, we set aside 9 smaller authorial corpora — those with a number of sentences below the dataset's median value of 517 (marked with * in Table 2). These texts were excluded from the dataset before our conventional 80/10/10 split and later added to the the test set. The idea is straightforward. If at the test stage the model encounters a text that does not belong to any of the authorial classes learned during the training, it is likely that the calibrated probability associated with the top prediction will be relatively low. If it falls below a certain threshold, the model is programmed to abstain from making a decision and assign an <UNK> label to the text in question. The samples with <UNK> label in the test split are necessary to monitor the model's capability to do.

An overview of the classification dataset is presented in the Table 2.

[1] We cannot publish the full texts with all the corresponding metadata. However, the shuffled chunks used in MLM fine-tuning and subsequent classifier training are made available on GitHub: https://github.com/glsch/rhetores_graeci.

Table 2
Classification dataset. Texts by authors marked with * were grouped under the <UNK> label. This label is present only in the test data to evaluate the model's ability to deal with uncertainty in an "open set" scenario.

Name	TLG	Date	Location
Aelius Aristides	284	II AD	Mysia
Aelius Herodianus & Pseudo-Herodianus	87	II AD	Alexandria
Aelius Theon*	607	I–II AD	Alexandria
Alciphron	640	II–III AD	Unknown
Alexander*	594	1st half II AD	Unknown
Anonymus Seguerianus*	2002	1st half III AD	Unknown
Cassius Longinus*	2178	mid III AD	Athens
Demetrius	613	I AD	Unknown
Dionysius Halicarnasseus	81	I BC	Halicarnassus
Eudemus	1376	II AD	Argos
Hermogenes	592	II–III AD	Tarsus
Lesbonax*	649	II AD	Miletus
Longinus*	560	I AD	Unknown
Marcus Cornelius Fronto*	186	II AD	Numidia
Menander	2586	III–IV AD	Laodicea
Minucianus Junior*	2903	III AD	Athens
Polyaenus	616	II AD	Macedonia
Polybius*	605	II AD	Sardis
Valerius Apsines	2027	III AD	Athens

5. Methodology

5.1. Base Transformers

To train our classifiers, we used three different pre-trained transformers as starting points:

(1) RoBERTa-sized GreBerta presented by Riemenschneider and Frank [3],
(2) Ancient Greek BERT trained by Singh et al. [14]
(3) Modern Greek BERT published by Koutsikakis et al. [16].

5.2. Masked Language Modeling Fine-Tuning

Fine-tuning pre-trained models on domain-specific corpora prior to further tuning them for a downstream task at hand is a common practice in NLP. It allows the model to adapt better to the unique linguistic features of the target domain. This intermediate step may enhance the model's ability to capture specific syntactical patterns and vocabulary, which in turn improves the performance on the final downstream task, such as classification. For this reason, before training classifiers for authorship attribution, we ran training with a masked language modeling objective. BERT-sized models were trained for 3 epochs with a learning rate 1×10^{-5} and warmup during the first 10% of training steps. RoBERTa-based model was trained for 1 epoch only with a learning rate 1×10^{-4} and without warmup steps. In both scenarios, the learning rate was decreasing linearly.

5.3. Sequence Classification

Authorship classifiers were trained on both out-of-the-box models and their MLM-fine-tuned versions. We employed a sliding window technique to segment the texts into chunks. The process was as follows:

1. Tokenization: We used the `bowphs/GreBerta` tokenizer to convert the entire corpus into tokens.[2]

[2]We did not repeat the experiment producing chunks with other available tokenizers.

2. Chunking: The tokenized corpus was then divided into chunks of 64 tokens, respecting the boundaries of works (and even chapters — in the case of the *ars*);
3. Overlap: To ensure continuity and capture context that might span chunk boundaries, we implemented an overlap between chunks. Each chunk overlapped with its adjacent chunks by 32 tokens (half of the chunk length).
4. Decoding: Finally, we decoded these token chunks back into text, resulting in our training data segments. By using a single tokenizer to chunk the entire corpus beforehand instead of splitting the texts with a tokenizer of the corresponding model, we ensured that all models were trained on the same segments of text.

The training was carried out for 700 steps by sampling batches containing 4 chunks per authorial class. Validation set was checked each 350 steps, i.e., twice during the training. Test set including <UNK>-labelled samples was checked upon the end of training. We report the results obtained on the test set.

6. Results and Discussion

6.1. General Performance

Table 3 summarizes the overall performance of the classifiers. Notably, additional MLM training proved beneficial only for the RoBERTa-sized bowphs/GreBerta model. For BERT-sized models, however, the inclusion of new data was detrimental. bowphs/GreBerta appears to be more stable, behaving more like general-purpose language models trained for well-resourced modern languages. This stands to reason: out of the three models with which we experimented, bowphs/GreBerta [3] is the largest and was trained on the riches and highest-quality Ancient Greek corpus.

Table 3
Performance metrics on the test split with the <UNK> category (not represented in the training data). The models were configured to assign <UNK> to samples with a calibrated top probability below 80%. (R) denotes models fine-tuned with an MLM objective on the same data that was used to train the classifiers.

Model	F1 Score	Accuracy
pranaydeeps/Ancient-Greek-BERT (R)	82.90%	80.96%
pranaydeeps/Ancient-Greek-BERT	83.68%	81.83%
nlpaueb/bert-base-greek-uncased-v1 (R)	78.34%	74.98%
nlpaueb/bert-base-greek-uncased-v1	79.22%	76.02%
bowphs/GreBerta (R)	**90.14%**	**90.12%**
bowphs/GreBerta	89.34%	89.27%

6.2. Authorship Attribution of the *ars*

The aim of this study was to get some fresh evidence about the authorship of the pseudo-Dionysian *ars*, a precious witness to the development of rhetorical theory during the High to Late Roman empire. Based on the *status quaestionis* surveyed in the section 2, we set up 3 research questions:

1. Can we further comfort or challenge the existing consensus opinion, according to which the attribution to Dionysius of Halicarnassus is incorrect?
2. How many works are discernible in the *ars* in the form we know it?
3. Can the model convincingly suggest an alternative attribution for the *ars* or any of its parts?

To address these questions, we applied the trained classifier to individual chapters of the ars, split into chunks following the described procedure. Table 4 summarizes the predictions made by the best-performing BERT-sized and RoBERTa-sized models.[3] For each chapter, we report the "majority vote"

[3] pranaydeeps/Ancient-Greek-BERT and bowphs/GreBerta (R)

Table 4
"Majority vote", share, and mean prediction probability for each chapter of the *ars*: Ancient Greek BERT vs GreBerta (R). "Rest" stands for the sum of all minor attributions. Sorted by the mean prediction probability.

	pranaydeeps/Ancient-Greek-BERT				bowphs/GreBerta (R)			
Ch.	**Author**	**Vote**	**Share**	**Prob.**	**Author**	**Count**	**Share**	**Prob.**
1	Menander	32	0.70	66.59	Menander	39	0.85	79.21
	Aelius Aristides	5	0.11	12.51	Aelius Aristides	3	0.07	7.39
	Dionysius H.	3	0.07	6.80	Dionysius H.	3	0.07	6.57
	Rest	6	0.13	14.10	Rest	1	0.02	6.83
2	Menander	30	0.59	57.46	Menander	36	0.71	68.80
	Dionysius H.	9	0.18	15.20	Aelius Aristides	5	0.10	9.32
	Aelius Aristides	9	0.18	14.48	Dionysius H.	6	0.12	9.23
	Rest	3	0.06	12.87	Rest	4	0.08	12.66
3	Menander	25	0.96	94.13	Menander	25	0.96	89.75
	Dionysius H.	1	0.04	2.05	Hermogenes	1	0.04	4.68
	Hermogenes	0	0.00	1.43	Dionysius H.	0	0.00	2.10
	Rest	0	0.00	2.40	Rest	0	0.00	3.47
4	Menander	15	0.79	69.45	Menander	16	0.84	79.77
	Hermogenes	3	0.16	11.68	Aelius Aristides	2	0.11	8.99
	Aelius Aristides	1	0.05	6.42	Demetrius	1	0.05	5.43
	Rest	0	0.00	12.45	Rest	0	0.00	5.81
5	Menander	21	0.49	48.00	Menander	28	0.65	60.73
	Aelius Aristides	12	0.28	25.16	Aelius Aristides	11	0.26	21.61
	Dionysius H.	3	0.07	7.34	Hermogenes	2	0.05	6.29
	Rest	7	0.16	19.50	Rest	2	0.05	11.37
6	Menander	34	0.61	52.98	Menander	34	0.61	56.71
	Hermogenes	9	0.16	15.58	Valerius Apsines	6	0.11	12.06
	Dionysius H.	5	0.09	8.78	Dionysius H.	7	0.12	11.11
	Rest	8	0.14	22.66	Rest	9	0.16	20.12
7	Menander	31	0.39	37.87	Menander	42	0.53	48.34
	Aelius Aristides	15	0.19	18.39	Aelius Aristides	12	0.15	14.65
	Dionysius H.	12	0.15	15.14	Valerius Apsines	8	0.10	10.85
	Rest	21	0.27	28.60	Rest	17	0.22	26.15
8	Hermogenes	49	0.21	21.63	Hermogenes	59	0.26	25.45
	Valerius Apsines	41	0.18	17.19	Aelius Aristides	58	0.25	21.86
	Aelius Aristides	45	0.19	16.92	Dionysius H.	49	0.21	19.96
	Rest	96	0.42	44.26	Rest	65	0.28	32.73
9	Hermogenes	60	0.20	17.95	Aelius Aristides	78	0.26	23.26
	Demetrius	45	0.15	15.37	Hermogenes	51	0.17	17.99
	Aelius Aristides	49	0.16	14.55	Dionysius H.	45	0.15	15.03
	Rest	145	0.48	52.14	Rest	125	0.42	43.72
10	Hermogenes	43	0.34	30.00	Hermogenes	52	0.42	38.21
	Dionysius H.	31	0.25	21.65	Dionysius H.	25	0.20	20.63
	Valerius Apsines	17	0.14	14.49	Valerius Apsines	16	0.12	11.48
	Rest	34	0.27	33.86	Rest	32	0.26	29.68
11	Hermogenes	41	0.37	31.65	Hermogenes	34	0.30	28.39
	Menander	23	0.21	20.22	Menander	23	0.21	19.30
	Dionysius H.	14	0.12	13.48	Dionysius H.	21	0.19	18.81
	Rest	34	0.30	34.64	Rest	34	0.30	33.51

(i.e., the number of chunks in the chapter attributed to a given author), the author's "share" (i.e., the proportion of chunks assigned to that author in the total number of chapter chunks), and the mean probability of the author across the chunks of the chapter. In the "majority vote", the attribution is defined by the top probability even if it falls below 80%.

6.3. No trace of Dionysius of Halicarnassus

In line with previous scholarship, although the name of Dionysius of Halicarnassus appears among the attributions, its weight is insignificant. Therefore, with regard to the first of the research questions, the evidence is overwhelming: stylistic affinity with Dionysius of Halicarnassus's writings is scarce, and the attribution to him cannot be supported by any of the two models.

6.4. *ars*'s association to the Menandrean corpus further strengthened

Apart from this rather predictable conclusion, our classifiers yield new insights into more complicated questions concerning the inner structure of the *ars* and the authorship of the texts, which constitute it. As clear from the Table 4, the attribution profiles for ch. 1–7 and 8–11 are drastically different. Even

Table 5
"Majority vote", share, and mean prediction probability for logical subdivisions within the *ars*, ch. 1–7: `GreBerta (R)`.

	Vote			Share (%)			Probability		
	1 & 7	2–4	5 & 6	1 & 7	2–4	5 & 6	1 & 7	2–4	5 & 6
Menander	81	77	62	0.68	0.83	0.65	59.70	76.64	58.45
Aelius Aristides	15	7	13	0.12	0.08	0.14	11.97	7.09	11.33
Hermogenes	7	2	7	0.06	0.02	0.07	7.87	3.77	9.01
Dionysius Halicarnassensis	9	6	7	0.08	0.06	0.07	7.6	5.7	7.06
Valerius Apsines	8	1	7	0.07	0.01	0.07	6.40	1.8	8.85

Table 6
"Majority vote", share, and mean prediction probability for logical subdivisions within the *ars*, ch. 1–7: Ancient Greek BERT.

	Vote			Share (%)			Probability		
	1 & 7	2–4	5 & 6	1 & 7	2–4	5 & 6	1 & 7	2–4	5 & 6
Menander	63	70	55	0.52	0.74	0.59	48.44	69.76	50.82
Aelius Aristides	20	10	15	0.17	0.11	0.16	16.23	9.22	15.89
Dionysius Halicarnassensis	15	10	8	0.12	0.11	0.09	12.07	9.47	7.87
Hermogenes	12	3	11	0.10	0.03	0.12	9.46	4.18	11.36
Valerius Apsines	10	2	4	0.08	0.02	0.04	7.81	3.13	6.49

when the probability is not high enough, Menander Rhetor is the top-ranked candidate in ch. 1–7. The signal is less clear in ch. 8–11. This difference goes in line with the *communis opinio* that the work is composite: a nearly identical attribution profile of ch. 1–7 being yet another argument in favour of its unity.

6.5. What does the model learn?

For the sake of explainability, DH specialists still widely use the bag-of-words model and corpus-specific manual feature engineering for various tasks involving writing style analysis, such as authorship attribution, authorship and self-authorship verification, clustering, etc. [43, 44, 45, 46]. Since deep learning methods lack this level of transparency, understanding exactly what our classifier learned is crucial. A thorough investigation of this matter will be the subject of a separate study, using explainable AI techniques such as integrated gradients and token attribution. Here, we limit our discussion to one insightful example, which seems to illustrate how the model works.

As previously mentioned in Section 2, all the genres addressed in ch. 2–5 are also discussed in the second treatise attributed to Menander Rhetor. Only the most prestigious of the epideictic genres, the panegyric — focused on in ch. 1 and 7 — does not correspond to any section in Menander's works. However, ch. 1, which provides introductory notes on panegyrics, often echoes the examples and some wording of the first treatise by Menander. Ch. 1 offers guidelines on how to appropriately praise gods ("leaders and name-givers of any festival"), cities where the festivals take place, and emperors who organize and preside over the festivals. All these topics are covered in Menander's first treatise.

Considering only ch. 2–5 or the fragments of ch. 1 that have clear parallels in Menander's work, one might argue that the classifier's decision was biased due to the significant content and semantic overlap, especially since such a tendency has been reported about the BERT-based classifiers [47]. However, the consistency of the attribution profile across the chapters by both models is reassuring, as it suggests that they capture more than just semantics.

Menandrean association appears all the stronger when the values for the logical subdivisions of the *ars*, ch. 1–7, are calculated. As Korenjak [35] has shown, in its current form, the order of the chapters is disorganized, and it is possible that the author intended to arrange them as follows: chapters

Table 7
Content overlap between the *ars* and the second treatise attributed to Menander. Chapter division for Menander's treatise follows Race [39].

Menander Rhetor Treatise II	*ars*
5, 6	2, 4
7	3
9	5
8, 10, 15	6

1 and 7 (panegyrics or appraisal speeches), chapters 2–4 (speeches related to family life occasions), and chapters 5–6 (speeches addressed to officials and epitaphs). In each of these sections, Menander maintains a stable leadership (Tables 5 and 6).

***ars*, ch. 8–11: multiple authors?** The discrepancy between the attribution profiles of ch. 8–9 and ch. 10–11 might suggest a division, albeit a less distinct one, than ch. 1–7 versus ch. 8–11. This result aligns with the assessment made by Usener [25], although it does not provide any further hint at the identity of the possible author.

However, the opposite hypothesis should still be considered seriously. In ch. 1–7, top two *single* attributions (i.e., Menander Rhetor and another author) in terms of "share" would cover *at least* 0.58–0.68 of the attributed chunks (ch. 7). In contrast, the top two attributions in ch. 8–11 provide, *at best*, 0.58–0.62 of the attributions (ch. 11 and 10), the attributions are more evenly distributed. Apparently, among the author classes present in our dataset, none is stylistically similar enough to the text of ch.8–11. This can be explained in two ways. Texts written in a comparable style are either completely absent from the dataset or are not appropriately distributed among author groups, making it challenging for the model to learn the features of this particular writing style. Keeping in mind the existing hypothesis about the relationship between the so-called Hermogenean canon and the works ascribed to Apsines, with extreme caution, we incline to the latter explanation.

Two works, which are part of the Hermogenean canon, *On Invention* and *On Method*, were already in Late Antiquity associated with the name of Hermogenes. In our dataset, therefore, following the TLG, we reproduce this conventional attribution. Yet, both are most likely inauthentic [48, 49, 50, 51, 52]. If the argumentation presented by Heath [53, 41] proves correct and these two texts can securely be ascribed to Apsines, the "new" writing style they would represent might possibly demonstrate a more pronounced affinity with the style of ch. 8–11. This and similar possibilities should be thoroughly checked in further experiments.

The scope of the much-needed detailed follow-up study becomes evident. A systematic and critical reassessment of attribution problems within the corpus of the *Rhetores Graeci* is necessary. Beyond merely reflecting on the attributions of individual works, it is important to establish the homogeneity of different rhetorical corpora within the framework of a pairwise authorship verification study.

But if we set aside the obscure case of ch. 8–11, should we conclude that ch. 1–7 were written by Menander Rhetor? Given the aforementioned limitations of our dataset, we would not go that far. However, our results suggest that the connection between the first part of the Pseudo-Dionysian *ars* and the Menandrean corpus likely extends beyond a theoretical affinity. Despite the obvious terminological discrepancies between the texts and their different levels of elaboration, the possibility of multiple authorship within the same school, or even common authorship, should be considered with all seriousness. The divergence between the *ars*, ch. 1–7, and the Menandrean corpus can also be explained, apart from the natural evolution of personal style and preferences, by the likelihood that those presenting complex rhetorical theory would probably follow the advice formulated by the author of ch. 11. The art of rhetoric involves presenting material in a way that convinces the audience. Thus, orators are similar to doctors who must not only select the right medication but also administer it in a manner acceptable to the patient [26, ch. 11, par. 9, p. 385, ll. 7–12]. In other words, multiple contextual

factors influenced the style of the presentation, and, in the cases when the stylistic affinity is clear, one should not probably overinterpret isolated differences.

7. Conclusion

This study uses transformer-based models to analyze ancient rhetorical texts for authorship attribution in classical philology. First, we adapted these models to handle the linguistic nuances of Ancient Greek texts from the 1st to the 4th century AD using masked language modeling. We then apply the fine-tuned models to identify authorship markers in *Ars Rhetorica*, a text possibly written by multiple ancient writers. This application not only reminds of benefits of modern AI techniques to classical studies but also deepens our understanding of ancient literary compositions through modern computational methods.

The results of BERT and RoBERTa classifiers do not support connection of the *ars* to Dionysius of Halicarnassus, going in line with the previous studies that question his authorship. They also strengthen the link of *ars* to the Menandrean corpus, particularly evident in the distinct attribution profiles between chapters 1–7 and 8–11, which suggests a composite nature of the work.

Despite the lack of transparency of MLM techniques compared to conventional methods, which prioritize human-interpretable features, the effectiveness and relevance of machine learning methods is noteworthy.

While neural networks are often criticized in digital humanities for their black-box nature [44], their ability to detect writing styles make them a valuable tool in the field of digital humanities. The use of these models promises significant advancements in authorship attribution and our understanding of ancient literary works.

8. Limitations

This study has several limitations that should be considered when interpreting the results.

Firstly, the issue of disputed authorship within the dataset is a significant challenge. For instance, the Hermogenean corpus and Menandrean treatises, both central to our analysis, have long-standing debates regarding their true authorship, see Section 4. These uncertainties could affect the attribution accuracy. We are currently working on a study intended to solve this issue, adopting an authorship verification approach.

Secondly, the use of transformer-based models like BERT and RoBERTa, come with limitations related to their opaque nature. The lack of interpretability in these models means that understanding the specific features and patterns the models use to make attributions is challenging. This limits our ability to provide a transparent rationale for the models' decisions, which is often critical in digital humanities research. Yet, the attempts were made to find way to make the results of pre-trained language models more interpretable, e.g., by means of the so-called integrated gradients [54]. These methods can perhaps be adapted for cases similar to ours.

Despite achieving notable accuracies with relatively short chunks (64 tokens), the models' performance still leaves room for improvement, particularly in terms of handling unbalanced corpora and downplaying the influence of the thematic clues. Nevertheless, their performance, comparable to state-of-the-art results for modern languages, demonstrates an ability to capture writing style. There clearly are instances where the models are overly confident, leading to incorrect authorship attribution. These errors could arise from factors such as the models' sensitivity to stylistic nuances and the complexity of the texts. Embracing more sophisticated methodologies for uncertainty-aware training would be an interesting avenue for further exploration.

Another potential avenue for future research is the development of chronological and regional classifiers. Texts from different regions and periods may exhibit unique linguistic and stylistic features that are not captured by a generalized model. Developing classifiers specific to historical periods or

geographical (and cultural) areas could enhance attribution accuracy and offer more detailed insights into the *ars* and many other texts.

Acknowledgments

We extend their gratitude to Jürgen Jost, Charlotte Schubert, Friedrich Meissner, Caroline Macé, and Mark de Kreij for welcoming this study and future collaboration between machine learning, history, and philology.

We would also like to thank Ben Nagy and two anonymous reviewers for the careful reading and insightful feedback.

We thank Shari Boodts and Sven Meeder, Principal Investigators of the ERC Proof of Concept project "ManuscriptAI" and the ERC Consolidator project "SOLEMNE". Without their support, this research would not have been possible.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, 2017. URL: https://arxiv.org/abs/1706.03762.

[2] R. Sprugnoli, M. Passarotti, F. M. Cecchini, M. Fantoli, G. Moretti, Overview of the EvaLatin 2022 evaluation campaign, in: R. Sprugnoli, M. Passarotti (Eds.), Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, European Language Resources Association, Marseille, France, 2022, pp. 183–188. URL: https://aclanthology.org/2022.lt4hala-1.29.

[3] F. Riemenschneider, A. Frank, Exploring Large Language Models for Classical Philology, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 15181–15199. doi:10.18653/v1/2023.acl-long.846.

[4] R. Sprugnoli, M. Passarotti (Eds.), Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024, ELRA and ICCL, Torino, Italia, 2024.

[5] M. Fabien, E. Villatoro-Tello, P. Motlicek, S. Parida, BertAA : BERT Fine-Tuning for Authorship Attribution, in: P. Bhattacharyya, D. M. Sharma, R. Sangal (Eds.), Proceedings of the 17th International Conference on Natural Language Processing (ICON), NLP Association of India (NLPAI), Indian Institute of Technology Patna, Patna, India, 2020, pp. 127–137. URL: https://aclanthology.org/2020.icon-main.16.

[6] J. Tyo, B. Dhingra, Z. C. Lipton, On the State of the Art in Authorship Attribution and Authorship Verification, 2022. URL: https://arxiv.org/abs/2209.06869. arXiv:2209.06869.

[7] J. Huertas-Tato, A. Huertas-Garcia, A. Martin, D. Camacho, PART: Pre-Trained Authorship Representation Transformer, 2022. URL: http://arxiv.org/abs/2209.15373.

[8] B. Ai, Y. Wang, Y. Tan, S. Tan, Whodunit? Learning to Contrast for Authorship Attribution, 2022. doi:10.48550/ARXIV.2209.11887.

[9] G. Storey, D. Mimno, Like Two Pis in a Pod: Author Similarity Across Time in the Ancient Greek Corpus, Journal of Cultural Analytics 5 (2020). doi:10.22148/001c.13680.

[10] B. Graziosi, J. Haubold, C. Cowen-Breen, C. Brooks, Machine Learning and the Future of Philology: A Case Study, TAPA 153 (2023) 253–284. doi:10.1353/apa.2023.a901022.

[11] T. Sommerschield, Y. Assael, J. Pavlopoulos, V. Stefanak, A. Senior, C. Dyer, J. Bodel, J. Prag, I. Androutsopoulos, N. De Freitas, Machine Learning for Ancient Languages: A Survey, Computational Linguistics 49 (2023) 703–747. doi:10.1162/coli_a_00481.

[12] T. Sommerschield, Y. Assael, J. Pavlopoulos, V. Stefanak, A. Senior, C. Dyer, J. Bodel, J. Prag, I. Androutsopoulos, N. de Freitas, Machine learning for ancient languages: A survey, Computational Linguistics (2023) 703–747. doi:10.1162/coli_a_00481.

[13] D. Bamman, P. J. Burns, Latin BERT: A Contextual Language Model for Classical Philology, 2020. URL: https://arxiv.org/abs/2009.10053.
[14] P. Singh, G. Rutten, E. Lefever, A Pilot Study for BERT Language Modelling and Morphological Analysis for Ancient and Medieval Greek, in: S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, S. Szpakowicz (Eds.), Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, Association for Computational Linguistics, Punta Cana, Dominican Republic (online), 2021, pp. 128–137. doi:10.18653/v1/2021.latechclfl-1.15.
[15] I. P. Yamshchikov, A. Tikhonov, Y. Pantis, C. Schubert, J. Jost, BERT in Plutarch's Shadows, 2022. URL: http://arxiv.org/abs/2211.05673.
[16] J. Koutsikakis, I. Chalkidis, P. Malakasiotis, I. Androutsopoulos, Greek-BERT: The Greeks Visiting Sesame Street, in: 11th Hellenic Conference on Artificial Intelligence, SETN 2020, Association for Computing Machinery, New York, NY, USA, 2020, p. 110–117. doi:10.1145/3411408.3411440.
[17] T. C. Burgess, Epideictic literature, volume 3, University of Michigan Library, 1902.
[18] G. W. Bowersock, Greek sophists in the Roman Empire, Clarendon Press, Oxford, 1969.
[19] G. A. Kennedy, Greek Rhetoric under Christian Emperors, volume 3, Wipf and Stock Publishers, 2008.
[20] B. E. Borg, Paideia: the World of the Second Sophistic, de Gruyter, 2008.
[21] C. Walz, Rhetores Graeci, 1834.
[22] L. Spengel, Rhetores Graeci, volume 1, Teubner, 1885.
[23] G. A. Kennedy, Some Recent Controversies in the Study of Later Greek Rhetoric, American Journal of Philology 124 (2003) 295–301.
[24] L. Sadée, De Dionysii Halicarnassensis scriptis rhetoricis quaestiones criticae, Teubner, Strasbourg, 1878.
[25] H. Usener, Dionysii Halicarnasei quae fertur Ars Rhetorica, Teubner, Leipzig, 1895.
[26] H. Usener, L. Radermacher (Eds.), Dionysii Halicarnasei quae exstant. Vol. 6: Opuscula, volumen secundum, volume 6, Teubner, Stuttgart–Leipzig, 1929.
[27] H. Rabe, Rhetoren-Corpora, Rheinisches Museum 67 (1912) 321–357.
[28] M. Heath, Pseudo-Dionysius Art of Rhetoric 8-11: Figured speech, Declamation, and Criticism, American Journal of Philology 124 (2003) 81–105.
[29] D. A. Russell, Classicizing Rhetoric and Criticism: the Pseudo-Dionysian Exetasis and Mistakes in Declamation, Le Classicisme à Rome aux 1ers siècles avant et après J.-C (1979).
[30] D. A. Russell, Rhetors at the Wedding, Proceedings of the Cambridge Philological Society 25 (1979) 104–117. doi:10.1017/S0068673500004156.
[31] H. Schott, TEXNH PHTOPIKH: quae vulgo integra Dionysio Halicarnassensi tribuitur, emendata, nova versione Latina et commentario illustrata, Sumtibus E.B. Suicquerti, 1804. URL: https://books.google.nl/books?id=SiYUAAAAYAAJ.
[32] F. Blass, De Dionysii Halicarnassensis scriptis rhetoricis, Max Cohen et fil., Bonn, 1863. URL: https://books.google.nl/books?id=k3g-AAAAcAAJ.
[33] K. Weismann, De Dionysii Halicarnassei vita et scriptis: Diss. inaug, Steuber, 1837. URL: https://books.google.nl/books?id=5XJSAAAAcAAJ.
[34] G. Thiele, Dionysii Halicarnasei quae fertur ars rhetorica rec. Hermannus Usener, Göttingische Gelehrte Anzeigen 159 (1897) 237–43.
[35] M. Korenjak, Ps.-Dionysius *Ars Rhetorica*I-VII: One Complete Treatise, Harvard Studies in Classical Philology 105 (2010) 239–254.
[36] J. Penndorf, De sermone figurato quaestio rhetorica, Leipziger Studien zur classischen Philologie 20 (1902) 169–194.
[37] K. Schöpsdau, Untersuchungen zur Anlage und Entstehung der beiden pseudodionysianischen Traktate περὶ ἐσχηματισμένων, Rheinisches Museum für Philologie 118 (1975) 83–123.
[38] H. Rabe, Hermogenis Opera, Teubner, 1985. URL: https://books.google.nl/books?id=WreAtwEACAAJ.
[39] W. H. Race, Menander Rhetor. Dionysius of Halicarnassus, Ars Rhetorica, volume 539 of *Loeb*

classical library, Harvard University Press, Cambridge (Mass.) London, 2019.
[40] D. A. Russell, N. G. Wilson, Menander Rhetor, Clarendon Press, Oxford, 1981.
[41] M. Heath, Menander: a Rhetor in Context, Oxford University Press, USA, 2004.
[42] K. Brodersen, Menandros. Abhandlungen zur Rhetorik, volume 88 of *Bibliothek der griechischen Literatur*, Anton Hiersemann, Stuttgart, 2019.
[43] S. Corbara, A. Moreo, F. Sebastiani, M. Tavoni, MedLatinEpi and MedLatinLit: Two Datasets for the Computational Authorship Analysis of Medieval Latin Texts, 2021. URL: http://arxiv.org/abs/2006.12289.
[44] T. Clérice, A. Glaise, Twenty-One* Pseudo-Chrysostoms and more: Authorship Verification in the Patristic World, in: Computational Humanities Research Conference 2023, Proceedings of the Computational Humanities Research Conference 2022, 2023. URL: https://inria.hal.science/hal-04211176.
[45] N. Manousakis, E. Stamatatos, Authorship Analysis and the Ending of Seven Against Thebes: Aeschylus' Antigone or Updating Adaptation?, Classical World 116 (2023) 247–274. doi:10.1353/clw.2023.0007.
[46] P. Beullens, W. Haverals, B. Nagy, The Elementary Particles: A Computational Stylometric Inquiry into the Mediaeval Greek-Latin Aristotle, Mediterranea. International Journal on the Transfer of Knowledge 9 (2024) 385–408. doi:10.21071/mijtk.v9i.16723.
[47] F. Brad, A. Manolache, E. Burceanu, A. Barbalau, R. Ionescu, M. Popescu, Rethinking the Authorship Verification Experimental Setups, 2022. URL: https://arxiv.org/abs/2112.05125.
[48] E. Bürgi, Ist die dem Hermogenes zugeschriebene Schrift Περὶ μεθόδου δεινότητος echt? I., Wiener Studien 48 (1930) 187–197.
[49] E. Bürgi, Ist die dem Hermogenes zugeschriebene Schrift Περὶ μεθόδου δεινότητος echt? II., Wiener Studien 49 (1931) 40–69.
[50] M. Patillon, Le *De Inventione* du Pseudo-Hermogène, volume 34/3 of *Aufstieg und Niedergang der römischen Welt*, De Gruyter, Berlin, Boston, 1997, pp. 2064–2172. doi:10.1515/9783110815146-003.
[51] M. Patillon, Pseudo-Hermogène, *L'Invention*. Anonyme, *Synopse des exordes*, volume 3, 1 of *Corpus rhetoricum*, Les Belles lettres, Paris, 2012.
[52] M. Patillon, Pseudo-Hermogène, *La méthode de l'habilité*. Maxime, *Lex objections irréfutables*. Anonyme, *Méthodes des discours d'adresse*, volume 5 of *Corpus rhetoricum*, Les Belles lettres, Paris, 2014.
[53] M. Heath, Hermogenes' Biographers, Eranos 96 (1998) 44–54.
[54] M. Sundararajan, A. Taly, Q. Yan, Axiomatic Attribution for Deep Networks, 2017. URL: https://arxiv.org/abs/1703.01365. arXiv:1703.01365.

A. Online Resources

The code and both models considered in detail in this study are accessible at:

- https://huggingface.co/glsch
- https://github.com/glsch/rhetores_graeci