

# A Bird’s Eye View of Early Modern Latin: Distant Reading, Network Analysis and Style Variation

(by Maciej Eder)

Classical philology—including Neo-Latin studies—is usually claimed to be rather conservative in adopting new trends and methodological paradigms to the study of ancient languages and literatures. Quite surprisingly, however, this claim does not really apply to the “digital turn” in the humanities, which from the very beginning was supported by historical corpora of ancient languages: the *Perseus Project* might serve as the most representative and the most comprehensive example (Crane 2014). Probably more surprising is the fact that Latin and Greek were the first languages to be assessed with statistical methods. Moreover, the same applies to *stylometry*—both the term and a non-traditional method of authorship attribution—which was, in fact, introduced by classical scholars.

As early as the fifteenth century, Leon Battista Alberti published a treatise on frequencies of particular vowels in different Latin genres. Having scrutinized occurrences of particular vowels in various Latin texts, he concluded that *a* and *e* are particularly frequent in poetry, while the other vowels are typical for rhetorical orations (Ycart 2013). In a study that became both a famous example of early modern *empirical philology* and a milestone in authorship attribution using quantitative methods, Lorenzo Valla dealt with the *Donation of Constantine*, a putative edict in which emperor Constantine transferred authority over Rome and the western part of the Roman Empire to the Popes. Valla’s *De falso credita et ementita Constantini donatione declamatio* (Valla 1440, translated 2007) was a detailed analysis of syntax, morphology and

lexis, showing without any doubt that the text could not have been written in the fourth century due to numerous anachronisms and grammatical idiosyncrasies typical for a much later period.

The nineteenth-century advent of non-traditional authorship attribution based on statistical analysis of language is also strongly connected to Greek and Latin philology—more than one could expect. Among the foundations of stylometric theory not only are Shakespearean studies published by Augustus de Morgan and Thomas Mandelhall, but also approaches to Pauline Epistles as conducted by William Benjamin Smith, also known as Conrad Mascol (cf. Holmes 1998, 112; Rudman 1998a, 354). As early as in 1867, Lewis Campbell, Professor of Greek at St. Andrews, devised a series of statistical tests for a new chronology of *Sophist* and *Politicus*; these included word order, rhythm, avoidance of hiatus, and “originality of vocabulary” as measured by the frequency of once-occurring words (*hapax legomena*). Using these methods, Campbell ascertained the two dialogues as late works by Plato. Campbell’s discovery went unnoticed for the next thirty years; it seems likely therefore, that Constantin Ritter’s study on Platonic chronology of 1888, which presented similar methods, was conceived independently of Campbell’s. Seminal founders of quantitative authorship attribution also include the inventor of the term *stylometry* and a scholar who proposed a novel method of inferring the chronology of Plato’s dialogues, Wincenty Lutosławski (1897; cf. Pawłowski and Pacewicz 2004). Even if Lutosławski’s works today are known only to the most sophisticated experts in Plato, the impact of the term *stylometry* has never been questioned.

While a full analysis of recent scholarship in Latin and Greek is beyond the scope of this essay, it is important to note that approaches to authorship cover many different genres and literary

periods. They include the Greek New Testament (Greenwood 1995), the collection of biographies of the Roman Emperors known as *Historia Augusta* (Gurney and Gurney 1998, Rudman 1998b), medieval visions ascribed to Hildegard of Bingen (Kestemont et al. 2013), the philosophical work entitled *Consolatio*—allegedly written by Cicero, and actually by Carlo Sigonio (Forsyth et al. 1999)—or John Milton’s *De doctrina Christiana* (Tweedie et al. 1998), to name but a few studies.

A vast majority of these approaches aimed to answer precisely defined questions: “who wrote a given work?” or “what was the exact chronological order in a set of extant texts?” For decades, literary scholars expected statistics to provide an objective verification of non-literary facts (e.g. detecting plagiarism); the questions specifically related to literature itself were somehow kept away from computational algorithms. However, even if the ability of solving mere yes/no problems does not really contribute to our knowledge of literature, statistical techniques might become an attractive extension to the repertoire of time-proven heuristic routines when they support *interpretation* of observed phenomena. As Hugh Craig asks: “If you can tell authors apart, have you learned anything about them?” (1999, 2004). Probably the most seminal study that brought stylometry from the field of language technology back to literature was John Burrows’s monograph on the novels by Jane Austen, in which the concept of “computation into criticism” was introduced (Burrows 1987), later extended to that of “algorithmic criticism” (Ramsay 2011). Further expansion of exact methods in the field of literary studies was substantially stimulated by Franco Moretti’s “distant reading” (Moretti 2007, 2013).

The present study is directly inspired by the above theoretical propositions. It aims to examine

the ways in which digital techniques can support scholarship on early modern Latin in its relation to earlier literary epochs. Particularly, the questions of imitation, intertextuality, stylistic differentiation affected by genre and/or chronology, the impact of authorial stylistic idiosyncrasies on the general picture of Latin literature, etc., will be addressed. To this end, a new stylometric method will be applied (Eder 2015b). The method, discussed below in detail, allows fitting numerous mutual text relations in a single network-like visualization. Like most of the stylometric procedures that have been introduced in recent decades, the method in question is closely related to authorship attribution, and so only a brief presentation of theoretical foundations in computer-assisted authorship inference is required here. Next, the idea of scaling up stylometric techniques will be presented. Finally, the new method will be tested on a relatively large collection of 150 Latin works covering the span of a dozen of centuries.

#### *From Attribution to Computational Stylistics*

Function words (articles, particles, prepositions, pronouns) are far more frequent than content words, and thus their frequencies can be reliably assessed using statistical procedures (Zipf 1949; Burrows 1987). Apart from their measurability, function words are used subconsciously (Chung and Pennebaker 2007), which makes them a perfect feature betraying individual writing habits. Since they are beyond authorial control, they cannot be easily plagiarized; it is difficult to manipulate the usage of the words that no one really notices. In a groundbreaking study on the collection essays—some of them written collaboratively—known as the *Federalist Papers*, Mosteller and Wallace counted the frequencies of selected function words and, using Bayesian inference, determined with a very high probability which of the disputed essays had been written by Alexander Hamilton, and which by James Madison (2007 [1964]). Since then, most frequent

words (MFWs) have become a classic way to distinguish authorial “fingerprint.” The list of MFWs is fairly similar across different languages. For English, these are: *the, and, to, of, a, I, in*, and so forth. For Latin, the frequency list begins with the following words: *et, in, non, est, ut, ad, cum, quod, ...* (a list of the first 250 MFWs in Latin is provided in Appendix B).

Needless to say, frequencies of the function words are a counter-intuitive indicator of stylistic differentiation, regardless of their actual strong discriminative power. It is counter-intuitive because style as defined by stylometry is entirely different from our usual understanding of this term. Being profoundly different, however, both points of view—computational and traditional—in fact represent two sides of the same coin. The style in its entirety is a multi-faceted phenomenon. Traditional stylistics tries to capture it holistically, while computational stylistics is focused on few aspects, or just one aspect: in most applications this is the usage of function words. Interestingly, this is the very aspect of style ignored by traditional scholarship. Thus, bearing in mind the obvious fact that, say, the elaborated Ciceronian style cannot be simply turned into countable numbers, the present study will apply statistical measures to find out whether stylometric similarities confirm a few time-proven stylistic hypotheses concerning the relations between Cicero and his early modern followers.

There are several statistical techniques used in authorship attribution. Some of them focus on one phenomenon carefully retrieved from a corpus—as do different indexes of vocabulary richness—while others rely on a large number of features (e.g. word frequencies) computed at once. The latter are called *multidimensional* and they are claimed to be much more sensitive to nuanced differences between samples. The reason for their attributive value is the fact that they aggregate

the impact of many linguistic features of individually weak discriminating strength (Nerbonne 2007, xvii). Minute differences in word usage are combined into an overall difference between two texts, or a *distance*; this measure of similarity is computed for every single pair of samples in the corpus. Now, if an anonymous text is compared against a set of samples written by known authors, the sample showing the closest (abstract) distance is considered to be stylistically the most similar to the anonymous one, and hence probably written by the same author. Such a concept of *nearest neighbors* is fundamentally important in authorship attribution.

Stylometric methodology, developed to solve authorship problems, can easily be extended and generalized to assess different questions in literary history. Explanatory multidimensional methods, relying on distance measures and supported with visualization techniques, are particularly attractive for this purpose. Namely, the underlying idea of tracing stylistic neighbors between texts can be extended to capture textual relations in the entire corpus: since attribution involves comparing an anonymous (or disputed) text against a selection of works written by known authors, it can be also used to compare the level of similarity between, say, poetry and prose. An example of such an approach to early modern Latin is a statistical comparison of epistolary style of Francesco Filelfo versus the style of Cicero (Deneire 2015).

The recent advances in stylometric methodology show that two general directions of extending the attribution techniques attract a good share of attention. Following the famous metaphor of digital humanities as a telescope for the mind, one might compare these two ways of stylometric investigations to two different optical instruments (McCarty 2012, 113). One direction can be compared to using a microscope, while the other is like using a telescope. The microscopic

approaches try to look inside a single literary work represented as a sequence of text chunks, in order to examine stylistic consistency of these chunks. This way of reasoning has been applied to reveal the nature of collaboration of Joseph Conrad and Ford Madox Ford (Rybicki et al. 2014), to find authorial takeovers in medieval Dutch poem entitled *Roman van Walewein* (van Dalen-Oskam and van Zundert 2007), and to inspect stylistic breaks in *Roman de la Rose* (Eder 2015a).

The telescopic way of extending stylometric methods involves scaling up the amount of input data many times, in the belief that a large-scale perspective will betray some new literary phenomena unnoticeable in close reading. Apart from Franco Moretti, the pioneers of this approach include Matthew Jockers with his “macroanalysis,” or massive comparison of 3,346 English novels from the 19<sup>th</sup> century (2013). In his own words, “the literary scholar of the twenty-first century can no longer be content with anecdotal evidence, with random ‘things’ gathered from a few, even ‘representative’, texts” (2013, 8).

The applicability of this vision to the study of early modern Latin will be examined below. However, since macroanalysis has somewhat underestimated the problem of validation of the obtained results, a novel method will be applied. One of its core features is the ability to evaluate particular links between analyzed texts. The underlying idea of the method is quite simple. If an anonymous text scrutinized using an attribution procedure turns out to be the most similar to one of the texts from the reference corpus, then, by extension, the same rule applies to any collection of texts of known authorship. In short: every single text from the corpus must have its own nearest

neighbor.<sup>1</sup> The neighbors of the neighbors are neighboring with further neighbors, and so on, resulting in a dense network of mutual links. These links can be then used to map the relations between groups of texts.

### *From Computational Stylistics to Network Analysis*

Distant-reading approaches to stylometry face at least three methodological issues. The most important one is the question whether a given method, reasonably effective for a collection of, say, 25 texts, can be scaled up to assess dozens or hundreds of texts without any significant side-effects. The performance of multidimensional methods depends on the number of frequent words taken into analysis (Koppel et al. 2009, Jockers and Witten 2010, Rybicki and Eder 2011); the results can be further affected by the choice of linkage algorithm (e.g. in hierarchical cluster analysis), and generally by a variety of other input parameters of the experiment. In attribution studies, the problem of unstable results must be even more inescapable in large datasets.

Secondly, fitting the relations between, say, a hundred texts becomes problematic for typical explanatory methods used in stylometry, such as multidimensional scaling or principal components analysis. When the number of texts exceeds a few hundred, the urgent need for new ways of visualization becomes obvious. Also, a balance between reliability, informativeness, and visual attractiveness should be kept in mind.

---

1 Note that “neighbor” means, in this case, the most similar text regardless of the absolute value of that similarity. Unnoticeable in a vast majority of cases, this issue might (sometimes) lead to biased results. To give an example: when one collects a random selection of works written in, say, different languages, these works will have their nearest neighbors anyway. Harvesting a corpus, one should be aware of this commonly-known drawback of nearest neighbor methods.

The third issue is even more important, at least from the viewpoint of stylometry beyond attribution; namely, when one tries to represent nearest neighbor relations between samples, authorial groupings are most likely to appear, but other stylistic signals will probably be filtered out. It has been shown that the authorial voice usually overwhelms other stylistic layers, related to genre, topic, translation and so forth (Craig 2004, Rybicki 2012, Schöch 2013). However, while it is interesting to see Bacon's *Historia regni Henrici Septimi* linked to his *Novum Organum*, and Erasmus' *Moriae encomium* recognized to be stylistically similar to *Institutio principis Christiani*, these results are neither novel nor surprising. Instead, one would like to know what is the relation (if any) between Erasmus and Bacon. In other words, computational stylistics should reveal not only the obvious connections between the works written by the same authors, but also subtle stylistic links that cannot be discovered with the naked eye.

The new method applied in this study aims to overcome the aforementioned three issues: unstable results, inability to create large-scale visualizations, and authorial voice overriding other stylometric signals. The technique combines the concept of *network* as a way to map large-scale literary similarities (Jockers 2013), the concept of *consensus* to find statistically significant relations (Lancichinetti and Fortunato 2012), and the assumption that textual similarities usually go beyond mere nearest neighborhood.

The new technique relies on the assumption that particular texts can be represented as nodes of a network, and their explicit relations as links between these nodes. The most significant difference, however, between the approaches applied so far and the present study is the way in which the nodes are linked. This new procedure of linking is twofold. One of the involved

algorithms computes the distances between analyzed texts and establishes, for every single node, a strong connection to its nearest neighbor (i.e. the most similar text) and two weaker connections to the 1<sup>st</sup> and the 2<sup>nd</sup> runner-up (i.e. two texts that get ranked immediately after the nearest neighbor). The second algorithm replicates the original test many times with slightly altered input parameters of the experiment, in order to generate multiple “snapshots” of the corpus (e.g. the test is replicated using 100 MFWs, then 200, and then, at increments of 100, all the way to 2,000 MFWs). Finally, all the connections produced in particular “snapshots” are added, resulting in a consensus network. The weights of these final connections tend to differ significantly: the strongest ones mean robust nearest neighbors, while weak links represent secondary and/or accidental similarities. The results are self-validated because they rely on the consensus of many single approaches to the same corpus sanitizes robust textual similarities and filters out spurious clusterings. The above two-fold procedure of linking text samples is implemented in the package “stylo,” an open-source stylometric library written in the R programming language (Eder et al. 2013).<sup>2</sup>

The next crucial step in network analysis is to arrange the nodes on a plane in such a way that they reveal as much information about linkage as possible. In the present study, one of the force-directed layouts was chosen, namely the algorithm ForceAtlas2 embedded in Gephi, an open-source tool for network manipulation and visualization (Bastian et al., 2009). Force-directed layouts perform gravity-like simulation and pull the most-connected nodes (i.e. the ones that have several links and/or their links are very strong) to the center of the network, while the least

---

<sup>2</sup> The package is available at the Computational Stylistics Group website (<https://sites.google.com/site/computationalstylistics/>) or directly from CRAN repository (<http://cran.r-project.org>). A detailed description of the whole technique of establishing and evaluating the links is provided in a separate study (Eder 2015b).

connected nodes are pushed outward. The two-dimensional plot representing the nodes—or particular texts—can be interpreted by a closer look at the network; any clear groupings and visual separations between clusters are meaningful.

### *Dataset*

Large-scale stylometric methods seem to be particularly applicable to the study of Latin literature in its long term, for several reasons. First and foremost, Latin is one of very few languages used for many centuries in a relatively fixed form. This gives us a unique opportunity to compare (stylistically) Latin literature from different epochs, and to trace development of style over time. Next, the situation of written Latin was quite different in (i) the Antiquity, where it was basically a high register of a spoken mother tongue, (ii) in the Middle Ages, where mastering Latin required some effort to escape one's own vernacular language, (iii) in the Renaissance, where the humanists undertook the task to artificially bring Latin back to its ancient flavor, and (iv) in subsequent two centuries, where Latin became a *lingua franca* in scholarly and scientific contexts. This gradual shift from a native to a foreign language offers the perfect fodder for stylometric investigation.

Another reason Latin is particularly appropriate for large-scale stylometric research is the relatively good availability of Latin texts on the internet in high-quality critical editions—especially when compared with other historical languages. Furthermore, unlike modern literatures, the entire corpus of Latin is fixed: the number of texts ever written can be reliably estimated, which makes any extrapolations feasible, even if numerous early modern works are

not digitized yet. Last but not least, Latin literature has been thoroughly researched by generations of scholars. Such abundant close-reading evidence is a great help for a stylometrist because it can be used to validate the results of a computational approach.

Ideally, a distant-reading study should involve large amounts of textual data. The present paper offers a first step towards such an experiment in Latin, and thus it is based on a subset of 150 texts rather than on an exhaustive corpus. The texts were harvested from three open-access databases: the *Perseus Project* (<http://www.perseus.tufts.edu>), *The Latin Library* (<http://www.thelatinlibrary.com>) and the *Bibliotheca Augustana* (<https://www.hs-augsburg.de/~harsch/augustana.html>). They include a selection of ancient writers (such as Seneca, Tacitus, Cicero), early Christian and medieval authors (such as Arnobius, Ambrosius, Thomas Aquinas), as well as early modern ones (such as Erasmus, Morus, Petrarca, Kepler). The complete list of authors and works is provided in Appendix A. It has to be emphasized, however, that the selection of texts is far from complete or fully representative. This is partly because the availability of Latin literature varies widely across the centuries (also for copyright issues), and partly because historical corpora will never be fully balanced.

Even if 150 texts is not a very big corpus, it shares some of the characteristics of big data collections. In particular, one cannot inspect all the texts manually and/or emend the transcription. Also, one cannot reliably exclude all external quotations, passages copied from other sources, and similar intertextual links; the same applies to any instances of (hidden) mixed authorship. Last but not least, one cannot normalize all spelling variants, e.g. *qua re* vs. *quare*, *quam ob rem* vs. *quamobrem*, *obedientia* vs. *oboedientia*, and similar instances. In such cases big

data means big noise. Some attempts to normalize the spelling variation were undertaken, though: the most important one included automatic replacement of all the letters *v* with *u* in the entire corpus, in order to neutralize the impact of different scribal and editorial traditions.

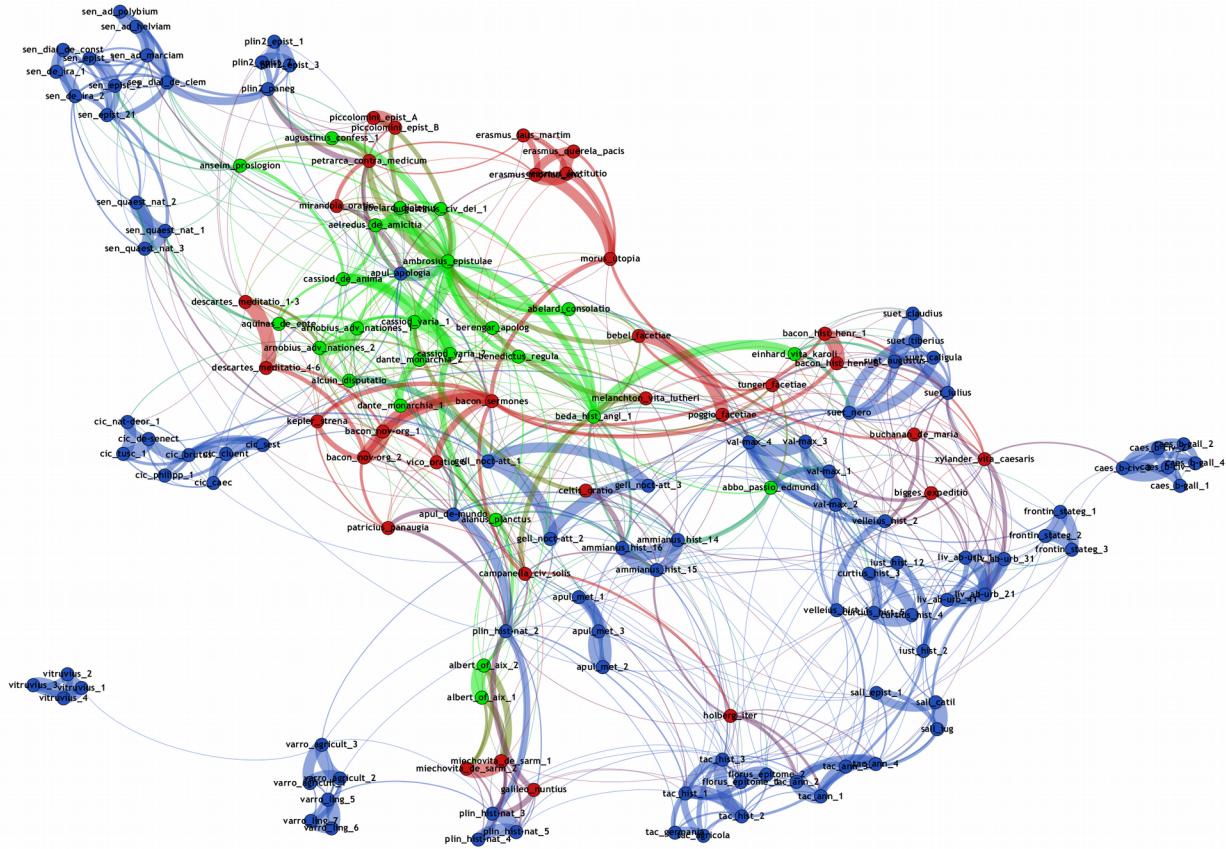
Analyzing Latin style with stylometric methods, one should also remember that the medieval authors relatively often cite the Bible and related sources, while the humanists' treatises are full of explicit and/or implicit quotations from classical literature. More importantly, the humanists consciously tried to avoid medieval vocabulary in favor of words that were used by Cicero. For that reason, a stylometric comparison of medieval and early modern Latin brings some additional issues, intensive text re-use being one of the most important (Eder 2013).

### *Results*

Using the consensus network procedure discussed above, a 150-node graph was computed (Figs. 1–4). A few groupings and separations can be identified at first glance: a distinct cluster for Seneca in the top-left corner, Caesar's oeuvre clearly distinguishable at the right side of the network, and so on. As expected, the strongest connections are within authorial clusters, which confirms a predominance of authorial signal in the dataset. Identifying these and similar connections might lead to interesting observations. However, the real strength of network analysis comes to the fore when a graph is contrasted against metadata (prior literary knowledge) and supported by literature-oriented research questions.

St. Jerome, early-Christian writer and translator of the Bible, claims that he had a dream in which God accused him: "You are a Ciceronian, not a Christian!" (*Epist. 22, 30*), because he had paid

too much attention to the beauty of the Ciceronian style. St. Jerome's famous dream reflects Christian antiquity's general attitude to classical literature: pagan texts were claimed to be generally dangerous, and thus they were rarely imitated (Bolgar 1954). Centuries later, Renaissance humanists "discovered" the classical authors again, and they intended to purge Latin language from medieval traces. It can be hypothesized, then, that these changes in the attitude to the classical literature were followed by style breaks measurable with stylometric methods.



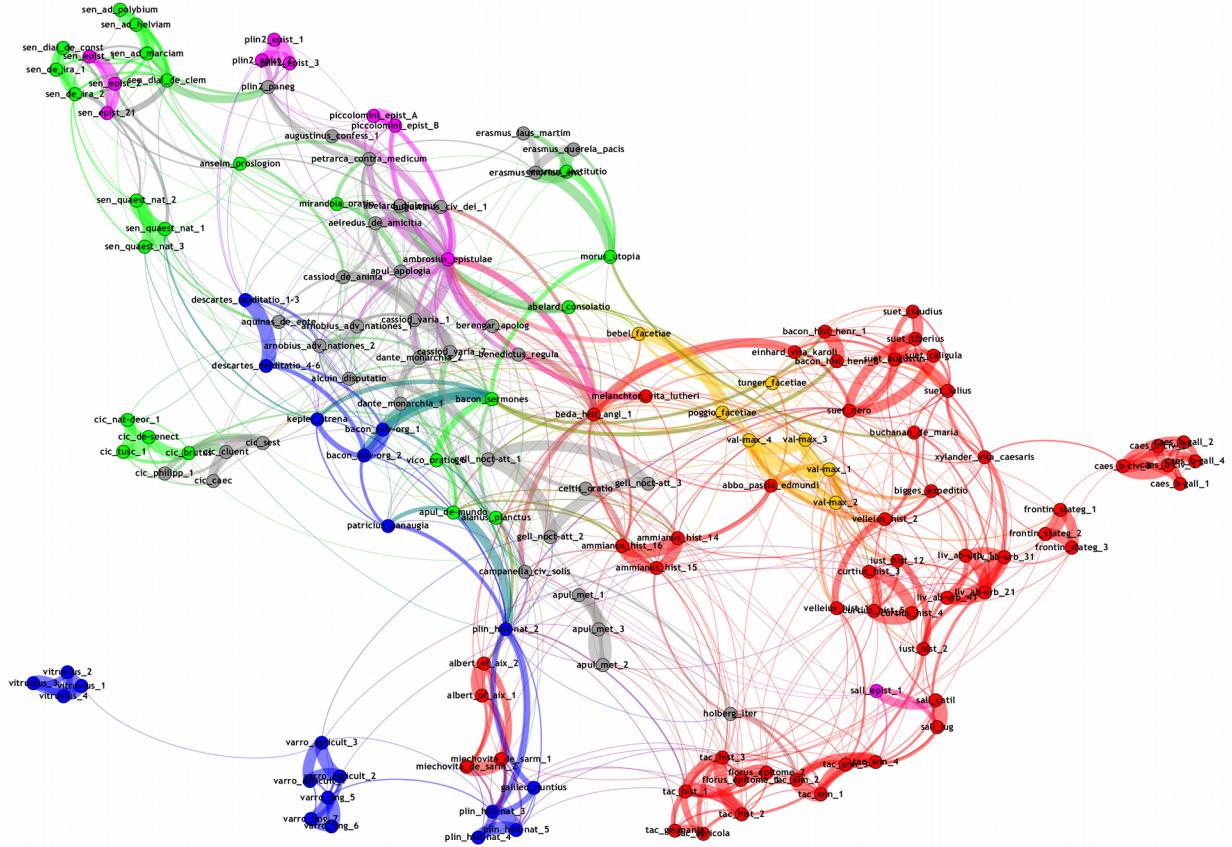
[Insert figure 1 here] Figure 1. Consensus network of 150 Latin texts, colored according to chronology: ancient texts are colored with blue, medieval texts with green, and early modern

ones with red<sup>3</sup>

On the network represented in figure 1, the nodes are colored according to chronology (ancient texts marked blue, medieval and early Christian works green, and early modern ones red). The first observation is that the colored samples—or the works written in temporal proximity—tend to form homogeneous regions on the graph. Medieval texts occupy the central area of the network, surrounded by early modern works; these are further surrounded ancient writers. However, a closer look reveals several exceptions to this apparently simple picture: *Apologia* by Apuleius (c. 124–c. 174) found its place in the center of the medieval cluster, while *Historia Hierosolymitana* by Albert of Aix (floruit c. 1100) is far away from the Middle Ages, and the same applies to *Passio sancti Edmundi* by Abbo of Fleury (c. 945–1004). Last but not least, several early modern texts appear scattered around the network. It means that the hypothesis of chronological development of style was not fully corroborated, even if some temporal patterns can be observed. Alternative stylistic signals should be examined, then.

---

<sup>3</sup> For a zoomable version of this figure, see <http://ems.itercommunity.org/essays/birds-eye-view-early-modern-latin-distant-reading-network-analysis-and-style/>



[Insert figure 2 here] Figure 2. Consensus network of 150 Latin texts, colored according to genre: history writers are colored with red, philosophical works with green, epistles with purple, collections of anecdotes with yellow, other non-fiction works with blue; some works are left unmarked due to their ambiguous genre classification<sup>4</sup>

In figure 2, the same network is colored according to genre: history (red), philosophy (green), epistles (purple), “technical” writings such as Vitruvius’s *De architectura* (blue), apophthegms and anecdotes (yellow). Some texts were left unmarked due to ambiguous classification or because they represented other genres, e.g. Apuleius’s *Metamorphoses* (literature) or Cicero’s *Philippic* 1 (rhetorical oration). The pattern seems to be much clearer than in the previous plot. In

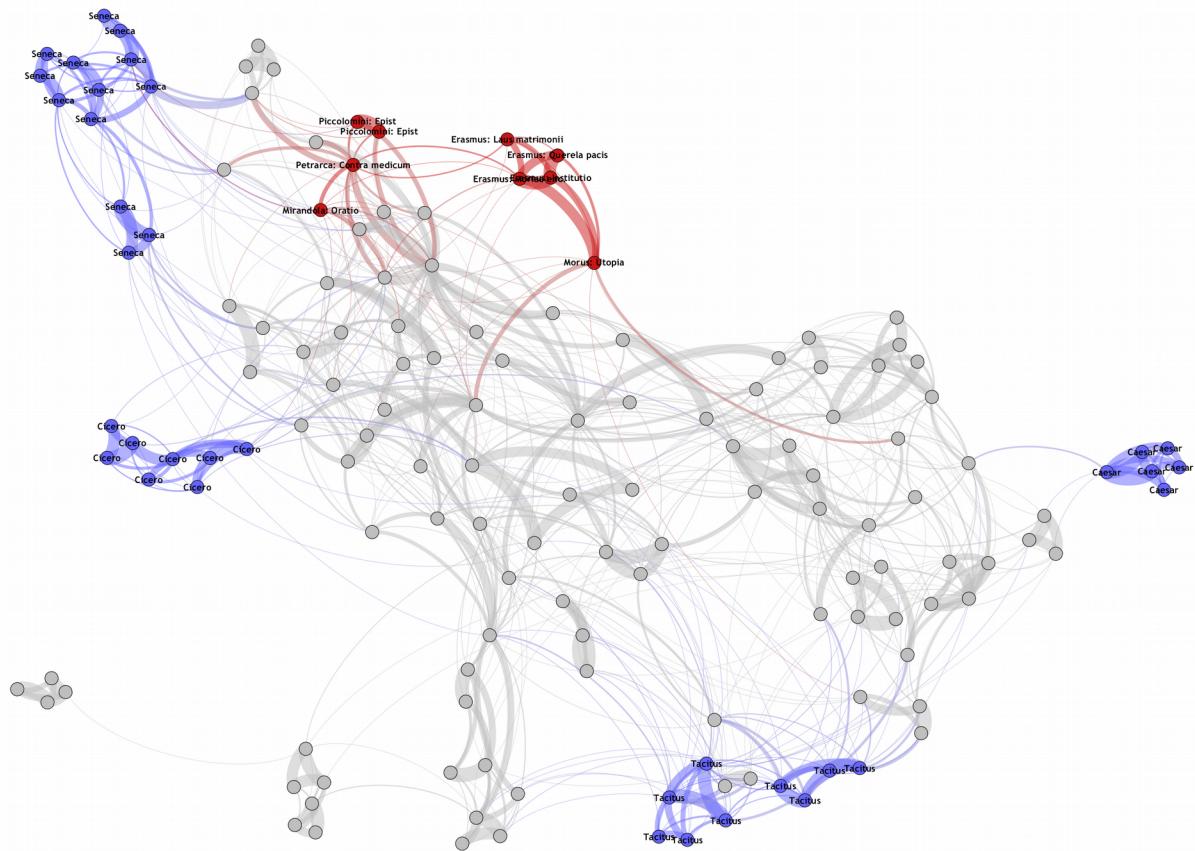
<sup>4</sup> For a zoomable online version of this figure, see <http://ems.ITERcommunity.org/essays/birds-eye-view-early-modern-latin-distant-reading-network-analysis-and-style>

particular, notice a large cluster of history writers that appeared on the right side. This highly homogeneous cluster containing historical works regardless of which period they represent suggests that historiography is stylometrically distinct and stable across centuries. Interestingly, the central part of this cluster is occupied by authors of *facetiae*, *apophthegms*, and *anecdotes*: Valerius Maximus, Poggio, Bebel, Tünger. They all seem to be very similar and—what is surprising—they seem to have been successful in imitating the style of the historians.

The other part of the plot is less distinct, even though one can notice that the “technical” writers tend to be linked together, and the same applies to the philosophers. At the same time, however, epistles do not form any cluster: Sallust’s letters are grouped with his two historical works, the letters by Pliny the Younger are close to his *Panegyricus*, Seneca’s *Epistulae ad Lucilium* are linked to his dialogues, and Ambrose’s epistolography is linked to a good share of other authors. The behavior of the epistles clearly indicates the predominance of the authorial signal over genre or topic. Paradoxically, this is also confirmed by the ambiguous results for Ambrose’s oeuvre: being a collection of letters both written by him and addressed to him, the text turns out to have a weak authorial signal, which is reflected by multiple yet unclear network connections.

The case of Ambrose prompts another research question: namely, how to interpret centrality and eccentricity of the network. Since the layout algorithm puts the heavily connected nodes towards the center and pulls outward nodes with only a few links, the actual question is how to interpret the density of connections. In figure 3, colors are applied to the works by the authors traditionally claimed to have mastered Latin style: Cicero, Seneca, Tacitus, Caesar, Petrarch, Pico della Mirandola, Erasmus, and More (early modern authors with red, ancient ones with blue). A

clear pattern emerges: these very authors occupy outer regions of the network. Varro is similarly eccentric, and Vitruvius even more so, i.e. two authors that have not been particularly appreciated for their style. A provisional interpretation of this phenomenon might be that a small number of connections indicate extraordinary authors—either great stylists, or bizarre writers. Or it could be interpreted the other way around: several connections might mean a less distinct author, with a weak authorial voice. This provisional explanation needs to be confirmed by stronger evidence, though.



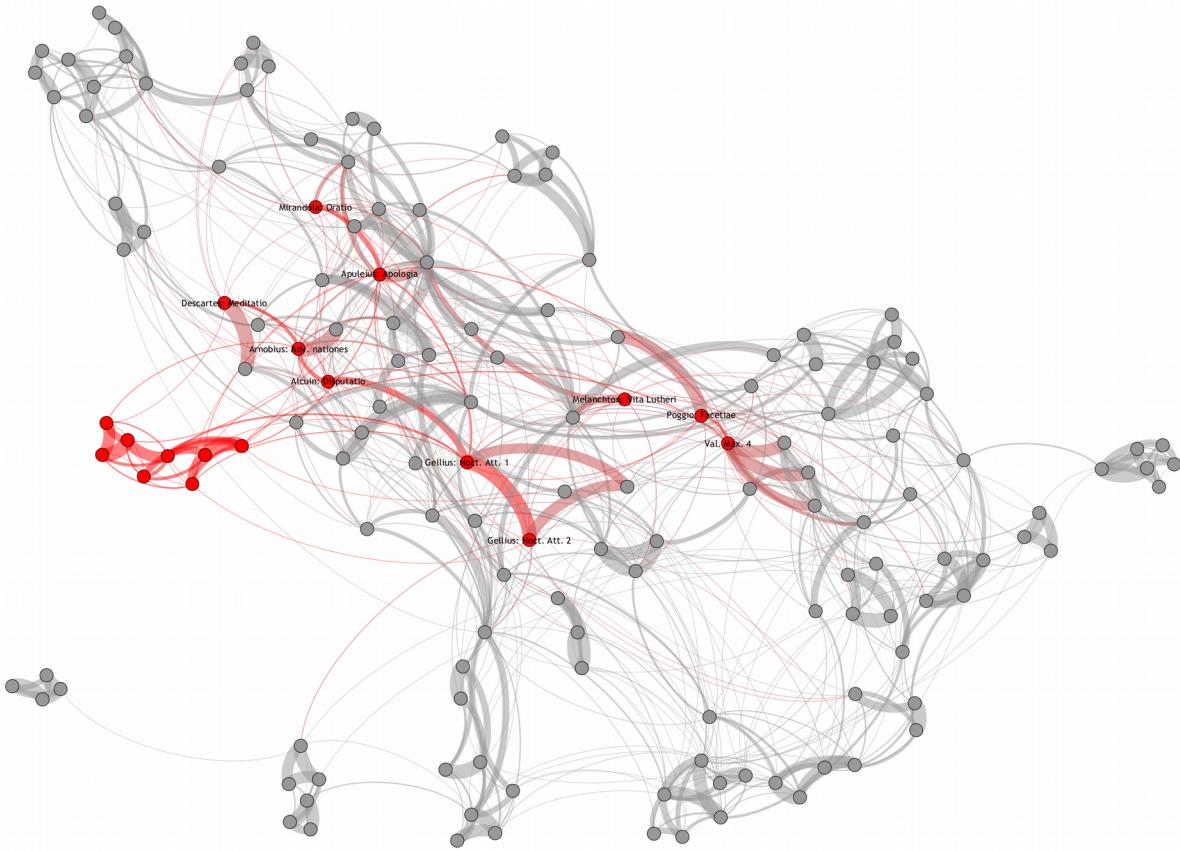
**Insert figure 3 here]** Figure 3. Consensus network of 150 Latin texts, with highlighted works of authors traditionally claimed to be great stylists; early modern authors colored with red, ancient

ones with blue<sup>5</sup>

The density of connections between different authors deserves yet another research hypothesis. Namely, it might reflect some traces of imitation, which is a fundamentally important aspect of early modern stylistic theories. It is a well-known fact that Latin style had a few flavors, the most discussed being the distinction between the style of Ciceronian and the Silver Latin style as represented by Seneca and Tacitus. In the Renaissance, a discussion of which of the two flavors should be imitated turned into the Ciceronian Quarrel, the single most important linguistic debate of that time (DellaNeva 2007). The debate had an immense effect on the development of Latin style. From the perspective of computational stylistics, a particularly interesting task might be an examination of the problem of “Attic” prose and the anti-Ciceronian movement of the late Renaissance and early Baroque, based on the analysis of the style of Justus Lipsius, Erasmus, and other writers such as Puteanus, Moretus, and Fredro (Croll 1924, 1996; Salmon 1980; Tunberg 1999). The questions to be answered stylometrically might be as follows: are traces of Seneca’s and Tacitus’ style indeed noticeable in this modern “Attic” way of writing? Did the “Attic” authors really escape stylistic Ciceronianism? And finally: how influential was Cicero, really?

---

<sup>5</sup> For a zoomable online version of this figure, see <http://ems.iltercommunity.org/essays/birds-eye-view-early-modern-latin-distant-reading-network-analysis-and-style>



[Insert figure 4 here] Figure 4. Consensus network of 150 Latin texts, with highlighted works by Cicero and all the works with which Cicero is directly connected<sup>6</sup>

While it is still too early to assess the problem of imitation in its entirety using computational techniques, the collected corpus of 150 prose texts allows us to examine the relations between Cicero and his followers. Since consensus networks rely on the concept of nearest neighbor as derived from authorship attribution, any two nodes that are linked are in consequence *very* similar stylistically. Now, if one identifies all the nodes connected directly to Cicero, it should

<sup>6</sup> For a zoomable online version of this figure, see <http://ems.itercommunity.org/essays/birds-eye-view-early-modern-latin-distant-reading-network-analysis-and-style>

give a relatively clear picture of the circle of his followers. In figure 4, the nodes linked directly to any of the works by Cicero were colored red. The highlighted texts include: Gellius' *Noctes Atticae*, Apuleius' *Apologia*, Valerius Maximus' *Facta et dicta* book 4, Arnobius' *Adversus nationes*, and Alcuin's *Disputatio*. This selection of ancient and medieval writers related to Cicero, being noteworthy from a general perspective, is slightly out of scope of this paper, though. From the point of view of Renaissance Ciceronianism, early modern authors and texts identified by the procedure are much more interesting: Descartes' *Meditationes*, Pico della Mirandola's *Oratio de dignitate hominis*, Melanchthon's *Vita Lutheri*, and Poggio's *Facetiae*. All these works are surprisingly non-surprising, as is the absence of Erasmus or Morus among the stylometric followers of Cicero. Pico della Mirandola was a well-known Ciceronianist, the same can be said about Poggio—even if his *Facetiae*, a collection of obscene and scatological jokes, is hardly comparable to the Ciceronian oeuvre at the first glance. Melanchthon, on the other hand, received a thorough humanistic education, which also included stylistic excellence based on imitating the Ciceronian style. The absence of Morus, and particularly Erasmus seems to suggest that their anti-Ciceronianism as a stylistic choice, were quite successful. In his well-known dialogue entitled *Ciceronianus*, Erasmus claims that imitating Cicero should be but the first step to achieve actual stylistic excellence; too close an imitation is a form of literary idolatry.

Being somewhat preliminary, the above results definitely deserve to be extended and systematically contrasted with traditional scholarship on Latin writers. Also, future studies should take into consideration a substantially bigger corpus of Latin literature.

### *Conclusions*

Stylometric methods based on the most frequent words turned out to be fairly successful, as evidenced in this reconnaissance study. When combined with attractive ways of visualizing the results—network analysis being one of the possible choices here—they allow for assessing large amount of textual data. More importantly, they allow for asking new research questions. Even if this study was focused on Latin literature, the method presented in this paper can be easily generalized to scrutinize other literary traditions. Also, very similar questions about chronology, genre, topic, sentiment, and other stylistic layers, can be formulated in the context of, say, Shakespeare canon or medieval French *Chansons de Geste*. The biggest added value of the method applied above, however, is the fact that it can be used to map a given literary tradition in its entirety—no matter how many texts are analyzed, they all can be handled by the method and represented in a form of a network (a map), in which some potentially interesting groups of texts are very likely to appear. Presumably, such a bird's eye view provides an insight to literary phenomena that is unavailable using close-reading approaches.

Computational stylistics proved a promising addition to the usual repertoire of explanatory methods. At the same time, however, an obvious gap between distant and close reading became embarrassingly evident: from a bird's eye view, one has no access to the texts themselves. It seems, then, that an optimal approach to literature might be a combination of these two perspectives.

In particular, traditional scholarship might take advantage of the by-products of stylometric experiments to pursue more detailed investigation of stylistic differentiation between authors. For instance, a list of word frequencies, obtained in split-seconds using computational

techniques, provides interesting material for further close reading and interpretation. To give an example: if one examines which of the words are shared across the entire corpus of 150 Latin texts, one gets a surprisingly low number of merely 24 shared words (cf. Appendix C). Certainly, the shortlist contains function words, but a careful comparison reveals that some very common function words are missing (Appendix C vs. Appendix B). In this point close reading comes to center stage, to reveal that the conjunction *si*, which is the 14<sup>th</sup> word in the entire corpus, is absent in Matthias of Miechow's *De duabus Sarmatiis tractatus*; the 17<sup>th</sup> word *aut* is not used by Apuleius in *Metamorphoses*; Thomas Aquinas (*De ente et essentia*) never uses *ac* and *atque*; the word *nec* is avoided by Sallust and Caesar. These and similar idiosyncrasies in word usage are usually overlooked by traditional approaches; stylometry helps to bridge the gap.

The relation goes both ways: hypotheses formulated by literary criticism, based on text-centric evidence, are a sine qua non to make stylometric experiments valid. State-of-the-art statistical procedures are very accurate, but they will fail when applied to answer spurious questions. Based on solid scholarship, however, stylometric techniques might be leveraged to answer distant-reading questions, with a belief that at some point, one will be able to produce a map of Latin literature in its entirety. The present study is intended to be a first step towards this vision.

## Works Cited

- Bastian, Mathieu, Sébastien Heymann and Mathieu Jacomy. 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *International AAAI Conference on Weblogs and Social Media*, <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- Bolgar, Robert R. 1954. *The Classical Heritage and its Beneficiaries*. Cambridge: Cambridge

University Press.

Burrows, John. 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.

Campbell, Lewis. 1867. *The Sophistes and Politicus of Plato. With a revised text and English notes by Lewis Campbell*. London: Clarendon Press. Reprinted New York: Arno Press, 1973.

Craig, Hugh. 2004. "Stylistic Analysis and Authorship Studies." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 273-88. Oxford: Blackwell.

Crane, Gregory R., ed. 2014. *Perseus Digital Library*. Tufts University.  
<http://www.perseus.tufts.edu>.

Chung, Cindy, and James W. Pennebaker. 2007. "The Psychological Functions of Function Words." In *Social Communication*, edited by K. Fiedler, 343-59. Psychology Press.

Croll, Morris W. 1924. "Muret and the History of 'Attic' Prose." *PMLA* 39: 254-309.

Croll, Morris W. 1996. "'Attic' Prose in the Seventeenth Century." In *Style, Rhetoric and Rhythm*, edited by J. M. Patrick and R. O. Evans, 51-101. Princeton NJ: Princeton University Press.

Dalen-Oskam, Karina van and Joris van Zundert. 2007. "Delta for Middle Dutch: Author and Copyist Distinction in Walewein." *Literary and Linguistic Computing* 22 (3): 345-62.

DellaNeva, JoAnn (ed.). 2007. *Ciceronian Controversies*. English translation by Brian Duvick. Cambridge, MA: Harvard University Press.

Deneire, Tom. 2015. "Filelfo, Cicero and Epistolary Style: A Computational Study." In *Filelfo: Man of Letters*, edited by Jeroen De Keyser. Leiden and Boston: Brill (in press).

Eder, Maciej. 2013. "Mind Your Corpus: Systematic Errors in Authorship Attribution." *Literary and Linguistic Computing* 28 (4): 603-14.

Eder, Maciej. 2015a. "Rolling Stylometry." *Digital Scholarship in the Humanities* 30, advance access published April 6, 2015.  
<http://dsh.oxfordjournals.org/content/early/2015/04/06/lle.fqv010>.

Eder, Maciej. 2015b. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 30 (forthcoming).

- Forsyth, Richard, David Holmes, and Emily Tse. 1999. "Cicero, Sigerio, and Burrows: Investigating the Authenticity of the *Consolatio*." *Literary and Linguistic Computing* 14 (3): 375-400.
- Greenwood, H. H. 1995. "Common Word Frequencies and Authorship in Luke's Gospel and Acts." *Literary and Linguistic Computing* 10 (3): 183-87.
- Gurney, Penelope and Lyman Gurney. 1998. "Authorship Attribution of the 'Scriptores Historiae Augustae'." *Literary and Linguistic Computing* 13 (3): 119-31.
- Holmes, David. 1998. "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing* 13 (3): 111-17.
- Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press.
- Jockers, Matthew and Daniela Witten. 2010. "A Comparative Study of Machine Learning Methods for Authorship Attribution." *Literary and Linguistic Computing* 25 (2): 215-23.
- Kestemont, Mike, Sara Moens, and Jeroen Deploige. 2013. "Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux." *Literary and Linguistic Computing* 29, advance access published October 26, 2013.
- Koppel, Moshe, Jonathan Schler and Shlomo Argamon. 2009. "Computational Methods in Authorship Attribution." *Journal of the American Society for Information Science and Technology* 60: 9-26.
- Lancichinetti, Andrea and Santo Fortunato. 2012. "Consensus Clustering in Complex Networks." *Scientific Reports* 2: 336: 1-7.
- Lutosławski, Wincenty. 1897. *The Origin and Growth of Plato's Logic: With an Account of Plato's Style and of the Chronology of his Writings*. London: Longmans.
- McCarty, Willard. 2012. "A Telescope for the Mind?" In *Debates in the Digital Humanities*, edited by Matthew Gold. Minneapolis: University of Minnesota Press.
- Moretti, Franco. 2007. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
- . 2013. *Distant Reading*. London: Verso.

- Mosteller, Frederick, and David Wallace. 2007 [1964]. *Inference and Disputed Authorship: The Federalist*. Reprinted with a new introduction by John Nerbonne. Stanford: CSLI Publications.
- Nerbonne, John. 2007. "The Exact Analysis of Text." Foreword. In Mosteller and Wallace (2007 [1964]), xi–xx.
- Pawłowski, Adam, and Artur Pacewicz. 2004. "Wincenty Lutosławski (1863–1954). Philosophe, helléniste ou fondateur sous-estimé de la stylométrie?" *Historiographia Linguistica* 21: 423–47.
- Ritter, Constantin. 1888. *Untersuchungen über Plato. Die Echtheit und Chronologie der platonischen Schriften, nebst Anhang: Gedankengang und Grundanschauungen von Platos Theätet*. Stuttgart: W. Kohlhammer.
- Rudman, Joseph. 1998a. "The State of Authorship Attribution Studies: Some Problems and Solutions." *Computers and the Humanities* 31: 351–65.
- . 1998b. "Non-traditional Authorship Attribution Studies in the Historia Augusta: Some Caveats." *Literary and Linguistic Computing* 13 (3): 151–57.
- Rybicki, Jan. 2012. "The Great Mystery of the (almost) Invisible Translator." In *Quantitative Methods in Corpus-Based Translation Studies*, edited by Michael Oakes and Meng Ji, 231–48. Amsterdam: John Benjamins.
- Rybicki, Jan and Maciej Eder. 2011. "Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?" *Literary and Linguistic Computing* 26 (3): 315–21.
- Rybicki, Jan, Mike Kestemont, and David Hoover. 2014. "Collaborative Authorship: Conrad, Ford and Rolling Delta." *Literary and Linguistic Computing* 29 (3): 422–31.
- Salmon, John Hearsey McMillan. 1980. "Cicero and Tacitus in Sixteenth-Century France." *The American Historical Review* 85: 307–31.
- Schöch, Christof. 2013. "Fine-Tuning our Stylistic Tools: Investigating Authorship, Genre, and Form in French Classical Theater." In: *Digital Humanities 2013: Conference Abstracts*. University of Nebraska-Lincoln, 383–86.
- Tunberg, Terence O. 1999. "Observations on the Style and Language of Lipsius's Prose: a Look

- at some Selected Texts." In *Iustus Lipsius: Europae lumen et columen*, edited by G. Tournoy, J. de Landsheer, and J. Papy, 169-78. Leuven: Cornell University Press.
- Tweedie, Fiona, David Holmes and Thomas Corns. 1998. "The Provenance of 'De Doctrina Christiana' Attributed to John Milton: A Statistical Investigation." *Literary and Linguistic Computing* 13 (2): 77-87.
- Valla, Lorenzo. 2007 [1517]. *On the Donation of Constantine*. Translated by G. W. Bowersock. Cambridge, MA: Harvard University Press.
- Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA.: Addison-Wesley.

#### Appendix A: Analyzed Works

- Abbo Floriensis, *Passio sancti Edmundi regis et martyris*  
 Abelardus Petrus, *Ad amicum suum consolatoria*  
 —, *Dialogus inter Philosophum, Iudaeum et Christianum*  
 Aelredus Rievallensis, *De amicitia*  
 Alanus de Insulis, *Liber de planctu naturae*  
 Albert of Aix, *Historiae Hierosolymitanae expeditionis liber i*  
 —, *Historiae Hierosolymitanae expeditionis liber ii*  
 Alcuin, *Disputatio de rhetorica et de virtutibus sapientissimi regis Karli et Albini magistri*  
 Ambrosius, *Epistulae variae*  
 Ammianus Marcellinus, *Historiae liber xiv*  
 —, *Historiae liber xv*  
 —, *Historiae liber xvi*  
 Anselmus Cantuariensis, *Proslogion*  
 Apuleius, *Apologia sive pro se de magia liber*  
 —, *De mundo*  
 —, *Metamorphoseon liber i*  
 —, *Metamorphoseon liber ii*  
 —, *Metamorphoseon liber iii*  
 Arnobius, *Adversus nationes liber i*  
 —, *Adversus nationes liber ii*  
 Augustinus, *De civitate Dei liber i*  
 —, *Confessionum liber i*  
 Bacon Francis, *Historia regni Henrici Septimi regis Angliae, capitulum i*  
 —, *Historia regni Henrici Septimi regis Angliae, capitulum viii*  
 —, *Novum Organum, liber i*  
 —, *Novum Organum, liber ii*  
 —, *Sermones fideles sive interiora rerum*

- Bebel Heinrich, *Liber facetiarum*  
 Beda Venerabilis, *Historiae ecclesiastica gentis Anglorum liber i*  
 Benedictus Nursinus, *Regula*  
 Berengarius Scholasticus, *Apologeticus contra sanctum Bernardum*  
 Bigges Walter, *Expeditio Francisci Draki in Indias Occidentales*  
 Buchanan George, *De Maria Scotorum regina*  
 Caesar, *Commentariorum de bello civili liber i*  
 —, *Commentariorum de bello civili liber ii*  
 —, *Commentariorum de bello civili liber iii*  
 —, *Commentariorum de bello Gallico liber i*  
 —, *Commentariorum de bello Gallico liber ii*  
 —, *Commentariorum de bello Gallico liber iv*  
 Campanella Thomas, *Civitas Solis*  
 Cassiodorus, *De anima*  
 —, *Variarum liber i*  
 —, *Variarum liber ii*  
 Celtis Conradus, *Oratio in gymnasio in Ingelstadio*  
 Cicero, *Brutus*  
 —, *Pro Caecina oratio*  
 —, *Pro Cluentio oratio*  
 —, *Cato Maior de senectute*  
 —, *De natura deorum liber i*  
 —, *In M. Antonium oratio Philippica prima*  
 —, *Pro Sestio oratio*  
 —, *Tusculanarum disputationum liber i*  
 Curtius, *Historiae Alexandri Magni liber iii*  
 —, *Historiae Alexandri Magni liber iv*  
 —, *Historiae Alexandri Magni liber v*  
 Dante Aligheri, *Monarchia liber i*  
 —, *Monarchia liber ii*  
 Descartes, *Meditationes 1–3*  
 —, *Meditationes 4–6*  
 Einhard, *Vita Karoli Magni*  
 Erasmus, *Institutio principis Christiani*  
 —, *Declamatio de laude matrimonii*  
 —, *Moriae encomium*  
 —, *Querela pacis*  
 Florus, *Epitomae liber i*  
 —, *Epitomae liber ii*  
 Frontinus, *Strategemata liber i*  
 —, *Strategemata liber ii*  
 —, *Strategemata liber iii*  
 Galileo Galilei, *Sidereus nuncius*  
 Gellius, *Noctes Atticae liber i*  
 —, *Noctes Atticae liber ii*  
 —, *Noctes Atticae liber iii*

- Holberg Ludvig, *Nicolai Klimii iter subterraneum*  
 Iustinus, *Historiarum Philippicarum liber ii*  
 —, *Historiarum Philippicarum liber xii*  
 Kepler Ioannes, *Strena seu de nive sexangula*  
 Livius, *Ab Urbe condita liber i*  
 —, *Ab Urbe condita liber xxi*  
 —, *Ab Urbe condita liber xxxi*  
 —, *Ab Urbe condita liber xli*  
 Matthias of Miechow, *De duabus Sarmatiis tractatus i*  
 —, *De duabus Sarmatiis tractatus ii*  
 Melanchthon Petrus, *Historia de vita et actis Lutheri*  
 Mirandola Giovani Pico, *Oratio de hominis dignitate*  
 Morus Thomas, *De nova insula Utopia*  
 Patricius Franciscus, *Panaugia*  
 Petrarca Franciscus, *Contra medicum quendam*  
 Piccolomini Eneas Silvius, *Epistulae* (part i)  
 —, *Epistulae* (part ii)  
 Plinius Minor, *Epistularum liber i*  
 —, *Epistularum liber ii*  
 —, *Epistularum liber iii*  
 —, *Panegyricus*  
 Plinius Maior, *Naturalis historiae liber ii*  
 —, *Naturalis historiae liber iii*  
 —, *Naturalis historiae liber iv*  
 —, *Naturalis historiae liber v*  
 Poggio Gian Francesco, *Facetiae*  
 Sallustius, *Bellum Catilinae*  
 —, *Bellum Iugurthinum*  
 —, *Epistulae*  
 Seneca, *Ad Helviam matrem de consolatione*  
 —, *Ad Marciam de consolatione*  
 —, *Ad Polybium de consolatione*  
 —, *Ad Novatum de ira liber i*  
 —, *Ad Novatum de ira liber ii*  
 —, *Ad Neronem de clementia*  
 —, *Ad Serenum de constantia*  
 —, *Epistularum ad Lucilium liber i*  
 —, *Epistularum ad Lucilium liber ii*  
 —, *Epistularum ad Lucilium liber xx*  
 —, *Questiones naturales liber i*  
 —, *Questiones naturales liber ii*  
 —, *Questiones naturales liber iii*  
 Suetonius, *Divus Augustus*  
 —, *Divus Claudius*  
 —, *Divus Iulius*  
 —, *Caligula*

- , *Nero*  
 —, *Tiberius*  
 —, *Agricola*  
 Tacitus, *Annalium liber i*  
 —, *Annalium liber ii*  
 —, *Annalium liber iii*  
 —, *Annalium liber iv*  
 —, *Germania*  
 —, *Historiarum liber i*  
 —, *Historiarum liber ii*  
 —, *Historiarum liber iii*  
 Thomas Aquinas, *De ente et essentia*  
 Tünger Augustin, *Facetiae*  
 Valerius Maximus, *Factorum et dictorum memorabilium liber i*  
 —, *Factorum et dictorum memorabilium liber ii*  
 —, *Factorum et dictorum memorabilium liber iii*  
 —, *Factorum et dictorum memorabilium liber iv*  
 Varro, *De agri cultura liber i*  
 —, *De agri cultura liber ii*  
 —, *De agri cultura liber iii*  
 —, *De lingua Latina liber v*  
 —, *De lingua Latina liber vi*  
 —, *De lingua Latina liber vii*  
 Valleius Paterculus, *Historiae Romanae liber i*  
 —, *Historiae Romanae liber ii*  
 Vico Gianbattista, *Oratio vi*  
 Vitruvius, *De architectura liber i*  
 —, *De architectura liber ii*  
 —, *De architectura liber iii*  
 —, *De architectura liber iv*  
 Xylander, *Vita caesaris*

## Appendix B: The Most Frequent Words

1	et
2	in
3	non
4	est
5	ut
6	ad
7	cum
8	quod
9	qui
10	sed
11	quam
12	a

13	quae
14	si
15	de
16	ex
17	aut
18	esse
19	ac
20	se
21	ab
22	sunt
23	enim
24	per
25	nec
26	atque
27	etiam
28	hoc
29	quo
30	quid
31	ne
32	sit
33	uel
34	autem
35	tamen
36	neque
37	ita
38	eius
39	id
40	inter
41	haec
42	quibus
43	me
44	eo
45	quidem
46	nihil
47	uero
48	nam
49	iam
50	erat
51	nisi
52	pro
53	ea
54	sibi
55	quoque
56	quia
57	te
58	tam

59	quem
60	esset
61	qua
62	ubi
63	omnia
64	tum
65	modo
66	his
67	apud
68	ipse
69	sic
70	eum
71	fuit
72	ille
73	mihi
74	sine
75	magis
76	illa
77	res
78	nunc
79	omnes
80	post
81	quos
82	at
83	tantum
84	ante
85	itaque
86	omnibus
87	quis
88	nos
89	e
90	eorum
91	omnium
92	igitur
93	rerum
94	ego
95	cuius
96	tibi
97	hic
98	sua
99	an
100	causa
101	potest
102	quasi
103	inquit
104	minus

- 105 dum
- 106 illi
- 107 suis
- 108 illud
- 109 ergo
- 110 sint
- 111 deinde
- 112 eos
- 113 nobis
- 114 uerum
- 115 ipsa
- 116 rebus
- 117 contra
- 118 suo
- 119 natura
- 120 cui
- 121 primum
- 122 maxime
- 123 sub
- 124 semper
- 125 rem
- 126 alia
- 127 re
- 128 eam
- 129 rei
- 130 unde
- 131 idem
- 132 ei
- 133 sui
- 134 ipsum
- 135 satis
- 136 illis
- 137 ibi
- 138 hanc
- 139 siue
- 140 tu
- 141 ob
- 142 inde
- 143 erant
- 144 quas
- 145 homines
- 146 quorum
- 147 simul
- 148 loco
- 149 hominum
- 150 is

- 151 tempore
- 152 posse
- 153 propter
- 154 fieri
- 155 parte
- 156 die
- 157 aliquid
- 158 hac
- 159 aliud
- 160 quidam
- 161 unum
- 162 suam
- 163 bellum
- 164 adeo
- 165 potius
- 166 suum
- 167 aliis
- 168 quantum
- 169 tunc
- 170 nulla
- 171 eadem
- 172 animi
- 173 uos
- 174 uti
- 175 sicut
- 176 habet
- 177 multa
- 178 ipsi
- 179 caesar
- 180 illum
- 181 hunc
- 182 alii
- 183 iis
- 184 fuisse
- 185 magna
- 186 possit
- 187 omnis
- 188 huius
- 189 una
- 190 solum
- 191 quaedam
- 192 suaे
- 193 quoniam
- 194 usque
- 195 eodem
- 196 omni

- 197 tempus
- 198 animo
- 199 genus
- 200 tot
- 201 bene
- 202 quippe
- 203 hinc
- 204 mox
- 205 fuerit
- 206 saepe
- 207 nomine
- 208 nemo
- 209 ista
- 210 ipso
- 211 habere
- 212 rex
- 213 dies
- 214 licet
- 215 uitae
- 216 primo
- 217 dei
- 218 nomen
- 219 super
- 220 locum
- 221 suos
- 222 illo
- 223 item
- 224 adhuc
- 225 erit
- 226 certe
- 227 haud
- 228 urbem
- 229 bello
- 230 naturae
- 231 uelut
- 232 circa
- 233 exercitus
- 234 partem
- 235 pars
- 236 quin
- 237 deus
- 238 corporis
- 239 aduersus
- 240 castra
- 241 populi
- 242 facile

243	ideo
244	fere
245	cur
246	terra
247	plus
248	opus
249	statim
250	denique

### Appendix C: Words Shared by All the Sample Works

1	et
2	in
3	non
4	est
5	ut
6	ad
7	cum
8	quod
9	qui
10	sed
11	quam
12	a
13	de
14	ex
15	se
16	ab
17	sunt
18	per
19	etiam
20	quo
21	quibus
22	quem
23	qua
24	omnia