

THE ELEMENTARY PARTICLES:
A COMPUTATIONAL STYLOMETRIC INQUIRY INTO THE
MEDIAEVAL GREEK-LATIN ARISTOTLE^{*}

PIETER BEULLENS
FWO VLAANDEREN – KU LEUVEN

WOUTER HAVERALS
PRINCETON UNIVERSITY

BEN NAGY
INSTITUTE OF POLISH LANGUAGE, POLISH ACADEMY OF SCIENCES (IJP PAN)



I. *Introduction*

The research presented in this article results from a pleasant and fruitful collaboration, for which Pieter Beullens provided the research question and the background knowledge about mediaeval Greek-Latin translation of Aristotle and other ancient authors, and Wouter Haverals and Ben Nagy developed, tested, and described the approach from computational stylometry. Our research started from the question of how computational stylistic analysis can help to identify Latin translators of ancient Greek texts and corroborate or contest previous attributions. While computational approaches to style have reached impressive results in the field of authorship recognition, the identification of translators faces the challenge that it must cleanly distinguish between the characteristics of the (Latin) translator and the features of the underlying (Greek) author. To corroborate the method's validity, a first case study focuses on the two preserved

* Some research for this article was carried out as part of Pieter Beullens' postdoctoral fellowship project *Mind Your Words! The Role of Medieval Translations in the History of Concepts*, funded by the Research Foundation – Flanders (12W5722N). Ben Nagy is supported by Polish Academy of Sciences Grant 2020/39/O/HS2/0291.

13th-century Greek-Latin versions of Aristotle's *Rhetoric*.¹ Since the translations were based on the same source text, they form an ideal litmus test to demonstrate the usefulness of the method to distinguish between translators rather than to identify the author. The attribution of one translation to William of Moerbeke is generally accepted, while the suggestion to ascribe the other on philological grounds to Bartholomew of Messina has led to mixed conclusions.

We first illustrate that the traditional stylistic approach to Aristoteles Latinus translations, as initiated and developed by Lorenzo Minio-Paluello and Fernand Bossier, shows an unacknowledged close connection with contemporary practices in authorship attribution studies – the work produced by those precursors was an invaluable basis for the design of our own computational method. After justifying the criteria governing the compilation of our corpus, and the origin and processing of the source texts, we explain the technical background of our stylometric methodology. Finally, the results of our investigation reveal some limitations and difficulties that will need further research and refinement.

II. Rhetorica: *State of the Art* ²

In the later Middle Ages, two Greek-Latin translations of Aristotle's *Rhetoric* circulated. William of Moerbeke's translation of the *Rhetoric* was the prevailing one, evidenced by its preservation in over one hundred manuscripts.³ Yet, the existence of a second, anonymous Greek-Latin version was already known to Amable Jourdain who quotes the opening paragraph from a Sorbonne manuscript.⁴ While the attribution to Moerbeke has never been challenged, the identity of the other translator is still undecided. Pieter Beullens, one of the authors of the present study, has previously argued for an attribution to Bartholomew of Messina. In this section, the existing debate on this topic is outlined.

Bernd Schneider's doctoral dissertation offered a comprehensive study of the anonymous translation, drawing upon the Sorbonne manuscript, a second

¹ We use *Rhetoric* when referring to Aristotle's work in general, but in line with the other cited titles we designate the translations of that text with the Latin form *Rhetorica*.

² This section was adapted from PIETER BEULLENS, « A Methodological Approach to Anonymously Transmitted Medieval Translations of Philosophical and Scientific Texts. The Case of Bartholomew of Messina », Ph.D. Diss., KU Leuven 2020.

³ RITA COPELAND, *Emotion and the History of Rhetoric in the Middle Ages*, Oxford University Press, Oxford 2021, p. 185, n. 90.

⁴ AMABLE JOURDAIN, *Recherches critiques sur l'âge et l'origine des traductions latines d'Aristote et sur des commentaires grecs ou arabes employés par les docteurs scolastiques*, Nouvelle édition, ed. CHARLES JOURDAIN, Fantin, Paris 1843, p. 444. Now MS Paris, Bibliothèque nationale de France, lat. 16673.

complete version from Toledo,⁵ and two fragmentary copies. The study eventually resulted in the critical edition for the *Aristoteles Latinus* series in 1978.⁶ Schneider's detailed analysis of the text led him to conclude that its translator was not very competent in the Greek language (« *interpres anonymus graeci sermonis non apprime sciens fuisse... videtur* »).⁷ The same can be said about his knowledge of Greek cultural heritage: the translator often misses proper names, which produce mix-ups and adventurous translations (« *Verwechslungen und abenteuerliche Übersetzungen* ») or are left out leaving open spaces in the text. Quotations from Greek poetry met the same fate.⁸ Still, our man is not entirely uncultivated. He demonstrates decent familiarity with prosody and scansion when he replaces Greek examples for metrical schemes with Latin illustrations where the mere translation of the words would have led to confusion and utter meaninglessness. With a similar feeling for cultural transposition, the opening verses from Homer's epic poems are rendered with lines from the *Ilias Latina* and Horace's *Ars poetica*.⁹

The suggestion to identify the anonymous translator of the *Rhetoric* as Bartholomew of Messina was first launched by Leonard Spengel in 1867, although he readily admitted that he had never seen the text for himself (« *mihi numquam visa* »).¹⁰ Spengel's attribution was briskly and without even an attempt at substantiation dismissed by Martin Grabmann, stating that « there are no positive leads for this » (« *Indessen fehlen hierfür die positiven Anhaltspunkte* »).¹¹

Schneider echoes Grabmann's view that Spengel's assumption cannot be justified by anything (« *daß sich Spengels Annahme durch nichts rechtfertigen läßt* »).¹² Unfortunately, Schneider's statement is poorly supported by limited evidence, consigned to a mere footnote (admittedly of Germanic size, taking up more than half a page). It can be split into three different focusses: (1) function words, (2) content words, and (3) syntax.

⁵ MS Toledo, *Biblioteca capitular*, 47.15. The final chapters of book III (from 1368a8 onward) are missing in both the Toledo and the Sorbonne manuscripts although the latter has an *explicit* stating: *Completus est*. Therefore, it seems logical to assume that the shared model, or even the autograph, was incomplete.

⁶ BERND SCHNEIDER, *Die mittelalterlichen griechisch-lateinischen Übersetzungen der aristotelischen Rhetorik*, de Gruyter, Berlin – New York 1971, and ID., *Rhetorica. Translatio anonyma sive Vetus et Translatio Guillelmi de Moerbeka*, Brill, Leiden 1978 (*Aristoteles Latinus*, 31/1–2).

⁷ SCHNEIDER, *Rhetorica*, p. xvi.

⁸ SCHNEIDER, *Die mittelalterlichen griechisch-lateinischen Übersetzungen*, p. 27–28.

⁹ SCHNEIDER, *Die mittelalterlichen griechisch-lateinischen Übersetzungen*, p. 28.

¹⁰ LEONARDUS SPENGLER, *Aristotelis Ars Rhetorica*, vol. I, Teubner, Lipsiae 1867, p. 117 and n. 2.

¹¹ MARTIN GRABMANN, *Forschungen über die lateinischen Aristoteles-Übersetzungen des XIII. Jahrhunderts*, Aschendorff, Münster 1916, p. 243.

¹² SCHNEIDER, *Die mittelalterlichen griechisch-lateinischen Übersetzungen*, p. 4 and n. 14.

(1) As far as function words are concerned, Schneider refers to three articles by Minio-Paluello and to Lorimer and Minio-Paluello's critical edition of *De mundo*.¹³ Schneider concludes that the comparison does not warrant confirmation of Spengel's intuitive attribution of the *Rhetorica anonyma* to Bartholomew.

(2) For the study of lexical choices, Schneider compares a very eclectic selection of seven terms from the *Rhetorica*, viz. ἀκμάζω, ἀκμή, ἀρμόττω, ἄσπίς, δικαστήριον, κεφαλαιωδῶς, and σεμνός, with their equivalents in *De mundo*, the only translation by Bartholomew for which he had a critical edition available at the time. Schneider does not provide his rationale for deeming these specific words as significant and conclusive evidence. Given that these words are translated differently in the two works, Schneider seems to imply that their translators cannot be identical.

There is little need to emphasise the restricted value of this very small sample and the lack of clarification as to why precisely these words are expected to represent Bartholomew's characteristic vocabulary. Beyond the small number of items and their seemingly arbitrary selection, an additional methodological issue is Schneider's choice of comparative material. It consists of only one text, *De mundo*, and to make matters worse, four of the selected terms appear just once in this text. Moreover, as opposed to the other translations by Bartholomew that we use in this article, *De mundo* is not attributed to him in any manuscript. Its acceptance as a translation by Bartholomew of Messina solely relies on the stylistic analysis by Minio-Paluello.

It is safe to assume from these observations that Schneider takes an overambitious leap by inferring from single occurrences in one particular work that this terminology reliably reflects Bartholomew's preferences, while at the same time overlooking potential alternative explanations. Implicitly, Schneider's approach relies on the shaky assumption that translators chose a specific Latin term for each Greek word at the start of their careers and consistently stuck to it, thereby disregarding the possibility that they could gain fresh or deeper insights, or even forget whatever bit of knowledge they earlier had.

It should therefore come as no surprise that Schneider's evidence does not stand the test of a more thorough comparison. Expanding the analysis to encompass more of Bartholomew's translations yields greater diversity in typical renderings and evidence of an evolving understanding. Let us return to what

¹³ LORENZO MINIO-PALUELLO « Guglielmo di Moerbeke, traduttore della 'Poetica' di Aristotele (1278) », *Rivista di Filosofia Neo-Scolastica*, 39 (1947), p. 1–17; ID., « Henri Aristippe, Guillaume de Moerbeke et les traductions latines médiévales des 'Météorologiques' et de 'De generatione et corruptione' d'Aristote », *Revue Philosophique de Louvain*, 45 (1947), p. 206–235; ID., « Boezio, Giacomo Veneto, Guglielmo di Moerbeke, Jacques Lefèvre d'Étaples e gli *Elenchi Sofistici* », *Rivista di filosofia neo-scolastica*, 44 (1952), p. 398–411; WILHELMUS L. LORIMER, LAURENTIUS MINIO-PALUELLO, *De mundo. Translationes Bartholomaei et Nicholai. Editio altera*, Desclée De Brouwer, Bruges – Paris 1965 (Aristoteles Latinus, 11/1–2); the three articles were reprinted in LORENZO MINIO-PALUELLO, *Opuscula. The Latin Aristotle*, Hakkert, Amsterdam 1972, p. 40–56; 57–86; 164–177.

Schneider labels as typical translations from the anonymous *Rhetorica*. He probably considered the translations for ἀκμή ('augmastica'), ἀκμάζω ('augmasticum fio'), and κεφαλαιωδῶς ('in specie capitis') clumsy in comparison with 'augmentum', 'augeor'/'cresco', and 'capitulatim' that he read in *De mundo*. However, interesting parallels can be found in Bartholomew's other translations to which Schneider had no access.

Schneider criticised the rendering of κεφαλαιωδῶς as *in specie capitis*, for which 'capitulatim' in *De mundo* (397b9) looks more appropriate. Yet in Hippocrates' *De natura pueri* (12.5 168.12), Bartholomew employs the same building scheme to latinize an adjective ending in -ειδής ('in specie panniculi' – ὑμενοειδής). Bartholomew's translation of δικαστήριον as 'curia' is consistent with its usage in the *Problemata* (950b5) and in the *Magna moralia* (1202a24), while 'iudicium' as found in *De mundo* (400b17, comparable with 'iudiciale' in the *Rhetorica*, 1416a32)¹⁴ also occurs in the *Problemata* (951a15), alongside 'pretorium' (952b36). A similar terminological wavering can be observed for ἀσπίς. Schneider contrasts Bartholomew's 'clipeus' from *De mundo* (399b3;35) with 'telum' in the anonymous *Rhetorica*, while he overlooks that his own critical edition of the latter records instances for 'clipeus', 'telum', 'scutum', and 'lancea' as renderings of ἀσπίς. In one passage (1413a5) the manuscripts document 'telum scutum' as a probable double reading from the translator's autograph.

There lies little weight in the distinction that Schneider makes between 'congruo' (allegedly Bartholomew's favourite) and 'convenio' (*Rhetorica*) as the Latin equivalents for ἀρμόττω. While 'congruo' is confirmed as Bartholomew's choice in the *Magna moralia* (1188a25; 1191a11), 'convenio' is found in *De mirabilibus auscultationibus* (837a33) and the *Problemata* (905b14; 922b24; 927a17; 933a21).¹⁵ Moreover, mirroring the oversight with the previous term, Schneider neglects to mention that the translation 'convenio' for ἀρμόττω is equally attested in the *Rhetorica*. His incomplete, or even incorrect reports of the translations for ἀσπίς and ἀρμόττω all but nullify the value of these particular content words as evidence.

As for the final example σεμνός, Schneider considers its translations 'humilis' and 'parvus' as indicative of the translator's poor knowledge of Greek, and contrasts them with Bartholomew's rendering as 'venerabilis' in *De mundo* (408b6).¹⁶ He could have added that the same Latin word is found in Bartholomew's version of Theophrastus' *De principiis*, ch. 1, for which Kley's edition was available.¹⁷

¹⁴ Mistakenly 1416b32 in Schneider's index.

¹⁵ Alongside single occurrences of 'pervenio' (928b14) and 'concordo' (919b8).

¹⁶ SCHNEIDER, *Die mittelalterlichen griechisch-lateinischen Übersetzungen*, p. 19, n. 21.

¹⁷ WALTER KLEY, *Theophrasts Metaphysisches Bruchstück und die Schrift περὶ σημείων in der lateinischen Übersetzung des Bartholomaeus von Messina*, Triltsch, Würzburg 1936.

In any case, Schneider argues that Bartholomew's understanding of the word was too sophisticated to reconcile with such an imprecise translation in the *Rhetorica*.

Nevertheless, Bartholomew's grasp of the term σεμνός is more complex than initially evident. In the *Magna moralia*, the adjective and its corresponding noun are transliterated as 'semnos' (1192b36) and 'semnotes' (1192b30), although the verb σεμνύω consistently becomes 'venero' (1195b19;22²;23). This fluctuation in translation choices may be indicative of a growing insight into the failure of previous attempts to adequately cover the original's semantic range.

(3) In his final argument, Schneider relies on syntax, referencing Dittmeyer's observation that Bartholomew was well aware of the exact meaning of the superlative preceded by ὅτι or ὡς, which he translated correctly in the *Problemata* (865b22: ὡς τάχιστα – 'quam velocissime').¹⁸ However, the identical expression is less expertly latinized as 'velocissime' (884a30) and 'cito' (966a17) elsewhere in the same work, while ὡς μάλιστα becomes 'ut maxime' (866a15). In the Hippocratic *De humana natura* 168.3, ὅτι ἐγγύτατα is latinized as 'adhuc proxima'.¹⁹ While Dittmeyer's claim of Bartholomew's syntactical proficiency may hold, Schneider's conclusion that the *Rhetoric's* inconsistent translations definitively argue against Bartholomew's translatorship is an overreach.

Overall, Schneider's footnote-long dismissal of Spengel's intuitive hypothesis relies on an eclectic range of function words, complemented by seven terms that he claims are uniquely translated, yet his reporting of these is notably incomplete. Moreover, these specific terms are contrasted with a single text, *De mundo*, attributed to Bartholomew solely on stylistic grounds. Finally, Schneider's analysis hinges on a single case of a correctly rendered syntactical concept, yet overlooks other dubious or blatantly incorrect interpretations of the same phenomenon. Schneider's inaccurately, or at least incompletely presented arguments are by no means sufficient to warrant the firm conclusions that the German scholar draws concerning Bartholomew's (lack of) involvement with the *Rhetorica*; neither is Pieter De Leemans's confirmation that Schneider has « convincingly refuted » Spengel's thesis.²⁰ Despite De Leemans's support for Schneider's view, Lisa Devriese, who edited Bartholomew's translation of the *Physiognomonica*, seems favourable to the contrary hypothesis without substantiating her claim: « It must

¹⁸ LEONHARD DITTMAYER, « Hat Bartholomaeus von Messina die *Rhetorica ad Alexandrum* übersetzt? », *Philologische Wochenschrift*, 58 (1938), p. 252–256 and 285–288, in particular p. 287.

¹⁹ Possibly as the result of a misreading as ἔτι ἐγγύτατα.

²⁰ PIETER DE LEEMANS, *Translating at the Court. Bartholomew of Messina and Cultural Life at the Court of Manfred, King of Sicily*, Leuven University Press, Leuven 2014, p. xxi, n. 23, but the author changed his opinion in the 2016 version of the *Aristoteles Latinus Database*, where he writes about the *translatio vetus* of the *Rhetoric*: « it seems, however, that its author belonged to the circle of Bartholomew of Messina ».

be noted that the anonymous translation bears many similarities to Bartholomew's idiom ». ²¹

If Schneider was convinced that the use of identical or similar content words as equivalents for Greek words is significant to attribute a Latin version to a known translator, he could have provided additional evidence of Bartholomew's fingerprints in the anonymous *Rhetorica*.

The adverb ἀπλῶς is predominantly and expectedly translated as 'simpliciter' in the *Rhetorica*, yet in four instances, 'universaliter' is preferred (1357a12;35; 1375b19; 1402a23), and in one other case both Latin equivalents appear together (1393a16). This particular equivalence of ἀπλῶς – 'universaliter' is also found in Bartholomew's translations of the *Magna moralia* (1193b5; 1199b32), and of the *Problemata* (882b6).

There is a notable tendency to translate ψηφίζομαι as 'numero', 'to count', which finds confirmations in the *Problemata* (950b6; and 951b1-4 where the verbs for 'to condemn' and 'to acquit' – καταψηφίζομαι and ἀποψηφίζομαι – are rendered as 'connumero' and 'enumero')²² and in the *Magna moralia* (ἀποψηφίζω – 'supernumero' 1206b21; σύμψηφος – 'connumeratus' 1203b27, 'connumerator' 1206b25), is echoed in the *Rhetorica* ('numero' 1384b35) and in 'numerus' as the rendering for ψήφισμα (1411a10).

Also in the *Rhetorica*, the translation that divides the compound Greek ἀκρόπολις into its etymological components as 'extremum urbis' (1400a33) is mirrored in *De mundo*, where 'extremum urbis' (399b33) occurs alongside the variant reading 'castrum' (the latter also in *De mirabilibus auscultationibus* 846a18). By contrast, in Moerbeke's version of the *Rhetorica*, the same word is rendered as 'extremum civitatis'. The peculiar translation for δρόμος as 'via', which is documented in the *Problemata* (945a24), gets two further confirmations in the anonymous *Rhetorica* (1361b9; 1406a23).

Even if Bartholomew of Messina were not to be identified as the translator of the anonymous *Rhetorica*, a few lexical choices strongly imply a Southern Italian or Sicilian origin of the translation. Consider the translation of σπένδομαι, meaning 'to make peace' (1411a10), rendered as 'facio treug(u)am' (the two main manuscripts of the text have different spellings). Although the noun apparently derives from a Germanic etymology and has left its traces in several languages (e.g. 'trève' in French, 'truce' in English), this particular form points to Italian or Spanish influence. A stronger indication is found in the use of the term 'cabellotus' for τελώνης (1397a25; cf. 'cabella' for τέλος in the same context, 1397a26), which refers to a tax collector. The word is documented in the provisions made by

²¹ LISA DEVRIESE, *Physiognomonica. Translatio Bartholomaei de Messana*, Brepols, Turnhout 2019 (Aristoteles Latinus, 19), p. LIII, n. 46.

²² Possibly an illustration of how the lack of knowledge about the cultural setting of the Athenian courts of law puzzled the translator and drove his attempt at equivalence to incomprehensibility.

emperor Frederick II of Sicily concerning this subject and it is used in similar legislation for Charles of Anjou as king of Sicily in 1277.²³ Although the term later became more widely spread, it was well enough established in 13th-century Sicily and South Italy to make it spring to mind when a local translator looked for a suitable Latin equivalent. In contrast, William of Moerbeke, lacking this geographical context, initially chose what looks like a transliteration ‘telonearius’, which he replaced by the periphrastic ‘qui tributa colligit’ when he revised the first version of his *Rhetorica* translation (see Schneider’s apparatus ad loc.).

From this summary, it should be clear that there are many evidentiary considerations that must be balanced and assessed. For the present, we set aside the historical context, and the style of translation in general, and focus on the new evidence arising from our stylometric study.

III. Stylometric Analysis

III.1. Historical Approaches to Stylometry

In the realm of literary studies, the quest for uncovering authorship and translator identity has long been guided by the nuanced observation of stylistic elements. While a comprehensive history of the stylistic method’s evolution falls beyond the scope of this article, it is pertinent to explore the intersections between traditional stylistic analysis and its modern counterpart which we will pursue in the following section: a stylometric computational approach. Stylometry represents a significant evolution in computer-aided textual analysis, offering a systematic approach to the analysis of textual styles. This evolution mirrors a broader shift in the humanities, where digital methods have started to augment traditional scholarship.

Tracing back to the 1940s, the scholarly endeavours of Lorenzo Minio-Paluello, and later Fernand Bossier in the 1990s, shed light on the intricacies of translating specific Greek words into Latin, particularly noting their frequencies as used by various translators.²⁴ Although rudimentary by today’s standards, their approach was groundbreaking in classical studies, demonstrating that straightforward counts of frequently used words could yield significant insights into

²³ *Fridericus II. Constitutiones pro Sicilia*, ed. WOLFGANG STÜRNER, cited in *Brepolis Cross Database Searchtool*, and the electronic version of Du Cange’s *Glossarium* (<http://ducange.enc.sorbonne.fr/cabellotus> – access 20 April 2020).

²⁴ See esp. MINIO-PALUELLO, « Guglielmo di Moerbeke », p. 1–17; ID., « Henri Aristippe »; FERNAND BOSSIER, « L’élaboration du vocabulaire philosophique chez Burgundio de Pise », in JACQUELINE HAMESSE (ed.), *Aux origines du lexique philosophique européen. L’influence de la Latinitas*, Fédération internationale des Instituts d’études médiévales, Louvain-la-Neuve 1997, p. 81–116.

translatorship. The pioneering methodology employed by Minio-Paluello bears a striking resemblance to modern-day stylometry, even though he did not explicitly refer to his method as stylometric.

Stylometry, as we understand it today, can be defined as the quantitative study of writing style through the analysis of linguistic patterns, structures, and word distributions within texts. This method has become an indispensable tool in various domains, from determining the authorship of disputed documents to unravelling the subtleties of translatorship and identifying distinctive features of textual genres. The concept of a 'stylistic DNA' or 'stylome', a term coined by Van Halteren et al., captures the essence of stylometry effectively.²⁵ Much like biological DNA that uniquely identifies individuals, a stylome encompasses the unique linguistic fingerprint of an author or translator.

The significance of stylometry was profoundly demonstrated by the landmark research of Mosteller and Wallace in 1964.²⁶ Their seminal work on the *Federalist Papers*, a collection of 85 political essays to promote the United States Constitution, represents an important moment in non-traditional authorship studies. By employing Bayesian statistical methods to analyse these 18th-century essays, which were published under the pseudonym 'Publius', Mosteller and Wallace sought to resolve the long-standing ambiguity surrounding the authorship of 12 of these essays. Focusing primarily on the frequency of function words – highly frequent words like articles, prepositions, and conjunctions – their analysis pointed convincingly towards James Madison as the author of the disputed essays. The study of translation style is a less crowded domain than the study of authorial style. For context, readers might refer to Jan Rybicki's 2012 article « The great mystery of the (almost) invisible translator » with references.²⁷ Rybicki's findings have important methodological implications for this study, but at this point we will simply note that it led us to examine (only) closed-class function words, as opposed to 'most frequent words' which are more commonly used in modern authorship attribution research.

The utility of function words in stylometric analysis is anchored in several key characteristics:²⁸ (1) High frequency: function words appear frequently, providing

²⁵ HANS VAN HALTEREN, HARALD BAAYEN, FIONA TWEEDIE, MARCO HAVERKORT, ANNEKE NEIJT, « New Machine Learning Methods Demonstrate the Existence of a Human Stylome », *Journal of Quantitative Linguistics*, 12/1 (2005), p. 65–77.

²⁶ FREDERICK MOSTELLER, DAVID WALLACE, *Inference and Disputed Authorship: The Federalist Papers*, Springer, New York 1964.

²⁷ JAN RYBICKI, « The Great Mystery of the (Almost) Invisible Translator. Stylometry in Translation », in MICHAEL P. OAKES, MENG JI (eds.), *Quantitative Methods in Corpus-Based Translation Studies. A Practical Guide to Descriptive Translation Research*, John Benjamins, Amsterdam – Philadelphia 1996, p. 231–248.

²⁸ JOSÉ NILO G. BINONGO, « Who Wrote the 15th Book of Oz? An Application of Multivariate Analysis to Authorship Attribution », *Chance*, 16 (2003), p. 9–17; MIKE KESTEMONT, « Function Words in

a rich dataset for analysis. (2) Closed set: the set of function words is relatively fixed and limited, allowing for a focused and comprehensive study. (3) Content-independence: their usage is largely independent of the text's subject matter or genre, offering insights that transcend specific themes. (4) Unconscious use: authors and translators use function words more or less unconsciously during the writing process, thus reflecting intrinsic patterns of language usage. The methodology employed by Minio-Paluello and Bossier aligns closely with these principles.

A prime example of Minio-Paluello's stylometric method *avant la lettre* is evident in his analysis of the highly frequent word 'autem' in the Latin translation of Aristotle's *Poetica*, found in MS Toledo, Biblioteca capitular, 47.10, and Eton College Library 129. He noted that William of Moerbeke uniquely used 'autem' to translate the Greek δέ. In contrast, Moerbeke's contemporaries varied between 'vero' or omitted the conjunction altogether.²⁹ Such detailed observations act as signposts in attributing translations, showcasing Minio-Paluello's intuitive sense for these linguistic frequencies. However, a notable distinction remains between his approach and modern computational stylometry, particularly in the methodological framework and the interpretative process. Minio-Paluello and Bossier's analyses, pioneering as they were, depend significantly on their *Gelehrtenintuition*, involving subjective decisions about which linguistic features to include in their analysis. In contrast, computational stylometry offers a more systematic approach, reducing the reliance on individual preferences. Despite this, Minio-Paluello's work profoundly embodies the spirit of the stylometric method, prefiguring the data-driven analysis of modern computational approaches, which is precisely the focus of the following section.

III.2. Context and Aims

In this section, we examine two translations of Aristotle's *Rhetoric* with respect to the style of the translators, based on their use of Latin function words. The texts are examined in the context of a larger corpus of mediaeval Latin translations, both to validate the methods as well as to better understand and contextualise the results as they apply to the uncertainly-attributed *Rhetorica*. The key objective of the stylometric investigation is to interrogate the attribution of the uncertain *Rhetorica* to Bartholomew of Messina. For the purposes of this section, it is putatively assigned to Bartholomew in the sigla, although as we will see, that attribution is not supported by the methods we applied.

Authorship Attribution. From Black Magic to Theory? », in *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, Association for Computational Linguistics, Gothenburg 2014, p. 59–66.

²⁹ MINIO-PALUELLO, « Guglielmo di Moerbeke », p. 8.

III.3. Corpus

As we delineate the corpus for our investigation, it is worthwhile to reflect on the early wisdom of Minio-Paluello, who astutely recognized the importance of a well-chosen comparative base in stylistic analysis:

The linguistic-stylistic proof of authorship for an anonymous translation demands a comparison with other undeniably attributed translations by the presumed author. This comparison should reveal a sufficient degree of similarity in consistent and frequently occurring characteristics between the anonymous version and those that are verified. Furthermore, it necessitates a comparison with the largest possible number of other translations to ascertain that these characteristics are specific to the author in question, and not common to other authors.³⁰

The reference corpus for our investigation consists of works by six translators to whom some Latin versions are firmly attributed. The attributions rely on explicit mentions in titles or colophons in manuscripts that transmit the translations, on the contexts and manuscripts in which the translations were transmitted, or on scholarly consensus based on traditional stylistic analysis. From the 12th century, we included James of Venice, Burgundio of Pisa, Henry Aristippus, and the anonymous translator of the *Physica Vaticana* and the *Metaphysica media*. The 13th century is represented by William of Moerbeke and Bartholomew of Messina. All translators in our investigation translated at least one (pseudo-)Aristotelian treatise, yet in order to widen the sample of translated texts as much as possible and to avoid possible interference caused by the use of similar source texts, treatises by other authors from philosophy, theology, and medicine were also selected. The corpus also includes two unattributed translations of treatises of the ancient Pyrrhonist Sextus Empiricus that were previously examined by the authors. The results of this investigation will be published in a separate article. The complete corpus is detailed in Appendix B.

It is appropriate to also mention what we deliberately did not include in the corpus. We did not consider any translations of logic, which were likely heavily tinged by the strong impact of Boethius' translations and commentaries on the Latin culture and vocabulary of education. Also omitted are Latin versions that

³⁰ MINIO-PALUELLO, « Guglielmo di Moerbeke », p. 3: « La prova linguistico-stilistica della paternità d'una traduzione anonima richiede un confronto con altre traduzioni indubbie del supposto autore, confronto che mostri un grado sufficiente di identità in caratteristiche costanti e frequenti tra la versione anonima e quelle sicure; ed oltre a questo, richiede un secondo confronto con il massimo numero possibile di altre traduzioni, da cui risulti che tali caratteristiche sono particolari all'autore in questione, e non ad altri autori ».

originated from the revision of previously existing translations since it would be difficult to adequately distinguish the input of the revising translator from the vocabulary of his older colleague. For the same reason, translations by Robert Grosseteste were ignored. There is insufficient evidence about his use of work done by predecessors, and even for some of the translations that were certainly made from scratch, it is debated how the workload was divided between him and his assistants.

III.4. General Methodology

Although this is not the right venue for a detailed discussion of the technical methods, it is important to note that the analysis process was designed to be as open and reproducible as possible. Due to copyright restrictions, we are unable to provide open-access copies of the raw corpus text. However, the processed data is available immediately after pre-processing and feature extraction (function word frequencies). Our analysis was performed using the Python programming language, scikit-learn, and subsidiary packages like UMAP.³¹ Data visualisations were created with the R language and ggplot2.³² Some analysis code was implemented from scratch as part of this project, and this is also made available under open licences. We have created a dedicated repository for the code and replication material for this article, which is freely accessible, and allows all figures and results to be precisely reproduced, or for the results to be further investigated.³³

The transformation of the corpus was relatively simple. After basic data-cleaning and character-set normalisation, every text was broken into sequential chunks of 1000 words, from which we extracted the function word frequencies. The chosen function words are presented in Appendix A. These words were selected based on domain expertise;³⁴ in some cases, they represent function

³¹ Most of the basic statistical analysis, scaling, feature extraction was performed using Scikit-learn: FABIAN PEDREGOSA et al., « Machine Learning in Python », *Journal of Machine Learning Research*, 12 (2011), p. 2825–2830. Unsupervised clustering is based on UMAP: LELAND MCINNES, JOHN HEALY, NATHANIEL SAUL, LUKAS GROßBERGER, « UMAP: Uniform Manifold Approximation and Projection », *Journal of Open Source Software*, 3/29 (2018), p. 861, <https://doi.org/10.21105/joss.00861> (accessed 27 March 2024).

³² HADLEY WICKHAM, *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York 2016.

³³ https://github.com/PieterBeullens/medtrans_stylo (accessed 24 March 2024).

³⁴ This is the point where we ought to clarify what might seem like a methodological anomaly. Modern authorship stylometry commonly employs analysis based on the frequencies of some number of most frequent words (anywhere from 100 to 5000), using distance measures based on Burrow's Delta (with assorted improvements). As Rybicki notes in his 2012 article, these approaches appear to submerge the style of the translator, and detect mainly authorial signal. By restricting the domain of analysis even more to expertly selected function words, much as Minio-Paluello and others have done with traditional methods, we hoped to suppress the

words where the mediaeval translators are known to differ, in others they are words that offer a choice to the translator in terms of the way Greek idiom is rendered in Latin. It is by these translator choices that the translator's individual stylistic fingerprints are revealed – it should be evident that when there is only one possible way to translate an expression then there is no room for individual style. In addition to the words listed in Appendix A, mediaeval orthographic variations were normalised, consolidating case-insensitive frequencies under the same headword – so ‘multotiens’, for example, encompasses the combined frequencies for ‘multotiens’, ‘multoties’, and the uppercase V/U spellings ‘MVLTOTIENS’, and ‘MVLTOTIES’.

All Works (plus two Rhetoricae)

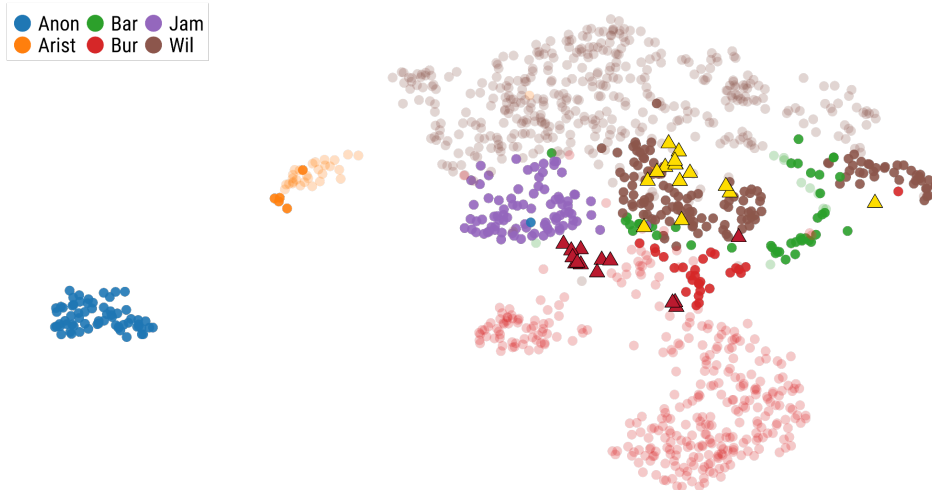


Fig. 1. UMAP cluster of the corpus. Colours indicate the translator (ground truth), with strong colours being the Aristotelian works. The two *Rhetorica* translations are emphasised as triangles, yellow for Moerbeke's *Rhetorica* (Wil-Rhet) and red for the anonymous Latin translation (Bar?-Rhet). For the sigla used here and in the following figures, see Appendix B.

III.5. Unsupervised Cluster Analysis

To begin, we performed a qualitative investigation of the stylometric properties of the various translations using unsupervised clustering methods. ‘Unsupervised’ in this sense is contrasted with ‘trained’ machine-learning methods in which an

authorial signal even more, to try and discern the very tenuous signal from the translator – and based on the results it seems that the method was effective.

algorithm learns the propensities of each translator from a training corpus and then provides an opinion on some unseen samples. Unsupervised methods simply arrange the data by self-similarity; thus, data points lying closer to each other are more similar. The class categories (the colours in the figures) are added post-analysis, based on empirical truth (which is never known to the algorithm). In preparing the corpus for this kind of clustering, some standard mathematical transformations are applied (z-scaling, normalisation), before passing the data to the algorithm. Recall that works are broken into 1000-word chunks to avoid distorting the statistics, and so one work of translation can appear in the figures as many individual points, depending on its size.

From the analysis of the overall corpus (Fig. 1), two things are clear: first, that translator style is certainly discernible by their use of function words; and second, that Aristotelian works (shown in stronger colours) generally cluster together, indicating that there is also an (unwanted, in this case) authorial signal. We can see from Fig. 1, however, that some translators have such an idiomatic style that their work does not group with the rest of the Aristotelian cluster – this is true particularly of Aristippus and the anonymous translator of the *Metaphysica Anonyma* and the *Physica Vaticana*.

With respect to the question at hand, we can see that the *Rhetorica* attributed to William of Moerbeke does indeed group with William's translations of Aristotle; this relationship can be seen in more detail in Fig. 2, in which we restricted the clustering to the various works of William. The stylistic pattern of the other *Rhetorica*, however, appears more isolated. If the case were to be made that the translation is by Bartholomew, it is not, *prima facie*, a strong one according to the cluster output, although two things must be noted: first, that we do not have a very large comparison corpus of Bartholomew's translations (and even of the ones we have, in the case of *De mundo* the attribution is not completely secure); and second, that Bartholomew's style appears to be considerably more varied compared to that of the other translators. Looking more closely at Bartholomew's cluster (Fig. 3), we can see distinct subclusters forming for e.g., the *Magna moralia* and the sample from the *Problemata*. This makes the distance of the *Rhetorica* from his other translations less compelling as negative evidence. Once again, we emphasise that unsupervised cluster visualisations are a qualitative technique; they give an excellent impression of the relationships in the data, but they cannot be interpreted in terms of concrete likelihoods. To look more closely at the claim that the second *Rhetorica* was translated by Bartholomew of Messina, we must turn to a different kind of statistical analysis.

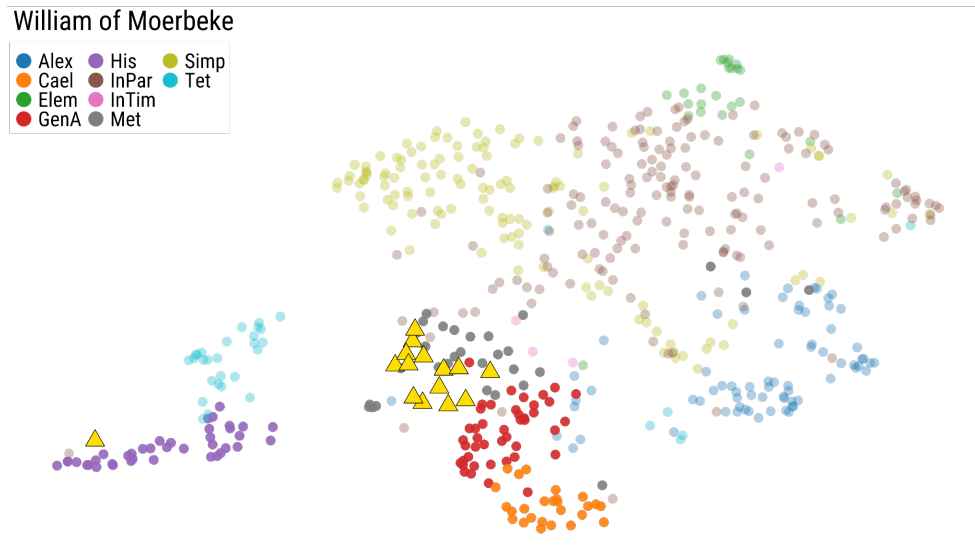


Fig. 2. UMAP cluster of William of Moerbeke's translations. Colours indicate works (see sigla in Appendix B), with strong colours being the Aristotelian works. Wil-Rhet is emphasised as yellow triangles.

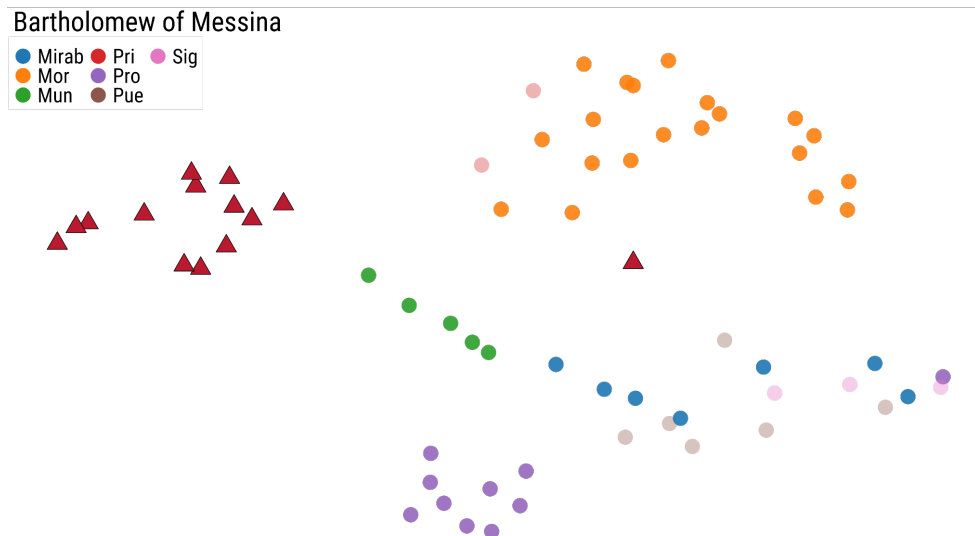


Fig. 3. UMAP cluster of Bartholomew of Messina's translations. Colours indicate works (see sigla in Appendix B), with strong colours being the Aristotelian works. Bar?-Rhet is emphasised as red triangles.

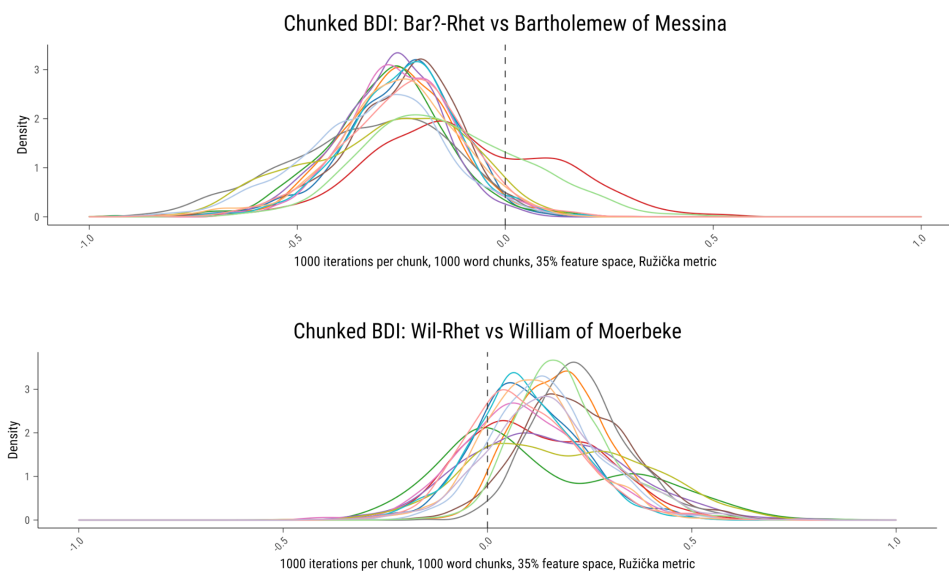


Fig. 4. Bootstrap Distance Imposters results for the two *Rhetorica* translations. Each line is the distribution of differences-in-distance for one chunk. Positive differences indicate a result closer to the candidate, negative distances a result closer to one of the imposters.

III.6. Bootstrap Distance Imposters

The statistical practice of bootstrapping is an important tool in modern descriptive (as opposed to inferential) statistics – in other words statistics that aims to quantify the typicality of an observation, as opposed to identifying or describing causes and effects. In this domain, bootstrap statistics are contrasted with distributional statistics, which apply specific mathematical models to observational data based on various assumptions (for example assuming the data is ‘normally distributed’, ‘homoscedastic’, ‘independent’, and so forth). Bootstrap statistics belong to the ‘parameter-free’ family of techniques that make no assumptions about the data, they instead use repeated random sampling to collect observations and measure typicality by empirical means. To (over)simplify, these techniques trade some precision for robustness. The style of bootstrapping applied here is a ‘regularising’ method that considers many observations where each observation is made on a random subset of the available features – we are measuring the frequency of 54 different function words, but each iteration considers only 20 of them. The effect is to reduce the influence of function words that might bias the outcome by being accidentally over-represented in the sample; this avoids ‘overfitting’, but again reduces accuracy. Bootstrap methods provide quantitative statistical results which tend to be conservative (so strong results are

usually reliable).

For this part of the analysis, we adapted and expanded work by Kestemont et al. on a method that is gaining acceptance as a ‘standard technique’ in computational stylometry, called ‘General Imposters’.³⁵ General Imposters is a quantitative method that works by comparing a question text to a set of candidate texts (texts by the proposed author) and a set of imposter texts (texts with similar characteristics, but not by the proposed author). Through bootstrapping, this method assesses whether the question text is statistically more similar to the candidate author than to the imposters. If the question text is, in fact, by a completely different author (neither the proposed candidate nor any of the imposters) then there would be no significant statistical difference observed in the analysis. The modifications to the method allow us to visualise the distribution of these differences, instead of recording a single summary statistic. These visualisations retain (and indeed highlight) the statistical uncertainty of the determinations, which can be concealed by other methods.

In the first place, we considered the issue at hand, comparing the attribution of William’s *Rhetorica* to the one with debated attribution to Bartholomew. The results can be seen in Fig. 4. The way the figure works is this: each *Rhetorica* is made up of 14–15 1000-word chunks. For each of those chunks, the bootstrap test is performed 1000 times, and the distribution of those results is shown. In each case, recall precisely what is being measured – the stylistic distance (i.e., difference) between the best matching chunks by the candidate author, and the best matching chunks by any ‘imposter’ author. If the distance is positive this means that the candidate is a better match; if negative, an imposter. In the case of William’s *Rhetorica*, the attribution to William seems fairly clear. In the bulk of cases, and for every chunk, the style is more like William than someone else. Regarding the text with uncertain attribution, the stylistic match with Bartholomew is very poor. In fact, the distances indicate that it is a better stylometric match to several of the ‘imposter’ authors than to Bartholomew. The results for each chunk can be summarised by the mass of the distribution that lies above 0 (the strength of the positive match). The summary chunk results can themselves be summarised by their median, although this reductive approach suppresses a great deal of the

³⁵ MIKE KESTEMONT, JUSTIN STOVER, MOSHE KOPPEL, FOLGERT KARSDORP, WALTER DAELEMANS, « Authenticating the Writings of Julius Caesar », *Expert Systems with Applications*, 63 (2016), p. 86–96. The new methods implemented for this article include the ‘ranking’ improvement introduced in NEKTARIA POTHAS, EFSTATHIOS STAMATATOS, « An Improved Imposters Method for Authorship Verification », in GARETH J. F. JONES, SÉAMUS LAWLESS, JULIO GONZALO, LIADH KELLY, LORRAINE GOEURIOT, THOMAS MANDL, LINDA CAPPELLATO, NICOLA FERRO (eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings*, Springer, Cham 2017, 138–144.

variation that is evident. In those terms, the attribution of William's *Rhetorica* to him is strong at around 79.4%. The attribution of the other *Rhetorica* to Bartholomew of Messina is around 3.3%. It was mentioned above that it is a better match to other candidates – for example 17.5% to William of Moerbeke, 15.7% to James of Venice and 33.2% to Burgundio of Pisa. It is important to note, though, that none of these figures should be construed as suggesting a possible attribution. Rather, they demonstrate how poorly the work matches our other examples of Bartholomew's style in terms of function words.³⁶

It is important to contextualise and interpret these results. To that end, we performed attributions across the entire corpus to determine the degree of certainty with which the methods attributed each work to its (assumed) translator. Those results are visible in Fig. 5. In each case, the distribution of chunk results is shown, summarised as above by the percentile above zero (the circular points show the precise values, the distribution curve is estimated), along with the median value (the central line in the distribution). From this we can see that match strength is quite variable, with some unquestioned translations displaying rather poor scores. This can be for many reasons – the work being translated might be in a genre with repetitive phrasing, the translator may have translated works from very different times, or the authors themselves might have very different styles in Greek, which naturally leads to variation in the translation. We can see that it is not simply that translators for whom we have more text are more easily attributed – the attributions to Henry Aristippus and the *Physica Vaticana* to Anonymous are strong; this occurs because the works by those translators are self-similar. Burgundio of Pisa, for whom we have a great deal of text, does not always attribute strongly, probably because of this internal stylistic variation. William of Moerbeke is both strongly self-similar as well as abundant. Which brings us to the perplexing results from Bartholomew's corpus. Not only is the *Rhetorica* a poor match, but so are other works generally considered to be his – in particular the *Magna moralia* and *De mundo*. Of course, the results from the *Rhetorica* are markedly worse as matches, so this does not give us cause to immediately assume misattribution in the case of the other two works, but it must be noted that the attribution of *De mundo* to Bartholomew is not iron-clad, as discussed above. What we must certainly accept, however, is that Bartholomew's style, as was already seen in the

³⁶ We did not include Burgundio as a potential translator of the *Rhetorica anonyma* in this study because he has not been suggested in existing scholarship and some of his stylistic trademarks, in particular his use of two or three alternative equivalents for difficult Greek terminology, are missing from this translation. It is true that the unsupervised clustering (Fig. 1) places the work near Burgundio's general cloud, and that the bootstrap match suggests a faint feasibility, but in our view this is simply because there is so much text from Burgundio in the corpus and his style of translation varies so widely (see Fig. 5).

clustering results, is much less self-similar than that of the other translators. But, to restate the obvious, the stylometric results *only include information drawn from the use of function words*. Function words as features are clearly effective (at least for the other translators) but there remain significant aspects of style which are completely ignored by these methods. This highlights both the value of quantitative stylometric approaches (in that they provide unbiased and reliable evidence based on, but extending, the techniques pioneered years ago by experts like Minio-Paluello and Bossier), but also the urgent need to contextualise the results and harmonise them with traditional philology.

III.7. Results and Challenges

Based on the results of our stylometric investigation, we continue to believe strongly that the use of function words is an effective means of identifying Greek-Latin translator style in general. In terms of overall performance, we have shown in previous work that machine-learning classifiers are able to correctly predict the translator of 1000-word chunks from our corpus with more than 90% accuracy.³⁷ This impressive performance, however, is no guarantee of accuracy when the methods are applied to an individual work – the performance varies both by translator and by work, with some cases being much more difficult and uncertain than others. In the case at hand, the *interpretation* of the analysis is reasonably clear; the stylometry suggests that the unattributed *Rhetorica* does not match the style of Bartholomew of Messina, *contra* our initial hypothesis. However, as mentioned above, the analysis is limited in scope (it only considered function words) and the translation style of Bartholomew is, according to all the evidence, more variable than that of the other translators in our corpus.

At this point, then, what we have is presumptively a negative result, but we argue that it should not be regarded as conclusive. Editors of the text, as well as editors of other translations by Bartholomew share an intuition that the work ‘feels’ like his style, and these scholarly intuitions are not to be set aside on the basis of a single computational study. There is considerable future work that is required before our rejection of the attribution to Bartholomew should be considered confirmed. The first task is to better understand the concrete differences in the use of function words, drawing upon the features of the text itself and comparing it to the authorial style of other Aristotelian translations from Bartholomew to see if the differences can be explained by the Greek originals. Failing this, the obvious next step would be to create an unbiased computational

³⁷ PIETER BEULLENS, WOUTER HAVERALS, BEN NAGY, « The Translators’ Touch. A Computational Stylometric Inquiry into Medieval Greek-Latin Translations » Long paper presented at *DH Benelux 2023*, Brussels.

methodology that could test the style in terms of key content words, operationalising the philological approach pursued by both Schneider and Beullens (who reached different conclusions based on the same text) that was discussed above.

Finally, though, we must also consider the historical landscape in the event that the translation is determined not to be by Bartholomew. In light of this result, and considering also our earlier suspicion, based on stylometric evidence, that the translations of Sextus Empiricus are neither by Bartholomew nor (probably) both by the same translator, we face the unsettling possibility that there is not just one more anonymous translator from Greek into Latin in the Italian thirteenth century, but perhaps at least three. This cannot help but reshape our understanding of the breadth of scholarship in late mediaeval Europe.

The Elementary Particles

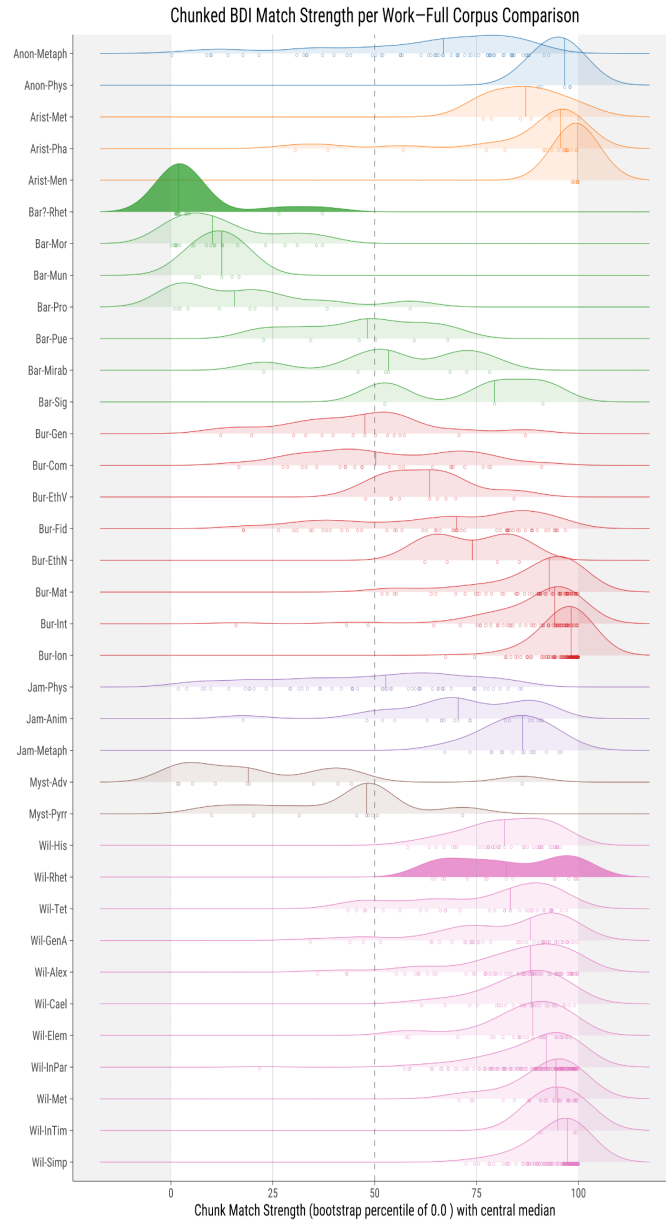


Fig. 5. Bootstrap Distance Imposters results for each work in the corpus when attributed to its own translator. 1000 iterations per chunk, bootstrapped at 35%, Ružička metric. Each chunk is shown as a circle in the rug plot, with an estimated distribution marked with the median. The ‘confidence’ score is the mean of the bootstrap confidence percentiles (percentage of distribution mass above 0). Higher numbers are more confident attributions.

Appendix A

FUNCTION WORDS USED

As described above, these are the 54 function words that were selected as a basis for the analysis:

atqui	ideo	quoniam
aut	igitur	quoque
autem	ita	precipue
certe	itaque	rursum
ceu	iterum	rursus
confestim	mox	sane
cum	multotiens	scilicet
dehinc	nam	sepe
deinceps	namque	sic
demum	nempe	sicut
denique	nimirum	siquidem
enim	puta	tamquam
ergo	quando	ut
etiam	quemadmodum	utique
exinde	quia	velut
forsan	quidem	veluti
forsitan	quippe	vero
fortassis	quod	videlicet

Appendix B

CORPUS OF TEXTS

Except for those stated below, all text files were kindly made available for the sole purpose of this investigation by Brepols Publishers and its Centre *Traditio Litterarum Occidentalium* from their databases *Library of Latin Texts* and *Aristoteles Latinus Database*, where the relevant bibliographical references can be accessed. The authors particularly acknowledge the help of Eddy Gouder, Tim Denecker, and Bart Janssens.

The following texts were acquired from other sources:

- Henricus Aristippus, *Meno*: VICTOR KORDEUTER, CARLOTTA LABOWSKY, *Meno. Interprete Henrico Aristippo*. *Plato Latinus*, vol. I, Warburg, London 1940, reprint Kraus, Nendeln 1973.

The Elementary Particles

- Henricus Aristippus, *Phaedo*: LAURENTIUS MINIO-PALUELLO, H.J. DROSSAART LULOFS, *Phaedo. Interprete Henrico Aristippo*. Plato Latinus, vol. II, Warburg, London 1950, reprint Kraus, Nendeln 1973.
- Bartholomew of Messina, *De natura pueri*: unpublished edition by Pieter Beullens.
- Burgundio of Pisa, *De complexionibus*: RICHARD J. DURLING, *Burgundio of Pisa's Translation of Galen's ΠΕΡΙ ΚΡΑΣΕΩΝ «De complexionibus»*, de Gruyter, Berlin-New York 1976.
- Burgundio of Pisa, *De interioribus*: RICHARD J. DURLING, *Burgundio of Pisa's Translation of Galen's ΠΕΡΙ ΤΩΝ ΠΕΠΙΟΝΘΟΤΩΝ ΤΟΠΩΝ «De interioribus»*, 2 vols, Steiner, Wiesbaden 1992.
- Johannes Chrysostomus, *In Iohannem*: text from his projected edition, kindly made available by professor Chris Nighman (clioproject.net).
- Johannes Chrysostomus, *In Matthaeum*: text from his projected edition, kindly made available by professor Chris Nighman (climo-project.wlu.ca).
- Sextus Empiricus, *Adversus Mathematicos*: unpublished edition by Pieter Beullens.
- Sextus Empiricus, *Pyrrhoniatarum informationum libri*: sample from his projected edition kindly made available by Dr Roland Wittwer.

The following table lists the texts used in the analysis, the short sigla used to refer to them in the reproduction code and Figures in this paper, and the size of the texts in words:

Translator	Siglum	Author	Text	Words
Anonymous	Anon-Metaph	Aristoteles	<i>Metaphysica Anonyma</i>	61295
	Anon-Phys	Aristoteles	<i>Physica Vaticana</i> (fragm.)	5850
Henricus Aristippus	Arist-Men	Plato	<i>Meno</i>	9469
	Arist-Met	Aristoteles	<i>Meteorologica</i> IV	6812
	Arist-Pha	Plato	<i>Phaedo</i>	20941
Bartholomew of Messina	Bar-Mir	Aristoteles	<i>De mirabilibus auscultationibus</i>	7306
	Bar-Mor	Aristoteles	<i>Magna moralia</i>	20739
	Bar-Mun	Aristoteles	<i>De mundo</i>	5640
	Bar-Pri	Theophrastus	<i>De principiis</i>	2914
	Bar-Pro	Aristoteles	<i>Problemata</i> 1, 11, 34–37	10324
	Bar-Pue	Hippocrates	<i>De natura pueri</i>	6149
	Bar?-Rhet	Aristoteles	<i>Rhetorica</i> I	14409
	Bar-Sig	Theophrastus	<i>De signis</i>	3386

Pieter Beullens, Wouter Haverals & Ben Nagy

Burgundio of Pisa	Bur-Com	Galenus	<i>De complexionibus</i>	24901
	Bur-EthN	Aristoteles	<i>Ethica Nova</i> (fragm.)	5066
	Bur-EthV	Aristoteles	<i>Ethica Vetus</i> (fragm.)	9697
	Bur-Fid	Johannes Damascenus	<i>De fide orthodoxa</i>	53755
	Bur-Gen	Aristoteles	<i>De generatione et corruptione</i>	14993
	Bur-Int	Galenus	<i>De interioribus I</i>	64149
	Bur-Ion	Johannes Chrysostomus	<i>In Iohannem</i> (33–88)	133001
	Bur-Mat	Johannes Chrysostomus	<i>In Matthaëum</i> (1–17)	66626
James of Venice	Jam-Anim	Aristoteles	<i>De anima</i>	19834
	Jam-Metaph	Aristoteles	<i>Metaphysica Vetustissima</i> (fragm.)	15961
	Jam-Phys	Aristoteles	<i>Physica Vetus</i>	53042
Unknown	Myst-Adv	Sextus Empiricus	<i>Adversus mathematicos</i> (fragm.)	10095
	Myst-Pyrr	Sextus Empiricus	<i>Pyrrhoniæ informationum libri</i> (sample)	9994
William of Moerbeke	Wil-Alex	Alexander of Aphrodisias	<i>In Meteorologica</i>	71996
	Wil-Cael	Aristoteles	<i>De caelo</i>	26961
	Wil-Elem	Proclus	<i>Elementatio theologica</i>	25782
	Wil-GenA	Aristoteles	<i>De generatione animalium</i>	44160
	Wil-His	Aristoteles	<i>De historia animalium I–V</i>	40000
	Wil-InPar	Proclus	<i>In Parmenidem Platonis</i>	161594
	Wil-InTim	Proclus	<i>In Timæum Platonis</i> (fragm.)	4775
	Wil-Met	Aristoteles	<i>Meteorologica</i>	30176
	Wil-Rhet	Aristoteles	<i>Rhetorica I</i>	15010
	Wil-Simp	Simplicius	<i>In De caelo I</i>	120399
	Wil-Tet	Ptolemaeus	<i>Tetrabiblos</i>	33712