

RAIJA VAINIO, REIMA VÄLIMÄKI, ALEKSI VESANTO, ANNI HELLA, FILIP
GINTER, MARJO KAARTINEN, TEEMU IMMONEN¹

RECONSIDERING AUTHORSHIP IN THE CICERONIAN CORPUS
THROUGH COMPUTATIONAL AUTHORSHIP ATTRIBUTION

The authorship of some texts related to Cicero or traditionally attributed to him has puzzled scholars for centuries. The most famous of these texts is *Rhetorica ad C. Herennium*, whose removal from the Ciceronian corpus was proposed as early as the fifteenth century. The other two (minor) texts are *Commentariolum petitionis*, usually attributed to Marcus Cicero's younger brother Quintus, and most recently *De optimo genere oratorum*. Sir Ronald Syme stated on the authenticity of old texts: «In every age the principal criteria of authenticity are the stylistic and the historical. They do not always bring certainty, for we do not know enough about either style or history. If a different approach can be devised, or a subsidiary method, so much the better»². In recent years, digital methods have offered promising results for the reattribution of classical texts. M. Kestemont, J.A. Stover and others have worked with some ancient Latin texts³, but although a computational analysis by R. Forsyth, D. Holmes, and E. Tse confirmed the consensus that *Consolatio Ciceronis* is indeed a sixteenth-century forgery⁴, until now these methods have had only a limited impact on the Ciceronian corpus itself. We attempt to take on this task with today's highly advanced computational methods and the use of high performance computing (CSC supercomputer, Kajaani, Finland). After a brief account on

¹ The main responsibility of the historical background and interpreting the results (chapters 1-3 and 5) lies on Raija Vainio, and that of the methods and respective parts in the results (ch. 4) on Aleksi Vesanto and Filip Ginter. All the text has been commented upon, revised and supplemented by Reima Välimäki, Anni Hella, Marjo Kaartinen and Teemu Immonen. Reima Välimäki was responsible for making the computational methods understandable for humanists. This study has been funded by the Academy of Finland, Academy Programme DIGIHUM 2015-2019: consortium Profiling Premodern Authors (PROPRAU), project number 293024.

² Syme 1947, 198.

³ Tweedie et al. 1998; Gurney-Gurney 1998; Forsyth et al. 1999; Kestemont 2012; Kestemont et al. 2015; Deploige-Moens 2016; Stover et al. 2016; Kestemont et al. 2016; Stover-Kestemont 2016a and 2016b; Richards et al. 2016; De Gussem 2017.

⁴ Forsyth et al. 1999.

the history of the texts, we will describe the methods and finally discuss our results in the historical context of the texts.

1. *Rhetorica ad C. Herennium*

Rhetorica ad C. Herennium, nowadays generally dated to the 80s BCE, was found in the fourth century. It was addressed to a Gaius Herennius but it is not clear which of the Herennii he was or even whether he was a Herennius at all. No reference to the text by any learned Roman is known of from the previous four centuries. Because of a close resemblance to Cicero's works, especially *De inventione*, this text was thought to be written by him: at the end of the fourth century Jerome referred to it as Cicero's⁵. In the twelfth century it was believed that Cicero had written the four books of the *Rhetorica* as an improved version of his early work *De inventione*. This provided an explanation as to why he would have authored two texts which have very similar content, except that the *Rhetorica* is larger⁶. However, there is no reference to Cicero in the *Rhetorica*, or to the *Rhetorica* in *De inventione* or in any of his other texts, as one would expect if he were the author. Because of the supposedly common origin, *De inventione* and the *Rhetorica* have usually been preserved in the same manuscripts, the former being called *Rhetorica prima* and the latter *Rhetorica secunda*. In the thirteenth century these became known as *Rhetorica vetus* and *nova* respectively⁷.

In the 1460s, Angelo Decembrio⁸ expressed doubts about the *Rhetorica*'s attribution to Cicero, and in 1491 Raffaele Regio was convinced of the false attribution, although he did not want to declare it in public, probably because of the sensitive nature of the question at that time⁹. He found stylistic differences between Cicero's works and the *Rhetorica* and

⁵ He gives the information twice: first in the prologue of *Commentarii in Abdiam*, and a few years later in *Contra Rufinum* 1, 16 (cf. Marx 1894, 1-8; Achard 1989, xiv).

⁶ Kennedy 1972, 113: «The first two and a half books of *ad Herennium* cover approximately the same ground as *de inventione*, though the organization and many details differ». See also Corbeill 2002, 31-34. For the role of *Rhetorica ad Herennium* in the Middle Ages, see Cox-Ward 2011.

⁷ Marx 1894, 52-53; Achard 1989, xiv.

⁸ *Angeli Decembrii de politia litteraria* was published as late as 1540 in Augsburg.

⁹ *Raphaelis Regii Quaestio. Vtrum Ars rhetorica ad Herenium falso Ciceroni inscribatur*, Venetiis, published 1492. Monfasani 1987, 112-115 has convincingly shown that Lorenzo Valla never denied the Ciceronian authorship, contrary to the general belief. See also Murphy-Winterbottom 1999, 77-80.

wondered why Quintilian, despite being a great admirer of Cicero, did not seem to know this text at all. Raffaele also proposed several other possible authors, one of which was Cornificius, mentioned by Quintilian in several passages¹⁰. Piero Vettori was certain of Cornificius's authorship in 1553¹¹. Nevertheless, in 1851 R. Klotz still placed the *Rhetorica* among Cicero's works. In addition, besides Cornificius, a number of possible authors have subsequently been suggested¹². The common feature of these candidates is that we do not possess any of their texts to which the *Rhetorica* could be compared.

Remarkable similarities on the one hand and noteworthy differences on the other have perplexed those assessing the relationship between the *Rhetorica* and *De inventione*¹³. At first, the author of the *Rhetorica* was assumed to have used Cicero as his source, then it was thought to have been the other way round, and nowadays most scholars support the theory of a common source, although they propose different routes to the texts we possess¹⁴. It is clear that these texts have a close connection as there are several passages with almost identical wording. G. Thiele argued that as students¹⁵ both authors had been taught by the same Latin teacher who had used a Greek *ars rhetorica*, but Cicero has made more additions to the text and used sources other than those of the author of the *Rhetorica*¹⁶. F. Marx described the nature of Cicero's *De inventione* as

¹⁰ Marx 1894, 61-69; Murphy-Winterbottom 1999, 81.

¹¹ *Petri Victorii Variarum lectionum libri XXV*, Florentiae 1553. Cf. also Calboli 1969, 3.

¹² Achard 1985, 57 n. 9 has listed suggestions offered: Cicero's teachers L. Aelius Stilo and M. Antonius Gniphio, Cicero's son, Cicero's freedmen M. Tullius Tiro and M. Tullius Laurea, L. Ateius Praetextatus (during the reign of Augustus), Papirius Fabianus (of Tiberius), Verginius Flavius (of Nero), Timolaus (of Aurelian). Achard himself brought out rhetorician L. Plotius Gallus, whom Cicero knew in his childhood (Achard 1985, 61, and 1989, xxiii), and L. Hirtuleius (Achard 1985, 65-68). Herrmann 1980 has proposed L. Annaeus Cornutus, and dated the *Rhetorica* to 57 CE.

¹³ See e.g. Corbeill 2002, 31-34.

¹⁴ Kennedy 1972, 126-128.

¹⁵ Cicero himself (*de orat.* 1, 5) refers to his early writing as «the unfinished and crude essays, which slipped out of the notebooks of my boyhood, or rather of my youth» (transl. by E.W. Sutton 1942 in Loeb CL). Cf. Kennedy 1972, 107. Cf. also Quintilian *inst.* 3, 6, 59: «But I have already mentioned what Cicero's own opinion about these books was. They were school lectures brought together, as it were, to form the treatises which he had composed as a young man; if there is anything wrong with them, it is the fault of his instructor, who may have been influenced either by the fact that, in this context, Hermagoras put examples from Legal Questions first, or else by the consideration that the Greeks call interpreters of law *pragmatikoi*» (transl. by D.A. Russell 2001 in Loeb CL).

¹⁶ Thiele 1889, 91-95. His view was supported by Kennedy 1972, 128-137, who thought that the author of the *Rhetorica* was a student of the same teacher a few years later than Cicero.

«copied or written down at dictation» and developed a complicated theory of two rhetoricians on Rhodes with different views. He hypothesized that the Latin teacher of the author of the *Rhetorica* had taken down the words of the earlier rhetorician almost verbatim, whereas Cicero's Latin teacher had used the more recent one¹⁷. His view was for the most part supported by H. Caplan, whereas G. Herbolzheimer and more recently G. Calboli have suggested that both used the same Latin manual, but Cicero altered the text more and additionally used the Greek rhetorician Hermagoras. J. Adamietz generally supported this view with the exception that he thinks Hermagoras was not used directly, his ideas being mediated through another Greek rhetorician. Without any information in our sources on their origin, these theories inevitably remain only speculation¹⁸. Although there is strong agreement among scholars that *Rhetorica ad Herennium* was not written by Cicero, we will use it as a test case for our analysis.

2. *Commentariolum petitionis*

In the later manuscript tradition, there is some confusion about the author of *Commentariolum petitionis*, and since the nineteenth century the question of its authorship has been a subject of heated discussion. The extant form of this text seems to be a letter from Quintus Cicero to his famous brother Marcus, since it has the typical greeting¹⁹ at the beginning, although some of the younger manuscripts, also considered inferior, seem to ascribe the text to Marcus Cicero²⁰. Since the «you» of the

¹⁷ Marx 1894, 78-84. Summary of his theory in Marx 1894, 161-162.

¹⁸ Caplan 1964, XXVIII-XXXII; Herbolzheimer 1926, 392-393; Calboli 1969, 26-29; Adamietz 1960, 96-97. For different theories, see Adamietz 1960, 1-7; Caplan 1964, XXVI-XXIX; Calboli 1969, 26-29; Achard 1989, XIV-XXXIV *passim*. Kennedy 1972, 117 has described Hermagoras's (flourished ca. 150 BCE) *stasis* theory as the biggest addition in Hellenistic rhetoric. For Hermagoras's role, see also Corbeill 2002, 29-30.

¹⁹ *Incipit commentarium consulatus petitionis. Quintus Marco fratri s(alutem) d(icit)*. «Here begins the handbook on how to campaign for the consulship. Quintus greets his brother Marcus» (transl. by Raija Vainio).

²⁰ For various readings in the manuscripts, see Bücheler 1869, 25; Nardo 1970, 162. Wikarjak 1966, 31-32 has noted that there are still a dozen instances in editions where the reading of these inferior manuscripts (*Lagomarsinianus* 38 and 117) has been preferred to the older ones. Likewise, he found it possible that especially the invective part of the text (*comm. pet.* 7-12), which he regarded worthy of Marcus Cicero's style, could indicate his participation in the writing.

text is undoubtedly Marcus Cicero — the text contains phrases such as *de petitione tua cogitanti* (*comm. pet.* 1) speaking of Marcus's campaign for the consulship (of the year 63) — he has been ruled out as a possible author.

A. Eussner in 1872 was the first to present doubts even about the authorship of Quintus Cicero. The argument proposed by him that has gained most support concerns several passages of the text which bear a close resemblance especially to Cicero's lost speech *In toga candida*, preserved partially by Asconius in his commentary from the first century CE²¹. Eussner's dissertation started a long debate for and against Quintus's authorship, combined with the question of the date of the text: if Quintus were the author, it would mean that Marcus had copied those passages of his speech from his younger and less experienced brother²². Many have found this scenario unacceptable²³, even though Quintus requested Marcus to make improvements to the text (*comm. pet.* 58)²⁴. J.P.V.D. Balsdon, however, remarked that the brothers shared their education in rhetoric²⁵, and warned about jumping to the conclusion that the question is inevitably about one text being an imitation of the other. As the chronological order of the composition of these texts and the direction of influence between them remain speculative, we will not extract these passages from the *Commentariolum* where they are an integral part of the text.

The question of authorship is, for obvious reasons, closely connected to that of the purpose of the writing, since it does not seem to be an ordinary private letter despite its conventional greeting²⁶. The author declares that his aim is to write at length all that has occurred to him regarding Cicero's campaign, and to bring together in one place and organize all the material that in real life seems disconnected

²¹ Similarities were noticed 150 years ago by Bücheler 1869, 8-9. Eussner 1872, 14-17 has also found resemblances with Cicero's speeches *Pro Murena* and *Pro Plancio*, as well as with his first letter to Quintus.

²² Eussner 1872, 4; Henderson 1950, 8 and 17.

²³ See e.g. Eussner 1872, 14; Gudeman 1894, 154; Nisbet 1961, 86. For summaries on different views, see Nardo 1970, 3-8; E. Deniaux in David et al. 1973, 248-256; Henderson 2002, 402-403.

²⁴ The request is not unusual, since we know that Quintus asked his brother to comment on his annals (Cic. *Att.* 2, 16, 4; cf. Cic. *ad Q. fr.* 2, 12, 4).

²⁵ Balsdon 1963, 250. But Fündling 2006 argues that the brothers did not have a common education in rhetoric; unfortunately he gives no evidence for his conclusion.

²⁶ Eussner 1872, 5; Henderson 1950, 8; Nisbet 1961, 84; Wikarjak 1966, 6; Nardo 1970, 30; E. Deniaux in David et al. 1973, 255-256.

and unspecified²⁷. At the end of the text he summarizes his intentions as follows:

Haec sunt quae putavi non melius scire me quam te sed facilius his tuis occupationibus colligere unum in locum posse et ad te perscripta mittere. Quae tametsi scripta ita sunt ut non ad omnes qui honores petant sed ad te proprie et ad hanc petitionem tuam valeant, tamen tu, si quid mutandum esse videbitur aut omnino tollendum, aut si quid erit praeteritum, velim hoc mihi dicas; volo enim hoc commentariolum petitionis haberi omni ratione perfectum (*comm. pet.* 58).

Here it is; I thought, not that I knew all this better than you, but that, considering how busy you are, I could more easily pull it together into one whole and send it to you in writing. Although it is written in such a way that it applies not to all who are seeking office but to you in particular and to this canvass, still, please tell me if you think that anything should be changed or struck out altogether, or if anything has been left out. For I want this handbook of electioneering to be considered perfect in every way (transl. by D.R. Shackleton Bailey 2002 in Loeb CL).

The *Commentariolum* has been characterized as a brochure or an essay, a handbook or manual, a treatise, a pamphlet, a report or memorandum, and a letter of instruction²⁸. D. Nardo has seen it as an open letter, although targeted at a restricted audience, belonging to Cicero's propaganda as a *homo novus*²⁹. According to those who have not accepted the text at face value and therefore regarded it as a later piece, it is either a rhetorical exercise or a forgery from the first century CE³⁰. M.C. Alexan-

²⁷ *Comm. pet.* 1, *Ad te perscribere ea quae mihi veniebant in mentem dies ac noctes de petitione tua cogitanti, non ut aliquid ex his novi addisceres, sed ut ea quae in re dispersa atque infinita viderentur esse ratione et distributione sub uno aspectu ponerentur.*

²⁸ E.g. Tyrrell 1877, 40; Wikarjak 1966, 12; Shackleton Bailey 2002, esp. 400-402; Syme 1947, 200; Henderson 1950, 397: «not a letter but a treatise»; Last 1932, 894; McDermott 1970, 385; Till 1962, 317 n. 5: «eine isagogische Studie in Briefform»; Sykutris 1962, 204: «Lehrbrief».

²⁹ Nardo 1970, e.g. 76, 112, 135-137. Cf. Clift 1945, 106; Wikarjak 1966, esp. 18-19. Lintott 1972, 218 agreed with Nardo in his review.

³⁰ Hendrickson 1892, 211-212, and again 1903, 78-82; Gudeman 1894, 154; Henderson 1950, 20-21; Watt 1958, 32 n. 2; Nisbet 1961, 85. One argument, used e.g. by Hendrickson 1892, 202-203, involves the tripartite division of the material, which is typical of rhetorical school exercises (cf. Quint. 3, 8, 15). Clift, however, has remarked that «elaborate division is [...] not adequate evidence of spuriousness, for the practice can be abundantly illustrated from Cicero's orations» (Clift 1945, 105 and n. 83, referring to *De imperio Cn. Pompei* and *Pro Quinctio*).

der, at the latest, sees it as a pseudepigraphic work that should be read ironically and that was written «to counter nostalgia for the old Republic»³¹. We will suggest that the *Commentariolum* is an authentic text written – at least for the most part – by Marcus Cicero.

3. De optimo genere oratorum

The short treatise *De optimo genere oratorum* has been considered a preface to Cicero's lost translation of two speeches by Greek masters, Demosthenes's *On the crown* and Aeschines's *Against Ctesiphon*³². The author gives the following information on the practices he has followed in his translation:

Converti enim ex Atticis duorum eloquentissimorum nobilissimas orationes inter seque contrarias, Aeschinis et Demosthenis; nec converti ut interpretes, sed ut orator, sentiis isdem et earum formis tamquam figuris, verbis ad nostram consuetudinem aptis. In quibus non verbum pro verbo necesse habui reddere, sed genus omne verborum vimque servavi (*opt. gen.* 14).

That is to say I translated the most famous orations of the two most eloquent Attic orators, Aeschines and Demosthenes, orations which they delivered against each other. And I did not translate them as an interpreter, but as an orator, keeping the same ideas and the forms, or as one might say, the "figures" of thought, but in language which conforms to our usage (transl. by H. M. Hubbell 1949 in Loeb CL).

Since there are neither traces of the actual translations nor any reference to them by Quintilian, it has been questioned whether they ever existed³³. They are first mentioned by Jerome (*epist.* 57, 5, 2-5) at the end of the fourth century and Sidonius Apollinaris (*epist.* 2, 9, 5) in the fifth. A. Dihle regarded it as likely that the preface has given rise to a conception of their existence in late antiquity and suggested that the *De optimo ge-*

³¹ Alexander 2018, 86 and 92-95.

³² Demosthenes's speech in defence of Ctesiphon (Περὶ τοῦ στεφάνου or *De corona*) and Aeschines's against him (Κατὰ Κτησιφώντος) in 330 BCE.

³³ So Hendrickson 1926, 122. — To be precise, the author does say, by using the perfect tense *converti*, that he **has** translated the speeches. Häfner 1928, 7-8 was convinced that in Cicero's time the writing of a preface was left till last. Quintilian mentions other translated texts in *inst.* 10, 5, 2, and Cicero's translations of Plato and Xenophon are among them, but not e.g. that of Aratus. Cf. also Rutilius Lupus below.

nerē should be removed from the Ciceronian corpus³⁴. Besides stylistic issues, e.g. «here and there an awkward style which Cicero would not use», he drew attention to some problems in the content³⁵. In addition, he interpreted Asconius's mention in his commentary on Cicero's speeches from the 50s CE, «a book which is assigned to Cicero»³⁶, to mean that it is spurious. On the other hand, D.H. Berry has shown that the prose rhythm of the *De optimo genere* supports Cicero's authorship³⁷. Regardless of the brevity of *De optimo genere oratorum*, which makes statistical analysis more difficult and consequently unreliable, we will propose that Cicero did indeed write this text.

4. Text corpora and methods

We have two corpora:

1. The corpus of Cicero's texts contains all texts from Cicero in prose. Major citations are extracted to avoid disturbance from other authors' style. The Ciceronian corpus contains approximately 900,000 words.
2. The background corpus includes almost all texts in prose from the first century BCE to the early fifth century CE, ninety-eight in total, obviously from numerous authors. Very short or fragmentary texts are excluded, as well as texts that consist largely of long citations (e.g. Varro's *De lingua Latina* and Aulus Gellius's *Noctes Atticae*). Asconius's commentary on Cicero's speeches is excluded. From the vast corpus of the most prolific Christian Latin author, Augustine, we have included only his three major works: *De civitate dei*, *De trinitate* and *Confessiones*. Inclu-

³⁴ Dihle 1955, 303. Kroll 1958, 1102 did not doubt Cicero's authorship, and also Ronconi has strongly supported it. He thought that the writing is a draft, and therefore lacks the *concinntas* (skilful joining of words and clauses) typical of Cicero, which is why Asconius would have doubted the authorship. This would also provide an explanation why Cicero and Quintilian do not mention the *De optimo genere*. Ronconi dated it between 52 and 46. Further, he believed that the translations remained a good intention (Ronconi 1998, 62-63).

³⁵ Dihle 1955, 306-309 e.g. found that the passage in *opt. gen.* 3 contradicts *orat.* 69.

³⁶ Ascon. *Mil.* 26, *ex libro apparet qui nomine Ciceronis inscribitur de optimo genere oratorum*.

³⁷ Berry 1996, 68-69 also presented a hypothesis of the course of events which led to the publication of this draft.

sion of all Augustinian texts (over four million words) would have resulted in an over-representation of his style in the corpus. The background corpus contains approximately 3.5 million words.

The texts are retrieved from public domain sources (mostly *The Latin Library*, partly *Bibliotheca Augustana*)³⁸. We are aware that the quality of the texts may not always be the best possible, but when forced to choose between the best critical editions and the reproducibility of data, we choose the latter. Sporadic misprints and other mistakes have a negligible impact when tremendous amounts of text are processed. The full list of authors and works is presented in the Appendix. The full data and code are released at GitHub³⁹.

Given the close history of Cicero's *De inventione* and *Rhetorica ad Herennium* and the probability that Cicero has written down mostly the words of his teacher, we decided to mark *De inventione* also as a test case and not as Cicero's text. This means we will attempt to predict whether Cicero is the author of these two texts as well as *De optimo genere oratorum* and *Commentariolum petitionis*.

To perform the authorship verification we use machine learning. Today, it is widely used and has been proven to be effective in authorship attributions and verifications⁴⁰. As we do not have preserved works from the other proposed authors, a closed-set authorship attribution is not possible. This would mean a set-up where one has x number of possible candidates and the work in question is attributed to one of them. An open-set verification is much more difficult to solve: is the work in question written by a given candidate (here Cicero) or by a different person or group? Recently, M. Koppel and Y. Winter have reduced the verification problem to the question whether a pair of documents is written by the same author⁴¹.

We use the two corpora combined as training data for the classifier, while the four test cases form the test data. Since we are not trying to attribute our test cases to a specific author but instead attempting to verify

³⁸ <http://www.thelatinlibrary.com>; https://www.hs-augsburg.de/~harsch/a_chron.html.

³⁹ <https://github.com/propreau/reconsideringciceronianauthorship> (data), <https://github.com/avjves/AuthAttHelper> (code and descriptions of the classifiers; this repository has a revised version of the CNN, the CNN architecture and setups used in this article).

⁴⁰ Stamatatos et al. 2014.

⁴¹ Koppel-Winter 2014.

whether or not they were written by Cicero, we treat all texts not written by Cicero as if they were written by a single author, thus creating one class for Cicero and another for the other authors⁴². From training data the classifier learns examples (in this case works of Latin literature) and a class for them (Cicero or non-Cicero). It learns a model, which is then used to classify an unlabelled text, in other words, to determine, whether a test case belongs to one or another class learned by the classifier.

For classification, each text is represented as a feature vector, where each element of the vector is the value of one specific feature. We conduct two different classifications, each using their own distinct features: word unigrams (one word) and bigrams (two consecutive words) for the first classification, and character five-grams (five consecutive characters, including the space character) for the second. These are commonly used features for document classification in general, and therefore used also for authorship attribution and verification tasks⁴³. When using the word n-grams, all content words (which primarily express lexical meaning) are masked by reducing them to their part-of-speech tags (*e.g.* “noun”, “verb”). The text documents are parsed using the UDPipe parser with a model trained from Latin-PROIEL treebanks (for examples, see Figure 1)⁴⁴. Consequently, the classifier will base its decisions on the part-of-speech tags and function words that have a general meaning (*e.g.* conjunctions). This forces the classifier to focus more on the author’s style and reduces the influence of the topic of the text on the result. Function words in particular have been successfully and repeatedly used in computational stylistics: their significant advantages are that they are high-frequency features and that their use is not genre-specific. Moreover, function words are less subject to an author’s conscious choice than content words. Although these characteristics have been subjected to criticism (for example, personal pronouns can be genre-specific)⁴⁵, several studies have shown the applicability of the function words also in the

⁴² This does not mean that the analysis would not recognize differences between authors in this class, see the cross-validation methods explained below.

⁴³ Stamatatos 2013.

⁴⁴ Straka et al. 2016; Universal Dependencies project «UDPipe is a trainable pipeline for tokenization, tagging, lemmatization and parsing of CoNLL-U files» <http://universaldependencies.org/tools.html#udpipe>; UDPipe homepage <http://ufal.mff.cuni.cz/udpipe>; Haug-Jøhndal 2008, <https://proiel.github.io/>.

⁴⁵ Kestemont 2014. See also Rybicki-Eder 2011 for the poorer performance of the most frequent words in Latin and Polish compared to English.

stylo-metric study of Latin texts⁴⁶. The automatically produced part-of-speech tagging is naturally not perfect; therefore, to establish its effect on the final results, we conduct different tests with and without tagging.

Figure 1. Feature extraction example using both plain and parsed text

Plain text:

Saepe et multum hoc mecum cogitavi, bonine an mali plus attulerit hominibus et civitatibus copia dicendi ac summum eloquentiae studium

Character 5-grams:

Saepe, aepe, epe e, pe et, e et, et m, et mu, t mul, mult, multu, ultum, ltum, tum h, um ho, m hoc, hoc, hoc m, oc me, c mec, mecu, mecum, ecum, cum c, um co, m cog, cogi, cogit, ogita, gitav, itavi, tavi, avi b, vi bo, i bon, boni, bonin, onine, nine, ...

Parsed text:

Saepe et multum hoc mecum VERB ADJ an ADJ plus VERB NOUN et NOUN NOUN VERB ac ADJ NOUN NOUN

Word 1-grams:

Saepe, et, multum, hoc, mecum, VERB, ADJ, an, ADJ, plus, VERB, NOUN, et, NOUN, NOUN, VERB, ac, ADJ, NOUN, NOUN

Word 2-grams:

Saepe et, et multum, multum hoc, hoc mecum, mecum VERB, VERB ADJ, ADJ an, an ADJ, ADJ plus, plus VERB, VERB NOUN, NOUN et, et NOUN, NOUN NOUN, NOUN VERB, VERB ac, ac ADJ, ADJ NOUN, NOUN NOUN

We train the classifier with the training data to recognize the features typical and untypical of Cicero's style. After that our test cases (*Rhetorica ad Herennium*, *De inventione*, *De optimo genere oratorum* and *Commentariolum petitionis*) can be presented to the classifier, which outputs a value indicating how much (positive) or how little (negative) each test case resembles the class (Cicero's texts). We rescale these values between -1 and 1 to obtain comparable results between different classifiers, as their prediction values have different scales. The rescaling is done by taking the maximum and minimum decision value from the classifier and retaining the same ratio in the new scale.

The value and the decision are largely useless in isolation if we cannot be certain that the classifications are valid overall. Here we apply the standard technique of leave-one-out cross-validation using the training

⁴⁶ Tweedie et al. 1998; Gurney-Gurney 1998; Forsyth et al. 1999; Kestemont et al. 2015; Deploige-Moens 2016; Stover-Kestemont 2016a; Richards et al. 2016; De Gussem 2017.

data, which provides us with an estimate of the classification accuracy and therefore the reliability of our results on the actual test documents. In cross-validation, we focus only on our training data, ignoring the actual test texts. We remove one document at a time from the training data and consider it as a new test case. Our current training data now consists of all texts but the new test case, and using it we subsequently train the classifier so that it gives a class and a value for the new test case. Since we know the authors of the texts included in the training data, these results show how accurately the classifier classifies data that it has not previously seen.

We apply two different types of classifiers. The first is a linear Support Vector Machine (SVM), which is a simple yet effective classifier that has traditionally been applied in text classification tasks⁴⁷. In particular, we use the scikit-learn LinearSVC implementation of SVM with L2 penalty and squared hinge as loss. The C-parameter of the classifier is set using cross-validation so as to avoid overfitting on the test data. In the following discussion, this is our Classification 1 (Cla1). The SVM works by learning a weight for every feature from the training data, so as to maximize the decision margin between the two classes. The weight being positive or negative is an indication of which class the feature is potentially associated with, although one needs to exercise caution when comparing features in isolation based on their weight. With the SVM, we use word and part of speech unigrams and bigrams as features, as these are, in the end, easier to interpret than the character n-grams which we will use for the neural network method. Each feature vector is normalized by turning them into their unit vectors, i.e. dividing them by their lengths, so that documents of different lengths can be accurately classified. We also perform an additional test where the SVM is used on plain text, without masking content words with their part-of-speech tags. In this way we attempt to establish the extent to which the vocabulary influences the result. This will be referred to as Classification 3 (Cla3).

Our Classification 2 (Cla2) is a Convolutional Neural Network (CNN), representing the current state of the art in text classification⁴⁸. The CNN uses a complex, non-linear classification function, which makes it difficult to extract a global weight of an individual feature. CNNs have been

⁴⁷ Cortes-Vapnik 1995; Chang-Lin 2011.

⁴⁸ LeCun et al. 2015.

used successfully in different authorship attribution tasks⁴⁹. CNNs slide a short window across the text and learn to recognize patterns (features) in this window so as to maximize the classification performance. In this manner, CNNs can learn a feature representation of the data that for SVMs needs to be given explicitly. In our case the sliding window consists of five overlapping subsequences, each being five characters long, *i.e.* ten characters in total (see Figure 1). Instead of presenting a full text document at a time to the network, we split each document into slices of around 1.000 words. Most importantly, this allows us to increase the number of training examples for the network and balances the sizes of the texts, which often differ substantially. Each slice is classified independently, and the results of all slices are averaged to get the value for the whole text document. It is important to remember that the absolute classification scores, especially when near the threshold between the two classes, are not decisive on their own. Instead of the numeric values, one should observe where a certain text document is placed with respect to others and how the results cluster.

As a pre-processing step, the texts were cleared of punctuation because it was not used in antiquity⁵⁰. When running classifications with SVM, the texts were also lower cased and all one-character long tokens were removed. This makes the features easier to interpret. With CNN, however, only the punctuation was cleared. We performed several tests using various setups of the classifier, for instance using both parsed and plain text, and here we present the results concentrating on two of the tests.

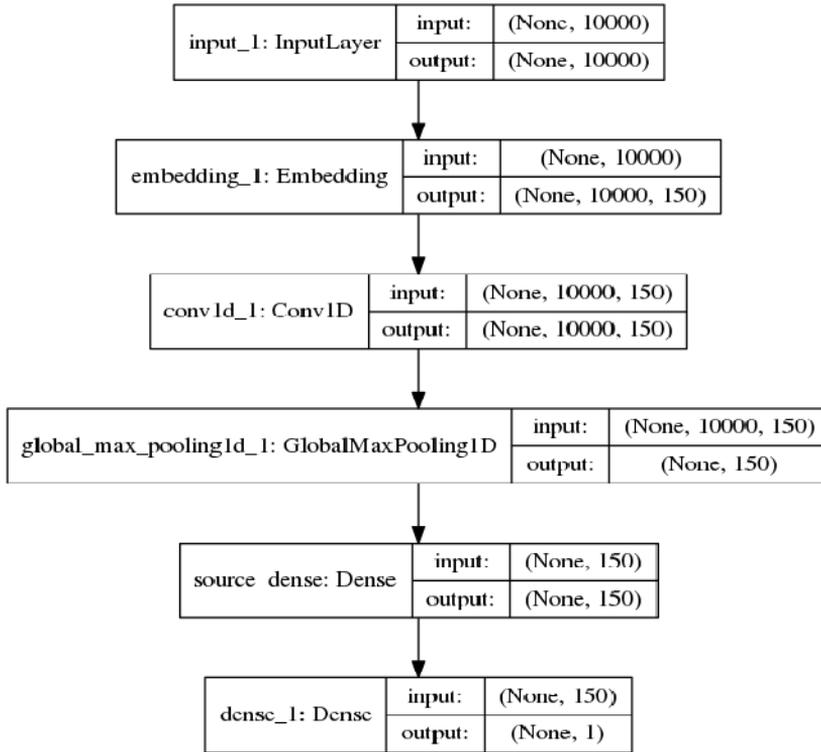
A detailed presentation of the CNN architecture is given below in Figure 2. The input layer has 10.000 nodes, each corresponding to a character five-gram in the input text slice. This corresponds to text slices of approximately 1.000 words. Each character five-gram is represented as a 150-dimensional embedding, and these are subsequently presented to a convolutional layer. We use 150 filters with kernel size set to 5. This gives kernels the ability to make decisions based on a maximum of 5 consecutive five-grams, *i.e.* 10 characters. We then apply max-pooling followed by a fully connected linear dense layer with 100 nodes. Finally, we

⁴⁹ Shrestha et al. 2017; Hitschler et al. 2017.

⁵⁰ Indeed, one of the first attempts to use computational authorship attribution to a classical text was marred by using the sentence length as a criterion for authorship. See Marriott 1979; Sansone 1990; Stover-Kestemont 2016, 140-141.

connect to a dense layer with sigmoid activation, giving us the binary classification result.

Figure 2. The CNN architecture.



The CNN can with good grounds be described as a “heavy gun” in authorship attribution problems. Indeed, simple classifiers and compression models outperformed more complex deep learning approaches in the most recent PAN-2018 competition⁵¹, which raises questions about the benefits of neural networks and other computation-heavy approaches. Based on our results, we cannot make definite claims about the superiority of the CNN, although it performed best with our data set. However, the use of two different classifiers using different feature

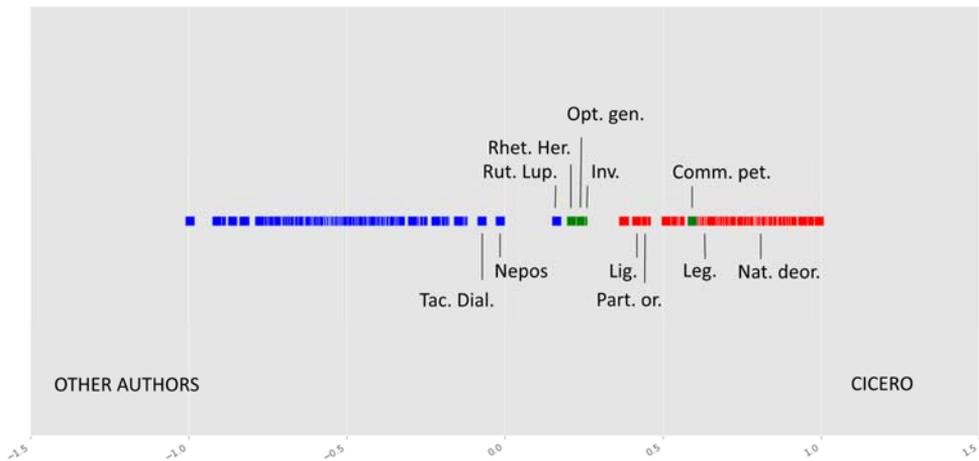
⁵¹ Kestemont et al. 2018.

sets is in accordance with P. Juola’s proposition of using independent analyses in verification problems to reduce the chance of false acceptance error⁵². Although simpler classifiers are available, we deemed that the CNN has added value, because its classifications are not only based on a completely different and more complex classification function than SVMs, but also on features it learns instead of being given explicitly. From the perspective of humanities, it is even further away from a human reader than analyses using frequencies of the most common words. At the same time, it is yet another, independent perspective on the problem we want to solve.

5. Results

Avoiding any absolute judgements at this point, we plot the texts on a scale by their classification score. Positive values indicate Ciceronian and negative non-Ciceronian authorship (zero marking the threshold).

Figure 3. Classification 1 (Cla1) – SVM Using Parsed Text



⁵² Juola 2015.

Figure 4. Classification 2 (Cla2) – CNN Using Plain Text

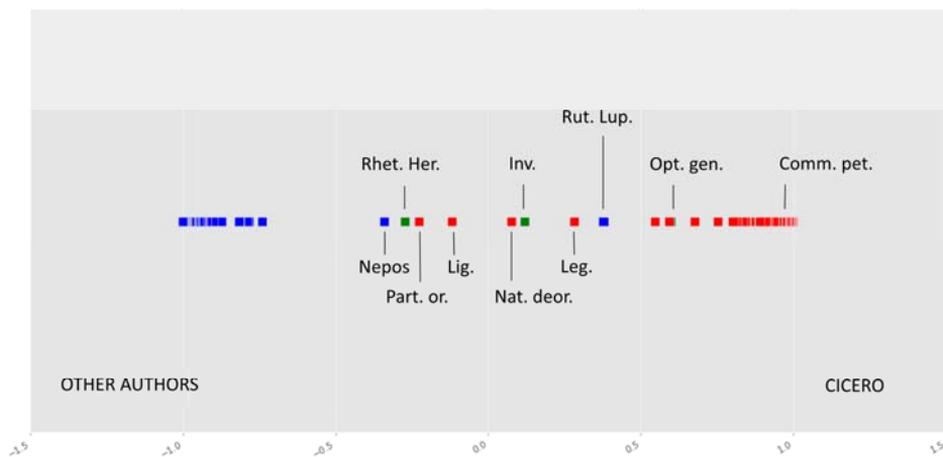


Table 1 presents the texts that the classifier has placed near the threshold in Cla1 and Cla2, or that are handled in our text for another reason, with their values received in the classifications.

Table 1. Values Received in the Classifications

Threshold 0.0	Classification 1 SVM, using parsed text	Classification 2 CNN, using plain text	Classification 3 SVM, using plain text
Quintilian <i>Declamationes maiores</i>	-0.42	-0.95	-0.34
Quintilian <i>Institutio oratoria</i>	-0.33	-0.81	-0.05
Tacitus <i>Dialogus de oratoribus</i>	-0.07	-0.94	-0.02
Nepos <i>Liber de Latinis historicis</i>	-0.01	-0.34	0.07
Rutilius Lupus <i>De figuris</i>	0.16	0.38	0.15
<i>Rhetorica ad Herennium</i>	0.21	-0.27	0.32
<i>De optimo genere oratorum</i>	0.24	0.60	0.16
<i>De inventione</i>	0.25	0.12	0.29
<i>Pro Ligario</i>	0.42	-0.12	0.20
<i>De partitione oratoria</i>	0.44	-0.23	0.33
<i>Commentariolum petitionis</i>	0.59	0.98	0.44
<i>De legibus</i>	0.62	0.28	0.47
<i>De natura deorum</i>	0.72	0.07	0.34
Accuracy	0.975	0.981	0.968
F1-score (Cicero)	0.967	0.974	0.959
F1-score (other texts)	0.979	0.985	0.975

Before discussing the actual test cases we will make some general remarks on the classifications and other texts. In Cla1 the UDPipe parser made mistakes in the assignment of grammatical cases, especially those with identical surface forms, and also with parts-of-speech. However, these errors, inevitable in the output of an automated parser, do not seem to have had a degrading effect, since the difference between the SVM using the parsed text and that using the plain text is in most cases relatively small.

In Table 1 we can also see the cross-validation accuracy and F1-scores, *i.e.* the harmonic mean of precision (the number of correct positive results divided by the number of all positive results) and recall (the number of correct positive results divided by the number of samples that should have been identified as positive). According to them, the CNN performs best of the three, though both the SVM classifications perform nearly as well. The SVM classification using parsed text seems to be slightly more accurate than that with plain text.

Owing to the more accurate neural network classifier, the background corpus clusters neatly in Cla2, forming a more compact cluster than in Cla1, as the values given to the text documents vary between -1.0 and -0.7, except for Cornelius Nepos's *Liber de Latinis historicis* and Rutilius Lupus's *De figuris sententiarum et elocutionis*. Also the Ciceronian corpus forms a cluster between 0.5 and 1.0, apart from four of his texts: *De partitione oratoria*, *Pro Ligario*, *De natura deorum* and *De legibus*⁵³.

As to individual texts, if the results of our two principal classifications differ significantly, the vocabulary may have an impact on the result in Cla2. In Cicero's case the presupposition is that the rhetorical vocabulary, which is much in use in his corpus, could attract other rhetorical texts in the classification. However, since Tacitus's *Dialogus de oratoribus* (-0.07 in Cla1, -0.94 in Cla2) and Quintilian's *Institutio oratoria* (-0.33 Cla1, -0.81 Cla2) and *Declamationes maiores* (-0.42 Cla1, -0.95 Cla2) show a discrepancy between the results of Cla1 and Cla2, and because it is Cla2 which gives the greater negative value, the CNN in Cla2 appears to base its decision on features other than vocabulary and there is a suggestion

⁵³ In our classifications Cicero's speeches *Post reditum in Quirites* and *Post reditum in senatu*, at times suspected of being someone else's (Berry 1996, 61 and n. 17), are placed without compunction among his works (0.54 Cla1 and 1.0 Cla2, 0.67 Cla1 and 0.99 Cla2 respectively).

that these texts include some strongly negative features not encountered in Cicero's texts. On the other hand, Rutilius Lupus's *De figuris sententiarum et elocutionis* (0.16 Cla1, 0.38 Cla2) is near the threshold in all classifications, which suggests that vocabulary is not the only criterion for classifying it as Ciceronian. We will return to this text below.

Cicero's *De natura deorum* (0.72 Cla1, 0.07 Cla2) is a good example of issues met at the threshold. In Cla3 we ran the corpora with the SVM using plain text to identify the features on which the CNN might be basing its decision. This test assigned the value 0.34 to *De natura deorum*, and a comparison with the result of Cla1 suggested that the vocabulary indeed has an impact, but this time in the negative direction: Some of the negative features received in Cla3 (such as *dei*, *spiritus*), are strongly represented in this text because of its topic. *De natura deorum* is a dialogue in which one person in each of the three books presents the theology of one philosophic school. It might be worth examining whether Cicero is deliberately attempting to change his style according to each speaker, since the values given to each slice of around 1.000 words in Cla2 vary considerably, between -0.82 and 0.92.

The results concerning Cicero's *Pro Ligario* (0.42 Cla1, -0.12 Cla2) and *De legibus* (0.62 Cla1, 0.28 Cla2) also point to the influence of their topics⁵⁴. In *De partitione oratoria* (0.44 Cla1, -0.23 Cla2), however, the vocabulary does not seem to have an influence (Cla3 gave approximately the same value, 0.33, as Cla1). Here the structure of the text may have an effect: it is written in the form of a dialogue between Cicero and his son. Especially the lines of Cicero's son usually comprise short questions whose brevity and scholastic manner is untypical of Cicero.

In Cornelius Nepos's *Liber de Latinis historicis* the vocabulary does not seem to have any major influence according to Cla3 (0.07 compared to -0.01 in Cla1), although most of this text concerns the life of Cicero's dearest friend, who features in his letters to Atticus. The CNN in Cla2 classifies the text as non-Ciceronian (-0.34).

Rutilius Lupus's *De figuris sententiarum et elocutionis* (early 1st c. CE) is placed among Cicero's texts (0.16 Cla1, 0.38 Cla2). The text is generally considered an incomplete Latin abridgement of a work on rhetorical fig-

⁵⁴ In *Pro Ligario* Cicero's defendant was accused of crimes in Africa, and Cicero gave the speech in front of C. Caesar in 46 BCE, repeatedly mentioning his name (cf. topic-wise Caesar's and Sallust's works). *De legibus* is incomplete (cf. Häfner 1928, 94-103; Gawlick 1994, 1035; see also Schmidt 1969).

ures by Gorgias, a rhetorician who was Cicero's son's teacher in Athens. Arguments have been presented that the *De figuris* would be a later epitome of Rutilius's treatise⁵⁵. There is no preface, and the extant text consists of short chapters, each pertaining to one figure named in Greek (41 in total). In each chapter there is a short definition of the figure, and relevant examples from speeches of Greek orators. Its importance lies on the number of these extracts, translated into Latin from the lost writings of Demosthenes and other Greeks⁵⁶. There is an implication that the vocabulary plays its part in Cla2. Still, there must be something else in common between this text and Cicero's writings, since it came close to the threshold in Cla1, and in Cla3 it was barely differentiated from them (0.15). As early as 1768, D. Ruhnken observed that a passage in which a Greek example does not function in Latin comes from Cicero, probably from his lost work⁵⁷. He also reported many other similarities with Cicero's writings, and G. Barabino has noted that in many passages Rutilius was strongly influenced by Cicero⁵⁸. In such a short text (around 4.300 words) such passages are likely to have an excessive weight in the analysis.

At least two of the twelve examples deriving from Demosthenes in Rutilius's text are from *De corona*, the very speech which Cicero translated, or planned to translate, into Latin (cf. *De optimo genere oratorum* above)⁵⁹. When we performed yet another extra test, outside our three classifications, and removed these three short passages from the text – one from Cicero's lost speech (in Rut. Lup. 1, 3) and two originally from *De corona* (in Rut. Lup. 1, 16 and 2, 1), the result of Cla2 changed so that the value given to this text shifted 0.13 points in the non-Ciceronian direction. Obviously, the removed passages include something Ciceronian, and it is probable that other similar passages remain in the *De figuris*. An explanation for these could be that Cicero made a translation of Demosthenes's *De corona* that is no longer extant or was only a draft, which then influenced Rutilius.

⁵⁵ See Brooks 1970, XIV. Not much is known about Rutilius's life, but apparently Quintilian knew the whole of his work (cf. Quint. 9, 2, 101-105 *passim*).

⁵⁶ Weissenberger 2006.

⁵⁷ Ruhnken 1768, 12-14. Rut. Lup. 1, 3 (ed. Barabino) on *paronomasia*: *Item. At huius sceleratissimi opera, qui fuit lucus religiosissimus, nunc erit locus desertissimus: nimirum quoniam traditam sibi publicorum custodiam sacrorum non honori, sed oneri esse existimavit.*

⁵⁸ To mention only the most obvious passages, Rut. Lup. 1, 2; 1, 17; 1, 18; 1, 21; 2, 7; 2, 9 (Ruhnken 1768 ad loc.; Barabino 1967, 78-80; 119-120).

⁵⁹ Rut. Lup. 1, 1 and 1, 16 (Brooks 1970, 50 and 72).

In the end we had four test cases. In Cla1 our analysis places the *Rhetorica* among Cicero's texts, albeit near the threshold (0.21), and *De inventione* quite close to it (0.25). In Cla2 the *Rhetorica* is regarded as a non-Ciceronian text by a narrow margin (-0.27). This is interesting, as Cla3 by its greater positive value (0.32) suggests that the vocabulary might have an influence here, and yet the CNN in Cla2 gives a negative value suggesting that there are some non-Ciceronian features in the text. *De inventione* is also placed near the threshold (0.12) in Cla2. This is a fair estimation, since most of the text derives from Cicero's teacher. On the other hand, our analyses give credence to the theory of a common origin of the two texts by placing them close to one another. The analyses do not tell us unequivocally whether or not Cicero wrote the *Rhetorica*, although Cla2 suggests that he did not. Therefore, it is reasonable to conclude that the work has been written by someone whose text is not preserved and hence not included in the background corpus⁶⁰. It is nevertheless interesting that the values given to each slice of around 1.000 words in Cla2 show a wide variation from -0.98 to 0.82. Similar variation is seen in the case of *De inventione* (between -0.87 and 0.94). The question arises as to why some parts of the *Rhetorica* are very Ciceronian. Has the author of the *Rhetorica* used Cicero as his source after all? (*De inventione* was extracted from the Ciceronian corpus, and therefore it has no influence here.)

In Cla1 *De optimo genere oratorum* is placed near the threshold (0.24) between the *Rhetorica* and *De inventione*, but in Cla2 it is assessed as strongly Ciceronian (0.60). Since the text comprises only around 1.600 words, there is room for suspicion that its brevity may have impeded a reliable statistical analysis. However, in Cla3 the result (0.16) was not very far from that in Cla1. This might imply that the vocabulary is not the reason why this text is placed clearly among Cicero's works in Cla2. The reason could lie in the greater accuracy of the CNN in Cla2, indicating that this text is indeed written by Cicero.

Dramatically, *Commentariolum petitionis* is placed explicitly among Cicero's texts in all the classifications. After receiving this result already in our first tests, we decided to extract it from the background corpus and mark it

⁶⁰ A computational study (Forsyth et al. 1999, 383-384) clearly separated the *Rhetorica* from the Ciceronian corpus. The results were based on a much smaller amount of data (25 works of Cicero compared with the *Rhetorica* and *Epistula ad Octavianum*) and a relatively simple analysis of word frequencies of the 46 most common words in these samples.

as our fourth test case. In Cla1 it is regarded as more Ciceronian (0.59) than five of his speeches and *De partitione oratoria*, *Topica*, *De fato* and *Academica*. In Cla2 the result is even more remarkable: the CNN regards the *Commentariolum* as a strongly Ciceronian text (0.98) along with some of his speeches and letters. These results give us good reason to attribute the *Commentariolum* to Cicero⁶¹. They cannot be caused by short passages that are similar to Cicero's speech *In toga candida* (which, according to some scholars, Quintus adapted from his brother's work, see above).

Since we only have four short letters written by Quintus Cicero⁶², we do not have enough comparative material (only ca. 440 words) to classify his text in a similar manner to his brother's, so we cannot unequivocally rule him out as the author. But if Quintus were the author, given the strongly Ciceronian result for the *Commentariolum*, the style of both brothers would be astoundingly similar. Stylometric studies on the Brontë sisters, who shared the same education and family background and who wrote in the same genre, have demonstrated marked differences in the sisters' style⁶³. Our two brothers had a common education and close relations, but Quintus was several years younger and a politician, not a full-time orator and a rhetorical theorist like his brother⁶⁴. It is highly improbable that the different circumstances in which they found themselves for most of their lives would have had no effect at all on their style. In addition, we know that the behaviour and temperament of the brothers was very different⁶⁵.

⁶¹ It may be worth noting that even Henderson 1950, 18, who has taken the *Commentariolum* as a fake or a later literary exercise, has conceded that «it is a remarkable attempt at a Ciceronian letter». Tyrrell-Purser 1877, 122 have noted «Many of these supposed defects would pass quite unnoticed if the work had been attributed to Marcus; indeed, many of them can actually be paralleled in the writings of the great orator». Hendrickson 1903, 87 has found the *clausulae*, the rhythmic structures at the end of a clause, to resemble Cicero's speeches. Cf. also Clift 1945, 102 n. 76; Wikarjak 1966, 31.

⁶² His letters are preserved in Cicero's collection: *fam.* 16, 8; 16, 16; 16, 26; 16, 27.

⁶³ Koppel et al. 2004 found the styles of the sisters hard, but possible, to distinguish. Another study using SVM and bigrams of syntactic labels succeeded in distinguishing the Brontë sisters with high accuracy even in relatively small text samples (200 – 1.000 words), see Hirst-Feiguina 2007.

⁶⁴ Münzer 1948, cols. 1287 and 1305.

⁶⁵ For instance Tyrrell-Purser 1904, 123 have commented that there is no sign of the irascibility (*iracundia*) of which Cicero (*ad Q. fr.* 1, 37-40) speaks at length as a typical characteristic of his brother. Tyrrell-Purser have regarded this as a proof of the genuineness of the text (i.e. being written by Quintus Cicero), arguing that the supposed later compiler would have made good use of this characteristic in the text. Cf. Münzer 1948, 1291.

Let us consider the possibility that Marcus Cicero is in fact the author. Why would he have written this treatise in such a way that the reader is left with the impression of him being the recipient of the text rather than its author? (The greeting, which is often confused in the manuscripts, could of course be a later addition by a copyist). Perhaps Cicero just wanted to leave a report of his campaign for the consulship, just as he planned to have his *res gestae* written in 55⁶⁶. He may have felt that he needed to distance himself from the text which deals with his own political ambitions during the campaign (cf. *fam.* 5, 12, 8 below, where he expresses his reluctance to write about his own achievements). Otherwise people could have considered him arrogant and impudent, and it might have had a damaging impact on his career⁶⁷. We know that he wrote such a report on his consulship in Greek and sent it to Atticus so that he could circulate it in Greece and make his accomplishments more widely known, and Cicero was planning a Latin version too⁶⁸. Perhaps *Commentariolum petitionis* is a part of this project, a draft of the early phases which led to his consulship. In his letter to Luceius (cf. n. 66) he also pondered the alternative that he himself might be driven to write the work he asked for, *commentarium consulatus* as he called it, should Luceius turn him down. Cicero found the idea of writing his own praises very problematic:

sed, quod te non fugit, haec sunt in hoc genere vitia: et verecundius ipsi de sese scribant necesse est si quid est laudandum et praetereant si quid reprehendendum est. accedit etiam ut minor sit fides, minor auctoritas, multi denique reprehendant et dicant verecundiores esse praecones ludorum gymnicorum, qui, cum ceteris coronas imposuerint victoribus eorumque nomina magna voce pronuntiarint, cum ipsi ante ludorum missionem corona donentur, alium praeconem adhibeant, ne sua voce se ipsi victores esse praedicent (*fam.* 5, 12, 8).

But I need not point out to you that this genre has certain disadvantages. An autobiographer must write over modestly where praise is due and pass over

⁶⁶ See Cic. *fam.* 5, 12, Cicero's letter to Luceius, whom he tries to persuade to write a monograph of his exploits in his consular year, especially the revealing of Catiline's conspiracy – which Luceius in fact agreed to do (cf. Cic. *Att.* 4, 6, 4). Cf. Häfner 1928, 63–64, who however dated the letter to 56: for the dating, see also Malaspina 2004, [s.v.](#)

⁶⁷ Cf. Nardo 1970, 76.

⁶⁸ Cic. *Att.* 1, 19, 10 *commentarium consulatus mei Graecae compositum misi ad te. [...]* *Latinum si perfecero, ad te mittam.* See also *Att.* 1, 20, 6 and 2, 1, 1–2, all three letters written in 60. Cf. Häfner 1928, 61–63.

anything that calls for censure. Moreover, his credit and authority are less, and many will blame him and say that heralds at athletic contests show more delicacy, in that after placing garlands on the heads of the winners and loudly proclaiming their names, they call in another herald when it is their turn to be crowned at the end of the games, in order to avoid announcing their own victory with their own lips. (transl. by Shackleton Bailey 2001 in Loeb CL).

There is a possibility that both brothers participated in the writing of the *Commentariolum*⁶⁹, if the concluding remark of Quintus in the text, the request to obtain Marcus's help in amending it, is genuine. That would account for the supposed irregularity of the style⁷⁰, but this could also be explained by the incompleteness of the text.

Our result makes the much-debated question of which of the texts was the earliest, the *Commentariolum* or Cicero's speech *In toga candida*, largely irrelevant if the author was the same. Scrutiny of the historical facts and their dates becomes pointless, since Cicero could have written the *Commentariolum*, or parts of it, at a later date.

To conclude, our analyses show that:

1. *Commentariolum petitionis* was written, at least largely, by Marcus Cicero. It is possible that his brother Quintus participated in writing it but we do not consider any major input by him probable, since it would be unlikely that the brothers had such a uniform style as that in the text, and since our analyses very strongly indicate a Ciceronian authorship.
2. *De optimo genere oratorum* is an authentic text by Cicero.
3. As to *Rhetorica ad Herennium*, we cannot draw a definite conclusion from our analyses, but if we take into account the text history it seems to be written by someone unknown from extant texts.
4. Rutilius Lupus's *De figuris sententiarum et elocutionis* has close similarities with Cicero's works. By extracting three short passages from the text, we discovered that they had impact in the analysis and included Ciceronian elements. It appears likely that the text includes other passages greatly influenced by Cicero. We

⁶⁹ Cf. Tyrrell 1877, 47 and 59. Nardo 1970, 136-137 has presupposed a close pragmatic and ideological collaboration between the brothers, with Quintus doing the actual writing but at Marcus's instigation.

⁷⁰ The opinion of Tyrrell 1877, 49 and Leo 1895, 448.

cautiously suggest that Cicero made a translation of Demosthenes's *De corona* and that Rutilius may have known it.

5. As to methodology, the possible influence of other texts, as well as the nature of the text in question, cannot be ignored. We have taken this into account *e.g.* by marking *De inventione* as a test case, even though its authorship is indisputable.
6. The CNN (Cla2) does seem to classify accurately. Vocabulary appears to influence the result only in certain respects, which is why it is useful to have the SVM (Cla1, and also Cla3) as support. Because rhetoric plays such an important part in Cicero's works, the rhetorical vocabulary is likely to be valued positive. Since Tacitus's and Quintilian's rhetorical writings are classified as very distinct from Cicero's, the conclusion must be that the rhetorical vocabulary does not weigh significantly: instead, the weight is laid on the negative features, which are perhaps forced more prominently forward. Another, but contrary, indication involves Cicero's *De natura deorum*, *Pro Ligario* and *De legibus*: the non-rhetorical vocabulary seems to weigh more than the positive features, and therefore these texts are drawn towards the background corpus. All in all, in very large corpora such as those we have used, it seems that the vocabulary has a limited effect, and the sheer number of texts makes the common function words the most significant features in the classification.
7. Study of old texts can certainly benefit from computational authorship attribution and verification, provided that philological and historical methods are not disregarded. As shown above, computational tests can confirm the results received by traditional humanistic methods, or alternatively produce totally new results by using analyses unfeasible for a human.

Appendix

Test cases*Commentariolum petitionis**De inventione**De optimo genere oratorum**Rhetorica ad C. Herennium***Corpus of Cicero's texts***Academica**Brutus**Cato maior de senectute**De divinatione**De domo sua**De fato**De finibus**De haruspicum responsis**De imperio Cn. Pompei**De lege agraria contra Rullum**De legibus**De natura deorum**De officiis**De oratore**De partitione oratoria**De provinciis consularibus**De re publica**Epistulae ad Atticum I–IV**Epistulae ad Atticum IX–XII**Epistulae ad Atticum V–VIII**Epistulae ad Atticum XIII–XVI**Epistulae ad Brutum**Epistulae ad familiares I–V**Epistulae ad familiares VI–XII**Epistulae ad familiares XIII–XVI**Epistulae ad Quintum fratrem**In Catilinam I–IV**In Pisonem**In Vatinius**In Verrem**Laelius de amicitia**Orator**Philippica**Post reditum in Quirites**Post reditum in senatu**Pro Archia**Pro Balbo*

Pro Caecina
Pro Caelio
Pro Cluentio
Pro Cn. Plancio
Pro Deiotaro
Pro Flacco
Pro Fonteio
Pro Ligario
Pro Marcello
Pro Milone
Pro Murena
Pro Quinctio
Pro Rabirio perduellionis reo
Pro Rabirio Postumo
Pro Roscio Amerino
Pro Roscio comoedo
Pro Scauro
Pro Sestio
Pro Sulla
Topica
Tusculanae disputationes

Background Corpus

Ambrosius, *De mysteriis*
 Ambrosius, *Epistulae variae*
 Ammianus Marcellinus, *Res gestae*
 Anonymus1, *De bello Alexandrino liber*
 Anonymus2, *De bello Africo*
 Anonymus3, *De bello Hispaniensi liber*
 Anonymus4, *Historia Apollonii regis Tyri*
 Apuleius, *De deo Socratis*
 Apuleius, *De dogmate Platonis*
 Apuleius, *De mundo*
 Apuleius, *Florida*
 Apuleius, *Metamorphoses*
 Arnobius, *Adversus nationes*
 Augustinus, *Confessiones*
 Augustinus, *De civitate Dei*
 Augustinus, *De trinitate*
 Augustus, *Res gestae*
 Aulus Hirtius, *De bello Gallico commentarius octavus*
 Aurelius Victor, *Liber de Caesaribus*
 Caesar, *De bello civili*
 Caesar, *De bello Gallico*
 Celsus, *De medicina*
 Columella, *De arboribus*
 Columella, *De re rustica*

Cornelius Nepos, *Liber de excellentibus ducibus exterarum gentium*
Cornelius Nepos, *Liber de Latinis historicis*
Curtius Rufus, *Historiarum Alexandri Magni libri*
Egeria, *Itinerarium*
Eutropius, *Breviarium historiae Romanae*
Florus, *Epitome de T. Livio*
Gaius, *Institutionum commentarii quattuor*
Hyginus, *De astronomia libri*
Hyginus, *Fabulae*
Hyginus, *De munitionibus castrorum*
Iunianus Iustinus, *Historiarum Philippicarum libri XLIV*
Livius, *Ab urbe condita*
Minucius Felix, *Octavius*
Plinius maior, *Naturalis historia II-XXVII*
Plinius minor, *Epistularum libri decem*
Plinius minor, *Panegyricus*
Pomponius Mela, *De chorographia*
Pseudo-Aurelius Victor, *Epitome de Caesaribus*
Pseudo-Aurelius Victor, *Liber de viris illustribus urbis Romae*
Pseudo-Tertullianus, *Spuria*
Quintilianus, *Declamationes maiores*
Quintilianus, *Institutio oratoria*
Rutilius Lupus, *De figuris sententiarum et elocutionis*
Sallustius, *Bellum Iugurthinum*
Sallustius, *De Catilinae coniuratione*
Seneca minor, *De beneficiis*
Seneca minor, *De consolatione*
Seneca minor, *De ira*
Seneca minor, *Dialogi 1-2; 6-12*
Seneca minor, *Epistulae morales ad Lucilium*
Seneca minor, *Quaestiones naturales*
Sextus Pomponius, *Liber singularis enchiridii*
Suetonius, *De vitis Caesarum*
Tacitus, *Agricola*
Tacitus, *Annales*
Tacitus, *Dialogus de oratoribus*
Tacitus, *Germania*
Tacitus, *Historiae*
Tertullianus, *Ad martyres*
Tertullianus, *Ad nationes*
Tertullianus, *Ad Scapulam*
Tertullianus, *Ad uxorem*
Tertullianus, *Adversus Hermogenem*
Tertullianus, *Adversus Iudaeos*
Tertullianus, *Adversus Marcionem*
Tertullianus, *Adversus Praxean*
Tertullianus, *Adversus Valentinianos*
Tertullianus, *Apologeticum*

Tertullianus, *De anima*
 Tertullianus, *De baptismo*
 Tertullianus, *De carne Christi*
 Tertullianus, *De corona militis*
 Tertullianus, *De cultu feminarum*
 Tertullianus, *De exhortatione castitatis*
 Tertullianus, *De fuga in persecutione*
 Tertullianus, *De idololatria*
 Tertullianus, *De ieiunio*
 Tertullianus, *De monogamia*
 Tertullianus, *De oratione*
 Tertullianus, *De paenitentia*
 Tertullianus, *De pallio*
 Tertullianus, *De patientia*
 Tertullianus, *De praescriptione haereticorum*
 Tertullianus, *De pudicitia*
 Tertullianus, *De resurrectione carnis*
 Tertullianus, *De spectaculis*
 Tertullianus, *De testimonio animae*
 Tertullianus, *De virginibus velandis*
 Tertullianus, *Scorpiace*
 Valerius Maximus, *Factorum et dictorum memorabilium libri*
 Varro, *Rerum rusticarum libri*
 Vegetius, *Epitoma rei militaris*
 Velleius Paterculus, *Historiae Romanae*
 Vitruvius, *De architectura*

Bibliography

- Achard 1985: G. Achard, *L'auteur de la Rhétorique à Herennius?* «REL» 63, 1985, pp. 56-68.
 Achard 1989: G. Achard, *Rhétorique à Herennius*, Texte établi et traduit par G. Achard, Paris 1989.
 Adamietz 1960: J. Adamietz, *Ciceros De inventione und die Rhetorik ad Herennium*, Marburg 1960.
 Alexander 2018: M.C. Alexander, [*Roman Amoralism Reconsidered: The Political Culture of the Roman Republic and Historians in an Era of Disillusionment*](#), Chicago 2018.
 Balsdon 1963: J.P.V.D. Balsdon, *The Commentariolum Petitionis*, «CQ» 13, 1963, pp. 242-250.
 Barabino 1967: G. Barabino, *P. Rutilii Lupi Schemata dianoeas et lexeos*, Saggio introduttivo, testo e traduzione, Genova 1967.

- Berry 1996: D.H. Berry, *The Value of Prose Rhythm in Questions of Authenticity: The Case of De Optimo Genere Oratorum Attributed to Cicero*, in *Papers of the Leeds International Latin Seminar*, «ARCA» 34, 9, Leeds 1996, pp. 47-74.
- Brooks 1970: E. Brooks, *P. Rutilii Lupi De figuris sententiarum et elocutionis*, ed. with Prolegomena and Commentary, Leiden 1970.
- Bücheler 1869: F. Bücheler, *Quinti Ciceronis reliquiae*, Lipsiae 1869.
- Calboli 1969: G. Calboli, *Cornifici Rhetorica ad C. Herennium*, Introduzione, testo critico, commento, Bologna 1969.
- Caplan 1964: H. Caplan, *[Cicero] ad C. Herennium de ratione dicendi (Rhetorica ad Herennium)*, with an English translation, Loeb CL, Cambridge MA 1964.
- Chang-Lin 2011: C.-C. Chang, C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, «ACM Transactions on Intelligent Systems and Technology (TIST)» 2 (3), 2011, pp. 1-27 [<https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>].
- Clift 1945: E.H. Clift, *Latin Pseudepigrapha: A Study in Literary Attributions*, Baltimore 1945.
- Corbeill 2002: A. Corbeill, *Rhetorical Education in Cicero's Youth*, in J.M. May (ed.), *Brill's Companion to Cicero. Oratory and Rhetoric*, Leiden 2002, pp. 23-48.
- Cortes-Vapnik 1995: C. Cortes, V. Vapnik, *Support-Vector Networks*, «Machine Learning» 20 (3), 1995, pp. 273-297 [http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf].
- Cox-Ward 2011: V. Cox, J.O. Ward (eds.), *The Rhetoric of Cicero in Its Medieval and Early Renaissance Commentary Tradition*, Leiden 2011.
- David et al. 1973: J.-M. David, S. Demougin, E. Deniaux, D. Ferey, J.-M. Flambard, C. Nicolet, *Le "Commentariolum Petitionis" de Quintus Cicéron. État de la question et étude prosopographique*, «ANRW» 1, 3, 1973, pp. 239-277.
- De Gussem 2017: J. De Gussem, *Bernard of Clairvaux and Nicholas of Montieramey: Tracing the Secretarial Trail with Computational Stylistics*, «Speculum» 92, 2017, pp. 190-225 [<https://www.journals.uchicago.edu/doi/10.1086/694188>].
- Deploige-Moens 2016: J. Deploige and S. Moens, *Visiones Hildegardis a Guiberto Gemblacensi exaratae*, in J. Deploige et al. (eds.), *Hildegardis Bingensis opera minora*, vol. II, CCCM 226A, Turnhout 2016, pp. 153-161.
- Dihle 1955: A. Dihle, *Ein Spurium unter den rhetorischen Werken Ciceros*, «Hermes» 83, 1955, pp. 303-314.
- Eussner 1872: A. Eussner, *Commentariolum petitionis examinatum atque emendatum*, Diss. Virceburgi 1872.
- Forsyth et al. 1999: R.S. Forsyth, D.I. Holmes, E.K. Tse, *Cicero, Sigonio, and Burrows: investigating the authenticity of the "Consolatio"*, «Literary and Linguistic Computing» 14, 3, 1999, pp. 1-26 [https://www.researchgate.net/publication/31407434_Cicero_Sigonio_and_Burrows_Investigating_the_authenticity_of_the_Consolatio].

- Fündling 2006: J. Fündling, *Tullii Cicerones*, in *Brill's New Pauly Online* [http://dx.doi.org/10.1163/1574-9347_dnp_e1222340].
- Gawlick 1994: G. Gawlick, *Cicero: Schriften, Werkbeschreibung*, in H. Flashar (ed.), *Grundriss der Geschichte der Philosophie. Die Philosophie der Antike* 4, 2, *Die hellenistische Philosophie*, Basel 1994, pp. 1015-1053.
- Gudeman 1894: A. Gudeman, *Literary Frauds among the Romans*, «TAPhA» 25, 1894, pp. 140-164.
- Gurney-Gurney 1998: P.J. Gurney, L.W. Gurney, *Authorship Attribution of the Scriptores Historiae Augustae*, «Literary and Linguistic Computing» 13, 3, 1998, pp. 119-131.
- Häfner 1928: S. Häfner, *Die Literarischen Pläne Ciceros*, Diss. Coburg 1928.
- Haug-Jøhndal 2008: D.T.T. Haug, M.L. Jøhndal, *Creating a Parallel Treebank of the Old Indo-European Bible Translations*, in C. Sporleder, K. Ribarov (eds.), *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, pp. 27-34.
- Henderson 1950: M.I. Henderson, *De commentariolo petitionis*, «JRS» 40, 1950, pp. 8-21.
- Henderson 2002: M.I. Henderson, *Introduction*, in Shackleton Bailey 2002.
- Hendrickson 1892: G.L. Hendrickson, *On the Authenticity of the Commentarium Petitionis of Quintus Cicero*, «AJPh» 13, 1892, pp. 200-212.
- Hendrickson 1903: G.L. Hendrickson, *The Commentariolum Petitionis Attributed to Quintus Cicero: Authenticity, Rhetorical Form, Style, Text*, «The Decennial Publications of the University of Chicago» 6, 1903, pp. 71-93.
- Hendrickson 1926: G.L. Hendrickson, *Cicero de Optimo Genere Oratorum*, «AJPh» 47, 1926, pp. 109-123.
- Herbolzheimer 1926: G. Herbolzheimer, *Ciceros rhetorici libri und die Lehrschrift des auctor ad Herennium*, «Philologus» 81, 1926, pp. 391-426.
- Herrmann 1980: L. Herrmann, *L. Annaeus Cornutus et sa rhétorique à Herennius Senecio*, «Latomus» 39, 1980, pp. 144-160.
- Hirst-Feiguina 2007: G. Hirst, O. Feiguina, *Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts*, in «Literary and Linguistic Computing» 22, 4, 2007, pp. 405-417 [<https://doi.org/10.1093/llc/fqm023>].
- Hitschler et al. 2017: J. Hitschler, E. van den Berg, I. Rehbein, *Authorship Attribution with Convolutional Neural Networks and POS-Eliding*, in *Proceedings of the Workshop on Stylistic Variation*, Copenhagen, Denmark, September 7-11 2017, pp. 53-58 [<http://aclweb.org/anthology/W17-4907>].
- Juola 2015: P. Juola, *The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions*, «Digital Scholarship in the Humanities» 30, Supplement 1, 2015, pp. i100-i113 [<https://doi.org/10.1093/llc/fqv040>].
- Kennedy 1972: G. Kennedy, *The Art of Rhetoric in the Roman World 300 B.C. – A.D. 300*, Princeton 1972.

- Kestemont 2012: M. Kestemont, *Stylometry for Medieval Authorship Studies: An Application to Rhyme Words*, «Digital Philology» 1, 2012, pp. 42-72 [[https://www.researchgate.net/publication/236706174 Stylometry for Medieval Authorship Studies An Application to Rhyme Words](https://www.researchgate.net/publication/236706174_Stylometry_for_Medieval_Authorship_Studies_An_Application_to_Rhyme_Words)]
- Kestemont 2014: M. Kestemont, *Function Words in Authorship Attribution From Black Magic to Theory?*, in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL) at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Gothenburg 2014, pp. 59-66. [[https://www.researchgate.net/publication/301404098 Function Words in Authorship Attribution From Black Magic to Theory](https://www.researchgate.net/publication/301404098_Function_Words_in_Authorship_Attribution_From_Black_Magic_to_Theory)].
- Kestemont et al. 2015: M. Kestemont, S. Moens, J. Deploige, *Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux*, «Literary and Linguistic Computing» 30, 2015, pp. 199-224 [[https://www.researchgate.net/publication/274151526 Collaborative authorship in the twelfth century A stylometric study of Hildegard of Bingen and Guibert of Gembloux](https://www.researchgate.net/publication/274151526_Collaborative_authorship_in_the_twelfth_century_A_stylometric_study_of_Hildegard_of_Bingen_and_Guibert_of_Gembloux)].
- Kestemont et al. 2016: M. Kestemont, J. Stover, M. Koppel, F. Karsdorp, W. Daelemans, *Authenticating the Writings of Julius Caesar*, «Expert Systems with Applications» 63, 2016, pp. 86-96 [<http://u.cs.biu.ac.il/~koppel/papers/caesar-paper-eswi.pdf>].
- Kestemont et al. 2018: M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, *Overview of the Author Identification Task at PAN-2018*, in *Working Notes Papers of the CLEF 2018 Evaluation Labs*, Avignon, France, September 10-14, 2018, CEUR Workshop Proceedings 2125, Avignon 2018 [http://ceur-ws.org/Vol-2125/invited_paper_2.pdf].
- Klotz 1851: R. Klotz, *M. Tullii Ciceronis Scripta quae manserunt omnia*, vol. I, Lipsiae 1851.
- Koppel et al. 2004: M. Koppel, J. Schler, D. Mughaz, *Text Categorization for Authorship Verification*, Fort Lauderdale 2004 [<http://rutcor.rutgers.edu/~amai/aimath04/SpecialSessions/Koppel-aimath04.pdf>].
- Koppel-Winter 2014: M. Koppel, Y. Winter, *Determining If Two Documents Are Written by the Same Author*, «Journal of the Association for Information Science and Technology» 65, 2014, pp. 178-187 [[https://www.researchgate.net/publication/259680069 Determining If Two Documents Are Written by the Same Author](https://www.researchgate.net/publication/259680069_Determining_If_Two_Documents_Are_Written_by_the_Same_Author)].
- Kroll 1958: W. Kroll, *Tullius 29. M. Tullius Cicero: Die rhetorischen Schriften*, in RE 7 A 1, 1958, cols. 1091-1103.
- Last 1932: H.M. Last, *The Lex Acilia*, in S.A. Cook, F.E. Adcock, M.P. Charlesworth (eds.), *The Cambridge Ancient History*, IX, Cambridge 1932, pp. 892-896.
- LeCun 2015: Y. LeCun, Y. Bengio, G. Hinton, *Deep Learning*, «Nature» 521, May 2015, pp. 436-444 [<https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf>].

- Leo 1895: F. Leo, *Die Publikation von Ciceros Briefen an Atticus*, in *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, Philologisch-historische Klasse* 1895, pp. 442-450.
- Lintott 1972: A.W. Lintott, Review of Nardo 1970, «JRS» 62, 1972, p. 218.
- Malaspina 2004: E. Malaspina, *Cronologia Ciceroniana in CD-Rom*, in N. Marinone, *Cronologia Ciceroniana*, Roma-Bologna 2004 [= [Ephemerides Tullianae](#)].
- Marriott 1979: I. Marriott, *The Authorship of the Historia Augusta: Two Computer Studies*, «JRS» 69, 1979, pp. 65-77.
- Marx 1894: F. Marx, *Incerti auctoris de ratione dicendi ad C. Herennium libri IV* [M. Tulli Ciceronis ad Herennium libri IV], Lipsiae 1894.
- McDermott 1970: W.C. McDermott, *Commentariolum Petitionis* 2, «Historia» 19, 1970, pp. 384-385.
- Monfasani 1987: J. Monfasani, *Three Notes on Renaissance Rhetoric*, «Rhetorica» 5, 1987, pp. 107-118.
- Münzer 1948: F. Münzer, *Tullius* 31. *Q. Tullius Cicero*, in *RE* 7 A 2, 1948, cols. 1286-1306.
- Murphy-Winterbottom 1999: J.J. Murphy, M. Winterbottom, *Raffaele Regio's 1492 Quaestio Doubting Cicero's Authorship of the Rhetorica ad Herennium: Introduction and Text*, «Rhetorica» 17, 1999, pp. 77-87.
- Nardo 1970: D. Nardo, *Il Commentariolum petitionis. La propaganda elettorale nella ars di Quinto Cicerone*, Padova 1970.
- Nisbet 1961: R.G.M. Nisbet, *The Commentariolum Petitionis: Some Arguments against Authenticity*, «JRS» 51, 1961, pp. 84-87.
- Richards et al. 2016: E.J. Richards, D.J. Wrisley, L. Dulac, *The Different Styles of Christine de Pizan. An Initial Stylometric Analysis*, «Le Moyen Français» 78-79, 2016, pp. 187-206 [<http://hdl.handle.net/2451/43475>].
- Ronconi 1998: F. Ronconi, *De optimo genere oratorum: storia di un abbozzo*, «Appunti romani di filologia» 1, 1998, pp. 43-68.
- Ruhnken 1768: *P. Rutilii Lupi De figuris sententiarum et elocutionis libri duo*, Recensuit et annotationes adjecit David Ruhnkenius, Lugduni Batavorum 1768.
- Rybicki-Eder 2011: J. Rybicki, M. Eder, *Deeper Delta across genres and languages: do we really need the most frequent words?*, «Literary and Linguistic Computing» 26, 3, 2011, pp. 315-321 [<https://doi.org/10.1093/llc/fqr031>].
- Sansone 1990: D. Sansone, *The Computer and the Historia Augusta: A Note on Marriott*, «JRS» 80, 1990, pp. 174-177.
- Schmidt 1969: P.L. Schmidt, *Die Abfassungszeit von Ciceros Schrift über die Gesetze*, Roma 1969.

- Shackleton Bailey 2002: D.R. Shackleton Bailey, *Cicero, Letters to Quintus and Brutus; Letter Fragments; Letter to Octavian; Invectives; Handbook of Electioneering*, Loeb CL, Cambridge MA 2002.
- Shrestha et al. 2017: P. Shrestha, S. Sicara, F.A. González, P. Rosso, M. Montes-y-Gómez, T. Solorio, *Convolutional Neural Networks for Authorship Attribution of Short Texts*, in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Vol. 2 Short Papers, Valencia, April 3-7 2017, pp. 669-674 [https://pdfs.semanticscholar.org/531c/a1835e3fa7c8733568531f9e3ac8e868c99c.pdf?_ga=2.138357938.300873119.1552651915-1491188483.1552651915].
- Stamatatos 2013: E. Stamatatos, *On the Robustness of Authorship Attribution Based on Character N-Gram Features*, «Journal of Law and Policy» 21, 2, 2013, pp. 421-439 [https://www.researchgate.net/publication/303443999_On_the_robustness_of_authorship_attribution_based_on_character_n-gram_features].
- Stamatatos et al. 2014: E. Stamatatos, W. Daelemans, B. Verhoeven, M. Potthast, B. Stein, P. Juola, M.A. Sanchez-Perez, A. Barrón-Cedeño, *Overview of the Author Identification Task at PAN 2014*, in L. Cappellato, N. Ferro, M. Halvey, W. Kraaij (eds.), *CLEF 2014 Evaluation Labs and Workshop – Working Notes Papers*, Sheffield, 15-18 September 2014), pp. 877-897 [<https://ceur-ws.org/Vol-1180/CLEF2014wn-Pan-StamatosEt2014.pdf>].
- Stover-Kestemont 2016a: J.A. Stover and M. Kestemont, *The Authorship of the Historia Augusta: Two New Computational Studies*, «BICS» 59, 2016, pp. 140-157.
- Stover-Kestemont 2016b: J.A. Stover, M. Kestemont, *Reassessing the Apuleian Corpus. A Computational Approach to Authenticity*, «CQ» 66, 2016, pp. 645-672.
- Stover et al. 2016: J.A. Stover, Y. Winter, M. Koppel, M. Kestemont, *Computational Authorship Verification Method Attributes a New Work to a Major 2nd Century African Author*, «Journal of the Association for Information Science and Technology» 67, 2016, pp. 239-242 [<http://u.cs.biu.ac.il/~koppel/papers/apuleius-paper-jasist.pdf>].
- Straka et al. 2016: M. Straka, J. Hajič, J. Straková, *UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing*, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, May 23-28, European Language Resources Association (ELRA) 2016, pp. 4290-4297 [http://www.lrec-conf.org/proceedings/lrec2016/pdf/873_Paper.pdf].
- Sykutris 1962: I. Sykutris, *Epistolographie*, in *RE Suppl.* 5, 1962, cols.185-220.
- Syme 1947: R. Syme, Review of Clift 1945, «JRS» 37, 1947, pp. 198-202.
- Thiele 1889: G. Thiele, *Quaestiones de Cornifici et Ciceronis artibus rhetoricis, Gryphiswaldiae* 1889.
- Till 1962: R. Till, *Ciceros Bewerbung ums Konsulat (Ein Beitrag zum Commentarium petitionis)*, «Historia» 11, 1962, pp. 315-338.

- Tweedie et al. 1998: F. Tweedie, D. Holmes, T. Corns, *The Provenance of De Doctrina Christiana, Attributed to John Milton: A Statistical Investigation*, «Literary and Linguistic Computing» 13, 1998, pp. 77-87.
- Tyrrell 1877: R.Y. Tyrrell, *The Letters of Quintus Cicero*, «Hermathena» 3, 5, 1877, pp. 40-59.
- Tyrrell-Purser 1904: R.Y. Tyrrell, L.C. Purser, *Appendix to the Introduction. The Commentariolum Petitionis*, in *The Correspondence of M. Tullius Cicero, Arranged According to Its Chronological Order*, vol. I, Dublin 1904 [reprint Hildesheim 1969], pp. 116-132.
- Watt 1958: W.S. Watt, *Notes on the Text of the Commentariolum Petitionis*, «CQ» 8, 1958, pp. 32-44.
- Weissenberger 2006: M. Weissenberger, *Gorgias [4]*, in *Brill's New Pauly Online* [http://dx.doi.org/10.1163/1574-9347_bnp_e426320].
- Wikarjak 1966: J. Wikarjak, *Brochure électorale de Quintus Cicéron*, «Archiwum filologiczne» 14, Warszawa 1966.