Clavius' edition of Euclid's *Elements*: A computational approach

Fabio Acerbi*

CNRS, UMR8167; fabacerbi@gmail.com

Andrea Bréard

Friedrich-Alexander-Universität Erlangen-Nürnberg; andrea.breard@fau.de

Florian Kessler

Friedrich-Alexander-Universität Erlangen-Nürnberg; florian.kessler@fau.de

Ramon Masià

Magna Societas Catalanorum; rmasia@gmail.com

Abstract

Christoph Clavius' Latin translation of Euclid's *Elements* is studied by applying standard and less standard tools of computational linguistics. Clavius' lexical choices are compared with the relevant Greek and Latin editions and translations. We shall thereby assess Clavius' treatment of Latin sources, as well as his specific, scholarly aims and his general strategy.

Keywords

Euclid's *Elements*, Christoph Clavius, Latin translations, computational linguistics

How to cite this article

Acerbi, Fabio, Andrea Bréard, Florian Kessler, Ramon Masià. "Clavius' edition of Euclid's *Elements*: a computational approach." *Galilæana* XXIII, 1 (2026): 35–80; doi: 10.57617/gal-99.

Acknowledgments

We are grateful to Veronica Gavagna, Pier Daniele Napolitani, and Bernard Vitrac for help and suggestions, to the referees for their constructive criticisms, which resulted in major improvements.

Article data

Date submitted: October 2025

Date accepted: April 2026

Basics on Christoph Clavius and his edition of the Elements

Christoph Clavius' (1538–1612)[1] edition of Euclid's *Elements* stands out as one of the achievements of the "Italian Renaissance of mathematics."[2] After the first edition of 1574,[3] this monumental work was repeatedly reprinted during Clavius' lifetime, sometimes substantially enhanced by the author, and was finally set to open his *Opera Mathematica* of 1612.[4] Clavius' edition comprises a Latin version of the Greek text and an overwhelming scholarly apparatus: additions to the original text in the form of definitions, lemmas, corollaries, and propositions, and again commentary, *praxeis*, and digressions on subjects that were not included in the original.

Clavius had a strategic aim in producing this kind of edition, and in fact in designing most of his works:[5] establishing mathematical sciences as a scholarly discipline, suited to being included in the curriculum of the Jesuits' academies and, more generally, consolidating the Society of Jesus as a leading academic authority within the early modern Republic of Letters.[6] Consequently, Clavius' aims in his Euclid edition were completeness and clar-

1 A detailed biographical account of Clavius can be read in Clavius, Christoph. *Corrispondenza*, ed. by Ugo Baldini and Pier Daniele Napolitani (Pisa: Università di Pisa, Dipartimento di Matematica, 1992), vol. I, parte 1, 33–58. For the scientific context in Rome, with special emphasis on Clavius' contributions to the Gregorian reform of the calendar, see Ugo Baldini, "Christoph Clavius and the Scientific Scene in Rome," in *Gregorian Reform of the Calendar. Proceedings of the Vatican Conference to Commemorate Its 400th Anniversary. 1582-1982*, ed. by George V. Coyne, Michael A. Hoskin, and Olaf Pedersen (Città del Vaticano: Pontificia Academia Scientiarum, 1983), 137–169. For Clavius' role in the last days of Ptolemaic astronomy see James M. Lattis, *Between Copernicus and Galileo. Christoph Clavius and the Collapse of Ptolemaic Cosmology* (Chicago and London: The University of Chicago Press, 1994). For overviews of Jesuit Science after Clavius see Mordechai Feingold, ed., *The New Science and Jesuit Science: Seventeenth Century Perspectives* (Amsterdam: Kluwer, 2003), and Michael John Gorman, *The Scientific Counter-Revolution. The Jesuits and the Invention of Modern Science* (London; etc.: Bloomsbury Academic, 2020).

2 This is the title of an excellent study published half a century ago: Paul L. Rose, *The Italian Renaissance of Mathematics* (Genève: Librairie Droz, 1975).

3 *Euclidis Elementorum libri XV. Accessit XVI de Solidorum Regularium comparatione. Omnes perspicuis demonstrationibus, accuratisque scholijs illustrati. Auctore Christophoro Clavio Bambergensi Societatis Iesu. Romae, Apud Vincentium Accoltum. 1574*. On Clavius' edition see Sabine Rommevaux, *Clavius : une clé pour Euclide au XVI^{e} siècle* (Paris: Librairie Philosophique J. Vrin, 2005), with focus on Book V.

4 *Christophori Clavii Bambergensis e Societate Iesu Opera Mathematica V. Tomis distributa [...]. Moguntiae Sumptibus Antonij Hierat excudebat Reinhardus Eltz [...] Anno M. DCXII*, 13–638.

5 A complete list of first editions is in Clavius, *Corrispondenza*, vol. I, parte 3, 5–11.

6 See the studies collected in Ugo Baldini, *Legem Impone Subactis. Studi su filosofia e scienza dei Gesuiti in Italia. 1540-1632* (Roma: Bulzoni, 1992); Romano Gatto, *Tra scienza e immaginazione. Le matematiche presso il collegio gesuitico napoletano (1552-1670 ca.)* (Firenze: Olschki, 1994); Rosario Moscheo, *I gesuiti e le matematiche nel secolo XVI. Maurolico, Clavio e l'esperienza siciliana*

ity of the proofs, exhaustiveness as to the proven results,[7] completeness of the scholarly apparatus,[8] emphasis on constructions (the *praxeis*) and on practical applications.[9]

It thus comes as no surprise that Clavius' Latin text of the *Elements* proper is not a translation from the Greek, but a complex mixture of (unavowed) borrowings and systematic rewriting.[10] It is not even clear whether Clavius was able to read Greek, or, if indeed he could read it,[11] whether he ever bothered to check Grynaeus' *princeps* of 1533, let alone any Greek manuscript.[12] This approach was not new, to some extent: in fourteenth- and fifteenth-century Byzantium, every copyist-scholar shaped his own edition of the *Greek* text of the *Elements* by conflating material drawn from several models and by more or less heavily modifying the text.[13] Grynaeus' edition – and in fact any edition of the Greek *Ele-*

(Messina: Società Messinese di Storia Patria, 1998); Antonella Romano, *La Contre-réforme mathématique. Constitution et diffusion d'une culture mathématique jésuite à la Renaissance* (Roma: École française de Rome, 1999), 35–178. Clavius' syllabus is edited and discussed in Romano Gatto, "Christoph Clavius' 'Ordo Servandus in Addiscendis Disciplinis Mathematicis' and the Teaching of Mathematics in Jesuit Colleges at the Beginning of the Modern Era," *Science & Education* 15 (2006): 235–258.

7 Read, at the beginning of the 1574 edition, the 48-page *Index problematum, ac theorematum, quae praeterea, quae continentur in Euclidis propositionibus, in his elementorum libris demonstrantur*. Following a mediaeval tradition, Clavius also proved all Euclidean "common notions," which included two postulates (see note 14 below): Vincenzo De Risi, "The Derivability Theory of Axioms. Logic and Mistranslation in the Middle Ages and the Renaissance," in *Pre-Modern Mathematical Thought. The Latin Discussion (13th–16th Century)*, ed. by Clelia V. Crialesi (Leiden: Brill, 2025), 237–288.

8 In this case, Clavius appears regularly to mention his sources, if only to contradict them. The issue is not irrelevant, for recent scholarship has sometimes regarded refuting the opposite claim as necessary: Eberhard Knobloch, "Sur la vie et l'œuvre de Christophore Clavius (1538–1612)," *Revue d'histoire des sciences* 41 (1988): 331–356, at 334–335, and Eberhard Knobloch, "L'œuvre de Clavius et ses sources scientifiques," in *Les jésuites à la Renaissance. Système éducatif et production du savoir*, ed. by Luce Giard (Paris: Presses Universitaires de France, 1995), 263–283, at 267–270. A blatant case of unavowed borrowing escaped Knobloch: see below.

9 On this feature see in particular Audrey Price, "Pure and Applied: Christopher Clavius's Unifying Approach to Jesuit Mathematics Pedagogy" (PhD diss., University of California, San Diego, 2017).

10 Yet, Clavius also calls it *interpretatio* (*Ordo*) and *traditio* (*Ordo brevior*): Gatto, "Christoph Clavius'," 249, 252–253. More on this crucial point in our Conclusion(s).

11 In his edition of the *Elements*, Books I–VI, Clavius mentions 3 Greek words: βεβηκέναι, ἀπολαμβάνω, ἀναλογία. The contexts of the mentions might suggest that Clavius' command of Greek was at best suboptimal. See also note 80 below.

12 *ΕΥΚΛΕΙΔΟΥ ΣΤΟΙΧΕΙΩΝ ΒΙΒΛ. ΙΕ'. ΕΚ ΤΩΝ ΘΕΩΝΟΣ ΣΥΝΟΥΣΙΩΝ. Εἰς τοῦ αὐτοῦ τὸ πρῶτον, ἐξηγημάτων Πρόκλου βιβλ. δ'. […]. Basileae Apud Ioan. Hervagium Anno M. D. XXXIII. Mense Septembri.*

13 On this widespread Byzantine practice of appropriation by rewriting see Fabio Acerbi, "Rewriting Mathematical and Astronomical Treatises," in *The Routledge Handbook for Rewriting in Byzantium*, ed. by Juan Signes Codoñer, Martin Hinterberger, and Inmaculada Pérez Martín (London: Routledge, 2025), 407–419.

ments until Peyrard's in 1814 – inherited a striking example of such a tampering with the text: the last two postulates were printed among the "common notions."[14]

Clavius could check plenty of Latin translations: the most successful Arabo-Latin translation, and the first that was printed, namely, Campanus';[15] the first Renaissance translation from the Greek, authored by Bartolomeo Zamberto,[16] first published in 1505 along with the translation of Euclid's minor works and subsequently, and aptly, reprinted propositionwise, intermingled with Campanus', by Lefèvre d'Étaples in 1516;[17] Federico Commandino's translation, printed two years before Clavius' edition.[18] Other translations, sometimes associated with the Greek text,[19] were available, and even abridged editions comprising only the principles and the enunciations of the propositions. Of the latter kind was Stephanus Gracilis' edition, printed in 1557.[20] The latter is exactly the only obvious source of Clavius' Latin text of the *Elements*: for reasons that escape us and without recognising his debt, he lifted everything he could lift from Gracilis' booklet, making small, cosmetic changes, if any, and modifying the wording in a few cases only.[21] As for the

[14] See Fabio Acerbi, Stefano Martinelli Tempesta, and Bernard Vitrac, "Gli interventi autografi di Giorgio Gemisto Pletone nel codice matematico Marc. gr. Z. 301," *Segno e Testo* 14 (2016): 411–456. This is the reason why, in Bolyai's groundbreaking work on the parallel postulate, this is referred to as *axioma XI euclidei*, and not as *postulatum V euclidei*.

[15] The colophon is *Opus elementorum euclidis megarensis in geometriam artem In id quoque Campani perspicacissimi Commentationes finiunt. Erhardus ratdolt Augustensis impressor solertissimus. venetiis impressit. anno salutis M.cccc.lxxxij. Octavis. Calendas. Junii.*

[16] *Euclidis megarensis philosophi platonicj Mathematicarum disciplinarum Janitoris […] Bar. Zamber. Vene. Interprete. […] Impressum Venetiis […] in aedibus Ioannis Tacuini librarii accuratissima diligentia recognitum. Anno reconciliatae divinitatis.M.D.V.VIII.Kalendas novembris*. For Zamberto's manuscript sources, see Bernard Vitrac, "A propos de l'histoire du texte des *Éléments* d'Euclide : Préalables à une nouvelle édition critique" (hal-03328161, 2022).

[17] *Euclidis Megarensis Geometricorum elementorum libri XV, Campani Galli transalpini in eosdem commentariorum libri XV, Theonis Alexandrini Bartholamaeo Zamberto veneto interprete, in tredecim priores, commentariorum libri XIII, […] Parisiis in officina Henrici Stephani e regione scholae Decretorum* (1516).

[18] *Euclidis Elementorum libri XV. Unà cum scholijs antiquis, À Federico Commandino urbinate nuper in latinum conversi, commentarijsque quibusdam illustrati, Pisauri, M D LXXII*. For Commandinus' manuscript sources see Bernard Vitrac, "La traduction latine des *Éléments* d'Euclide par Federico Commandino : sources, motivations" (hal-03328386, 2021).

[19] See the website Méδée at https://thamous.univ-rennes1.fr/docs/Medee/gb/index.html# (Bernard Vitrac and Alain Herreman).

[20] *Euclidis Elementorum libri XV. Graecè & Latiné, Quibus, cùm ad omnem Mathematicae scientiae partem, tùm ad quamlibet Geometriae tractationem, facilis comparatur aditus […]. Lutetiae, Apud Gulielmum Cavellat […] 1557.*

[21] See Francesca Salvi, "Le fonti degli *Elementi* di Euclide di Cristoforo Clavio" (Tesi di Laurea, Università degli Studi di Firenze, 2005), part 2, and, for Book V, Enrico Giusti, "La théorie des proportions au XVI^{e} siècle : entre philologie et mathématiques," in Liber amicorum *Jean Dhom-*

constructions and the proofs, Clavius made eclectic choices:[22] in most cases he loosely followed the deductive progression of the Greek text while adding cases; sometimes, he reorganised the proof; in a few cases, he radically changed it.[23] In every instance, Clavius did not translate the text (read his explicit statement in our Conclusion(s)), but provided a new formulation of almost all sentences. He likewise felt free to redraw almost all diagrams; his propositions may also adopt a different lettering (first example in Prop. I.3).

The aims of the present contribution

In our perspective,[24] Clavius' choice of lifting the principles and the enunciations from Gracilis' edition while translating the rest *suo Marte* has a capital consequence: the lexicon of his edition is redundant. There are two striking examples. One comes from Prop. I.11, where the enunciation has Gracilis' *excitare* for ἀγαγεῖν ("to raise"), whereas 2 lines later, the setting-out, which is intended to replicate the lexical content of the enunciation and is absent in Gracilis' edition,[25] has *erigere*. The other is from Book V, where λόγος is Gracilis' *ratio* in the enunciations but (often, not always!) Clavius' own preferred translation *proportio* in the proofs.

Well, what is our perspective, and what are our aims? We shall automatically process the lexical content of Euclid's text in Clavius' edition and compare it with the lexical content of some reference texts by using standard and less standard stylometric tools. We

bres, ed. by Patricia Radelet-de Grave (Turnhout: Brepols, 2008), 173–193, at 185 and 193.

22 See, by clicking on "Clavius 1574," the structural correlations listed in the Méδée website mentioned in note 19 above.

23 The partitions into these three categories of the proofs of the propositions in Books I and XI are (28,19,1) and (27,11,2), respectively: Salvi, "Le fonti," 93 and 129, and the detailed comparisons at 93–128 and 129–151, respectively.

24 The stress on "our perspective" also means that the lexical redundancy of Clavius' Latin text of the *Elements* does not affect the mathematical soundness and readability of the latter, nor does it undermine Clavius' work as a mathematician. We shall return on this point in the Conclusion(s). An extreme example of lexical redundancy of a (mediaeval) Latin translation from Greek is given by Henry Aristippus' version of Plato's *Phaedo*: Lorenzo Minio-Paluello, "Henri Aristippe, Guillaume de Moerbeke at les traductions latines médiévales des Météorologiques et du De Generatione et Corruptione d'Aristote," *Revue Philosophique de Louvain* 45 (1947): 206–235, at 213–214 = Lorenzo Minio-Paluello, *Opuscula. The Latin Aristotle* (Amsterdam: Adolf M. Hakkert, 1972), 64–65.

25 For the proof-theoretical terminology used in this paper, and in general for all features of the demonstrative stylistic code of Greek mathematics, see Fabio Acerbi, *The Logical Syntax of Greek Mathematics* (New York; Heidelberg: Springer, 2021). See our Conclusion(s) on Clavius programmatically disregarding all features – even such obvious features as the exact and systemic correspondence between enunciation and setting-out – of the Euclidean demonstrative stylistic code.

conceived and realised both the database (xml coding, tagging, and lemmatisation of all works mentioned below – in fact, of the entire Greek and Byzantine mathematical and astronomical corpus – supplemented by a three-tier morphosyntactic categorisation of the lemmas)[26] and, unless otherwise stated, the Python scripts that allow creating and automatically processing it. We also conceived the several notions introduced and the scripts that implement them. Neither our database nor our scripts are in the public domain.

The reference texts we shall use are the Greek text of the *Elements* in Heiberg's standard edition, the 12th-century Graeco-Latin translation in Busard's edition, and Commandinus' translation.[27] We did not include any Arabo-Latin translation because the lexical record is blurred by the passage through Arabic. We did not include any other Renaissance translation of the Greek *Elements* because the task of producing and checking an OCR of any such text is time-consuming, life is short, and skilled contractors apparently hide behind Pluto's orbit.[28] In particular, both editions of Zamberto's translation are an OCR nightmare, with the additional hassle that one must disentangle the thousands of references to previous results, by book and proposition number, that Zamberto introduced in the text. For the same reasons we limited our investigation to Books I–VI of the *Elements*.

Our primary aim is to provide a quantitative, multi-focus assessment of the aforementioned lexical redundancy of Clavius's text. Our strategic aim is to establish a protocol for automatically processing corpora made of highly regularised scientific texts written in natural language, possibly involving translations. One must stress that stylometry as applied to this kind of texts cannot provide anything that is not obvious to any well-informed and perceptive reader. It does, however, provide a quantitative approach – and hence numer-

26 According to our forms-to-lemma database, which amounts to a perfect parser by its very conception, the standard Greek parsers applied to technical texts have a failure rate of about 30%. The performance rate for technical Latin is higher but still far from optimal.

27 The Graeco-Latin translation had almost no circulation: two extant witnesses only are known, one of them incomplete – but recall that the quotes from the *Elements* in Fibonacci's *Liber abbaci* (1228) are taken form this translation: Menso Folkerts, "Leonardo Fibonacci's Knowledge of Euclid's *Elements* and of Other Mathematical Texts," *Bollettino di Storia delle Scienze Matematiche* 24 (2004): 93–113. The Graeco-Latin translation is edited in Hubert L. L. Busard, *The Mediaeval Latin Translation of Euclid's* Elements *made directly from the Greek* (Stuttgart: Franz Steiner Verlag, 1987); it is an output of the so-called "Sicilian school" of translation (a bit more on this in note 76 below). The e-text of this translation has been downloaded from the website https://www2.hf.uio.no/polyglotta/index.php, checked (just a few mistakes have been spotted), and converted to an e-text suitable for our purposes. Commandino's translation and Clavius' edition have been OCR'd, checked, and converted.

28 The production of the regularised e-texts of Commandinus' translation and Clavius' edition was partly funded by the project Sin-aps (Alexander von Humboldt-Professorship, FAU Erlangen-Nürnberg). Getting regularised e-texts of all Latin translations of the *Elements* is one of the goals of the ANR-DFG project EUCLIDES (PI Vincenzo de Risi), see https://sphere.cnrs.fr/en/projets-en-cours/euclides-anr-project-2024-2027/.

ical data and powerful visualisation tools – to such obvious yet vaguely apprehended features: in our case, that Clavius' text of the *Elements* is not a translation. In a sense, then, the notions, the tables, the graphs, the diagrams, the bulk and intensive indicators presented in what follows are not simply a tool, but the goal, and for this reason they are a good many and varied: they *are* a significant quantitative description of the involved texts according to a carefully devised protocol, and hence they speak for themselves and are, at least in our present perspective, more important than any moral we may derive from them – we shall derive one in the final conclusion, don't worry.

As said above, the entire lexical record of the above-listed four works has been recorded and tagged via lemmatisation. As the frequency distributions of word tokens and types in a text are statistical ensembles, we shall frequently use notions borrowed from statistical thermodynamics and define a number of statistical features and indicators accordingly.[29] In particular, we shall systematically use suitably defined "entropy" and "temperature" parameters as indicators of lexical disorder,[30] and a set of matrices and associated numerical indicators that display and measure, respectively, the lexical reorganisation that occurs in translation. The discussion focuses in succession on defining, computing, presenting, and comparing the relevant statistical features and indicators of the lexicon of the selected works. These features and indicators fall into five broad categories:

- *General data related to lexical content*. We discuss the progressive creation of the dictionary of a given work; the extent to which a power law is a good fit to the frequency distribution of the lemmas; the distribution of *hapax*, one-form lemmas, and most frequent lemmas; the form spread within one and the same lemma; the distribution of lemmas, forms and occurrences among the parts of speech (POS).
- *Location and density of the information content within a proposition and a Book, by morphosyntactic types*. We introduce the intensive indicators X-entropy and X-cloud for categorial types X that generalise the standard POS; some representative computations of these indicators are carried out.
- *Lexical reorganisation consequent on translation*. We discuss in detail the lexical spread consequent on translation by introducing the correlation tables and the correlation and dispersion matrices of a translation, along with the notion of "metalemma." On these grounds, we measure the lemmatic reorganisation that has occurred in a given translation by introducing several indicators that process the data set out in the correlation tables and in the dispersion matrix.

29 In the same perspective, see the breakthrough study Pierluigi Contucci, Claudio Giberti, Godwin Osabutey, and Cecilia Vernia. "Statistical properties of the rooted-tree encoding of ℕ," *Physica A* 686 (2026), art. 131361, https://doi.org/10.1016/j.physa.2026.131361.

30 Entropy was famously introduced in information theory, and thereby and later in computational linguistics, in Claude E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* 27 (1948): 623–656.

- *Textual proximity*. After this, standard stylometric tools are used to determine the relations of proximity between the selected works: cluster analysis, consensus trees, network analysis, and principal component analysis (PCA).[31]
- *Syntactic and deductive depth*. A dedicated Section goes deeper into syntax by displaying and discussing the "diagrams of subordination" of Book I and by comparing a few bulk and intensive indicators of deductive depth.

A warning is in order before beginning with the lexical analysis. Even strictly analogous texts in Greek and Latin – even more so if the Latin texts are translations of a Greek text – need not display similar lexical data. The main reason, and a crucial point in the process of naturalisation of a rigid stylistic code as the one employed in Greek demonstrative texts,[32] is the absence of the definite article in Latin. This may be partly compensated by an increase of the (pro)nominal content of the Latin translations,[33] for the nominalising article in Greek technical texts also allows omitting plenty of nouns. As we shall see in the next Section, the relative frequency of the article in the Greek *Elements* is very high. However, denotative letters are an essential feature of the demonstrative code and at least one article must figure in each designation. Consequently, most of the articles can be safely omitted in translation,[34] even if they may be replaced by nouns and pronouns.[35] Add that the verb morphology is richer in Greek than in Latin, and that ambiguous word forms are less frequent in Greek than in Latin: think of the ambiguity of such widespread forms as *quam* and *quod*, a real nightmare for someone engaged in lemmatising a Latin text.

31 We used the `stylo` package for dendrograms, consensus trees and PCA plots. See the descriptions of the package in Maciej Eder, Jan Rybicki, and Mike Kestemont, *'Stylo': a package for stylometric analyses* (2015); Maciej Eder, Jan Rybicki, and Mike Kestemont, "Stylometry with R: A Package for Computational Text Analysis," *R Journal* 8 (2016): 107–121. The package can be downloaded at https://github.com/computationalstylistics/stylo.

32 On the meaning of "naturalisation" in this context see Abdelhamid I. Sabra, "The Appropriation and Subsequent Naturalization of Greek Science in Medieval Islam: A Preliminary Statement," *History of Science* 25 (1987): 223–243.

33 In the Graeco-Latin translation, the article is frequently replaced by conjoined forms of *is* and *qui*; Commandinus often employs forms of *ipse* to the same end. As a matter of fact, these pronouns "translate" the case of the denotative letter, as the nominalising article in Greek has first and foremost this function. For this reason, we did not regard Commandinus' *ipse* as a translation of the Greek article. On William of Moerbeke's "Latin" translation of the Greek nominalising article with the French article *le* – one of the highest points of his linguistic genius – see Lorenzo Minio-Paluello, "Guglielmo di Moerbeke traduttore della *Poetica* di Aristotele (1278)," *Rivista di Filosofia Neo-Scolastica* 39 (1947): 1–17, at 5–7 = Id., *Opuscula*, 44–46.

34 In the *Elements*, only 23% of the articles are not attached to denotative letters.

35 Just check the frequency of forms of *recta* in the first propositions of Book I in the Graeco-Latin translation.

Lexical analysis

By applying the tools of computational linguistics described in the previous Section, we now compare the lexical content of Books I–VI of the Greek *Elements* (*Elem.* henceforth), of the Graeco-Latin translation (*Gr.-Lat.*),[36] of Commandinus' translation (*Comm.*), and of Euclid's text in Clavius' 1574 edition (*Clavius*).[37] These texts will be collectively referred to as "the four works;" the three Latin texts as "the Latin translations" – even if we have seen that *Clavius* is not a real translation.

Some definitions are in order. Every continuous sequence of signs in a text is an "occurrence;" identical occurrences are a "form;" the set of forms related to one and the same term is a "lemma;" the set of lemmas is the "dictionary" of the text. A lemma is designated by writing one of its forms without diacritics, if any is required in the intended language: ἀπὸ, ἀπό, ἀπ', and ἀφ' are forms of the Greek lemma απο; *quotcunque* and *quotcumque* are forms of the Latin lemma *quotcumque*. The headwords of any lexicon follow the same principle, while keeping diacritics. Words resulting from a crasis such as κἀκεῖνον are split into καὶ + ἐκεῖνον, which are recorded separately as forms of the lemmas και and εκεινος; a Latin *word* to which enclitic *que* is attached is split as *andque* + *word*, which are assigned to different lemmas. In order to eliminate lexical noise, the denotative letters, the numbers, and the transliterated Greek words in Latin texts count as one single lemma each.

Raw data, creating the dictionary

The following table compares the raw lexical data of the four works (a heading like, e.g., "% l/o" means the ratio lemmas/occurrences in percentage).

	lemmas	forms	occurrences	% l/o	% f/o	% l/f
Elem.	348	1773	52894	0.66	3.35	19.63
Gr.-Lat.	415	1743	46067	0.90	3.78	23.81
Comm.	380	1665	45176	0.84	3.69	22.82
Clavius	619	2400	47371	1.30	5.07	25.79

36 This translation merely transliterates 29 forms and wavers between transliterating and translating 28 forms; see Busard, *The Mediaeval*, 12–15. Busard does not mention *esto*, which in no instance is the Latin future imperative of *sum*, but a transliteration of the Greek present imperative ἔστω; we shall treat this as a Latin form, but normally we count the transliterated forms that figure in Books I–VI as one single lemma, see just below.

37 Disentangling Euclid's text from the commentary in *Comm.* and *Clavius* is easy, since the commentary is expressly marked as such. Happily, both editions also put the book-and-proposition references to previous results in the margins. We did not include these references in the lemmatised text.

The lower number of occurrences in the Latin translations is entirely due to the absence of the article. *Elem.* is famously the lexically driest piece of literature ever written; *Comm.* is more restrained than *Gr.-Lat.*; the lexicon explodes with *Clavius*, which also engages its lemmas in a wider range of syntactic configurations (check the ratio forms/occurrences). The best linear regression of the Latin data with respect to size (occ.) is given by the ratio lemmas/forms ($\times 10^4$), with slope – 0.1365 and a coefficient of determination $R^2 = 0.9933$.

We may trace the progressive creation of the dictionary of an assigned work by computing the relative (that is, normalised to the size of the dictionary) number of lemmas engaged up to any given textual unit. In our case, "textual unit" means either a principle (definition, postulate, or common notion, the latter two in Book I only) or a proposition (regular proposition, lemma, or corollary).

The patterns of the progressive creation of the dictionary in the four works are displayed as Figures 1a–d of the Appendix; the textual units are counted sequentially starting from def. 1 of Book I; the red vertical bars mark the beginning of a Book. The steep increase in *Clavius*, textual unit 61, coincides with the beginning of the propositions of Book I. Otherwise, the curve in *Clavius* displays, more than the three other works, the sublinear behaviour typical of texts written in natural language, as codified in Heaps' law.[38] Conversely, the increase that coincides with the beginning of Book V is less marked in *Clavius* than in the three other works. As the words that denote a mathematical object – whose first occurrence is normally found in a definition – are a small fraction of the entire dictionary, there is no sharp step at the beginning of each Book.

Zipf's law

The most important statistical item associated with the lexicon of a text is the distribution of the number of occurrences referred to each lemma (this is the "frequency"). We can study the rank distribution of the lemmas as a function of their frequency,[39] in order to check to what extent they follow Zipf-Mandelbrot's law.[40] This law states that, for any

[38] Harold S. Heaps, *Information Retrieval. Computational and Theoretical Aspects* (New York; San Francisco; London: Academic Press, 1978), sect. 7.5.

[39] The least frequent lemma has rank 1, the second most frequent one has rank 2, etc.; consecutive ranks are associated with lemmas that have the same number of occurrences. Note that the literature is divided between using either the frequency or the rank as the independent variable. We adopt the same convention as in Newman's papers cited in the next note; this convention is to be preferred on theoretical grounds. Zipf and Mandelbrot put the rank as the independent variable.

[40] See Benoît Mandelbrot, "Information Theory and Psycholinguistic," in *Scientific Psychology*, ed. by Benjamin B. Wolman (New York: Basic Books, 1965), 550–562; Mark E. J. Newman, "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics* 46 (2005): 323–335; Aaron Clauset, Cosma R. Shalizi, and Mark E. J. Newman, "Power-law distributions in empirical data,"

sufficiently large linguistic *corpus*, the rank r of each lemma as a function of its frequency f (very nearly) follows an inverse power law like $r = Af^{-\alpha}$, where A and α are positive real numbers and $\alpha = 1$ in this case. If we take the logarithm of the two variables, the distribution of the data in a log-log graph is (very nearly) a straight line. Modifications made by Benoît Mandelbrot to Zipf's law include the introduction of a rank shift that amounts to setting a lower bound f_{min} to the scaling range and the generalisation to scaling exponents α other than 1.

The rank distributions of the lemmas as a function of the frequency in the four works are plotted in the log-log graphs displayed as Figures 2a–e of the Appendix (the scale is the same); the elongated "dots" in the graphs represent lemmas that have the same frequency. We also included the data of the complete text (that is, including all sorts of commentary) of Clavius' 1574 edition, Books I–VI; this text is noted *Clavius**. Two distributions appear to merge in each graph: an almost-linear yet sparsely populated high-frequency regime (normally dominated by function words, see below) merges with a less neatly delineated middle- and low-frequency regime. This means that middle-frequency lemmas are over-represented in the geometric sectorial lexicon, a fact that is also confirmed by the sparseness of the highest-frequency lemmas. As a mathematical treatise has two naturally embedded scales (the proposition and the Book), the superposition of at least two regimes is the behaviour that is to be expected.[41] The distribution of *Clavius** appears to be less hunched than the others; this will be confirmed by the numerical data introduced and set out in the next paragraph.

The table below sets out the following statistical parameters, in the hypothesis that a power law does fit the actual frequency data: the frequency f_{min} that is the lower bound of the scaling range; the maximum likelihood estimate of the scaling exponent α; the expected error σ on it; the p value of the goodness-of-fit KS test taken as a confirmatory parameter, which means that, in the indicated range, the power law is a good fit if $p < S$ (and the lower is p, the better is the fit), where the significance level S is usually taken to be 0.1 or 0.05. The hypothesis that the data follows a power law is thus in general weakly supported. A comparison with other candidate distributions by means of a suitably

SIAM Review 51 (2009): 661–703. In the computations described below, we used Jeff Alstott's package `powerlaw` (https://nbviewer.org/github/jeffalstott/powerlaw/blob/master/manuscript/Manuscript_Code.ipynb), presented in Jeff Alstott, Ed Bullmore, and Dietmar Plenz, "powerlaw: a Python package for analysis of heavy-tailed distributions" (arXiv:1305.0215v3, 2014); see the resources in https://aaronclauset.github.io/powerlaws/.

41 For a similar phenomenon occurring in large corpora (> 2000 lemmas) see Ramon Ferrer i Cancho and Ricard V. Solé, "Two Regimes in the Frequency of Words and the Origins of Complex Lexicons: Zipf's Law Revisited," *Journal of Quantitative Linguistics* 8 (2001): 165–173; Marcelo A. Montemurro, "Beyond the Zipf–Mandelbrot law in quantitative linguistics," *Physica A* 300 (2001): 567–578, which appears largely to draw on the previous.

defined p value shows that a power distribution is a better candidate ("Y" in the table) than an exponential distribution in all cases, whereas it is a worse candidate ("N" in the table) than a lognormal distribution in all cases bar *Clavius**. The high value of f_{min} for any text except *Clavius** confirms the fact that, in point of style, *Clavius** is more akin to a literary text than to mathematical text.

					power vs.	
	f_{min}	α	σ	p	exp	lognorm
Elem.	147	2.108	0.146	0.063	Y	N
Gr.-Lat.	182	2.049	0.158	0.070	Y	N
Comm.	187	2.257	0.176	0.049	Y	N
Clavius	139	2.217	0.149	0.054	Y	N
*Clavius**	1	1.398	0.010	0.044	Y	Y

Uniqueness of use: 1-occurrence and 1-form lemmas

Let us now look closely at the head of the frequency distribution of lemmas and forms. The data related to the lemmas that have 1 occurrence only ("*hapax*" without qualification, henceforth) or 1 form only (indeclinable words are included, the *hapax* are not) is as follows:

	lemmas	*hapax*	%	1 form	%	tot.	%
Elem.	348	51	14.66	88	25.29	139	39.94
Gr.-Lat.	415	78	18.80	126	30.36	204	49.16
Comm.	380	56	14.74	136	35.79	192	50.53
Clavius	619	131	21.16	199	32.15	330	53.31

This data is not easy to model. The relative frequency (lemm.) of *hapax* is fairly stable, as its spread, which is 44.61%, is lower than the size (lemm.) spread, which amounts to 77.59%. *Comm.* has fewer *hapax* than the other Latin texts but more 1-form lemmas.

The absolute and relative frequencies (lemm.) of the *hapax* by POS are as in the following table; the columns shaded grey contain the relative frequency of the overall lemmas referred to each POS (see the sub-Section *Distribution of lemmas, forms, and occurrences by POS* for the complete data). *Clavius*' metadiscursive drive is made obvious by the high relative frequencies of verbs and adverbs.

	Elem.			*Gr.-Lat.*			*Comm.*			*Clavius*		
	#	%		#	%		#	%		#	%	
adjective	15	29.41	25.86	19	24.36	26.02	15	26.79	25.00	30	22.90	20.84
adverb	7	13.73	10.92	14	17.95	12.29	14	25.00	15.00	33	25.19	18.42
noun	11	21.57	19.54	16	20.51	18.80	8	14.29	12.31	21	16.03	16.48
particle	4	7.84	8.62	6	7.69	8.43	2	3.57	2.46	5	3.82	8.89
preposition	1	1.96	4.60	4	5.13	5.30	1	1.79	1.26	1	0.76	4.20
pronoun	1	1.96	4.60	0	0.00	3.37	1	1.79	4.38	2	1.53	2.58
verb	12	23.53	25.29	19	24.36	25.06	15	26.79	25.29	39	29.77	28.27

We also give the distribution of the absolute and relative frequencies of the 1-form lemmas bar *hapax* by POS. The partly filled column gives the number of Greek indeclinable words in *Elem*. that exhibit at least two forms because of accentual phenomena (barytonesis, enclisis accent) or elision.

cat.	*Elem.*				*Gr.-Lat.*			*Comm.*			*Clavius*		
	#		%		#	%		#	%		#	%	
adjective	15		17.05	25.86	18	14.29	26.02	17	12.50	25.00	16	8.04	20.84
adverb	25	6	28.41	10.92	37	29.37	12.29	43	31.62	15.00	80	40.20	18.42
noun	12		13.64	19.54	11	8.73	18.80	9	6.62	12.31	10	5.03	16.48
particle	13	13	14.77	8.62	29	23.02	8.43	37	27.21	2.46	46	23.31	8.89
preposition	7	8	7.95	4.60	16	12.70	5.30	16	11.76	1.26	25	12.56	4.20
pronoun	5		5.68	4.60	4	3.17	3.37	3	2.21	4.38	4	2.01	2.58
verb	11		12.50	25.29	11	8.73	25.06	11	8.09	25.29	18	9.04	28.27

High-frequency lemmas

If the *hapax* are the head of the rank/frequency distribution of lemmas, the most frequent lemmas are its tail. The 40 most frequent lemmas of the four works are listed in the tables below. The lemma "o" is the Greek article, "A" the lemma "denotative letter;" the colour shadings identify the POS.[42] We set out the data of *Elem*. first:

[42] The lemmas "article" and "denotative letter" are excluded from shading because they always occupy the first places; the adverbs are also unshaded. For the correspondence colour-POS, see the first table after this sequence of four.

lemma	occ.	%	lemma	occ.	%	lemma	occ.	%	lemma	occ.	%
ο	11446	21.64	ευθεια	623	1.18	μεν	350	0.66	οτι	248	0.47
Α	9584	18.12	απο	611	1.16	ορθος	338	0.64	δεω	247	0.47
ειμι	2622	4.96	αυτος	583	1.10	γαρ	320	0.60	οσπερ	236	0.45
και	2190	4.14	τριγωνος	570	1.08	ουτως	313	0.59	λοιπος	228	0.43
ισος	1899	3.59	κυκλος	483	0.91	πλευρα	285	0.54	εαν	228	0.43
υπο	1350	2.55	δυο	416	0.79	σημειον	269	0.51	εκατερος	227	0.43
προς	1174	2.22	επι	389	0.74	δη	264	0.50	διδωμι	223	0.42
αρα	1154	2.18	ως	377	0.71	δεικνυμι	263	0.50	βασις	217	0.41
δε	858	1.62	επει	359	0.68	εχω	260	0.49	κεντρον	214	0.40
γωνια	781	1.48	μειζων	352	0.67	τετραγωνος	250	0.47	ελασσων	208	0.39
tot.	33058	62.50	tot.	37821	71.50	tot.	40733	77.01	tot.	43009	81.31

Then the data of *Gr.-Lat.*:[43]

lemma	occ.	%	lemma	occ.	%	lemma	occ.	%	lemma	occ.	%
A	9620	20.88	is	896	1.95	punctum	336	0.73	oporteo	232	0.50
sum	3041	6.60	quoniam	618	1.34	quidem	330	0.72	habeo	228	0.50
rectus	2783	6.04	a	579	1.26	autem	324	0.70	do	227	0.49
et	2368	5.14	in	558	1.21	enim	314	0.68	tetragonum	222	0.48
qui	2243	4.87	trigonus	548	1.19	si	312	0.68	ipse	221	0.48
aequalis	1795	3.90	vero	485	1.05	ita	307	0.67	multiplex	210	0.46
ergo	1465	3.18	quantitas	480	1.04	idem	290	0.63	reliquus	209	0.45
ad	1136	2.47	circulus	471	1.02	ut	290	0.63	uterque	209	0.45
angulus	933	2.03	duo	464	1.01	latus	268	0.58	basis	206	0.45
sub	908	1.97	maior	373	0.81	quam	254	0.55	parallelo-grammum	202	0.44
									dico	202	
tot.	26292	57.07	tot.	31764	68.95	tot.	34789	75.52	tot.	36955	80.22

[43] An interesting quirk introduces a bias in the Latin data: Latin *rectus* (adj.) corresponds to both ευθεια (noun) and ορθος (adj.). See note 49 below.

Then the data of *Comm.*:

lemma	occ.	%	lemma	occ.	%	lemma	occ.	%	lemma	occ.	%
A	9841	21.78	triangulum	622	1.38	habeo	366	0.81	minor	265	0.59
sum	3182	7.04	igitur	611	1.35	quoniam	356	0.79	a	258	0.57
et	2000	4.43	linea	611	1.35	latus	354	0.78	is	256	0.57
aequalis	1912	4.23	circulus	484	1.07	si	348	0.77	reliquus	251	0.56
ad	1507	3.34	ex	442	0.98	idem	347	0.77	proportio	245	0.54
angulus	1347	2.98	maior	432	0.96	ita	342	0.76	oporteo	241	0.53
ipse	1118	2.47	quadratus	419	0.93	ergo	310	0.69	parallelogram-mum	240	0.53
qui	1097	2.43	duo	406	0.90	quam	302	0.67	do	230	0.51
rectus	941	2.08	ut	395	0.87	enim	287	0.64	vero	224	0.50
in	658	1.46	autem	387	0.86	punctum	280	0.62	basis	222	0.49
tot.	23603	52.24	tot.	28412	62.89	tot.	31704	70.18	tot.	34136	75.56

And finally the data of *Clavius*:

lemma	occ.	%	lemma	occ.	%	lemma	occ.	%	lemma	occ.	%
A	10132	21.39	ut	554	1.17	circulus	379	0.80	centrum	249	0.53
sum	3516	7.42	duo	517	1.09	linea	376	0.79	totus	247	0.52
aequalis	1651	3.49	maior	504	1.06	cum	329	0.69	minor	241	0.51
rectus	1513	3.19	latus	486	1.03	ita	316	0.67	hic	231	0.49
ad	1454	3.07	idem	478	1.01	-que	315	0.66	duco	231	0.49
et	1419	3.00	ipse	457	0.96	ex	296	0.62	autem	231	0.49
angulus	1308	2.76	igitur	454	0.96	dico	293	0.62	inter	230	0.49
qui	961	2.03	si	413	0.87	multiplex	264	0.56	parallelogram-mum	229	0.48
in	723	1.53	quadratus	407	0.86	enim	262	0.55	do	219	0.46
triangulum	609	1.29	quam	382	0.81	se	261	0.55	proportio	218	0.46
tot.	23286	49.16	tot.	27938	58.98	tot.	31029	65.50	tot.	33355	70.41

Clavius' lexical drift towards low frequency lemmas highlights again his metadiscursive drive. The distribution by POS of the 40 most frequent lemmas bar article and denotative letter shows the increase of the pronominal content in the Latin translations, consequent on the absence of the article (the colour shadings of the POS are the same as in the previous tables):

	Elem.			*Gr.-Lat.*			*Comm.*			*Clavius*		
	#	%		#	%		#	%		#	%	
adjective	7	18.42	25.86	7	17.50	26.02	6	15.38	25.00	7	17.95	20.84
adverb	1	2.63	10.92	1	2.50	12.29	1	2.56	15.00	1	2.56	18.42
noun	9	23.68	19.54	9	22.50	18.80	10	25.64	12.31	9	23.08	16.48
particle	10	26.32	8.62	10	25.00	8.43	10	25.64	2.46	9	23.08	8.89
preposition	4	10.53	4.60	4	10.00	5.30	4	10.26	1.26	4	10.26	4.20
pronoun	2	5.26	4.60	4	10.00	3.37	4	10.26	4.38	5	12.82	2.58
verb	5	13.16	25.29	5	12.50	25.06	4	10.26	25.29	4	10.26	28.27

We may measure the distance between the frequency distributions (lemm.) of hapax, 1-form lemmas, and most frequent lemmas, as given in the previous table and in the two similar tables above, and the rescaled relative frequency of the overall occurrences by using standard distances such as Eder's Simple and Canberra.[44] The results are as in the table below; the *hapax* are a sample more representative of the overall distribution than the other two:

	Elem.	*Gr.-Lat.*	*Comm.*	*Clavius*	
hapax	2.766	3.163	3.013	3.717	*Eder*
	1.116	1.340	1.165	1.590	*Canb*
1-form lemmas	7.023	10.186	12.896	11.435	*Eder*
	1.801	2.456	3.476	2.933	*Canb*
40 m. f. lemmas	7.767	9.003	13.418	10.829	*Eder*
	2.157	2.580	3.729	2.993	*Canb*

The following table compares the 10 most frequent lemmas of the four works (the colour shadings identify identical lemmas in different works, and also see the Sankey diagram displayed as Figure 3 of the Appendix):

[44] See note 65 below for these distances.

Elem.			*Gr.-Lat.*			*Comm.*			*Clavius*		
lemma	occ.	%	lemma	occ.	%	lemma	occ.	%	lemma	occ.	%
ο	11446	21.64	A	9620	20.88	A	9841	21.78	A	10132	21.39
A	9584	18.12	sum	3041	6.60	sum	3182	7.04	sum	3516	7.42
ειμι	2622	4.96	rectus	2783	6.04	et	2000	4.43	aequalis	1651	3.49
και	2190	4.14	et	2368	5.14	aequalis	1912	4.23	rectus	1513	3.19
ισος	1899	3.59	qui	2243	4.87	ad	1507	3.34	ad	1454	3.07
υπο	1350	2.55	aequalis	1795	3.90	angulus	1347	2.98	et	1419	3.00
προς	1174	2.22	ergo	1465	3.18	ipse	1118	2.47	angulus	1308	2.76
αρα	1154	2.18	ad	1136	2.47	qui	1097	2.43	qui	961	2.03
δε	858	1.62	angulus	933	2.03	rectus	941	2.08	in	723	1.53
γωνια	781	1.48	sub	908	1.97	in	658	1.46	triangulum	609	1.29
tot.	33058	62.50	tot.	26292	57.07	tot.	23603	52.24	tot.	23286	49.16

This is a fairly good index of lexical dispersion. Three general facts must be taken into account when reading this data. First, and obviously, the absence of the article in Latin accounts for at least 8000 missing occurrences, if we want to be generous with allowing the article to be replaced by pronouns in some instances (lemmas in red; *Clavius* did not try to translate the Greek article). This explains the flattening of the distribution of the occurrences in all Latin translations:[45] 39.76% of the lexical content of *Elem.* is just made of articles and denotative letters. Second, the constant presence of the article and of particles and prepositions is a structural trait of the Greek language, which is partly lost in Latin: one counts (6,4,3,3) such items among the 10 most frequent lemmas of the four works. The relative frequency flow between the 20 most frequent lemmas is represented by the Sankey diagram displayed as Figure 3 of the Appendix, where the height of the coloured rectangles is proportional to a suitable rounding of the relative frequency.[46] The loops are attached to lemmas whose position in the ranking of the other works is lower than the 20th place. Likewise, contributions to non-looping lemmas in the diagram that come from lemmas lower than the 20th place are neglected; this problem arises conspicuously for the Greek particles δε and αρα, and to a lesser extent for και and προς. The fictitious boxes "new occurrences" and "lost occurrences" take also account of the fact that the relative frequencies are not normalised to a common scale.

45 For comparison's sake, the 10 most frequent lemmas of Archimedes' *De sphaera et cylindro* and of Pappus' *Collectio* represent 57.65% and 60.21% of their lexical content, respectively.

46 The temporal ordering of the Latin works adopted in the table and in the chart is lexically irrelevant, for they are all independent editions or translations of *Elem.* So, the Sankey does not depict a real frequency flow between the Latin works, but the rearrangement of the ranking.

Form spread

As Greek and Latin are inflectional languages, it is interesting to study the form spread within one and the same lemma. The 26 lemmas of *Elem*. that have at least 10 forms are as follows (see below for the colour shadings):

lemma	# forms	lemma	# forms	lemma	# forms
A	511	επιζευγνυμι	16	συνιστημι	13
ειμι	33	εφαπτομαι	16	λαμβανω	11
τεμνω	23	λοιπος	15	μειζων	11
αυτος	21	ος	15	ορθος	11
ο	19	διδωμι	14	πλευρα	11
ελασσων	18	ισος	14	αφαιρεω	10
εχω	18	αγω	13	εγγραφω	10
περιεχω	18	ποιεω	13	ευθεια	10
αναγραφω	16	προσπιπτω	13		

The 17 lemmas of *Gr.-Lat.* and of *Comm.* that have at least 10 forms are as follows:

Gr.-Lat.

lemma	# forms	lemma	# forms
A	526	is	12
seco	16	rectus	12
habeo	14	facio	11
contineo	13	ipse	11
sum	13	qui	11
contingo	12	coapto	10
describo	12	constituo	10
divido	12	sumo	10
idem	12		

Comm.

lemma	# forms	lemma	# forms
A	527	sum	13
seco	21	constituo	12
describo	17	contingo	12
contineo	15	habeo	12
qui	15	rectus	12
is	14	duco	11
sumo	14	reliquus	11
idem	13	uterque	11
ipse	13		

The 41 lemmas of *Clavius* that have at least 10 forms are as follows:[47]

lemma	# forms	lemma	# forms	lemma	# forms	lemma	# forms
A	514	hic (pron.)	16	qui	14	efficio	11
divido	26	idem	16	alius	13	exsisto	11
seco	23	sumo	16	comprehendo	13	quicumque	11
counter	22	describo	15	compono	12	reliquus	11
sum	20	habeo	15	ille	12	applico	10
constituo	19	is	15	ipse	12	cado	10
aufero	17	ostendo	15	oppono	12	consequor	10
tango	17	subtendo	15	pono	12	do	10
contineo	16	dico	14	rectus	12	produco	10
duco	16	facio	14	demonstro	11	totus	10
						uterque	10

One must not forget that accents and elisions tend to increase the number of forms: 5 lemmas of *Elem.* enter the ranking thanks to this feature of the Greek language: αφαιρεω, ευθεια, ορθος, μειζων, πλευρα.

The colour shadings identify lemmas that are identical or such that the Latin lemma is a translation of the Greek lemma; the colours in the column of the number of forms mark two lemmas that translate a single lemma: αυτος > *idem*, *ipse*; εφαπτομαι > *contingo*, *tango*; περιεχω > *comprehendo*, *contineo*; τεμνω > *divido*, *seco*; ευθεια, ορθος < *rectus*. Both uniquely related pairs of Greek and Latin lemmas and each of these complexes will be called "metalemmas;" more on them below. With this qualification, there are 12 core metalemmas common to the four texts, (0,1,2,1) metalemmas common to three texts, (0,0,2,0,0,1) common to two texts only.[48] *Elem.* has 8 unique lemmas with at least 10 forms (30.77%), *Gr.-Lat.*, 1 (5.88%), *Comm.*, 0 (0.00%), *Clavius*, 19 (46.34%). Clavius' rhetorical education is apparent from this data: lexical redundancy, form spread.

It comes as no surprise that the 12 core metalemmas comprise the denotative letters, the noun and the adjective for "straight," "right" (ευθεια, ορθος / *rectus*),[49] the most com-

47 The lemma "counter" collects the ordinals that feature in the headings of principles, propositions, and Books, which are part of the text in *Clavius*.

48 The entries in each parenthesis give the numbers that correspond, in a suitable order, to the 4 combinations of 3 objects out of 4 and to the 6 combinations of 2 objects out of 4, respectively.

49 In principle, the forms of the lemma ευθεια (noun) should be assigned to the adjective ευθυς, which, however, is scarcely used. In Latin, *recta* etc. are indeed forms of *rectus*, which is frequently used in the masculine for the "right" angle. So, we decided to separate ευθεια from ευθυς but not *recta* from *rectus*. Otherwise, we would have lost the noun of the most important math-

mon pronouns (αυτος / *idem, ipse*; ος / *qui*), 5 verbs with an obvious geometric connotation (αναγραφω / *describo*; εφαπτομαι / *contingo, tango*; λαμβανω / *sumo*; περιεχω / *comprehendo, contineo*; συνιστημι / *constituo*; τεμνω / *divido, seco*), and the all-purpose verbs "to be" and "to have."

We represent the form spread of the core metalemmas bar denotative and numeral letters in the charts displayed as Figures 4a–b of the Appendix, where the core metalemmas are located in the tick marks of the abscissae and the numbers of forms have been normalised to the total number of forms in each text;[50] we took the arithmetic mean of the form numbers of the two lemmas that translate a single lemma, if both of the former are present in the list. In chart 4a, the order of the metalemmas in *Elem.* is also adopted for the other works; this highlights an initial peak of the Latin translations (form spread of *sum* much lower than the form spread of ειμι), and the oscillations of *Comm.* and *Clavius*. In chart 4b, the order is by decreasing number of forms in each text separately; the form spread of *Elem.* and the striking similarity of the patterns of *Comm.* and *Clavius* are apparent.

Let us also compare the above overall data:

	≥ 10 forms	% lemmas	occ.	tot. occ.
Elem.	26	7.47	30046	52894
Gr.-Lat.	17	4.10	20086	46067
Comm.	17	4.47	18640	45176
Clavius	41	6.62	21109	47371

These parameters are an index of lexical concentration: the same lemma is used in several syntactic situations. The occurrences engaged by the lemmas with the highest number of forms are linearly correlated with the size (occ.) of the text. The correlation is very good: the best linear regression of the data, with slope 0.6703, has a coefficient of determination $R^2 = 0.9936$. Of course, there is a natural upper bound to the number of forms a lemma can display in a sectorial lexicon: mainly pronouns or adjectives and, above all, verbs naturally admit a high number of forms – yet the limited number of semantic fields accessible to a sectorial lexicon introduces a bias in the number of available POS.

Distribution of lemmas, forms, and occurrences by POS

The following table sets out the distribution of the occurrences, in our reference texts, among categories that by and large coincide with the standard POS. As said, specific items

ematical object of Greek mathematics, or categorised (Latin) two crucial mathematical objects in an inconsistent way.

50 That is, for ειμι / *sum*, (33, 13, 13, 20) > (33⁄1774, 13⁄1743, 13⁄1665, 20⁄2400) = (0.0186, 0.0075, 0.0078, 0.0083).

as numerals, denotative letters, and Greek words are each assigned a separate category. The column shaded yellow gives the frequencies of *Elem.* rescaled as if the text did not contain any article.

	Elem.			*Gr.-Lat.*		*Comm.*		*Clavius*	
	occ.	%		occ.	%	occ.	%	occ.	%
adjective	6308	11.93	15.23	8631	18.74	7548	16.71	8683	18.33
adverb	1313	2.48	3.15	1179	2.56	1255	2.78	1911	4.03
article	11446	21.64	0.00	0	0.00	0	0.00	0	0.00
denot. lett.	9584	18.12	23.12	9620	20.88	9841	21.78	10132	21.39
Greek	0	0.00	0.00	8	0.02	0	0.00	0	0.00
noun	5089	9.62	12.28	4957	10.76	6188	13.70	6382	13.47
number	0	0.00	0.00	10	0.02	0	0.00	84	0.18
particle	7385	13.96	17.82	7515	16.31	6640	14.70	6110	12.90
preposition	4416	8.35	10.66	3903	8.47	3399	7.52	3774	7.97
pronoun	1248	2.36	3.01	3875	8.41	3143	6.96	2713	5.73
verb	6105	11.54	14.73	6369	13.83	7162	15.85	7582	16.01

Even compared with the rescaled relative frequencies of *Elem.*, it is apparent that the pronominal content of the Latin texts increases as a partial counterbalance to the absence of the definite article. The lower relative frequency (occ.) of particles in Latin texts points to a loosening of the deductive structure more than to structural facts underlying the two languages, for the relative frequencies (lemm.) of this POS are nearly identical in the four works, see the tables just below. The relative frequency data of the above table is displayed in the chart displayed as Figure 5 of the Appendix, where the lemmas "Greek" and "number" have been eliminated.

Let us finally set out a complete table of lemmas, forms, and occurrences distributed among POS. The reader is urged to spend some time comparing the numbers in the tables. The first table sets out the data of *Elem.*:

	lemmas	%	forms	%	occ.	%	% l/o	% f/o	% l/f
adjective	90	25.86	336	18.95	6308	11.93	1.43	5.33	26.79
adverb	38	10.92	44	2.48	1313	2.48	2.89	3.35	86.36
article	1	0.29	19	1.07	11446	21.64	0.01	0.17	5.26
denot. lett.	1	0.29	511	28.82	9584	18.12	0.01	5.33	0.20
noun	68	19.54	235	13.25	5089	9.62	1.34	4.62	28.94

particle	30	8.62	48	2.71	7385	13.96	0.41	0.65	62.50
preposition	16	4.60	32	1.80	4416	8.35	0.36	0.72	50.00
pronoun	16	4.60	68	3.84	1248	2.36	1.28	5.45	23.53
verb	88	25.29	480	27.07	6105	11.54	1.44	7.86	18.33
total	348	100.00	1773	100.00	52894	100.00	0.66	3.35	19.63

The same table for *Gr.-Lat*:

	lemmas	%	forms	%	occ.	%	% l/o	% f/o	% l/f
adjective	108	26.02	337	19.33	8631	18.74	1.25	3.90	32.05
adverb	51	12.29	52	2.98	1179	2.56	4.33	4.41	98.08
denot. lett.	1	0.24	526	30.18	9620	20.88	0.01	5.47	0.19
Greek	1	0.24	7	0.40	8	0.02	12.50	87.50	14.29
noun	78	18.80	237	13.60	4957	10.76	1.57	4.78	32.91
number	1	0.24	7	0.40	10	0.02	10.00	70.00	14.29
particle	35	8.43	35	2.01	7515	16.31	0.47	0.47	100.00
preposition	22	5.30	24	1.38	3903	8.47	0.56	0.61	91.67
pronoun	14	3.37	66	3.79	3875	8.41	0.36	1.70	21.21
verb	104	25.06	452	25.93	6369	13.83	1.63	7.10	23.01
total	415	100.00	1743	100.00	46067	100.00	0.90	3.78	23.81

The same table for *Comm.*:

	lemmas	%	forms	%	occ.	%	% l/o	% f/o	% l/f
adjective	95	25.00	320	19.22	7548	16.71	1.26	4.24	29.69
adverb	57	15.00	57	3.42	1255	2.78	4.54	4.54	100.00
denot. lett.	1	0.26	527	31.65	9841	21.78	0.01	5.36	0.19
noun	66	17.37	205	12.31	6188	13.70	1.07	3.31	32.20
particle	40	10.53	41	2.46	6640	14.70	0.60	0.62	97.56
preposition	19	5.00	21	1.26	3399	7.52	0.56	0.62	90.48
pronoun	12	3.16	73	4.38	3143	6.96	0.38	2.32	16.44
verb	90	23.68	421	25.29	7162	15.85	1.26	5.88	21.38
total	380	100.00	1665	100.00	45176	100.00	0.84	3.69	22.82

The same table for *Clavius*:

	lemmas	%	forms	%	occ.	%	% l/o	% f/o	% l/f
adjective	129	20.84	442	18.42	8683	18.33	1.49	5.09	29.19
adverb	114	18.42	115	4.79	1911	4.03	5.97	6.02	99.13
denot. lett.	1	0.16	514	21.42	10132	21.39	0.01	5.07	0.19
noun	102	16.48	313	13.04	6382	13.47	1.60	4.90	32.59
number	1	0.16	22	0.92	84	0.18	1.19	26.19	4.55
particle	55	8.89	57	2.38	6110	12.90	0.90	0.93	96.49
preposition	26	4.20	28	1.17	3774	7.97	0.69	0.74	92.86
pronoun	16	2.58	107	4.46	2713	5.73	0.59	3.94	14.95
verb	175	28.27	802	33.42	7582	16.01	2.31	10.58	21.82
total	619	100.00	2400	100.00	47371	100.00	1.31	5.07	25.79

Information content by categorial type: X-entropy

We have also introduced an intensive indicator, which measures how much lexical information of a specific kind is provided in a given text or in any of its relevant subdivisions. In order to do this, we have categorised each lemma by means of a three-tier typology that is a refinement of the standard partition into POS. The broadest categories are the POS as given above, the finer ones are partly semantically oriented.[51] From this categorisation we have extracted a cross-section of categorial types,[52] which are as follows (capital letters refer to the sigla in the charts displayed in the Appendix): INFerential and coordinant (among particles), METadiscursive (among any POS), OPErational (among nouns and verbs), RELational (among nouns and verbs), NOMinal (all nouns + pronouns), DENotative letters, QUAlifiers (among adjectives), GENerality determiners (among adjectives). To exemplify the notions we are going to introduce, the computations in this and the following subsection mainly refer to the textual unit made by a Book.

In each proposition of each Book of the four works, we counted all the occurrences related to the lemmas that belong to any given categorial type among those in the list above, and we normalised to the total number of words in the assigned proposition. For

51 An almost complete list of these sub-categories is given in Fabio Acerbi and Ramon Masià, *George of Trebizond, Introduction to the Almagest. Greek and Latin Text* (Leiden: Brill, 2026) Introduction, sect. 4.

52 The selected cross-section does not realise a partition of the lemmas: the categorial types given below may not be exclusive, and do not cover all lemmas (for instance, the prepositions are not included).

any given categorial type X and any given proposition *i*, we further normalised the output of the previous step to the sum of the outputs associated with all the propositions of the given Book. For each Book *j*, the resulting numbers p_{ij}^{X} are a probability distribution, for their sum is 1 by definition. Consequently, we define the intensive indicator X-entropy of the categorial type X in the assigned Book as $H_j^X \equiv -\Sigma_i p_{ij}^X \log p_{ij}^X$.[53] The results are as in the X-entropy vs. Book charts displayed as Figures 6a–h of the Appendix.

For all categorial types and almost all Books, the X-entropy confirms the wider lexical spread of *Clavius* (an exception is GEN in Book II, for *Clavius* omits all the generalising determiners that qualify the way a given straight line is cut into segments) while highlighting the compartmentation of the lexicon by Book and the monothematic character of Books II and IV. The data of *Elem.* and those of *Gr.-Lat.* are normally hard to distinguish; note, however, the discrepancy as for NOM: *Gr.-Lat.* may have a lower entropy because it tends to "saturate" the nominal lexicon by frequently supplying nouns like *recta*, *angulus*, *ratio*.

Location and density of categorial types: X-clouds

Within the word space of a mathematical proposition, word-tokens belonging to specific POS or to categorial types as those defined in the previous sub-Section may tend to cluster or, conversely, to display a wide spread. We measured the location and density of some categorial types by introducing a four-leg indicator, which we call "X-cloud." This indicator is constructed as follows.

Compute the position, in number of word-tokens (the first word has position 1, the second has position 2, etc.), of each of the occurrences of lemmas belonging to an assigned categorial type X. Divide each of these positions by the number of words contained in the assigned proposition: these are the "normalised positions" (hence, the last word has normalised position 1). Take the arithmetic mean of these normalised positions: this is the "X-centroid" of the assigned proposition. Take the minimum of the normalised positions: this is the "X-beginning" of the assigned proposition. Take the maximum of the normalised positions, and subtract from it the X-beginning as computed in the previous step: this is the "X-range" of the assigned proposition.[54] Count the number of occurrences of lemmas belonging to the categorial type X in the assigned proposition. Divide this number by the overall number of words counted from the first to the last occurrence (inclusive of the extremes) of lemmas belonging to the categorial type X in the assigned proposition: this is the "X-density" within the X-range of the assigned proposition.

53 In our entropy computations here and below, we use the base 10 logarithm.

54 Thus, the X-range of a unique occurrence belonging to categorial type X is 0. On the one hand, this makes the definition of "X-density" cumbersome, and yields X-range $\neq 1$ for a proposition that begins and ends with a word-token belonging to categorial type X. On the other hand, including the X-beginning in the range would have made the X-range of an X-singleton depend on the size of the proposition, which we deemed a drawback more annoying than the one just pointed out.

All these proposition indicators can be lifted to a Book or to a set of Books by taking suitable means. The bookwise behaviour of the four indicators for the categorial types X = DEN, INF, NOM is plotted in the charts displayed as Figures 7a–l of the Appendix.

We also selected the categorial types X = INF, NOM and plotted three of the above indicators propositionwise (the propositions are numbered sequentially from prop. 1 of Book I) in the charts displayed as Figures 8a–f of the Appendix; the gaps in the interpolating lines correspond to autonomous deductive units – mainly corollaries and additional propositions – that do not figure in the associated text. The mean values for the entire segment of Books I-VI are as follows:

	INF-cen	INF-beg	INF-ran	NOM-cen	NOM-beg	NOM-ran
Elem.	0.5211	0.1159	0.7526	0.5087	0.0557	0.9280
Gr.-Lat.	0.5490	0.1246	0.7900	0.5030	0.0478	0.9371
Comm.	0.5244	0.1527	0.6977	0.4985	0.0524	0.9054
Clavius	0.5830	0.2574	0.6325	0.4995	0.0622	0.9110

The first three charts show that *Clavius* starts deducing later and goes on doing this less long. The second three charts highlight the considerable extension of the NOM-range; the systematic exception is Book V, where adjectives normally replace nouns in the enunciations and in the setting-outs.

Correlation table and correlation and dispersion matrix of a translation

The difficulty of comparing the lexicon of a Greek and a Latin text that are one the translation of the other lies in the fact that there is no natural one-to-one correlation between Greek and Latin lexemes, and that many possible correlations are context-dependent. Of course, there are obvious instances of bijective correlation (think of the adjective “two”), and their number increases in technical texts, which tend to make the lexicon univocal.[55] However, a number of facts blurs the correlation:

– A one-to-one correlation is lost between many categories of lexemes just because the Latin translators did not replicate the rigid conventions of the Greek stylistic code. Take coordinant and inferential particles: αρα and δη, whose use is strictly differentiated in Greek, more or less randomly correspond to *ergo, idcirco, ideo, quare, quocirca* (just check the proof of Prop. I.1 in *Clavius*); δε can be *autem, vero, verum*; γαρ can be *enim, nam, namque, quia*; και and τε can be *et, ac, atque,* enclitic *que,* adverbial και can be *etiam* and *quoque*.

55 Technical texts written in natural language normally fall short of this requirement. The *Elements* are no exception: see the discussion in Acerbi, *The Logical Syntax*, 35–36.

- Verbs with two preverbal prepositions are common in Greek and exceedingly rare in Latin: the only specimens in *Gr.-Lat.* are *adimpleo* and *adinvenio*, the latter is the only specimen in *Clavius*; no specimen in *Comm*. On the other hand, if *Gr.-Lat.* and *Clavius* have both *invenio* and *adinvenio*, this occurs because *Elem.* has both ευρισκω and προσευρισκω. This example shows that the dilution of the technical lexicon already took place in the Greek language.
- A lemma in Greek may correspond to a form in Latin; a case in point are the verbal adjectives (Greek), which correspond to a verb form (Latin). Likewise, some correlated lemmas may belong to different grammatical categories, for instance, adjective vs. noun.
- Most of *Clavius*'s Euclid is not a translation, so strictly speaking we cannot compare its lexicon with the lexicon of *Elem.*: we might more properly state that we draw a comparison between Gracilis' partial translation and *Elem*. It remains that the setting-outs in *Clavius* reproduce most of Gracilis' lexicon – and we are interested in the evolution of the lexical record as a witness to the evolution of the way Greek mathematics was appropriated and naturalised.

Paying due attention to these features, we have separately established univocal and multivocal correlations between lemmas of *Elem.* and lemmas of *Gr.-Lat.*, *Comm.*, and *Clavius*. The lemmas and occurrences (overall numbers and percentage with respect to the entire lexical content) engaged for each of the three combinations Greek/Latin are given in the following tables.[56] *Clavius*' huge lexical spread is obvious, for it employs a lesser amount of its own lexicon in the "translation."

# lem.	*Gr.-Lat.*	*Comm.*	*Clavius*	# occ.	*Gr.-Lat.*	*Comm.*	*Clavius*
	380	342	328		45820	45034	45943
Elem.	334	323	294	*Elem.*	41254	41083	41078

% lem.	*Gr.-Lat.*	*Comm.*	*Clavius*	% occ.	*Gr.-Lat.*	*Comm.*	*Clavius*
	91.57	89.74	52.99		99.46	99.69	96.99
Elem.	95.98	92.82	84.48	*Elem.*	77.99	77.67	77.66

[56] This means, for instance, that *Gr.-Lat.* uses 380 lemmas to translate 334 lemmas of *Elem.*, and the occurrences these lemmas engage in the two works are 45820 and 41254, respectively. The second table gives the percentages. At least 96.87% of the percentage (occ.) loss of *Elem.* with respect to the Latin translations comes from the absence of the article.

As several Latin lemmas may translate one and the same Greek lemma,[57] we can compute how many Latin lemmas are ranked in a given position in the sequences that translate one and the same Greek lemma.[58] The distribution, which is of course independent of the actual ordering of any such sequence of Latin lemmas, exhibits a typical Pareto-law-like behaviour, see Figure 9 of the Appendix.

Complex networks of correlation normally come to be established in translation for some key categories of lemmas, often comprising function words. For instance, the "forward" Greek inferential particles give rise in *Clavius* to the following *correlation matrix*, which includes all (source) Greek lemmas that are translated by a given maximal set of (target) Latin lemmas.[59]

δε	*autem*	*vero*			
αρα	*igitur*	*ergo*	*quocirca*	*quare*	*itaque*
δη	*igitur*	*autem*			
ουν	*itaque*				
μην	*vero*				

A further example comes from *Gr. Lat.*: a few constellations of verbs that are naturally related – namely, the verbs meaning tangency, motion, generation and intersection of lines – are also linked to the indefinite adjective οποιοσουν "whichever" by the several translations of the determiners of arbitrariness involving forms of the verb τυγχανω "to happen."

57 At most 6 Latin lemmas in the case of *Gr.-Lat.*, 9 for *Comm.*, 7 for *Clavius*. The percentages of lemmas in *Gr.-Lat.*, *Comm.*, and *Clavius* that correspond to one single lemma in *Elem.* are 58.55, 65.91, and 33.17, respectively. This confirms that *Clavius*' lexical richness is mainly a consequence of its rewriting the proofs.

58 In the first correlation table below, these numbers are 5, 3, 1, 1, 1.

59 That is, if a Greek lemma in the correlation matrix is translated by a number of Latin lemmas, all these Latin lemmas and all Greek lemmas they translate are also included in the matrix. The correlation matrices are easily constructed by picking a Greek lemma and applying the definition recursively.

τυγχανω	*accido*	*fortuitus*	*quicumque*	*quilibet*	*contingo*	*quomodolibet*
προσπιπτω	*accido*	*venio*				
προσαναγραφω	*adscribo*					
παραβαλλω	*adscribo*	*constituo*	*describo*	*educo*		
βαινω	*consisto*	*constituo*	*insido*			
συνιστημι	*consisto*	*constituo*	*sisto*			
απτομαι	*contingo*					
εφαπτομαι	*contingo*					
αναγραφω	*describo*	*adscribo*				
αγω	*duco*	*protraho*				
διαγω	*educo*	*protraho*	*produco*			
εκβαλλω	*emitto*	*educo*	*ejicio*			
προσεκβαλλω	*produco*	*educo*				
ερχομαι	*transeo*	*venio*				
οποιοσουν	*unusquislibet*	*quicumque*				
ηκω	*venio*					

Setting all target lemmas as headers of a suitable number of columns and reordering the rows and the columns, a correlation matrix can be regularised to give a perspicuous representation of the correlation between the source and target lemmas, and thereby as it were "diagonalised," where by "diagonal" we mean here the cell diagonal coming down from the upper left corner. Depending on the distribution of the target lemmas, there may not exist an ordering of the source and target lemmas that allows filling the entire cell diagonal. In the present instance, only 15 of the 16 possible cells can be filled. The reason is that the two verbs απτομαι and εφαπτομαι are translated by one and the same lemma, namely, *contingo*. A diagonalised form of the just-presented correlation matrix – which in general is not unique – is set out below.

	fortuitus	*accido*	*adscribo*	*constituo*	*insido*	*sisto*	*contingo*	*describo*	*duco*	*protraho*	*emitto*	*produco*	*transeo*	*unusquislibet*	*venio*	*quicumque*	*quilibet*	*quomodolibet*	*consisto*	*educo*	*ejicio*
τυγχανω	*fortuitus*	*accido*					*contingo*									*quicumque*	*quilibet*	*quomodolibet*			
προσπιπτω		*accido*													*venio*						
προσαναγραφω			*adscribo*																		
παραβαλλω			*adscribo*	*constituo*				*describo*												*educo*	
βαινω				*constituo*	*insido*														*consisto*		
συνιστημι				*constituo*		*sisto*													*consisto*		
απτομαι							*contingo*														
αναγραφω			*adscribo*					*describo*													
αγω									*duco*	*protraho*											
διαγω										*protraho*		*produco*								*educo*	
εκβαλλω											*emitto*									*educo*	*ejicio*
προσεκβαλλω												*produco*								*educo*	
ερχομαι													*transeo*		*venio*						
οποιοσουν														*unusquislibet*		*quicumque*					
ηκω															*venio*						
εφαπτομαι							*contingo*														

The correlations between Greek and Latin lemmas can be classified as in the *correlation table* below, where numerical pairs such as n/m mean that n different lemmas of *Elem.* are correlated with m different lemmas of the intended Latin translation. The first row gives the total number of lemmas involved in a given kind of correlation. The second row gives the kind of the correlation; we shall give this correlation a name later. The third row gives the scaling factor between the pairs located in the first two rows, namely, the number of the correlations of a given kind. For instance, from the second column of the first table one gets that *Gr.-Lat.* contains 19 correlations (third row) in which 2 Greek lemmas are translated by 1 Latin lemma (second row), hence these 19 correlations involve 38 Greek lemmas and 19 Latin lemmas (first row). To any of the pairs in the second row there corresponds a correlation matrix of the kind described above. We set out the correlation table for *Gr.-Lat.* first:

161/161	38/19	6/2	4/1	32/64	3/9	1/4	1/5	30/30	12/18	4/8	6/4	3/3	3/4	3/6	5/4	6/10	16/21
1/1	2/1	3/1	4/1	1/2	1/3	1/4	1/5	2/2	2/3	2/4	3/2	3/3	3/4	3/6	5/4	6/10	16/21
161	19	2	1	32	3	1	1	15	6	2	2	1	1	1	1	1	1

The same table for *Comm.*:

154/154	34/17	15/5	17/34	5/15	18/18	2/3	4/8	3/3	3/4	3/5	4/6	5/2	4/10	5/7	7/16	8/7	10/14	22/22
1/1	2/1	3/1	1/2	1/3	2/2	2/3	2/4	3/3	3/4	3/5	4/6	5/2	4/10	5/7	7/16	8/7	10/14	22/22
154	17	5	17	5	9	1	2	1	1	1	1	1	1	1	1	1	1	1

The same table for *Clavius*:

151/151	26/13	6/2	15/30	3/9	1/4	1/5	1/6	12/12	4/6	2/4	3/2	3/3	3/4	3/5	4/2	4/8	5/2	5/4	5/7	7/9	7/10	9/10	14/20
1/1	2/1	3/1	1/2	1/3	1/4	1/5	1/6	2/2	2/3	2/4	3/2	3/3	3/4	3/5	4/2	4/8	5/2	5/4	5/7	7/9	7/10	9/10	14/20
151	13	2	15	3	1	1	1	6	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1

We may visualise the data set out in a correlation table by introducing an $r \times s$-cell *dispersion matrix D*, where r and s are the numbers of lemmas of the texts R and S that are involved in the translation, respectively. In our case, R = *Elem.* and S = *Gr.-Lat.*, *Comm.*, or *Clavius*. The lemmas index the rows and the columns of the dispersion matrix according to their order in the associated correlation table; the right lower submatrix is occupied by lemmas that do not have a correlate. A cell D_{ij} of the dispersion matrix is filled by lemma j if this is the translation of lemma i. As seen above, given a maximal set of Latin lemmas that translate a set of Greek lemmas, their ordering can be arranged in such a way as to maximise the number of cells filled along any block diagonal. The result is a block matrix of sorts,[60] whose larger blocks coincide with the diagonalised correlation matrices introduced above. The results are displayed as Figure 10 of the Appendix; the black cells contain words, the cells in colour are empty and contribute to identifying different diagonalised correlation matrices. The order from top to bottom is *Gr.-Lat.*, *Comm*, *Clavius*; zoom in to see details.

We measure the lemmatic reorganisation that has occurred in a given translation by introducing five indicators that process the data set out in the correlation table and in the dispersion matrix. The idea is to measure the extent to which the lexicon is reorganised by the interplay of lexical sinks (several source lemmas are translated by fewer target lemmas) and sources (the opposite occurs). The indicators are:

- The "temperature" T_t of the given translation, defined, as is customary in statistical thermodynamics, by means of the Equipartition Theorem, but read the other way around.[61] We select the first row of a correlation table and compute $\delta_i = |n_i - m_i|$ for n_i and m_i in each pair n_i/m_i, where i is the position of the pair in the sequence of the table. We measure the "mean thermal energy" associated with the dynamics of lemmatic diversity occurring in translation by taking the expectation value of the square of δ_i w.r.t. a discrete Maxwell-Boltzmann probability distribution featuring the same square in the mass function:[62] $T_t \equiv \Sigma_i \delta_i^2 exp(-\delta_i^2)$. A bijective translation has $i = 1$ only, $\delta_i = 0$, and hence $T_t = 0$; the temperature of a silly translation in which all source lemmas are translated by one and the same target lemma rapidly converges to 0 when the number of source lemmas increases.[63] A translation that is "hotter" than another has carried out a more extensive lemmatic reorganisation.
- A "translation entropy," $H_t \equiv -\Sigma_i \partial_i \log \partial_i$, where ∂_i is δ_i normalised to $\Sigma_i \delta_i$. Here again, a bijective translation has $H_t = 0$, and a translation having a greater entropy than another also carried out a more extensive lemmatic reorganisation. Because of the small sample size, most of the contribution to the first two indicators comes from small δ_i.
- The "mixing index," which is a bulk indicator of lemmatic reorganisation. We select

60 It is easy to see that the result cannot be a real block matrix.

61 Check for instance Ravinder R. Puri, *Modern Thermodynamics and Statistical Mechanics. A Comprehensive Foundation* (Cham: Springer, 2024), sect. 3.2. The approach in Sasuke Miyazima and Keizo Yamamoto, "Measuring the Temperature of Texts," *Fractals* 16 (2008): 25–32, takes the underlying physics wrong; it just amounts to computing the slope of a trivial reshaping of the log-log graph of Zipf's law.

62 There is no need to introduce constants of normalisation because we set a definition and the temperature scale is arbitrary.

63 A totally fuzzy translation, in which each source lemma is translated by all target lemmas, can be impossible to achieve, depending on the size of the text and on the distribution of the occurrences.

the second row of a correlation table and compute $\Delta_i = |n_i - m_i|$ for n_i and m_i in each pair n_i/m_i, where i is the position of the pair in the sequence of the table. We thus define the mixing index $M_t = (\Sigma_i\Delta_i)/$(average of the numbers of source and target lemmas). Clearly, $0 \leq M_t \leq 1$, the upper bound being approximated by the silly translation and attained by an empty translation. Here again, a bijective translation has $M_t = 0$.

- The normalised value S_t of the lexical span of the given translation as a subset of the dispersion matrix D. The idea is to compute the normalised cell size of the minimal rectangle spanned by an n/m correlation between lemmas. Taking for example *Gr.-Lat.*, this means that the lexical span of a 1/1 correlation is 1 (one single cell is involved) times 161 (the number of such cells), the lexical span of a complex correlation, for instance 16/21, is 16×21 (the rectangle spanned by the lemmas involved in the correlation, whatever their distribution within the rectangle, see the next indicator) times 1 (just one correlation is involved). We normalise to the cell size of the entire dispersion matrix. With reference to the three correlation tables above, set u_i/v_i for a pair in the second row and w_i for the corresponding number in the third row, where i is the position of the pair in the sequence. The normalised lexical span is defined as $S_t \equiv \Sigma_i u_i v_i w_i / rs$. One has $1/_{\min\{r,s\}} \leq S_t \leq 1$.
- As the rectangle spanned by the lemmas involved in a given correlation may be partly empty, we compute the degree of lexical reorganisation F_t of the given translation as the ratio between the sum of the filled cells within all rectangles spanned by all n/m correlations such that both m and $n \neq 1$ and the sum of the cell sizes of the rectangles spanned by the same correlations.[64] Clearly, $0 \leq F_t \leq 1$, where we set $F_t = 1$ whenever there are no correlations such that both m and $n \neq 1$.

The values of these indicators are set out in the following table. *Clavius* is hotter and more mixing than the other translations; within the rectangles spanned by the correlation matrices, its lexicon is nearly as tightly reorganised as in *Comm.*

	T_t	H_t	M_t	S_t	F_t
Gr.-Lat.	0.812	0.958	0.0962	0.006983	0.584
Comm.	1.325	1.015	0.1159	0.011031	0.504
Clavius	1.914	1.181	0.1576	0.009654	0.511

Measuring the textual distances among the four texts: first- and second-order metalemmas

As the (ordered) set of the relative frequencies of the lemmas of any text is a vector, we may measure the distance between two texts by defining a suitable distance between the

64 The excluded correlations cannot give rise to empty cells.

relative frequency vectors associated with them.[65] In order to compute the distance between *Elem*. and any of its Latin translations, we must preliminarily reduce the multivocal correlations between lemmas described in the preceding sub-Section to one-to-one numerical correlations between occurrences of single lemmas. This necessarily entails creating fictitious lemmas, any of which we call (first-order) "metalemma;" the number of metalemmas for any given correlation in the second row of each of the three correlation tables above are those set out in the third row of the same tables. The reduction procedure that gives the occurrences of a metalemma is as follows:

- if the correlation is 1/1, the metalemma coincides with the lemma on either side;
- if the correlation is n/m, with $n > 1$ and $m > 1$, we create the metalemma by adding the occurrences on the Greek side and on the Latin side, respectively.

In this way, a set of Greek-Latin metalemmas is associated with each of the three Latin translations. These three sets of metalemmas need not coincide; this raises a problem, because measuring textual distances requires operating on a stable lexicon. To overcome this problem, we introduced second-order metalemmas, defined as the set-theoretical union, over all translations, of all Greek lemmas that are translated by a given maximal set of Latin lemmas. Clearly, the number of second-order metalemmas is smaller than the number of metalemmas associated with each translation; conversely, there exist huge second-order metalemmas.[66] Note that a given second-order metalemma need not be shared by all translations, but there is – in fact, there must be – a core of shared second-order metalemmas. The outcome of the reduction procedure just outlined are 188 second-order metalemmas, 167 of which are shared by the four works.

Cluster analysis, consensus trees, networks

We first perform a cluster analysis applied to second-order metalemmas using a number of standard distances, even if stylometric tools grounded on textual distances are not tailored to process a corpus comprising four texts only. For some of the distances available in the package `stylo`, the results are charted in dendrograms as the ones displayed as Figures

[65] If the relative frequency vectors of the two texts do not refer to the same set of lemmas, just upgrade them to the set-theoretical sum of the sets of lemmas and put 0 for the frequency of a lemma that is absent in either set. Recall that a distance between two vectors is a non-negative symmetric function of them which is zero iff the two vectors coincide and such that the triangle inequality holds. A summary of the currently used textual distances can be found in Eder, Rybicki, and Kestemont, *'Stylo'*, 15–17. An assessment of some stylometric techniques is in Maciej Eder, "Visualization in stylometry: Cluster analysis using networks," *Digital Scholarship in the Humanities* 32 (2017): 50–64; see also Stefan Evert et al. "Understanding and explaining Delta measures for authorship attribution," *Digital Scholarship in the Humanities* 32 (2017): ii4–ii16.

[66] The richer second-order metalemma, which swallows most of the verb content both in Greek and in Latin, contains 54 Greek lemmas and 102 Latin lemmas.

11a–d of the Appendix. MFW are the Most Frequent Words (in fact, second-order meta-words) in each work used to compute the distance, and the "culling" is the percentage of texts that must contain a given lemma in order for the latter to be included in the computation; a 0% culling means that no culling is operative, so all lemmas are included; "Culled @ 0–100%" means that culling is weighted out by consensus. The dendrograms confirm that the organisation of the lexicon in *Clavius* is farther from the organisation in *Elem*. than *Gr.-Lat*. and *Comm*. are, and that the latter two translations compete for being the most faithful rendering of the Greek original.

As is easy to see, the outcome of the cluster analysis depends on the selected distance, on the number of MFW, and on culling, sometimes in ways that may be bewildering. To overcome this problem, which can be perceived as a drawback, a "consensus tree" providing the most likely outcome of the dendrograms that result from varying the number of MFW and the culling can be constructed for each distance. The caption "Consensus 0.5" in the graphs means that a node is validated in the consensus tree as soon as at least 50% of the dendrograms have it; the number of MFW ranges from 100 to 160, culling from 0% to 100%. The consensus trees for the 10 distances available in the package `stylo` highlight a partition: 7 distances make *Gr.-Lat.*, and 3 *Comm.*, as the translation nearest to *Elem*. Two examples are displayed as Figures 12a–b of the Appendix.

The data displayed in a dendrogram can be also visualised by means of a weighted network, which represents a kind of "lexical interaction" between texts in the form of a network whose edges have a weight that is proportional to lexical proximity; this is measured, in the networks displayed as Figures 13a–b of the Appendix and in the same order from left to right, by the same two distances and parameters as in the two dendrograms displayed as Figures 11a–b of the Appendix.

Principal component analysis

A different approach, the "principal component analysis" (PCA) as applied to frequency distributions of lemmas in a set of works,[67] looks at a representative low-dimensional cross-section of the multidimensional space distribution of the relative frequency vectors f_{ij}, $i = 1 \dots n$, $j = 1 \dots k$, of n lemmas associated with the k works. "Representative" means here that most of the variance of the distribution is captured by the cross-section. This amounts to diagonalising the covariance matrix or the correlation matrix of the relevant vectors and by selecting the span of the eigenvectors that have the highest eigenvalues; each of the latter is the variance of a "principal component." For the purpose of representing the result in a graph, two such vectors are selected. This bidimensional cross-section projects the distribution cloud, which is located in the k-dimen-

[67] On PCA see Ian T. Jolliffe, *Principal Component Analysis* (New York; Berlin; Heidelberg: Springer, 2002).

sional space of the selected lemmas or in the n-dimensional space of the works,[68] onto the plane that contributes most to the lexical variation. The number of MFW is kept fixed, culling can be weighted out by consensus. The variations along the two principal components PC1 and PC2 are represented in a plane graph, which plots both the MFW and the titles of the works, each referred to axes lying on opposite sides of the plot square. A separate histogram in the plot gives the ratio of the variance "explained" – that is, projected onto the associated eigenvectors – by the several PCs. The works can also be set out in a separate graph.

The four plots displayed as Figures 14a–b and 15a–b of the Appendix select two MFW numbers (lower numbers give less crowded plots) and accordingly the two ways of presentation just mentioned, and operate independently on the covariance matrix and on the correlation matrix.[69] Metalemmas on the border of the cloud contribute to the variance of the distribution more than tokens near its center; the same applies to works. The plot for the covariance matrix of the metalemmas is extremely crowded towards the center, as is natural for a work and its translations – the plot for the correlation matrix better displays the fine structure of the central core. PCA as applied to the covariance and correlation matrices exhibits an extremely flattened distribution of lemmas and works along the two principal components: PC1 and PC2 "explain" a very high ratio of the variance.

Syntactic analysis

The deductive progression of the four works has been quantified by assigning a level of "logico-syntactic subordination," or "syntactic depth," to each autonomous clause of each proposition of a segment of text as representative as that of the entire Book I. The "autonomous clauses" are main and subordinate clauses, genitive and ablative absolutes, non-attributive participles; in terms of discourse logic, postposed explanations and identifications of objects introduced by "that is" are counted as subordinates.

The main clauses of each proposition of Book I are assigned level 0; each subordinate raises the level by one unit; nested subordinates allow reaching higher levels. The result is a numerical sequence that can be represented by a "subordination diagram," where the dots are joined by line segments. In principle, a Greek mathematical proposition in pure demonstrative style should exhibit a subordination diagram flattened on level 0, with the following exceptions:[70]

68 To a lemma is attached the k-component vector of its relative frequencies in the k works, to a work is attached the n-component vector of the relative frequencies in it of the n lemmas.

69 In the PCA plots, every second-order metalemma gets an abridged designation that is made of the first two lemmas it includes, taken in alphabetic order Greek-Latin and preceded by letter M.

70 On all the parts of demonstrative discourse listed below see Acerbi, *The Logical Syntax*, sects. 1.1, 4.1, 4.4, 4.5.3, 5.2.1, 5.3.2.

- the protasis of the conditional clause that formulates the enunciation of most theorems;
- the substantive clause that, in diorisms, states in instantiated form what is to be demonstrated;
- the causal subordinate that opens the proof or a sub-proof;
- the protasis of the conditional clause that opens a *reductio ad absurdum*, and the relative clause – ὅπερ ἄτοπον "which is absurd" or the like – that closes it;
- the QED formulas ὅπερ ἔδει δεῖξαι and ὅπερ ἔδει ποιῆσαι at the end of the conclusion of a proposition; this is also an example of a nested subordination, which reaches to level 2 (this fact will be further assessed presently).

One must not forget that a canonical text like the *Elements* is already an exception to the just-described optimal behaviour: in the whole treatise as it is handed down by the manuscript tradition, there are more than 200 postposed explanations introduced by γάρ (main clause, which is assimilated to a subordinate in point of discourse logic), 40 by ἐπειδήπερ (causal subordinate), and 36 by διὰ τό + infinitive (causal subordinate). Each of these clauses raises the level by 1.

We decided to process two sets of data: (1) as described above; (2) the subset of (1) that results from eliminating all conclusions, inclusive of the QED formula. The reasons for this choice are that (a) the conclusions – which by default repeat the enunciation word-for-word or are just a truncation of it – and the QED formulas do not have any deductive import; (b) *Gr.-Lat.* almost randomly curtails both the conclusions and the QED formulas (writing, for instance, just *quod oportebat* and omitting *facere*, which amounts to eliminating a level-2 subordinate), whereas Heiberg's edition of the Greek text homogenises the manuscript evidence and prints all of the QED formulas in full; (c) perfectly well-formed Latin QED formulas like *quod demonstrandum erat* count as one single level-1 subordinate; they are used almost randomly in *Clavius*. All of this shows that the fully-fledged conclusions introduce a bias in the measurement of syntactic depth.

The subordination diagrams of the four works are displayed as Figures 16a–h of the Appendix; for each work, the graph for dataset (1) precedes that for dataset (2).

The raw data set out in the subordination diagram may also be processed to give a number of bulk or intensive indicators of syntactic depth and information availability; all these indicators bar the second and the next-to-last in the list below increase with syntactic depth.

- The overall number of autonomous clauses in Book I, which is an index of deductive dilution, for the same result is established in a greater number of steps.
- The overall and relative (that is, normalised to the overall number of autonomous clauses) numbers of level 0 clauses in Book I, which is an index of adherence to the standard stylistic code.

- The overall and relative (that is, divided by the overall number of level ≠ 0 clauses) sum of the levels of level ≠ 0 clauses in Book I,[71] which are the area under the subordination diagram, if this were set out as a histogram, and its mean height, respectively, and measure the strength of the "explanatory brake."[72]
- The mean value for a proposition of the indicators defined in the previous item.
- The first intensive indicator is the "temperature" T_A of text *A*, defined by means of the Equipartition Theorem. We measure the "mean thermal energy" associated with the level fluctuations of the clause sample by computing the discrete first derivative of the level function and by taking the expectation value of the square of this derivative w.r.t. a discrete Maxwell-Boltzmann probability distribution featuring the same square in the mass function: if a_{ij} is the *i*-th autonomous clause in the *j*-th proposition of work *A* starting from a_{0j}, for proposition *j* we compute $\delta_{ij} = |a_{ij} - a_{i-1,j}|$, where $i \geq 1$, and we set $T_A \equiv \sum_j \sum_i \delta_{ij}^2 exp(-\delta_{ij}^2)$. As is to be expected, a stronger explanatory brake makes a text "hotter" than another. There is no need to introduce constants of normalisation because we set a definition and the temperature scale is arbitrary; see the next indicator. Any text flattened to one single level – which is necessarily level 0 – has $T_A = 0$.
- The "temperature" as defined in the previous item, but normalised to $T_A = 100$ for *Elem*.
- The second intensive indicator is a kind of entropy, which measures the degree to which information is "buried" in deep syntactic levels; if a_i is the relative (that is, divided by the overall sum of the levels of the clauses in Book I) sum of the levels of the clauses in proposition *i* of Book I of work *A*, we define its "proposition entropy" $H_p \equiv -\sum_i a_i \log a_i$. We set $H_p = 0$ for any text flattened to one single level.
- Likewise, if k_i is the relative (that is, divided by the number of clauses in Book I) frequency of level $i \geq 0$ clauses in Book I of work *A*, we define its "book entropy" as $H_b \equiv -\sum_i k_i \log k_i$. We set $H_b = 0$ for any text flattened to one single level.

The following tables set out the indicators of dataset (1) first, then those of dataset (2).

(1)	# cl.	# level 0		sum levels		mean prop.		T	T_n	H_p	H_b
Elem.	1332	796	59.760	628	1.172	13.083	2.441	253.243	100.000	1.6393	0.3741
Gr.-Lat.	1266	773	61.058	562	1.140	11.708	2.375	228.523	90.239	1.6344	0.3606
Comm.	1486	808	54.374	816	1.204	17	2.507	321.302	126.875	1.6269	0.4024
Clavius	1518	670	44.137	1157	1.364	24.104	2.842	363.306	143.461	1.5845	0.4639

[71] Of course, this coincides with the sum of the levels of all clauses in Book I.
[72] This metaphor means that syntactic depth, frequently in the form of postposed explanations, slows down the deductive progression.

(2)	# cl.	# level 0		sum levels		mean prop.		T	T_n	H_p	H_b
Elem.	1118	732	65.474	423	1.096	8.812	2.283	190.851	100.000	1.6173	0.3281
Gr.-Lat.	1129	732	64.836	440	1.108	9.167	2.309	195.194	102.275	1.6175	0.3354
Comm.	1217	738	60.641	555	1.159	11.562	2.414	238.898	125.175	1.5984	0.3673
Clavius	1396	632	45.272	1065	1.394	22.187	2.904	337.040	176.598	1.5687	0.4683

We may also compute the relative proposition entropy $H(A||B) \equiv \sum_i a_i \log(a_i/b_i)$ for both datasets, with a_i and b_i defined as for the proposition entropy. This yields the following values:

(1)	*Elem.*			(2)	*Elem.*		
Gr.-Lat.	0.0061	*Gr.-Lat.*		*Gr.-Lat.*	0.0043	*Gr.-Lat.*	
Comm.	0.0130	0.0140	*Comm.*	*Comm.*	0.0185	0.0208	*Comm.*
Clavius	0.0796	0.0676	0.7291	*Clavius*	0.0973	0.0936	0.9500

The first table portrays *Gr.-Lat.* as "cooler" and syntactically tighter than *Elem*. The second table shows that the above-mentioned abbreviations of QED formulas in *Gr.-Lat.* by and large explain away this bewildering state of affairs: the syntax of *Gr.-Lat.*, Book I dataset (2), is marginally "hotter" and less tight than the syntax of *Elem*. The subordination diagrams and all indicators make it obvious that *Clavius* slams on the explanatory brake and locates more information in deep syntactic levels. *Comm.* always occupies an intermediate position.

Conclusion(s)

After this full immersion in technicalities, we would like to make two final points, touching on translation theory and on a literary-theoretic assessment of Clavius' edition of the *Elements*.

First, we believe that the present paper might contribute to showing that a strictly technical, digital humanities-oriented approach to (early modern) scientific translation may lead to rigorous results that corroborate factually- and historically-rooted assessments. One must not forget that early translators very seldom indulge in metadiscurse about their own translation methods;[73] present-day historians must normally extract such methods –

[73] A celebrated exception is Burgundio's praise of the *de verbo ad verbum* method in the preface of his own translation of John Chrysostomos' homilies on John's Gospel: Peter Classen, *Burgundio von Pisa. Richter – Gesandter – Übersetzer*. (Heidelberg: Carl Winter Univeritätsverlag, 1974), 84–102, English translation of the relevant passages in Dimitri Gutas, Charles Burnett,

if any can be perceived – from selected features of a given translation, as for instance the way specific classes of function words are translated.[74] These focused investigations may also settle problems of attribution. In the case of the mediaeval translators of Aristotle, this approach was brought to perfection by Lorenzo Minio-Paluello and is currently pursued by the Leuven school that is in charge of the editions of the *Aristoteles Latinus*.[75] As we have seen on the example of *Gr.-Lat.*, such an approach was also used by the editors of the mediaeval Latin translations of Greek scientific works.[76] On the other hand, translation studies appear to have started systematically interacting with digital humanities only very recently; we hope that our study will contribute to widening this interaction.[77]

and Uwe Vagelpohl, eds., *Why Translate Science? Documents from Antiquity to the 16th Century in the Historical West (Bactria to the Atlantic)* (Leiden; Boston: Brill, 2022), 493–497; this book conveniently collects, both in the original and in English translation, preliminary statements of translators that operated before the modern era (neither of the statements to be read below is reported). Burgundio lists Boethius' version of Nicomachus' *Introductio arithmetica* among the *de verbo ad verbum* translations, but Boethius states exactly the contrary: Classen, "Burgundio," 91.139–92.141 = Gutas, Burnett, and Vagelpohl, *Why Translate Science?*, 495 versus Boethius, *De institutione arithmetica*, 4.27–5.4. In his philosophical translations, Boethius indeed acted as a *fidus interpres* (as Horace, *Ars poetica* 133–134, calls the author of a word-for-word version): Gutas, Burnett, and Vagelpohl, *Why Translate Science?*, 32, 44. One might also wish to compare Burgundio's praise with Vladimir Nabokov's brilliant defence of his own translation of Pushkin's *Onegin*, see Vladimir Nabokov, "Problems of Translation: *Onegin* in English," *Partisan Review* 22 (1955): 496–512.

74 The obvious fact that content words cannot be used to establish filiations between the Latin translations of the *Elements* (by the way, these filiations were already established, to the satisfaction of everyone, in the Méδée website) has apparently escaped the author of Benjamin Wardhaugh, "Euclidean terms in European languages, 1482–1703," *Historia Mathematica* 68 (2024): 22–37.

75 See Minio-Paluello, *Opuscula*, and the *prolegomena* to the several volumes of the *Aristoteles Latinus* series, listed at https://hiw.kuleuven.be/dwmc/research/al.

76 For *Gr.-Lat.*, see already John E. Murdoch, "Euclides Graeco-Latinus: A Hitherto Unknown Medieval Latin Translation of the Elements Made Directly from the Greek," *Harvard Studies in Classical Philology* 71 (1967): 249–302. Lexical analyses of other products of the "Sicilian school" of translation (on which see at least Charles H. Haskins, *Studies in the History of Mediaeval Science* (Cambridge, MA: Harvard University Press, 1924), 155–193) can be read in Wilfred R. Theisen, "*Liber De Visu*: The Greco-Latin Tradition of Euclid's Optics," *Mediaeval Studies* 41 (1979): 44–105, at 51–54; Shuntarō Itō, *The Medieval Latin Translation of the Data of Euclid* (Tokyo; Boston; Basel; Stuttgart: University of Tokyo Press; Birkhäuser, 1980), 23–38.

77 See Julie McDonough Dolmaya, *Digital Research Methods for Translation Studies* (London and New York: Routledge, 2024); Hilary Brown, Regina Toepfer, and Jörg Wesche, *Early Modern Translation and the Digital Humanities* (Berlin: J. B. Metzler, 2025); Raluca Tanasescu, "Translation Studies and Digital Humanities," in *The Routledge Handbook of Translation Technology and Society*, ed. by Stefan Baumgarten and Michael Tieber (London and New York: Routledge, 2025), 192–205.

Second, our analysis appears to support a historiographical conclusion. To get there, we first remark that the Latin mathematical lexicon did not get stabilised by a sustained tradition as the Greek lexicon did. By Clavius' times, the mediaeval translation practice *de verbo ad verbum* had fallen out of fashion; as for the geometric lexicon, it was the result of the conflation of the lexica stemming from the mediaeval Arabo-Latin tradition and from the much more recent Graeco-Latin tradition. Prospective scholars received a *trivium*-based education, with glimpses, if any, of quadrivial territory.[78] This entails, among other things, that a varied lexicon and a complex syntax were employed as a matter of course and even valued, thereby diluting the lexical and syntactic rigidity of technical Greek.

Clavius' well-thought, mathematically sound, far-reaching educational project aims at blending the theoretical mathematics handed down from ancient Greece (Tome I of his *Opera Mathematica*), the practical mathematics that characterised the Middle Ages and the early modern era (Tome II), and the mathematics traditionally applied to astronomy and chronology in spherical and plane trigonometry (Tomes I and III), astrolabica (Tome III), gnomonica (Tome IV), and Easter Computi (Tome V). The intended readership of this project mainly point to *trivium*-trained, young students. Let us first read how Conrad Dasypodius – the author with Christian Herlinus of a translation of *Elem*. I–VI featuring a "translation" of all proofs into Aristotelian syllogisms – put it a few years before Clavius published his own edition:

> [an Euclidean proof] *ex multis constat syllogismorum argumentis seu mediis, ad tollendam itaque obscuritatem & difficultatem: syllogismorum propositiones seiunguntur: & quæ à Theone acutissima breuitate sunt connexa: hic fusius sunt dilatata, & apertiora reddita. [...] Erunt fortassis qui dicent hæc nimis scholastica & puerilia esse: fateor & apertè atque candidè dico pueris me hæc scribere, & nostris discipulis qui ex classibus post cognitionem linguarum, adhæc Rhetoricæ & Dialecticæ deducuntur ad geometras, & descendunt in arenam geometrarum, ut ibi se exerceant: & iudicium quod tenerum adhuc est, confirment: animum quoque suum hisce contemplationibus iuuent: denique ipsam memoriam in qua multum est situm corroborent.*[79]

To summarise: this is trivial stuff devised for our students: we needed to give the Euclidean proofs a form (apparently, no Euclidean form was perceived): this form is the Aristotelian syllogism, by means of which we also make explicit all deductive steps that the original text keeps implicit.

[78] Clavius, a self-taught mathematician, certainly did not get any glimpse: he was famously flashed by geometry when attending a course on Aristotle's *Posterior Analytics* (Clavius, *Corrispondenza*, vol. I, parte 1, 38).

[79] *Analyseis geometricæ sex librorum Euclidis. Primi et quinti factæ à Christiano Herlino : reliquæ unà cum commentariis, & Scholiis perbreuibus in eosdem sex livros Geometricos: à Cunrado Dasypodio. [...] Pro schola Argentinensi [...] Excudebat Iosias Rihelius. M.D.LXVI*, aiii.

Clavius probably did not have students he could call *pueri*, but he also needed to give his Euclid a form suited to such a readership and attendance: a form in which the tightly intertwined steps of the logical fabric of a proof *fusius sunt dilatata, & apertiora reddita*. Let us read the statement of Clavius' programme, which includes a statement of the method underlying his version of the *Elements*:[80]

> *Demonstrationes aliorum, maxime Theonis, quas quidem ipsius esse Euclidis, non leuibus argumentis adducti quidam asseuerant, & Proclus etiam testatur, breuiores, quantum per rei difficultatem licuit, uel certe planiores, quando illud non potuimus, delucidioresque reddere conati sumus. Non enim illas nude, ac totidem uerbis, quot erant scripte, proposuimus. Etenim ea est interdum illarum breuitas, ut illud accidat, quod ab elegantissimo poeta dictum est.* Breuis esse laboro, obscurus fio*:*[81] *Interdum etiam, cum breuius, atque succinctius efferri possint, magna, ob longiorem, quam satis est, sermonem, affertur molestia legenti. Quare utrunque uitantes, eas, uelut παραφραστικῶς, atque ad eum fere modum tradidimus, quem, cum publice Euclidem interpretaremur, obseruauimus; hac etiam re auditorum desiderio, & uoluntati, quantum est in nobis, satisfacere cupientes. Ita enim, nostra sententia, Euclides facilius a studiosis, ijs præsertim, qui ceu tyrones, hæc Mathematica studia nunc primum auspicantur, ac maiore uoluptate, utilitateque cognoscetur.*

Thus, Clavius' edition of the *Elements*, and in fact most of his works, are first and foremost scholarly products tuned to a readership of students.

In the perspective of the present study, the result of all of this is the dissolution of one of the stylistic codes that made Greek mathematics a distinctive literary product and the *Elements* one of its gems. Starting with Clavius, and despite his boasting about adopting an intermediate approach, this code was replaced by a mix of pedantry and rhetorical flourish, both witnessed, as highlighted in the previous Sections, by the conspicuous lexical spread and by the increasing syntactic complexity of Clavius' text – not a real translation,

[80] *Euclidis Elementorum libri XV, Ad Lectorem*. A similar statement can be read in the short text introducing the five riddles in form of epigrams contained in the *Appendix* of Clavius' *Algebra* (*Algebra Christophori Clavii Bambergensis e Societate Iesu. Romae, Apud Bartholomæum Zannettum. Anno M. DC. VIII.*, 378; the enthusiastic praise of algebra on pages 1–3 is worth a reading): Clavius states that they are followed by *eorum interpretatione, non quidem de verbo ad verbum, sed ut res clarior fieret, expresso illorum senso dumtaxat*. Note that this does not prove that Clavius himself translated these poems; the fact that he proposed three versions of the last epigram suggests the opposite. Clavius' preface to his own edition of the *Elements* was slightly modified in the subsequent reprints: most notably, Commandinus was downgraded from *geometra peritissimus* (1574) to *non vulgaris* (1589 and later).

[81] Horace, *Ars poetica* 25–26. This quote is also an allusion to a broader debate, for the topos of the *fidus interpres*, amply attested in Latin sources as far as late antiquity, frequently alludes to Horace's programmatic poem (see note 73 above).

let us stress this point once more – of the *Elements*. The dissolution of the demonstrative stylistic code paved the way to establishing the modern historiographical assessment of Greek mathematical works through their "mathematical content" only, and to entrenching the belief that scientific texts are lowest-brow paraliterature amply deserving the contempt of purportedly high-brow classicists.

References

Early printed sources

Algebra Christophori Clavii Bambergensis e Societate Iesu. Romae, Apud Bartholomæum Zannettum. Anno M. DC. VIII.

Analyseis geometricæ sex librorum Euclidis. Primi et quinti factæ à Christiano Herlino : reliquæ unà cum commentariis, & Scholiis perbreuibus in eosdem sex livros Geometricos: à Cunrado Dasypodio. […] Pro schola Argentinensi […] Excudebat Iosias Rihelius. M.D.LXVI, aiii.

Christophori Clavii Bambergensis e Societate Iesu Opera Mathematica V. Tomis distributa […]. Moguntiae Sumptibus Antonij Hierat excudebat Reinhardus Eltz […] Anno M. DCXII.

Euclidis Elementorum libri XV. Accessit XVI de Solidorum Regularium comparatione. Omnes perspicuis demonstrationibus, accuratisque scholijs illustrati. Auctore Christophoro Clavio Bambergensi Societatis Iesu. Romae, Apud Vincentium Accoltum. 1574.

Euclidis Elementorum libri XV. Graecè & Latiné, Quibus, cùm ad omnem Mathematicae scientiae partem, tùm ad quamlibet Geometriae tractationem, facilis comparatur aditus […]. Lutetiae, Apud Gulielmum Cavellat […] 1557.

Euclidis Elementorum libri XV. Unà cum scholijs antiquis, À Federico Commandino urbinate nuper in latinum conversi, commentarijsque quibusdam illustrati, Pisauri, M D LXXII.

Euclidis Megarensis Geometricorum elementorum libri XV, Campani Galli transalpini in eosdem commentariorum libri XV, Theonis Alexandrini Bartholamaeo Zamberto veneto interprete, in tredecim priores, commentariorum libri XIII, […] Parisiis in officina Henrici Stephani e regione scholae Decretorum.

Euclidis megarensis philosophi platonicj Mathematicarum disciplinarum Janitoris […] Bar. Zamber. Vene. Interprete. […] Impressum Venetiis […] in aedibus Ioannis Tacuini librarii accuratissima diligentia recognitum. Anno reconciliatae divinitatis.M.D.V.VIII.Kalendas novembris.

ΕΥΚΛΕΙΔΟΥ ΣΤΟΙΧΕΙΩΝ ΒΙΒΛ. ΙΕ'. ΕΚ ΤΩΝ ΘΕΩΝΟΣ ΣΥΝΟΥΣΙΩΝ. Εἰς τοῦ αὐτοῦ τὸ πρῶτον, ἐξηγημάτων Πρόκλου βιβλ. δ'. […]. Basileae Apud Ioan. Hervagium Anno M. D. XXXIII. Mense Septembri.

Opus elementorum euclidis megarensis in geometriam artem In id quoque Campani perspicacissimi Commentationes finiunt. Erhardus ratdolt Augustensis impressor solertissimus. venetiis impressit. anno salutis M.cccc.lxxxij. Octavis. Calendas. Junii.

Secondary literature

Acerbi, Fabio. *The Logical Syntax of Greek Mathematics*. New York; Heidelberg: Springer, 2021.

Acerbi, Fabio. "Rewriting Mathematical and Astronomical Treatises." In *The Routledge Handbook for Rewriting in Byzantium*, edited by Juan Signes Codoñer, Martin Hinterberger, and Inmaculada Pérez Martín, 407–419. London: Routledge, 2025. https://doi.org/10.4324/9781003274339-31

Acerbi, Fabio, Stefano Martinelli Tempesta, and Bernard Vitrac. "Gli interventi autografi di Giorgio Gemisto Pletone nel codice matematico Marc. gr. Z. 301." *Segno e Testo* 14 (2016): 411–456. https://hal.science/hal-03596976v1

Acerbi, Fabio, and Ramon Masià, *George of Trebizond, Introduction to the Almagest. Greek and Latin Text*. Leiden: Brill, 2026.

Alstott, Jeff, Ed Bullmore, and Dietmar Plenz. "powerlaw: a Python package for analysis of heavy-tailed distributions." arXiv:1305.0215v3, 2014. https://doi.org/10.1371/journal.pone.0085777

Baldini, Ugo. "Christoph Clavius and the Scientific Scene in Rome." In *Gregorian Reform of the Calendar. Proceedings of the Vatican Conference to Commemorate Its 400th Anniversary. 1582-1982*, edited by George V. Coyne, Michael A. Hoskin, and Olaf Pedersen, 137–169. Città del Vaticano: Pontificia Academia Scientiarum, 1983.

Baldini, Ugo. *Legem Impone Subactis. Studi su filosofia e scienza dei Gesuiti in Italia. 1540-1632*. Roma: Bulzoni, 1992.

Boethius, *De institutione arithmetica libri duo*. Edited by Gottfried Friedlein. Lipsiae: In aedibus B. G. Teubneri, 1867.

Brown, Hilary, Regina Toepfer, and Jörg Wesche. *Early Modern Translation and the Digital Humanities*. Berlin: J. B. Metzler, 2025.

Busard, Hubert L. L. *The Mediaeval Latin Translation of Euclid's* Elements *made directly from the Greek*. Stuttgart: Franz Steiner Verlag, 1987.

Classen, Peter. *Burgundio von Pisa. Richter – Gesandter – Übersetzer*. Heidelberg: Carl Winter Univeritätsverlag, 1974.

Clauset, Aaron, Cosma R. Shalizi, and Mark E. J. Newman. "Power-law distributions in empirical data." *SIAM Review* 51 (2009): 661–703. https://doi.org/10.1137/070710111

Clavius, Christoph. *Corrispondenza*. Edited by Ugo Baldini and Pier Daniele Napolitani. Pisa: Università di Pisa, Dipartimento di Matematica, 1992.

Contucci, Pierluigi, Claudio Giberti, Godwin Osabutey, and Cecilia Vernia. "Statistical properties of the rooted-tree encoding of N." *Physica A* 686 (2026), art. 131361. https://doi.org/10.1016/j.physa.2026.131361

De Risi, Vincenzo. "The Derivability Theory of Axioms. Logic and Mistranslation in the Middle Ages and the Renaissance." In *Pre-Modern Mathematical Thought. The Latin Discussion (13th–16th Century)*, edited by Clelia V. Crialesi, 237–288. Leiden: Brill, 2025.

Eder, Maciej. "Visualization in stylometry: Cluster analysis using networks." *Digital Scholarship in the Humanities* 32 (2017): 50–64. https://doi.org/10.1093/llc/fqv061

Eder, Maciej, Jan Rybicki, and Mike Kestemont. *'Stylo': a package for stylometric analyses*. 2015.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *R Journal* 8 (2016): 107–121. https://doi.org/10.32614/RJ-2016-007

Evert, Stefan et al. "Understanding and explaining Delta measures for authorship attribution." *Digital Scholarship in the Humanities* 32 (2017): ii4–ii16. https://doi.org/10.1093/llc/fqx023

Feingold, Mordechai, ed. *The New Science and Jesuit Science: Seventeenth Century Perspectives*. Amsterdam: Kluwer, 2003.

Ferreri Cancho, Ramon, and Ricard V. Solé. "Two Regimes in the Frequency of Words and the Origins of Complex Lexicons: Zipf's Law Revisited." *Journal of Quantitative Linguistics* 8 (2001): 165–173. https://doi.org/10.1076/jqul.8.3.165.4101

Folkerts, Menso. "Leonardo Fibonacci's Knowledge of Euclid's *Elements* and of Other Mathematical Texts." *Bollettino di Storia delle Scienze Matematiche* 24 (2004): 93–113.

Gatto, Romano. *Tra scienza e immaginazione. Le matematiche presso il collegio gesuitico napoletano (1552-1670 ca.)*. Firenze: Olschki, 1994.
Gatto, Romano. "Christoph Clavius' 'Ordo Servandus in Addiscendis Disciplinis Mathematicis' and the Teaching of Mathematics in Jesuit Colleges at the Beginning of the Modern Era." *Science & Education* 15 (2006): 235–258. https://doi.org/10.1007/s11191-004-5607-8
Giusti, Enrico. "La théorie des proportions au XVI[e] siècle : entre philologie et mathématiques." In Liber amicorum *Jean Dhombres*, edited by Patricia Radelet-de Grave, 173–193. Turnhout: Brepols, 2008.
Gorman, Michael John. *The Scientific Counter-Revolution. The Jesuits and the Invention of Modern Science*. London; etc.: Bloomsbury Academic, 2020.
Gutas, Dimitri, Charles Burnett, and Uwe Vagelpohl, eds. *Why Translate Science? Documents from Antiquity to the 16th Century in the Historical West (Bactria to the Atlantic)*. Leiden; Boston: Brill, 2022.
Haskins, Charles H. *Studies in the History of Mediaeval Science*. Cambridge MA: Harvard University Press, 1924.
Heaps, Harold S. *Information Retrieval. Computational and Theoretical Aspects*. New York; San Francisco; London: Academic Press, 1978.
Itō, Shuntarō. *The Medieval Latin Translation of the Data of Euclid*. Tokyo; Boston; Basel; Stuttgart: University of Tokyo Press; Birkhäuser, 1980.
Jolliffe, Ian T. *Principal Component Analysis*. New York; Berlin; Heidelberg: Springer, 2002.
Knobloch, Eberhard. "Sur la vie et l'œuvre de Christophore Clavius (1538–1612)." *Revue d'histoire des sciences* 41 (1988): 331–356. https://doi.org/10.3406/rhs.1988.4100
Knobloch, Eberhard. "L'œuvre de Clavius et ses sources scientifiques." In *Les jésuites à la Renaissance. Système éducatif et production du savoir*, edited by Luce Giard, 263–283. Paris: Presses Universitaires de France, 1995.
Lattis, James M. *Between Copernicus and Galileo. Christoph Clavius and the Collapse of Ptolemaic Cosmology*. Chicago and London: The University of Chicago Press, 1994.
Mandelbrot, Benoît. "Information Theory and Psycholinguistic." In *Scientific Psychology*, edited by Benjamin B. Wolman, 550–562. New York: Basic Books, 1965.
McDonough Dolmaya, Julie. *Digital Research Methods for Translation Studies*. London and New York: Routledge, 2024.
Minio-Paluello, Lorenzo. "Guglielmo di Moerbeke traduttore della *Poetica* di Aristotele (1278)." *Rivista di Filosofia Neo-Scolastica* 39 (1947): 1–17.
Minio-Paluello, Lorenzo. "Henri Aristippe, Guillaume de Moerbeke at les traductions latines médiévales des Météorologiques et du De Generatione et Corruptione d'Aristote." *Revue Philosophique de Louvain* 45 (1947): 206–235. https://doi.org/10.3406/phlou.1947.4104
Minio-Paluello, Lorenzo. *Opuscula. The Latin Aristotle*. Amsterdam: Adolf M. Hakkert, 1972.
Miyazima, Sasuke, and Keizo Yamamoto. "Measuring the Temperature of Texts." *Fractals* 16 (2008): 25–32. https://doi.org/10.1142/S0218348X08003776
Montemurro, Marcelo A. "Beyond the Zipf–Mandelbrot law in quantitative linguistics." *Physica A* 300 (2001): 567–578. https://doi.org/10.1016/S0378-4371(01)00355-7
Moscheo, Rosario. *I gesuiti e le matematiche nel secolo XVI. Maurolico, Clavio e l'esperienza siciliana*. Messina: Società Messinese di Storia Patria, 1998.

Murdoch, John E. "Euclides Graeco-Latinus: A Hitherto Unknown Medieval Latin Translation of the Elements Made Directly from the Greek." *Harvard Studies in Classical Philology* 71 (1967): 249–302. https://doi.org/10.2307/310767

Nabokov, Vladimir. "Problems of Translation: *Onegin* in English." *Partisan Review* 22 (1955): 496–512.

Newman, Mark E. J. "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46 (2005): 323–351. https://doi.org/10.1080/00107510500052444

Price, Audrey. "Pure and Applied: Christopher Clavius's Unifying Approach to Jesuit Mathematics Pedagogy." PhD diss., University of California, San Diego, 2017.

Puri, Ravinder R. *Modern Thermodynamics and Statistical Mechanics. A Comprehensive Foundation*. Cham: Springer, 2024.

Romano, Antonella. *La Contre-réforme mathématique. Constitution et diffusion d'une culture mathématique jésuite à la Renaissance*. Roma: École française de Rome, 1999.

Rommevaux, Sabine. *Clavius : une clé pour Euclide au XVIe siècle*. Paris: Librairie Philosophique J. Vrin, 2005.

Rose, Paul L. *The Italian Renaissance of Mathematics*. Genève: Librairie Droz, 1975.

Sabra, Abdelhamid I. "The Appropriation and Subsequent Naturalization of Greek Science in Medieval Islam: A Preliminary Statement." *History of Science* 25 (1987): 223–243. https://doi.org/10.1177/007327538702500301

Salvi, Francesca. "Le fonti degli *Elementi di Euclide* di Cristoforo Clavio." Tesi di Laurea, Università degli Studi di Firenze, 2005.

Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (1948): 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Tanasescu, Raluca. "Translation Studies and Digital Humanities." In *The Routledge Handbook of Translation Technology and Society*, edited by Stefan Baumgarten and Michael Tieber, 192–205. London and New York: Routledge, 2025.

Theisen, Wilfred R. "*Liber De Visu*: The Greco-Latin Tradition of Euclid's Optics." *Mediaeval Studies* 41 (1979): 44–105. https://doi.org/10.1484/J.MS.2.306242

Vitrac, Bernard. "La traduction latine des *Éléments* d'Euclide par Federico Commandino : sources, motivations." hal-03328386, 2021.

Vitrac, Bernard. "A propos de l'histoire du texte des *Éléments* d'Euclide : Préalables à une nouvelle édition critique." hal-03328161, 2022.

Wardhaugh, Benjamin. "Euclidean terms in European languages, 1482–1703." *Historia Mathematica* 68 (2024): 22–37. https://doi.org/10.1016/j.hm.2024.07.002