



PROJECT MUSE®

*Authorship Analysis and the Ending of Seven Against Thebes :
Aeschylus' Antigone or Updating Adaptation?*

Nikos Manousakis, Efstathios Stamatatos

Classical World, Volume 116, Number 3, Spring 2023, pp. 247-274 (Article)

Published by Johns Hopkins University Press



➔ For additional information about this article

<https://muse.jhu.edu/article/896430>

Authorship Analysis and the Ending of Seven Against Thebes: Aeschylus' Antigone or Updating Adaptation?

NIKOS MANOUSAKIS AND EFSTATHIOS STAMATATOS

ABSTRACT: The present paper revisits the discussion concerning the authenticity of a crucial part in Aeschylus' *Seven Against Thebes*: the highly controversial ending of the play. Much has been written on the subject by various scholars, and even though there is now a general consensus that at some point in antiquity the ending of the play was "touched" by an author other than Aeschylus, the problem still remains unresolved in its devilish details. The question is of critical importance for classicists and theatre practitioners but also for anyone interested in classical literature, since, if the ending in the manuscripts is in fact Aeschylean, then Aeschylus could have been the first dramatist—long before Sophocles—to put on stage a defiant Antigone, eager to bury her brother Polyneices despite the civic prohibition. If the ending is spurious, then this will decisively affect how the play in question is read, studied, and staged. To address the problem, we used various tried and tested computer authorship attribution methods: *Common n-grams*, *Support Vector Machines*, and *n-gram tracing*. Thus, this study sheds new, interdisciplinary light on an old and perplexing philological question.

KEYWORDS: Aeschylus, *Seven Against Thebes*, Ending, Antigone, Sophocles, Authenticity, Authorship Attribution, Algorithmic Models, *Machine Learning*

In 467 BCE Aeschylus staged a thematically connected tetralogy on the myth of the Labdacids, winning first prize at the City Dionysia. The tetralogy consisted of *Laius*, *Oedipus*, *Seven Against Thebes* (henceforward *Seven*), and the satyr-drama *Sphinx*.¹ The only extant of these plays,

¹ On Aeschylus' Theban tetralogy see Hutchinson 1985: xvii–xl; Sommerstein 2010: 84–90.

Seven, and one of the earliest surviving tragedies, is about the mutual killing of Oedipus' sons, Eteocles and Polyneices. At the beginning, Thebes has been under siege for some time: Polyneices, forced into exile by his brother, went to Argos and talked king Adrastus into attacking his homeland, to restore him into his father's inheritance. During the play a scout lets Eteocles know that seven enemy leaders will soon attack the seven gates of Thebes, each one allotted to one gate.² Eteocles then makes his own defensive dispositions. In a scene that occupies nearly one-third of the play (369–676), he stations (in seven pairs of speeches) seven Theban champions to face the Argive leaders.³ He announces that he will defend the seventh gate, in which, though, the enemy leader is none other than his brother, and the Chorus of Theban maidens attempts to stop the fratricide to no effect. A messenger brings the news that Eteocles has killed Polyneices but has also been slain himself by the latter. From that point on, the play has most likely been tampered with at some later time, and the transmitted text is not (exactly) what Aeschylus composed.⁴

Oedipus' two daughters, Antigone and Ismene, are never mentioned before the culmination of the play (after Eteocles' final exit and the news about the fratricide), in which the utter destruction of the royal family is suggested⁵—a fate which is sealed with the blood of Eteocles and Polyneices. However, just before the end of the drama the girls arrive on stage, and join the women of the Chorus to sing part of an elaborate lament over the dead bodies of the princes. At the closure of the lament, the question of burial is raised, and is suggested—only to be rejected (see Sommerstein 2008: 265n152)—that the two brothers should lay at the royal tombs, next to their father. And then, something even more extraordinary happens: a herald enters to announce the decision of the Theban people's council⁶: while Eteocles is to be buried with full honors, Polyneices is to be cast out unburied as prey to the dogs. Antigone refuses to accept that

² For the myth of the Seven see, in detail, Natanblut 2005. For the Seven in art see Armantrout 1990.

³ For the various aspects of these seven pairs of speeches see, concisely, Hutchinson 1985: 103–106. See further, Zeitlin 1982: especially from p. 53 onwards; Vidal-Naquet 1990; Catenacci 2004; and D.W. Berman 2007: 33–86.

⁴ "At present, probably the majority of specialists would agree that the ending [of *Sept.*] as we have it is not Aeschylus' work, though its genuineness still occasionally finds defenders." See Sommerstein 2010: 91.

⁵ See *Sept.* 689–691, 812–813, 878–879, 951–956. Cf. 720–726, 876–877, 881–885.

⁶ Cf. the decree of the infuriated Argive people in Euripides' *Or.* 46(–50) with the note by Willink ad loc.

and declares that she will bury her brother herself. The herald argues with her to no avail and leaves without defining a penalty for burying Polyneices. The Chorus then divides into two groups. One, following Ismene's lead, escorts Eteocles to his grave, and the other, following Antigone's lead, escorts Polyneices. The authenticity of this part of the play, especially Antigone's confrontation with the herald, is highly suspect. In fact, this scene, according to Sommerstein (2008: 147),

ruins an ending which till then had stressed, over and over again, the equality of the brothers in death, and leaves the action of the play, and therefore the trilogy, lacking any closure; at one point, moreover (1039 [. . .]), the text can hardly be understood without prior acquaintance with Sophocles' *Antigone*. It is overwhelmingly probable, therefore, that this scene was added for a restaging of *Seven*, at a time when Sophocles' play had made it impossible to think of the mutual slaughter of Eteocles and Polyneices without also thinking of the tragic heroism of their sister.⁷

As is the case with *Prometheus Bound*, a play traditionally ascribed to Aeschylus but in all likelihood not by him,⁸ and unlike what is the case with *Rhesus* attributed to Euripides,⁹ the authenticity of *Seven* was, so far as one can know, never questioned in antiquity. The suggestion of a spurious ending came about in the mid-nineteenth century.¹⁰ In 1948 the

⁷ Hutchinson 1985: 209(–211) maintains that “[t]he final scene of [Sept.] is closely related to the final scene of [Euripides’] *Phoen*.”—not (at least primarily) Sophocles’ *Ant*. Yet the *exodos* in Euripides’ drama (1582–1766) is also suspect. See, in detail, Mastronarde 1994: 591–94, 627–28, 635–37. Diggle 1994 and Kovacs 2002 reject the *exodos* as a whole. Mastronarde 1994 rejects 1596, 1634, 1637–1638 and 1737–1766 in his text. If the final part of *Phoen*. is indeed spurious, then the popularity of Sophocles’ *Ant*. may have triggered both interpolations, in *Phoen*. and *Sept*. For the athetesis of the scene in *Phoen*. see further, Natanblut 2003: 87–91.

⁸ See, concisely, Sommerstein 2010: 228–32; in detail Griffith 1977; and most recently Manousakis 2020 for a book-length traditional and non-traditional (computer-based) *authorship analysis* of this play.

⁹ For the authorship of *Rh*. see Liapis 2012: lxxvii–lxxv; Fries 2014: 22–47; Manousakis and Stamatatos 2018 for a detailed traditional and non-traditional (computer-based) *authorship analysis* of the play; and most recently Fantuzzi 2020: 16–23.

¹⁰ See Lloyd-Jones 1959: 80. This scholar does not exactly defend the *exodos* of *Sept*. (see 113–114). As Winnington-Ingram 1977: 4n4 suggests, he seeks to demonstrate “that the objective evidence adduced against [it] falls short of establishing that [it is] spurious.” For the authenticity question see further, Koenigsbeck 1891; Platt 1912; Page 1934: 30–32; Nicolaus 1967; Mellon 1974; Brown 1976: 206–19; Hutchinson 1985: 209–11; and Sommerstein 2010: 90–93. Other discussions are by Fraenkel 1964; Cameron 1971: 49–56; Winnington-Ingram 1977: 5–4; Taplin 1977: 169–91; Thalmann 1978: 137–41; Dawe

didaskalia to *Seven* in the Medicean manuscript was first published, and it became clear that this is the third and concluding tragedy in Aeschylus' Theban tetralogy—not the second, as it had been generally supposed until then. Yet this is a very peculiar concluding play, leaving some major loose ends tied to Polyneices' burial and Antigone's disobedience. If the current ending of *Seven* is sound, then at the final part of his Theban tetralogy Aeschylus chose to raise a crucial issue, to then hardly treat it.¹¹ “[H]istorically speaking, this argument lies at the root of the whole modern discussion of the problem” (Lloyd-Jones 1959: 80).

Besides the “dramatic absurdity” (Sommerstein 2010: 92) of Antigone's scene, the disputed status of the ending of *Seven* also poses various other questions:

1. Is the brevity of the messenger's speech announcing the mutual killing of the brothers (792–821: *rhēsis* and dialogue with the Chorus) associated with some (major) textual disruption?¹²
2. Is the anapaestic introduction to the third stasimon (822–831) from Aeschylus' hand?¹³
3. How is one to account for the few anapaests (861–874) introducing Antigone and Ismene into the play and at the same time abruptly disrupting the Chorus' lament over the dead bodies of Eteocles and Polyneices?¹⁴

1976 and 1978; Conacher 1996: 71–74; West 2000: 351–52; Barrett 2007; all against Aeschylean authorship of the ending of *Sept*. Flintoff 1980; Orwin 1980; Ryzman 1983; and Tsantsanoglou 2010 are in favor. See also the bibliography by Grilli 2018: 78n37, specifically on dating the interpolation.

¹¹ See especially, Thalmann 1978: 141; Lupaş and Petre 1981: 282; and Sommerstein 2010: 92. For the expedition of the Epigonoi and Aeschylus' Theban tetralogy see further, Cameron 1971: 53ff; Taplin 1977: 181n2; Hutchinson 1985: 167, 195–96; Conacher 1996: 73–74.

¹² See Taplin 1977: 84, 167–69; and Hutchinson 1985: 173–74. However, see also Verrall 1887: 93 who notes that “[f]rom [805 to 821] there is scarcely a verse which has not been excised, displaced, and corrected in various ways.” See, in detail, Willink 1968. Verrall 1887: 96–97 draws attention to the peculiarities in vocabulary and metre in the passage (see especially the four resolutions in 14 lines, all at position six (803, 806, 810, 816), and 3 of them (803, 806, 816) of the same word-shape: $\cup\cup\cup-$). See the table by Schein 1979: 82. Von der Mühlh 1964 suggests a lacuna, conjecturing that the original, the Aeschylean, messenger speech was longer.

¹³ For abnormalities of language, metre, and sense in the passage see Dawe 1978: 88–89. However, see also Hutchinson 1985: 184–86.

¹⁴ After the dead bodies are brought on stage, the women of the Chorus are self-identified as “rowers” with the wind of lamentation in their sails, seeing off the two princes to

4. How is one to account for the silence of the sisters during the first part of the lament: “are [they] introduced at this point [. . .] to take a quite unnecessary part in a lament belonging to the Chorus or to do nothing at all”? (Brown 1976: 207)¹⁵
5. How is one to account for the fact that the rest of the play is clearly designed for two actors and the ending for three?¹⁶
6. How is one to account for a tragedy that seems to be pointing to the total extinction of the house of Laius—who should have died without issue (see *Seven* 742–749)—only to show that this is not the case in the end?
7. How is one to account for an ending—if the two bodies are separated in burial—distorting what seems to be the playwright’s intention concerning the fate of the two brothers: to be, eventually, reconciled in death.¹⁷

These and other, more or less specific, questions make one wonder about the (precise) extent and structure of tampering the play has

a place of no return. This quite apposite opening of the choral lament is, out of the blue, interrupted by the two sisters, who come on stage to join the lament, but stay (inconceivably) silent for some time, since “we [(the women of the Chorus)], <having heard> the news first, should raise the unpleasing sound of the Fury’s hymn, and sing of the hateful paean of Hades.” See 866–869. The translation is by Sommerstein 2008. Hutchinson 1985: 191 suggests that the interpolator “did not wish to have the sisters appear before the corpses, and he did not wish to expel the authentic lines 848–860: he was [thus] driven to this result.” See, further, Taplin 1977: 169–79. Murray 1955 makes the sisters sing the lament from the beginning, 875 onwards, while Page 1972 makes them sing 961–974 and 989–1004. Hutchinson 1985, West 1990, and Sommerstein 2008 assign the dirge to the two semi-Choruses.

¹⁵ For the silence of the sisters see, in detail, Taplin 1972: 84–89 and 1977: 179–80. Furthermore, Hutchinson 1985: 192; Thalmann 1978: 140; and Sommerstein 2010: 90–91.

¹⁶ See Brown 1976: 207. Yet see also Tsantsanoglou 2010. “An additional motive for the new ending may have been to provide employment for a third actor. Normally, in revivals of Aeschylean plays which had originally used only two actors, it would be possible to satisfy a three-man troupe by giving each of them at least one speaking part. [*Sept.*] is unique among surviving Greek plays in having originally had a total of only two speaking parts, and the third actor in the troupe might well not have been too pleased at having to sit out a complete play; the added ending gives him a role (and indeed the chance to play Antigone!).” See Sommerstein 2010: 95n24.

¹⁷ See *Sept.* 766–767, 884–885, 908–909, 941. Further, as Sommerstein 2010: 83 aptly puts it, “[no] distinction [is ever] made between the brother who died saving Thebes and the brother who died sworn to destroy it. The only difference between Eteocles and Polyneices is that the latter had been in exile (979, 991); everything else that is said of them is both alike.”

suffered, and if it is possible to tell which lines exactly are by Aeschylus beyond 792 (the messenger's entrance). Some of the earliest scholars to address this issue, Bergk and Wilamowitz, differ about retaining 961–1004. They both remove the two sisters, rejecting 861–874 (–873 in West's 1990 edition), but the former maintains that the original text ends at 960, whereas the latter at 1004 (see Lloyd-Jones 1959: 80–81). Even though he rejects the herald's scene, 1005–1078, Murray 1955 does retain Antigone and Ismene in the preceding lament. Page 1972 retains the sisters in the (final part of the) lament as well as the herald's scene. Taplin 1977, after Fraenkel 1964, also suggests that the play "seems almost complete at 1004."¹⁸ Dawe 1978: 101 indicates that 822–831, 848–874, 996–997 and 1005–1078 are not Aeschylean, while Brown 1976 considers a middle ground, arguing that 1026–1053 (Antigone's answering speech about Polyneices' burial and her argument with the herald) is interpolated, while 1005–1025 (the herald's speech) and 1054–1078 (the choral finale) are authentic. West 1990 and Sommerstein 2008 reject both the sisters' participation in the Choral lament and the herald's scene.

The present study attempts to approach anew the authenticity question, using (non-traditional) *authorship analysis*. Following Manousakis and Stamatatos 2018 and Manousakis 2020, applying *authorship analysis* on Greek drama, we employed authorship attribution algorithmic models to ascertain if the text of *Seven*, as it has come down to us, is tampered with and/or interpolated; and if so, to determine (as accurately as possible) the extent and structure of the corruption. The models we employed are: the *Common n-grams (CNG)* method, *Support Vector Machines (SVM)*, and *n-gram tracing*, and they were chosen in order to not only show whether the disputed lines in *Seven* are by Aeschylus but also to track down their possible affinity to the work of other relevant authors (comparative corpus). As regards the specific target/suspect text, the process, as one can infer from the discussion above, is puzzling. There is no long, uniform piece to examine (e.g. a complete play or an actually extensive excerpt), and one has to try to authenticate a series of (more or less) short passages within a work of considerable length (1078 lines).

Following the scholarly evidence, we can safely suggest that there are three spots in the text one should focus on—130 lines overall:

¹⁸ Taplin 1977: 180n1. Cf. Thalmann 1978: 137; Conacher 1996: 71–74.

- a) the messenger's speech about the mutual fratricide and the Chorus' response to it: 792–821
- b) the anapaests opening the lament, the anapaests introducing Antigone and Ismene during the lament, and two personalized dochmiac lines in the main lament: 822–831, 861–874, 996–997
- c) the herald vs. Antigone scene, along with the concluding anapaests of the play: 1005–1078

In all three pieces there may lie adapted Aeschylean material. Only for the anapaests introducing the sisters, the iambic confrontation between Antigone and the herald, and the final anapaests (861–874 and 1005–1078), 88 lines overall, can one suggest an interpolation. In view of these observations, we feel that the play as a whole should be brought under the scrutiny of *authorship analysis*, with 792–1078, 287 lines overall, being its focus. Lines 792–1078 constitute five thematic sections: 792–821—the messenger's interaction with the Chorus about how Eteocles and Polyneices perished; 822–860—the Chorus lamenting before the arrival of Antigone and Ismene; 861–1004—the Chorus lamenting along with the two sisters; 1005–1053—the herald's interaction with Antigone; 1054–1078—the Chorus' final interaction with the (mute) sisters.

The comparative corpus used in the present study consists of five secure plays by Aeschylus, and five by Sophocles, Euripides, and Aristophanes respectively. The choice of plays, whenever possible, was made on chronological grounds, in order to cover—as widely as possible—all phases in the career of each of these dramatists. Overall, the corpus, in principle put together in terms of generic affinity,¹⁹ consists of the 20 plays given in the table below.²⁰

¹⁹ See Manousakis 2020: 156–57. Tragic language is largely artificial—as is also the case with the epic language (about which see, briefly, Willi 2011)—and stands at some (defamiliarising) distance from the (natural) Attic Greek of 5th century BCE. The language of tragedy displays various peculiarities in morphology (dialectal and other), syntax, and vocabulary, through which the linguistic capture of the current and the eternal takes place at the same time. This is by no means to say that there is no formal common ground between spoken Attic Greek and this idiom. As Dik 2007: 2 wonders, “how can a poet effectively characterize Antigone and Creon, or how does he expect us to follow Oedipus' interrogation of the shepherd, unless he draws on some common core of Greek grammar that is shared by the spoken language and by written prose?” On tragic language see, in detail, Else 1965: 72–73; Rutherford 2012: chap. 3; most importantly Silk 1996; Mastronarde 2002: 81–96; and Allison 2003: 15 (especially n62). The distancing quality of the artificial tragic idiom becomes quite clear in paratragedy (see Rutherford 2012: 59–61).

²⁰ For the methodological decisions in processing this corpus see, in greater detail, Manousakis 2020: 158–64.

Table 1: Comparative Corpus

<i>Aeschylus</i>	<i>Sophocles</i>	<i>Euripides</i>	<i>Aristophanes</i>
<i>Persae</i> (472 BCE)	<i>Antigone</i> (about 442 BCE)	<i>Alcestis</i> (438 BCE)	<i>Clouds</i> (423 BCE, and 420/17/14 for the extant version) ²¹
<i>Suppliants</i> (463 BCE?) ²²	<i>Electra</i> (± 420 BCE)	<i>Medea</i> (431 BCE)	<i>Birds</i> (414 BCE)
<i>Agamemnon</i> (458 BCE)	<i>Oedipus the King</i> (± 420 BCE)	<i>Heracles</i> (± 414 BCE)	<i>Lysistrata</i> (411 BCE)
<i>Libation Bearers</i> (458 BCE)	<i>Philoctetes</i> (409 BCE)	<i>Ion</i> (± 413 BCE)	<i>Frogs</i> (405 BCE)
<i>Eumenides</i> (458 BCE)	<i>Oedipus at Colonus</i> (405/1 BCE)	<i>Bacchae</i> (± 408/6 BCE) ²³	<i>Wealth</i> (388 BCE)

Authorship attribution is much like a law case. A researcher attempts to “uncover” the truth about the authorial provenance of a text: to remain unbiased in proving or disproving a traditional judgment about the authorship of some text, or even to show that a text is the work of an author who was not even in the picture at the outset of the research. Yet in technical terms authorship attribution can be seen as a computer science problem of optimal algorithmic *representation* of the text(s) under analysis on the basis of specific (textual) features, and the classification of this/these text/s to the right class (group-category): the attribution of a text to its author, or its removal from the corpus of an author who never composed it. Various textual—*trace*²⁴—indicators of authorship and (all the more sophisticated) methods of exploiting them have been proposed for solving this problem. Such attempts are mainly based on lexical (distribution of words or sequences of words, vocabulary richness, word or sentence length, etc.), grammatical–syntactic (distribution of parts of

²¹ For the revised version of *Clouds* see Dover 1968: lxxx–xcviii; Hubbard 1986; Kopff 1990; Storey 1993; Meineck and Storey 2000: 115–19.

²² See Garvie 2006²: ix–xv, 9 for the dating of the play.

²³ In Euripides’ case the aim was to compile a corpus covering as broadly as possible the four decades of his career.

²⁴ For the term *trace*—the “imprint” an author unintentionally and unconsciously leaves in her/his writings—see, in detail, Manousakis 2020: 8–14, *passim*.

speech or syntactic structures, etc.), character (distribution of character sequences), or semantic features (e.g. distribution of synonyms and semantic dependencies). Lexical and character *trace* indicators are, most conveniently, language-independent. These features can be extracted to their full extent—to produce the machine *representation* of the text—and examined thoroughly. The distribution of character sequences (letters/sounds) has been proven by far the most effective *trace* indicator in automated authorship attribution studies. Character *n*-grams, i.e. strings/sequences of consecutive characters by one (unigrams), two (bigrams), three (trigrams), four (tetragrams) etc., yield the most accurate results in classification. The length of *n* (bigrams, trigrams, etc.) depends on the needs of each individual study (the main criterion being the relative length of words in the language the study concerns). It is evident, of course, that the larger the *n* becomes, the closer the character *n*-grams will come to representing the thematically charged, and thus biased, content words—passing from the morphological and grammatical level of language to the purely lexical. Even though they capture both morphemic and lexical information, character *n*-grams are neither morphemes nor words, but are in an intermediate state—which also provides us with contextual information for morphemes and words (see, in detail, Manousakis 2020: 153–56). All authorship attribution methods used in the present study are based on character *n*-grams.

I. The *Common n*-grams (CNG) Method

The *CNG profile* authorship attribution method²⁵ is based on extracting the most frequent character *n*-grams of size *n* in the training data

²⁵ *Supervised Machine Learning* authorship attribution models fall into two categories, “according to whether they treat each training text individually or cumulatively (per author). In more detail, some [models, called *profile-based* models,] concatenate all the available training texts per author in [a single] file and extract a cumulative representation of that author’s [*trace*] [. . .] from this concatenated text [(the author’s *profile*)]. That is, the differences between texts written by the same author are disregarded [. . .]. [A]nother family of [models, the *instance-based*,] requires multiple training text samples per author [. . .]. In this case, each training text is individually represented as a separate *instance* of authorial [*trace*] [. . .]. [E]ach training text sample [is] a unit that contributes separately to the attribution model. In other words, each text sample of known authorship is an *instance* of the problem in question [. . .]. [The majority of current models for authorship attribution are of this kind. There are also hybrid models,] combin[ing] characteristics of *profile-based* and *instance-based* [algorithms].” See Stamatatos 2009: 545–46, 548. The

(candidate authors). These *n*-grams are sorted by their *normalized* frequency, and the *L* most-frequent *n*-grams, a parameter tuned by the researcher, as well as the value of *n*, define an author's *profile*. Given a testing document (text of unknown authorship), the testing *profile* is produced in the same way, and then the distances between the testing *profile* and the *profiles* of various training authors are calculated. The testing document is finally classified using the *k*-nearest neighbors algorithm: it is attributed to the training author whose *profile* is closest to the testing *profile*.²⁶ More technically, the *profile*, *P*, of some text is a set of *L* pairs $\{(g_1, f_1), (g_2, f_2), \dots, (g_L, f_L)\}$, where g_1, g_2, \dots, g_L are the *L* most frequent *n*-grams of the text (decreasing order) and f_1, f_2, \dots, f_L their *normalized* (over text length)²⁷ frequencies of occurrence. A testing text is assigned/ascribed to a training author through a dissimilarity function comparing the testing text *profile* with the *profiles* of all the training authors. Let *A* be the set of the candidate authors and T_a the training text (the concatenation of all text samples) of author *a*. For a given *n* and *L* (trigrams and 3,000, for instance) consider $P(x)$ as the *profile* of a testing text and $P(T_a)$ as the *profile* of an author *a*. If $f_x(g)$ and $f_{T_a}(g)$ are the frequencies of the *n*-gram *g* in the testing text and in *a*, an author's training text, respectively, the distance (dissimilarity) measure between $P(x)$ and $P(T_a)$ is defined as follows:²⁸

CNG model has proved very effective, see Kešelj, Peng, Cercone, and Thomas 2003; and Stamatatos 2007: 238. See further, Juola 2006: 290–96.

²⁶ See Juola 2006: 293–94. For the *k*-nearest neighbors algorithm see, concisely, De Bruyne 2010: 17–18; and in more detail (the step by step analysis by) Bramer 2016: 29–36; also the discussion by Aggarwal 2015: 160–70.

²⁷ *Normalization (unity-based)* is used to regulate the range of independent variables or features of the data: to bring all values into the range [0,1]. It can be generalized to restrict the range of values in a dataset between any arbitrary points. Since the range of values of raw data varies widely, objective functions will not work properly in a series of (*Machine Learning*) algorithms without *normalization*.

²⁸ This is Stamatatos' 2007 version of the original distance metric, addressing a problem the original formula does not address: the size of *profiles* in a dataset. The original formula calculates the distance for character *n*-grams in the *profile* of a training author and the document of unknown authorship. Each *n*-gram not present in the testing *profile* adds a constant to the metric. Stamatatos' distance function accounts only for items present in the document of unknown authorship. Thus, the distances between the testing and training *profiles* are all calculated on the basis of the same number of items, and therefore a short training *profile* does not affect the overall accuracy of the method. If an author's *profile* is very short, shorter than the predefined *profile* length even, the original formula gives this author an important, biased, advantage over others, since s/he will falsely appear to have proportionately more *n*-grams in common with the document of unknown

$$d_1(P(x), P(T_a)) = \sum_{g \in P(x)} \left(\frac{2(f_x(g) - f_{T_a}(g))}{f_x(g) + f_{T_a}(g)} \right)^2$$

Simply put, the *CNG* method calculates the dissimilarity between the collective/overall *trace* of a training (candidate) author and that of a testing text (of unknown authorship) by the amount of character *n-grams* they have in common. The training author whose collective character *n-grams trace* is closer to that of the testing text is the author most likely to have written it. “Experimental results in a variety of corpora have shown that $3 \leq n \leq 5$ and $1,000 \leq L \leq 5,000$ [are the *n* and *L* values which perform best] in most [. . .] cases.”²⁹ As regards *L*, taking into account the shortness of the suspect passages, and also the evidence that this value is effective for short dramatic pieces (see Manousakis 2020: 202), we tuned the algorithm to 1,000. As regards *n*, we employed only trigrams and tetragrams, in order for the model to provide results based on morphemic, lexical, and contextual information, but at the same time to avoid, as much as possible, lexical pervasiveness (content words) in the analysis.

We used *CNG* to study the highly suspect passages in *Seven* within their context: we tested the authorial *trace(s)* of 792–821 (the messenger’s speech), consisting of 165 words, 822–60 and 875–1004 (the lament for Eteocles and Polyneices save the anapaestic introduction of their two sisters (861–874)), consisting of 624 words, and 861–874 along with 1005–1078 (the anapaestic introduction of Antigone and Ismene and the final scene of the play (herald, Antigone, Chorus), consisting of 453 words. To test the attribution accuracy of *CNG* on such short passages

authorship. The modified formula, on the contrary, is stable and unbiased in an *imbalanced* (length) setting. It is non-symmetrical, for it assumes that the questioned *profile* is possibly shorter (or much shorter) than that of known authorship. For the *class imbalance* problem see further, Stamatatos 2008; Stańczyk 2016. Even though *imbalance* is not crucial for our study (the difference in length between a typical Aeschylean and a typical Sophoclean or Euripidean play is ≈ 300 metrical lines), it is a matter of concern. The comparative corpus (training) was in fact put together in view of the Aeschylean material shortage. Aeschylus’ available *profile* (counting *Sept.* up to line 860) is about 18,000 words shorter than a corresponding (made up of six tragedies) Sophoclean or Euripidean *profile*. That is, the original *CNG* distance metric could have been, to some extent of course, biased towards Aeschylus, simply due to his *profile* being shorter. Hence, we employed only Stamatatos’ formula here.

²⁹ See Stamatatos 2007: 238. More specifically, *L* belongs to 1,000, 2,000, 3,000, 4,000, 5,000, and *n* belongs to 3, 4, 5.

deriving from ancient Greek drama, we first applied this method to several passages of *mutatis mutandis* similar length, deriving from plays of secure authorship not included in the comparative corpus. The 14 randomly selected passages (some of them compiled) containing—as do the suspect ones—spoken and sung speech, are given in table 2 below.

The results of this first *CNG* analysis were very encouraging. More specifically, all validation passages (2 by Sophocles, 8 by Euripides, and 4 by Aristophanes) were correctly attributed to their authors. Applying the same analysis to the suspect pieces in *Seven*, one is surprised (or not) to see that the philological consensus about the spuriousness of certain parts of the text is proven quite right. Examined with both trigrams and tetragrams, *Seven* 861–874 and 1005–1078 (the anapaestic introduction of the two sisters and the whole Antigone vs. herald scene with the final anapaests), consisting of 2,869 characters overall, are authorially associated with Sophocles, not Aeschylus. On the contrary, the disputed by some

Table 2: Validation Passages

<i>Sophocles</i>	<i>Euripides</i>	<i>Aristophanes</i>
<i>Ajax</i> 1–120	<i>Electra</i> 1–111	<i>Thesmophoriazousae</i> 1–100
<i>Ajax</i> 134–262	<i>Electra</i> 112–212	<i>Thesmophoriazousae</i> 101–129, 312–330, 351–371, 434–442, 459–464, 520–530, 663–687, 707–727, 785–845, 953–989, 990–1000, 1015–1055, 1065– 1098 & 1136–1159
	<i>Helen</i> 1–120	<i>Ecclesiazousae</i> 1–100
	<i>Children of Heracles</i> 120–287	<i>Ecclesiazousae</i> 298–310, 483–503, 513–519, 571–581, 893–899, 900–923, 938–945 & 952–976
	<i>Children of Heracles</i> 353–380 & 748–783	
	<i>Orestes</i> 1–139	
	<i>Orestes</i> 1353–1502	
	<i>Iphigenia in Tauris</i> 392–455 & 1089–1152	

scholars 822–831,³⁰ and the rather secure 832–860 and 875–1004—parts of the lament consisting of 3,962 characters overall—are authorially associated with Aeschylus. Probably the most surprising aspect of the results has to do with the messenger’s speech and the Chorus’ reaction to the double fratricide. Examined with trigrams *Seven* 792–821, consisting of 1,106 characters, is almost equally associated in its authorial *trace* with both Aeschylus and Sophocles. Nevertheless, when examined with tetragrams Aeschylus’ *trace* prevails in the passage. This makes one think that (it is not impossible that) the “arranger” of the ending of *Seven* may have somewhat “reshaped” the original messenger’s speech.

II. Support Vector Machines (SVM)

After using character *n*-grams with a *profile-based* method to test the *trace(s)* in *Seven*, we used the same *representation* with an *instance-based* method: *Support Vector Machines (SVM)*. This is a robust, effective, and widely tested *supervised* authorship attribution method (see Juola 2006: 285–86, 321). *SVM*, based on the *statistical learning theory*, was first developed in 1963.³¹ Given a dataset of training examples (texts), each marked as belonging to one or the other of (specifically) two categories (authors), an *SVM* algorithm builds a model that assigns new examples (unclassified texts) to one category or the other, making it a binary linear classifier. The first objective of such an algorithm is to find a *hyperplane* (a partition of the feature space into two categories) in an *n*-dimensional space (*n* here is the number of features used in the *representation*) that has the maximum margin, i.e. the maximum distance between data points (texts) of both categories *a priori* assigned to a certain class. In fact, the *hyperplane* is a subspace (decision boundary) one dimension lower than its ambient space: in a two-dimensional space the *hyperplane* would be a line; if the number of input features is 3, the *hyperplane* becomes a two-dimensional plane, etc. Maximizing the distance that segregates the

³⁰ E.g. Sommerstein 2008 puts them in brackets. Actually, though, Hutchinson 1985: 184–86 has shown that these lines are much less suspect than they were thought to be.

³¹ For the *SVM* classification see, concisely, the entry by Zhang 2011. See further, Clarke, Fokoué, and Zhang 2009: 262–94; also, Schölkopf and Smola 2002; Campbell and Ying 2011; Stoean and Stoean 2014. For the *SVM* in text categorization and authorship attribution see Joachims 2002. See further, the bibliography provided by Juola 2006: 285–86; and Stamatatos 2009: 549–50; for *SVM* classification applied to a series of challenging literary texts see Can, Can, Duygulu, and Kalpakli 2012; and Kaur and Verma 2015.

two categories (authors) provides reinforcement that unseen data points (unclassified texts) will be classified with confidence. The two or more data points—the vectors determining the position and orientation of the *hyperplane*—marking the maximum margin between the two sides are the *Support Vectors*. Through this process the model is trained to “know” how to match disputed texts with one of the two authors whose *trace* is given in advance.

Simply put, *SVM* aims at solving a *supervised* binary classification problem. That is, it uses a *hyperplane*, a subspace, to most effectively separate, to sort out in two classes, objects of known status in a vector space and then categorizes new, unseen, objects into these two groups, based on their properties and on the set of known examples already categorized. The model builds binary classifiers which distinguish (i) between one of the categories and the rest (one-versus-all), or (ii) between every pair of classes (one-versus-one). The classification of new *instances* (texts) in the one-versus-all case is done through a winner-takes-all strategy: the classifier with the highest output function assigns the class. The one-versus-one classification is done by a max-wins voting strategy: every classifier assigns the *instance* to one of the two classes, then the vote for the assigned class is increased by one, and the class with most votes determines the *instance* classification. The *SVM* algorithm used in the present study is based on the one-versus-one approach. The (text *representation*) parameters that have to be tuned in this model are the *n-gram* length and the minimum number of times the *n-grams* used in the analysis are to occur in the training corpus (texts of known authorship that train the algorithm to recognize the *trace* of the given authors).³² For instance, one can use in the analysis all trigrams occurring at least 100 times in the training corpus. Evidently, the lower this number is, the more *n-grams* will be taken into consideration—and if one chooses to analyze, for example, trigrams occurring at least once in the training corpus, one is practically analyzing all trigrams in the corpus. This would lead to a major increase of *dimensionality* in the data (mass and perplexity of information), since the texts would be *represented* by thousands of *n-grams*. However, *SVM* is a powerful model that can effectively handle high-dimensional and sparse data.

³² The *SVM* algorithm also has its own hyper-parameters that should be tuned to achieve optimal classification results. In the present case, default values are used (a linear kernel and $C=1$). On this we follow the practice of previous studies, see Stamatatos 2013 and 2018; also, Sapkota, Bethard, Montes, and Solorio 2015.

As indicated, the comparative corpus used in this study consists of 20 plays: five by each of the chosen four dramatists. This makes for six training pairs in the SVM classification model. That is, Sophocles–Aeschylus; Sophocles–Euripides; Sophocles–Aristophanes; Euripides–Aeschylus; Aristophanes–Aeschylus; Euripides–Aristophanes. Let us suppose, for example, that texts by these pairs of authors are to be analyzed on the basis of trigrams that occur in the training corpus at least 50 times. At first, the SVM model is trained to tell apart the texts by these authors in pairs: based on character trigrams occurring at least 50 times in the training corpus, it should be able to correctly conclude (find the right *hyperplane* between the two given authors) if some text is by Aeschylus or Sophocles, Euripides or Aristophanes, Sophocles or Euripides, etc. When training is over, the model tests the knowledge it has acquired. This process is called *leave-one-out-cross-validation* and it tests how the results of the SVM analysis will generalize to independent data: an evaluation of the consistency of the classification. In *leave-one-out-cross-validation* the model is trained in all *instances* (texts) in the training corpus but one. When training is over, the *instance* left out is tested (classified) as being of unknown authorship. For example, all plays in the training corpus save *Antigone* could be used in training. The training would then be tested in assigning this tragedy to Euripides, Sophocles, Aristophanes, or Aeschylus. The process would be repeated for all plays in the training corpus. Hence, the accuracy of the model is validated internally on texts homologous to the one actually of unknown authorship.

After the *leave-one-out-cross-validation* is complete, one may further test the accuracy of the model externally, using texts that were not part of the comparative corpus—e.g. Sophocles' *Ajax*, Euripides' *Trojan Women* etc. If both the internal (*leave-one-out-cross-validation*) and external validation processes show that the SVM model accurately assigns the validation texts to their actual authors, then one can proceed with some confidence in testing a disputed text. This text, as is the case with all texts so analyzed, will be assigned to the author who gets the most votes by the classifiers. For example, if the Sophocles-Aeschylus, Sophocles-Euripides, Sophocles-Aristophanes classifiers all assign a text to Sophocles, then this text will eventually be attributed to that dramatist. Due to the notable shortness of the disputed piece, in the present case we analyzed all training and validating texts (full plays) not in their complete form but segregated into parts of 50 lines. This segregation was also applied to the whole text of *Seven*, which was thus tested bit by bit.

More specifically, the 1078-line play was tested in 22 parts of 50 lines each: 1–50, 51–100, 101–150, 151–200, 201–250, etc. The last part, 1051–1078, was evidently somewhat shorter than the rest (28 lines: 133 words) but without this having any biasing effect on the process.

The *SVM* method has already been shown to be very effective in correctly attributing quite short pieces of ancient Greek poetry,³⁵ and the results of the—also segregated—*leave-one-out-cross-validation* conducted on the comparative corpus were quite encouraging about the robustness and accuracy of the model and enlightening as regards the most effective minimum values of occurrence of character *n*-grams. The texts used here in external validation, i.e. the 14 validation pieces we employed earlier with *CNG*, were all correctly attributed to their authors. However, in a certain validation test one of the pieces was almost misattributed. This document consists of all lyrics in Aristophanes' *Thesmophoriazousae*: apart from the comic poet, its authorial *trace* was also strongly associated with Euripides. Strangely, instead of showcasing a defect, this “error” brings out the robustness of the method. More specifically, an extensive paratragic³⁴ piece in *Thesmophoriazousae* is in fact molded from Euripidean material. Lines (lyrics) 1015–1055 and 1065–1072(–1097, 74 lines, 303 words \approx 1/5 of the lines in our sample) in this play are parodically adapted by Aristophanes from Euripides' lost *Andromeda*. Aristophanes' pastiche is so close, it turns out, to Euripides' original, that a sensitive attribution model such as *SVM* remains ambivalent about which of the two poets the piece should be assigned to.³⁵

To test the *trace(s)* in *Seven* we employed trigrams and tetragrams occurring 7 to 300 times in the training corpus; since these were shown to be the most robust in the whole validation process. The attribution results for trigrams and tetragrams occurring 7 to 300 times in the training corpus confirm the results of the *CNG* model concerning the controversial parts, and are given in the following table.

³⁵ See Manousakis 2020: 208 for the relevant analysis of a 21-line (125 words) piece from Hesiod's *Theog.*

³⁴ For Aristophanic paratragedy see Silk 1993 and 2000: 42–97 (Aristophanes' engagement with comedy and tragedy); Rosen 2005: 255–61; Farmer 2017. See also Robson 2009: 103–19. Paratragedy in Old Comedy is not an exclusively Aristophanic trait. See Miles 2009 for paratragedy in Strattis. For the common and uncommon ground between the two genres see further, Nelson 2016; Jendza 2020.

³⁵ See, in detail, Manousakis 2020: 210–12. For the *Andromeda* parody in *Thesm.*, see also Sells 2019: 167–75.

Table 3: SVM Results for the Suspect Passages

<i>Part of the text</i>	<i>Attributed to</i>	<i>Degree of Affinity</i>	<i>Following Author</i>	<i>Degree of Affinity</i>
Part 17: lines 801–850	Aeschylus	≈ 70%	Sophocles	≈ 25%
Part 18: lines 851–900	Sophocles	≈ 55%	Aeschylus	≈ 40%
Part 20: lines 951–1000	Aeschylus	≈ 70%	Sophocles	≈ 20%
Parts 21 & 22: lines 1001–1078	Sophocles	≈ 85%	Aeschylus	≈ 10%

That is, parts 21 and 22—*Seven* 1001–1078—of the segregated analysis are both attributed to Sophocles with ≈ 85% degree of affinity on average, and to Aeschylus with ≈ 10%. The remaining parts, lines 901–950 (part 19) from the concluding choral dirge among them, are all attributed to Aeschylus with ≈ 90% degree of affinity on average, save for parts 17, 18 (containing the anapaestic introduction of the sisters: 861–874) and 20. The results show that the final part, Antigone confronting the herald about Polyneices’ burial, must be an interpolation in the Aeschylean text. The anapaestic introduction of the two sisters must also, unsurprisingly, be an interpolation; possibly by the same hand as the herald scene and plausibly triggered by the same dramatic necessity. Furthermore, according to the SVM model results, the messenger’s speech and part of the lament succeeding it, parts 17, 18 and 20, may have suffered minor alterations at the hands of an author other than Aeschylus.

III. *n*-gram Tracing

“On the 21st of November 1864, only 5 months before he was assassinated, Abraham Lincoln, the 16th President of the USA, sent a short letter of condolence to Lydia Bixby of Boston, a widow whose five sons were believed to have died in the Civil War. [. . .] Th[is] [l]etter would [. . .] become one of America’s most famous pieces of correspondence, praised for its sentiment and style and counted among Lincoln’s greatest texts” (Grieve, Clarke, Chiang, Gideon, Heini, Nini, and Waibel 2019: 493). However, the authorship of this historic note, consisting of 139

words, has been long disputed, with John Hay, Lincoln's young assistant and Secretary of State under McKinley and Roosevelt, being a plausible alternative. There are interesting (traditional) stylistic arguments on both sides, yet none of them gets even close to being conclusive. In a 2019 study, Grieve, Clarke, Chiang, Gideon, Heini, Nini, and Waibel, despite its rather discouraging shortness, managed to show that the disputed letter was actually written by Hay, not Lincoln. To do so, these scholars devised a new technique for short-text attribution: *n-gram tracing*. In its general outlines the process "involves first extracting all sequences of linguistic forms (i.e. characters and words) that occur in the questioned document, and then finding the possible author who uses the highest percentage of these forms" (see Grieve, Clarke, Chiang, Gideon, Heini, Nini, and Waibel 2019: 496).

The *n-gram tracing* algorithm takes character and word *n-grams* as its input. In *n-gram tracing* the percentage of all *n-grams* occurring in a questioned text and also at least once in a candidate author's writing sample is calculated. This is done for all candidates, and the text is then attributed to the author whose writing sample contains the highest percentage of *n-grams* in the questioned document. By extracting the complete number of *n-grams* occurring in a suspect text one obtains a broad and unbiased feature set. In actual fact, the algorithm measures the percentage of the *n-gram* types found in the suspect document that also occur at least once in equal-sized, to the total number of features, samples from each candidate author: only the presence or absence of these *n-grams* in the suspect document and the candidate author's writing samples is considered, not their relative frequencies, so as to avoid extracting relative frequencies from very short texts. To make sure the results will not be dependent on limited sampling, the analysis can be repeated on many different random sequences of texts, and the average cumulative percentages of *n-grams* detected there can be calculated and plotted at regular intervals of the total features detected. The controversial document is attributed to the candidate author with the highest *overlap coefficient*: measuring similarity between two sets by dividing the size of their intersection (number of shared features) by the size of the smaller of the two sets (its total number of features).

Overall, *n-gram tracing* was shown, through a *leave-one-out-cross-validation* approach, to be effective in telling apart the writings of Lincoln and Hay with a very high degree of accuracy. This evaluation was based on twenty-five different *n-gram* types, including one- to five-word and

one- to twenty-character *n*-grams, and each analysis aggregated over ten randomized sequences of texts per author. The most efficacious *n*-gram types and values emerging from the analysis were character pentagrams to ten-grams—the best results obtained with character seven-grams and eight-grams—, while word *n*-grams were shown to be less accurate than character *n*-grams. At any rate, the results are remarkable given the meagre linguistic data. Bearing the above in mind, we employed *n*-gram tracing to ascertain which parts in the rather short disputed ending of *Seven* are by Aeschylus.

To begin, we used the 14 validation texts employed earlier for *CNG* and *SVM* methods to test the attribution accuracy of *n*-gram tracing in short ancient Greek dramatic texts. We used character six-grams and, as Grieve et al. 2019 did for sentences, we prevented the analysis from spanning metrical lines (minimum 15 syllables for trochees, 12 for iambs, approximately 10 for anapaests (non-lyric dimeters), and at times even fewer for the lyrics).³⁶ The affinity of the validation texts to those of secure authorship was calculated, with no overlap, on the basis of 150,000 characters from randomly selected metrical lines in the secure texts. The results showed that *n*-gram tracing is a very effective model for accurately attributing short texts of ancient Greek drama. All 14 validation pieces were correctly attributed. Yet, for reasons previously explained, the lyrics in Aristophanes' *Thesmophoriazousae* (plotted in fig. 1: the upper line stands for the attribution author, the line below that for the author second in line, etc.), even though correctly attributed to the comic poet, were also distinctly connected to Euripides' *trace*. Hence, one can suggest with some confidence that *n*-gram tracing, as also *SVM*, is effective not only in rightly attributing short ancient Greek dramatic texts, but also in further detecting rather latent authorial ties between these texts.

Our next step was to apply *n*-gram tracing, maintaining the *n*-gram type and value successfully employed in the validation tests, to the three suspect passages constituting the ending of *Seven*. The results of this analysis were in line with both the *CNG* and *SVM* results. That is, lines 822–831, 832–860, and 875–1004 (plotted in fig. 3 below)—the choral lament for the princes save the introduction of their sisters—are primarily

³⁶ Technically, lyric lines are the most challenging to determine in any automated analysis due to period-end. Yet even when used as laid out in a particular standard edition (e.g. West 1990 for Aeschylus), attribution results are shown to remain the same.

authorially associated with Aeschylus, while Sophocles is the second most likely candidate for the authorship of the piece. Still, there is no way whatsoever to ascertain, based on this or any other current authorship attribution model, if 996–997, the two personalized (and hence attachable to Antigone and Ismene) suspect dochmiac lines in the lament (πρὸ πάντων δ' ἐμοί. / καὶ τὸ πρόσω γ' ἐμοί.) are by Aeschylus' hand. Lines 861–874 and 1005–1078 (plotted in fig. 4 below)—the introduction of Ismene and Antigone, and the latter's confrontation with the herald—are distinctly associated with Sophocles' authorial *trace*. Lines 792–821 (plotted in fig. 2 below)—the messenger's speech and interaction with the Chorus—are primarily associated with Sophocles' *trace*, while Aeschylus' closely follows. This does not mean, of course, as is also the case with the results of all attribution models employed in the present study, that Sophocles (or Euripides etc.) composed the final part of *Seven*. It only means that this is the best solution the algorithmic model(s) could find to the problem of having to (definitely) ascribe the suspect passages to some dramatist in the training corpus.

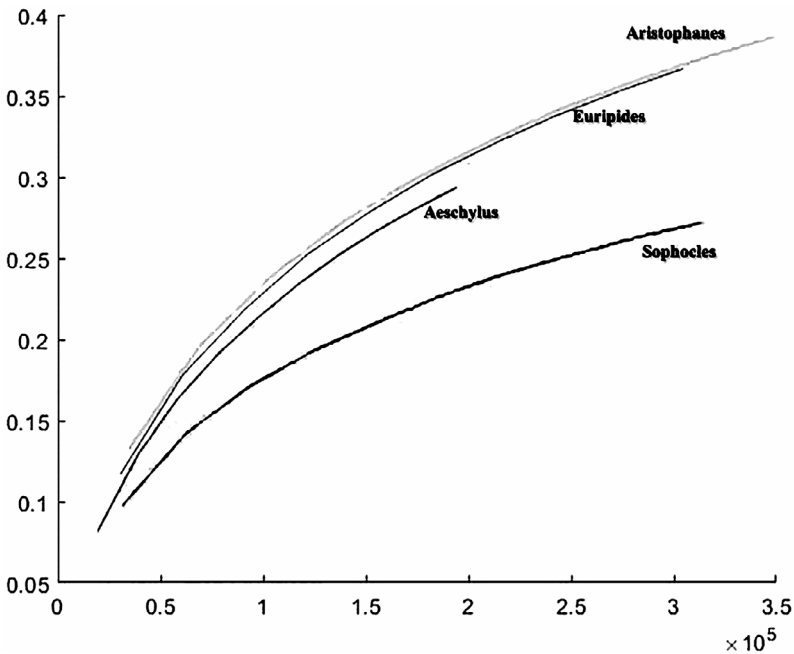


Figure 1: Character six-grams analysis of the lyrics in *Thesmophoriazousae*
 x-axis: total *n*-grams, y-axis: ratio of common *n*-grams

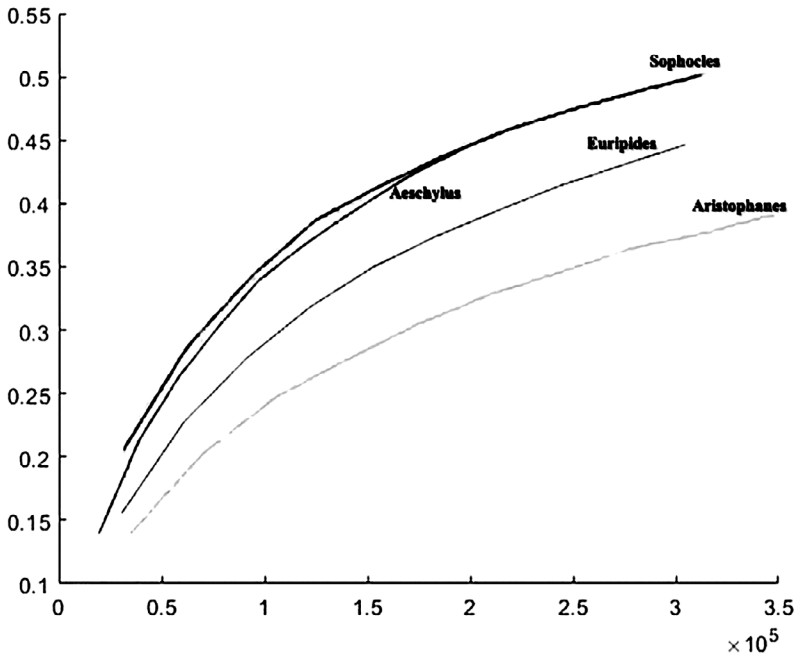


Figure 2: Character six-grams analysis of *Seven* 792–821
 x-axis: total n -grams, y-axis: ratio of common n -grams

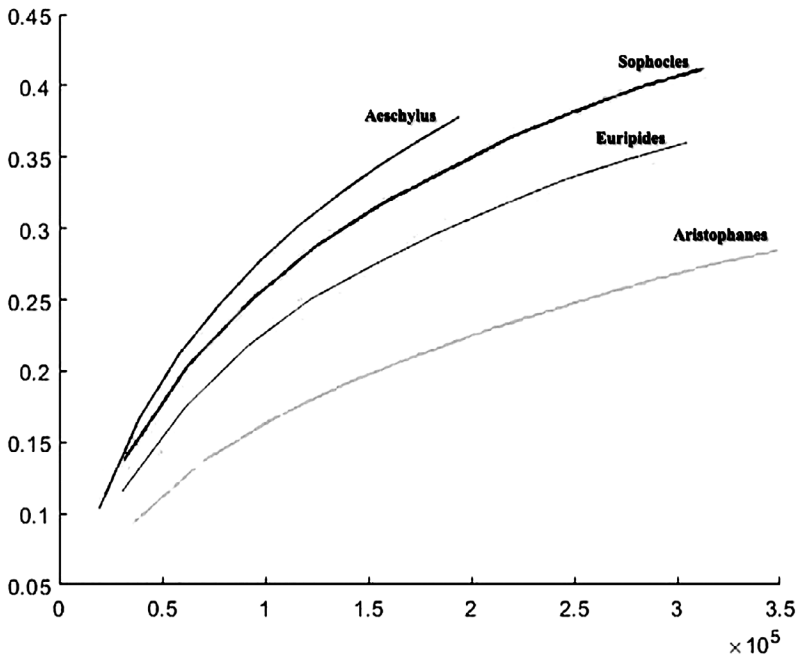


Figure 3: Character six-grams analysis of *Seven* 822–831, 832–860, & 875–1004
 x-axis: total n -grams, y-axis: ratio of common n -grams

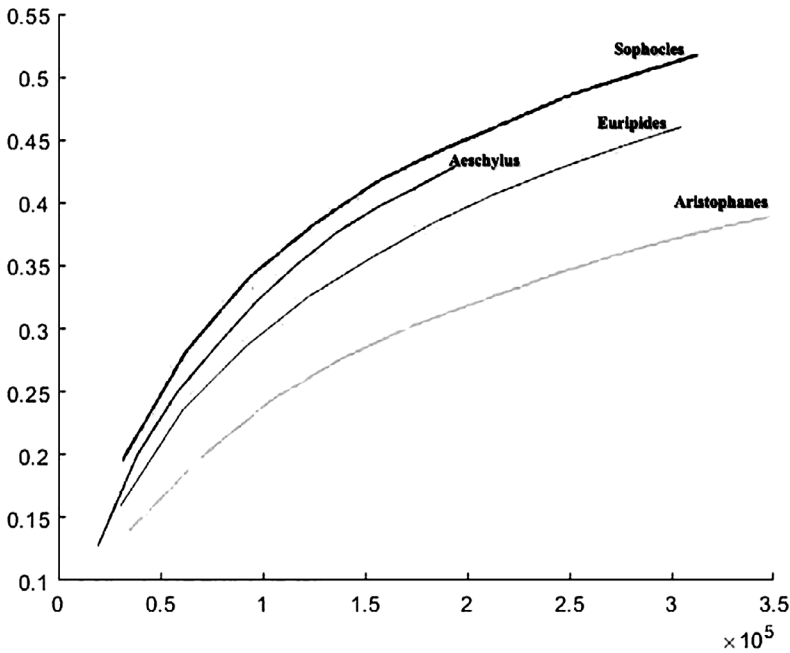


Figure 4: Character six-grams analysis of *Seven* 861–874 & 1005–1078
 x-axis: total *n*-grams, y-axis: ratio of common *n*-grams

IV. Conclusions

The present (non-traditional) *authorship analysis* study concerning the ending of *Seven* seems to be reinforcing the philological consensus that this part of the text is a later adaptation. It is shown that character *n*-grams, studied with *CNG*, *SVM*, and *n*-gram tracing, indicate *trace* inconsistencies in the passage under discussion, in all likelihood tied to tampering and interpolation. As Sommerstein (2010: 93) observes,

The original [Aeschylean] play, then, [might have] ended as follows: choral reflections on the news of the brothers' death (822–847); arrival of the two bodies (848–860); choral lament (875–960); antiphonal dirge for two solo voices and Chorus (961–1004—except for 996–997—plus a few further lines now lost); *exeunt omnes*. [. . .] [An adaptation of this play] is not surprising, [especially] if [*Seven* of 467 BCE] was thought unlikely to appeal to later audiences, [. . .] being produced on its own, without the two preceding plays. [. . .] [S]ome later producer commissioned a poet of uneven talent to [possibly reconfigure

the messenger's speech,] write some anapaests to cover the entrance of Antigone and Ismene, a couple of lyric lines [engaging the sisters] to insert in the dirge, a confrontation between Antigone and a herald, and a split exit for the Chorus, with the object of spicing up the conclusion of the play with reminiscences of one of the most famous tragedies ever written, Sophocles' *Antigone*.

Luckily, this interpolator seems to have been rather reluctant to damage the text by doing major deletions, and it is unlikely that he has removed more than “some two hundred lines of Aeschylus” (Page 1934: 32).

If the ending of *Seven* as it has come down to us is introduced by someone other than Aeschylus, this significantly affects our knowledge about the *Antigone* tradition. “[N]o source prior to the fifth century ever mentions Antigone [. . .]. Both she and Ismene emerge at the end of [Aeschylus' *Seven*] to lament the loss of their brothers, and this is [their] first appearance in literature or art” (see Gantz 1993: 519–20). If the ending of *Seven* is a later, updating adaptation, then Sophocles' *Antigone* “is the first surviving work of literature to treat Antigone's action in defying Creon's prohibition in order to bury Polyneices [. . .] [Yet] there must be a strong suspicion that this is because it was in fact the first work ever to do so” (see Cairns 2016: 9). Hence, the ending of *Seven* should be “of interest for us [. . .] as evidence for the “reception” of [*Antigone*], not for pre-Sophoclean treatments of the myth” (see Griffith 1999: 7). About who may have composed the “Aeschylean” version of *Antigone*, very little (if anything) could be said. He may have been “a fourth-century Athenian” (see Barrett 2007: 323)—and it is also not impossible that, as Martin West ingeniously suggested, he was none other than the author of *Prometheus Bound*.³⁷

The final part of *Seven* as it came down to us is indeed, in all likelihood, a later reworking of the original tragedy, crafted to bring the Aeschylean plot into line with the Labdacids myth as it was used predominantly by Sophocles in *Antigone*, and also possibly by Euripides in *Phoenissae* (see note 7 here). As for the language of the extant ending, one cannot help but notice, following Barrett 2007: 323, that the “linguistic

³⁷ See West 2000: 352, suggesting that “the points of similarity between the [*Sept.*] passages and [*Pr.*] seem to be more than fortuitous.” This claim can now be supported by the fact that, as Manousakis 2020: 198–221 shows in detail, the *n-grams trace* of *Pr.* as a whole is markedly tied to Sophocles, not Aeschylus, both by *CNG* and *SVM*, as is also shown in the present paper to be the case with the ending of *Sept.*

incompetence” of the author “manifest[s] itself not in solecisms but in straining the language in what he mistakenly supposes to be the manner of Aeschylus, or in falling flat where Aeschylus would have risen.” The text of *Seven* is linguistically, dramaturgically and also scenically (the presence of an(y) Antigone on stage is attractive in modern theatrical terms, yet an irrelevant Antigone, such as the one in *Seven*, quickly turns out to be nothing more than a flash in the pan) far more coherent and effective if the sisters and the herald are altogether removed in performance. This suggestion is now based not only on connoisseurship but also on hard data.

ACADEMY OF ATHENS AND UNIVERSITY OF THE AEGEAN

nmanousakis@academyofathens.gr

stamatatos@aegean.gr

Works Cited

- Aggarwal, C.C. 2015. “Instance-Based Learning: A Survey.” In C.C. Aggarwal, ed. *Data Classification: Algorithms and Applications*. Boca Raton. 157–85.
- Allison, T.B. 2003. *Aeschylean Stylistics: A Study of Linguistic Variation*. PhD Dissertation—University of Michigan.
- Armantrout, G.L. 1990. *The Seven Against Thebes in Greek Art*. PhD Dissertation—University of Michigan.
- Barrett, W.S. 2007. “Seven Against Thebes: The Final Scene.” In M. L. West, ed. *Greek Lyric, Tragedy, and Textual Criticism: Collected Papers [of W.S. Barrett]*. Oxford. 322–50.
- Berman, D.W. 2007. *Myth and Culture in Aeschylus’ Seven Against Thebes*. Rome.
- Brown, A.L. 1976. “The End of the Seven against Thebes.” *CQ* 26: 206–19.
- Bramer, M. 2016. *Principles of Data Mining*. London.
- Cairns, D. 2016. *Sophocles: Antigone*. London.
- Campbell, C. and Ying, Y. 2011. *Learning with Support Vector Machines*. San Rafael.
- Can, E.F., Can, F., Duygulu, P., Kalpakli, M. 2012. “Automatic Categorization of Ottoman Literary Texts by Poet and Time Period.” In E. Gelenbe, R. Lent, and G. Sakellari, eds. *Computer and Information Sciences II: 26th International Symposium on Computer and Information Sciences*. London. 51–57.
- Catenacci, C. 2004. “Realtà e immaginario negli scudi dei *Sette contro Tebe* di Eschilo.” In A. Bernardini, ed. *La città di Argo: Mito, storia, tradizioni poetiche: Atti del Convegno Internazionale—Urbino 13–14–15 Giugno 2002*. Rome. 163–76.
- Clarke, B., Fokoué, E., and Zhang, H.H. 2009. *Principles and Theory for Data Mining and Machine Learning*. Dordrecht.
- Conacher, D.J. 1996. *Aeschylus: The Earlier Plays and Related Studies*. Toronto.
- Dawe, R.D. 1976. “The End of the *Seven against Thebes*.” *CQ* 17: 16–28.

- . 1978. "The End of the *Seven against Thebes* Yet again." In R. D. Dawe, J. Diggle, and P.E. Easterling, eds. *Dionysiaca: Nine Studies in Greek Poetry*. Cambridge. 87–103.
- De Bruyne, S. 2010. *Process, Data and Classifier Models for Accessible Supervised Classification Problem Solving*. Brussels.
- Diggle, J. 1994. *Euripides Fabulae III: Helena, Phoenissae, Orestes, Bacchae, Iphigenia Aulidensis, Rhesus*. Oxford.
- Dik, H. 2007. *Word Order in Greek Tragic Dialogue*. Oxford.
- Dover, K.J. 1968. *Aristophanes: Clouds*. Oxford.
- Else, G.F. 1965. *The Origin and Early Form of Greek Tragedy*. Cambridge.
- Fantuzzi, M. 2020. *The Rhesus Attributed to Euripides*. Cambridge.
- Farmer, M. 2017. *Tragedy on the Comic Stage*. Oxford.
- Flintoff, E. 1980. "The Ending of the *Seven against Thebes*." *Mnemosyne* 33: 244–71.
- Fraenkel, E. 1964. "Zum Schluß der *Sieben gegen Theben*." *MH* 21: 58–64.
- Fries, A. 2014. *Pseudo-Euripides: Rhesus*. Berlin.
- Gantz, T. 1995. *Greek Early Myth: A Guide to Literary and Artistic Sources*. Baltimore MD.
- Garvie, A.F. 2006². *Aeschylus' Supplikes: Play and Trilogy*. Bristol.
- Grieve, J., Clarke I., Chiang E., Gideon H., Heini A., Nini A., and Waibel E. 2019. "Attributing the *Bixby Letter* using n-gram tracing." *Digital Scholarship in the Humanities (DSH)* 34: 493–512.
- Griffith, M. 1977. *The Authenticity of Prometheus Bound*. Cambridge.
- . 1999. *Sophocles: Antigone*. Cambridge.
- Grilli, A. 2018. "The Semiotic Basis of Politics in *Seven Against Thebes*." *Skenè – Journal of Theatre and Drama Studies* 4: 55–86.
- Hubbard, Th.K. 1986. "Parabolic Self-Criticism and the Two Versions of Aristophanes' *Clouds*." *ClAnt* 5: 182–97.
- Hutchinson, G.O. 1985. *Aeschylus: Septem Contra Thebas*. Oxford.
- Jendza, C. 2020. *Paracomedie: Appropriations of Comedy in Greek Tragedy*. Oxford.
- Joachims, T. 2002. *Learning to Classify Text Using Support Vector Machines*. Boston.
- Juola, P. 2006. "Authorship Attribution." *Foundation and Trends in Information Retrieval* 1: 233–334.
- Kaur, N. and Verma, A. 2015. "Authorship Attribution of Punjabi Poetry using SVM Classifier." *International Journal of Advanced Research in Computer Science and Software Engineering* 5: 1055–1061.
- Kešelj, V., Peng, F., Cercone, N., and Thomas, C. 2003. "n-gram-based Author Profiles for Authorship Attribution." In V. Kešelj and T. Endo, eds. *Proceedings of the 6th Pacific Association for Computational Linguistics Conference*. Halifax. 255–64.
- Koenigsbeck, J. 1891. *De Septem contra Thebas exitu*. Danzig.
- Kopff, E.C. 1990. "The Date of Aristophanes, *Nubes* II." *AJP* 111: 318–29.
- Kovacs, D. 2002. *Euripides: Helen, Phoenician Women, Orestes*. Cambridge, MA.
- Liapis, V. 2012. *A Commentary on the Rhesus Attributed to Euripides*. Oxford.
- Lloyd-Jones, H. 1959. "The End of the *Seven against Thebes*." *CQ* 9: 80–115.
- Lupaş L. and Petre L. 1981. *Commentaire aux Sept Contre Thèbes d'Eschyle*. Bucharest.

- Manousakis, N. 2020. *Prometheus Bound: A Separate Authorial Trace in the Aeschylean Corpus*. Berlin.
- Manousakis, N. and Stamatatos, E. 2018. "Devising *Rhesus*: A Strange Collaboration between Aeschylus and Euripides." *Digital Scholarship in the Humanities (DSH)* 33: 347–61.
- Mastrorarde, D. J. 1994. *Euripides: Phoenissae*. Cambridge.
- . 2002. *Euripides: Medea*. Cambridge.
- Meineck, P. and Storey, I.C. 2000. *Aristophanes: Clouds*. Indianapolis.
- Mellon, P. S. 1974. *The Ending of Aeschylus' Seven Against Thebes and its Relation to Sophocles' Antigone, and Euripides' Phoenissae*. PhD Dissertation—Stanford University.
- Miles, S.N. 2009. *Strattis, Tragedy, and Comedy*. PhD Dissertation—University of Nottingham.
- Murray, G. 1955. *Aeschyli Septem quae Supersunt Tragoedias*. Oxford.
- Natanblut E. 2003. *The Blessedness of Thebes and the Accursed Family of Oedipus: an Interpretation of Euripides' Phoenissae*. PhD Dissertation—University of Virginia.
- . 2005. *The Seven Against Thebes Myth in Greek Tragedy*. Montreal.
- Nelson, S. 2016. *Aristophanes and His Tragic Muse: Comedy, Tragedy and the Polis in 5th century Athens*. Leiden.
- Nicolaus, P. 1967. *Die Frage nach der Echtheit der Schlusszene von Aischylos' Sieben gegen Theben*. Tübingen.
- Orwin, Cl. 1980. "Feminine Justice: The End of the *Seven against Thebes*." *CP* 75: 187–96.
- Page, D. L. 1934. *Actors' Interpolations in Greek Tragedy—Studied with Special Reference to Euripides' Iphigenia in Aulis*. Oxford.
- . 1972. *Aeschyli Septem quae Supersunt Tragoedias*. Oxford.
- Platt, A. 1912. "The Last Scene of the *Seven against Thebes*." *CR* 26: 141–44.
- Robson, J. 2009. *Aristophanes: An Introduction*. London.
- Rosen, R.M. 2005. "Aristophanes, Old Comedy, and Greek Tragedy." In R. Bushnell, ed. *A Companion to Tragedy*. Malden, MA. 251–68.
- Rutherford, R.B. 2012. *Greek Tragic Style: Form, Language, and Interpretation*. Cambridge.
- Ryzman, M. 1983. "The authenticity of the final scene of Aeschylus *Septem contra Thebas*—A psychological perspective." *Revue Belge de Philologie et d'Histoire* 61: 87–115.
- Sapkota, U., Bethard, S., Montes, M., and Solorio, T. 2015. "Not All Character *n-grams* are Created Equal: A Study in Authorship Attribution." In R. Mihalcea, J. Chai, and A. Sarkar, eds. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver. 93–102.
- Schein S.L. 1979. *The Iambic Trimeter in Aeschylus and Sophocles—A Study in Metrical Form*. Leiden.
- Schölkopf, B. and Smola, A. 2002. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge.
- Sells, D. 2019. *Parody, Politics and the Populace in Greek Old Comedy*. London.
- Silk, M.S. 1993. "Aristophanic Paratragedy." In A.H. Sommerstein, S. Halliwell, J. Henderson, and B. Zimmermann, eds. *Tragedy, Comedy and the Polis: Papers from the Greek Drama Conference (Nottingham 18–20 July 1990)*. Bari. 477–504.

- . 1996. “Tragic Language: The Greek Tragedians and Shakespeare.” In M.S. Silk, ed. *Tragedy and the Tragic: Greek Theatre and Beyond*. Oxford. 458–96.
- . 2000. *Aristophanes and the Definition of Comedy*. Oxford.
- Sommerstein, A. H. 2008. *Aeschylus: Persians, Seven Against Thebes, Suppliants*, Prometheus Bound. Cambridge MA.
- . 2010. *Aeschylean Tragedy*. London.
- Stamatatos, E. 2007. “Author Identification Using Imbalanced and Limited Training Texts.” In A.-M. Tjoa and R. Wagner, eds. *Proceedings of the 18th International Conference on Database and Expert Systems Applications: 4th International Workshop on Text-based Information Retrieval (Regensburg)*. Los Alamitos. 237–41.
- . 2008. “Author Identification: Using Text Sampling to Handle the Class Imbalance Problem.” *Information Processing and Management* 44: 790–99.
- . 2009. “A Survey of Modern Authorship Attribution Methods.” *Journal of the Association for Information Science and Technology (JASIST)* 60: 538–56.
- . 2013. “On the Robustness of Authorship Attribution Based on Character *n*-gram Features.” *Journal of Law and Policy* 21: 421–439.
- . 2018. “Masking Topic-related Information to Enhance Authorship Attribution.” *Journal of the Association for Information Science and Technology (JASIST)* 69: 461–73.
- Stańczyk, U. 2016. “The Class Imbalance Problem in Construction of Training Datasets for Authorship Attribution.” In A. Gruca, A. Brachman, S. Kozielski, and T. Czachórski, eds. *Man-Machine Interactions 4: 4th International Conference on Man-Machine Interactions (Kocierz Pass)*. Cham. 535–47.
- Stoean, C. and Stoean, R. 2014. *Support Vector Machines and Evolutionary Algorithms for Classification: Single or Together?* Cham.
- Storey, I.C. 1993. “The Dates of Aristophanes’ *Clouds* II and Eupolis’ *Baptai*: A Reply to E.C. Kopff, The Date of Aristophanes’ *Nubes* II.” *AJP* 114: 71–84.
- Taplin O. 1972. “Aeschylean Silences and Silences in Aeschylus.” *HSCP* 76: 57–97.
- . 1977. *The Stagecraft of Aeschylus: The Dramatic Use of Exits and Entrances in Greek Tragedy*. Oxford.
- Thalmann, W.G. 1978. *Dramatic Art in Aeschylus’s Seven Against Thebes*. New Haven.
- Tsantsanoglou, K. 2010. “The Lament of *Seven Against Thebes* and the Role of Telestes.” In S. Tsitsirides, ed. *Παραχորήγημα: μελετήματα προς τιμήν του καθηγητή Γρηγόρη Μ. Σηφάκη*. Heraklion. 31–56.
- Verrall A. W. 1887. *The Seven Against Thebes of Aeschylus*. London.
- Vidal-Naquet, P. 1990. “The Shields of the Heroes: Essay on the Central Scene of the *Seven Against Thebes*.” In J. P. Vernant and P. Vidal-Naquet, eds. *Myth and Tragedy in Ancient Greece* (tr. J. Lloyd). New York. 273–300.
- Von der Mühl, P. 1964. “Der Zweikampf der Oidipussöhne im dritten Epeisodion der *Septem*.” *MH* 21: 225–27.
- West, M.L. 1990. *Aeschyli Tragoediae—cum Incerti Poetae Prometheo*. Stuttgart.
- . 2000. “*Iliad* and *Aethiopsis* on the Stage: Aeschylus and Son.” *CQ* 50: 338–52.
- Willi, A. 2011. “Homeric Language.” In M. Finkelberg, ed. *The Homer Encyclopedia* (Vol. II). Chichester. 458–64.

- Willink C.W. 1968. "A Problem in Aeschylus' *Septem*." *CQ* 18: 4–10.
- . 1986. *Euripides: Orestes*. Oxford.
- Winnington-Ingram, R. P. 1977. "*Septem Contra Thebas*." *YCS* 25: 1–45.
- Zeitlin, F. 1982. *Under the Sign of the Shield: Semiotics and Aeschylus' Seven Against Thebes*. Rome.
- Zhang, X. 2011. "Support Vector Machines." In Sammut C. and Webb G.I., eds. *Encyclopedia of Machine Learning*. New York NY. 1214–20.