RESEARCH ARTICLE

Are the Hexameters and Hendecasyllables of Classical Poetry a Mathematical Imprint of Short-Term Memory?

Emilio Matricciani[1,*]

[1] Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, 20133 Milan, Italy

Abstract

The hexameter of Greek and Latin epic and the Italian hendecasyllable are usually described in terms of syllables, feet and accents. This paper tests a different hypothesis: that both verses were shaped, without the poets being aware of it, by the number of words a line can hold, a number close to the central value of Miller's 7 ± 2 law. On this view the two metres are an imprint of short-term memory, the stage through which information must pass before it is stored in long-term memory and later recalled, as was required in societies where poetry was learned and recited by heart. The hypothesis is examined on Virgil's *Aeneid* and Lucretius' *De Rerum Natura*, both written in Latin hexameters, on their classical Italian translations in unrhymed hendecasyllables by Annibale Caro and Alessandro Marchetti, and on Dante's *Divina Commedia*. For each book, characters, words, sentences, punctuation marks and lines were counted; from these counts I computed word length, sentence length, the word interval (words between two punctuation marks), the number of word intervals per sentence and the number of words per line. A uniformity index is introduced to measure how evenly each parameter is distributed across books. Word length and words per line are almost constant in every text (about 6.4-6.6 words per hexameter and 6.8-7.0 per hendecasyllable), whereas sentence length and word interval change from book to book and, in the *Aeneid*, between the two hexads. The results suggest that the poets controlled punctuation and sentence structure deliberately, while the length of the line in words was set by short-term memory rather than by syllable counting alone.

Keywords: *Aeneid*, *De Rerum Natura*, hendecasyllable, hexameter, Italian poetry, linguistic parameters, Miller's law, short-term memory.

1 Introduction

In a recent paper [1], I proposed that the hendecasyllabic verse–used by Dante in the *Divina Commedia*, by Petrarch and by many other Italian authors–was most likely not devised merely by counting syllables, but by unconsciously counting words, since each line (verse) contains approximately 7 words, the central value of Miller's 7 ± 2 law [2]. The hendecasyllable, with its "alchemy" and complex rules, appears to be the product–unconscious on the author's part–of the way in which short-term memory (STM) manifests itself to the outside world.

Submitted: 28 September 2026
Accepted: 29 October 2026
Published: 10 November 2026

Vol. 1, **No.** 1, 2026.
10.62762/JMSL.2026.283501

***Corresponding author:**
✉ Emilio Matricciani
emilio.matricciani@polimi.it

Citation
Matricciani, E. (2026). Are the Hexameters and Hendecasyllables of Classical Poetry a Mathematical Imprint of Short-Term Memory?. *Journal of Mathematical Studies of Literature*, 1(1), 28–48.

© 2026 by the Author. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Another classical verse is the hexameter, used by Greek and Latin poets [3–8]. Virgil's *Aeneid* is written in dactylic hexameter, the six-foot metre of Greek epic poetry, as in Homer's *Iliad* and *Odyssey* [9]. Each of the first four feet can be either a dactyl (one long and two short syllables) or a spondee (two long syllables); the fifth foot is usually a dactyl, and the last foot is a spondee or a trochee, because the final syllable is *anceps*. Virgil used this varying mix of fast dactyls and slow spondees to match the mood and action of his narrative.

Another Latin masterwork written in hexameters is Lucretius' *De Rerum Natura* (*On the Nature of Things*), a didactic poem that explains Epicurean philosophy to a Roman audience. I have examined it because it deals with many topics that today we would call scientific, across a wide range of disciplines, so it is worth asking whether it resembles or differs from a narrative poem such as the *Aeneid*.

The aim of this paper is to show that the hexameter, too–despite its complex structure–conforms to the central value of Miller's 7 ± 2 law, just as the hendecasyllable does. My thesis is that in classical societies (and beyond), where learning and recitation relied also on memory, the "invention" of the hexameter and of the hendecasyllable was in fact a product of human STM itself. Hexameters and hendecasyllables were, and still are, its mathematical "imprint", since the passage through STM is a necessary step for storing information in long-term memory, with a view to its subsequent retrieval.

The STM is needed for storing information in long-term memory even when dealing with written texts. When writing, an author introduces pauses of different lengths–today marked by punctuation added to ancient texts at a later stage by skilled readers [10]–because, in that moment, the author is also acting as a reader and is subject to the limits of his own STM. Hexameters, and later hendecasyllables, served precisely this function. Readers interested in STM can consult the references cited in [1].

After this introductory section, Section 2 presents the *Aeneid* and *De Rerum Natura*; Section 3 recalls the deep-language parameters used in the paper; Section 4 recalls the equivalent buffers of the extended short-term memory (E-STM); Section 5 reports an exploratory data analysis of the *Aeneid*; Section 6 compares the books of the two hexads by means of a "uniformity index"; Section 7 discusses which parameters were likely consciously designed; Section 8 reports the exploratory data analysis and statistics of *De Rerum Natura*; Section 9 reports results on classical Italian translations of the two poems in hendecasyllables; Section 10 reports the uniformity index in Italian poetry; finally, Section 11 draws the conclusions. Appendix A gives information on the raw textual data and their processing; Appendix B lists the mathematical symbols; Appendix C proves a formula; Appendix D reports additional scatterplots of *De Rerum Natura*.

2 *Aeneid* and *De Rerum Natura*

I briefly introduce the two poems–without attempting to describe all their literary and poetic features, to which scholars have devoted entire libraries–and focus only on the aspects relevant to this study.

2.1 *Aeneid*

The *Aeneid*, written by Virgil (70-19 BC) between 29 and 19 BC, is the national epic of ancient Rome [11, 12]. It tells the journey of Aeneas and the Trojans from Troy to central Italy, and their trials and victories after they reached their "promised land". The poem illustrates the archaeology and topography of Latium and Rome and, through prophecies and foreshadowing, presents key events of Roman history and the achievements of the Emperor Augustus (63 BC-AD 14), a contemporary of Virgil [13].

The *Aeneid* is divided symmetrically into two parts. The first part–books 1-6, the first hexad–is inspired by the *Odyssey*: it recounts Aeneas' flight from Troy and his wanderings, his tragic love for Dido and his descent into the Underworld. The second hexad–books 7-12–is inspired by the *Iliad*: it recounts the landing in Latium and Aeneas' war against the local peoples to conquer his new homeland.

This bipartition is a sort of *Odyssey* followed by an *Iliad*, in reverse order. The first hexad is about adventure and the discovery of Aeneas' destiny; the second about armed conflict and the fulfilment of the mission, the founding of a new Troy, the future Rome. Style and themes shift accordingly, while the quality of the language remains high throughout.

Books 1-6 (the "Odyssean" part) have a more melancholic and introspective tone. The language dwells on wandering, the unknown and personal emotion, and the subjective, psychological element prevails: think of the love and suicide of Dido in book 4, or of the descent into the Underworld in book 6.

Books 7-12 (the "Iliadic" part) move to an epic, martial register. The language becomes denser in archaic Latin terms, detailed descriptions of battle and military vocabulary, and the narration is more detached. Battles, catalogues of local heroes and duels prevail, and Aeneas sheds the role of the grieving refugee to become the predestined commander.

Mathematical and statistical analyses of the *Aeneid* have been based mostly on word frequency, such as the data of the Dickinson College Commentary (https://dcc.dickinson.edu/vergil-aeneid/vergil-caesar-vocab-data, last accessed 1 July 2026). These studies show that in the first hexad there is a peak of words related to the sea and to inner pain and emotion; in the second hexad these words decrease sharply, while martial terms increase strongly. The frequency of proper names and place names also changes: names linked to the Trojan past prevail in the first hexad, whereas new names of Italic peoples and heroes appear in the second.

Besides word frequency, the other large-scale mathematical analysis is due to the philologist George E. Duckworth (1903-1972), who conducted a well-known study on the numerical "symmetry" of the *Aeneid*, identifying proportions and patterns [14]–none linked to the linguistic parameters considered below–based on the golden ratio $\varphi = (1 + \sqrt{5})/2 \approx 1.618$.

According to Duckworth: "Virgil's superb poetic power, as seen in imagery, sound effects (such as alliteration and assonance), and complex metrical patterns, contributes to the greatness and splendor of the poem. The *Aeneid* is one of the most consciously planned and carefully constructed epics of world literature. In a work of such magnitude, with so many threads inextricably and harmoniously interwoven, it should occasion no surprise that one or more basic designs of symmetry and variety, of parallelism and contrast, underlie the composition of the poem, both as a whole and in its separate parts. All readers of the *Aeneid* are conscious of these qualities to a degree, but the extent to which structural pattern and architectonic design dominate the epic has not been fully realized. Moreover, these structural features are not merely an adornment for their own sake but are devised to emphasize and make more significant the meaning of the poem; structure and content go hand in hand, and Virgil's conscious artistry in combining the two is an additional proof of his supreme achievement as an epic poet."

Furthermore, he writes: "Virgil composed the *Aeneid* on the basis of mathematical proportion; each book reveals, in small units as well as in the main divisions, the famous numerical ratio known variously as the Golden Section, the Divine Proportion, or the Golden Mean ratio. The existence of this particular mathematical proportion everywhere in the *Aeneid* may seem to some readers, if not fantastic, at least highly controversial, and yet it must be accepted as a fact."

Duckworth discusses Virgil's alternation of the books, the division of the poem into two hexads, with the resulting parallelism between the books of the hexads, and the *Aeneid* as a trilogy–three sections of four books each, the first and third framing the central portion, which contains the material of greatest historical and patriotic interest.

In summary, Duckworth identifies three major structural patterns: (a) the alternation between the less serious odd-numbered books (1, 3, 5, 7, 9, 11) and the more significant and tragic even-numbered books (2, 4, 6, 8, 10, 12); (b) the division of the poem into two hexads (1-6 and 7-12), with the books of the second hexad paralleling those of the first, both by similarity and by contrast; (c) the tripartite division of the poem, with the tragedy of Dido (books 1-4) and the tragedy of Turnus (books 9-12) enclosing the more national and Augustan portion (books 5-8). Besides these major patterns, he finds many smaller ones.

However, Duckworth's categories and comparisons are based on "themes" rather than on objective quantitative linguistic parameters such as those used here. The only quantitative datum he finds is the ubiquitous golden ratio, which appears–according to his investigation–no fewer than 1,044 times.

Although Virgil certainly conceived the two hexads to mirror the *Odyssey* and the *Iliad*, including parallel sub-themes in both, I doubt that he could have "managed" the distribution of themes throughout the poem so as to obtain the golden ratio 1,044 times, while at the same time taking care of the structure of the hexameters, the length of each book, the parallelism of themes, and so forth. In my investigation, reported below, I have never found the golden ratio.

2.2 *De Rerum Natura*

De Rerum Natura (*On the Nature of Things*), Lucretius' only known work, was composed in the 50s BC and was circulating by about 54 BC, shortly after the poet's death (c. 99-c. 55 BC), a generation before Virgil's poem [15].

The poem consists of six books in dactylic hexameters. The first three books (first triad) give a fundamental account of being and nothingness, matter and space, atoms and their motion, the infinity of the universe in time and space, the regularity of reproduction, the nature of mind and spirit as material bodily entities, and their mortality, since they and their functions (consciousness, pain) end with the bodies that contain them.

The last three books (second triad) give an atomistic and materialistic explanation of phenomena such as vision and the senses, sex and reproduction, natural forces and agriculture, the heavens and diseases. The argument moves from things imperceptible to the senses (atoms) to overwhelming sensory phenomena (earthquakes, plagues).

While each book has its own unity, the poem as a whole can be regarded as made of three pairs of books, or of two halves of three books each (triads). Apart from the single hexameter, the most clearly defined structural unit of the work is the book: its length, internal articulation and rhetorical shape are carefully planned [16]. As the poem moves from physics to anthropology and human history, Lucretius draws more explicitly on the historical, political and mythological registers of late Republican Rome.

De Rerum Natura made a fundamental and lasting contribution to the language of Latin philosophy. Its style is marked by the challenge of forging a technical philosophical vocabulary in Latin [16–19].

As for the *Aeneid* and other classical works, such as the *Iliad* and the *Odyssey*, mathematical and statistical analyses of *De Rerum Natura* have been based mostly on word frequency. Scholars of classical philology and corpus linguistics have indexed the vocabulary of the text to study its Epicurean concepts, its formulas and its poetic style. An analysis of Lucretius' vocabulary can be found in the book by Donald C. Swanson [20], which provides an itemized analysis of the complete word stock of the poem, classifying words by frequency, morphology and word formation. Louis Roberts compiled a complete alphabetical concordance of every word occurrence in the poem [6]. The site https://www.intratext.com/Catalogo/Autori/Aut238.HTM (last accessed 1 July 2026) provides an open digital text of *De Rerum Natura* with word lists arranged alphabetically, by frequency, by reversed spelling and by token length.

2.3 Mathematical analysis

My analysis of the *Aeneid* and *De Rerum Natura* is based on linguistic parameters–recalled in Section 3–of which a writer may be unaware; for this reason they can reveal textual structures that escaped the author. I developed these mathematical tools in recent years and have applied them to several case studies [1, 25, 27]. They differ from those used by other scholars engaged in the quantitative analysis of texts, who usually count words (lexical frequency), word clusters and the like [21–24]. As far as I know, none of these studies has investigated the mathematical structure underlying alphabetical texts and its likely connection with short-term memory, which is the purpose of the present approach.

3 Deep-language parameters

I recall some fundamental deep-language parameters defined in References [25, 26]. They are useful for comparing texts because they can assess similarity objectively, beyond the writer's awareness. These parameters refer to the "surface" structure of texts, not to the "deep" structure of cognitive theory, as discussed in my previous papers.

Let n_C, n_W, n_S, n_I and n_L be, respectively, the number of characters, words, sentences, interpunctions (punctuation marks) and lines counted in disjoint blocks of text–in this paper, the books. Four deep-language parameters can be defined (Appendix A gives information on the raw textual data and their processing; Appendix B lists the mathematical symbols):

Number of characters per word, C_P:

$$C_P = \frac{n_C}{n_W} \tag{1}$$

Number of words per sentence, P_F:

$$P_F = \frac{n_W}{n_S} \tag{2}$$

Number of words per interpunction, referred to as the word interval, I_P:

$$I_P = \frac{n_W}{n_I} \tag{3}$$

Number of word intervals per sentence, M_F:

$$M_F = \frac{n_I}{n_S} \tag{4}$$

Notice that I_P, P_F and M_F are not independent, because:

$$I_P = \frac{P_F}{M_F} \tag{5}$$

For poetry, I also define the number of words per line, $I_{P\ell}$, referred to as the word interval per line [1]:

$$I_{P\ell} = \frac{n_W}{n_L} \tag{6}$$

4 Equivalent buffers of the extended short-term memory

To memorize a sentence, human short-term memory can be modelled by three *equivalent* buffers in series (Figure 1) [27, 28]. The first buffer is linked to the number of syllables (or characters) per word. The second buffer is linked to the number of words between two consecutive interpunctions, i.e., the word interval I_P, which seems to follow Miller's 7 ± 2 law [2] and to model the output of short-term memory properly. The third buffer is linked to the number of word intervals contained in a sentence, described by M_F.

Figure 1 shows the capacity of the buffers and how they are filled, not the actual temporal sequence in which the mind processes textual information, which is largely unknown (some processing may occur in parallel). In other words, Figure 1 shows an "equivalent" flow of information across the three buffers, not a precise temporal correspondence.

I refer to the overall model as the "extended" short-term memory (E-STM), not to be confused with the working memory of cognitive psychology [29]. The model, however simple, shows how the generation of characters, words and sentences can be perceived externally and therefore studied. As previously noted [1, 27, 28], the underlying cognitive process remains largely unknown, yet its "output" can be observed in alphabetical languages and can provide a quantitative model to work with.

In conclusion, the number of characters per word, C_P, the word interval, I_P, and the number of word intervals in a sentence, M_F, are useful for revealing how the writer has unconsciously shaped the text, under the constraints of his own E-STM and of the demands placed on the reader's E-STM to process meaning correctly. We can now study the texts of Virgil and Lucretius.

5 Exploratory data analysis: *Aeneid*

I present a series of scatterplots that show similarities and differences between the two hexads. Not only do they agree with the findings of specialists in classical literature, but they can also reveal new information.

Figure 2 shows characters, words and C_P versus book, and the scatterplot between words and characters. In the scatterplot, the ratio between ordinate and abscissa gives C_P. Characters and words are strongly correlated, and C_P fluctuates around its mean in a wave-like pattern, with increasing values in the triads 1-2-3, 6-7-8 and 9-10-11. No quantity follows the golden ratio.

The three regression lines–the one for the whole poem and those for the two hexads–merge. Because the same language, Latin, is used throughout, word length cannot distinguish the hexads; only the relative frequency of words reveals differences, as discussed in Section 2. In other words, Virgil "automatically" uses the same language throughout the poem.

Figure 3 shows sentences, words and P_F versus book number, and the scatterplot between sentences and words. In the scatterplot, the ratio between ordinate and abscissa gives P_F. The regression lines of the two hexads differ and do not merge with that of the whole poem, although the mean values of P_F differ by only about 4% (Table 2); the difference lies mainly in the spread of sentence length, which is larger in the second hexad. A single regression line for the whole poem would therefore not represent it accurately.

Figure 4 shows the scatterplot between words and punctuation marks; the ratio between ordinate and abscissa gives I_P. The two hexads clearly differ and the three regression lines do not merge: Virgil "designed" the length of the word interval quite differently in the two hexads, and the regression line of the whole poem would not represent the poem accurately.

Figure 5 shows the scatterplot between sentences and punctuation marks; the ratio between ordinate and abscissa gives M_F. Again the hexads differ and the three regression lines do not merge, so Virgil "designed" the number of word intervals per sentence differently in the two hexads.

Figure 6 shows P_F versus M_F. The hexads clearly differ, and the same remarks apply to the regression lines.

Figure 7 shows M_F versus I_P. The hexads clearly differ. More important, the size of the second buffer (I_P) and that of the third buffer (M_F) are uncorrelated: the word interval and the number of such intervals in a sentence appear to have been set without regard to

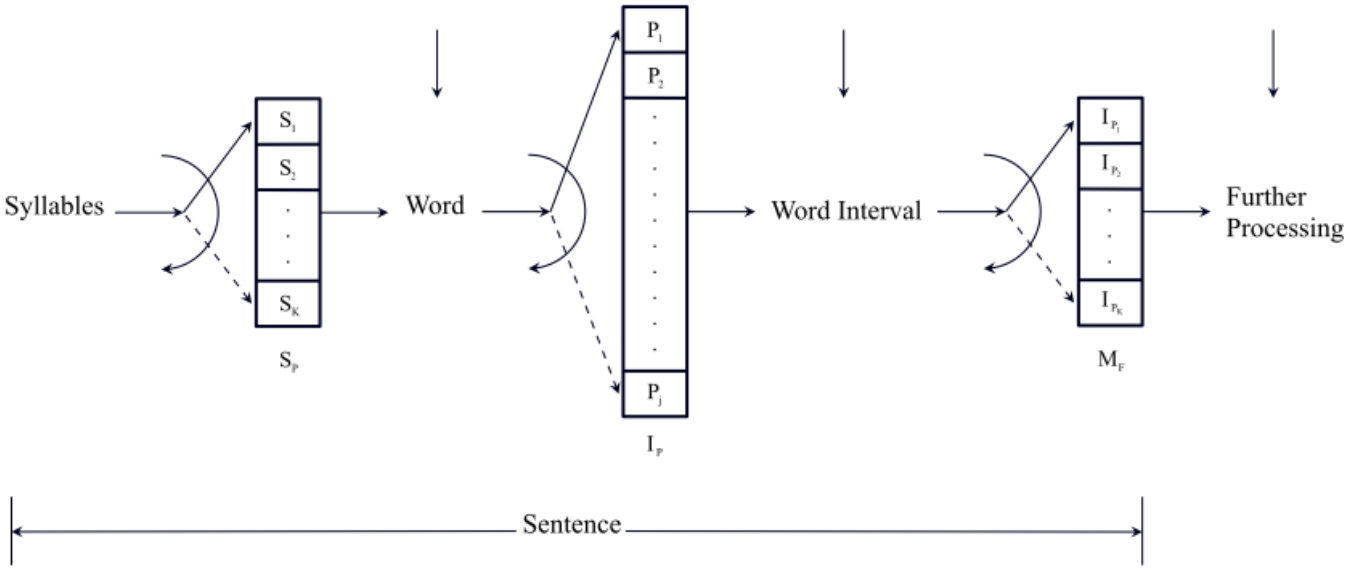


Figure 1. Modelling of the extended short-term memory (E-STM) with three equivalent buffers in series [27, 28]. Syllables $S_1, S_2, \ldots, S_k$ are stored in the first buffer, from 1 to about 3-4 items, until a space or an interpunction is introduced (vertical arrow) to fix the length of the word; S_P denotes the number of syllables per word. Words $P_1, P_2, \ldots, P_j$ are stored in the second buffer–approximately within Miller's range–until an interpunction (vertical arrow) is introduced to fix the length of I_P. The word interval I_P is then stored in the third buffer, from about 1 to 6 items, until the sentence ends with a full stop, a question mark or an exclamation mark (vertical arrow). The process is then repeated for the next sentence. The first buffer is linked to the length of words; the second to the number of words between two consecutive interpunctions, I_P (word interval); the third to the number of word intervals in a sentence, M_F. All these buffers convey meaning; syllables do not.

each other.

Figure 8 shows I_P versus P_F. This relationship is typical of the behaviour of STM [25, 28]: if sentences become longer, the writer–who is also a reader–must introduce punctuation marks, that is, he must decrease I_P (and therefore increase M_F for a fixed sentence length), because his STM is limited by Miller's law. The differences between the two hexads are again large.

Finally, Figure 9 shows I_P, $I_{P\ell}$ and M_F versus book number. I_P and M_F change between the two hexads, whereas $I_{P\ell}$ is practically constant, as expected from the structure of the hexameter.

6 Hexads: similarities and differences

In the previous section, I have shown how the main linguistic parameters vary across the books. In this section, I first summarize the similarities and differences between the two hexads with sample statistics, and then propose an index to assess the symmetry between books.

6.1 Descriptive statistics

Table 1 reports the totals for the whole poem and the two hexads. Table 2 shows the mean value and standard deviation (in parentheses) of the parameters.

Table 1. Totals of characters, words, sentences and lines, *Aeneid*.

Books	Characters	Words	Sentences	Lines
All (1-12)	367 317	63 731	3 283	9 899
1-6	176 289	30 573	1 603	4 756
7-12	191 028	33 158	1 680	5 143

Means and standard deviations have been calculated by weighting each book with its number of words, so that shorter books do not count as much as longer ones (Figure 2); the statistical weight of a book is the ratio between its number of words and the total number of words of the *Aeneid*. It should also be noted that the "average" values of these parameters, calculated from the totals (Table 1), do not theoretically coincide with arithmetic or statistical means— which are always greater, as established in Appendix D of reference [30] —although in Aeneid they turn out to be very close, owing to the similar lengths of the books.

From these findings, the hexads appear fairly symmetric, but not for all parameters. Symmetry can be evaluated with a "uniformity index", defined below.

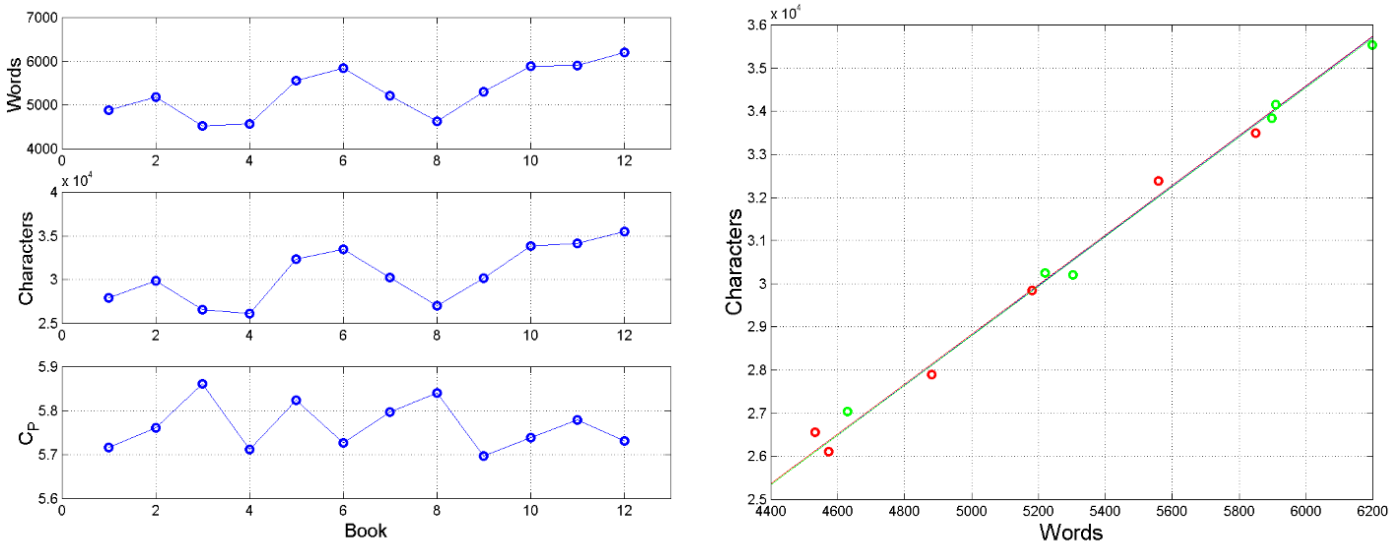


Figure 2. *Aeneid*. Left panel: characters, words and C_P versus book number. Right panel: scatterplot between words and characters. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem. The three lines merge and cannot be distinguished.

Table 2. Mean value and standard deviation (in parentheses) of the indicated parameters, *Aeneid*.

Books	C_P	P_F	I_P	$I_{P\ell}$	M_F
All (1-12)	5.76 (0.05)	19.51 (1.42)	5.47 (0.77)	6.44 (0.06)	3.62 (0.45)
1-6	5.77 (0.06)	19.12 (0.94)	4.71 (0.12)	6.43 (0.06)	4.06 (0.14)
7-12	5.76 (0.05)	19.87 (1.67)	6.17 (0.35)	6.45 (0.06)	3.22 (0.21)

6.2 Uniformity index

Let us consider a variable $x(k)$–in our case, its mean value in book k–and define the following distribution $d(k)$ versus book k:

$$d(k) = \frac{x(k)}{\sum_{j=1}^{N} x(j)} \tag{7}$$

Eq. (7) is the ratio between the mean value $x(k)$ in book k and the sum of the means in the N books ($N = 12$ for the *Aeneid*). If $x(k) = x_0$, that is, if the value were the same in every book, then $d(k) = 1/N$, a uniform distribution across books, and the symmetry would be perfect. Otherwise some asymmetry would be present.

To establish whether an experimental distribution is uniform, we could use the traditional tools of statistics, but Eq. (7) does not define a standard "frequency" distribution, so another tool is needed. I propose a "uniformity" index $U(k)$ that measures how far the experimental distribution is from the theoretical uniform value $d(k) = 1/N$:

$$U(k) = \frac{d(k)}{1/N} = N\, d(k) \tag{8}$$

If $U(k) = 1$ for every k, then $d(k) = 1/N$; otherwise $U(k) \neq 1$. The closer $U(k)$ is to 1, the more uniform the experimental distribution.

Table 3. Standard deviation of $U(k)$ for the indicated parameters, *Aeneid*.

W	S	C_P	P_F	I_P	$I_{P\ell}$	M_F
0.109	0.134	0.020	0.076	0.149	0.010	0.130

Because of its definition, the arithmetic mean of $U(k)$ is exactly 1 (Appendix C); therefore, deviations from this value can be described by its standard deviation. Moreover, Eq. (8) does not depend on the absolute values of the variable, so different parameters can be compared on the same basis.

Figure 10 shows $U(k)$ for the linguistic parameters, and Table 3 reports the standard deviation of $U(k)$.

From Figure 10 and Table 3 we observe–however trivial it may seem–that the only parameters distributed practically uniformly across the 12 books are C_P and $I_{P\ell}$, namely word length and hexameter length in words. The other parameters depend on the book and some, markedly, also on the hexad.

After these findings, it is worth discussing whether,

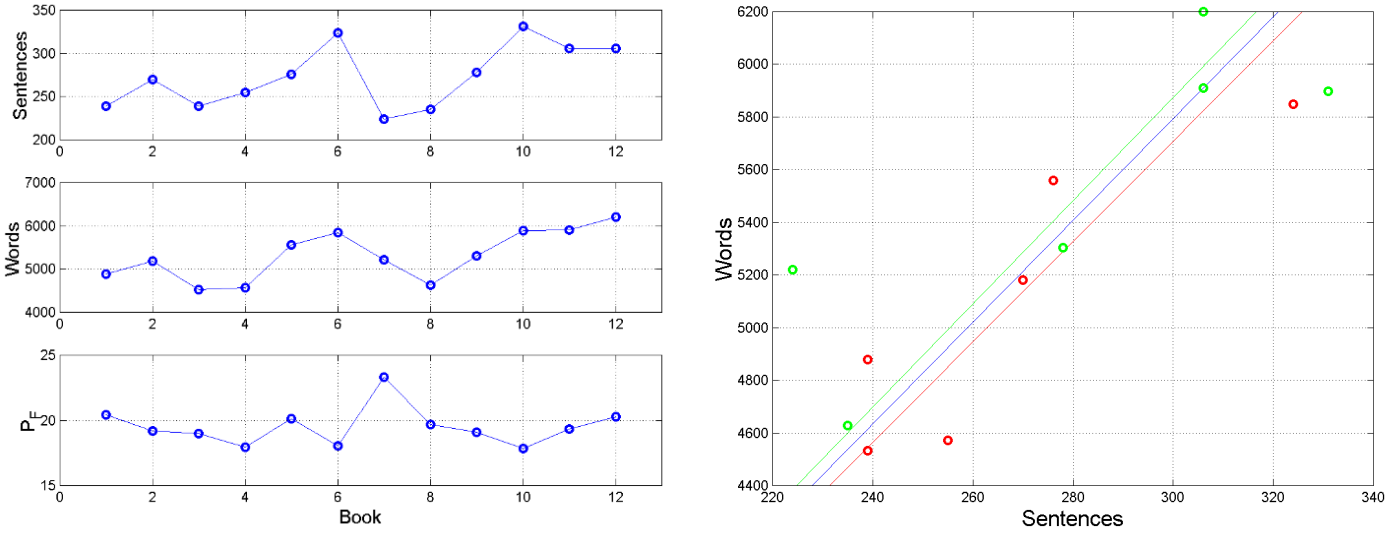


Figure 3. *Aeneid*. Left panel: sentences, words and P_F versus book number. Right panel: scatterplot between sentences and words. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem.

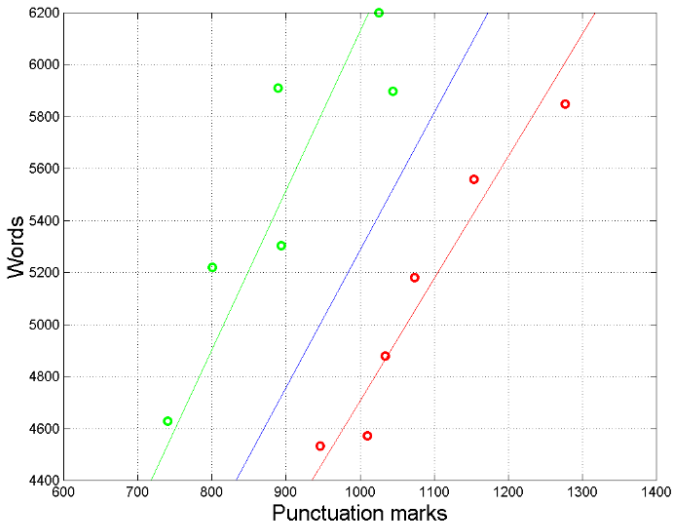


Figure 4. *Aeneid*. Scatterplot between words and punctuation marks. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem.

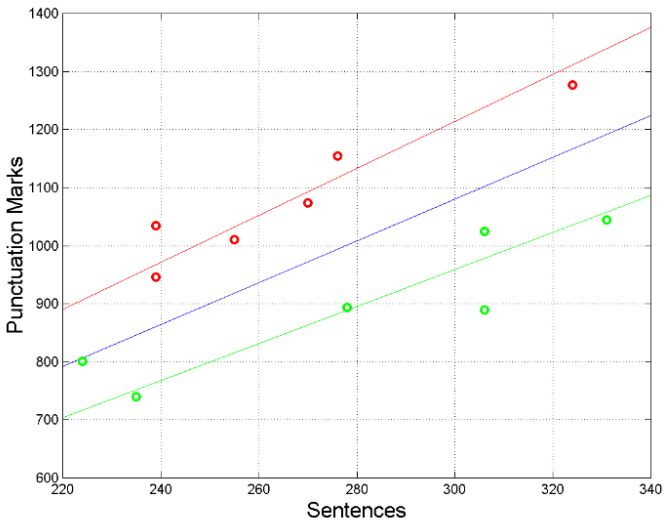


Figure 5. *Aeneid*. Scatterplot between sentences and punctuation marks. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem.

and which, linguistic parameters were consciously "designed" by the author, and to what extent. The following section discusses this issue.

7 Parameters consciously "designed"? A discussion

Were the linguistic parameters consciously "designed" by Virgil? If so, which ones, and to what extent? In the light of the observations above, the following considerations can be made.

Regarding C_P, no significant variation is observed between the hexads, because $U(k) \cong 1$ (Table 3 and Figure 10). Even though scholars of Latin literature note changes in vocabulary between the hexads (Section 2), word length does not change significantly. Virgil writes in Latin and cannot significantly alter the length of the linguistic "codewords" used to convey meaning.

Regarding P_F, the hexads are only slightly different, since their mean values fall within $\pm 2\%$ of the global mean. However, the standard deviation is much larger in the second hexad (Table 2), where sentences can

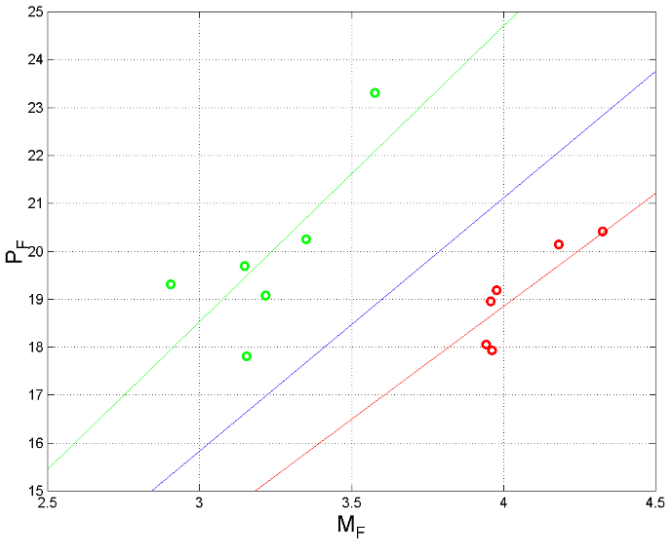


Figure 6. *Aeneid*. Scatterplot between P_F and M_F. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem.

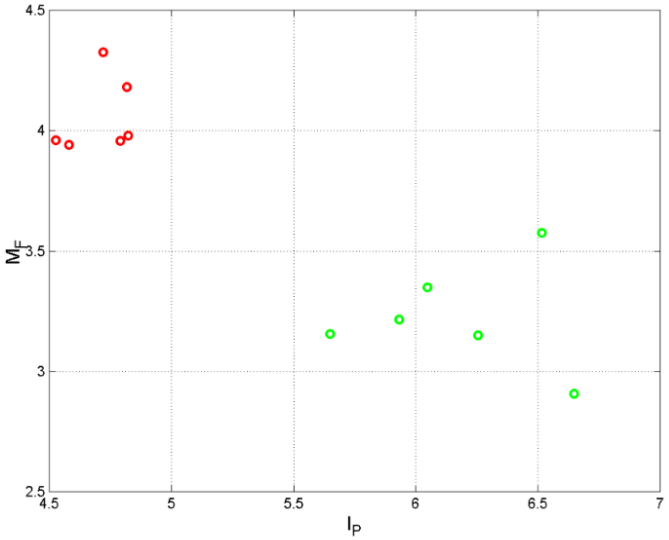


Figure 7. *Aeneid*. Scatterplot between I_P and M_F. Red circles and red line: first hexad; green circles and green line: second hexad. The blue line is the regression line calculated for the whole poem.

become much longer than in the first hexad.

Regarding I_P, the hexads differ markedly, implying different capacities of the second buffer (STM) of the E-STM model (Figure 1). The first hexad has a significantly smaller value, which imposes a lighter load on the reader's STM.

Regarding M_F, the two hexads are also markedly different, implying different capacities of the third buffer of the E-STM. For this buffer, however, the situation is reversed: the first hexad needs a larger capacity to accommodate more word intervals in a sentence. This follows from the data: in the first hexad, sentences have about the same length as in the second, but the word interval is significantly shorter, so a sentence contains more word intervals.

The differences in I_P and M_F between the hexads seem to arise from the different "stories" told in the first ("Odyssey") and in the second hexad ("Iliad"), because these parameters model the second and third buffers of the E-STM (Figure 1).

Regarding $I_{P\ell}$, the hexads are practically identical, because its distribution is practically uniform, $U(k) \cong 1$ (Table 3 and Figure 10). This is understandable: each line is a hexameter, whose length in words Virgil cannot significantly change because of its structure.

In conclusion, three facts emerge:

a) The hexads differ in the parameters that describe the second (I_P) and third (M_F) equivalent buffers of the E-STM. The author acts consciously by introducing interpunctions (punctuation marks) to convey meaning better.

b) The hexads do not differ in the (uniform) distribution of the parameters the author cannot design, such as C_P (Latin) and $I_{P\ell}$ (hexameter; STM).

c) In the whole poem, each line contains $I_{P\ell} \approx 6.4$ words, with a very small standard deviation (≈ 0.06). This value is very near Miller's central value whereas, on the contrary, the word interval I_P is near Miller's lower bound (Table 2), with a large standard deviation (0.77), more than ten times that of $I_{P\ell}$.

The observation about $I_{P\ell}$ seems obvious, because each line is a hexameter, yet it is important. As with hendecasyllables ([1] and Section 9), it suggests that the verse was "shaped" unconsciously. Metrical theory assumes that the poet counted syllables, not words; however, his STM "planned" the length of the hexameter in words, because–I believe–only words convey meaning. Counting words appears to be imposed by STM to fit the second buffer of Figure 1, so that the information can be stored in long-term memory and retrieved later, perhaps for public recitation. Conversely, I_P, M_F and P_F do not appear to be constrained in this way.

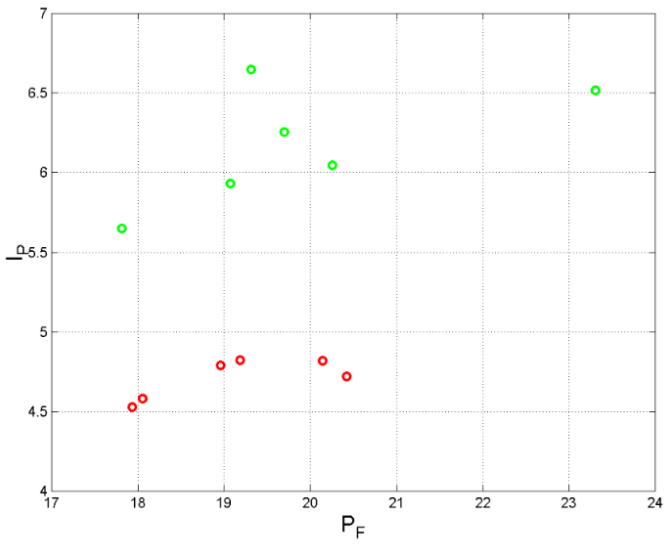


Figure 8. *Aeneid*. Scatterplot between I_P and P_F. Red circles: first hexad; green circles: second hexad.

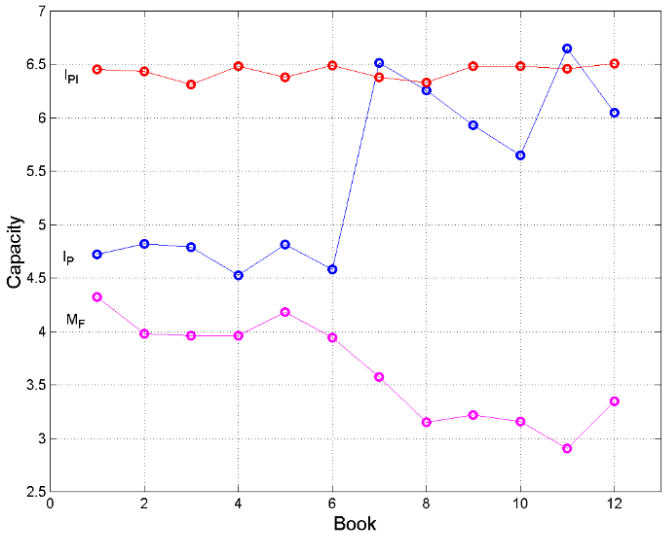


Figure 9. *Aeneid*. Buffer capacity versus book number. Blue circles: I_P; magenta circles: M_F; red circles: $I_{P\ell}$.

8 Exploratory data analysis and statistics of *De Rerum Natura*

As with the *Aeneid*, I present scatterplots between parameters to show similarities and differences between the two triads of *De Rerum Natura*. Since many remarks made on the *Aeneid* also apply here, I show only some data and refer the reader to Appendix D (Figures A1-A5) for further details.

Figure 11 shows characters, words and C_P versus book number, and the scatterplot between words and characters. Characters and words are strongly correlated. Note also the larger values in the books of the second triad. As in the *Aeneid*, the three regression lines–the one for the whole poem and those for the two triads–merge.

Figure 12 shows sentences, words and P_F versus book number, and the scatterplot between sentences and words. Here the triads are markedly different, and the three regression lines do not merge.

Figure 13 shows I_P, $I_{P\ell}$ and M_F versus book number. I_P and M_F vary across the triads, whereas $I_{P\ell}$ remains practically constant, as expected from the structure of the hexameter.

Table 4 presents the totals for the whole poem and the triads. Table 5 shows mean values and standard deviations (in parentheses).

Figure 14 shows $U(k)$ for the linguistic parameters, and Table 6 reports its standard deviations.

Finally, Figure 15 compares the standard deviation of $U(k)$ in the *Aeneid* and *De Rerum Natura*.

Table 4. Totals of characters, words, sentences and lines, *De Rerum Natura*.

Books	Characters	Words	Sentences	Lines
All (1-6)	274 693	49 030	1 782	7 415
1-3	124 984	22 347	747	3 376
4-6	149 709	26 683	1 035	4 039

From these results, Lucretius, too, cannot design the length of either words (C_P)–he is writing in Latin–or hexameters ($I_{P\ell}$)–he is counting syllables.

It is interesting that $I_{P\ell}$ (6.61) is very close to that of the *Aeneid* (6.44), near the central value of Miller's law. This supports the thesis that the hexameter was "designed" by STM to convey meaning by memorizing words, not syllables. In contrast, I_P is not tied to the hexameter: its mean value (7.94) lies above Miller's central value, towards the upper part of the range. The author designs word intervals, as he does with sentence length–where he is freer–because he deliberately introduces interpunctions. Note, however, that in *De Rerum Natura* the word interval is also fairly uniform across books (0.037 in Table 6, against 0.149 in the *Aeneid*), so the distinction between designed and non-designed parameters is less sharp than in Virgil's poem.

The triads are clearly asymmetric in the number of words, sentences and lines. Compared with the *Aeneid*, sentences are markedly longer (P_F), probably because of the elaboration needed to convey complex scientific and philosophical concepts.

Table 5. Mean value and standard deviation (in parentheses) of the indicated parameters, *De Rerum Natura*.

Books	C_P	P_F	I_P	$I_{P\ell}$	M_F
All (1-6)	5.60 (0.05)	27.70 (2.25)	7.94 (0.28)	6.61 (0.09)	3.49 (0.30)
1-3	5.59 (0.05)	29.95 (0.94)	7.87 (0.11)	6.62 (0.10)	3.80 (0.06)
4-6	5.61 (0.04)	25.81 (0.90)	7.99 (0.35)	6.61 (0.08)	3.23 (0.11)

Table 6. Standard deviation of $U(k)$ for the indicated parameters, *De Rerum Natura*.

W	S	C_P	P_F	I_P	$I_{P\ell}$	M_F
0.107	0.180	0.009	0.090	0.037	0.015	0.094

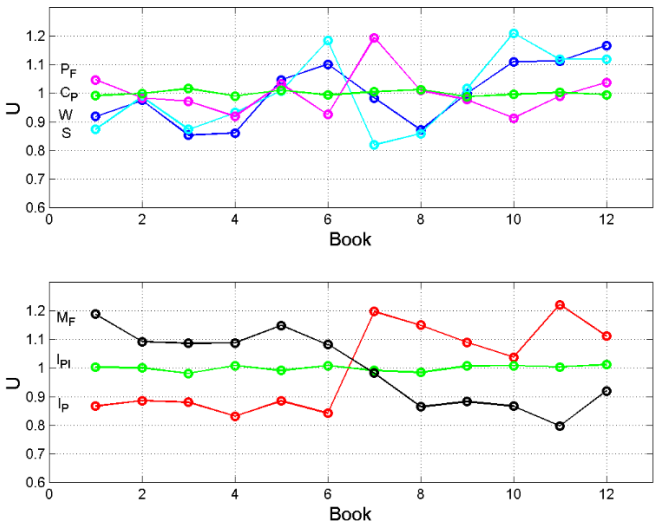


Figure 10. Uniformity index $U(k)$ for the indicated linguistic parameters, *Aeneid*. Upper panel: words (W), blue; sentences (S), cyan; characters per word (C_P), green; words per sentence (P_F), magenta. Lower panel: word intervals per sentence (M_F), black; words per interpunction (I_P), red; words per line ($I_{P\ell}$), green.

Finally, according to the universal readability index [31], *De Rerum Natura* would have been more difficult to read than the *Aeneid*, because of its larger values of P_F and I_P.

In the following section, I examine some classic translations of these poems into Italian–and into hendecasyllables.

9 Classical Italian translations

I present some findings on classic Italian translations in hendecasyllables of the *Aeneid* (*Eneide*) and of *De Rerum Natura* (*Sulla natura delle cose*), texts read in Italian high schools. The aim is not to compare the original texts with their translations in full–a study I intend to pursue in the future–but to examine the transformation of hexameters into hendecasyllables.

Here too, as in the *Divina Commedia* [1], the hendecasyllables contain about seven words: a translator using hendecasyllabic verse, even without rhyme as in these cases, cannot escape the central value of Miller's law.

I first summarize the historical transition from hexameters to hendecasyllables, and then present the classic Italian translations of the *Aeneid* and *De Rerum Natura*.

9.1 From hexameters to hendecasyllables

The historical and formal link between the Greek and Latin dactylic hexameter and the Italian hendecasyllable does not stem from a direct structural evolution, but from a process of cultural adaptation, translation and rhythmic transposition.

Although both the hexameter and the hendecasyllable established themselves as the classic metres of epic poetry, the hendecasyllable–nominally an eleven-syllable line, though it can extend a little [1]–originated when medieval and Renaissance Italian poets sought to transpose the solemnity of the classical hexameter, based on syllable quantity and structured in six feet, into a language governed by stress and by the number of syllables, such as Italian.

In Greek and Latin, poetry was based on syllable duration. As early as late antiquity, however, the western vernaculars lost the distinction between long and short vowels, and with it the duration of syllables. With the rise of vernacular languages such as Italian, poets could no longer compose true hexameters: read according to the natural stress of the words rather than vowel length, a classical hexameter loses its rhythm. Writers needed an equivalent based on a fixed number of syllables, solemn and flexible enough for long narrative, like the hexameter of Homer or Virgil. That equivalent was the hendecasyllable.

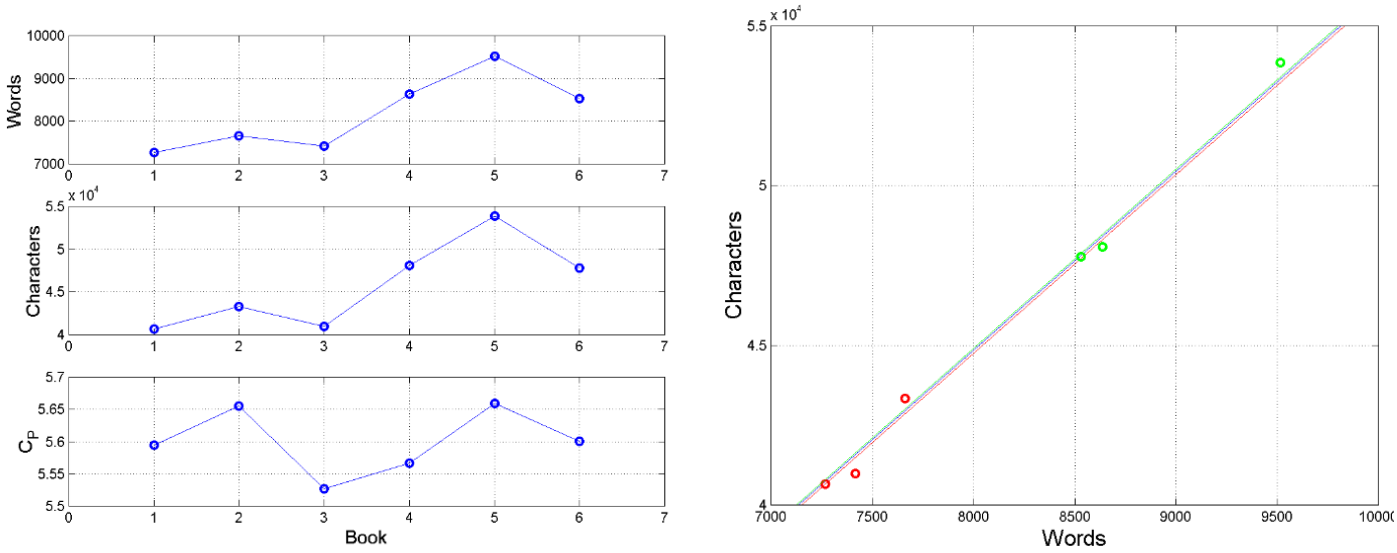


Figure 11. *De Rerum Natura*. Left panel: characters, words and C_P versus book number. Right panel: scatterplot between words and characters. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem. The three lines merge and cannot be distinguished.

9.2 Italian classical translations into hendecasyllables

The classical Italian translation of the *Aeneid* is by Annibale Caro (1507-1566), poet and playwright. It is one of the masterpieces of Italian Renaissance literature, a very free and personal version in unrhymed hendecasyllables, written in an elegant and musical language typical of the Renaissance [32–35].

The classical Italian translation of *De Rerum Natura* is by Alessandro Marchetti (1633-1714). It is the first complete Italian verse translation and one of the most celebrated of Italian literature. Marchetti, a mathematician of the Galilean school, also used unrhymed hendecasyllables. His version has long been regarded as a model of fluency and elegance, and influenced later poets such as Ugo Foscolo (1778-1827).

Table 7 presents the totals of the linguistic parameters, and Table 8 shows mean values and standard deviations (in parentheses).

As shown in Table 7, the number of lines (hendecasyllables), words and sentences differs from that of the Latin texts. I believe this is due to two main factors. First, Latin uses a complex system of noun endings (cases) to mark whether a word is a subject or an object; Italian has dropped this system and relies on word order (subject-verb-object) and prepositions, although this order is often disregarded, both in marked sentences and in poetry. Second, Latin has no definite or indefinite articles, whereas Italian has. In this respect, any translator has no choice. I will examine these aspects in more detail in a future paper.

Table 8 highlights the following points. Italian words (C_P) are on average shorter than Latin words, because they do not carry case endings and because the frequent definite and indefinite articles lower the average word length. Sentences (P_F) are decidedly longer than in Latin. The word interval (I_P) tends to be larger than in Latin, especially in *De Rerum Natura*, placing a heavier burden on the second buffer of the E-STM. The number of word intervals per sentence (M_F) is larger than in Latin, indicating a greater burden also on the third buffer. The most interesting aspect–however "trivial" it may seem–is that $I_{P\ell}$ is close to 7, the central value of Miller's law: the hendecasyllables of Caro and Marchetti are quantitatively similar to those of Dante [1], and their length in words is about the same as that of the Latin hexameters.

Figure 16 shows $U(k)$ for the linguistic parameters of the Italian translation of the *Aeneid*, and Table 9 reports its standard deviations.

From these results, the translator cannot design the length of either words (C_P)–he is writing in Italian–or hendecasyllables ($I_{P\ell}$)–he is apparently counting syllables. These two parameters have practically the same standard deviation, 0.008-0.009. The other parameters are not uniformly distributed.

Figure 17 shows $U(k)$ for the Italian translation of *De Rerum Natura*, and Table 10 reports its

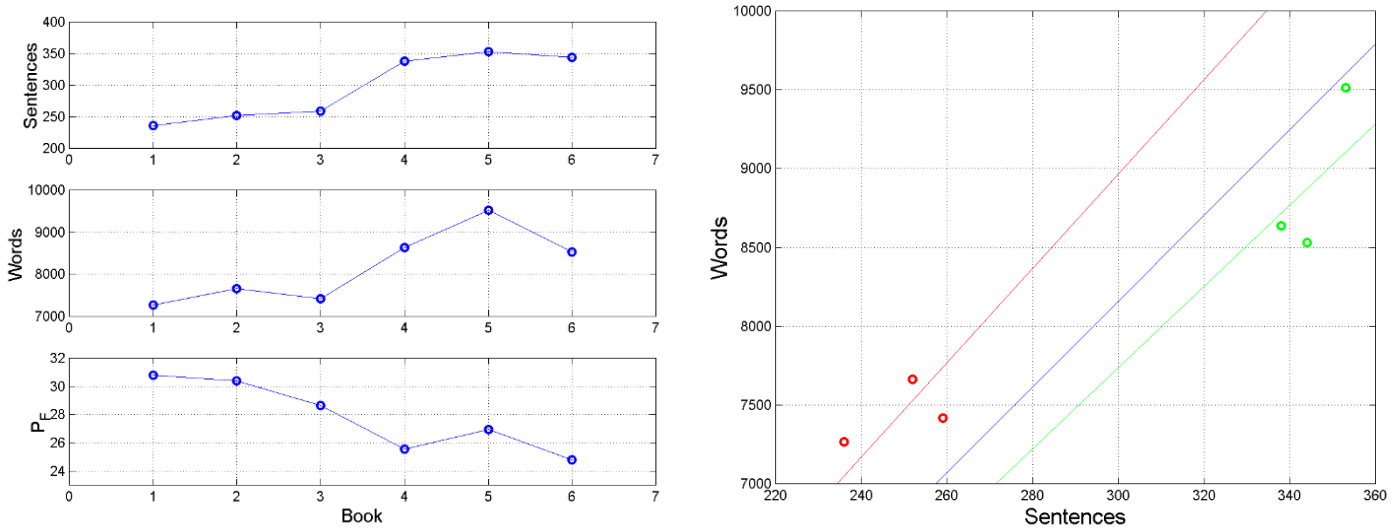


Figure 12. *De Rerum Natura*. Left panel: sentences, words and P_F versus book number. Right panel: scatterplot between sentences and words. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem.

Table 7. Totals of characters, words, sentences and lines in the classical Italian translations of the *Aeneid* and *De Rerum Natura*.

Poem	Characters	Words	Sent.	Lines
Eneide (*Aeneid*)	459 539	107 639	4 179	15 369
Sulla natura delle cose (*De Rerum Natura*)	327 410	72 769	1 677	10 699

standard deviations. Also in this case, the translator cannot design the length of either words (C_P) or hendecasyllables ($I_{P\ell}$), which have the same standard deviation, 0.007. The other parameters are not uniformly distributed.

In conclusion, translating poems written in hexameters into poems written in hendecasyllables can alter many linguistic parameters, but not $I_{P\ell}$. This finding is consistent with the thesis that hexameters and hendecasyllables were a product of human STM, needed for storing information in long-term memory for subsequent retrieval.

In the next section, I compare the uniformity index of Dante's *Divina Commedia* [1] with that found in the *Eneide* and *Sulla natura delle cose*.

10 Uniformity index in Italian poetry

The basic textual unit of the *Divina Commedia* is the *canto* (plural *canti*). The poem consists of 100 canti, grouped into three *cantiche* (plural of *cantica*), namely *Inferno*, *Purgatorio* and *Paradiso*. Each cantica contains 33 canti, plus an additional canto at the beginning of *Inferno* that introduces the whole poem. The canti are written in interlocking rhymed hendecasyllables, organized into tercets.

The tercet is an Italian metrical form of three lines. In the interlocking rhyme scheme (*terza rima*), the first and third lines of each tercet rhyme, and the second line rhymes with the first and third lines of the following tercet. The canto ends with a single line that rhymes with the second line of the last tercet. Dante was the first poet to use the tercet for a long poem; hence it is also known as Dante's tercet.

I discuss the symmetry of the poem with the uniformity index, as a follow-up of my study of the *Divina Commedia* [1]. Here the unit is the canto, so $N = 100$ in Eqs. (7) and (8). Figure 18 shows $U(k)$ for the linguistic parameters, and Table 11 reports the standard deviation of $U(k)$.

These results show a high uniformity of word length (C_P) and of words per line ($I_{P\ell}$), another indication that STM shapes these parameters. The number of words per canto (W) is also very uniform, because of the tercet structure and the similar length of the canti

Table 8. Mean value and standard deviation (in parentheses) of the indicated parameters in the classical Italian translations of the *Aeneid* and *De Rerum Natura*.

Poem	C_P	P_F	I_P	$I_{P\ell}$	M_F
Eneide (*Aeneid*)	4.27 (0.03)	25.94 (2.18)	6.26 (0.20)	7.00 (0.06)	4.15 (0.36)
Sulla natura delle cose (*De Rerum Natura*)	4.50 (0.03)	43.56 (2.74)	9.53 (0.43)	6.80 (0.04)	4.57 (0.17)

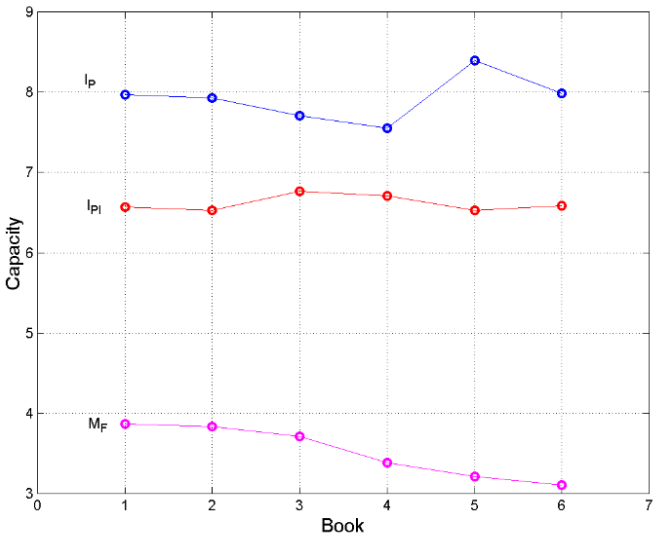


Figure 13. *De Rerum Natura*. Buffer capacity versus book number. Blue circles: I_P; magenta circles: M_F; red circles: $I_{P\ell}$.

Table 9. Standard deviation of $U(k)$ for the indicated parameters, *Eneide* (*Aeneid*).

W	S	C_P	P_F	I_P	$I_{P\ell}$	M_F
0.115	0.169	0.009	0.089	0.035	0.008	0.091

measured in tercets.

Finally, Figure 19 compares the standard deviations in the *Eneide*, *Sulla natura delle cose* and *Divina Commedia*.

In conclusion, Figure 19 shows that Dante, Caro and Marchetti did not choose word lengths (C_P) at their own discretion throughout their texts: they use Italian, a language that has not changed significantly since the 13th century [25]. The *Divina Commedia* is still read in Italian high schools as it was written, with only moderate difficulty. It is also evident–a seemingly trivial result–that they could not vary the length of the line in words, $I_{P\ell}$ (Table 8 and [1]): the standard deviation of $U(k)$ for $I_{P\ell}$ is very small in all three texts (0.007-0.029, Tables 9-11), as it is for C_P. The somewhat larger value of the *Divina Commedia* is consistent with its much finer subdivision (100 canti instead of 12 or 6 books).

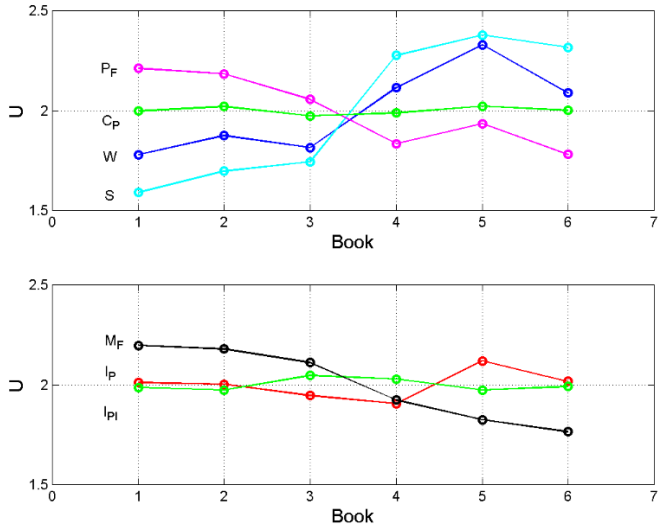


Figure 14. Uniformity index $U(k)$ for the indicated linguistic parameters, *De Rerum Natura*. Upper panel: words (W), blue; sentences (S), cyan; characters per word (C_P), green; words per sentence (P_F), magenta. Lower panel: word intervals per sentence (M_F), black; words per interpunction (I_P), red; words per line ($I_{P\ell}$), green.

Table 10. Standard deviation of $U(k)$ for the indicated parameters, *Sulla natura delle cose* (*De Rerum Natura*).

W	S	C_P	P_F	I_P	$I_{P\ell}$	M_F
0.138	0.140	0.007	0.071	0.048	0.007	0.041

11 Conclusion

The analysis shows that the hexameter of Greek and Latin poetry, crafted according to complex rules that required a deep knowledge of the syllabic structure of words, was designed–in terms of the E-STM model (Figure 1)–with explicit attention only to the first buffer. The hexameter was adopted throughout the Greco-Roman world because a poem composed in hexameters could be easily memorized and recited. The same holds–centuries later and in vernacular languages–for the hendecasyllable, the verse of Dante's *Divina Commedia* and of many other poets.

In this article, as in [1], I have argued that the structure of both hexameters and hendecasyllables reflects the unconscious work of short-term memory in producing word sequences that could easily be stored

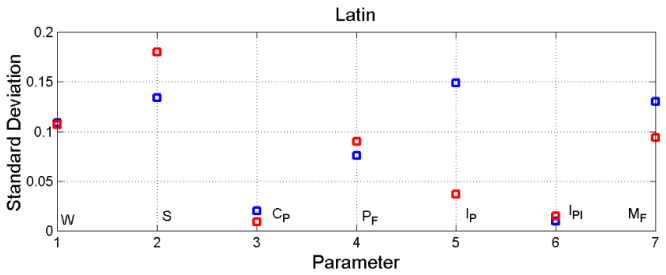


Figure 15. Standard deviation of $U(k)$ for the linguistic parameters indicated on the x-axis. *Aeneid*: blue squares; *De Rerum Natura*: red squares. W: words; S: sentences.

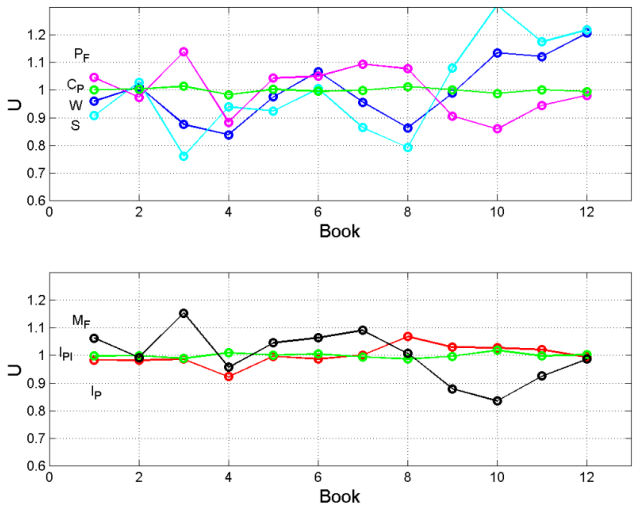


Figure 16. Uniformity index $U(k)$ for the indicated linguistic parameters, *Eneide* (*Aeneid*). Upper panel: words (W), blue; sentences (S), cyan; characters per word (C_P), green; words per sentence (P_F), magenta. Lower panel: word intervals per sentence (M_F), black; words per interpunction (I_P), red; words per line ($I_{P\ell}$), green.

in long-term memory for recitation.

The architects of hexameters and hendecasyllables studied rhythm by counting syllables, which do not convey meaning by themselves. I believe that they did not realize that, for these verses to succeed, they had first to pass through STM, which operates according to Miller's law and sets the length in words of both verses. In the texts examined here, it is most likely the central capacity of the second buffer of the E-STM (Figure 1) that determined the true metrics of the hexameter and of the hendecasyllable, namely about seven words, in Latin as in Italian.

In short, the creators of these metres dressed in elegant attire what their short-term memory required for long-term retention; to them, it always seemed that they were playing–and they certainly did–with syllables, accents and long or short vowels.

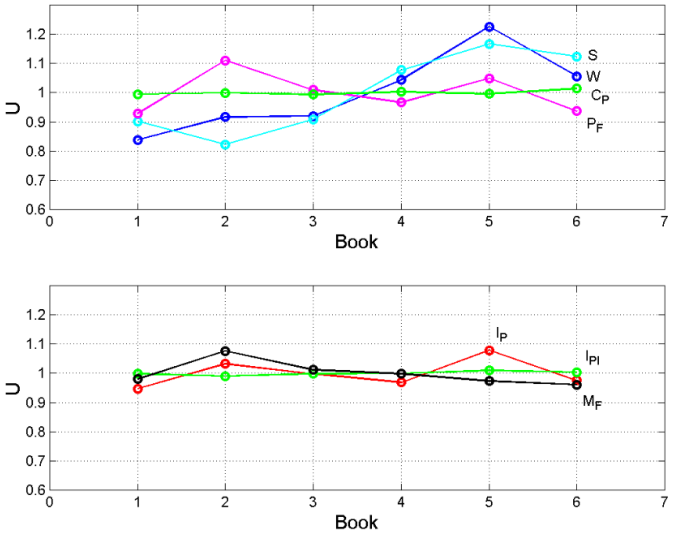


Figure 17. Uniformity index $U(k)$ for the indicated linguistic parameters, *Sulla natura delle cose* (*De Rerum Natura*). Upper panel: words (W), blue; sentences (S), cyan; characters per word (C_P), green; words per sentence (P_F), magenta. Lower panel: word intervals per sentence (M_F), black; words per interpunction (I_P), red; words per line ($I_{P\ell}$), green.

Table 11. Standard deviation of $U(k)$ for the indicated parameters, *Divina Commedia*.

W	S	C_P	P_F	I_P	$I_{P\ell}$	M_F
0.056	0.173	0.023	0.168	0.093	0.029	0.116

Data Availability Statement

The Latin texts and Italian translations are publicly available at the websites listed in Appendix A. The processed data are available from the author upon request.

Funding

This research received no external funding. In addition to the customary citation of the sources of the digital texts, the author thanks the colleagues who, with dedication, make the texts of works foundational to Western culture freely available to everyone.

Conflicts of Interest

Emilio Matricciani served as an Editor-in-Chief of the *Journal of Mathematical Studies of Literature* at the time of manuscript submission. To ensure the integrity of the peer-review process, Emilio Matricciani was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor.

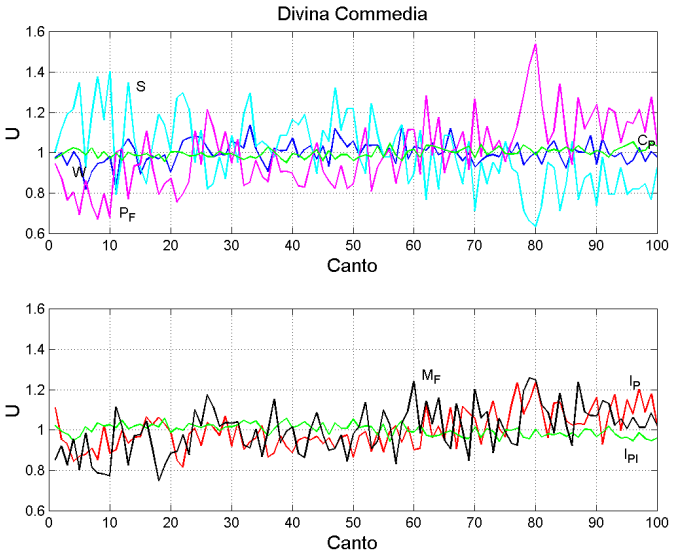


Figure 18. Uniformity index $U(k)$ for the indicated linguistic parameters, *Divina Commedia*. Upper panel: words (W), blue; sentences (S), cyan; characters per word (C_P), green; words per sentence (P_F), magenta. Lower panel: word intervals per sentence (M_F), black; words per interpunction (I_P), red; words per line ($I_{P\ell}$), green.

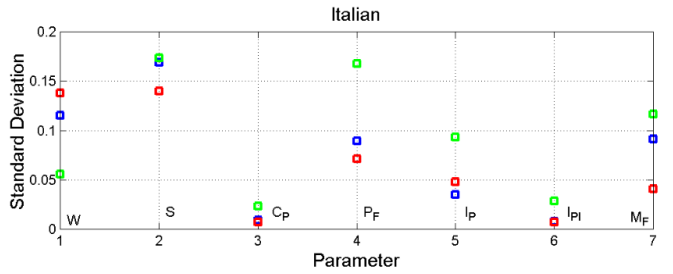


Figure 19. Standard deviation of $U(k)$ for the linguistic parameters indicated on the x-axis. *Eneide*: blue squares; *Sulla natura delle cose*: red squares; *Divina Commedia*: green squares. The values of C_P and $I_{P\ell}$ of the *Eneide* and *Sulla natura delle cose* practically coincide. W: words; S: sentences.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable. This study analyses publicly available literary texts and involves no human participants or animals.

References

[1] Matricciani, E. (2025). A multidimensional mathematical analysis of Dante's *Divina Commedia*. *Journal of Mathematical Studies of Literature*, 1(1), 4-20. [CrossRef]

[2] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. [CrossRef]

[3] Nagy, B. (2022). Rhyme in classical Latin poetry: Stylistic or stochastic? *Digital Scholarship in the Humanities, 37*(4), 1097-1118. [CrossRef]

[4] Duckworth, G. E. (1969). *Vergil and Classical Hexameter Poetry: A Study in Metrical Variety*. Ann Arbor, MI: University of Michigan Press. [CrossRef]

[5] Young, A. M. (1932). Schematized word order in Vergil. *The Classical Journal, 27*(7), 515-522. https://www.jstor.org/stable/3289987?seq=1

[6] Roberts, L. (1977). *A Concordance of Lucretius*. New York: Garland. https://search.catalog.loc.gov/instances/ac86362e-be55-5ade-a232-de94d5f949b7?option=lccn&query=77070759

[7] Allen, W. S. (1973). *Accent and Rhythm: Prosodic Features of Latin and Greek. A Study in Theory and Reconstruction*. Cambridge: Cambridge University Press. https://archive.org/details/accentrhythmpros0000alle/page/n3/mode/2up

[8] Duffell, M. J. (2012). Syllabic verse. In R. Greene & S. Cushman (Eds.), *The Princeton Encyclopedia of Poetry and Poetics* (4th ed., pp. 1388-1390). Princeton, NJ: Princeton University Press. https://www.oxfordreference.com/display/10.1093/acref/9780190681173.001.0001/acref-9780190681173-e-1101

[9] West, M. L. (1982). *Greek Metre*. Oxford: Clarendon Press. https://ancientlibrary.yatsishin.com/download/28/pdf/28.pdf

[10] Parkes, M. B. (1993). *Pause and Effect: An Introduction to the History of Punctuation in the West*. Berkeley: University of California Press. [CrossRef]

[11] Feeney, D. (1993). The *Aeneid* as a poem of history. *The Classical Outlook, 70*(3), 94-96 (first published in *Omnibus*, 23, 1992, 19-21). https://www.jstor.org/stable/43937473?seq=1

[12] Fendel, V. (2025). Switchpoints: Half-lines in Vergil's *Aeneid* from an attention perspective. *Rosetta, 30*, 9-27. [CrossRef]

[13] Thomas, R. F. (2001). *Virgil and the Augustan Reception*. Cambridge: Cambridge University Press. [CrossRef]

[14] Duckworth, G. E. (1962). *Structural Patterns and Proportions in Vergil's Aeneid: A Study in Mathematical Composition*. Ann Arbor, MI: University of Michigan Press. [CrossRef]

[15] Kenney, E. J. (2007). Lucretian texture: Style, metre and rhetoric in the *De rerum natura*. In S. Gillespie & P. Hardie (Eds.), *The Cambridge Companion to Lucretius* (pp. 92-110). Cambridge: Cambridge University Press. [CrossRef]

[16] Farrell, J. (2007). Lucretian architecture: the structure and argument of the De rerum natura. *The Cambridge Companion to Lucretius*, (88), 76-91. https://repository.upenn.edu/server/api/core/bitstreams/831447c7-1dc2-41e6

-a310-b043b70d68f5/content

[17] Taylor, B. (2020). *Lucretius and the Language of Nature*. Oxford: Oxford University Press. [CrossRef]

[18] Sedley, D. (1998). *Lucretius and the Transformation of Greek Wisdom*. Cambridge: Cambridge University Press. https://archive.org/details/lucretiustransfo0000sedl

[19] Sedley, D. (1999). Lucretius' use and avoidance of Greek. In J. N. Adams & R. G. Mayer (Eds.), *Aspects of the Language of Latin Poetry* (Proceedings of the British Academy, 93, pp. 227-246). Oxford: Oxford University Press. https://www.thebritishacademy.ac.uk/documents/3903/93p227.pdf

[20] Swanson, D. C. (1962). *A Formal Analysis of Lucretius' Vocabulary*. Minneapolis, MN: Perine Book Company.

[21] Robey, D. (1993). Scanning Dante's *Divine Comedy*: A computer-based approach. *Literary and Linguistic Computing, 8*(2), 81-84. [CrossRef]

[22] Dexter, J. P., Tripuraneni, N., Dasgupta, T., Kannan, A., Brofos, J. A., Bonilla Lopez, J. A., ... & Chaudhuri, P. (2017). Quantitative criticism of literary relationships. *Proceedings of the National Academy of Sciences, 114*(16), E3195-E3204. [CrossRef]

[23] Perveen, T., Munawar, H., Zahra, M., & Lodhi, M. A. (2025). Corpus stylistic exploration of lexical patterns, semantic fields, and contextual meaning in Dante's *Divine Comedy*. *Contemporary Journal of Social Science Review, 3*(4), 114-131. https://www.researchgate.net/figure/Review-of-the-previous-researches_tbl2_398932256

[24] Re, A. (2026). Applicazioni alle lingue e alle letterature classiche delle tecniche di analisi digitale. *Umanistica Digitale, 24*, 263-288. [CrossRef]

[25] Matricciani, E. (2019). Deep language statistics of Italian throughout seven centuries of literature and empirical connections with Miller's 7 ± 2 law and short-term memory. *Open Journal of Statistics, 9*, 373-406. [CrossRef]

[26] Matricciani, E. (2020) A Statistical Theory of Language Translation Based on Communication Theory. *Open Journal of Statistics, 10*, 936–997. [CrossRef]

[27] Matricciani, E. (2025). Revealing Short-Term Memory Communication Channels Embedded in Alphabetical Texts: Theory and Experiments. *Information, 16*(10), 847. [CrossRef]

[28] Matricciani, E. (2025). Equivalent processors modelling the short-term memory. *International Journal of Pure and Applied Mathematics Research*, 1—28. [CrossRef]

[29] Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. [CrossRef]

[30] Matricciani, E. (2024). Multi–Dimensional Data Analysis of Deep Language in JRR Tolkien and CS Lewis Reveals Tight Mathematical Connections. *AppliedMath, 4*(3), 927-949. [CrossRef]

[31] Matricciani, E. (2023) Readability Indices Do Not Say It All on a Text Readability. *Analytics, 2*, 296-314. [CrossRef]

[32] De Sanctis, F. (2006). *Storia della letteratura italiana* (first published 1870). Milano: Rizzoli. https://skypescuola.wordpress.com/wp-content/uploads/2015/05/de-sanctis-storia-della-letteratura-italiana.pdf

[33] Ferroni, G. (2021). *Storia della letteratura italiana. Dalle origini al Quattrocento*. Milano: Mondadori Università. https://www.mondadoriuniversita.it/libri-universitari/storia-della-letteratura-italiana-giulio-ferroni/

[34] Frare, P., & Brenna, S. (2023). *Dalle origini a Leopardi. La letteratura italiana e le sue grandi opere*. Milano: Pearson. https://publicatt.unicatt.it/handle/10807/230632

[35] Maiden, M. (1995). *A Linguistic History of Italian*. London: Longman. [CrossRef]

Appendix A. Raw textual data and processing

The Latin texts were extracted from the best available digital editions of the two poems at https://www.thelatinlibrary.com (last accessed 1 July 2026). The classical Italian translations of the *Aeneid* (*Eneide*) and *De Rerum Natura* (*Sulla natura delle cose*) were downloaded from https://liberliber.it/ (last accessed 1 July 2026).

The processing, although time consuming, consists of two steps:

1) Extraction of the raw data of each book: number of lines, characters, words, sentences and interpunctions; the procedure was repeated 12 times for the *Aeneid* and 6 times for *De Rerum Natura*.

2) Further processing to calculate the deep-language parameters, their scatterplots and various statistics.

For the first step, I did not use specialized software, but a tool freely available to any reader and writer, namely Microsoft Word. Before processing, I manually deleted titles, footnotes and other extraneous material, leaving only the plain text.

For each text block (a book), Word directly provides the number of characters and words. The other raw data are easy to obtain. The number of sentences is calculated in three steps: first, every full stop is replaced with a full stop, an action that does not change the text but returns the number of substitutions, i.e., the number of full stops; the same procedure is then repeated for question marks (second step) and exclamation marks (third step). The sum of the three totals gives the number of sentences. The same procedure gives the number of commas, colons and semicolons; adding these to the number of sentences gives the number of interpunctions. The number of lines is indicated explicitly in the original files.

For the second step, I used software codes developed for my previous papers to obtain scatterplots, regression lines and statistics.

Appendix B. List of mathematical symbols

Symbol	Definition
C_P	Characters per word
C, W	Total number of characters and of words in a text
$d(k)$	Normalized distribution of $x(k)$ across books, Eq. (7)
I_P	Word interval (words per interpunction)
$I_{P\ell}$	Word interval per line (words per line)
M_F	Word intervals per sentence
N	Number of books (or canti)
n_C	Characters
n_I	Interpunctions (punctuation marks)
n_L	Lines
n_S	Sentences
n_W	Words
P_F	Words per sentence
S_P	Syllables per word (Figure 1)
$U(k)$	Uniformity index, Eq. (8)
$x(k)$	Mean value of a parameter in book k

Appendix C. Arithmetic mean of $U(k)$

Starting from the definition, Eq. (8), we get:

$$U(k) = \frac{x(k)/\sum_{j=1}^{N} x(j)}{1/N} = N\,\frac{x(k)}{\sum_{j=1}^{N} x(j)}$$

By setting $\sum_{j=1}^{N} x(j) = C_x$, the arithmetic mean $\langle U \rangle$ is given by:

$$\langle U \rangle = \frac{1}{N} \sum_{k=1}^{N} U(k) = \frac{1}{N} \frac{N}{C_x} \sum_{k=1}^{N} x(k) = \frac{C_x}{C_x} = 1$$

Appendix D. Further scatterplots of *De Rerum Natura*

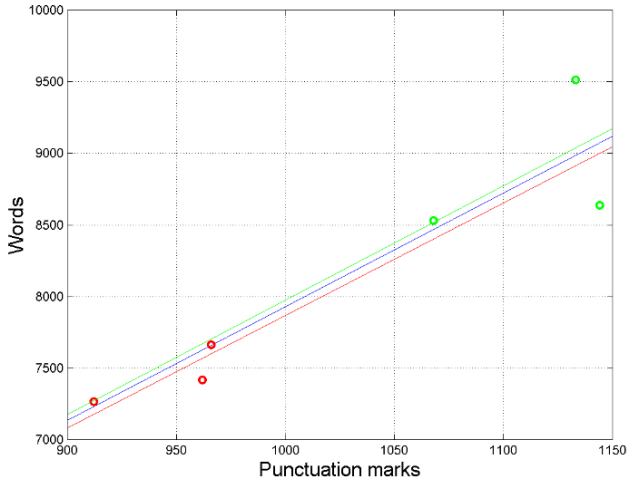


Figure A1. *De Rerum Natura*. Scatterplot between words and punctuation marks. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem.

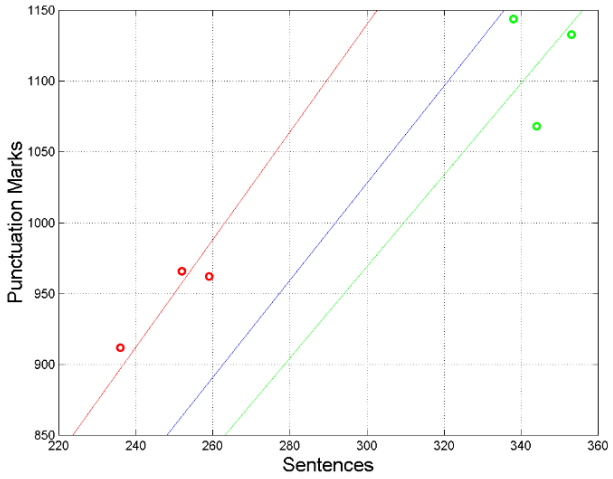


Figure A2. *De Rerum Natura*. Scatterplot between punctuation marks and sentences. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem.

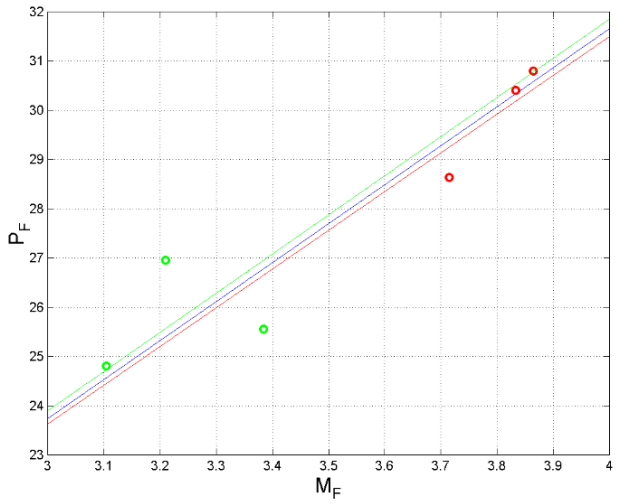


Figure A3. *De Rerum Natura*. Scatterplot between P_F and M_F. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem.

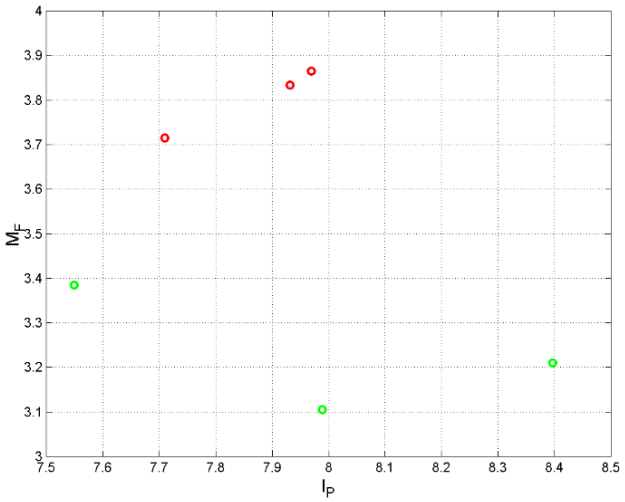


Figure A4. *De Rerum Natura*. Scatterplot between M_F and I_P. Red circles and red line: first triad; green circles and green line: second triad. The blue line is the regression line calculated for the whole poem.

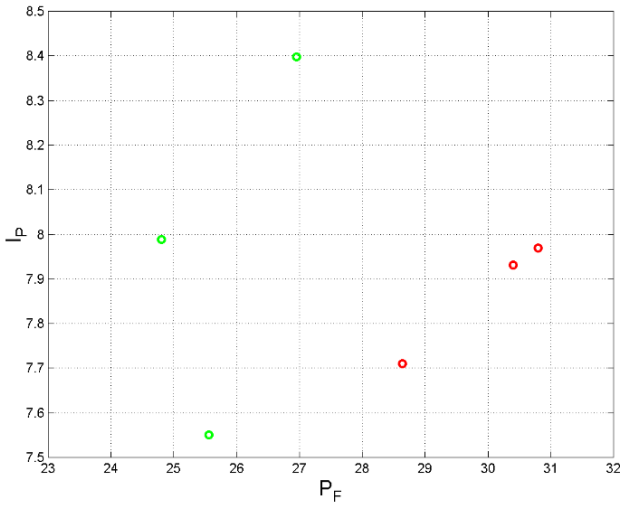


Figure A5. *De Rerum Natura*. Scatterplot between I_P and P_F. Red circles: first triad; green circles: second triad.



Emilio Matricciani was Professor of Telecommunications at Politecnico di Milano, Milan, Italy, until his retirement in 2022, where he taught Information Theory, Probability, and Satellite and Terrestrial Communication Systems to graduate students. He still teaches Scientific Writing to PhD students at Sapienza University of Rome and at the University of Naples Federico II. He remains active in research on satellite and deep-space communications and on radio propagation at millimetre waves. He has also studied literary texts with the tools of communication theory, proposing a mathematical theory of the deep-language structure of alphabetical texts and identifying multiple linguistic channels in literary texts and their connection with human short-term memory. He is included in the Stanford University list of the world's top 2% scientists. He serves on the Editorial Boards of *Information* and *Open Journal of Statistics* and is the Editor-in-Chief of the *Journal of Mathematical Studies of Literature*. (Email: emilio.matricciani@polimi.it)