Examining the Authenticity of Plato's Epistle VII through Deep Learning

Examining the Authenticity of Plato's *Epistle VII* through Deep Learning

A thesis presented by
Jordan Bliss Perry
to
The Departments of Computer Science and the Classics

in partial fulfillment of the requirements
for the degree with Honors of Bachelor of Arts
in the subject of Languages and Literatures

Directed by Stuart Shieber and Mark Schiefsky

Harvard University
Cambridge, Massachusetts
March 2021

Thesis advisors: Professors Stuart Shieber and Mark Schiefsky Jordan Bliss Perry

Examining the Authenticity of Plato's *Epistle VII* through Deep Learning

Abstract

Plato's *Epistle VII*, a text in which the famous Athenian philosopher describes his political involvement in the affairs of 4th-century B.C.E Syracuse, has long been considered dubious by classical philologists. In particular, scholars have scrutinized two sections of the letter, in the first of which Plato gives political advice contrary to other claims made in his other works, and in the second of which Plato digresses from his political narrative to discuss a philosophical doctrine known as the Theory of Forms. Specifically, some scholars have raised the possibility of textual interpolation, whereby inauthentic passages might have been added to an otherwise authentic text.

This paper sets out to apply computational methodology from deep learning to provide further insight on such a long-standing problem in Platonic scholarship. As such, I developed a bidirectional long-short-term memory (LSTM) recurrent neural network (RNN) with trainable word embeddings to classify units of roughly 100 words of Ancient Greek text as belonging to Plato or one of six other Ancient Greek prose authors. Given Ancient Greek's rich morphology, special care was taken to formulate an optimal pre-processing approach: of four methods — plaintext, lemmatization, byte-pair encoding (BPE), and a lemmatization-BPE ensemble — the ensemble exhibited the highest test accuracy (89.28%), improving significantly upon a Naïve Bayes baseline model (70.93%). Applied to *Epistle VII*, this model reveals that the letter seems mostly authentic, except for two markedly more spurious sections, one of which corresponds nearly perfectly with the boundaries of the section consisting of political advice to the Sicilians. Such a result provides further support to the pre-existing claim that this section is an interpolation by a non-Platonic author within an otherwise Platonic text.

Contents

Listing of figures

Acknowledgments

This thesis would not have been possible without the guidance of my wonderful advising team. I'd like to thank Stuart Shieber, Mark Schiefsky, and Nadav Asraf for encouraging me at every phase of this process and always pushing me to take my research one step further. Writing a paper in both computer science and Classics at the same time was no simple feat, and I couldn't have asked for more helpful support. Furthermore, I must express my appreciation for the helpful feedback given by Chris Tanner and Said Esteban Belmehdi along the way, specifically during my editing rounds.

Beyond this thesis, it behooves me to recognize all of the teachers throughout my entire education who have fostered my journey in Classics thus far: Paul Langford, who first introduced me to Plato; Nick Unger, who encouraged me to appreciate the nuances of the Greek language; Sally Morris, who guided my baby steps in Greek grammar; Mark Hauser, whose rapid-fire middle school Latin class first introduced me to ancient languages; and Joy Hurd, who convinced me at a young age that Ancient Greek could indeed be cool.

Last but not least, I must thank my family for their support, encouragement, and direction, and for always rooting for me as I have pursued my passions and dreams. This project would not have occurred without the lifelong love of learning that you have striven above all to instill in me. To conclude with an obligatory Greek quote from Socrates, it is through you that I have learned that ὁ ἀνεξέταστος βίος οὐ βιωτὸς ἀνθρώπῳ — that the unexamined life is not worth living.

1

Introduction

One of the most longstanding debates in Platonic scholarship concerns the authenticity of the *Epistles*, a collection of letters traditionally included as part of the Platonic corpus alongside famous texts such as the *Republic*, *Symposium*, and *Apology*. Accordingly, this paper concerns itself with using modern computational methods from natural language

processing and deep learning to shed further light on this debate over authorship. Before developing such methodology, however, I begin with one simple question: what makes a text Platonic?

1.1 Questions of Authenticity in the Platonic Corpus

That is, if centuries of scholarship have questioned whether the famous Athenian philosopher Plato actually wrote many of the texts attributed to him, why are those works still soundly included nowadays in the Platonic corpus? Perhaps one answer can be found in the literary opinions expressed by a relatively obscure Egyptian Greek intellectual from the beginning of the first century C.E. Among various other pursuits, from astrology and music theory to serving in the court of the Roman emperor Tiberius, this scholar, Thrasyllus of Mendes, is perhaps most famous in modern times for having organized Plato's writings into an arrangement which has survived to this day. Specifically, Thrasyllus divided the 36 works he considered part of the Platonic canon into tetralogies — groups of four, a pattern reminiscent of traditional Athenian theater, according to which three plays would be grouped together with one satyr play. The nine tetralogies devised by Thrasyllus form the organization of the surviving medieval manuscripts through which Plato's works were preserved into modern times, and still remain the basic division of Plato's text in many authoritative editions, translations, and commentaries today, such as John Burnet's multi-volume edition in the Oxford Classical Texts series. That is, the 36 texts considered authentic by Thrasyllus have remained under Plato's name, while the nine texts he rejected have been relegated to the label of the "spurious dialogues of Plato," not even appearing in some modern editions.

Nevertheless, the fact that Thrasyllus' arrangement has endured does not mean that it has never been questioned. In particular, various scholars have expressed skepticism regarding the authenticity of many of the works included. Thrasyllus' enumeration of the nine "spurious dialogues" alongside Plato's thirty-six canonical works implies his consideration of the latter as authentic, yet modern scholars have raised various concerns about the authorship of ten of these thirty-six texts (*Alcibiades*, *Second Alcibiades**, *Hipparchus**, *Rival Lovers**, *Theages**, *Greater Hippias*, *Clitophon*, *Minos**, *Epinomis**, *Epistles*), even reaching consensus on the inauthenticity of the works marked with asterisks (Cooper in Plato 1997, vi). These philologists, philosophers, and stylometrists have extensively debated the content as well as the style of these works, using tactics ranging from considering the compatibility of the philosophical positions taken in the different texts to analyzing the linguistic features of the Greek used to express these ideas. The scholar John Cooper, for example, rejects the *Second Alcibiades* as Plato's because the dialogue's Greek displays traces of Northern Greek features from the 3rd century B.C.E. (inconsistent with Plato's biography as an Athenian from the 5th-4th century B.C.E), whereas he doubts the authenticity of the *Epinomis* because of its conflicts with the opinions expressed in the Laws, the work to which it is explicitly presented as an appendix (Cooper in Plato 1997, 597; 1617). Nevertheless, despite this recurring pattern of scholarly concern regarding many of the included works, Thrasyllus' arrangement has remained intact until now, probably because the subjective and inconclusive character of questions of authenticity has forced later Platonic editors to include all of the works despite suspicions that certain texts are spurious (Cooper in Plato 1997, x).

1.2 Plato's *Epistles* and Their Significance

Nine of the ten texts of dubious authenticity listed above belong to the literary genre of dialogue — that is, descriptions of philosophical conversations usually featuring Plato's teacher Socrates as an interlocutor — as do all of the works considered authentic. The one exception is the *Epistles*, or Letters, a collection of thirteen letters purportedly written by Plato himself to a set of various intended recipients. Unlike the dialogues, in which Plato never features as a character and in which the line between Plato's own thought and that of Socrates can be notoriously unclear, the *Epistles* feature Plato narrating his own biography and expressing personal opinions on various events in his lifetime. The most recurring theme, appearing in eight of the thirteen letters, is Plato's involvement in the political affairs of Syracuse, a Greek city in Sicily. These letters paint a picture of three visits by Plato to the city, from which one can piece together the following story, as paraphrased from Robert Bury's commentary on the *Epistles*:

Plato first visited Sicily at approximately forty years old (388–387 B.C.E.), when he struck up a friendship centered around similar philosophical opinions with Dion, brother-in-law of Dionysius I, the tyrant of Syracuse. Over the next twenty years, while Dionysius I successfully governed Syracuse and oversaw an intellectual blossoming in the city, his successor, Dionysius II, showed himself a more deficient leader. On Dion's urging, Plato made his second visit to Sicily (367 B.C.E.) to help Dionysius II grow into the famous ideal of the philosopher-king proposed in Plato's *Republic*, yet Dionysius II soon became disillusioned with the reforms urged by the two philosophers and drove Dion into exile, prompting Plato to leave shortly thereafter. Nevertheless, Dionysius II still desired the

reputation of an intellectually enlightened ruler and continued to court Plato, culminating in Plato's third visit (361–360 B.C.E.), which ended in failure as Dionysius II doubled down on his refusal to restore Dion and showed only superficial interest in Plato's philosophical teachings. After Plato escaped — essentially having been put under house arrest in the palace — he met with Dion at the Olympic Games, whereupon Dion decided to overthrow Dionysius II militarily. The intervention succeeded, yet power struggles between Dion and the other leaders escalated, eventually culminating in the assassination of Dion and the plunging of Sicily into civil war (Bury in Plato 1999, 386–389).

Scholarly opinion on the authenticity of these letters as a whole has been notoriously fickle. A survey of the particular trains of argumentation will be presented later in this paper; what is notable is that there are some scholars who reject all of the letters as spurious, some who reject some of them, and a (much) smaller group who accept all of them. Furthermore, the level of confidence ascribed to each letter's authenticity also ranges significantly, with *Epistles VII* and *VIII* considered most the likely of the collection to be genuine, *I, V, IX* and *XII* considered most likely inauthentic, and the others subject to more disagreement among scholars (Bury in Plato 1999, 391–392). Indeed, a full consideration of the relations between the individual letters with regard to authorship could occupy many more studies and will not be attempted in this paper.

Instead, I will concern myself solely with the question of the authorship of *Epistle VII*, by far the longest and most substantive of the letters in the collection. Written as an open letter providing advice to the supporters of Dion after his assassination, *Epistle VII* more or less takes the literary form of an *apologia* recounting and justifying Plato's political involvement in Sicily. The letter proceeds as a detailed autobiographical account of Plato's three visits to

Sicily and of his relationships with Dion and Dionysius II, with two notable digressions — the first, holding true to the stated purpose of the letter, his counsel to Dion's supporters (330c–337d); the second, a philosophical *excursus* expanding a doctrine known as the "Theory of Forms" observed in many of the genuine Platonic dialogues (341b–345c). The Stanford Encyclopedia of Philosophy defines this theory as stating that "the world that appears to our senses is in some way defective and filled with error, but there is a more real and perfect realm, populated by entities (called "forms" or "ideas") that are eternal, changeless, and in some sense paradigmatic for the structure and character of the world presented to our senses" (Kraut 2017). In this specific context, Plato lays out this doctrine and uses it to explain why Dionysius II's efforts to write philosophical treatises were misguided, because, as a corollary of this theory, true knowledge cannot be expressed in writing.

As such, *Epistle VII* has been the letter whose authenticity has been examined with the most scrutiny, because of the significant ramifications its content has for our knowledge of Plato's biography and philosophical doctrine. Regarding the former, the seventh letter provides a rare portrayal of a pragmatic, rather than theoretical, approach to politics by Plato. In his dialogues, Plato is famous for theoretical expositions on political philosophy — notably, in the *Republic*, in which his portrayal of the character Socrates calls for the installation of "philosopher-kings" as the most noble form of government — yet his dialogues purposefully avoid describing his practical political involvement or philosophical opinion in general, instead voicing arguments through characters such as Socrates. The seventh letter, in contrast, paints a picture of Plato the statesman, involving himself in Syracusan power struggles and actually attempting to carry out the ideals of governance

preached in his dialogues; if true, this portrayal provides a pragmatic counterweight to the "ivory tower," removed-from-society image commonly held of his Athenian Academy. In general, evidence that *Epistle VII* is inauthentic would cast significant doubt upon the veracity of the surviving account of Plato's time in Sicily, because it is our main source for those adventures. Furthermore, the question of *Epistle VII*'s authenticity also has strong implications for our conceptions of Plato's Theory of Forms as described in the previous paragraph. The ancient philosophy scholar Anthony Kenny states that "the clearest short statement of the [Theory of Forms] is found not in the dialogues but in the seventh of the letters traditionally attributed to Plato" (Kenny 2004, 75). Indeed, the letter is the only place where Plato explains the idea directly in his own voice, as opposed to the dialogues, where the theory is conveyed indirectly through the speech of interlocutors such as Socrates. Therefore, decisive evidence against the authenticity of this letter would change our understanding of one of Plato's key philosophical principles, including, for example, the claim made in the letter that writing philosophical treatises is a useless activity because it cannot accurately capture the essence of the knowledge one has acquired. Indeed, this opinion, if true, provides a concrete reason for Plato's decision to use the dialogue instead of the treatise as the literary mechanism through which to convey and record his ideas and those of his teacher, Socrates.

1.3 A Stylometric Approach

While countless philologists have debated the content and style of the seventh letter using qualitative methods for years, other scholars have sought to take a more quantitative approach — known as stylometry — to gain a more precise, low-level analysis of Plato's

style of Greek. Although the application of this approach to the Platonic corpus extends over a century back, with stylometrists such as Wincenty Lutosławski (1897) gathering hand-counted textual statistics from Plato's works and deriving various metrics from them, the field has been revolutionized by developments in computational power, text digitization, and machine learning, enabling more robust studies such as those by Gerard Ledger (1989) and Thomas Koentges (2020). These studies have shown strong promise in their ability to model what one might refer to as the "subconscious" features of Plato's style beyond particularly salient patterns such as the occurrence of rare words or the avoidance of certain phrases. Indeed, outside of Ancient Greek literature, stylometric methods have been key to solving such authorship problems as the authenticity of the Federalist Papers, with Frederick Mosteller and David Wallace demonstrating the usage of filler words such as "upon" as discriminating features between text written by Madison and by Hamilton, a study which shows the feasibility of a content-agnostic approach to settling questions of disputed authorship (Mosteller and Wallace 1963).

Specifically, the goal of my research is to enhance this stylometric branch of the scholarship concerning the authenticity of *Epistle VII* through the introduction of state-of-the-art techniques from natural language processing (NLP) and specifically deep learning (DL) to this age-old debate. Recent years have seen significant advances in the power of neural networks and the computing resources needed to employ these models to practical effect, and there have arisen numerous potential applications in tasks such as machine translation, sentiment analysis, text generation, and — of course — authorship attribution. One particularly promising feature of these recent advances, as will be discussed at length in later chapters, is the ability for neural models to essentially learn the

most important stylistic features in a given text on their own, providing an alternative to hand-curated lists of textual features (the methodology employed by most previous stylometrists discussed herein) and enabling analysis of many more underlying patterns in the text than a human could monitor. At the same time, improved digitization of the Ancient Greek textual corpus has enabled digital humanists to analyze these texts at much greater scale. While these advances will surely enable the production of new results concerning the authenticity of the *Epistles*, this study will give equal priority to the development of sound heuristics regarding the application of these recent advances in NLP (which were mostly developed for English or other modern languages) to the Ancient Greek language, given the relatively small degree of neural applications in Ancient Greek compared to other languages and its morphological complexity. Indeed, I hope that my work will serve another purpose as a more general reference for those attempting to apply modern machine learning tactics to questions of interest in the Ancient Greek textual corpus, beyond Platonic authorship attribution.

From here, I will proceed with an overview of some of the most common arguments in traditional Platonic scholarship regarding the authenticity of *Epistle VII*, specifically focusing on two sections of the letter often considered dubious. From there, I will describe the development of the subset of that scholarship employing computational methodology through the 20th century until today, culminating with my motivation for trying out the neural methods used in this paper. Before analyzing *Epistle VII*, however, it is necessary to develop some best practices for the usage of deep learning on Ancient Greek text given the dearth of previous examples. Thus, I will describe a few different pre-processing approaches for handling Ancient Greek's rich morphology, demonstrating that an ensemble approach

using lemmatization and byte-pair encoding results in the multi-class authorship classifier with the highest accuracy (89%). Using that model, I will proceed to analyze the seventh letter section by section, presenting the digital results in conjunction with the traditional scholarship and ultimately concluding that the letter is largely authentic with the exception of one or two interpolations. Such a multidisciplinary approach underscores my strong belief that computational analysis is most valuable when it engages with traditional scholarship — whether enhancing or undermining pre-existing claims about the authenticity of the letters.

2

Prior Philological and Stylometric Analysis of the Seventh Letter

Before presenting my computational methodology for evaluating the authorship of Plato's seventh epistle, however, I will review some previous lines of argumentation – both philological and statistical — with regard to this debate. Over the several centuries of

scholarly treatment of this question of Platonic authenticity, both sides have employed many distinct styles of argument towards their respective claims, ranging from noticing the difficulty of preserving letters in the ancient world to pointing out irregularities in the letters' content and style compared to the genuine Platonic dialogues. Hence, it would be impossible to outline all of these arguments in the scope of this paper. Rather, my goal is to provide a high-level survey of some of the recurring arguments — literary, political, and philosophical — against Platonic authorship of *Epistle VII*. This summary will be useful for contextualizing the results of my model on the letter later in this paper. In addition, I have included some prior examples of the usage of stylometry and eventually machine learning to analyze the Platonic corpus and thus the authorship question, in order to provide greater context for the evolution towards the neural methodology described in subsequent sections of this paper.

2.1 Philological Arguments

2.1.1 The Authenticity of Ancient Letters in General

One of the most basic arguments against Platonic authorship of *Epistle VII* is that the authenticity of ancient letters in general, particularly those containing philosophical content, is inherently dubious. As early as the 17th century, scholars such as Richard Bentley (1874) have argued that certain collections of ancient letters (in his case, those of Phalaris, a Sicilian tyrant from the 6th century B.C.E.) are spurious. Recently, the famous scholar of ancient philosophy Michael Frede has extended this general claim with a list of letter collections attributed to philosophers (e.g., Heraclitus, Democritus, Diogenes), all of which he claims are inauthentic. In his view, the first verifiably authentic set of Ancient

Greek philosophical letters was written by Epicurus, who lived and wrote nearly a century after Plato (Burnyeat and Frede 2015, 7-8). These two arguments, according to Frede's reasoning, should lead one to doubt, rather than presuppose, the authenticity of Plato's *Epistles*. In addition, Frede provides a few specific reasons why such a consistent trend of inauthenticity exists among ancient letter collections. For example, letters are often not written by the purported author but rather a scribe, just as speeches even in modern times are sometimes written by speechwriters — similarly, letters, like speeches, could theoretically be produced as part of rhetorical education in an attempt to emulate the style of other authors (Burnyeat and Frede 2015, 11-12). One must additionally consider the institutions of Plato's era: unlike the centralized political situation in turn-of-the-millenium Rome enabling the preservation of the famous letters written by authors such as Cicero, 4th century B.C.E. Athens, he claims, does not seem to have possessed the infrastructure of secretaries, clerks, libraries, and other bureaucratic institutions necessary for preserving written correspondence (Burnyeat and Frede 2015, 15).[1]

Those who argue in favor of the seventh letter's authenticity, however, point out a key distinction between private and public, or "open," letters, claiming that the latter are exempt from the doubt explained in the previous paragraph. For example, Robert Bury, the editor of the Loeb edition of Plato's *Epistles*, agrees that the difficulties of preservation justify an *a priori* skepticism of ancient letters, yet contends that the seventh letter (as well as the eighth) does not fit into this category because it is an "open" letter (Bury in Plato 1999, 390–391). Specifically, even though the letter is addressed specifically to the people

1. Nevertheless, one must consider the longevity of Plato's Academy, which survived for centuries — if there was any institution in Athens possessing the necessary bureaucracy to preserve letters, it would have been the Academy.

of Sicily, a few scholars have postulated that the dismissive attitude shown towards the Sicilian way of life (Pl. *Ep. VII.* 326b–326d), the long re-counting of recent events in Sicily with which any local would be familiar, and the unlikeliness of a direct appeal by Syracusan citizens to Plato for political advice all indicate that the letter is, as described by J. Harward, a "literary document" meant for consumption in Athens and across all of Greece rather than an actual address to a narrower Sicilian audience (Harward 1928, 144; 148). Bury echoes such arguments himself and juxtaposes this status of an "open" letter with the letter's function as an apologia of Plato's actions in Sicily (Bury in Plato 1999, 473). Given that "open" letters have a much larger chance of preservation due to their wider audience and typical function as manifestos or essays — essentially, speeches in written form — this group of scholars tends to consider the possibility of Platonic authorship for the seventh (and eighth) letter much higher than for the other letters, and argues at least that these letters should be treated by default as authentic given their "open" status (Bury in Plato 1999, 391–392). Indeed, most arguments in favor of the authenticity of *Epistle VII* follow this format – that is, theoretically justifying, rather than empirically proving, the authenticity of the letter.

2.1.2 Plato's Opinions on Political Involvement

Some scholars, however, maintain that the types of arguments above — namely, those relating to the authenticity and intended audience of ancient letters in general — do not necessarily rule out or confirm Platonic authorship of *Epistle VII*, but simply place the burden of proof on those who seek to claim its genuineness, instead of assuming *a priori* Platonic authorship. Frede is of this belief, as is Ludwig Edelstein, who devoted a whole book, *Plato's Seventh*

Letter, to evaluating the authenticity of the letter. For Edelstein, the question now turns to what type of evidence would provide decisive proof of authenticity. After dismissing the soundness of historical correspondence — stating that an astute forger could also possess knowledge of the general historical context and even Plato's personal biography — Edelstein claims that Platonic authorship can only be evaluated on the basis of the letter's content and its relation to the ideas put forth in other Platonic texts, namely, the dialogues (Edelstein 1966, 2).

In particular, one main focus of this type of investigation has been Plato's opinions on politics as expressed in the seventh letter vis-à-vis the claims made in other genuine Platonic works. The crux of this question revolves around Plato's intervention in the internal affairs of Syracuse in *Epistle VII*, and the description of this intervention as measured against the ideals presented in other texts such as the *Republic* and the *Laws*. At the beginning of the letter, it is stated that "wherefore the classes of mankind (I said) will have no cessation from evils until either the class of those who are right and true philosophers attains political supremacy, or else the class of those who hold power in the States becomes, by some dispensation of Heaven, really philosophic" (Pl. *Ep. VII.* 326a-b).[2] In other words, the letter's author claims that the only government which can ensure the happiness of its citizens is one comprised of philosopher-kings, an idea which essentially restates a section from Plato's signature work the *Republic* (Pl. *Resp. V.* 473c-d). At first, this compatibility does not suggest inauthenticity of authorship in the slightest. However, such a perspective changes when one considers the dating of these works (established through chronological

2. Translations of the following texts are taken from the Loeb Classical Library editions: *Epistle VII* (Bury in Plato 1999), the *Laws* (Bury in Plato 1926), and the *Republic* (Emlyn-Jones and Preddy in Plato 2013a), (Emlyn-Jones and Preddy in Plato 2013b).

information presented in the letter).[3] It is thought likely that *Epistle VII* was written at more or less the same time (354/353 B.C.E.) as the *Laws*, another famous work by Plato, which was certainly composed in a later period of Plato's career than the *Republic* (Burnyeat and Frede 2015, 51). However, unlike the *Republic*, the *Laws* specifically does not endorse a regime of philosopher-kings as the only viable government, but rather proposes a "mixed constitutional" government featuring an assembly, council, and elected officials (Burnyeat and Frede 2015, 51). Frede claims that this idea indicates a marked evolution in Plato's thought since the composition of the *Republic* and sees this dichotomy between the language of philosopher-kings in *Epistle VII* and that of constitutional government in the *Laws* as irreconcilable, rejecting the possibility that such different ideas could have been written by the same author at the same time (Burnyeat and Frede 2015, 55-56). Given that the authenticity of the *Laws* is not disputed by scholars, proponents of this argument consider Platonic authorship of *Epistle VII* as dubious.

Moreover, there exist statements in the seventh letter that conflict with both the *Laws* and the *Republic*. At one point, for example, Plato advises Dion's followers that one "ought not to apply violence to his fatherland in the form of a political revolution, whenever it is impossible to establish the best kind of polity without banishing and slaughtering citizens, but rather he ought to keep quiet and pray for what is good both for himself and for his State" (Pl. *Ep. VII.* 331d). Such a renouncement of violence is remarkable, considering that Plato states in the *Republic* that the philosopher-kings, upon gaining power, would "take a state ... and people's way of life, rather as you do a writing tablet ... [and] wipe it clean — not an easy task" (Pl. *Resp. VI.* 501a), and in the *Laws* that "the best purge is painful, like all medicines of a drastic

3. The letter references as a recent event the death of Dion, which is known to have occurred in 354 B.C.E.

nature – the purge which hales to punishments by means of justice linked with vengeance, crowning the vengeance with exile or death" (Pl. *Leg. V.* 735d-e). For this dichotomy between the Plato of the seventh letter and the Plato of the dialogues to be reconcilable, Edelstein states that Plato must have rejected the beliefs he previously espoused when writing the *Republic* and the *Laws*, a possibility he considers unlikely given the overt allusions to those texts and the ideals of a "philosopher-king" in the letter (Edelstein 1966, 24). Thus, according to Edelstein, those who promote the letter's authenticity must take on the burden of proof in decisively reconciling these conflicting positions (Edelstein 1966, 2).

Other scholars have focused not so much on the external conflicts between the letter and other Platonic texts such as the *Laws* as on the internal discord between different sections of the letter itself regarding political philosophy. In one section of the letter, the author claims, alongside Dion, to have advised Dionysius II that "if he pursued the course we describe, and made himself right-minded and sober-minded, then, if he were to re-people the devastated cities of Sicily and bind them together by laws and constitutions so that they should be leagued both with himself and with one another against barbarian reinforcements, he would thus not merely double the empire of his father but actually multiply it many times over" — essentially, that Dionysius should adopt constitutional government for the good of his kingdom and his people, a proposition which, notably, does not mention a regime of philosophers in any sense (Pl. *Ep. VII.* 332e-333a). Such an idea has a precedent in the *Laws*, where it is claimed that the quickest way to bring an effective constitution to a state, and thus make its citizens happiest, would be to install an enlightened or virtuous monarch (Pl. *Leg. IV.* 710b). This thought, however, clashes with the claim referenced two paragraphs prior that only a philosopher-king could rule a state effectively, since it states

that any virtuous ruler who abides by a constitutional system — not necessarily a philosopher — could ensure the happiness of the citizens. For this reason, in his work "Plato and Practical Politics," the scholar Malcolm Schofield refers to *Epistle VII* as a "confused amalgam of [the] *Republic* and *Laws*," recognizing the contradiction between alternating endorsements of philosopher-kings, and constitutional rule (Schofield 2000, 301). Schofield considers this discrepancy as evidence for lack of complete Platonic authorship for the letter, specifically claiming that the relevant section (Plato and Dion's advice to Dionysius) was inserted sometime after the original composition of the letter by a later writer to provide more explicit political advice to the Sicilians to whom the letter was addressed (Schofield 2000, 301).

Indeed, Schofield is not the only scholar to hypothesize the possibility of textual interpolation, or the later addition of non-Platonic writing into an original draft by Plato. In his 1906 dissertation titled *Quaestionum de Septima et Octava Platonis Epistola Capita Duo*, the German scholar Maximilian Odau also supported such a possibility, claiming that *Epistle VII* is actually a re-edited version of an actual letter which Plato purportedly sent to Sicily (Harward 1928, 144). Specifically, Odau states that the original letter consisted of portions of *Epistle VII* and *Epistle VII*, and that the letter passed down as part of the canonical collection of Platonic letters was rather an early draft of this letter revised by a later editor, who removed the section from the eighth letter and added a few new sentences (Harward 1928, 144). While Odau's specific conjecture has been treated with skepticism by later scholars such as Harward, the fact that multiple individuals have hypothesized the interpolation of non-Platonic text within an otherwise authentic document illustrates the suitability of this possibility for further examination through digital means in this paper.

2.1.3 The Philosophical Digression

Aside from the question of Plato's political views, another group of philologists has focused on conflicting philosophical ideas in the digression on the Theory of Forms. The scholar Myles Burnyeat further advances the discrepancy argument formulated in previous paragraphs, when he claims not only that the philosophical digression in *Epistle VII* on the Theory of Forms conflicts with other works of Plato, but also that it contains a clumsily stated argument uncharacteristic of such a celebrated thinker (or even his students or critics) (Burnyeat and Frede 2015, 122). The argument in question proceeds as follows: the letter's author claims that individual words derive meaning through convention, stating that "none of the objects, we affirm, has any fixed name, nor is there anything to prevent forms which are now called 'round' from being called 'straight,' and the 'straight' 'round'" (Pl *Ep. VII.* 343a-b). From this premise, it is concluded that "while there are two separate things, the real essence and the quality, and the soul seeks to know not the quality but the essence, [an individual word] proffers to the soul either in word or in concrete form that which is not sought" — essentially stating that it is impossible for a word (or a definition of a word, since a definition is simply other words put together) to exhibit not only the "quality" of the concept denoted by that word, but also its "essence" itself (Pl *Ep. VII.* 343c).

Burnyeat considers this line of argumentation flimsy because it contrasts with numerous other examples within the Platonic dialogues of spirited attempts to encapsulate the "essence" of a concept within words. Examples of this phenomenon include moments such as when Socrates chides the eponymous character of the *Euthyphro* for describing attributes of "holiness" rather than the very meaning of "holiness," or when Socrates scolds Polus in the *Gorgias* for critiquing (at least, at first) the art of Gorgias rather than describing the

actual nature of the art (Pl. *Euthphr.* 10e-11a; *Grg.* 448e). Moreover, Burnyeat also notes that the author of *Epistle VII* essentially jumps directly from premise to conclusion, committing a rather unsupported non-sequitur uncharacteristic of the authentic Plato's well-reckoned arguments (Burnyeat and Frede 2015, 132). Such a line of reasoning has strong implications not only for the seventh letter's authenticity but also for its wider historical and literary significance. That is, were it the case that the letter was written not necessarily by Plato but even by one of his philosophically competent colleagues, one could more fully trust the novel historical and philosophical information put forth in the letter despite the ambiguity of direct Platonic authorship; on the flip side, should one accept Burnyeat's argument, they should no longer assume the genuineness of such reasoning or any information provided by the letter for that matter.

Furthermore, some scholars have even gone so far as to claim that not only the content but also the very language of the philosophical digression shows little correspondence with Plato's other writings. George Boas, another doubter of the seventh letter's authenticity, claims that one can find many un-Platonic phrases in the letter's discussion of the Theory of Forms – particularly, ἐπιχείρησιν ("undertaking") and τὸ ποιόν τι ("quality"), both of which he considers to be soundly Aristotelian (Boas 1948, 456). While he concedes that perhaps Plato seldom used these phrases in his own writing yet somehow handed them down to his student Aristotle, he considers this scenario unlikely given the fact that Plato specifically draws attention to rare terminology in his other works. In the *Theaetetus*, for example, Plato writes once that "ἴσως οὖν ἡ 'ποιότης' ἅμα ἀλλόκοτόν τε φαίνεται ὄνομα" ("perhaps, [the term] 'quality' seems to be an appellation of unusual nature") (Pl. *Tht.* 182a). Burnyeat echoes Boas' concern regarding τὸ ποιόν τι – stating that it is only used thrice in the Platonic

corpus, but as an interrogative and not as a nominal phrase answering that interrogative – and adds his skepticism of the expression τὸ τί, which he claims never appears in the indisputably authentic portion of the Platonic corpus (Burnyeat and Frede 2015, 128). While Boas attributes these expressions to Aristotle and Burnyeat decisively rejects such a hypothesis, they both agree that this terminology further undermines the possibility of Platonic authorship.

2.2 Stylometric Analysis

This last argument departs from most of the other claims presented so far in that it considers the style rather than content of the seventh letter. That being said, the methodology shown by Boas and Burnyeat for analyzing style is still fairly qualitative, focusing on discussion of a few crucial terms and their conflict with other texts. In contrast, there has been another group of scholars who have sought to analyze Platonic style and its implications on questions of authenticity in the Platonic corpus through more quantitative forms of analysis. While many of these scholars predate the modern advances in machine learning which inspired this paper, a discussion of the historical development of quantitative stylometry in the context of Platonic authorship should prove a useful background for my motivation in choosing the neural models used in this study.

2.2.1 Lutosławski: Early Attempts at Quantifying Style

One of the first stylometrists relevant in this tradition was the Polish intellectual Wincenty Lutosławski, who developed a quantitative method of describing Plato's style in his 1897 book *The Origin and Growth of Plato's Logic*. Although he does not tackle the particular

question of the authorship of the *Epistles*, his study is nevertheless relevant to this paper as an early example of a statistical attempt to quantify Plato's specific usage of Greek. The goal of this study was to trace the development of logic as a concept within Plato's writings, a task which relies on the presence of an effective chronology of the Platonic corpus. In order to re-assemble such a timeline, Lutosławski looks to style as a potential indicator of the particular stage of Plato's life in which a particular text might have been written. Specifically, he attempts to create groups of stylistically "similar" texts, claiming that such similarity would imply composition within more or less the same period of Plato's writing career.

Lutosławski expresses his guiding principle for creating such groups as a "law of stylistic affinity," claiming that, "of two works of the same author and of the same size, that [one] is nearer in time to a third, which shares with it the greater number of stylistic peculiarities, provided that their different importance is taken into account, and that the number of observed peculiarities is sufficient to determine the stylistic character of all the three works" (Lutosławski 1897, 152). Essentially, he states that stylistic similarity between two works can be quantified by a measure of shared textual features, and that such a "score" of similarity between one text and a reference can be benchmarked against "scores" of another text with the reference text. Proceeding from this position, Lutosławski formulated a methodology for calculating such a "score." Compiling from the works of other philologists 500 stylistic textual features ("peculiarities," as he refers to them) – ranging from the lexical (e.g. #200: "ὥσπερ less frequent than καθάπερ"), to the morphological (#317: "εἶπον prevailing over ἔλεγον") and syntactic (#391: "πέρι c. accus. prevailing over πέρι c. genitive), to even the rhetorical (#451: "rhetorical interrogations between 10 and

20% of all interrogations") and the content-based (#13: "partial prevalence of other teachers over Socrates") (Lutosławski 1897, 149–150) — he divides these features into four groups:

1. **Accidental**: "peculiarities" which appear at most once in one specific text (Lutosławski 1897, 146).

2. **Repeated**: "peculiarities" which appear twice in a "small" dialogue or up to four times in a "large dialogue," as well as certain "peculiarities" designated as more important than the previous category (Lutosławski 1897, 147).

3. **Important**: "peculiarities" which appear more than twice in a "small" dialogue or more than four times in a "large dialogue," as well as certain "peculiarities" designated as more important than the previous category (Lutosławski 1897, 148).

4. **Very important**: Certain "peculiarities" designated as more important than the previous category (Lutosławski 1897, 150).

This approach itself is not particularly unique. For example, contemporaries of Lutosławski such as Constantin Ritter and Rans Raeder claimed authenticity for the *Epistles* on the basis of correspondence in the distribution of features ranging from the minute or subconscious (particle usage) to the deliberate (rare words) (Levison, Morton, and Winspear 1968, 310). Lutosławski's innovation is his combination of the counts of each of his compiled features into a single metric serving as a quantification, or discriminant, of style. That is, for each text considered, he counts the number of each category of features, and calculates a weighted sum of these counts (the score which he refers to as the "total units of affinity") by using the following formula, known as a

discriminant function (where x_i is equivalent to the number of peculiarities present from the ith category):

$$A = x_1 + 2x_2 + 3x_3 + 4x_4$$

From this "affinity" score, Lutosławski claims that one can establish the relative stylistic, and thus chronological, similarity of various documents in the Platonic corpus. While certainly a rather crude approach — particularly on consideration of the large number of hand-counted features and the arbitrary categorization and weighting of each category — Lutosławski's methodology is significant as an early attempt at quantifying style for its attention to feature engineering and weighting different features differently. Indeed, this work would influence future scholars seeking to perform more sophisticated analyses of Plato's style.

2.2.2 Levison et al.: Frequency Analysis

Another approach to stylometry, however, relies less on linguistic or syntactic features than on analysis of the distribution of statistics such as sentence length, word frequency, and word position. Building upon earlier work by W.C. Wake, who claimed tentative authenticity of *Epistle VII* given correspondences in its distribution of sentence length with that of the *Laws*, the scholars M. Levison, A.Q. Morton, and A.D. Winspear advance this idea by comparing word frequency distributions in *Epistle VII* with the corresponding distributions from the Platonic dialogues (Levison, Morton, and Winspear 1968, 313–314). Their argument proceeds as follows. First, they analyze the distribution of

sentence lengths in *Epistle VII* and in another work by Plato, the *Apology*. Reporting that the mean value is 26.7 in the former and 17.1 in the latter, they claim that one of these works must not have been written by Plato on the grounds that sentence length distribution should be more or less constant across different works by the same author. Moreover, they respond to the potential objection that such a claim would not hold in the case of temporal or genre differences between two works by stating that the *Apology* displays genre correspondence to *Epistle VII* given that they both are long prose works with an apologetic character (the former for Socrates, the latter for Plato himself), and that the *Apology* is an atypical dialogue since it mostly consists of Socrates' self-defense in court. To determine which of the two works is inauthentic — disregarding the traditional scholarly assumptions confirming the authenticity of the *Apology* — Levison et al. compiled a list of Platonic works, and, for each work, calculated the proportion of sentences which contain the particle δε as the second or third word; in their opinion, this test is an appropriate measure of authorship given its lack of correlation with sentence length and thus resilience against genre differences between dialogue (short sentences) and prose (long sentences) (Levison, Morton, and Winspear 1968, 315-316). The relevant proportion was 33.7% for *Epistle VII*, 16.4% for the Apology, and, on average, around 15% for the remaining Platonic works examined, a result which implies the inauthenticity of *Epistle VII*. Of course, one might wonder whether this result can be replicated across many more similar features, or whether it is purely a convenient finding from the δε test alone.

Although Levison and his colleagues do not formulate as intricate a potpourri of linguistic, syntactical, and lexical features as Lutosławski, their approach to "quantifying" the authenticity problem of the *Epistles* is significant in its focus on "subconscious" stylistic

features such as sentence length and particle positions, rather than the more overt features such as rare words which have traditionally been considered marks of style and thus authorship. As referenced in the previous paragraph in the context of the "δε" test, such an approach can perhaps be considered more resilient against variation in genre, since one can assume that certain preferences such as particle frequencies are so deeply ingrained in the author's mind as to be impervious to deliberate variation of other stylistic features across different works by the same author. This notion is particularly relevant to the question of the authenticity of *Epistle VII*, given that most of the remaining works by Plato are dialogues, not letters. Furthermore, their usage of statistical methodology such as Pearson's chi-squared test shows a mathematical sophistication more refined than previous analysts. This increased rigor and emphasis on less stylistically deliberate and more low-level features would continue to appear in the work of future stylometrists.

2.2.3 Ledger: Advances in Discriminant-Based Clustering

One of these future scholars was Gerard Ledger, who unites the types of textual features seen in Levison's study with more advanced versions of the discriminant functions formulated by Lutosławski (Ledger 1989). Ledger's methodology relies at a basic level on curating low-level features of Plato's text by hand, and passing these features to a discriminant function to classify texts by time period, genre, author, or other categories. Nevertheless, unlike Lutosławski's *The Origin and Growth of Plato's Logic*, Ledger's work *Re-Counting Plato* is dedicated wholly to a computational analysis of Plato's style and its implications for questions of not only chronology but also authenticity — a topic which inevitably leads to the *Epistles*.

Ledger's methodology is as follows. Doing away with Lutosławski's burgeoning list of 500 features, Ledger selects 37 features for each text processed, divided into three categories. Although focusing on the character, rather than word, level, these features show strong resemblance to the positional word frequencies formulated by Levison et al.:

- 19 "ALET" variables, corresponding to the percentage of words in a given text containing each of the following letters: α, γ, δ, ε, η, θ, ι, κ, λ, μ, ν, ο, π, ρ, σ, τ, υ, β/ζ/ξ/φ/χ/ψ, ω (Ledger 1989, 7).
- 9 "BLET" variables, corresponding to the percentage of words in a given text ending with each of the following letters: α, ε, η, ι, ν, ο, σ, υ, ω (Ledger 1989, 8).
- 9 "CLET" variables, corresponding to the percentage of words in a given text having each of the following letters as the penultimate character: α, δ, ε, η, ι, ο, τ, υ, ω (Ledger 1989, 8).

After creating 37-dimensional representations of 1000-word long samples from each text generated using the features above, Ledger labels each sample by the work from which it originates. With all of the samples assigned to a class corresponding to a given Platonic work, he conducts multivariate discriminant analysis on these samples using the SAS Discrim algorithm (employing the Mahalanobis distance — a metric quantifying the proximity of a sample set of features to the mean of the distribution underlying that sample — as a measure of proximity between texts) to calculate the optimal discriminant function between the different clusters of samples (Ledger 1989, 35-36). One key difference from Lutosławski's methodology is that Ledger uses statistical learning techniques to find an optimal discriminant function, rather than selecting an arbitrary formula as described

above. A useful result of this procedure is data regarding the Mahalanobis distance between all of the works analyzed (generated from the average discriminant score over all of the samples from one given work), allowing comparison of the relative similarity between different works in the Platonic corpus (Ledger 1989, 146). Indeed, Ledger analyzes these scores with the question of the authenticity of the *Epistles* in mind, resulting in the following claims:

- *Epistle VII* is authentic because it is most similar to the *Laws*, and, moreover, it is the text to which the *Laws* itself shows the most similarity. This result is interesting especially in the context of the content differences between the two texts discussed earlier in this section. As Ledger claims, it is unlikely that *Epistle VII* would be inauthentic because, if that were the case, it should not have been closer in proximity to the *Laws* than other texts of certain authenticity from the same time-period. Furthermore, the Mahalanobis distance between the two texts, 3.35, is smaller than the corresponding statistic between highly connected pairs of texts such as the *Timaeus* and the *Critias* (Ledger 1989, 148-149).

- *Epistles III* and *VIII* are also authentic because they are in close proximity with *Epistle VII* (Ledger 1989, 151-152).

- *Epistles II* and *XIII* are ambiguous.

- The other *Epistles* do not contain enough text to be appropriate for such analysis (Ledger 1989, 153).

2.2.4 Koentges: (Some) Machine Learning for Feature Engineering

While certainly more rigorous than Lutosławski's, Ledger's methodology can still be expanded in a few ways. Particularly ripe for improvement is Ledger's selection of features, which are drawn exclusively from letter frequency statistics, contrasting with Lutosławski's wide assortment of lexical, morphological, syntactical, and rhetorical phenomena. Thirty years later, Thomas Koentges, in his 2020 study The *Un-Platonic Menexenus*, would reconcile Ledger's more powerful statistical methods with Lutosławski's more diverse range of textual features, recognizing that Ledger's methodology was limited by the capability of 20th-century computing technology (Koentges 2020). Although the object of this study was to provide a digital analysis of the authenticity of Plato's *Menexenus* — a research question tangential to that of this paper — a summary of Koentges' work is helpful towards understanding the current state of digital methodology for authorship attribution of Ancient Greek texts.

While Ledger relies on letter frequency statistics which were easily computable, Koentges takes advantage of modern NLP libraries such as `stylo` (Eder, Rybicki, and Kestemont 2016) to create 25 different feature sets of up to 1000 dimensions each — that is, 25 different digital representations of each textual sample. Some of these feature groups consist of frequency counts of the most commonly occurring words in the corpus; others consist of frequency counts of the most commonly occurring length four sequences of words (4-grams) (Koentges 2020, 234-236). Furthermore, Koentges uses data mining algorithms such as J48 (decision trees) and Naïve Bayes, accessed through the package WEKA, to prune some of the feature sets above, creating additional feature sets optimized for classification tasks. For each feature set, Koentges analyzes the similarity of different samples in two different ways:

1. Reducing the dimension of each feature vector from up to 1000 to 2 using t-distributed stochastic neighbor embedding (t-SNE), and then using Euclidean distance as a similarity metric to cluster the resulting two-dimensional samples into groups (Koentges 2020, 237).

2. Forgoing dimensionality reduction and using a variety of similarity metrics (e.g. Euclidean distance, Jaccard distance, cosine distance) to cluster the original, high-dimensional feature vectors into groups (Koentges 2020, 238).

As described above, Koentges' approach incorporates more modern techniques such as data mining for feature engineering and t-SNE for dimensionality reduction, constituting a refinement of the basic clustering approach taken by Ledger. However, the core of his approach is still hand-selected feature engineering — while these features are pruned using machine learning algorithms, they are ultimately rooted in various human-understandable observations about the text, such as the most frequently occurring words, or sentence length. One issue with this approach is that, in any given set of texts, there are low-level features in the data or combinations thereof which could potentially be solid indicators of style (and, hence, authorship) but undetectable by humans. Hand-crafted feature engineering will fail to recognize these underlying trends given their sheer number. Instead, one could look towards various other, more powerful classification models which can essentially automate the process of finding an optimal set of textual features; in particular, the neural approach considered later in this paper is one such solution.

This is not to say, however, that hand-selected features are necessarily inferior in modeling stylistic differences between different authors. While authors do possess certain habits in their writings which are so "subconscious" as not to be imagined by those

stylometrists who hand-craft feature vectors, there are other stylistic differences known to many generations of philologists between authors which are indeed deliberate choices. Accordingly, explicitly including these known discriminating features can better model conscious decisions that authors made regarding the style of their writing, potentially even to differentiate themselves from other authors. Furthermore, it is usually easier to interpret results of digital analysis conducted with hand-engineered features — specifically, to infer which features might serve as a strong indicator of a particular author's style, an ability not provided by machine-created vectors composed of intricate combinations of obscure and low-level textual statistics. Given these advantages, most previous digital work concerning Platonic authorship questions relies mostly on hand-selected feature vectors, and thus, to experiment with a new approach, I will focus my methodology on models which can learn optimal feature sets themselves, such as the recurrent neural networks formulated in the next chapter.

2.3 Conclusion

One salient result from the above survey of stylometric analyses of *Epistle VII* is that the majority of the scholars mentioned — indeed, all of them, except for Levison et al. — claim that their results imply Platonic authorship of the letter, a trend particularly noteworthy in the context of an increasing consensus among philologists and philosophers *against* that very conclusion, as described in the previous section of this paper. A possible explanation for this phenomenon has to do with the nature of the arguments traditionally levied against the letter's authenticity. With the exception of a few points of contention regarding specific philosophical terminology, most of the arguments rely on attacking the content, rather

than the style, of the letter as dubious. One aspect of content-based arguments is that they are notably a one-edged sword — that is, to re-phrase Edelstein's opinion presented earlier in this paper, discords in content are a much stronger indication *against* authenticity than similarities in content are an indication *for* authenticity. Therefore, it is reasonable that the majority of scholars who compare the letter's content with other works by Plato find that the letter is inauthentic. That is, proponents of its Platonic authorship would have needed to pursue other methods to verify their arguments given the inefficacy of content similarity as a mark of authenticity.

Examinations based on the letter's style, however, are subject to the opposite handicap — as shown by the work on the Federalist Papers, stylistic agreement can be seen as decisive evidence towards Platonic authorship, while, given the genre difference between *Epistle VII* and the Platonic dialogues, a difference in style cannot be attributed solely to different authorship. Essentially, it would be much more difficult to forge an author's subconscious language habits than to ensure concordance with that author's other writings and the general historical record. Thus, content-based examinations should naturally tend to skew towards claiming inauthenticity, while style-based arguments should more often lean towards attempting to show authenticity. Examined in the context of this paper, this heuristic suggests that the machine-learning approach taken in later sections should only provide decisive evidence *in favor of* authenticity. Should my models show strong stylistic dissimilarities with the remainder of Plato's corpus, this discrepancy would be significant yet nevertheless still inconclusive because of the possibility of other reasons for this discrepancy such as genre difference between letters and dialogues. Such results would not stand on their own and need to be supported by other evidence, perhaps from the

traditional scholarship presented in this chapter concerning the authenticity of the letter. Simply put, there are no letters known to have been written by Plato against which *Epistle VII* can be benchmarked, yet the accuracy of the predictions of supervised models such as those presented in this paper depends on the similarity of the training data to the data to be used for prediction.

3

Analyzing the Seventh Letter through Neural Methods

As mentioned in the last chapter, most stylometric analysis of the Plato's Epistles has hitherto been conducted using methods which rely on hand-gathered features. Thus, I will explore a different class of models — recurrent neural networks — which can essentially

learn optimal feature representations themselves given their large number of trainable parameters corresponding to different combinations of lexical features. Indeed, it is these models which are behind many of the rapid developments in text-based artificial intelligence over the last ten years. Given the dearth of prior usage of neural models for digital analysis of Ancient Greek, however, it is necessary before examining the Epistles to settle on some best practices for such applications, with the morphological richness of the Greek language vis-à-vis English specifically in mind. Thus, in this chapter, I will evaluate a few different variants of the same basic long-short-term memory (LSTM) model to classify chunks of Ancient Greek prose as belonging to a certain author, and then use the best of these variants to analyze *Epistle VII* alongside other allegedly spurious Platonic texts.

3.1 Developing a Neural Authorship Classifier for Ancient Greek

3.1.1 Challenges of Greek Morphology for NLP

Ancient Greek is a language with a rich inflected morphology. While languages such as English derive much syntactic meaning from word order, Ancient Greek has the liberty to employ a more flexible word order because the function of most words can be determined quite precisely by their morphological form. Specifically, Ancient Greek nouns, adjectives, and participles take on different forms for number, case, and gender, while Ancient Greek verbs conjugate over number, person, voice, tense, and mood. This expression of many grammatical attributes in morphology creates dozens, if not hundreds, of inflected forms for many words in the Greek language. Figure 3.1 demonstrates some basic examples of Greek morphemes.

The existence of several distinct forms for any given semantic root can add complications

(1) ὁ Σωκράτης ὁρᾷ.
the.N.MASC.SG Socrates.N.SG see.3.SG.PRS.IND.ACT
'Socrates sees.'

(2) τὸν Σωκράτην ὁρῶ.
the.ACC.MASC.SG Socrates.ACC.SG see.1.SG.PRS.IND.ACT
'I see Socrates.'

(3) οἱ φίλοι τοῦ Σωκράτους ὁρῶσιν.
the.N.MASC.PL friend.N.PL the.GEN.MASC.SG Socrates.GEN.SG see.3.PL.PRS.IND.ACT
'Socrates' friends see.'

Figure 3.1: Examples of Greek morphology

to NLP models that treat each word as a distinct type. For example, many models would consider the first two sentences in Figure 3.1 completely different on the grounds that they share none of the same words, despite the fact that, semantically, their overarching meanings and connotations are similar. Given this inability to establish connections between different forms of the same word and different words with the same morphology — that is, an inability to generalize — such models risk losing valuable information about the texts upon which they are operating.

For this paper, the specific application in question is that of authorship identification, in which an NLP model predicts the most likely candidate for authorship of a given text from a predetermined set of authors. Traditionally, these approaches have relied on examining lexical details, with models attempting to learn which particular words or sequences thereof (from obscure, low frequency words to highly common words such as particles and adverbs) are most revealing of an author's style. However, morphological distinctions could

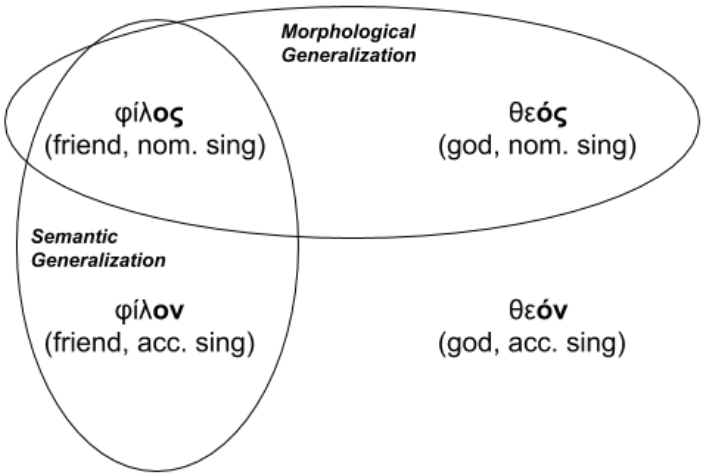


Figure 3.2: Baseline word representation offers neither semantic generalization (grouping together different morphological forms of the same lexical root) nor morphological generalization (recognizing similarity between different roots exhibiting the same morphology). One could expect that an optimal model would improve in one or both of these respects.

also be seen as a marker of a given author's style, since the morphological forms most frequently employed by such author could indicate a particular dialect or timeframe of the Greek language, an adherence to certain alternate forms, or a preference for certain grammatical constructions — all clues that can distinguish one author's writing from another's. Therefore, one might postulate that authorship classifier models with the ability to account for variations in morphology as well as lexicon would prove to be more accurate; thus, different approaches to encoding Ancient Greek morphology are evaluated in the context of an authorship classification task.

3.1.2 Methods Evaluated

Plaintext

The baseline plaintext model will take as input the words of the original text with no alterations. With regard to morphology, this method ensures the separate treatment of different morphological forms of a given word, at the expense of the model's ability to generalize semantically across these different forms of the same ultimate lexeme, or to generalize morphologically across different lemmata with the same morphology.

Lemmatization

The most basic response to the trade-off of the previous approach is lemmatization, or reducing each word to its lemma or underlying dictionary form. In other words, nouns are reduced to the nominative singular, verbs are reduced to the first person singular present active indicative, and adjectives are reduced to the nominative singular masculine. Thus, the sentence Ἐν δὲ τούτῳ προσάγουσι τῷ Κύρῳ τοὺς αἰχμαλώτους δεδεμένους, τοὺς δέ τινας καὶ τετρωμένου becomes ἐν δε ούτος προσαγω ὁ κυρος ὁ αἰχμαλωτος δεω1[1] ὁ δε τις και τιτζωσκω. (stripping words of their accents, as in this experiment — see "Text Preprocessing" below). An English analogue would be the reduction of the sentence "Plato's works are seen as foundational to Western philosophy's development" to "Plato work be see as foundational to Western philosophy development." Greek lemmatization can be performed using a pre-trained lemmatizer from the open-source Classical Language Toolkit (CLTK) (Johnson

1. The "1" in the lemma distinguishes δέω (I bind, tie) from δέω (I lack, need), which have identical dictionary forms.

et al. 2014–2020).[2]

(4) ὁ Σωκράτης τὸν Πλάτονα ὁρᾷ.
the.N.MASC.SG Socrates.N.SG the.ACC.MASC.SG Plato.ACC.SG see.3.SG.PRS.IND.ACT
'Socrates sees Plato.'

(5) τὸν Σωκράτη ὁ Πλάτων ὁρᾷ
the.ACC.MASC.SG Socrates.ACC.SG the.N.MASC.SG Plato.N.SG see.3.SG.PRS.IND.ACT
'It is Socrates whom Plato sees.'

Figure 3.3: The loss of semantic difference through lemmatization.

Given the large number of inflected nominal and verbal forms in the Greek language, lemmatization allows a model to recognize multiple inflected forms as the same word in the vocabulary set. This reduction also has the side effect of significantly reducing the vocabulary size of the dataset under consideration. Nevertheless, the trade-off is that lemmatization removes all morphological information from a word. Consider the sentences in Figure 3.3. Given Ancient Greek's flexible word order, the meanings of these two sentences derive different meaning only through their morphology. The result of lemmatization on both of these two sentences would be ὁ Σωκράτης ὁ Πλάτων ὁρῶ, leveling the semantic difference between the two sentences. Intuitively, lemmatization can thus be seen as prioritizing semantic generalization between words at the expense of morphological information.

2. This particular lemmatizer operates context-free, meaning that its assignment of a particular lemma to any input is not conditional on the context of the sentence in which that word appears. In general, such lemmatizers are innately imperfect in their inability to choose the one most correct lemma for words with multiple morphological interpretations, as they will return all possible correct lemmata for ambiguous forms.

Byte Pair Encoding

One approach that might ameliorate such a limited treatment of morphological generalization is byte pair encoding (BPE), a data compression technique developed by Philip Gage in 1994 (Gage 1994). This technique has recently found use in many NLP applications (Sennrich, Haddow, and Birch 2016). The underlying algorithm, is to break up each example at the character level (appending a special word-break token '</w>', at the end of every word), and to replace the most frequent pair of characters that appear adjacent to each other with a combined character, for a pre-specified number of iterations. A simple example of the algorithm can be seen in Figure 3.4.

Iteration	Sentence	Combination
1	b i g g e r </w> i s </w> b e t t e r </w>, b o l d e r </w> i s </w> b e t t e r </w>	
2	b i g g er </w> i s </w> b e t t er </w>, b o l d er </w> i s </w> b e t t er </w>	e, r
3	b i g g er</w> i s </w> b e t t er</w>, b o l d er</w> i s </w> b e t t er</w>	er, </w>
4	b i g g er</w> is </w> b e t t er</w>, b o l d er</w> is </w> b e t t er</w>	i, s

Figure 3.4: Initial iterations of byte-pair encoding on an English sentence.

Byte pair encoding is useful because it allows the model to recognize frequently occurring subword units. Intuitively, the most frequent subword units should be morphemes that accomplish common grammatical functions — this conjecture can be seen above in the few steps of the encoding in Figure 3.4, which learns the word ending "er</w>" (comparative adjective marker) and the common verb form "is</w>". Extending

this idea to Greek, the byte pair encoding used in this experiment learns within its first 50 iterations the common morphemes seen in Figure 3.5.

Morpheme	Function	Iteration Found
αι</w>	Nominative Plural Feminine Suffix Nominative Plural Feminine Article	1
το</w>	Nominative/Accusative Neuter Singular Article 3rd Singular Medio-passive Verbal Suffix	2
ων</w>	Genitive Plural Suffix Nominative Singular Masculine Suffix	3
και</w>	"and"	5
ιν</w>	Accusative Singular Feminine Suffix	32
ειν</w>	Active Infinitive Suffix	33

Figure 3.5: Byte-pair encoding as method of identifying common Ancient Greek morphemes.

Thus, given its ability to discover common morphemes — particularly inflectional suffixes — and treat them as separate tokens from their stems, one might hypothesize that byte pair encoding possesses a greater ability to generalize morphologically across different root words with the same prefix or suffix than the other representation methods discussed above. Nevertheless, this advantage does not necessarily come at the expense of semantic generalization. Especially if allowed more iterations, byte pair encoding should also discover common word stems, effectively resulting in a version of the text that splits each word into its base stem plus inflectional suffixes. Thus, although primarily able to generalize morphologically across words, byte pair encoding also enables a degree of semantic generalization across different inflectional forms of the same word (albeit less so than lemmatization, given that only the most frequently-occuring word roots will be learned by

the encoding and that some Ancient Greek words have multiple stems[3]).

Lemmatization-BPE Ensemble

It has been hypothesized above that lemmatization and BPE can achieve semantic and morphological generalization, respectively. The seemingly orthogonal nature of these respective improvements over the baseline plaintext model, however, raises the possibility that both of these methods can be combined to achieve both semantic and morphological generalization. Therefore, as a final model, I propose an *ensemble method*[4] combining the probability outputs generated by the lemmatization and BPE models individually. That is, for a given unit of text, both of those models output a probability distribution over all of the authors considered, and label a text with the author whose probability is the highest. To create an ensemble method, one can sum the corresponding probabilities for each author and take the author whose *aggregated* probability is the highest, thus incorporating both the lemmatization and the BPE into the final prediction. This method should prove more robust and thus potentially more accurate than lemmatization or BPE individually, since it can reap the benefits of both semantic and morphological generalization.

3.1.3 Experiment Setup

Task and Corpus

To evaluate the impact of these methods, I processed the same text corpus in the different ways defined above and used each processed dataset as input to a sequence-based, neural

3. e.g. φέρω (I carry), οἴσω (I will carry), ἤνεγκα (I carried).

4. An ensemble method aggregates the predictions of multiple individual and slightly different models to calculate a more robust output.

classification model, given that the ultimate goal of the experiment is to evaluate these representations in the context of authorship classification. In addition, I created one ensemble model which combined the individual lemmatization and BPE models. The specific classification task at hand is to assign chunks of text to one of the following seven authors, who all lived during the 5th to 4th centuries B.C.E. and wrote prose in the Attic Greek dialect typical of Athens. (The choice of authors was inspired by Ledger, who used the same seven in his digital analysis of authenticity in the Platonic corpus):

1. **Plato** (428 – 348 B.C.E), one of the pre-eminent Athenian philosophers, whose writings, other than the *Epistles*, consisted nearly entirely of dialogues portraying philosophic conversations and featuring his teacher Socrates or other interlocutors.

2. **Xenophon** (430 – 354 B.C.E.), a historian well-known for his historical works but who also wrote philosophical dialogues and treatises as well.

3. **Thucydides** (460 – 400 B.C.E.), a historian who wrote the famous *History of the Peloponnesian War* seen as a text foundational to modern political science.

4. **Demosthenes** (384 – 322 B.C.E.), one of the Attic Orators, who is considered most famous for his speeches in opposition to Philip II of Macedon.

5. **Lysias** (445 – 380 B.C.E.), another of the Attic Orators.

6. **Isocrates** (436 – 338 B.C.E.), another of the Attic Orators.

7. **Aeschines** (389 – 314 B.C.E.), another of the Attic Orators.

A complete list of the works included can be found in Appendix A. The text of these seven authors — drawn from Brennan Nicholson's cleaned version of the Perseus Digital Library

Canonical Greek Literature corpus (Nicholson 2020) — has been divided into chunks of roughly 100 words and split into three subsets: a training set (on which the model learns its parameters), a validation set (on which a model's parameters can be tuned and optimized), and a test set (on which the model is evaluated), using an 80-10-10 split. This process ensured the proportionate representation of units from all works in each subset. In addition, the potentially spurious works of Plato (the Epistles and the ten dubious dialogues), have been held out for later analysis.

Dataset	Pl.	Xen.	Dem.	Thuc.	Isoc.	Lys.	Aeschin.
Training	3573	2222	2022	972	777	396	299
Validation	447	278	252	121	97	50	38
Test	447	278	253	122	97	49	37
Spurious Platonic	496	0	0	0	0	0	0
Seventh Epistle	73	0	0	0	0	0	0

Figure 3.6: Distribution of authors across datasets (100-word units).

Text Pre-processing

During the pre-processing stage, each document is separated into groups of roughly 100 words, formed by concatenating subsequent sentences in the same work until the combined length of the concatenation exceeds 100 words, at which point a new group is created. Indeed, one advantage of RNNs is their ability to handle input sequences of varying length.

From these remaining sentences, punctuation and accentuation marks are removed, given that these notations are not original components of Ancient Greek text and were added later by textual commentators. As an exception to this rule, two sets of diacritics are kept: aspiration marks (ἑ, ἐ) and iota subscripts (ῳ, ᾳ), because these diacritics can

meaningfully distinguish between pairs of different morphological forms (e.g. θεά – goddess, nom. sg. vs. θεᾷ – goddess, dat. sg.). Furthermore, although a distinction between capital and lowercase letters did not exist at the time these works were penned, capitalization is kept in place to distinguish proper nouns, since, unlike English, Ancient Greek orthography does not capitalize words at the beginning of each sentence. Implementations for many of these text cleaning functionalities were provided by the CLTK (Johnson et al. 2014–2020).

One last phrase of preprocessing involved identifying and replacing rare words — that is, topical vocabulary or proper nouns that might serve as obvious giveaways for particular authorship. The intention was to force the models to distinguish between authors on the basis of their morphological and and lexical habits, rather than on the various characters and topics about which the seven authors wrote, since I assume that the former is a stronger and more robust indicator of the subconscious "style" of an author than the latter. For each of the three preprocessing methods discussed above, I calculated the term frequency-inverse document frequency (tf-idf) score of each word in the combined texts for each author, a metric which represents the word's importance to the given author's combined text in the training set.[5] In computing these scores, the tf-idf vectorizer was provided a list of frequently occurring Greek stop words (provided by the CLTK) to ignore, as well as a list of words that appear in more than one author's works (i.e., it retained features having a document frequency score (df) below 1/7); this modification prioritized the identification and subsequent removal of words strongly characteristic of particular authors, as opposed

5. The tf-idf score seeks to measure how important a given word is to a certain document - in this case, the complete text in the training set of one given author. It is calculated by dividing the term frequency (number of times the word appears in the document) by the document frequency (proportion of documents in which the word appears), thus reflecting both the word's frequency in the document and its rarity in other documents.

to words appearing frequently in the text of *all* authors.

Accordingly, any word found to have a tf-idf score above a certain threshold — 0.025 for the original text, 0.05 for the lemmatized text (a greater threshold due to the smaller number of features in a lemmatized text and thus greater tf-idf values for each feature) — was replaced in the text with the token `<imp>` for 'important.' These threshold values were selected by manual inspection of the words filtered out at such a threshold to confirm that they were primarily topical words or proper nouns. Since the byte pair encoding algorithm ran for only 2500 iterations, it had limited ability to "discover" rarer words and there were no character sequences which appeared solely in one author's text.

Model Architecture

The classification architecture of choice is a bidirectional LSTM (Hochreiter and Schmidhuber 1997).[6] After preprocessing as defined above, each word in the input sequence is passed through a trainable embedding layer before entering the LSTM. The hidden state outputs for both directions of the LSTM are concatenated and then fed through a fully connected layer, employing dropout with a probability of 0.2 during training for regularization purposes (Srivastava et al. 2014). After the hidden layer, a softmax activation function generates a probability distribution across the seven authors mentioned above. The model employs a cross-entropy loss function which weights examples of different classes inversely to their share of all examples, in order to combat the

6. An LSTM is a variant of a recurrent neural network (RNN), a model which operates on a sequence of inputs — in this case, words (or character sequences discovered by BPE). It is initialized with a hidden state vector and updates that hidden state for each input in the sequence. The final hidden state is mapped to a seven-dimensional output vector, with each element of that output corresponding to the probability that the input sequence belongs to one of the seven authors examined. The parameters specifying the transformations are continually optimized to produce the most accurate classifications on the training set visible to the model.

Embedding Size	Hidden Layer Size	Batch Size	Epochs	Learning Rate
512	128	32	30	0.0001

Figure 3.7: Model hyperparameters

strong class imbalance between different authors in the dataset. Training is conducted using an Adam optimizer for a fixed number of epochs, using checkpointing based on mean validation loss (Kingma and Ba 2015). Figure 3.7 contains the hand-tuned hyperparameters.

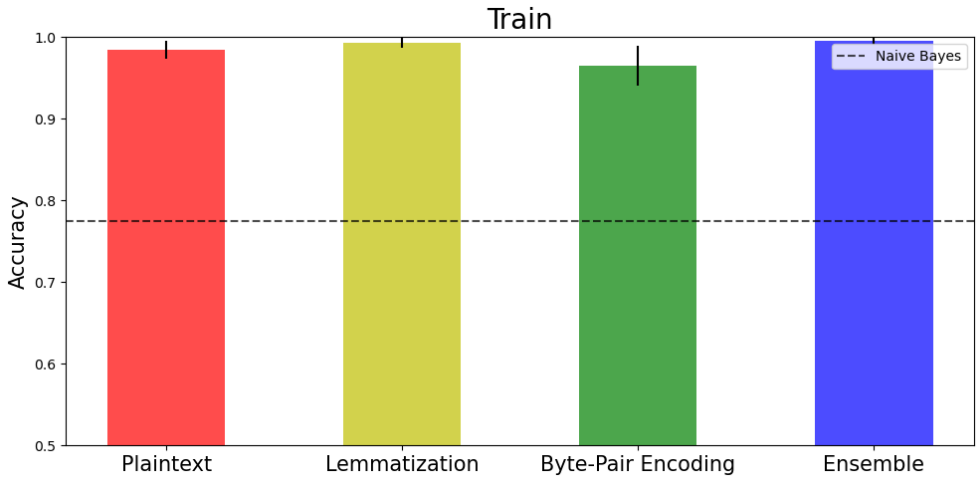
3.1.4 Results

Due to the randomness of weight initialization and dropout regularization, each model was trained and evaluated ten times. Accordingly, the mean and standard deviation accuracies across all trials are presented below. For non-plaintext models (lemmatization, byte-pair encoding, lemmatization-BPE ensemble), I evaluated the statistical significance of the change in accuracy using a two-tailed Welch's t-test to evaluate the null hypothesis that the given form of preprocessing does not impact the accuracy of the authorship classifier relative to the plaintext version.[7] Finally, as a baseline against which to benchmark the general efficacy of neural methods in this context, I have included the results of a Naive Bayes classifier,[8] which takes as input a tf-idf vector for each input unit of the plaintext.

7. Tests such as the t-test seek to measure the statistical significance of a certain claim as expressed through the data. In this case, I seek to prove that lemmatization/byte-pair encoding have a noticeable impact on model accuracy vis-à-vis plaintext — in other words, to disprove the *null hypothesis* stating the contrary. The p-value corresponds to the probability of this null hypothesis given the data measured (in other words, that any differences can be attributed to noise rather than a different underlying data distribution); thus, a small p-value ($< .05$) serves as stronger proof of the refutation of the null hypothesis.

8. Naive Bayes classifiers are deterministic models which estimate the probability of particular authorship given the text, from the probabilities of the individual words in a text given that author. It is through models from this family that Mosteller and Wallace (1963) determined the authorship of the *Federalist Papers*.

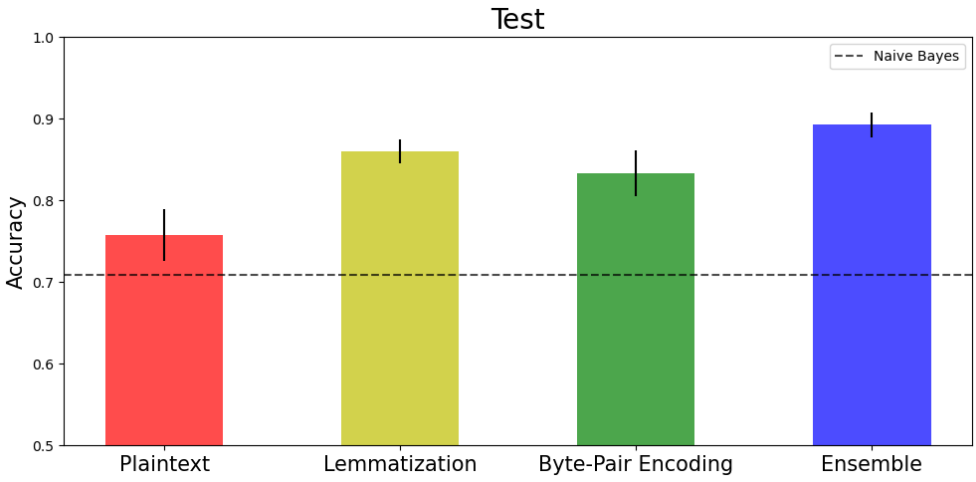
Training Set



Model Type	Accuracy Mean	Accuracy Standard Deviation
Naive Bayes	0.7744	
Plaintext	0.9844	0.0112
Lemmatization	0.9933	0.0062
Byte-Pair Encoding	0.9652	0.0244
Ensemble	0.9952	0.0048

Model Type	t-statistic	p-value
Lemmatization	2.0866	0.0557
Byte-Pair Encoding	−2.1426	0.0523
Ensemble	2.6847	0.0199

Test Set



Model Type	Accuracy Mean	Accuracy Standard Deviation
Naive Bayes	0.7093	
Plaintext	0.7580	0.0319
Lemmatization	0.8602	0.0150
Byte-Pair Encoding	0.8329	0.0281
Ensemble	0.8928	0.0150

Model Type	t-statistic	p-value
Lemmatization	8.6901	1.0007×10^{-6}
Byte-Pair Encoding	5.2858	5.3004×10^{-5}
Ensemble	11.5711	5.0931×10^{-8}

Class Breakdown (Lemmatization-BPE Ensemble)

	Pl.	Xen.	Dem.	Thuc.	Isoc.	Lys.	Aeschin.
Pl.	430	11	0	2	0	4	0
Xen.	21	250	3	1	0	2	1
Dem.	3	4	221	0	3	11	11
Thuc.	0	4	0	118	0	0	0
Isoc.	1	0	5	1	86	2	2
Lys.	1	3	13	0	1	30	1
Aeschin.	1	2	14	0	0	2	18

Figure 3.8: Confusion matrix for one of the lemmatization-BPE ensemble models. Each cell denotes the number of 100-word text units belonging to the author of its row and predicted to be the author of its column. Thus, any texts not on the diagonal are misclassifications.

Author	**Mean Test Accuracy**	**Test Accuracy SD**
Pl.	0.9398	0.0190
Xen.	0.9058	0.0145
Dem.	0.8881	0.0432
Thuc.	0.9533	0.0238
Isoc.	0.9041	0.0288
Lys.	0.5694	0.0573
Aeschin.	0.4595	0.0793

3.1.5 Analysis

First, one can observe that all of the four model variants presented above outperform the baseline plaintext Naive Bayes classifier. In particular, the plaintext LSTM (75.80%) surpasses the corresponding plaintext Naive Bayes model (70.93%) in test accuracy by nearly 5 percentage points, illustrating the advantages gained by using a neural approach to this problem. This result can potentially be explained by the fact that the Naive Bayes model does not account for word order, a significant disadvantage given the fact that preferring or avoiding certain specific sequences of words can be seen as a mark of style and thus authorship. Furthermore, Naive Bayes models tend to perform sub-optimally on imbalanced datasets, unlike the neural approaches here which counteract this problem through weighting the different classes.

The major conclusion to be drawn from the data above, however, is that the representations which allow for greater generalization across words — whether semantically (lemmatization), or morphologically (byte-pair encoding) — tend to produce models with a statistically significant increase in test accuracy, reflecting a greater ability to classify unseen text. Specifically, these two models have an average test accuracy between 7 and 13 percentage points greater than the baseline plaintext model, which does not enable such generalization. This result is statistically significant given the fact that the associated p-values are well below 0.05. Such a result is concordant with the initial hypothesis that treatment of morphological features in a text will increase the performance of authorship classifiers.

Furthermore, one should also note the differences between the non-plaintext models: the lemmatization model outperforms the BPE model by nearly 2.5 percentage points in

average test accuracy, and the lemmatization-BPE ensemble outperforms the lemmatization by another 3 percentage points on test accuracy. The first result implies that semantic generalization is more important than morphological generalization for authorship classification in Ancient Greek texts — perhaps because, as mentioned earlier in this paper, that BPE fails to generalize for words which possess multiple stems, a phenomenon which appears in many of the most frequent verbs in Ancient Greek. Nevertheless, the second result implies that morphological generalization is still important to the accuracy of the classifier as well. Combining BPE with lemmatization as part of the ensemble method results in an additional increase in accuracy versus lemmatization alone — implying that this increase in accuracy is due to the morphological generalization of the BPE model absent from the lemmatization model. In other words, the ensemble classifier is indeed able to take advantage of the more or less orthogonal improvements in semantic and morphological generalization offered by lemmatization and BPE, respectively.

Finally, having established that the lemmatization-BPE ensemble model is the most accurate and thus promising of the classifiers analyzed, one can also see that it not only performs at high accuracy in general but also on each of the authors individually. The confusion matrix in figure 3.8 illustrates the number of correctly and incorrectly classified units in the test corpus for each author, along with the incorrect author label for any misclassified units. The strong signal along the diagonal, and the class accuracy statistics indicated below the confusion matrix, indicate that, across the seven different authors examined — particularly Plato (at nearly 94% mean class accuracy on the test set) — the model performs at a high accuracy (with the exception of perhaps Aeschines and Lysias given their relative infrequency in the corpus). It is an interesting fact to note that Plato is

most frequently misclassified as Xenophon, perhaps because they both wrote Socratic dialogues, and that Lysias and Aeschines are confused with Demosthenes, perhaps because all three wrote orations.

3.2 Examining *Epistle VII*

The above section demonstrates that the lemmatization-BPE ensemble has the most potential for accuracy amongst all the different models presented for Ancient Greek authorship classification. Therefore, in this section, I will use the 10 ensemble models produced above to conduct analysis on *Epistle VII* at a section-by-section level. In addition, for comparison, I will also apply these models to two other allegedly spurious texts in the Platonic corpus, the *Epinomis* and *Menexenus*. This process simply entailed running the ten models on all of the sections from these texts (collecting both the predicted labels and the probability assigned to Platonic authorship), and, in some cases, aggregating the results across those ten models.

3.2.1 Other Spurious Platonic Texts

As a control on the forthcoming analysis of *Epistle VII*, it is helpful to gain a measure of the model's performance on other allegedly spurious Platonic texts. The two texts selected are the *Epinomis* and *Menexenus*, the authenticity of which scholars have doubted for distinct reasons.[9] Although one could devote an entire thesis to the corresponding authenticity question of either of these two texts, the goal here is to quickly present the two works and to

9. Unfortunately, it was not methodologically appropriate to conduct a similar analysis on a reliably authentic Platonic text given that all such texts were used as training data for my models.

confirm my model's concordance with previous classical scholarship and digital studies regarding them.

The *Epinomis*

In the case of the *Epinomis*, a text presented as an appendix to the *Laws*, editors such as Cooper have attacked its authenticity not on stylistic grounds but on content-based grounds, claiming that its arguments are inconsistent with the *Laws* (Cooper in Plato 1997, 1617). In fact, Ledger claims in his digital analysis of Plato's writings that the *Epinomis* seems consistent with Platonic style (Ledger 1989, 150). My models further support these two claims, indicating that the *Epinomis* largely seems to be Platonic on the basis of style, with the mean probability of Platonic authorship hovering around 0.9 throughout the letter — an expected result which bolsters claims regarding the efficacy of my approach.

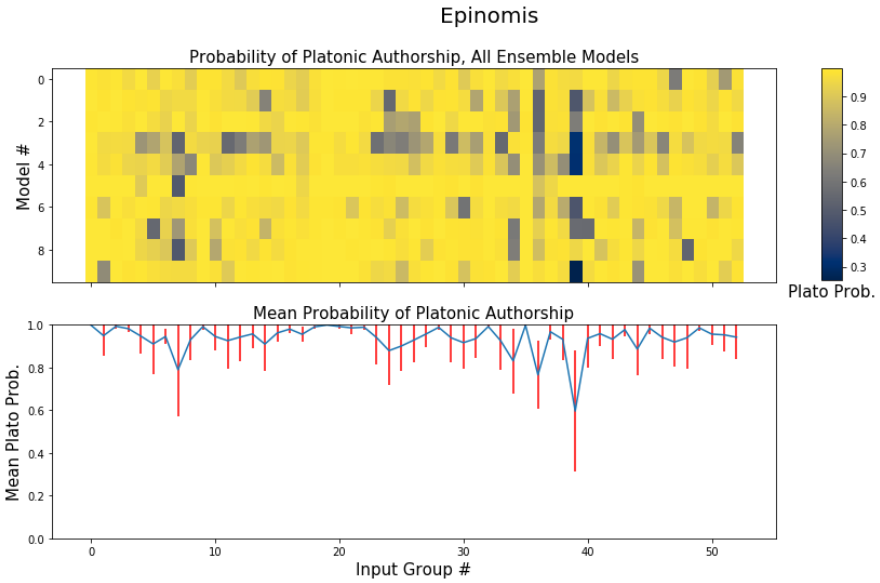


Figure 3.9: Assigning probabilities of Platonic authorship to the groups of Plato's *Epinomis*.

The *Menexenus*

The next work examined is the *Menexenus*, a dialogue which nevertheless takes the form of a funerary oration around a third of the way into the text. Given this rather unconventional form for one of Plato's dialogues, this text has also been considered spuriously Platonic. As such, this document was the focus of Koentges' study (2020) mentioned in the previous chapter, which claimed that the document was stylistically inconsistent with the rest of Plato. Again, the results of my digital analysis is concordant with both traditional scholarship as well as Koentges' claims. The lemmatized-BPE ensemble is successfully able to demarcate the portion of the work corresponding to the funerary oration and recognize its tenuous stylistic connection to Plato's known writings, with the mean probability of Platonic authorship falling from the 0.8–1 range to < 0.2 after this textual boundary.

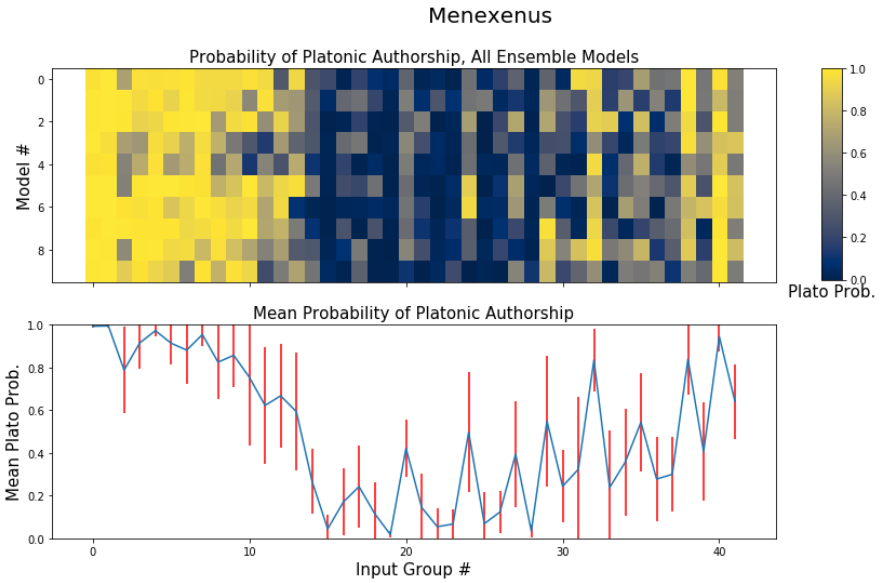


Figure 3.10: Assigning probabilities of Platonic authorship to the groups of Plato's *Menexenus*.

Both of these results — namely, that the *Epinomis* shows some stylistic similarities with Platonic texts, and that the portions of the *Menexenus* containing the funeral oration show dissimilarities with Platonic style — agree with other philological and digital analysis conducted on these two works. Thus, this concordance bolsters the credibility of my neural approach. Specifically, my models are successfully able to classify groups of text both traditionally considered as Platonic and conventionally doubted as spurious. Furthermore, it is promising that the models can successfully identify meaningful breaks in the text such as the transition in the Menexenus from dialogue to oratory form. Thus, having established some level of reliability in the results of my approach, I now turn to *Epistle VII*.

3.2.2 *Epistle VII* Results

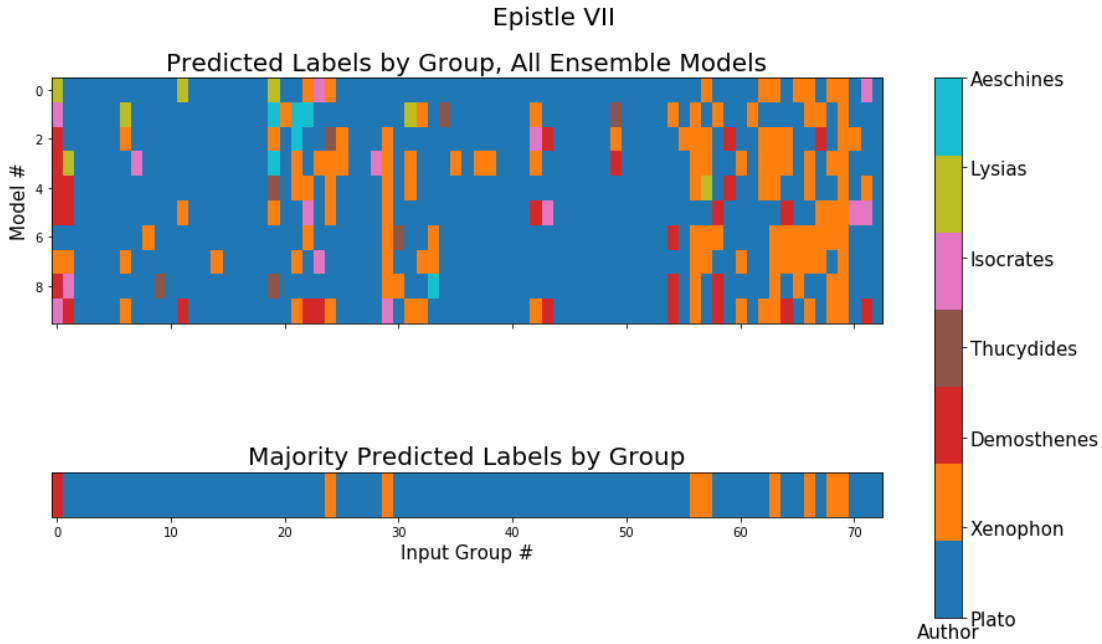


Figure 3.11: Labeling the groups of Plato's *Epistle VII*.

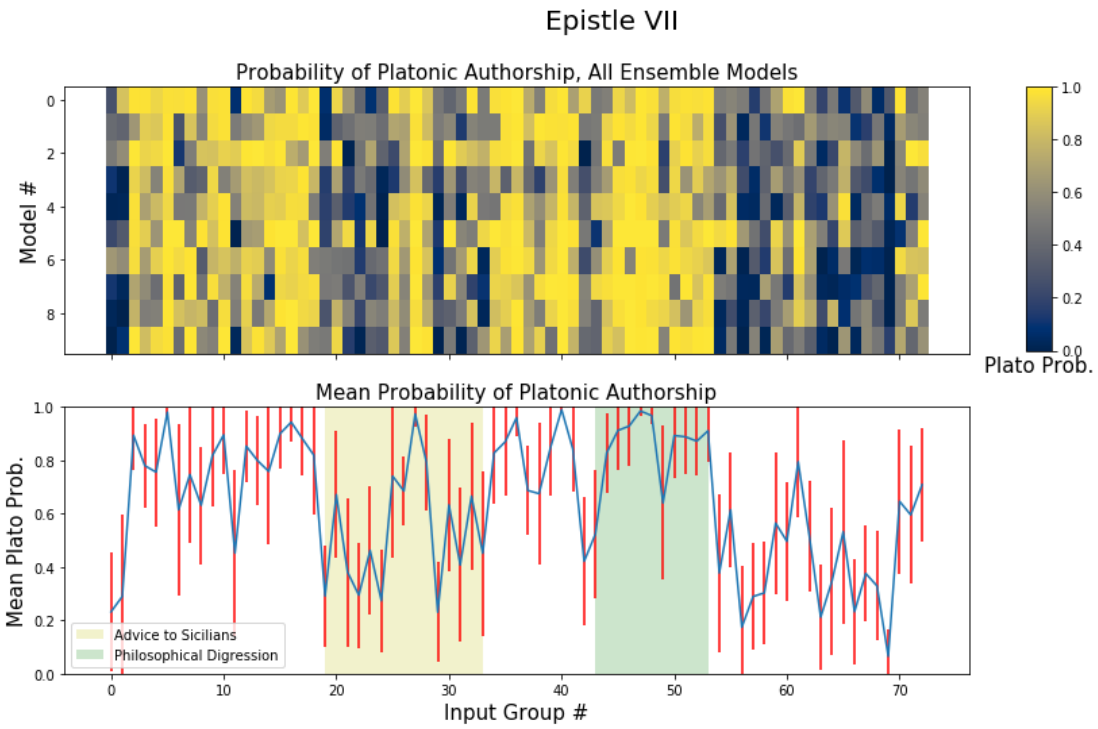


Figure 3.12: Assigning probabilities of Platonic authorship to the groups of Plato's *Epistle VII*.

3.2.3 Analysis

The charts above underscore the importance of considering each section of *Epistle VII* individually. Such granular analysis of the different parts of the letter reveals that some sections are consistently considered by the different models to be much more indicative of Platonic authorship than others. Specifically, a glance at the graphs above indicates that the letter can be demarcated into distinct segments which all display different patterns of predicted authorship:

1. **Groups 1–18** (corresponding to sections 323d–331e), with the exception of the opening to the letter, are labeled by most of the models as being highly indicative of Platonic authorship. This section is mostly narrative, recounting Plato's first and second visits to Sicily. **LIKELY AUTHENTIC**

2. **Groups 19–33** (corresponding to sections 332a–337c) consist of two segments of text labeled as dubiously Platonic (19–24 and 29–33 = 332a–334b, 335e–337c) surrounding an isolated segment of text seen as highly Platonic (25–28 = 334c–335d). This section corresponds closely to the portion of the text consisting of Plato's political advice to the supporters of Dion examined in the first chapter of this paper. **LIKELY SPURIOUS**

3. **Groups 33–41** (337d–340e)and 43–53 (341c–345b) are considered to be largely consistent with Platonic style, especially the latter, although they are separated by group 42 (341a–b), which is seen by the models as less likely to have been written by Plato. This segment corresponds to Plato's account of his efforts to educate Dionysius II, and groups 43–53 match up almost exactly with the extent of the philosophical digression on the Theory of Forms. **LIKELY AUTHENTIC**

4. **Groups 54 onward** (345c–352a) exhibit a much weaker link to Plato than the rest of the letter. However, the final portion of this group (70–72 = 350d–352a) exhibits relatively less discrepancy with Platonic style than the rest of this portion of the letter. In terms of content, this segment describes Plato's ignominious third visit to Sicily and his subsequent interactions with Dion as the latter conspires to overthrow Dionysius II. **LIKELY SPURIOUS**

The fact that the models were successfully able to segment the text of *Epistle VII* in a manner corresponding to traditional divisions of its content provides further evidence of its ability to successfully learn meaningful patterns in Ancient Greek texts. It is indeed remarkable that the model nearly exactly identifies peculiar sections of the text such as the political advice to the Sicilians and the philosophical digression which have been analyzed by traditional text-based scholarship in the context of the letter's authenticity. As such, I will briefly discuss both of these sections in greater detail to further connect the results of this computational analysis with the pre-existing scholarship on the letter discussed in the first chapter of this paper.

Advice to Sicilians

As mentioned above, groups 19-33 (332a–337c) reveal an interesting pattern: two segments of groups considered relatively un-Platonic (19–24 and 29–33 = 332a–334b, 335e–337c) which surround an isolated portion with high likelihood of Platonic authorship (25–28 = 334c-335d). Specifically, for the two blocks seen as spurious, the mean probability of Platonic authorship frequently dips below 0.5, with some of the models predicting a probability of less than 0.2. This result is notable in the context of the arguments (summarized in the first chapter of this paper) made by Frede and Schofield concerning the dubiousness of Platonic authorship for this portion of the letter, given that it makes claims inconsistent with other Platonic writings such as the *Laws*. In particular, Schofield has singled out views in this very section as inconsistent with other political opinions expressed in the letter, from which he proceeds to argue that this section was a later interpolation by a non-Platonic author (Schofield 2000, 301). The fact that the style of this portion is

considered by the models to be inconsistent with other Platonic examples provides further support for Schofield's content-based arguments for the spuriousness of this section.

However, one must not ignore the portion in the middle of this segment (25–28 = 334c–335d) which is considered by the models to be highly Platonic (with the mean probability of Platonic authorship exceeding 0.8 and exhibiting a comparatively low standard deviation). This specific section, unlike many other parts of Plato's advice to the Sicilians, happens to provide advice consistent with Plato's political commentary in the *Laws*, specifically, that "neither Sicily, nor yet any other State ... should be enslaved to human despots but rather to laws." (Pl. *Ep. VII*. 334c) As such, since arguments regarding the spuriousness of this section on the grounds that its claims clash with the content of the *Laws* would not hold in this case, this section can be considered more likely to be Platonic on the grounds of content. Thus, the fact that the models agree with this claim on the basis of style further supports the hypothesis that sections 334c-335d constitute authentic text in the middle of a non-Platonic interpolation.

Philosophical Digression

Another portion of the letter frequently discussed by philologists is the philosophical digression on the Theory of Forms (roughly corresponding to groups 43–53). As discussed earlier in this paper, scholars such as Burnyeat have identified this portion of the letter as dubiously Platonic given the tenuous philosophical argumentation and certain philosophical terminology not seen in other Platonic writings. (Burnyeat and Frede 2015, 122) The results of this model, however, indicate that this section is perhaps the portion of the letter which shows the greatest likelihood of having been written by Plato, with mean

probability of Platonic authorship frequently exceeding 0.8. Furthermore, the models tend to be in strong agreement about these predictions of non-Platonic authorship, with the standard deviation among the predicted probabilities of Platonic authorship lower than the that of rest of the letter. Such results seem to uphold the authenticity of this section despite the arguments made by Burnyeat. One might claim, however, that this section was labeled as Platonic less because of its style than because of its greater affinity to the philosophical character of other Platonic works than the primarily political and biographical content of the rest of the letter. Nevertheless, the presence of other segments in the letter that are classified as strongly Platonic (groups 2–18, for example) yet do not focus heavily on philosophical content should reduce this concern.

The Conclusion of the Letter

Besides the political advice to the Sicilians, the models also label the conclusion of the letter (345c–352a) as, for the most part, dubiously Platonic. For the units of text belonging to this section, the mean probability of Platonic authorship tends to remain below 0.5, sometimes dipping as low as 0.1. Although this portion of the letter has not been discussed as extensively in traditional scholarship with respect to its authorship, one might find clues in its content focusing on Plato's relations to Dion during the time directly preceding the latter's decision to rebel against Dionysius II, ultimately plunging all of Sicily into civil war. Given the high political stakes of Plato's relations with such a pivotal figure, one could theoretically imagine the impetus for a potential forgery of this section to portray Plato's involvement in this situation with a certain spin. This potential reason for inauthenticity, coupled with the consistent and sharply defined low likelihood of Platonic authorship as

predicted by the model, supports the possibility that this section is another spurious interpolation. This hypothesis, however, is weaker for this section than for the section containing Plato's advice to the Sicilians given that this claim has not been supported by the scholarship of prior philologists and philosophers.

Alternate Authorship

If the political advice to the Sicilians and the conclusion were later interpolations to a previously authentic text, the natural question arises as to the identity of the actual author (or authors) of these sections. This is a problem, unfortunately, which cannot be solved within the scope of this paper, because the six other authors analyzed were selected not as alternate candidates for authorship of *Epistle VII*, but rather as other writers with large corpora of preserved works who lived at roughly the same time as Plato and wrote in the same dialect of Greek. Indeed, scholarship has focused much more closely on the question of *whether* Plato wrote the letters than on the related question of *who*, other than Plato, might have written them. Consequently, no strong alternate candidates have been identified in the scholarship. In any case, the most likely candidate for alternate authorship probably would not be another famous Ancient Greek author, but rather a more obscure student or associate of Plato with personal knowledge of his involvement in Sicily and sufficient motive to alter the account of such involvement. Nevertheless, the fact that such a large amount of the text classified as not belonging to Plato is labeled as characteristic of Xenophon is interesting — perhaps this is to be expected, given that Xenophon also wrote Socratic dialogues similar (relative to the other authors, at least) in genre and content to those of Plato's.

3.3 Conclusion

In short, the above analysis of the different sections of *Epistle VII* shows that, for the most part, the letter is likely authentic, with two primary exceptions: the political advice to the Sicilians (332a–337c) and the ending (345c–352a). Such an explicitly delineated pattern of sections considered authentic interrupted by sections considered spurious strongly supports the possibility of interpolation(s) originally proposed by philologists such as Schofield. It is worth noting, however, that the philosophical digression on the Theory of Forms, doubted by scholars such as Burnyeat, appears, on the other hand, to be quite consistent with Platonic style — even if its arguments have been characterized as clumsily expressed and thus un-Platonic. Nevertheless, it is a positive sign for the significance of this analysis that the model was able to identify different content-based divisions of the letter, a result which confirms the ability of such models to effectively analyze Ancient Greek texts in general.

4

Closing Remarks

The results of this study add further credence to the hypothesis of non-Platonic textual interpolation in Plato's *Epistle VII*. Specifically, I view it as likely on the basis of the models' predictions that the majority of the letter is authentic, with the portion consisting of political advice to the Sicilians and perhaps the conclusion having been added to the letter by some other author. Notably, the much-discussed digression on the Theory of Forms

appears authentic. Such a conclusion, however, relies just as much on traditional scholarship as on computational analysis, with arguments such as Schofield's regarding the potential interpolation of the political advice and the results of this paper's models mutually informing each other. In essence, the results of this experiment correspond nicely to the interplay between style-based and content-based arguments described at the end of the first chapter — namely, my analysis claims on stylistic grounds that most of the letter is authentic, yet uses content-based arguments to augment the neural model's predictions that certain portions are inauthentic.

The methodology and results of this study raise many questions regarding both the authenticity question of the *Epistles* and the application of compuatational methods to Ancient Greek. Concerning the former, now that it seems likely that some portions of the letter were not written by Plato, *who*, then, might the alternative author be? Was it one person or rather multiple different interpolators who added to the text? An attempt to answer these questions would surely begin with a closer examination of Plato's associates in the Academy and their potential motivation for writing text under Plato's name. Concerning the computational methodology, although the results of this paper show that the lemmatization-BPE ensemble leads to the greatest accuracy for Ancient Greek authorship classification models, should this result always be the case for any task in Ancient Greek NLP? Furthermore, the models presented are by no means the only possibilities to consider. Would it be suitable to use more powerful models such as transformers for Ancient Greek given the limited size of the corpus? Finally, how can one know for sure that the models are truly learning Platonic style, and not just the genre differences between the different authors?

Finally, the fact that a neural authorship classification model was able not only to accurately label texts 89% of the time but also to identify the different sections of *Epistle VII* with precision shows the strong promise for the application of similar models to other longstanding debates in classical philology. The applications for deep learning in the field of Classics are vast in number, ranging from similar problems of authorship in other Platonic works to a diverse set of undertakings such as text restoration, sentiment analysis, and named entity recognition. It is my hope that the methodology outlined in this paper, such as the model architecture and examination of pre-processing methods, will serve as a precedent for subsequent and more ambitious studies. Ultimately, this paper shows that a computational approach to text analysis not only stands on itself, but also supports arguments made in traditional philological scholarship. This powerful partnership holds tremendous potential for solving other problems of note in Ancient Greek literature.

A

Texts Included in the Authorship Classification Dataset

Work	Author
Parmenides	Plato
Letters	Plato

Work	Author
Timaeus	Plato
Sophist	Plato
Euthyphro	Plato
Theaetetus	Plato
Republic	Plato
Statesman	Plato
Hipparchus	Plato
Phaedrus	Plato
Meno	Plato
Gorgias	Plato
Protagoras	Plato
Hippias Major	Plato
Alcibiades 1	Plato
Alcibiades 2	Plato
Critias	Plato
Epinomis	Plato
Crito	Plato
Phaedo	Plato
Cratylus	Plato
Apology	Plato
Laws	Plato

Work	Author
Minos	Plato
Symposium	Plato
Lovers	Plato
Cleitophon	Plato
Lysis	Plato
Ion	Plato
Charmides	Plato
Hippias Minor	Plato
Laches	Plato
Euthydemus	Plato
Theages	Plato
Menexenus	Plato
Philebus	Plato
Agesilaus	Xenophon
Cyropaedia	Xenophon
Hellenica	Xenophon
Anabasis	Xenophon
Hiero	Xenophon
On the Cavalry Commander	Xenophon
On the Art of Horsemanship	Xenophon
On Hunting	Xenophon

Work	Author
Economics	Xenophon
Symposium	Xenophon
Apology	Xenophon
Memorabilia	Xenophon
Ways and Means	Xenophon
Constitution of the Lacedaemonians	Xenophon
History of the Peloponnesian War	Thucydides
For The Soldier	Lysias
Against Philon	Lysias
Defense in the Matter of the Olive Stump	Lysias
On the Murder of Eratosthenes	Lysias
Against Andocides	Lysias
Against Nicomachus	Lysias
Accusation of Calumny	Lysias
Against Alcibiades 2	Lysias
Against Eratosthenes	Lysias
On The Refusal Of A Pension	Lysias
Against Pancleon	Lysias
Against The Corn-Dealers	Lysias
Defense Against a Charge of Subverting the Democracy	Lysias
Against Agoratus	Lysias

Work	Author
Against Alcibiades 1	Lysias
Against Diogeiton	Lysias
Against Simon: Defense	Lysias
On A Wound By Premeditation	Lysias
For Callias	Lysias
Funeral Oration	Lysias
Against The Subversion of the Ancestral Constitution	Lysias
Olympic Oration	Lysias
Against Theomnestus 2	Lysias
In Defense of Mantitheus	Lysias
Against Philocrates	Lysias
For Polystratus	Lysias
Against Epicrates and his Fellow-envoys	Lysias
On The Property Of The Brother Of Nicias: Peroration	Lysias
On the Scrutiny of Evandros	Lysias
On the Property of Aristophanes	Lysias
Defence Against A Charge Of Taking Bribes	Lysias
On The Property Of Eraton	Lysias
Against Ergocles	Lysias
Against Theomnestus 1	Lysias
Against Timarchus	Aeschines

Work	Author
Against Ctesiphon	Aeschines
The Speech on the Embassy	Aeschines
Helen	Isocrates
To Demonicus	Isocrates
Against Euthynus	Isocrates
Aegineticus	Isocrates
Against the Sophists	Isocrates
Evagoras	Isocrates
Plataicus	Isocrates
To Archidamus	Isocrates
To the Children of Jason	Isocrates
To Dionysius	Isocrates
To the Rulers of the Mytilenaeans	Isocrates
To Nicocles	Isocrates
Nicocles or the Cyprians	Isocrates
Against Lochites	Isocrates
Concerning the Team of Horses	Isocrates
Trapeziticus	Isocrates
Against Callimachus	Isocrates
Panegyricus	Isocrates
Archidamus	Isocrates

Work	Author
To Antipater	Isocrates
To Philip	Isocrates
Areopagiticus	Isocrates
To Timotheus	Isocrates
Antidosis	Isocrates
Panathenaicus	Isocrates
On the Peace	Isocrates
To Alexander	Isocrates
Busiris	Isocrates
Apollodorus Against Nicostratus	Demosthenes
Exordia	Demosthenes
Third Philippic	Demosthenes
For Phormio	Demosthenes
Against Onetor II	Demosthenes
Against Nausimachus and Xenopeithes	Demosthenes
On Halonnesus	Demosthenes
Letters	Demosthenes
Apollodorus Against Callipus	Demosthenes
Against Callicles	Demosthenes
First Olynthiac	Demosthenes
Against Boeotus I	Demosthenes

Work	Author
Second Philippic	Demosthenes
Against Onetor I	Demosthenes
On the Chersonese	Demosthenes
Against Pantaenetus	Demosthenes
For the Liberty of the Rhodians	Demosthenes
Philip's Letter	Demosthenes
Against Timocrates	Demosthenes
Against Aristocrates	Demosthenes
Against Olympiodorus	Demosthenes
Apollodorus Against Stephanus II	Demosthenes
Against Spudias	Demosthenes
Against Androtion	Demosthenes
Against Aristogeiton I	Demosthenes
On Organization	Demosthenes
On the Navy-Boards	Demosthenes
Against Boeotus II	Demosthenes
Against Evergus And Mnesibulus	Demosthenes
Apollodorus Against Timotheus	Demosthenes
Against Zenothemis	Demosthenes
Against Lacritus	Demosthenes
Third Olynthiac	Demosthenes

Work	Author
First Philippic	Demosthenes
Apollodorus Against Polycles	Demosthenes
Against Eubulides	Demosthenes
The Erotic Essay	Demosthenes
Apollodorus Against Neaera	Demosthenes
On the Peace	Demosthenes
Second Olynthiac	Demosthenes
Against Phormio	Demosthenes
Against Apaturius	Demosthenes
Against Theocrines	Demosthenes
The Funeral Speech	Demosthenes
Against Dionysodorus	Demosthenes
On The Trierarchic Crown	Demosthenes
Against Phaenippus	Demosthenes
Apollodorus Against Stephanus I	Demosthenes
Answer to Philip's Letter	Demosthenes
For the People of Megalopolis	Demosthenes
Against Aphobus III	Demosthenes
Against Leptines	Demosthenes
Against Aphobus I	Demosthenes
On the Crown	Demosthenes

Work	Author
Against Leochares	Demosthenes
Against Macartatus	Demosthenes
Against Aristogeiton II	Demosthenes
On the Embassy	Demosthenes
Against Meidias	Demosthenes
On the Treaty with Alexander	Demosthenes
Against Aphobus II	Demosthenes
Fourth Philippic	Demosthenes

Bibliography

Bentley, Richard. 1874. *Dissertations upon the Epistles of Phalaris, Themistocles, Socrates, Euripides, and upon the Fables of Aesop.* Edited by Wilhelm Wagner. Berlin: Calvary.

Boas, George. 1948. "Fact and Legend in the Biography of Plato." *The Philosophical Review* 57 (5): 439–457.

Burnyeat, Myles, and Michael Frede. 2015. *The Pseudo-Platonic Seventh Letter: A Seminar.* Edited by Dominic Scott. Oxford: Oxford University Press.

Edelstein, Ludwig. 1966. *Plato's Seventh Letter.* Leiden: Brill.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text analysis." *R Journal* 8 (1): 107–121. https://journal.r-project.org/archive/2016/RJ-2016-007/index.html.

Gage, P. 1994. "A New Algorithm for Data Compression." *The C Users Journal Archive* 12:23–38.

Harward, J. 1928. "The Seventh and Eighth Platonic Epistles." *Classical Quarterly* 22 (3–4): 143–154.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation.* 9 (8): 1735–1780.

Johnson, Kyle P., Patrick Burns, John Stewart, and Todd Cook. 2014–2020. *CLTK: The Classical Language Toolkit.* https://github.com/cltk/cltk.

Kenny, Anthony John Patrick. 2004. *Ancient Philosophy.* Oxford: Clarendon Press.

Kingma, Diederik, and Jimmy Ba. 2015. "Adam: A Method for Stochastic Optimization." In *Proceedings of the 3rd International Conference on Learning Representations.* San Diego, December.

Koentges, Thomas. 2020. "The Un-Platonic *Menexenus*: A Stylometric Analysis with More Data." *Greek, Roman, and Byzantine Studies* 60:211–241.

Kraut, Richard. 2017. "Plato." In *The Stanford Encyclopedia of Philosophy,* Fall 2017, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/plato/.

Ledger, Gerard R. 1989. *Re-Counting Plato: A Computer Analysis of Plato's Style.* Oxford: Clarendon Press.

Levison, M., A. Q. Morton, and A. D. Winspear. 1968. "The Seventh Letter of Plato." *Mind* 77 (307): 309–325.

Lutosławski, Wincenty. 1897. *The Origin and Growth of Plato's Logic with an Account of Plato's Style and of the Chronology of his Writings.* London: Longmans / Green.

Mosteller, Frederick, and David L Wallace. 1963. "Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of the Disputed Federalist Papers." *Journal of the American Statistical Association* 58 (302): 275–309.

Nicholson, Brennan. 2020. *Ancient-Greek-Char-Bert.* https://github.com/brennannicholson/ancient-greek-char-bert.

Plato. 1926. *Laws, Volume I: Books 1-6.* Translated by R. G. Bury. Loeb Classical Library 187. Cambridge, MA: Harvard University Press.

———. 1997. *Complete Works of Plato.* Edited by John M. Cooper and D. S. Hutchinson. Cambridge, MA: Hackett Publishing.

———. 1999. *Timaeus; Critias; Cleitophon; Menexenus; Epistles.* Translated by R. G. Bury. Loeb Classical Library 234. Cambridge, MA: Harvard University Press.

———. 2013a. *Republic, Volume I: Books 1-5.* Translated by Christopher Emlyn-Jones and William Preddy. Loeb Classical Library 237. Cambridge, MA: Harvard University Press.

———. 2013b. *Republic, Volume II: Books 6-10.* Translated by Christopher Emlyn-Jones and William Preddy. Loeb Classical Library 276. Cambridge, MA: Harvard University Press.

Schofield, Malcolm. 2000. "Plato and Practical Politics." In *The Cambridge History of Greek and Roman Political Thought,* edited by Christopher Rowe, 293–302. Cambridge: Cambridge University Press.

Sennrich, Rico, Barry Haddow, and Alexandra Birch. 2016. "Neural Machine Translation of Rare Words with Subword Units." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* 1715–1725. Berlin: Association for Computational Linguistics, August.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research* 15 (56): 1929–1958.