Statistics and linguistics: Can we tell something more about Pliny the Elder?

0. Introduction

The language and style of Pliny the Elder have been studied since the 19th century with a strong comparative approach. Grasberger,[[1]](#footnote-1) Müller,[[2]](#footnote-2) and Gaillard[[3]](#footnote-3) have analyzed and recorded which features distinguished Pliny’s style from the standard (or Ciceronian) literary one. We are offered long lists of examples of ‘unclear’ and ‘obscure’ sentences, with little or no attention to the specific characteristics of Pliny’s work. The Natural History, indeed, treated an amount of arguments and information in the field of ancient science that could hardly be found in any other Latin author. This approach has led to severe judgments from historians of Latin literature,[[4]](#footnote-4) and only in the second half of the twentieth century have scholars drawn attention to the particularities of Pliny’s Natural History, pointing out the necessity of finding a new methodological path to explore its language and style. Works by Önnerfors,[[5]](#footnote-5) Beaujeu,[[6]](#footnote-6) Healy,[[7]](#footnote-7) and Capponi,[[8]](#footnote-8) and a number of other articles,[[9]](#footnote-9) have discussed the extent to which ‘irregularities’ in Pliny’s language or lexicon are due to the unusual contents of his books and to his attitude towards the information he conveyed, which, in addition, was often taken from Greek treatises. Pinkster[[10]](#footnote-10) has finally stated how pragmatics, thanks to a wide range of studies concerning the characteristics of technical writings, may represent the key to understanding the priorities that have shaped Pliny’s language.

The aim of this paper is to show that, after having taken into account all the methodological approaches of the last fifty years, the use of digital and statistical tools can help orient the scholar in the ample spectrum of morphological and syntactical phenomena which give a faithful representation of Pliny’s language and style. The information obtained through these statistical tools is certainly meaningful only if interpreted in light of the original text: this paper will show how the statistical analysis of classical texts can accompany the traditional linguistic and literary analysis, helping the scholar to identify phenomena which are not always evident through direct reading. My work, as will be explained later in detail, focuses on the second book of the Natural History, which treats mainly astronomy.

This paper is structured in three sections. Section one will explain in detail the nature of my study, introduce the statistical analysis of classical languages as a theme, describe the statistical and digital tools that have been chosen for this work, and outline which kind of results might be expected from these analyses. Section two will deal with the comparison between two texts—Natural History II and Seneca’s Natural Questions VII[[11]](#footnote-11)—from a syntactic point of view. I will then briefly sketch how Natural History II is positioned in a corpus representing different literary genres (section three), though without carrying out a deep textual and historical analysis.

1. Pliny the Elder and statistical tools

1.1 Pliny the Elder and the language of science

Pliny’s work stands out as an unicum in classical literature: as Schilling[[12]](#footnote-12) has shown, Pliny has proven capable of great originality in content (showing personal opinions and criticism towards his sources), in structure (following often an ‘empiric approach’ more than an abstract disposition of the arguments), and in history of science, being at the same time a scrupulous preserver of previous knowledge but still a curious investigator of every subject studied. From a stylistic point of view, many scholars (first of whom, Önnerfors[[13]](#footnote-13)) have highlighted that Pliny’s language cannot be simply described as technical, since it is characterized by frequent shifts from technical to literary and even almost vernacular language. This complexity might frustrate the attempt to give a systematic description of Pliny’s language and style. Moreover, Pliny was confronted with two different scopes: on one side to provide a summary of all the knowledge of his time, while on the other to try to give a moral presentation of this knowledge, without necessarily renouncing a philosophical frame.

A second element of complexity is that, as many scholars have pointed out (latest of whom, Pinkster[[14]](#footnote-14)), every thematic unity in the Natural History must be considered as autonomous from stylistic and linguistic points of view: not only, obviously, the vocabulary, but also the structure of the sentences is strongly influenced both by the subject and the sources, which are necessarily different in every section. The decision to focus on the second book stems from the consideration of its thematic unity; moreover, it is the only book dealing with astronomical subjects, allowing for a focus on the Plinian language of astronomy. We shall add that book II is an interesting mélange of styles: starting with a stylistically factual introduction concerning the mysteries of the universe, it continues with a very complicated description of the movement of the sun, stars, and planets (a purely technical section), and ends with more descriptive paragraphs concerning terrestrial phenomena. It is therefore a good example of the complexity of Plinian style.

The first studies of Pliny’s language are based on the view that Pliny’s syntax is anomalous in comparison to other (mainly literary) writers. As already discussed, for instance, by Évrard and Mellet,[[15]](#footnote-15) this kind of approach actually precludes a statistical model, and, since it is based not on the computation of phenomena but on the impressions of the scholar, it might lead to inexact statements. An objective and systematic approach, with the help of statistical data, is necessary to avoid such interpretations. This paper will focus on two kinds of comparison: direct comparison with another astronomy treatise, in order to evaluate whether the two styles are similar or which features distinguish them, and comparison with other literary texts, in order to see if the scientific aspects of the text are prevalent or if it shares meaningful syntactic features with other genres.

The comparisons will be carried out by statistical calculations handled by specific programs, which will be described in the next paragraph.

1.2 The L.A.S.L.A., HyperbaseWeb, and the second book of the Natural History.

This work takes place among the research of the L.A.S.L.A. (Laboratoire d’Analyse Statistique des Langues Anciennes), founded in 1961 by Louis Delatte[[16]](#footnote-16) at the University of Liège and now under the direction of Prof. Dominique Longrée. From its beginning, the L.A.S.L.A. has been involved in the time-consuming but unavoidable duty of lemmatizing and morphosyntactically tagging Latin and Greek texts, a necessary step to statistically analyze a corpus of classical authors. Despite the development of automatic lemmatizers and taggers, the L.A.S.L.A’s lemmatization is handled by specialized philologists. Each philologist is responsible for, at least, an entire text (book, work, tragedy, etc.), in order to guarantee a certain consistency to the tagging within that text. The L.A.S.L.A. ensures coordination among scholars, stating general rules (which are listed in the handbook of lemmatization[[17]](#footnote-18)), and answering the most frequent questions; constant communication among scholars working on a text’s lemmatization is also provided, both remotely and through personal meetings. This manual work can rely with a good degree of certainty on the data extracted from L.A.S.L.A. treated texts, though human errors are unavoidable. Having been active for more than fifty years now, a significant amount of Latin literature has been treated (almost two million words), and work continues to be done. L.A.S.L.A. lemmatization provides morphological information (part of speech, declension, conjugation, mood, number, gender etc.) and some syntactical (is a verb independent or subordinate? To which kind of subordinate sentence does it belong?).[[18]](#footnote-19) In fact, the L.A.S.L.A. has contributed to methodological reflections about the use of statistical analysis in classical studies, and more specifically, in Latin and Greek linguistics. The L.A.S.L.A. has both produced its own works based on this approach and developed digital tools[[19]](#footnote-20) used to broaden the amount of information that can be drawn from a text.

Nowadays, L.A.S.L.A. lemmatized and tagged texts are available online, on the website hyperbase.unice.fr, which is an online version of the previous CD-ROM program Hyperbase Textes Latins. This has been made possible by the close collaboration between the L.A.S.L.A. and the laboratory BCL - «Bases, Corpus, Langages», Université de Nice-CNRS. HyperbaseWeb provides the researcher with many statistical tools, which I will quickly list, providing the technical bibliography.[[20]](#footnote-21)

A recent study by Poudat and Landragin[[21]](#footnote-22) offers a complete description of methods and instruments available for corpus-based research, showing the number of options available to every scholar, and giving indications about which methods are preferable depending on the nature of the research. I therefore recommend this reading for a more complete description of corpus-based research, while I will focus only on tools which are useful for this specific study.

* Search

The “Search” instrument allows users, not only to find a specific form surrounded by a certain span of text (that can be selected by the user), but also to search for all the forms corresponding to a certain morphological analysis (for instance all the substantives of the second declination) and all the forms deriving from a certain lemma.[[22]](#footnote-23) The user can also look for sequences combining forms, lemmata, codes, and unspecified words.

* Theme or specific co-occurrents

This function allows the user to find the co-ocurrents of a certain form, lemma, or morphological code. The user can indicate the span of text considered for co-occurrence: a paragraph might be chosen for thematic research, while a sentence might be more appropriate for a strictly linguistic analysis. The user can choose as well if the co-ocurrence will be calculated considering the lemmata, the forms, or the code of the words included in the span. Finally, the user can decide to filter the results, taking into account—for the calculations—only some grammatical categories (for example the verbs, or the substantives, etc.).[[23]](#footnote-24) HyperbaseWeb shows also the co-ocurrents of second degree, i.e. words that are co-occurent of those being considered.[[24]](#footnote-25)

* Distribution

The distribution tool combines different kind of functions whose aim is to show how linguistic features are distributed in a corpus formed by several texts. In particular, through the calculation of the z-score, it is possible to show which grammatical or lexical features differentiate each part of the corpus.[[25]](#footnote-26). The results can be visualized as a histogram, which can display the z-score, the absolute frequency or the relative frequency of a certain form, lemma, or code. The program can also generate a Correspondence Analysis (CA), representing on a Cartesian graph the relative positions of text and features (or terms) in order to highlight oppositions or, on the contrary, correlations among parts of the corpus,[[26]](#footnote-27) based on the words or grammatical categories chosen by the scholar. Another available graphical representation is the tree-analysis, which organizes in branches either the texts of the group or the categories chosen on the basis of the proximity to/distance from every other element. The number indicated on the node shows the priority in the grouping of elements. The distance dividing one element from another (measured by following the branches) indicates the distance between the twos.[[27]](#footnote-28).

The ‘search’ tool is useful, as will be shown later, for generating a list of passages containing a certain linguistic feature (not only a certain form), allowing the philologist to look more deeply at the analyzed phenomena. The distribution tool helps to situate Pliny among other texts, highlighting which features effectively distinguish Pliny’s prose. The analysis of co-ocurrents, finally, helps to nuance the role of certain linguistic elements in a text, providing some hints for the effective comprehension of the use of such elements in a certain context. It can also show, when comparing the co-ocurrents of the same element in different texts, how different contexts influence the role of this element. However, it will not be used in this specific article.

2. Seneca’s Natural Questions and Pliny’s Natural History

Seneca’s work Natural Questions and Pliny’s second book have often been associated, since, in a span of a few years, they deal with the same subject matter.[[28]](#footnote-29) It is important to underline that, despite a superficial similarity, the two works are deeply different, since they are written at different scopes: Quaestiones examines each presented opinion in depth, and Historia, though very complex to analyze, hints at a wider approach to all human knowledge.[[29]](#footnote-30) The two works convey a different view of nature, clearly observed by Ramos-Maldonado 2000–2002. In order to compare texts of similar length, let us take the seventh book of the Natural Questions. Dealing with cometae, it supports the thesis that they are regular and not accidental celestial phenomena, facing the same effort of Pliny when describing the movements of celestial bodies.[[30]](#footnote-31)

Seneca provides the reader with an ample doxography concerning every theme approached in the books. Afterwards, he adds his own opinion, showing therefore an active engagement with his predecessors. Much more concerned by the description of phenomena than by their ‘obscure’ causes,[[31]](#footnote-32) Pliny does not take a doxographical approach to the matter, even though he is reporting others’ opinions. Previous authors are regarded much more as sources of data and information than for their scientific theories, and even if Pliny shows some criticism in regards to the contents of the works he consults, he prioritizes conveying as much knowledge as his age can provide over debating scientific issues.

The specific vocabulary of these two texts is quite similar. But the question remains: are the different natures of the works reflected in the language of the two authors, or does the common subject lead to similar expression?

2.1 Statistical Data

The corpus, formed by Natural History II and Natural Questions VII, is constituted by 25,293 words, about 18,000 of which come from Natural History and 7,000 from Natural Questions. For the following histograms, the exact numerical values will be listed in the Appendix. A first interesting glance, provided by the ‘Distribution’ tool, shows how the discourses are distributed between the two texts[[32]](#footnote-33) (Figure 1).

 [[Figure 1 shall be inserted here]]

Figure : Histogram of the z-score of the parts of speech in the second book of the Natural History and the seventh book of the Natural Questions.

Only values higher than two (or lower than negative two) are statistically meaningful. The first striking evidence is that many elements stand above the limit-value of 5, which indicates that they differ significantly in the use of language. Summarizing the data, Pliny’s language is much more nominal than Seneca’s, which, on the contrary, is distinguished by the use of verbs. Pliny makes also an abundant use of numerals, which is a consequence of his accuracy in describing the movement of the planets and the length of celestial phenomena. It is also noticeable that Pliny makes a wider use of prepositions.

Focusing on the nouns, it might be interesting to see whether Pliny’s outstanding z-score is due to a specific declension of nouns, and therefore linked on some semantic or morphological basis, or to a specific case, which might lead to syntactical considerations, or to some gender or number, which might highlight the preference for collective nouns, or abstract nouns, etc. (Figure 2).

[[Figure 2 shall be placed here]]

Figure : Histogram of the z-score of the different cases, numbers and declensions in the use of substantives in the second book of the *Natural History* and the seventh book of the Natural Questions.

Looking at the cases, the ablative is by far (17.8) the most characteristic feature of Pliny’s use of nouns, followed by the genitive (10.0). Highly meaningful is also the use of singular nouns, and, morphologically speaking, the frequency of nouns of the second declension.

The important use of the ablative case in Pliny has already been observed.[[33]](#footnote-34) It is even more interesting to look at the absolute frequency of the different cases in the two authors (Figure 3). Seneca employs more frequently nominative, accusative, and vocative cases (883 total attestations) than dative, ablative, and genitive (540). On the contrary, Pliny prefers the latter (3274) to the former (2479). This is quite a clear indication, in my opinion, of a completely different way of structuring a sentence.

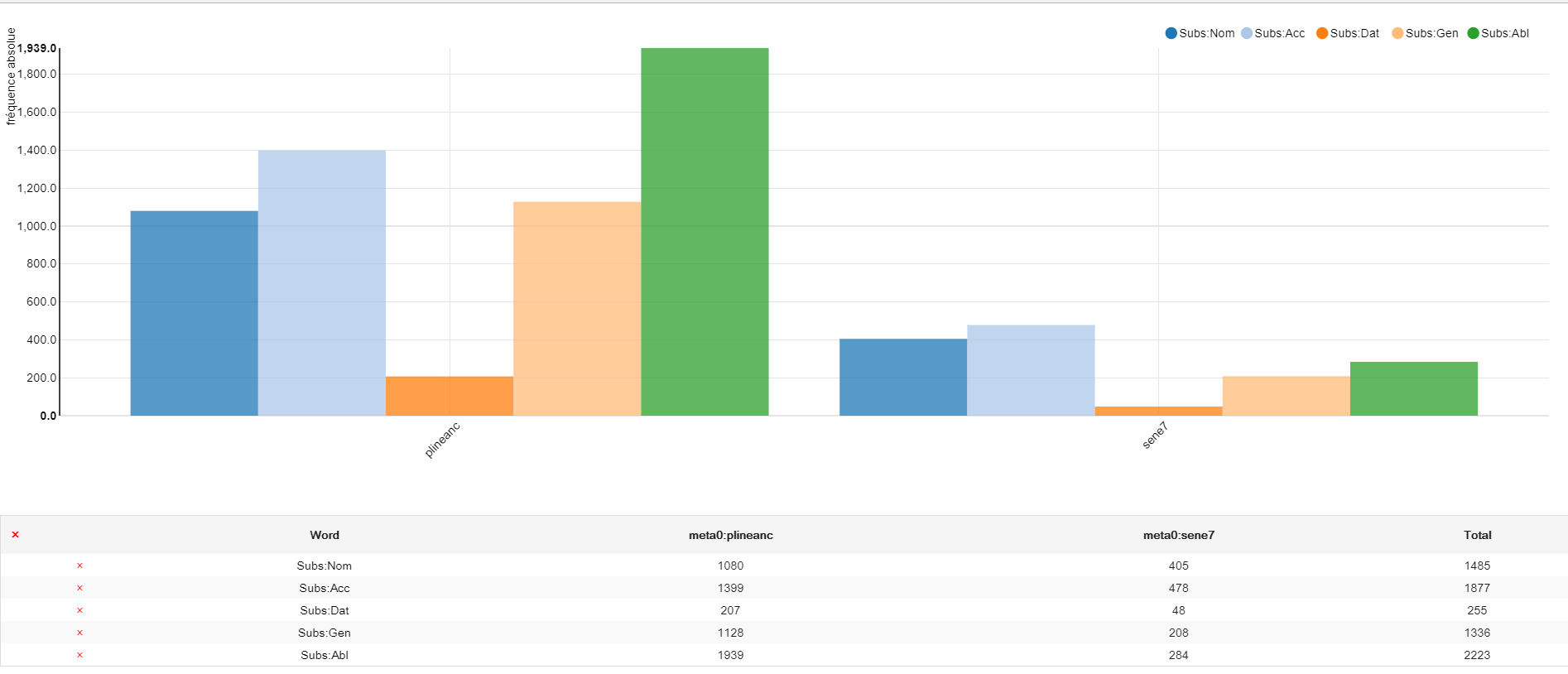
[[Figure 3 shall be inserted here]]

Figure : Histogram of the absolute frequency of the substantives distributed in the different cases in the second book of Natural History and the seventh book of the Natural Questions.

Let us compare, as an example, two paragraphs in order to sketch out the differences between the two authors. In order to identify comparable sentences, we will choose a passage in which both authors deal with the movement of rapid winds. Even though both the paragraphs treat the atmospheric phenomenon of “accidental winds,” the reasons why the subject is brought up in the text are distinct. While Pliny provides a systematic description of all of the atmospheric phenomena, which includes typhoons, etc., Seneca, focuses on the confutation of Epigenes theory of the origin of comets, which states that they might stem out of cyclones. Pliny insists on the actual description of the landscapes and the natural elements that cause the formation of storms; Seneca’s paragraph is, on the contrary, focused on the necessity of showing that the evolution of cyclones prevents the possibility that comets might originate from them:

Dicebam modo non posse diu uerticem permanere nec supra lunam aut usque in stellarum locum crescere. Nempe efficit turbinem plurium uentorum inter ipsos luctatio. Haec diu non potest esse: nam cum uagus et incertus spiritus conuolutatus est, nouissime uni uis omnium cedit. 3. Nulla autem tempestas magna perdurat: procellae, quanto plus habent uirium, tanto minus temporis; uenti, cum ad summum uenerunt, remittuntur. Omnia uiolenta necesse est ipsa concitatione in exitum sui tendant. Nemo itaque turbinem toto die uidit, ne hora quidem. Mira uelocitas eius et mira breuitas est. Praeterea uiolentius celeriusque in terra circaque eam uoluitur; quo excelsior, eo solutior laxiorque est, et ob hoc diffunditur. 4 Adice nunc quod, etiamsi in summum pertenderet, ubi sideribus iter est, utique ab eo motu qui uniuersum trahit solueretur. Quid enim est illa conuersione mundi citatius? Haec omnium uentorum in unum congesta uis dissiparetur et terrae solida fortisque compages, nedum particula aeris torti.[[34]](#footnote-35)

Seneca Natural Questions VII, IX.2–4

Seneca’s sentences are short and tend to convey one detail at a time. Spatial connotations are not essential, since the description is carried out, not for the sake of a precise depiction of cyclones, but with the aim of affirming the evidence that cyclones do not have the right characteristics to create comets. Indeed, when a certain ‘moment’ is mentioned, or a place, it is because they represent a step of the argumentation, and are therefore expanded in a sentence (cum uagus et incertus spiritus conuolutatus est introduces the end of the cyclone; cum ad summum venerunt anticipates the dissolving of winds; ubi sideribus iter est underlines the point that stronger forces act in the higher region of the sky).

We also find in Seneca a rhetorical question, which is clearly a way of convincing the reader. Another striking element is the absence, in Seneca, of relative clauses. Now let us see now how Pliny deals with the subject:

Simili modo uentos uel potius flatus posse et arido siccoque anhelitu terrae gigni non negauerim; posse et aquis aëra exspirantibus, qui neque in nebulam densetur nec crassescat in nubes; posset et solis inpulsu agi, quoniam uentus haud aliud intellegatur quam fluctus aëris, pluribusque etiam modis. Namque et e fluminibus ac niuibus et e mari uidemus, et quidem tranquillo, et alios, quos uocant altanos, e terra consurgere; qui, cum e mari redeunt, tropaei uocantur, si pergunt, apogei. Montium uero flexus crebrique uertices et conflexa cubito aut confracta in umeros iuga, concaui uallium sinus, scindentes inequalitate ideo resultantem aëra (quae causa etiam uoces multis in locis reciprocas facit), sine fine uentos generant. Iam quidem et specus, qualis in Dalmatia ore uasto, praeceps hiatu, in quem deiecto leui pondere, quamuis tranquillo die, turbini similis emicat procella; nomen loco est Senta. Quin et in Cyrenaica prouincia rupes quaedam austro traditur sacra, quam profanum sit attrectari hominis manu, confestim austro uoluente harenas. In domibus etiam multis madefacta inclusa opacitate conceptacula auras suas habent. Adeo causa non deest.[[35]](#footnote-36)

Pliny Natural History II 114–115

Pliny’s sentences are longer, and the terms in ablative inform the reader of places, moments, and manners in which phenomena take place. Pliny also uses relative clauses in order to ‘add’ more information to the sentence: the characteristics of vapor coming out from the sea, the name of the winds, and the origin of echo. These two paragraphs show a concrete example of the phenomena reflected in the statistical data: Pliny’s sentences are long and articulated, mainly expanded by relative or participial sentences, and convey as much information as possible. Seneca prefers short sentences, without many complements, and gives specifications of place and time through subordinate sentences, factors which explain the low rate of genitives and ablatives.[[36]](#footnote-37) The importance of participles in Pliny’s prose is confirmed by statistical data. Indeed, even though Seneca employs proportionally more verbs than Pliny, as we have seen, this does not apply to the participles. The distribution tool tells us that Verb:Part is the only z-score of verbal categories which is meaningfully positive in Pliny and negative in Seneca (data are reported in the Appendix, table 4). The wide use of participles has already been mentioned,[[37]](#footnote-38) and this, combined with the previous observations, helps to sketch out the structure of the longer and less clearly articulated sentences in Pliny vs. the short and strictly argumentative sentences in the Natural Questions. It might be interesting to filter the data concerning nouns and verbs considering the notion of the absolute ablative, but this would require a deeper level of analysis that would go beyond the purpose of this article.[[38]](#footnote-39)

A last observation concerning the comparison between Seneca and Pliny deals with the distribution of the invariable parts of speech in the two texts. This important field of research has already proven to be very enlightening about the purposes and the nature of the texts considered,[[39]](#footnote-40) and has been successfully applied to the Natural History.[[40]](#footnote-41) With this in mind, we will explore what kind of adverbs and coordinating conjunctions are found more frequently in the two texts. For this we will need to combine the morphological and the semantic registers, using the distribution tool to determine the z-score of the 30 more frequent adverbs in the two texts (Figure 4). As we know, this category is less defined in Latin syntax, and it includes words with very different functions[[41]](#footnote-42).

[[Figure 4 shall be inserted here]]

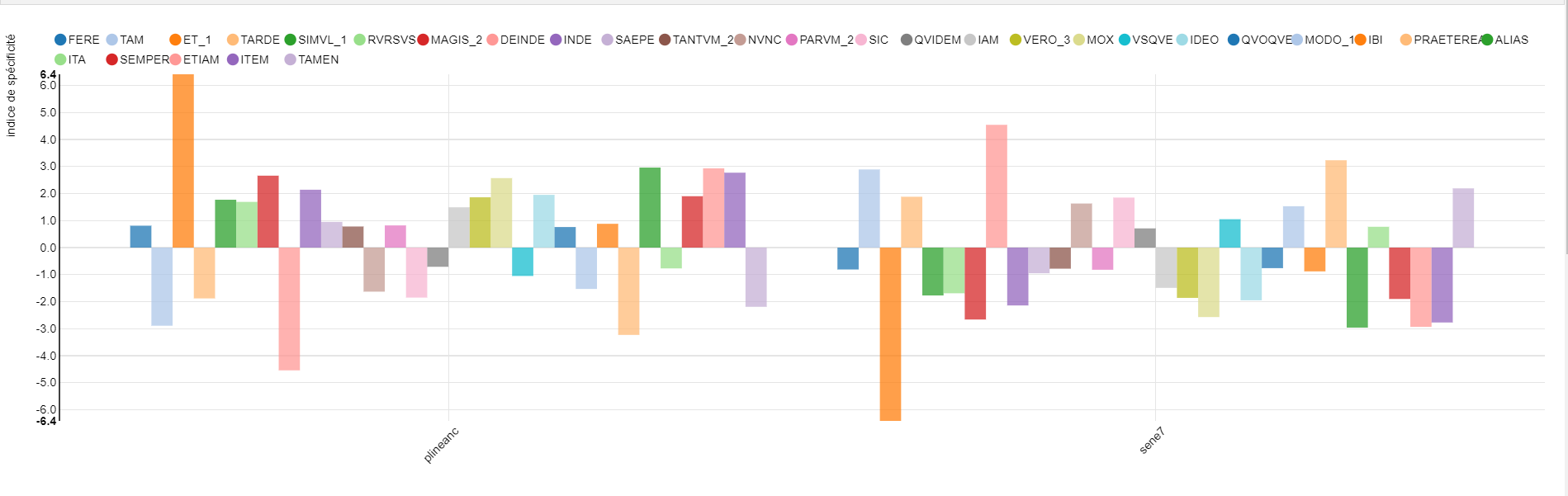


Figure : Histogram of the z-score of the 30 most frequent adverbs in the second book the Natural History and the seventh book of the Natural Questions.

Pliny makes an intensive use of the adverbial et (and, in a less marked way, of etiam), while deinde has a positive z-score in Seneca. The use of the adverbial et seems to be a typical feature of Natural History II (at least), since, even when compared with other authors, we still find a positive z-score (16. 1), as briefly shown by the data reported in the Appendix (table 6[[42]](#footnote-43)).

The TLL proposes four categories to describe how the adverbial et can be used: additive, cumulative, iuncturae, singularia. Because the research tool enables search for a specific lemma, we can find directly all occurrences of the adverbial et (LEM: ET\_1) in Pliny’s text. Many examples of et follow a conjunction or an adverb. In this case the first particle determines the ‘role’ of the added element: for instance, in sed et (uero et) the added element contrasts, somewhat, with the previous one,[[43]](#footnote-44) or quin et adds an element that emphasizes what has just been written;[[44]](#footnote-45) ideo et announces that the added element is a consequence of what was stated in the previous sentence.[[45]](#footnote-46) Contrary to Pliny’s usage, we never find such a combination of words in Seneca. The TLL considers iuncturae to include also the group ‘et + possessive pronoun/adjective’, which we found in our text. This group is specialized in the expression of one concept: that an event, phenomenon, took place also in contemporary times:[[46]](#footnote-47)

Nam ut XV diebus utrumque sidus quaereretur, et nostro aeuo accidit, imperatoribus Vespasianis patre III. filio II. consulibus[[47]](#footnote-48).

Non minus mirum ostentum et nostra cognouit aetas anno Neronis principis supremo [...][[48]](#footnote-49).

Amnes retro fluere et nostra uidit aetas Neronis principis supremis[[49]](#footnote-50).

This schema is interesting because it reinforces the credibility and meaning of the notions just described: that Pliny’s epoch also witnessed such events, on one side, stands as a proof of what he says, and, on the other, cues the reader to consider such ‘abstract’ material in the frame of his own experience.

Another recurring expression, used to express always the same information, is the linguistic group unde et:

Martis stella, ut proprior, etiam ex quadrato sentit radios, a XC partibus, unde et nomen accepit motus primus et secundus nonagenarius dictus ab utroque exortu[[50]](#footnote-51).

Percussae in qua diximus parte et triangulo solis radio inhibentur rectum agere cursum et ignea ui leuantur in sublime; hoc non protinus intellegi potest uisu nostro, ideoque existimantur stare, unde et nomen accepit statio[[51]](#footnote-52).

In Falisco omnis aqua pota cadidos boues facit, [...] rursus nigras Penius rufasque iuxta Ilium Xanthus, unde et nomen amni[[52]](#footnote-53).

Pliny always uses unde et to introduce an etymology, adding information (the name), then linking it to the previous phrase by means of the etymology. His intent is similar to what we have already seen: the name of the phenomenon (or the river) guarantees the validity of what has just been said and links the information to something familiar to the reader. This expression is typical of the entirety of Natural History, representing therefore a “plinian feature” (we found the expression unde et nomen at Natural History IV 65, V 73, VIII 218, etc., especially in botanical books[[53]](#footnote-54)). While Pliny is the first to employ the expression unde et nomen to introduce an etymology, the expression would subsequently be used by different authors (Cyprian, Ambrose, Augustine, Cassiodorus), and it is regularly found in Isidorus’ Etymologiae. We see therefore how the high rate of adverbial et is partly explained by the use of some recurring expressions necessary to Pliny’s informative aim. The expressions become part of the technical language, a kind of formula for introducing certain information.

In the remaining occurrences, Pliny mainly uses the adverbial et “additive”, as described in the TLL: in this case it can be substituted by etiam. For instance, after having listed some ty of solar and lunar eclipses, he adds that we have even more information about those phenomena: Intra ducentos annos Hipparchi sagacitate compertum est et lunae defectum aliquando quinto mense a priore fieri [...].[[54]](#footnote-55) We also find examples of the et used cumulative, vel c. augendi notione i. q. ‘vel’,[[55]](#footnote-56) as in the sentence: Veneris tantum stella excedit eum binis partibus, quae causa intellegitur efficere ut quaedam animalia et in desertis mundi nascantur.[[56]](#footnote-57)

The importance of the use of this et is particularly evident in the section already quoted in 2.1 above (Natural History II, 114–115), where we find three times in a row the expression posse et; in the previous paragraph, not quoted here, the same expression is repeated another three times. Posse et first introduces three possible origins of lightning and thunderbolts, then, in our paragraphs, three possible causes of winds. This anaphora gives a stylistic connotation to these lists of different possibilities, and, at the same time, indicates to the reader that they are all equally probable in Pliny’s eyes. We see here how the analysis of a linguistic feature must take into consideration also stylistic aspects: dealing with a long list of elements, Pliny uses the additive et because he is not interested in distinguishing the elements in a hierarchy of importance; however, he somehow creates an anaphoric effect, giving a sort of rhythm to the section, and orienting the reader by clearly announcing every new item of the list.

To conclude, the necessity of informing the reader of all the available knowledge requires the wide use of the adverbial et, which helps present multiple elements without establishing a specific link or hierarchy between them. However, this necessity is transformed by Pliny into a stylistic feature, spanning from the creation of fixed adverbial expressions, dedicated to one specific role, to anaphorae that guide the reader through the structure of the text.

3. Conclusions: Natural History II, Natural Questions VII, and other literary genres

The data analyzed up to this point show how Pliny’s second book of the Natural History and Seneca’s seventh book of the Natural Questions differ from one another. It might be interesting, however, to position them in a larger context: when confronted with texts of other literary genres, do the differences between them level off (because their scientific aspects remain prevalent), or do their differences stay meaningful, showing that other elements (such as style and authorial intent) are more important? A quick way to check might be to insert these two texts in the corpus already mentioned in the Introduction,[[57]](#footnote-58) and, using AFC and tree-analysis, see which texts are similar. The corpus has been created to represent texts in the main prose literary genres (historiography, biography, novel, philosophy, rhetoric, epistolography, scientific and technical prose) and didactic and scientific poems; authors of the republican period accompany authors of the ‘silver age’ in order to give a faithful description of the eventual chronological differences. Comparing, for instance, Pliny the Younger’s letters with Seneca’s *Consolationes* and Cicero’s discourse could uncover whether directly addressing the reader has consequences for the structure of sentences, and show how distant Pliny is from this dialogical structure. This could then refine the concept of Encyclopaedia and the educational aspects of the Natural History. The presence of Cato’s De agricultura could serve the purpose of comparing our astronomical text with Latin technical literature.

The texts that have been chosen are comparable in size to Pliny’s Natural History II and Seneca’s Natural Questions VII. This helps keep the focus on the influence of the genre on the style, and prevents losing pertinent data in data sets that are too large and that exceed the parameters of our study.

When the distance between the partitions of the corpus is evaluated on the grounds of lemmata, forms, and codes present in the texts,[[58]](#footnote-59) the scientific texts we have selected are grouped together, and the thematic division of the database is coherent with the subjects of the works (Figure 5):

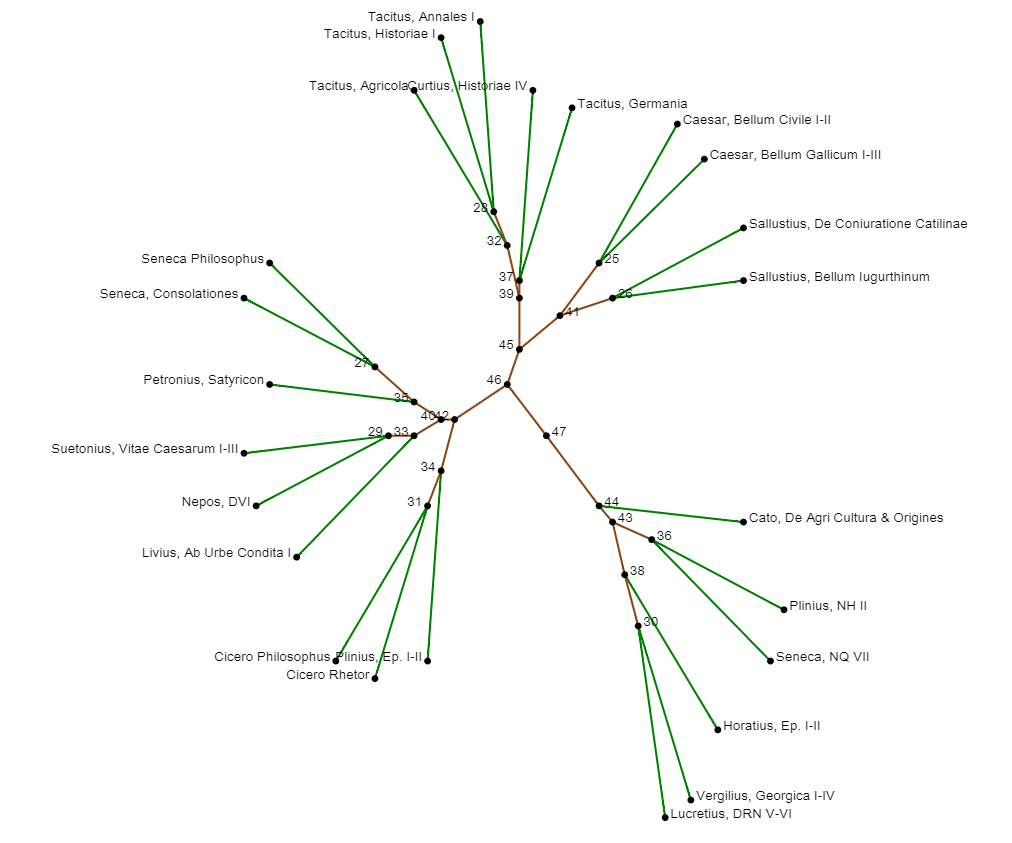
[[Figure 5 here]]

Figure : Tree analysis based on the distribution of codes, forms, and lemmata on the corpus constituted by various Latin texts.

Focusing only on the parts of speech, i.e. only on the codes, the tree-analysis gives a completely different result (Figure 6):

[[Insert here Figure 6]]

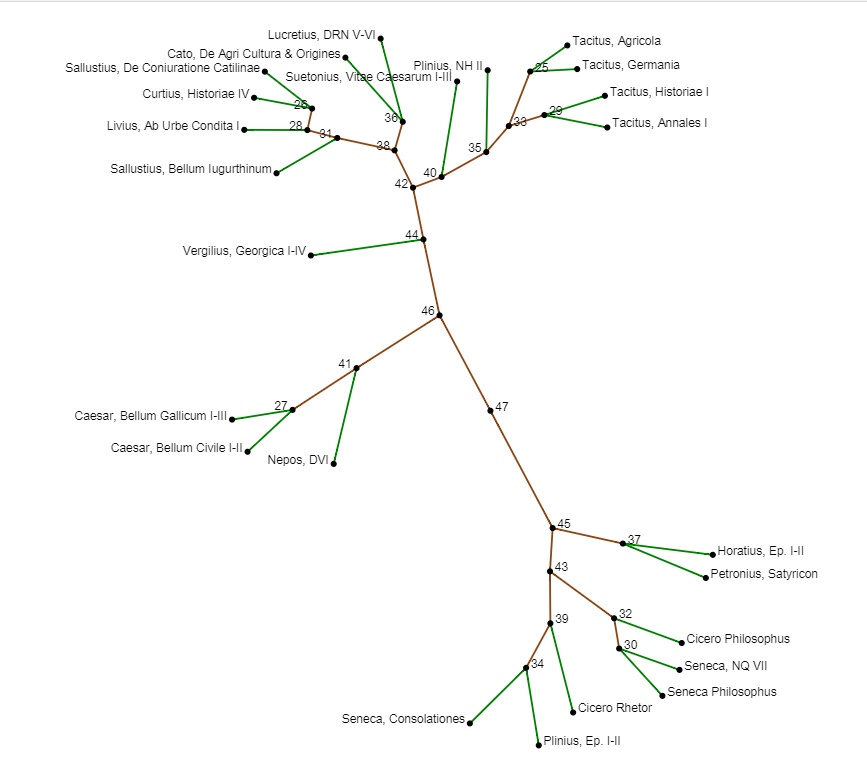


Figure : Tree analysis on the same corpus as Figure 5, based only on the z-score of the parts of speech.

The tree is very clearly divided in two parts: on one side, we have historical texts (Tacitus, Livy, Curtius, Sallust, Suetonius, Cornelius Nepos Origins) and scientific ones (Natural History, On Agriculture, On the Nature of Things), on the other side, philosophical, rhetorical, and epistolary texts, with the addition of Natural Questions VII. Caesar, Cornelius Nepos, and Virgil, even if closer to other historical and scientific texts, form an isolated unit. Such a sharp division hints at two different considerations.

First, the approaches used by Seneca (argumentation, debate, hypotheses) influence the language of the book in a stronger way than its subject matter, aligning it, not only to other works by Seneca, but also to philosophical texts written by other authors. This underlines the importance of considering authorial intent when analyzing a work, since many linguistic features may be explained simply by the proximity to a literary genre that is not necessarily suggested by the contents of that work. Applying this reasoning to Pliny’s works reconfirms the point that Pliny’s language cannot be considered as purely technical, since, as we have seen, the intent of Pliny’s encyclopaedia goes beyond the intent of a focused technical text.

The second point implies the re-consideration of Pliny’s title Natural History. We have already mentioned the complexity behind the term Historia, but an interesting analysis carried out by P. Jal[[59]](#footnote-60) underlines how the title might hint at a new conception of history:

[...] écrire une histoire nouvelle, plus pratique, embrassant l’ensemble des activités humaines et voulant faire connaître le plus grand nombre des aspects et des manifestations de la nature dans laquelle nous vivons.[[60]](#footnote-61)

We have seen that the linguistic features that we have analyzed are often explained by the necessity to convey as much information as possible drawn from a wide range of sources; to provide the reader with non-hierarchical data, leaving them to the users’ interpretation; to link ‘remote material’ to the readers’ present reality. These concerns were partly shared by historians; on the contrary, both philosophers and rhetoricians are interested in convincing the audience through a precise process of reasoning, which brings strict organization to the speech or selection of material. This AFC (Figure 7), indeed, shows that the elements which characterize the oppositions along the principal axis are those mentioned before: for historical texts and the Natural Questions, substantives (which have an important weight in determining the first axis, 30%), adjectives, adverbs, coordinating particles; for philosophical and rhetorical texts, verbs and all the elements providing an interaction with the reader (interrogative particles, interjections, etc.)

[[Figure 7 shall be inserted here]]

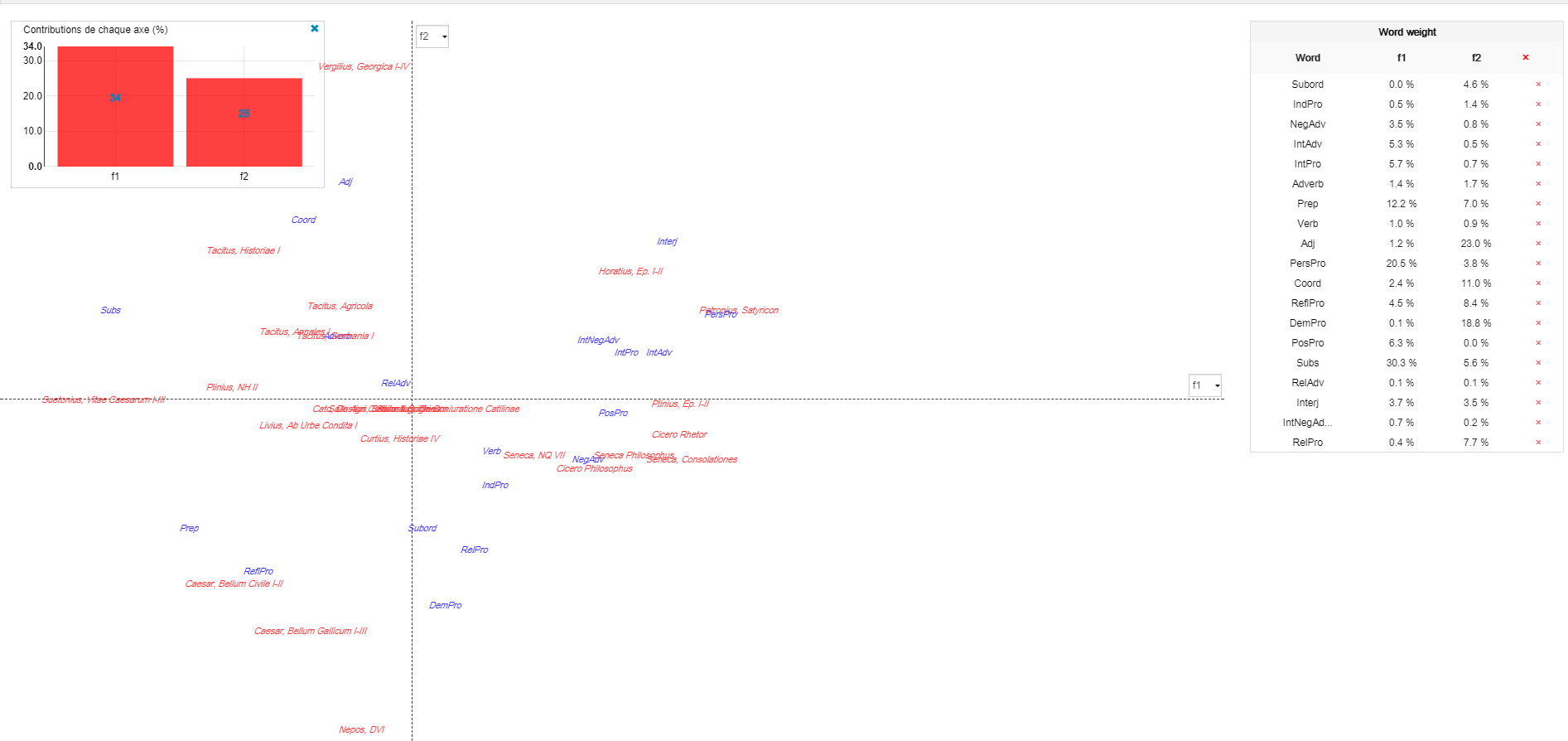


Figure : AFC of the parts of speech on the corpus formed by various Latin texts.

Of course, such indications encourage further study, but this is an interesting example of a fruitful combination of statistical and linguistic analyses, applied to an author whose place in the scholarship has been strongly compromised by a complete misrepresentation of the complex and peculiar nature of his text.

The information generated in the examples above would not have been available to even the most attentive reader: the data points considered are too numerous to be maintained mentally, and this sharp division is particularly meaningful because it can take into account all grammatical categories. This prevents the risk of focusing only on a singular aspect and potentially neglecting meaningful pieces of information. Of course, without a closer reading of the text, datum cannot be fully interpreted. However, the objective confirmation provided by numbers is necessary to give the right direction to further investigation. Therefore, statistical analysis can be a powerful tool for the scholar. Indeed, it not only confirms or belies intuitions, but also reveals information that would not be available with traditional linguistic analysis.

Bibliography

Alfonsi, L., and A. Ronconi, eds. 1982. Plinio il Vecchio sotto il profilo storico e letterario. Atti del Convegno di Como, 5–6–7 ottobre 1979. Atti della Tavola rotonda nella ricorrenza centenaria della morte di Plinio il Vecchio, Bologna 16 dicembre 1979. Como.

Beajeu, J. 1982. “La langue de l'astronomie dans l'Histoire naturelle de Pline l'Ancien.” In Alfonsi and Ronconi 1982:83–95.

Benzécri, J. P. 1983–1987. Pratique de l’analyse des données. 3 vols. Paris.

Braccesi, L. “Plinio storico.” In Alfonsi and Ronconi 1982:53–82.

Brunet, E. 2003. “Peut-on mesurer la distance entre deux textes?” Corpus 2:47–70.

Capponi, F. 1991. “Per uno studio della sintassi pliniana.” Sileno 17:171–184.

Conte, G. B. 1982. “L'inventario del mondo. Ordine e linguaggio della natura nell'opera di Plinio il Vecchio.” In Gaio Plinio Secondo. Storia Naturale 1: cosmologia e geografia, eds. G. B. Conte, A. Barchiesi, and G. Ranucci, xvii–xlvii. Turin.

Corcoran, T. H. 1972. Seneca. Naturales Quaestiones II. Cambridge, MA.

Cova, P. V. 1986. “L’ablativo assoluto nella Naturalis Historia. Studi sulla lingua di Plinio il Vecchio.” In Studi sulla lingua di Plinio il Vecchio, eds. P. V. Cova, R. Gazich, G. E. Manzoni, and G. Melzani, 13–142. Milano.

Denooz, J. 2007. “Langues classiques et informatique: les premiers travaux.” Journal of Latin Linguistics 9.3:11–20.

Gaillard, F. 1904. De breviloquentia pliniana Quaestiones. Marburg.

Grasberger, L. 1860. De usu Pliniano. Würzburg.

Gross, N. 1989. Senecas Naturales Quaestiones. Komposition, naturphilosophische Assaugen une ihre Quellen. Stuttgart.

Healy, J. F. 1987. “The Language and Style of Pliny the Elder.” In Filologia e forme letterarie. Studi offerti a Francesco Della Corte IV, 1–27. Urbino.

Hine, H. M. 1996. L. Annaeus Seneca. Naturalium Quaestionum libri. Stuttgart.

Jal, P. 1987. “Pline et l'historiographie latine.” Helmantica 28:171–186.

Kroon, C. 2004. “Scales of involvement and the use of Latin casual connectives.” In Registros lingüísticos en las lenguas clásicas, eds. Eire and A. Ramos Guerreira, 65–86. Salamanca.

———. 2011. “Latin Particles and the Grammar of Discourse.” In A companion to the Latin language, ed. J. Clackson, 176–195. Oxford.

Lausdei, C. 1987. “Problemi esegetici e linguistici de la Naturalis Historia.” Helmantica 38:255–271.

Lébart, L., and A. Salem. 1994. Statistique textuelle. Paris.

Longrée, D., G. Purnelle, and C. Philippart De Foy. 2010. “Structures phrastiques et analyse automatique des données morphosyntaxiques : le projet LatSynt.” In Proceedings of the JADT 2010, 10th International Conference on the Statistical Analysis of Textual Data, Rome, 9–11 June 2010, ed. S. Bolasco, 433–441. Rome.

Longrée, D., and S. Mellet. 2012. “Asymétrie de la coocurrence et contextualisation. Le rôle de la flexion casuelle dans la structuration des réseaux coocurrentiels d’un mot-pôle en latin.” Corpus 11:1–31.

Longrée, D., and G. Purnelle. 2014. “Le programme ‘LatSynt’, un instrument pour mesurer des distances entre prosateurs latins classiques ?” LEC 82:47–63.

Luong, X., M. Juillard, S. Mellet, and D. Longrée. 2007. “Trees and after: The concept of text topology: Some applications to verb-form distributions in language corpora.” In Literary and Linguistic Computing 22.2:167–186.

Mayhoff, C. 1906. C. Plini Secundi Naturalis Historiae libri XXXVII: libri I–IV. Leipzig.

Müller, J. 1883. Der Stil des älteren Plinius. Innsbruck.

Muller, C. 1977. Initiation aux méthodes de la statistique linguistique. Paris.

Née, E. 2017. Méthodes et outils informatiques pour l’analyse du discours. Rennes.

Norden, E. 1909. Die antike Kunstprosa. Leipzig.

Önnerfors, A. 1956. Pliniana. Plinii Maioris Naturalis Historia. Studia Grammatica, Semantica, Critica. Uppsala.

Philippart de Foy, C. 2014. LASLA – Nouveau manuel de lemmatisation du latin. http://cipl93.philo.ulg.ac.be/LaslaEncodingInitiative/Files/manlem.pdf

Pinkster, H. 2005. “The language of Pliny the Elder.” In Aspects of the language of latin prose, eds. T. Reinhardt, N. Lapidge, and J. N. Adams, 239–256. Oxford.

Pinkster, H. 2015. Oxford Latin syntax. Oxford.

Poudat, C., and F. Landragin. 2017. Explorer un corpus textuel. Méthodes – pratiques – outils. Louvain-la-Neuve.

Rackham, H. 1938. Pliny. Natural History I. Cambridge, MA.

———. 1940. Pliny. Natural History III. Cambridge, MA.

Ramos-Maldonado, S. 2002–2002. “Plinio el Viejo y Séneca ante la naturaleza.”. Excerpta Philologica 10–12:391–404.

Rehm, A. 1975. “Die siebente Buch der Naturales Quaestiones des Seneca und die Kometentheorie des Poseidonios.” In Seneca als Philosoph, ed. G. Maurach, 228–263. Darmstadt.

Schilling, R. 1978. “La place de Pline l’Ancien dans la littérature technique.” Revue de Philologie 52.2:272–283.

Täckholm, U. 1952. “Zum Gebrauch des Genitivus Qualitatis in Plinius' Naturalis Historia.” Eranos 50:98–110

Waiblinger, F. P. 1977. Senecas Naturales Quaestiones. Grieschische Wissenschaft und römische Form. Munich.

Appendix: statistical data

1. Distribution (z-score) of parts of speech between Natural History II and Natural Questions VII (Figure 1).

|  |  |  |
| --- | --- | --- |
| Word | meta0:plineanc | meta0:sene7 |
| [RelAdv](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=RelAdv) | 0.86 | -0.86 |
| [Subs](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs) | 16.34 | -16.34 |
| [Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Adverb) | 3.77 | -3.77 |
| [IntAdv](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=IntAdv) | -8.83 | 8.83 |
| [PersPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=PersPro) | -5.8 | 5.8 |
| [IntPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=IntPro) | -5.9 | 5.9 |
| [Subord](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subord) | -4.9 | 4.9 |
| [Coord](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Coord) | -2.28 | 2.28 |
| [IntNegAdv](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=IntNegAdv) | -1.1 | 1.1 |
| [IndPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=IndPro) | -4.36 | 4.36 |
| [Verb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb) | -9.6 | 9.6 |
| [RelPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=RelPro) | -4.62 | 4.62 |
| [Interj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Interj) | -0.77 | 0.77 |
| [Prep](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Prep) | 4.42 | -4.42 |
| [ReflPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=ReflPro) | -3.61 | 3.61 |
| [RPosPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=RPosPro) | -3.89 | 3.89 |
| [Num](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Num) | 11.55 | -11.55 |
| [Adj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Adj) | 2.09 | -2.09 |
| [PosPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=PosPro) | -1.46 | 1.46 |
| [DemPro](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=DemPro) | -4.51 | 4.51 |

2. Distribution (z-score) of nouns categories between Natural History II and Natural Questions VII (Figure 2).

|  |  |  |
| --- | --- | --- |
| Word | meta0:plineanc | meta0:sene7 |
| [Subs:Voc](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Voc) | 0.32 | -0.32 |
| [Subs:Nom](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Nom) | -1.31 | 1.31 |
| [Subs:Neutre](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Neutre) | 0.57 | -0.57 |
| [Subs:Abl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Abl) | 17.78 | -17.78 |
| [Subs:Commun](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Commun) | 0.81 | -0.81 |
| [Subs:MascFem](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:MascFem) | 0.32 | -0.32 |
| [Subs:MascNeutr](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:MascNeutr) | 1.02 | -1.02 |
| [Subs:5Decl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:5Decl) | 1.67 | -1.67 |
| [Subs:1Decl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:1Decl) | 5.39 | -5.39 |
| [Subs:2Decl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:2Decl) | 9.3 | -9.3 |
| [Subs:Plur](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Plur) | 8.14 | -8.14 |
| [Subs:Masculin](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Masculin) | 0.57 | -0.57 |
| [Subs:Sing](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Sing) | 12.23 | -12.23 |
| [Subs:Dat](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Dat) | 2.95 | -2.95 |
| [Subs:Acc](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Acc) | 1.25 | -1.25 |
| [Subs:Feminin](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Feminin) | 1.22 | -1.22 |
| [Subs:Gen](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Gen) | 10.04 | -10.04 |
| [Subs:3Decl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:3Decl) | 6.59 | -6.59 |
| [Subs:4Decl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:4Decl) | 6.39 | -6.39 |
| [Subs:GrDecl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:GrDecl) | 1.04 | -1.04 |
| [Subs:AnomDecl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:AnomDecl) | 2 | -2 |

3. Distribution (absolute frequency) of nouns cases between Natural History II and Natural Questions VII (Figure 3).

|  |  |  |  |
| --- | --- | --- | --- |
| Word | meta0:plineanc | meta0:sene7 | Total |
| [Subs:Nom](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Nom) | 1080 | 405 | 1485 |
| [Subs:Acc](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Acc) | 1399 | 478 | 1877 |
| [Subs:Dat](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Dat) | 207 | 48 | 255 |
| [Subs:Gen](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Gen) | 1128 | 208 | 1336 |
| [Subs:Abl](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Subs:Abl) | 1939 | 284 | 2223 |

4. Distribution (z-score) of verbal categories between Natural History II and Natural Questions VII.

|  |  |  |
| --- | --- | --- |
| Word | meta0:plineanc | meta0:sene7 |
| [Verb:PqPerfPeri](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:PqPerfPeri) | 0 | 0 |
| [Verb:Sup-u](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Sup-u) | 2.13 | -2.13 |
| [Verb:Inf](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Inf) | -1.56 | 1.56 |
| [Verb:Act](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Act) | -13.53 | 13.53 |
| [Verb:Ind](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Ind) | -13.41 | 13.41 |
| [Verb:VerbAdj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:VerbAdj) | -1.17 | 1.17 |
| [Verb:Subj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Subj) | -7.37 | 7.37 |
| [Verb:1Conj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:1Conj) | 2.1 | -2.1 |
| [Verb:AnomConj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:AnomConj) | -10.54 | 10.54 |
| [Verb:Dep](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Dep) | 1.78 | -1.78 |
| [Verb:Perf](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Perf) | 1.09 | -1.09 |
| [Verb:Pres](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Pres) | -8.78 | 8.78 |
| [Verb:3Conj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:3Conj) | -2.74 | 2.74 |
| [Verb:Imper](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Imper) | -2.04 | 2.04 |
| [Verb:MixConj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:MixConj) | -1.12 | 1.12 |
| [Verb:2Conj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:2Conj) | -7.38 | 7.38 |
| [Verb:Imp](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Imp) | -6.85 | 6.85 |
| [Verb:PqPerf](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:PqPerf) | -1.1 | 1.1 |
| [Verb:Fut](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Fut) | -4.37 | 4.37 |
| [Verb:4Conj](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:4Conj) | -1.21 | 1.21 |
| [Verb:FutPerfPeri](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:FutPerfPeri) | 0 | 0 |
| [Verb:2Pers](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:2Pers) | -5.88 | 5.88 |
| [Verb:Part](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Part) | 8.58 | -8.58 |
| [Verb:3Pers](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:3Pers) | -14.33 | 14.33 |
| [Verb:Gerund](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Gerund) | 1.26 | -1.26 |
| [Verb:PerfPeri](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:PerfPeri) | 0.32 | -0.32 |
| [Verb:1Pers:1Pers](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:1Pers:1Pers) | -4.73 | 4.73 |
| [Verb:Sup-um:Sup-um](file:///C:\\hyperbase\\core\\controller\\action\\show_spec.php%3fword=Verb:Sup-um:Sup-um" \o "Verb:Sup-um:Sup-um) | 0.32 | -0.32 |
| [Verb:SemiDep](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:SemiDep) | -0.99 | 0.99 |
| [Verb:FutPerf](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:FutPerf) | -2.04 | 2.04 |
| [Verb:Pas](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=Verb:Pas) | 2.98 | -2.98 |

5. Distribution (z-score) of 30 more frequent adverbial lemmata between Natural History II and Natural Questions VII (Figure 4).

|  |  |  |
| --- | --- | --- |
| Word | meta0:plineanc | meta0:sene7 |
| [LEM:FERE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:FERE:Adverb) | 0.81 | -0.81 |
| [LEM:TAM:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:TAM:Adverb) | -2.89 | 2.89 |
| [LEM:ET\_1:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:ET_1:Adverb) | 6.41 | -6.41 |
| [LEM:TARDE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:TARDE:Adverb) | -1.88 | 1.88 |
| [LEM:SIMVL\_1:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:SIMVL_1:Adverb) | 1.77 | -1.77 |
| [LEM:RVRSVS:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:RVRSVS:Adverb) | 1.69 | -1.69 |
| [LEM:MAGIS\_2:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:MAGIS_2:Adverb) | 2.66 | -2.66 |
| [LEM:DEINDE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:DEINDE:Adverb) | -4.54 | 4.54 |
| [LEM:INDE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:INDE:Adverb) | 2.14 | -2.14 |
| [LEM:SAEPE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:SAEPE:Adverb) | 0.95 | -0.95 |
| [LEM:TANTVM\_2:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:TANTVM_2:Adverb) | 0.78 | -0.78 |
| [LEM:NVNC:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:NVNC:Adverb) | -1.63 | 1.63 |
| [LEM:PARVM\_2:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:PARVM_2:Adverb) | 0.82 | -0.82 |
| [LEM:SIC:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:SIC:Adverb) | -1.85 | 1.85 |
| [LEM:QVIDEM:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:QVIDEM:Adverb) | -0.71 | 0.71 |
| [LEM:IAM:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:IAM:Adverb) | 1.49 | -1.49 |
| [LEM:VERO\_3:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:VERO_3:Adverb) | 1.86 | -1.86 |
| [LEM:MOX:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:MOX:Adverb) | 2.57 | -2.57 |
| [LEM:VSQVE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:VSQVE:Adverb) | -1.05 | 1.05 |
| [LEM:IDEO:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:IDEO:Adverb) | 1.95 | -1.95 |
| [LEM:QVOQVE:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:QVOQVE:Adverb) | 0.76 | -0.76 |
| [LEM:MODO\_1:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:MODO_1:Adverb) | -1.53 | 1.53 |
| [LEM:IBI:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:IBI:Adverb) | 0.88 | -0.88 |
| [LEM:PRAETEREA:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:PRAETEREA:Adverb) | -3.23 | 3.23 |
| [LEM:ALIAS:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:ALIAS:Adverb) | 2.96 | -2.96 |
| [LEM:ITA:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:ITA:Adverb) | -0.77 | 0.77 |
| [LEM:SEMPER:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:SEMPER:Adverb) | 1.9 | -1.9 |
| [LEM:ETIAM:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:ETIAM:Adverb) | 2.93 | -2.93 |
| [LEM:ITEM:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:ITEM:Adverb) | 2.77 | -2.77 |
| [LEM:TAMEN:Adverb](file:///C:\hyperbase\core\controller\action\show_spec.php%3fword=LEM:TAMEN:Adverb) | -2.19 | 2.19 |

6. Distribution (z-score) of adverbial et in literary database.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Plinius, Ep. I-II | 0.97 |  | Nepos, DVI | -9 |
| Horatius, Ep. I-II | -2.44 |  | Suetonius, Vitae Caesarum I-III | 12.05 |
| Curtius, Historiae IV | -1.24 |  | Sallustius, De Coniuratione Catilinae | -4.43 |
| Petronius, Satyricon | 2.49 |  | Sallustius, Bellum Iugurthinum | -6.54 |
| Cicero Rhetor | -6.1 |  | Tacitus, Agricola | -0.72 |
| Cicero Philosophus | -5.33 |  | Tacitus, Historiae I | 0.85 |
| Seneca, Consolationes | -0.88 |  | Tacitus, Annales I | -2.08 |
| Seneca Philosophus | -0.83 |  | Tacitus, Germania | 6.03 |
| Livius, Ab Urbe Condita I | 1.37 |  | Cato, De Agri Cultura & Origines | -6.49 |
| Caesar, Bellum Civile I-II | -6.18 |  | Plinius, NH II | 16.07 |
| Caesar, Bellum Gallicum I-III | -6.33 |  | Seneca, NQ VII | -1.23 |
| Vergilius, Georgica I-IV | 0.86 |  | Lucretius, DRN V-VI | -5.13 |

1. Grasberger 1860. [↑](#footnote-ref-1)
2. Müller 1883. [↑](#footnote-ref-2)
3. Gaillard 1904. [↑](#footnote-ref-3)
4. The most famous is certainly Norden’s: “Sein Werk gehört, stilistisch betrachtet, zu den schlechtesten, die wir haben” (1909:314). [↑](#footnote-ref-4)
5. Önnerfors 1956. [↑](#footnote-ref-5)
6. Beaujeu 1979. [↑](#footnote-ref-6)
7. Healy 1987. [↑](#footnote-ref-7)
8. Capponi 1991. [↑](#footnote-ref-8)
9. As an example, Täckholm 1952 or Serbat 1973. [↑](#footnote-ref-9)
10. Pinkster 2005. [↑](#footnote-ref-10)
11. Pliny’s and Seneca’s texts will be quoted from the Teubner editions, Mayhoff 1906 and Hine 1996. [↑](#footnote-ref-11)
12. Schilling 1978. [↑](#footnote-ref-12)
13. Önnerfors 1956. [↑](#footnote-ref-13)
14. Pinkster 2005. [↑](#footnote-ref-14)
15. Évrard and Mellet 1998. [↑](#footnote-ref-15)
16. For a history of the foundation of the L.A.S.L.A., in the frame of the evolution of statistical studies on ancient languages, cf. Denooz 2007:11–20. [↑](#footnote-ref-16)
17. Philippart de Foy 2014. [↑](#footnote-ref-18)
18. For a clearer description of the information stored in L.A.S.L.A. lemmatized texts, Cf. Longrée et al. 2010. [↑](#footnote-ref-19)
19. Cf. for instance the Website Opera Latina (http://cipl93.philo.ulg.ac.be/OperaLatina/), and the related bibliography (Denooz 2007), or the syntactic parser LatSynt (Longrée and Purnelle 2014). [↑](#footnote-ref-20)
20. The basic reference for the application of statistics to textual studies is Muller 1977, followed by Lebart and Salem 1994. [↑](#footnote-ref-21)
21. Poudat and Landragin 2017. Even more recent, Née 2017. [↑](#footnote-ref-22)
22. Cf. Poudat and Langin 2017:184–200 on the importance of this tool for the study of concordance and repeated sequences. [↑](#footnote-ref-23)
23. An example for clarification: looking for the lemmata co-ocurrent in a sentence with a certain form (for instance exercitum) will tell us which lemmata tend to appear ‘frequently’ (this term hides the more complex concept of specificity) in the same phrase as the word exercitum. We might expect for instance the lemma DVX or the lemma VINCERE as a result. If the user applies the filter of nouns, the calculations will take into consideration only the nouns appearing in the sentences. [↑](#footnote-ref-24)
24. An example: searching for the lemmata co-occurrent with the lemma PRINCEPS\_1 in Seneca’s work, we see that the lemma CLEMENTIA is a co-occurrent and that the lemma MITIS is co-occurrent with the pairing PRINCEPS\_1-CLEMENTIA. For a general explanation of the hypergeometric distribution which stands behind the calculations of co-occurrence, and for a description of HyperbaseWeb options, cf. Poudat and Landragin 2017:200–209, which also quotes the most important bibliography on this subject. For the specific case of Latin, considerations on how to use this tool (for instance the different results obtained by considering the lemma or the declined forms), cf. Longrée and Mellet 2012:1–31. [↑](#footnote-ref-25)
25. That means that, taking into the consideration the size of each text in the corpus, we are to see if one phenomenon appears more or less frequently that what we should expect considering the mean of the entire corpus; and to check if the difference between the expected value and the observed one is statistically meaningful or not. [↑](#footnote-ref-26)
26. The CA was first developed by Benzécri, who was trying to find a visual representation of possible correlations between multiple factors. It is an approximation, since it is a bidimensional projection of a tridimentional space, therefore the scholar should pay particular attention to the actual meaning of the position of the points in the graph (for a complete explanation of the factorial analysis of a corpus, see Poudat and Landragin 2017:103–115; on the CA, Poudat and Landagin 2017:115–122, and Benzécri 1983). [↑](#footnote-ref-27)
27. The tree-analysis, mainly developed by Luong, is explained in Poudat and Landragin 2017:135–140. See also Luong et al. 2007. [↑](#footnote-ref-28)
28. See, for example, Ramos-Maldonado 2000–2002:391–393. [↑](#footnote-ref-29)
29. Cf. Ramos-Maldonado 2000–2002:394–395 and Conte 1991. [↑](#footnote-ref-30)
30. On Seneca’s opinions about comets, cf. Rehm 1975. On the Natural Questions in general, cf. Gross 1989, Waiblinger 1977 and Gauly 2004. [↑](#footnote-ref-31)
31. Cf. for instance Natural History XI 8: [...] nobis propositum est naturas rerum manifestas indicare, non causas indagare dubias. (“but our purpose is to point out the manifest properties of objects, not to search for doubtful causes,” Rackham 1940: 437). [↑](#footnote-ref-32)
32. The use of the Distribution tool with only two texts shows a symmetrical result: if the z-score is positive in one author, it will be negative with the same absolute value in the other, since the corpus is only constituted by the two texts. Therefore the Distribution tool might be more interesting to use with more than two partitions of the corpus. However, the result indicates whether the different distribution of phenomena between the two texts is random or not, and this information is useful in the frame of this study. [↑](#footnote-ref-33)
33. Cf. Müller 1883, sparsim, Capponi 1991, sparsim, Täckholm 1952, Grasberger 1860:40. [↑](#footnote-ref-34)
34. “I was saying recently that turbulent air cannot continue very long nor rise above the moon or grow up to the place of the stars. Obviously, it is the struggling of many winds among themselves which produces a whirlwind. Such a struggle cannot last long. For when wandering and irregular moving air has become convoluted the force of all the winds eventually yields to one wind. Moreover, a large storm does not last. The more strength squalls have, the less time they have. Winds diminish when they reach their maximum. It needs be that all violence by its very impetuosity tends towards its own destruction. Consequently, no one has seen a whirlwind last an entire day or even an hour; its speed is startling and its brevity is amazing. Besides, on the earth and around it a whirlwind flies more violently and swiftly. The higher it is the less cohesive and compact it is, and for this reason it is dissipated. Add now the fact that even if it reached the highest region, where the stars have their path, a whirlwind would be especially broken up by the motion which carries along the universe. For what is more rapid than the turning of the world? By means of this rotation the force of all winds, even if gathered in one place, would be dissipated, and so would the solid and powerful structure of the earth, to say nothing of a little bit of twisted air” (Corcoran 1972:245–247). [↑](#footnote-ref-35)
35. 35“Similarly I am not prepared to deny that it is possible for winds or rather gusts of air to be produced also by a dry and parched breath from the earth, and also possible when bodies of water breathe out a vapour that is neither condensed into mist or solidified into clouds; and also they may be caused by the driving force of the sun, because wind is understood to be nothing else than a wave of air; and in more ways as well. For we see winds arising both from rivers and bays and from the sea even when calm, and others, called altani, arising from the land; the latter when they come back again from the sea are called turning winds, but if they go, offshore winds. The windings of mountains and their clustered peaks and ridges curved in an elbow or broken off into shoulders, and the hollow recesses of valleys, cleaving with their irregular contours the air that is consequently reflected from them (a phenomenon that in many place causes words spoken to be endlessly echoed) are productive of winds. So again are caverns, like the one with an enormous gaping mouth on the coast of Dalmatia, from which, if you throw some light object in it, even in calm weather a gust like a whirlwind bursts out; the name of the place is Senta. Also it is said that in the province of Cyrenaica there is a certain cliff, sacred to the South wind, which it is sacrilege for the hand of man to touch, the South wind immediately causing a sand-storm. Even manufactured vessels in many houses if shut up in the dark have peculiar exhalations. Thus there must be some cause for this” (Rackham 1938:116–118). [↑](#footnote-ref-36)
36. In this regard it is interesting to notice that the full stop is characteristic of Seneca’s book: indeed it has a positive z-score of 6.2. This datum of course depends heavily on the editor’s choices, but is still rather revealing. [↑](#footnote-ref-37)
37. See for instance Müller 1883:23, Capponi 1991, Lausdei 1987:261, where they are listed as a characteristic of the whole Natural History. Cf. also Pinkster 2005:248. [↑](#footnote-ref-38)
38. For an ample discussion of the role of the ablative absolute in the Natural History, cf. Cova 1986. [↑](#footnote-ref-39)
39. Kroon 2011. [↑](#footnote-ref-40)
40. Kroon 2004. [↑](#footnote-ref-41)
41. For a classification of the non-verbal and non-nominal parts of speech (the so-called particles) in Latin, see the dedicated section in Pinkster 2015:65–70. [↑](#footnote-ref-42)
42. This corpus is formed by various texts covering different literary genres: history, philosophy, rhetoric, epistolary, science, biography. More detail is given in the last section of the article. The partitions are chosen in order to be balanced with the size of Natural History II. [↑](#footnote-ref-43)
43. Cf. for instance Natural History II 22: [...] Fortuna sola invocatur ac nominatur, una accusatur, rea una agitur [...]volubilis, a plerisque vero et caeca existimata (“Fortune alone is invoked and named, alone accused, alone impeached [...], deemed volatile and indeed by most men blind as well,” Rackham 1938:183). [↑](#footnote-ref-44)
44. Cf. Natural History II 106: Nec meantium modo siderum haec uis est, sed multorum etiam adhaerentium caelo, quotiens errantium accessu inpulsa aut coniectu radiorum exstimulata sunt [...]. Quin et sua sponte quaedam statisque temporibus [...] (“Nor does this power belong to the moving stars only, but also to many of those that are fixed to the sky, whenever they are impelled forward by the approach of the planets or goaded on by the impact of the rays [...]. Indeed some stars move of themselves and at fixed times [...],”Rackham 1938:251) [↑](#footnote-ref-45)
45. Cf. Natural History II 39: Ideo et peculiaris horum siderum ratio est [...] (“Consequently the course of these stars also is peculiar,” Rackham 1938:193). Many other iuncturae are to be found in the text : unde et, nec non et, tum et, namque et, namque et, non modo, uerum et, itaque et, sic et, quoniam et. [↑](#footnote-ref-46)
46. The expression et nostra aetas is present also at Natural History II 99, even though here it is a case of correlative et: Trinos soles et antiqui saepius uidere, sicut Sp. Postumio Q. Mucio et Q. Marcio M. Porcio et M. Antonio P. Dolabella et M. Lepido L. Planco cos., et nostra aetas uidit Diuo Claudio Principe [...](“In former times three suns have often been seen at once, for example in the consulships of Spurius Postumius and Quintus Mucius, of Quintus Marcius and Marcus Porcius, of Marcus Antonius and Publius Dolabella and of Marcus Lepidus and Lucius Plancus; and our generation saw this during the principate of his late Majesty Claudius [....],” Rackham 1938:243). [↑](#footnote-ref-47)
47. Natural History II 57: “For the eclipse of both sun and moon within 15 days of each other has occurred even in our time, in the year of the third consulship of the elder Emperor Vespasian and the second consulship of the younger” (Rackham 1938:207). [↑](#footnote-ref-48)
48. Natural History II 199: “Our generation also experienced a not less marvelous manifestation in the last year of the Emperor Nero [...] (Rackham 1938:331). [↑](#footnote-ref-49)
49. Natural History II 232: “Even our generation has seen rivers flow backward at Nero’s last moments” (Rackham 1938:359). [↑](#footnote-ref-50)
50. Natural History II 60: “The planet Mars being nearer feels the sun’s rays even from its quadrature, at an angle of 90 degrees, which has given to his motion after each rising the name of ‘first’ or ‘second ninety-degree” (Rackham 1938:209). [↑](#footnote-ref-51)
51. Natural History II 70: “This cannot be directly perceived by our sight, and therefore they are thought to be stationary, which has given rise to the term ‘station’” (Rackham 1938:217). [↑](#footnote-ref-52)
52. Natural History II 230: “In the district of Falerii all the water makes oxen that drink it white [...], the Peneu again makes them black, and the river Xanthus at Ilium red, which gives the river its name” (Rackham 1938: 357). [↑](#footnote-ref-53)
53. The search for similar expressions (such as unde et [...] vocant, unde et [...] appellata, unde et [...] appellantur) increases the number of results. [↑](#footnote-ref-54)
54. Natural History II 57: “Less than 200 years ago the penetration of Hipparchus discovered that an eclipse of the moon also sometimes occurs four months after the one before” (Rackham 1938:205). [↑](#footnote-ref-55)
55. TLL, v. 5.2, c. 870. [↑](#footnote-ref-56)
56. Natural History II 66: “Only the planet Venus goes two degrees outside the zodiac; this is understood to be the reason that causes some animals to be born even in the desert places of the word” (Rackham 1938:213). [↑](#footnote-ref-57)
57. The corpus is constituted by the following partitions: Pliny the Younger Letters I and II; Horace Epistles I and II, Curtius Histories of Alexander the Great III; Petronius Satyricon; Cicero Rhetor (On his House); Cicero Philosophus (On Friendship, Cato the Elder on Old Age); Seneca Consolationes (To Marcia, To Mother Helvia, To Polybium); Seneca Philosophus (On the Shortness of Life, On the Firmness of the Wise Person, On Clemency) ; Livy History of Rome I; Caesar Civil War I and II; Caesar Gallic Wars I–III; Nepos The Lives of famous men; Suetonius The Lives of the Twelve Caesars I–III; Sallust The Conspirancy of Catiline; Sallust Iugurthine War; Tacitus The Life of Agricola; Tacitus Histories I; Tacitus Germania; Tacitus Annales I; Cato the Elder On Agriculture and Origins; Pliny Natural History II; Seneca Natural Questions VII; Virgil Georgics; Lucretius On the Nature of Things V and VI. [↑](#footnote-ref-58)
58. For a wide discussion of the notion of intertextual distance, see Brunet 2003. [↑](#footnote-ref-59)
59. Jal 1987. [↑](#footnote-ref-60)
60. Jal 1987:177. A similar opinion is expressed in Braccesi 1982:56. [↑](#footnote-ref-61)