

A Mathematical Analysis of Texts of Greek Classical Literature and Their Connections

Emilio Matricciani

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano, Milan, Italy

Email: Emilio.Matricciani@polimi.it

How to cite this paper: Matricciani, E. (2025) A Mathematical Analysis of Texts of Greek Classical Literature and Their Connections. *Open Journal of Statistics*, 15, 1-34.

<https://doi.org/10.4236/ojs.2025.151001>

Received: December 26, 2024

Accepted: February 4, 2025

Published: February 7, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A multi-dimensional mathematical theory applied to texts belonging to the classical Greek Literature spanning eight centuries reveals interesting connections between them. By studying words, sentences, and interpunctuations in texts, the theory defines deep-language variables and linguistic channels. These mathematical entities are due to writer's unconscious design and can reveal connections between texts far beyond writer's awareness. The analysis, based on 3,225,839 words contained in 118,952 sentences, shows that ancient Greek writers, and their readers, were not significantly different from modern writers/readers. Their sentences were processed by a short-term memory modelled with two independent processing units in series, just like modern readers do. In a society in which people were used to memorize information more often than modern people do, the ancient writers wrote almost exactly, mathematically speaking, as modern writers do and for readers of similar characteristics. Since meaning is not considered by the theory, any text of any alphabetical language can be studied exactly with the same mathematical/statistical tools and comparisons are possible, regardless of different languages and epochs of writing.

Keywords

Alphabetical Languages, Deep-Language Variables, Extended Short-Term Memory, Greek Literature, Iliad, Linguistic Channels, New Testament, Odyssey, Universal Readability Index

1. Introduction

A multi-dimensional mathematical theory of alphabetical texts can reveal interesting connections between authors, between texts of the same author or even between texts in different languages, including translations, regardless of the epoch

of writing. In recent years we have developed, in a series of papers [1]-[10], what we believe is a mathematical/statistical theory that fits the purpose of studying texts in a multi-dimensional mathematical framework by using linguistic variables authors are not aware of. For example, this analysis has recently [11] revealed strong connections between *The Lord of the Rings* (J.R.R. Tolkien) and *The Chronicles of Narnia* and *The Space Trilogy* (C.S. Lewis), therefore confirming both the conclusions reached by scholars of English Literature and the power of the mathematical theory, based on simple, understandable and easily calculable variables. The theory can also reveal connections with the short-term memory (STM) and the so-called extended short-term memory (E-STM) of readers and writers as well [9] [10], since writers are also readers of their own texts.

The theory considers the number of words, sentences and interpunctuations (punctuation marks). It defines deep-language variables and linguistic channels within texts which are due to writer's unconscious design and, therefore, can reveal connections between texts far beyond writer's awareness.

Since meaning is not considered by the theory, any text of any alphabetical language can be studied exactly with the same mathematical/statistical tools. Today, many scholars are working hard to arrive at a "semantic communication" theory, or "semantic information" theory which should include some rudiments on meaning, but the results are still in their infancy and far from useful applications [12]-[19]. These theories, as those concerning the short-term memory [20]-[47], have not considered most of the main "ingredients" of our theory, which can be very easily retrieved in alphabetical texts of any epoch.

The aim of this paper is to apply the theory to important texts of the classical Greek Literature and New Testament (NT). The analysis will indicate that these ancient writers, and their readers, were not significantly different from the modern writers/readers. This find is quite interesting because, in a society in which most people were illiterate and used to memorize oral information more than modern people do, the ancient writers wrote almost exactly, mathematically speaking, as the modern writers do and for readers with similar characteristics, therefore underlining the long-term persistence of human mind processing tools. Of course, differences are found, as in modern texts, only because of the genre of the text.

After this introductory section, Section 2 presents the database of the classical Greek Literature texts studied. Section 3 defines the deep-language variables and establishes some inequalities in calculating their mean values; Section 4 applies a useful graphical tool, namely a vector representation of texts; Section 5 recalls the theory of linguistic channels; Section 6 reports and discusses the performance of important linguistic channels; Section 7 recalls and calculates a universal readability index for each text and compares them on a common ground; Sections 8 and 9 study the short-term memory of ancient Greek writers/readers, and show that it is just like that of modern writers/readers. Finally, Section 10 draws a conclusion. Several Appendices report numerical data useful for applying the theory in each section and guide scholars who wish to apply the theory to their own texts.

2. Database of Ancient Greek Literary Texts

In this section, we introduce the database of classical Greek literature texts that have been mathematically studied. The choice of these texts was dictated by their well-known importance in cultural history, and also by the availability of digital versions. **Table 1** lists authors and books concerning history, geography and philosophy (referred to as Greek-1 texts), poetry and theatre (Greek-2 texts). This is a large sample of classical Greek Literature. Notice that *Iliad* and *Odissey*, although traditionally attributed to the mythical Homer, are studied separately because they were likely written by different authors. **Table 2** lists the texts of the New Testament.

Table 1. Total number of characters, words, sentences and interpunctons contained in the indicated texts of authors belonging to history and other disciplines (Greek-1) and to Poetry and Theatre (Greek-2).

	Texts	Characters	Words	Sentences	Interpunctons
History and other disciplines (Greek-1)					
Aeneas Tactitian (IV century BC) military communications	Poliocertica	75,266	13035	579	1714
Aeschines (389-314 BC) statesman, orator	Against Ctesiphon, Against Timarchus, On the Embassy	398,924	69,764	2555	11,381
Aristides (530-462 BC) statesman, orator	Orationes	1,205,412	222,272	8731	30,771
Aristotle (384-322 BC) philosopher	De Partibus Animalium, Historia Animalium, Physica, Metaphysica, Politica, De Caelo, Politica, Meteorologica, Topica	2,386,790	509,646	17,790	65,252
Demosthenes (384-322 BC) Statesman, orator	Phylippics 1-4; Adversus Leptinem, In Midiam, Adversus Androtonem, In Aritocratem, In Timocratem, In Aristogitonem 1-2, In Aphobum 1-2, Contra Onetorem 1-2, Olyntiaches	560,697	111,179	4351	16,812
Flavius Josephus (37AD-c. 100 AD) historian	The Jewish War, Antiquities of the Jews	2,333,545	424,482	13,272	40,910
Herodotus (484-425 BC) historian and geographer	Histories 2-9	820,761	157,490	5945	19,082
Pausanias (110-180 AD) geographer	Description of Greece 1-10	987,016	176,864	6272	20,502
Plato (428-348 BC) philosopher	The Republic, The Apology of Socrates	547,962	111,125	6566	20,591
Plutarch (48-125 AD) historian	Parallel Lives	2,750,711	499,683	17,905	64,365

Continued

Polybius (206-124 BC) historian	Histories	1,530,968	256,495	8830	28,997
Strabo (60 BC-21 AD) geographer	Geographica	821,855	158,993	5301	18,356
Thucydides (460-404 BC) historian	Histories	814,309	151,906	4410	17,158
Xenophon (430-354 BC) historian	Anabasis	297,161	57,186	2420	7634
Poetry and theatre (Greek-2)					
Aeschylus (525-456 BC) playwright	Agamemnon	43,088	8250	611	1451
Aesop (620-564 BC) fabulist	Fables	204,913	39,122	2172	7437
Euripides (480-406 BC) playwright	Medea, Iphigenia in Aulis	88,964	17,970	1392	3455
Homer (IX or VIII century BC) poet	Iliad	548,830	111,878	3830	15,719
Homer (IX or VIII century BC) poet	Odyssey	427,148	87,282	3591	15,259
Pindarus (518-438 BC) poet	Isthmean Odes, Nemean Odes, Olympian Odes, Pythian Odes	114,732	21,140	941	3299
Sofocles (497-406 BC) playwright	Electra, Oedipus at Colonus	95,532	20,077	1488	3809
All		17,054,584	3,225,839	118,952	413,954

Table 2. Total number of characters, words, words, sentences and interpunctons contained in the indicated books of the New Testament. The genealogies in *Matthew* (verses 1.1 - 1.17) and in *Luke* (verses 3.23 - 3.38) have been deleted, as done in [4] [6] [7] [48], for not biasing the statistical analyses.

Text	Characters	Words	Sentences	Interpunctons
Matthew	88,605	18,121	914	2546
Mark	56,452	11,393	612	1595
Luke	95,180	19,384	964	2763
John	70,418	15,503	848	2310
Acts	95,647	18,757	760	2163
Hebrews	26,317	4940	164	711
Apocalypse	45,970	9870	333	1280

We used digital texts (WinWord digital files) and counted the number of characters, words, sentences, and interpunctons (punctuation marks). Before doing so, we deleted the titles, footnotes, and other extraneous material in the digital

texts. The count is very simple, although time-consuming. Winword directly provides the number of total words and their characters. The number of total sentences is calculated by using WinWord to replace every full stop with a full stop; this replacement gives the number of full stops. The same procedure was repeated for question marks and exclamation marks. The sum of the three totals gives the total number of sentences. The same procedure gives the total number of commas, colons and semicolons. The sum of these latter values with the total number of sentences gives the total number of interpunctuations. The same procedure was applied to the New Testament books. These latter data were also used for previous studies [4] [6] [7] [49]. The original Greek texts of **Table 1** were downloaded from <https://www.perseus.tufts.edu/> (last accessed on 19 October 2024). The New Testament books were downloaded as indicated in [4] [6] [7] [49].

Punctuation marks (interpunctuations) were introduced in the *scriptio continua* by ancient well-educated readers acting as “editors” [48] [50]-[57], respectful of the original text and its meaning. Very likely they maintained the correct subdivision in sentences and introduced interpunctuations within sentences for not distorting meaning and emphasis. In other words, we can hypothesize that the author introduced interpunctuations. The mathematical theory, however, is very robust against slightly different versions of the Greek texts because it never considers meaning. If a word is missing or substituted with another, or if a short text is missing, it does not affect the statistical analysis. This applies also to the quality of the Greek used. This a point of force of the theory.

In the next section, we recall the theory of deep-language variables.

3. Deep-Language Variables of Texts, Statistical Means and Minimum Values

In this section, we first define and then explore four linguistic variables, which are termed deep-language variables [1] [2]. These variables are very useful because they are not consciously controlled by writers, therefore, they can reveal connections between texts/authors and can also indicate likely influence of an author on another one, as shown by Lewis and Tolkien [11]. To avoid possible misunderstandings, these variables refer to the “surface” structure of texts, not to the “deep” structure mentioned in cognitive theory. We first recall their definition and then prove useful inequalities.

3.1. Deep-Language Variables

Let n_C , n_W and n_I be respectively the number of characters, words and interpunctuations (punctuation marks) calculated in disjoint blocks of texts, such as chapters or any other subdivisions, then four deep-language variables are defined (Appendix A lists the mathematical symbols used in the present paper).

The number of characters per word, C_P :

$$C_P = \frac{n_C}{n_W} \quad (1)$$

The number of words per sentence, P_F :

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

The number of interpunctuations per word, referred to as the word interval, I_P :

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

The number of word intervals, n_{I_p} , per sentence, M_F :

$$M_F = \frac{n_{I_p}}{n_S} \tag{4}$$

Equation (4) can be written also as $M_F = P_F / I_P$. **Table 3** & **Table 4** report the

Table 3. Mean values of deep-language variables C_P , P_F , I_P , M_F in the indicated authors and texts of Greek Literature.

	$\langle C_P \rangle$	$\langle P_F \rangle$	$\langle I_P \rangle$	$\langle M_F \rangle$	$\langle G_U \rangle$	Years	Multiplicity factor α	Mismatch index I_M
Greek-1								
<i>Aeneas Tactitian</i>	5.77	23.18	7.71	3.01	43.1	9.6	0.352	-0.450
<i>Aeschines</i>	5.72	28.03	6.14	4.56	50.7	7.8	0.048	-0.909
<i>Aristides</i>	5.42	26.42	7.26	3.63	47.3	8.4	0.906	-0.049
<i>Aristoteles</i>	4.68	29.29	7.84	3.72	48.7	8.0	1.085	0.041
<i>Demosthenes</i>	5.04	25.80	6.62	3.90	54.3	7.4	0.384	-0.445
<i>Flavius Josephus</i>	5.50	32.17	10.43	3.09	25.2	15	1.802	0.286
<i>Herodotus</i>	5.21	26.56	8.26	3.22	42.6	9.6	1.184	0.084
<i>Pausanias</i>	5.58	28.40	8.64	3.28	36.5	11.5	0.825	-0.096
<i>Plato</i>	4.93	18.63	5.49	3.32	68.0	5.2	4.538	0.639
<i>Plutarch</i>	5.50	29.35	7.81	3.73	42.2	9.7	1.060	0.029
<i>Polybius</i>	5.97	29.19	8.88	3.30	31.5	12.5	0.996	-0.002
<i>Strabo</i>	5.17	30.94	8.75	3.55	38.7	10.9	0.311	-0.525
<i>Thucytides</i>	5.36	35.10	8.90	3.96	34.9	11.7	0.097	-0.823
<i>Xenophon</i>	5.20	24.62	7.59	3.25	48.1	8.2	0.612	-0.241
Greek-2								
<i>Aeschylus</i>	5.22	14.34	5.75	2.48	68.5	5.3	3.117	0.514
<i>Aesop</i>	5.24	18.29	5.28	3.46	65.6	5.6	1.360	0.153
<i>Euripides</i>	4.95	13.54	5.23	2.57	74.9	4.0	7.733	0.771
<i>Homer's Iliad</i>	4.91	29.61	7.13	4.15	50.9	7.9	0.104	-0.812
<i>Homer's Odissey</i>	4.89	24.37	5.72	4.26	61.5	6.2	0.214	-0.647
<i>Pindarus</i>	5.43	23.13	6.45	3.61	53.7	7.5	0.180	-0.694
<i>Sofocles</i>	4.76	14.26	5.31	2.68	75.1	4.0	6.279	0.725
<i>All</i>	5.29	28.51	8.06	3.56	42.9	--	--	--

Table 4. Mean values of deep-language variables C_P , P_F , I_P , M_F in the indicated book of the New Testament.

Book	$\langle C_P \rangle$	$\langle P_F \rangle$	$\langle I_P \rangle$	$\langle M_F \rangle$	$\langle G_U \rangle$	Years	Multiplicity factor α	Mismatch index I_M
<i>Matthew</i>	4.91	20.27	7.18	2.83	55.61	7.3	20.66	0.908
<i>Mark</i>	4.96	19.14	7.17	2.68	56.14	7.2	18.35	0.897
<i>Luke</i>	4.91	20.47	7.11	2.89	55.68	7.3	20.21	0.906
<i>John</i>	4.54	18.56	6.79	2.74	62.21	6.1	25.75	0.925
<i>Acts</i>	5.10	25.47	8.77	2.91	41.35	9.8	9.41	0.808
<i>Hebrews</i>	5.33	32.00	7.02	4.53	47.71	8.4	0.05	-0.912
<i>Apocalypse</i>	4.66	30.70	7.79	3.97	48.95	8.1	0.38	-0.448

mean values of these variables, the other variables in [Table 3](#) and [Table 4](#) will be discussed in the following sections.

Notice that all mean values have been calculated by weighing each text with its number of words to avoid that shorter texts weigh statistically as much as long ones. In other words, any text considered weighs the number of its words, compared to the total number of words. We have also used this method to calculate the mean values of the data bank of Greek-1 plus Greek-2 (last line in [Table 3](#)). In this case, for example, the statistical weight of *Aeneas Tactitian* is $13035/3,225,839 \approx 0.004$ (see [Table 1](#)) while the weight of *Aristotle* is $509646/3,225,839 \approx 0.1580$.

The mean values of these variables can be calculated from the sample totals listed in [Table 1](#) and [Table 2](#). However, do not be misled; these values are not equal to the arithmetic or to the statistical means, as we prove now.

3.2. Inequalities

Let M be the number of samples (*i.e.*, number of disjoint blocks of text, such as chapters or books), then, for example, the statistical mean value $\langle P_F \rangle$, is given by

$$\langle P_F \rangle = \sum_{k=1}^M P_{F,k} \times (n_{W,k} / W) \quad (5)$$

where $W = \sum_{k=1}^M n_{W,k}$ is the total number of words. Notice that

$\langle P_F \rangle \neq \frac{1}{M} \sum_{k=1}^M P_{F,k} \neq \sum_{k=1}^M n_{W,k} / \sum_{k=1}^M n_{S,k} = W/S$, where S is the total number of sentences.

For example, for Aristotle $W = 509646$ and $S = 17790$, [Table 1](#). These values would give the average $P_F = W/S = 509646/17790 = 28.65$, while the statistical mean (calculated on the nine books listed in [Table 1](#)) $\langle P_F \rangle = 29.29 > 28.65$.

In general, the average values calculated from sample totals are always smaller than their statistical means, therefore they give lower bounds, as we prove in the following.

Let us consider, for example, the parameter P_F . Because of Chebyshev's inequality ([\[58\]](#), inequality 3.2.7), we can write Equation (5) as:

$$\langle P_F \rangle = \sum_{k=1}^M \frac{n_{W_k}}{n_{S_k}} \frac{n_{W_k}}{W} \geq \frac{1}{M} \sum_{k=1}^M \frac{n_{W_k}}{n_{S_k}} \sum_{k=1}^M \frac{n_{W_k}}{W} = \frac{1}{M} \sum_{k=1}^M \frac{n_{W_k}}{n_{S_k}} \tag{6}$$

Equation (6) states that the mean calculated with samples weighted $1/M$ (arithmetic mean) is smaller than (or equal to) the mean calculated with samples weighted n_{W_k}/W .

Now, again for Chebyshev’s inequality, we get:

$$\sum_{k=1}^M \frac{n_{W_k}}{n_{S_k}} \geq \frac{1}{M} \sum_{k=1}^M n_{W_k} \sum_{k=1}^M \frac{1}{n_{S_k}} = \frac{W}{M} \sum_{k=1}^M \frac{1}{n_{S_k}} \tag{7}$$

Further, for Cauchy-Schwarz’s inequality (or by the fact that the harmonic mean is less than, or equal to, the arithmetic mean), we get:

$$\sum_{k=1}^M \frac{1}{n_{S_k}} \geq \frac{M^2}{\sum_{k=1}^M n_{S_k}} \tag{8}$$

Finally, by inserting these inequalities in (6), we get:

$$\langle P_F \rangle \geq \frac{W}{M^2} \frac{M^2}{\sum_{k=1}^M n_{S_k}} = \frac{W}{S} \tag{9}$$

Equation (9) establishes that the statistical mean calculated with samples weighted n_{W_k}/W is greater than (or equal to) the average calculated with sample totals. The values given by these three methods of calculation coincide only if all texts are perfectly identical, *i.e.*, with the same number of characters, words, sentences and interpunctons, a case improbable.

The mean values of **Table 3** and **Table 4** (or their minimum values directly calculated from the totals, as discussed above) are useful for a first assessment of how “close”, or similar, mathematically, texts are by defining linear combinations of deep-language variables [1]. Texts are modeled as vectors, the representation of which is briefly recalled in the next section.

4. Vector Representation of Texts

Let us consider the six vectors of the indicated components of deep-language variables, $\mathbf{R}_1 = (\langle C_p \rangle, \langle P_f \rangle)$, $\mathbf{R}_2 = (\langle M_f \rangle, \langle P_f \rangle)$, $\mathbf{R}_3 = (\langle I_p \rangle, \langle P_f \rangle)$,

$\mathbf{R}_4 = (\langle C_p \rangle, \langle M_f \rangle)$, $\mathbf{R}_5 = (\langle I_p \rangle, \langle M_f \rangle)$, and $\mathbf{R}_6 = (\langle I_p \rangle, \langle C_p \rangle)$, and their resulting vector sum:

$$\mathbf{R} = \sum_{k=1}^6 \mathbf{R}_k \tag{10}$$

By considering the coordinates x and y of Equation (10), the scatterplot of their ending points is shown in **Figure 1**, where X and Y are normalized coordinates so that Sofocles (black triangle) is at the origin ($X = 0, Y = 0$) and Flavius Josephus (blue triangle) is at ($X = 1, Y = 1$), through the linear transformations:

$$X = \frac{x - x_{Sofocles}}{x_{Flavius} - x_{Sofocles}} \tag{11a}$$

2) *Iliad* and *Odyssey* are significantly distant, although they are traditionally attributed to *Homer*.

3) The three synoptic gospels (*Matthew*, *Mark* and *Luke*) are each other very close. *John* almost coincides with *Aesop*. The gospels are nearer to Greek-2 than to Greek-1. *Acts* is nearer to historians (e.g., *Herodotus*) than to the Synoptics. *Hebrews* and *Apocalypse* are near each other and are clearly distinct from the other NT books, likely indicating they were written either by the same writer or by writers belonging to the same Christian group [6]. More studies, connections and details on these NT books can be found in [4] [6] [7].

5. Linguistic Channels and Signal-to-Noise Ratio

The representation of texts as vectors gives a necessary but not sufficient condition of possible connections and influence of authors on each other, e.g., see in [6] the discussion about the couple *Aesop-John*. The linguistic channels, always present in texts [3], can further assess similarity and likely dependence because they provide a “fine-tuning” analysis of authors/texts’ connections.

In this section, we first recall the most important linguistic channels in alphabetical texts; secondly the general theory of linear channels concerning the processing of two experimental scatterplots; and thirdly the theory applied to the particular case of a single scatterplot.

Since the theory deals with linear regressions, it can be applied and is useful in any science field in which linear relationships fit experimental data. The “performance” of a channel is measured by a suitable signal-to-noise ratio.

5.1. Linguistic Channels

In texts, we can always define at least four linguistic linear channels [3] [11], namely:

- 1) Sentence channel (S-channel)
- 2) Interpunctuations channel (I-channel)
- 3) Word interval channel (WI-channel)
- 4) Characters channel (C-channel).

In S-channels, the number of sentences of two texts is compared for the *same* number of words. Notice that, as far as we know, only the theory of linguistic channels allows this comparison. These channels describe how many sentences the author of text j writes, compared to the writer of text k (reference text), by using the same number of words. Therefore, these channels are more linked to P_F than to the other variables. Very likely they reflect the style of the writer.

In I-channels, the number of word intervals I_p 's of two texts is compared for the *same* number of sentences. These channels describe how many short texts between two contiguous punctuation marks (of length I_p words) two authors use; therefore, these channels are more linked to M_F than to the other variables. Since M_F is connected to the E-STM, I-channels are more related to the second buffer of readers’ E-STM than to the style of the writer.

In WI-channels, the number of words (*i.e.*, I_p) contained in a word interval is compared for the *same* number of interpunctuations. These channels are more linked to I_p than to other variables, therefore WI-channels are more related to the first buffer of readers' E-STM than to the style of the writer.

In C-channels, the number of characters of two texts is compared for the same number of words. These channels are more related to the language used, e.g. Greek in this case, than to the other variables.

5.2. Theory of Linguistic Channels

An independent (reference) variable x (e.g., n_w in S-channels) and a dependent variable y (e.g., n_s) can be related by a regression line (slope m) passing through the origin:

$$y = mx \quad (12)$$

Let us consider two texts Y_k and Y_j . For we can write Equation (12) for the same couple of parameters. In both cases, Equation (12) does not give their full relationship because it only connects the mean conditional values. More general linear relationships must also consider the scattering of the data—measured by the correlation coefficients r_k and r_j —around the regression lines (slopes m_k and m_j):

$$y_k = m_k x + n_k \quad (13)$$

$$y_j = m_j x + n_j$$

Equation (12) connects y to x only on the average. Equation (13) introduces additive “noise” n_k and n_j , with zero mean value. The noise is due to the correlation coefficient $|r| \neq 1$.

In these channels, which consider two scatterplots, we compare two texts by eliminating x , e.g., the number of sentences in two texts—for an equal number of words—by considering not only the mean relationship but also the scattering of the data.

As recalled before, we refer to this communication channel as the “sentences channel” and to this processing as “fine tuning” because it deepens the analysis of the data and provides more insight into the relationship between two texts. The mathematical theory follows.

By eliminating x , from Equation (13) we obtain the linear relationship between—now—the sentences in text Y_k (reference, input text) and the sentences in text Y_j (output text):

$$y_j = \frac{m_j}{m_k} y_k - \frac{m_j}{m_k} n_k + n_j \quad (14)$$

Compared with the independent (input) text Y_k , the slope m_{jk} is given by

$$m_{jk} = \frac{m_j}{m_k} \quad (15)$$

The noise source that produces the correlation coefficient between Y_k and Y_j

is given by

$$n_{jk} = -\frac{m_j}{m_k} n_k + n_j = -m_{jk} n_k + n_j \quad (16)$$

The “regression noise-to-signal ratio”, R_m , due to $|m_{jk}| \neq 1$, of the channel is given by:

$$R_m = (m_{jk} - 1)^2 \quad (17)$$

The unknown correlation coefficient r_{jk} between y_j and y_k is given by:

$$r_{jk} = \cos \left| \arccos(r_j) - \arccos(r_k) \right| \quad (18)$$

The “correlation noise-to-signal ratio”, R_r , due to $|r_{jk}| < 1$, of the channel that connects the input text Y_k to the output text Y_j is given by:

$$R_r = \frac{1 - r_{jk}^2}{r_{jk}^2} m_{jk}^2 \quad (19)$$

Because the two noise sources are disjoint, the total noise-to-signal ratio of the channel connecting text Y_k to text Y_j is given by:

$$R = (m_{jk} - 1)^2 + \frac{1 - r_{jk}^2}{r_{jk}^2} m_{jk}^2 \quad (20)$$

Finally, the total signal-to-noise ratio is given by

$$\begin{aligned} \gamma &= 1/R \\ \Gamma &= 10 \times \log_{10} \gamma \end{aligned} \quad (21)$$

Γ is in dB.

Notice that no channel can yield $|r_{jk}| = 1$ and $|m_{jk}| = 1$ (*i.e.*, $\Gamma = \infty$), a case referred to as the ideal channel, unless a text is compared with itself (self-comparison, self-channel). In practice, we always find $|r_{jk}| < 1$ and $|m_{jk}| \neq 1$. The slope m_{jk} measures the multiplicative “bias” of the dependent variable compared to the independent variable; the correlation coefficient r_{jk} measures how “precise” the linear best fit is. The slope m_{jk} is the source of the regression noise of the channel, the correlation coefficient r_{jk} is the source of the correlation noise. Finally, notice that since the probability of finding $|r| = 1$ and $m = 1$ is practically zero, all channels are always noisy.

5.3. The Channel with a Single Scatterplot: One-to-One Correspondence

To clarify what we mean by a single scatterplot and one-to-one correspondence, let us consider the translation of a text in which we can draw a scatterplot for each linguistic variable. For example, we can display the number of words per chapter in the translated text versus that of the original text. Now, if there were a perfect ideal translation, the translated text would have the same number of words per chapter of the original text, therefore in this case $m = 1$ and $r = 1$ because the scatterplot would collapse to the deterministic 45° linear relationship between

x and y :

$$y_k = x \quad (22)$$

Now, if in the expressions of the previous sub-section we set $m_k = 1$ and $r_k = 1$ (hence, $n_k = 0$), we study a special case of the general theory for which we get the following expressions:

$$m_{jk} = m_j \quad (23)$$

$$R_m = (m_j - 1)^2 \quad (24)$$

$$r_{jk} = \cos \left| \arccos(r_j) - 0 \right| = r_j \quad (25)$$

$$R_r = \frac{1 - r_j^2}{r_j^2} m_j^2 \quad (26)$$

$$R = (m_j - 1)^2 + \frac{1 - r_j^2}{r_j^2} m_j^2 \quad (27)$$

$$\gamma = \frac{1}{R} \quad (28)$$

In other words, we study how the deterministic relationship (22), which describes a noiseless channel, is transformed into the experimental relationship of the noisy channel, because now $m_j \neq 1$ and $|r_{jk}| < 1$.

In conclusion, we can study a single scatterplot with the tools of the general theory, therefore the signal-to-noise ratio γ is still a single index that synthetically describes the relationship between x and y .

In the next section we study the four channels mentioned above.

6. Performance of Linguistic Channels

In this section, we first apply the general theory of linguistic channels to the texts of Greek-1 and Greek-2 (Sections 5.1 and 5.2), then we show how to apply the single scatterplot theory (Section 5.3) by considering, as an example, the translation of *Iliad* from Greek to Italian. Slope and correlation coefficients of the regression lines are reported in Appendices B and C.

6.1. Greek-1

Table 5 reports, for example, Γ in the S-channels. Appendices B and C report Γ for the other three channels. **Table 5** is interpreted as follows. The author/text in the first row is the reference author/text, *i.e.*, the channel input author/text Y_k of the theory; the author/text in the first column is the channel dependent output author/text Y_j . For example, if *Aristides* is the input and *Demosthenes* is the output, then $\Gamma = 22.59$ dB ($\gamma = 181.55$); viceversa, if *Demosthenes* is the input and *Aristides* is the output, then $\Gamma = 23.16$ ($\gamma = 207.01$), a small asymmetry always found in linguistic channels [3]. In other words, for the same number of words, the number of sentences in *Aristides* is transformed into the number of sentences, for the same number of words in *Demosthenes* with a high Γ , and viceversa. This

finding means the two texts share a common style very much, as far as sentences are concerned. The channel is little noisy, the regression line that relates n_s of *Demosthenes* (dependent variable) to n_s of *Aristides* (independent variable) has $m_{jk} = 1.0392$ and $r_{jk} = 0.9982$. Now, in this example since *Aristides* lived before *Demosthenes* the large Γ may indicate that *Aristides* influenced *Demosthenes'* style. In any case, the two texts are much correlated in the S-Channel.

The red and blue colours in **Table 5** highlight the channels with $\Gamma \geq 15$ dB ($\gamma \geq 31.62$), with the following meaning: blue indicates not only that the number of sentences of the input and output texts are much correlated but also that the input author might have influenced the output author because he lived before. Red indicates a large correlation, as in the blue cases, but no likely influence can be supposed because the input author lived after the output author. Similar observations can be made for the other authors/texts and linguistic channels (see Appendix B).

Figure 2 synthesizes the results of the four channels by showing the average Γ calculated by considering the input author (left panel, arithmetic average of the values reported in the corresponding column of **Table 5**) or the output author

Table 5. Greek-1. Average Γ , S-Channel. The author/text in the first row is the reference, *i.e.* the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j . For example, if *Aristides* is the input and *Demosthenes* is the output, then $\Gamma = 22.59$ dB, viceversa, if *Demosthenes* is the input and *Aristides* is the output, then $\Gamma = 23.16$ dB, a small asymmetry always found in linguistic channels. Cases with $\Gamma \geq 15$ dB are highlighted in color: blue indicates not only that the number of sentences of the input and output texts are significantly very similar—for the same number of words—but also that the input author might have influenced the output author because he lived before; red indicates a large similarity but no likely influence can be invoked because the input author lived after the output author. Largest Γ : *Aristides-Herodotus*, $\Gamma = 27.56$ dB; minimum Γ : *Thucydides-Plato*, $\Gamma = -7.06$ dB.

Author	<i>Aeneas</i>	<i>Aeschi</i>	<i>Aristi</i>	<i>Aristo</i>	<i>Demo</i>	<i>Flavius</i>	<i>Hero</i>	<i>Paus</i>	<i>Plato</i>	<i>Plut</i>	<i>Poly</i>	<i>Strabo</i>	<i>Thuc</i>	<i>Xen</i>
<i>Aeneas</i>	∞	5.99	15.31	8.68	16.11	7.42	14.45	11.24	10.10	10.42	8.79	3.39	-1.57	19.05
<i>Aeschines</i>	9.18	∞	8.68	15.81	7.43	8.04	9.64	10.20	5.79	15.63	4.84	18.96	7.60	11.55
<i>Aristides</i>	16.69	7.75	∞	11.95	23.16	12.85	27.31	18.72	7.78	13.43	14.71	5.19	-0.07	16.88
<i>Aristoteles</i>	11.60	16.33	13.32	∞	11.22	13.56	15.01	17.15	6.33	25.57	8.76	12.37	4.85	14.39
<i>Demosthenes</i>	17.48	5.86	22.59	9.26	∞	10.32	18.44	14.07	8.35	10.56	15.00	3.49	-1.48	15.02
<i>Flavius</i>	10.51	10.30	14.69	15.17	12.78	∞	15.63	19.09	5.76	14.27	14.35	8.55	3.14	11.63
<i>Herodotus</i>	15.96	8.99	27.56	13.89	19.27	14.05	∞	22.44	7.52	15.57	13.59	6.26	0.77	17.65
<i>Pausanias</i>	13.39	10.64	19.74	17.05	15.68	18.06	23.12	∞	6.77	17.79	13.39	7.95	2.14	15.38
<i>Plato</i>	6.71	-1.22	3.16	0.02	4.17	-0.70	2.65	1.23	∞	1.01	1.09	-3.10	-7.06	4.50
<i>Plutarch</i>	12.84	15.26	13.98	25.08	11.74	12.24	15.84	17.03	6.83	∞	8.48	10.91	3.83	16.59
<i>Polybius</i>	11.67	5.82	16.21	9.20	16.51	13.00	15.03	13.93	6.59	9.81	∞	3.82	-1.01	11.16
<i>Strabo</i>	7.69	19.97	7.47	13.26	6.39	7.54	8.28	8.95	4.99	12.58	4.28	∞	11.32	9.47
<i>Thucydides</i>	4.89	10.69	4.45	7.94	3.66	4.57	5.01	5.44	3.47	7.62	1.86	13.35	∞	6.06
<i>Xenophon</i>	19.85	9.19	15.48	12.39	13.95	8.95	16.31	13.73	8.78	15.07	8.24	6.01	0.48	∞

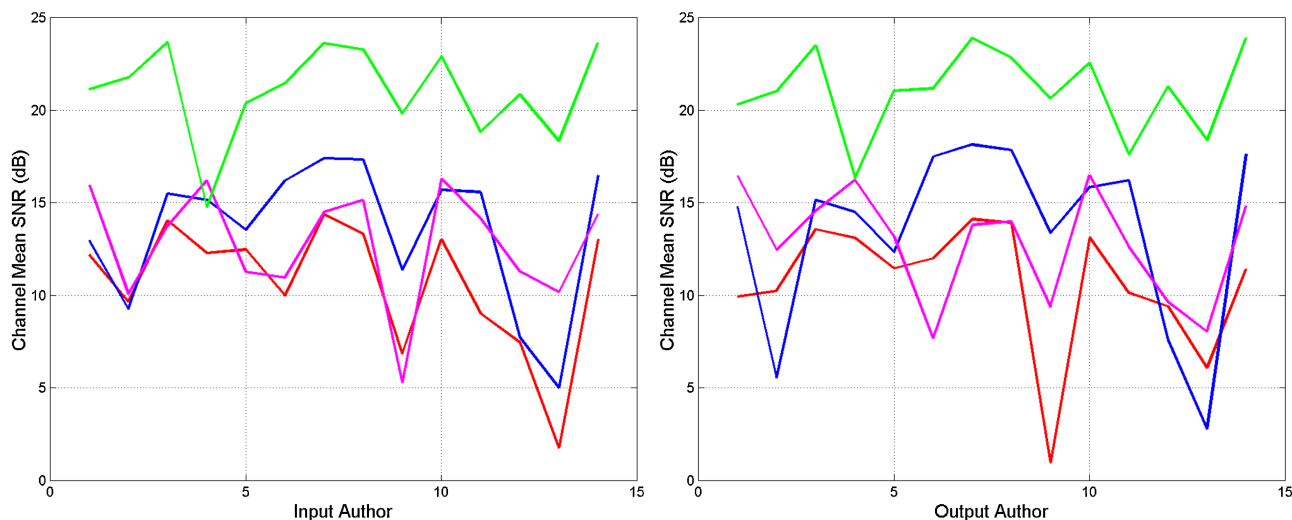


Figure 2. Greek-1. Average Γ calculated by considering the input author (left panel, average of the values reported in the corresponding column of **Table 5**) or the output author (right panel, average of the values reported in the corresponding row). S-Channel: red line; I-channel: blue line; WI-channel: magenta line; I-Channel: green line. *Aeneas Tactician* 1; *Aeschines* 2; *Aristides* 3; *Aristotle* 4; *Demosthenes* 5; *Flavius Josephus* 6; *Herodotus* 7; *Pausanias* 8; *Plato* 9; *Plutarch* 10; *Polybius* 11; *Strabo* 12; *Thucydides* 13; *Xenophon* 14.

(right panel, arithmetic average of the values reported in the corresponding row). The asymmetry typical of linguistic channels is clearly evident.

For example, *Aristides* (no. 3) has large Γ both when he is the input author (left panel) and when he is the output author (right panel). The authors who are very uncorrelated, along with all others, are *Plato* (no. 9) and *Thucydides* (no. 13).

From **Figure 2**, we can conclude that:

1) C-Channels (green line) give large Γ for all authors, in any case. These large values are just saying that all authors use the same language because C_p changes little from author to author. The minimum is found with *Aristotle* (no. 4) which is not a historian or geographer like the other authors. These channels are not very apt to distinguish or assess large differences between texts or authors [11].

2) S-Channels (red line) and WI-Channels (magenta line) are the most similar. This may be due to the fact that both are linked to the E-STM capacity (see Section 8).

3) I-Channels (blue line) give Γ just smaller than that of C-Channels. I-Channels deal with I_p , therefore the word interval used by all authors is not very different (see **Table 3** and Section 8).

6.2. Greek-2

Table 6 shows the results in the S-channel for Greek-2, Appendix C reports Γ for the other three channels. We can notice that the cases of similarity or likely dependence are very few. *Sofocles* may be influenced by *Aeschylus*, and *Pindarus* by the writer of *Odissey* therefore confirming their closeness in **Figure 1**.

Notice that *Iliad* and *Odissey* have significant different Γ in the three channels able to distinguish better authors/texts. They are also distant in **Figure 1**. Now,

modern scholars generally agree that *Homer* composed the *Iliad* most likely relying on oral traditions, and at least inspired the composition of the *Odyssey* but *did not write it* [59].

Figure 3 synthesizes the results of the four channels of Greek-2. We notice that these channels are less correlated that those of Greek-1, therefore texts are significantly different, therefore the literary genre affects these channels (details are reported in Appendix C). C-Channels (green line) give the largest Γ , in the same range of Greek-1 because the authors use the same language.

Table 6. Greek-2. Average Γ , S-Channel. The author/text in the first row is the reference, *i.e.* the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j . Cases with $\Gamma \geq 15$ dB (*i.e.*, $\gamma = 31.6$) are highlighted in colour: blue indicates not only that the number of sentences of the input and output texts are significantly very similar—for the same number of words—but also that the input author might have influenced the output author because he lived before; red indicates a large similarity but no likely influence can be invoked because the input author lived after the output author. Largest Γ : *Aeschylus-Sophocles*, $\Gamma = 24.83$ dB; minimum Γ : *Iliad-Euripides*, $\Gamma = -3.29$ dB.

Author	<i>Aeschylus</i>	<i>Aesop</i>	<i>Euripides</i>	<i>Iliad</i>	<i>Odyssey</i>	<i>Pindarus</i>	<i>Sofocles</i>
<i>Aeschylus</i>	Inf	8.04	9.75	-1.70	0.73	2.48	24.49
<i>Aesop</i>	10.95	Inf	8.77	4.58	7.13	9.31	11.43
<i>Euripides</i>	9.41	4.43	Inf	-3.29	-2.80	-1.61	10.86
<i>Iliad</i>	5.21	8.61	4.79	Inf	12.54	10.71	5.36
<i>Odyssey</i>	6.56	10.52	5.09	10.08	Inf	20.47	6.60
<i>Pindarus</i>	7.55	11.86	5.47	7.39	19.61	Inf	7.54
<i>Sofocles</i>	24.83	8.71	11.56	-1.40	0.66	2.32	Inf

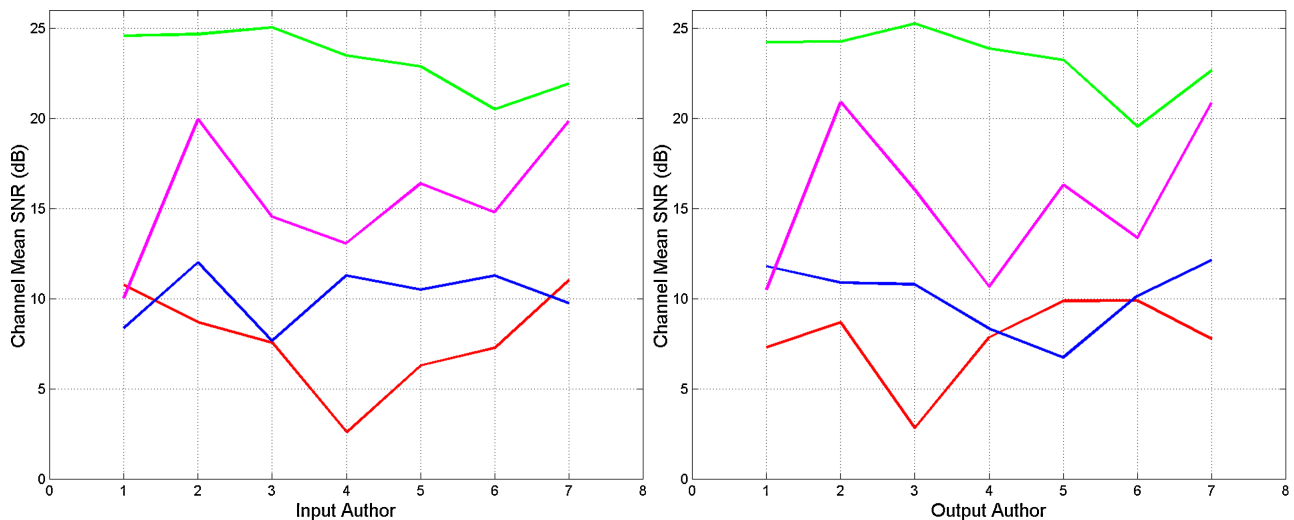


Figure 3. Greek-2. Average Γ calculated by considering the input author (left panel) or the output author (right panel)). S-Channel: red line; I-channel: blue line; WI-channel: magenta line; I-Channel: green line. *Aeschylus* 1; *Aesop* 2; *Euripides* 3; *Iliad* 4; *Odyssey* 5; *Pindarus* 6; *Sophocles* 7.

6.3. Single Scatterplot: Translation of a Text

In this sub-section, we show how to apply the single scatterplot theory (Section

5.3) by considering, as an example, the translation of *Iliad* from Greek to Italian in the classical translation done by Ippolito Pindemonte (1753-1828), an Italian poet. Specifically, we compare the number of words per Book, 24 samples. **Figure 4.** Shows the scatterplot, the linear best fit, with $r_j = 0.9836$ and $m_j = 1.2215$, therefore $\Gamma = 10.03$ dB, a rather poor value due equally to the correlation ($R_r = 0.0502$) and regression ($R_m = 0.0491$) noise.

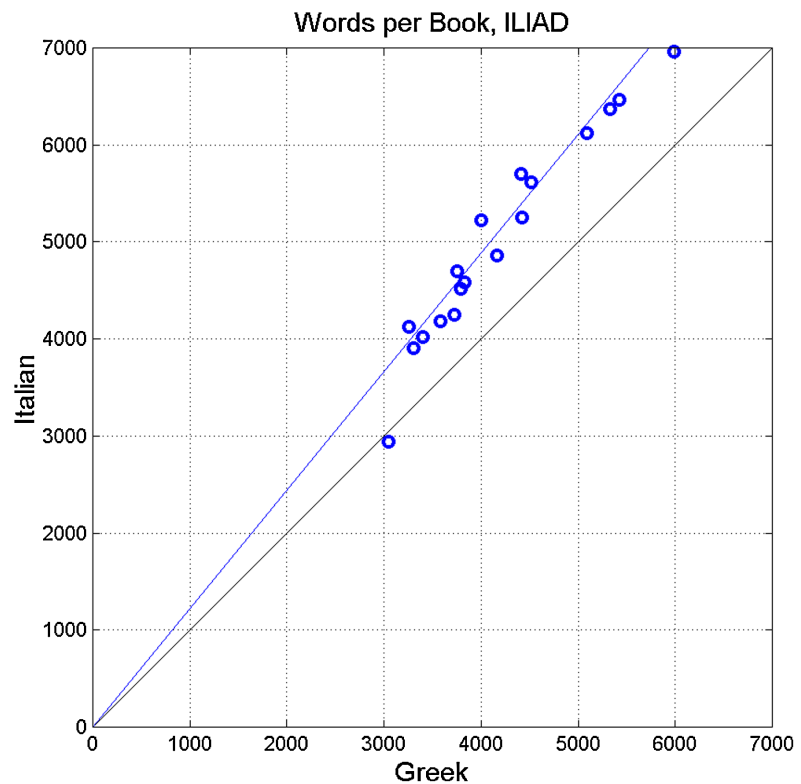


Figure 4. Blue circles and blue line: Number of words per Book in *Iliad* in the Italian translation versus the original Greek text. The 45°—black line represents the noiseless channel.

We can notice that the number of words in Italian is always larger than in Greek, therefore the translator seems to need more words to explain the original meaning, with an average 22.15% per Book. The total number of words in Italian is 136,050 against 111,878 in Greek (21.6% increase).

In the next section, we will estimate the readability of these authors by considering a universal readability index.

7. Universal Readability Index

In Reference [8], we proposed a universal readability index given by:

$$G_U = 89 - 10kC_p + 300/P_F - 6(I_p - 6) \quad (29)$$

$$k = \langle C_{p,ITA} \rangle / \langle C_{p,Lan} \rangle \quad (30)$$

In Equation (30), $\langle C_{p,ITA} \rangle = 4.48$, $\langle C_{p,Lan} \rangle$ is the mean statistical value in the language considered. By using Equations (29) and (30), the mean value $\langle kC_p \rangle$ of

any language is forced to be equal to that found in Italian, namely 4.48. The rationale for this choice is that C_p is a parameter typical of a language which, if not scaled, would bias G_U without really quantifying the reading difficulty of readers, who in their own language are used, on average, to read shorter or longer words than in Italian. This scaling, therefore, avoids changing G_U only because a language has, on the average, words shorter (as English) or longer (as classical Greek) than Italian. In any case, C_p affects Equation (29) much less than P_F or I_p [1]. In this paper, from **Table 3**, $\langle C_{p,Lan} \rangle = 5.29$.

Table 3, **Table 4** report the mean value $\langle G_U \rangle$ of each author/text. Notice that $\langle G_U \rangle$ is always larger (more optimistic) than the value calculated by inserting in Equations (29), (30) the mean values $\langle P_F \rangle$, $\langle I_p \rangle$ (proof in Appendix A of Ref. [11]).

It is interesting to “decode” these mean values into the minimum number of school years, Y , necessary to assess that a text/author passes from being “very difficult” to being only “difficult” to read, according to the modern Italian school system, assumed as a common reference, see **Figure 1** of Ref. [8]. The results are listed in **Table 3**, **Table 4**. Of course, this assumption does not mean that ancient Greek readers attended school for the same number of years of the modern students, but it is only a way to do relative comparisons, otherwise difficult to assess from the mere values of $\langle G_U \rangle$. In other words, we should consider Y as an “equivalent” number of school years.

Figure 4 (left panel) shows $\langle G_U \rangle$ versus Y . An inverse proportionality is clearly evident: The more the readability index decreases, the more school years are required for reading the text “with difficulty”. The author with the greatest readability index (74.9) is *Euripides*, whose readers require only 4 years of school, therefore, “elementary” school; the author with the smallest readability index ($\langle G_U \rangle = 25.2$, due to the large values of both I_p and P_F) is Flavius Josephus, whose readers require about 15 years of school, therefore, “university” level.

The synoptic gospels have very similar readability indices: *Matthew* and *Luke* practically coincide (55.61 and 55.68); *Mark* is very near (56.14). These gospels are more similar to the texts of Greek-2 than to those of Greek-1. *John* is the most readable book ($\langle G_U \rangle = 62.21$), *Acts* is the least readable ($\langle G_U \rangle = 41.35$) and requires more school years (about 10 years) than *John* (6.1 years, about like *Aesop*, 5.6 years, see their vicinity in **Figure 1**). Notice that *Acts* is more similar to the texts of Greek-1 (e.g., *Herodotus*) than to those of Greek-2 (see also [4]).

The readability indices of *Hebrews* and *Apocalypse* are very similar ($\langle G_U \rangle = 47.1$ and $\langle G_U \rangle = 48.95$) and both require about 8 years of school. See [6] for the possibility that both texts were written either by the same author or by two authors of the same early Christian group.

Figure 5 (right panel) shows $\langle G_U \rangle$ versus the distance $d = \sqrt{X^2 + Y^2}$ from the origin (0,0) in the vector plane (**Figure 1**); the “outlier” point is due to *Odyssey*. An inverse proportionality is also clearly evident: The more $\langle G_U \rangle$ decreases the more d increases, therefore, as anticipated in Section 4, the distance from a

reference text/author is a relative measure of readability.

The remarks in Section 4 on the NT books can be reiterated, because *Matthew* and *Luke* are each other superposed ($d = 0.48$), *Mark* is very near ($d = 0.44$). *John* is the nearest gospel to the origin ($d = 0.31$). *Acts*, *Hebrews* and *Apocalypse* are the most distant texts. *Hebrews* and *Apocalypse* are each other close.

Figure 6 shows Y versus $\langle I_P \rangle$ (left panel) and versus $\langle P_F \rangle$ (right panel). In both cases Y increases as $\langle I_P \rangle$ and $\langle P_F \rangle$ increase. The authors/texts that use long word intervals and sentences engage more readers' E-STM and, for this reason, are better matched to readers with longer schooling.

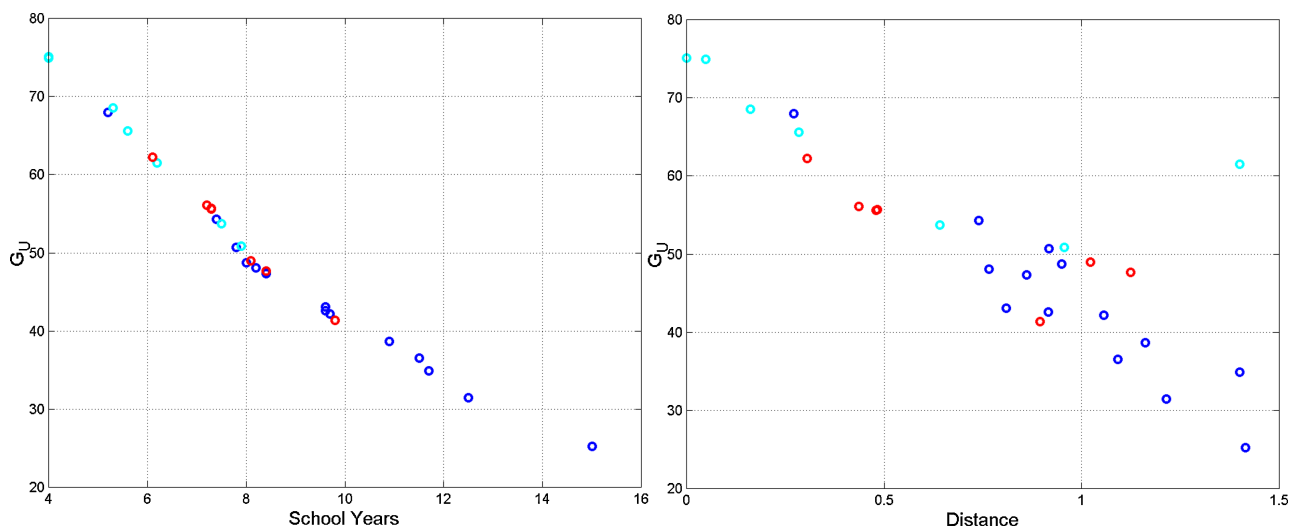


Figure 5. Left panel: $\langle G_U \rangle$ versus Y in passing from “very difficult” to “difficult” to read. Greek-1: blue circles; Greek-2: cyan circles; NT: red circles. Right panel: $\langle G_U \rangle$, versus distance d from the origin (0,0) in the vector plane (Figure 1). Greek-1: blue circles; Greek-2: cyan circles; NT: red circles. The “outlier” text is due to *Odyssey*.

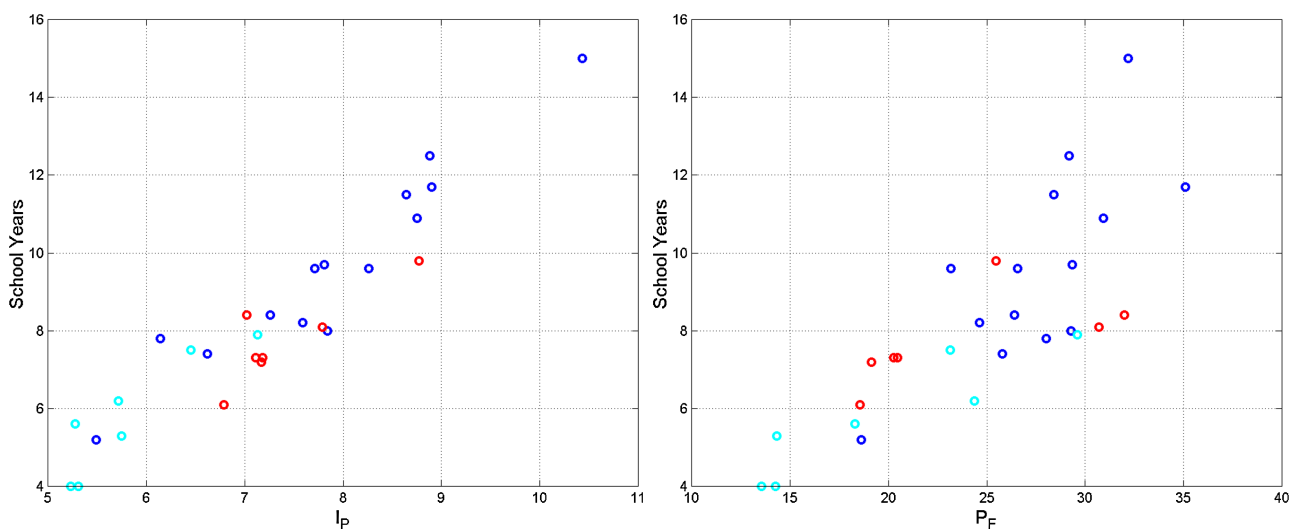


Figure 6. Left panel: Y —passing from “very difficult” to “difficult”—versus $\langle I_P \rangle$, Greek-1: blue circles; Greek-2: cyan circles; NT: red circles. Right panel: Y versus $\langle P_F \rangle$: Greek-1: blue circles; Greek-2: cyan circles; NT: red circles. The largest $Y = 15$ is due to *Flavius Josephus*.

In the next section, we use $\langle P_F \rangle$, $\langle I_p \rangle$ and $\langle M_F \rangle$ to calculate interesting indices connected to the E-STM of readers and writers, as well.

8. Short-Term Memory of Writers/Readers

Recently, we have proposed and applied a well-grounded conjecture that a sentence—read or pronounced, the two activities are similarly processed by the brain [9]-[16]—is elaborated by the E-STM with two independent processing units in series, with similar buffers of similar capacity. The clues for conjecturing this model have emerged by considering a large number of novels belonging to the Italian and English Literatures. We have shown that there are no significant mathematical/statistical differences between the two literary corpora, according to deep-language variables. In other words, the mathematical surface structure of alphabetical languages—a creation of human mind—seems to be deeply rooted in humans, independently of the particular language used. In this section, we show that this is true also for the ancient readers of Greek Literature.

A two-unit E-STM processing is justified according to how a human mind seems to memorize “chunks” of information written in a sentence. Although simple and related to the surface of language, the model seems to describe mathematically the input-output characteristics of a complex mental process largely unknown.

The first processing unit is connected to the number of words between two contiguous interpunctuations, variable indicated by I_p —termed the word interval—approximately ranging in Miller’s 7 ± 2 law range [1] [21]. The second processing unit is connected to the number of word intervals contained in a sentence, M_F , ranging approximately from 1 to 6. The capacity (expressed in words) required to process a sentence ranges from 8.3 to 61.2 words, values that can be converted into time by assuming a reading speed. This conversion gives the range 2.6 - 19.5 seconds for a fast-reading reader [31], and 5.3 - 30.1 seconds for a common reader of novels, values well supported by experiments [22]-[47].

Notice that the two buffers are linked, mathematically, to both syntax and semantics, because the word interval I_p refers to a single chunk of information memorized in the STM, and M_F refers to how many chunks make a full sentence, memorized in the E-STM. In other words, these two variables should be among the elements to be considered at the foundation of a future theory of semantic information mentioned in Section 1.

The E-STM must not be confused with the intermediate memory [60] [61]. It is not modelled by studying neuronal activity, but by studying only surface aspects of human communication, such as words, sentences and interpunctuations, whose effects writers and readers have experienced since the invention of writing. In this section we show that these two independent units are also present in ancient Greek texts.

8.1. E-STM First Buffer (Linked to I_p)

Figure 7 shows $\langle I_p \rangle$ versus $\langle P_F \rangle$ and the non-linear best-fit regression curves

for Greek-1, Greek-2 and NT. As we have already established in modern languages and Latin [1] [2] [4], if $\langle P_F \rangle$ increases $\langle I_p \rangle$ tends to approach a horizontal asymptote. In other words, even if a sentence gets longer, $\langle I_p \rangle$ cannot become larger than about the upper limit of 7 ± 2 Miller's law (namely about 9), because of the constraints imposed by the E-STM capacity of readers and writers.

The coincidence of $\langle I_p \rangle$ with the bounds of Miller's law is clearly evident in **Figure 6**, just like in modern languages as the best-fit curves found in Italian and English novels [9], in modern languages [4]—also drawn in **Figure 6**—clearly show.

From **Figure 6**, we can draw the following conclusions.

- 1) There is a marked distinction between the regression curves concerning Greek-1, Greek-2 and NT.
- 2) The regression curves of Italian and English, which refer only to novels, agree very well with the regression curves of Greek-2 and NT.
- 3) Greek-1 is clearly mathematically different of Greek-2. The difference between novels and other types of writings, such as essays, was clearly found also in Italian writers as well [1].

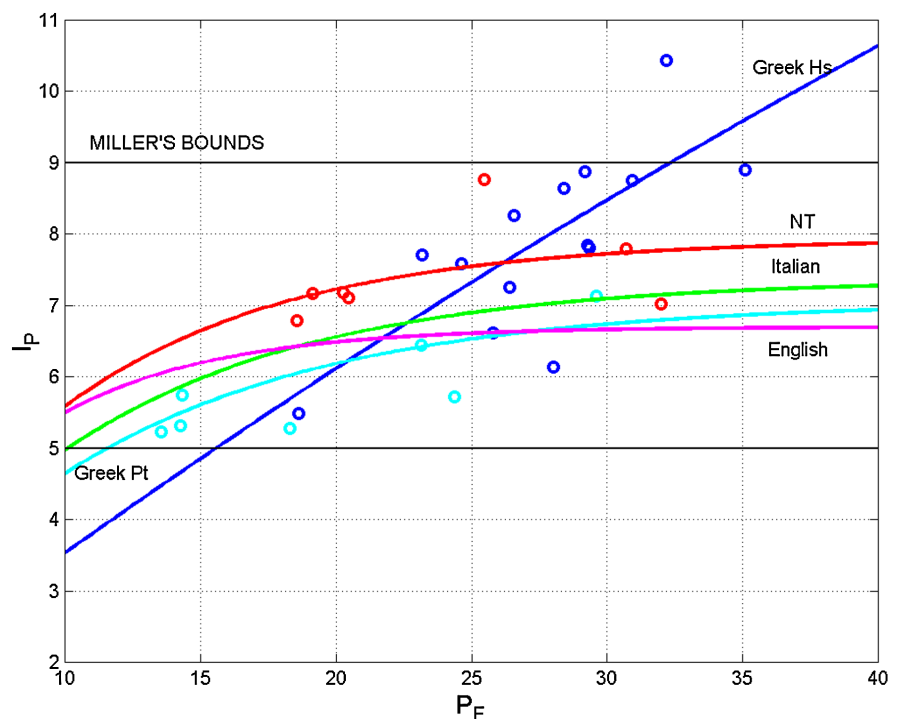


Figure 7. $\langle I_p \rangle$, versus $\langle P_F \rangle$. The continuous lines are non-linear best fit curves. Greek-1 texts: blue circles and blue line; Greek-2: cyan circles and cyan line; NT: red circles and red line; Italian Literature best fit: green line. English Literature best fit: magenta line [9].

8.2. E-STM Second Buffer (Linked to M_F)

Figure 8 shows the scatterplot between M_F and I_p for the samples of the entire data bank used to calculate the statistical means of **Figure 6**. The horizontal

green line reports the unconditional statistical mean $\langle M_F \rangle$, the black line reports the conditional mean versus I_p .

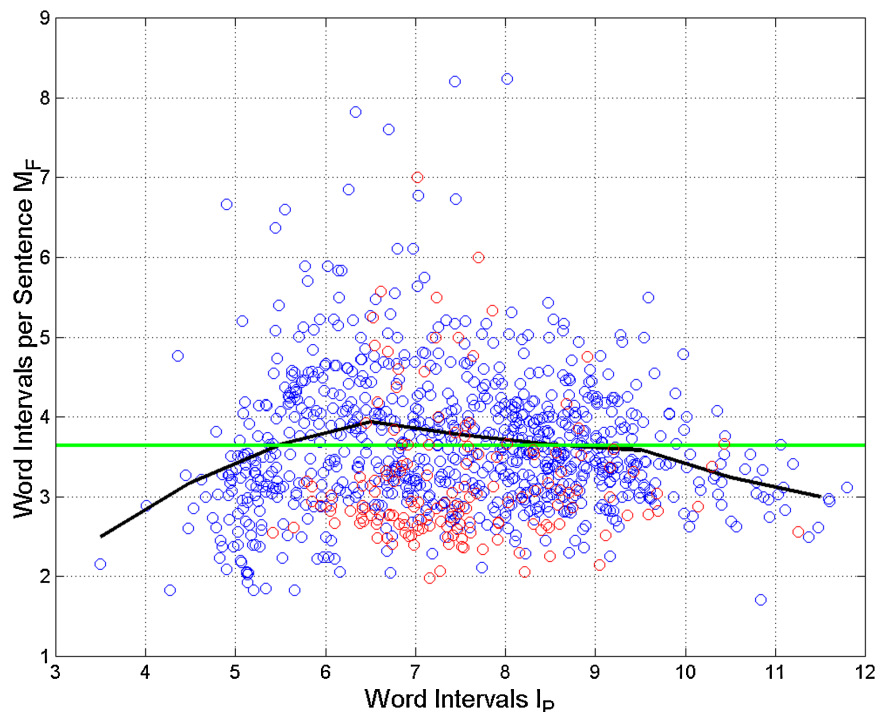


Figure 8. Scatterplot between M_F and I_p in the Greek Literature (Greek-1 plus Greek-2, blue circles)—this is the entire data samples used to calculate statistical means of [Table 3](#), [Table 4](#)—and in NT (red circles). The green horizontal line reports the statistical mean $\langle M_F \rangle$; the black line reports the conditional mean of M_F versus I_p , in 1-unit steps of I_p .

Now, the correlation coefficient between I_p and M_F in [Figure 8](#) is practically zero (namely 0.03). The probability density of I_p samples ([Figure 9](#), left panel) and M_F samples ([Figure 9](#), right panel) can be modelled with a three-parameter log-normal density function—because $I_p \geq 1$, $M_F \geq 1$ —as in Italian and English [\[9\]](#). Since a bivariate log-normal density function can be a sufficiently good model for the joint density of $\log(I_p)$ and $\log(M_F)$, at least in the central part of the marginal distributions, it follows that, if the correlation coefficient is zero, $\log(I_p)$ and $\log(M_F)$ are not only uncorrelated but also independent in the Gaussian case. Therefore, I_p and M_F are also independent and the two processing units of the E-STM work sufficiently independently, as with modern readers.

The size of the second E-STM buffer is in the same range of modern languages, as the bulk of the data in [Figure 8](#) is the range from $M_F \approx 2$ to $M_F \approx 6$ word intervals per sentence.

In conclusion, these texts were conjecturally processed by a two-unit E-STM very similar to the E-STM of modern readers, even though these ancient readers were more accustomed to memorize oral information than modern ones. The specific size of the two buffers required in reading a text depended only on the genre, as for modern readers.

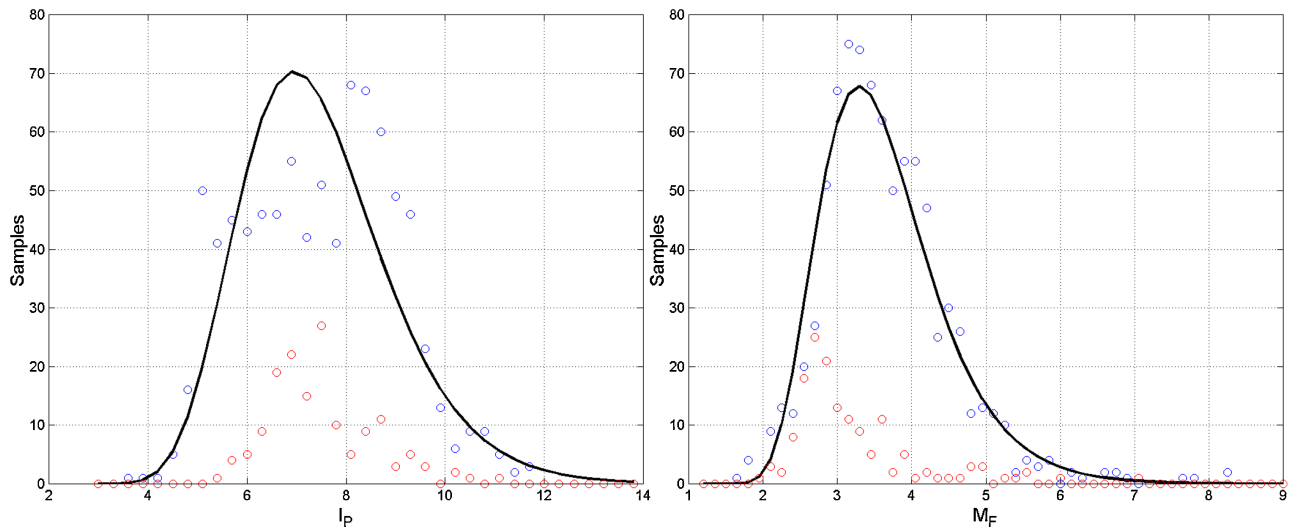


Figure 9. Probability density of I_p (Left panel) and M_F (Right panel). Greek Literature (Greek-1 plus Greek-2): blue circles; NT books: red circles. The continuous black curves model the Greek Literature samples with a three-parameter log-normal density function.

9. Multiplicity Factor and Mismatch Index

In [10], we studied the number of sentences that can be theoretically recorded in the E-STM, and compared them with those of Italian and English novels. We found that most authors write for readers with short E-STM buffers and, consequently, are forced to reuse sentence patterns to convey multiple meanings. This behavior is measured by the multiplicity factor α , defined as the ratio between the number of sentences in a text and the number of sentences theoretically allowed by the E-STM.

We found that $\alpha > 1$ is more likely than $\alpha < 1$ and often $\alpha \gg 1$. In the latter case, writers reuse many times the same pattern of number of words in sentences. Few novels show $\alpha < 1$; in this case, writers do not use some or most of them.

Another useful index is the mismatch index, I_M , in the range ± 1 , which measures to what extent a writer uses the number of sentences theoretically available, defined by:

$$I_M = \frac{\alpha - 1}{\alpha + 1} \quad (31)$$

If $\alpha = 1$ then $I_M = 0$, therefore the number of sentences in a text equals the number of sentences theoretically allowed by the STM, a perfect match. If $\alpha > 1$ then $I_M > 0$ therefore the number of sentences in a text is greater than that theoretically allowed (overmatching, the authors repeats patterns); if $\alpha < 1$ then $I_M < 0$, the number of sentences in a text is smaller than that theoretically allowed (undermatching, the authors use fewer patterns than those available).

Table 3, Table 4 report α and I_M for each author. From these results, we find that the authors who show practically perfect match are only *Aristides*, *Aristote*, *Plutarch* and *Polybius*. No book of the NT shows a perfect match.

Figure 10 shows α versus $\langle I_p \rangle$ (left panel) and versus $\langle M_F \rangle$ (right panel).

We can see that $\log(\alpha)$ and $\langle I_P \rangle$ (first STM buffer) are substantially uncorrelated, while $\log(\alpha)$ and $\langle M_F \rangle$ (second E-STM buffer) are significantly correlated. This latter finds mean that the number of sentence patterns is due only to the second E-STM buffer. The finds concerning Italian and English Literatures [10] are scattered just like in Greek, therefore underlining no significant changes in more than 2000 years.

Figure 11 shows α versus $\langle P_F \rangle$ (left panel) and the mismatch index $\langle I_M \rangle$ (right panel). We can see that $\log(\alpha)$ and $\langle P_F \rangle$ are correlated because large values of $\langle P_F \rangle$ can contain many word intervals I_P , therefore large values of

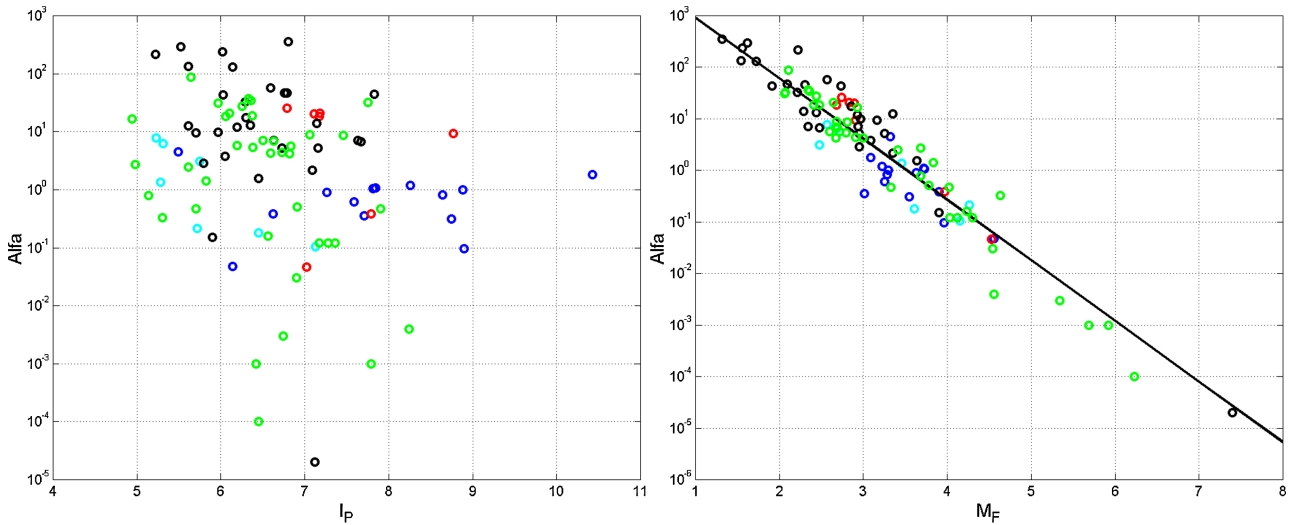


Figure 10. Left panel: α versus $\langle I_P \rangle$ (first E-STM buffer). Greek-1: blue circles; Greek-2: cyan circles; NT books: red circles; Italian: green circles; English: black circles. Right panel: scatterplot of α versus $\langle M_F \rangle$ (E-STM, second buffer). Greek-1: blue circles; Greek-2: cyan circles; NT books: red circles; Italian: green circles; English: black circles.

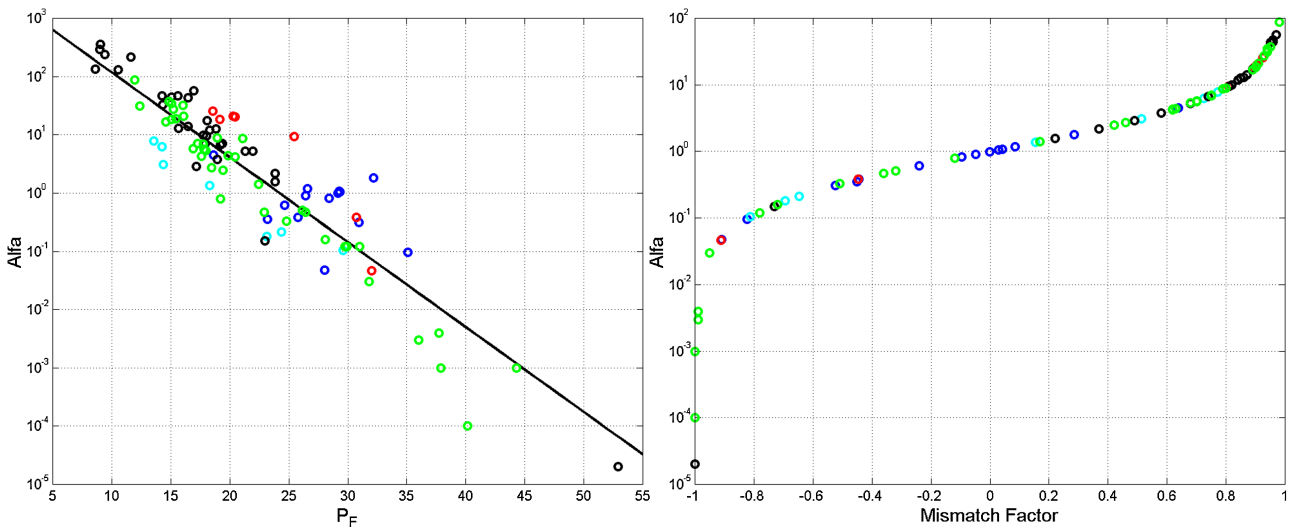


Figure 11. Left panel: α versus $\langle P_F \rangle$. Greek-1: blue circles; Greek-2: cyan circles; NT books: red circles; Italian: green circles; English: black circles. Right panel: scatterplot of α versus the mismatch index I_M . Greek-1: blue circles; Greek-2: cyan circles; NT books: red circles; Italian: green circles; English: black circles.

$\langle M_F \rangle$. The mismatch index follows, of course, Eq. (24) and clearly indicates where texts/authors are located, including Italian and English ones.

10. Conclusions

After the discussion of the findings reported in each section, we can conclude that the multi-dimensional mathematical theory applied to texts of the classical Greek Literature—spanning eight centuries—reveals likely connections between authors/texts far beyond writers' awareness—just like it does in modern literatures—and with the extended short-term memory of ancient readers.

The analysis, based on 3,225,839 words contained in 118,952 sentences, has shown that ancient Greek writers, and their readers, were not significantly different from modern writers/readers. Their sentences were processed by extended short-term memory and modelled with two independent processing units in series, just like in modern readers. This finding is very interesting because, in a society in which people are used to memorize information more than modern people do, authors write almost exactly, mathematically speaking, as modern writers and for readers of similar characteristics. Since meaning is not considered, any text of any alphabetical language can be studied exactly with the same mathematical/statistical tools and, therefore, comparisons can be made, regardless of different languages and epochs of writing.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] 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. <https://doi.org/10.4236/ojs.2019.93026>
- [2] Matricciani, E. (2020) A Statistical Theory of Language Translation Based on Communication Theory. *Open Journal of Statistics*, **10**, 936-997. <https://doi.org/10.4236/ojs.2020.106055>
- [3] Matricciani, E. (2022) Multiple Communication Channels in Literary Texts. *Open Journal of Statistics*, **12**, 486-520. <https://doi.org/10.4236/ojs.2022.124030>
- [4] Matricciani, E. (2022) Linguistic Mathematical Relationships Saved or Lost in Translating Texts: Extension of the Statistical Theory of Translation and Its Application to the New Testament. *Information*, **13**, Article No. 20. <https://doi.org/10.3390/info13010020>
- [5] Matricciani, E. (2023) Capacity of Linguistic Communication Channels in Literary Texts: Application to Charles Dickens' Novels. *Information*, **14**, Article No. 68. <https://doi.org/10.3390/info14020068>
- [6] Matricciani, E. (2023) Linguistic Communication Channels Reveal Connections between Texts: The New Testament and Greek Literature. *Information*, **14**, Article No. 405. <https://doi.org/10.3390/info14070405>
- [7] Matricciani, E. (2023) Readability across Time and Languages: The Case of Matthew's

- Gospel Translations. *AppliedMath*, **3**, 497-509.
<https://doi.org/10.3390/appliedmath3020026>
- [8] Matricciani, E. (2023) Readability Indices Do Not Say It All on a Text Readability. *Analytics*, **2**, 296-314. <https://doi.org/10.3390/analytics2020016>
- [9] Matricciani, E. (2023) Is Short-Term Memory Made of Two Processing Units? Clues from Italian and English Literatures down Several Centuries. *Information*, **15**, Article No. 6. <https://doi.org/10.3390/info15010006>
- [10] Matricciani, E. (2024) A Mathematical Structure Underlying Sentences and Its Connection with Short-Term Memory. *AppliedMath*, **4**, 120-142.
<https://doi.org/10.3390/appliedmath4010007>
- [11] Matricciani, E. (2024) Multi-Dimensional Data Analysis of Deep Language in J.R.R. Tolkien and C.S. Lewis Reveals Tight Mathematical Connections. *AppliedMath*, **4**, 927-949. <https://doi.org/10.3390/appliedmath4030050>
- [12] Calvanese Strinati, E. and Barbarossa, S. (2021) 6G Networks: Beyond Shannon towards Semantic and Goal-Oriented Communications. *Computer Networks*, **190**, Article ID: 107930. <https://doi.org/10.1016/j.comnet.2021.107930>
- [13] Shi, G., Xiao, Y., Li, Y. and Xie, X. (2021) From Semantic Communication to Semantic-Aware Networking: Model, Architecture, and Open Problems. *IEEE Communications Magazine*, **59**, 44-50. <https://doi.org/10.1109/mcom.001.2001239>
- [14] Xie, H., Qin, Z., Li, G.Y. and Juang, B. (2021) Deep Learning Enabled Semantic Communication Systems. *IEEE Transactions on Signal Processing*, **69**, 2663-2675.
<https://doi.org/10.1109/tsp.2021.3071210>
- [15] Luo, X., Chen, H. and Guo, Q. (2022) Semantic Communications: Overview, Open Issues, and Future Research Directions. *IEEE Wireless Communications*, **29**, 210-219. <https://doi.org/10.1109/mwc.101.2100269>
- [16] Yang, W., Du, H., Liew, Z.Q., Lim, W.Y.B., Xiong, Z., Niyato, D., et al. (2023) Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges. *IEEE Communications Surveys & Tutorials*, **25**, 213-250.
<https://doi.org/10.1109/comst.2022.3223224>
- [17] Bellegarda, J.R. (2000) Exploiting Latent Semantic Information in Statistical Language Modeling. *Proceedings of the IEEE*, **88**, 1279-1296.
<https://doi.org/10.1109/5.880084>
- [18] D'Alfonso, S. (2011) On Quantifying Semantic Information. *Information*, **2**, 61-101.
<https://doi.org/10.3390/info2010061>
- [19] Zhong, Y. (2017) A Theory of Semantic Information. *China Communications*, **14**, 1-17. <https://doi.org/10.1109/cc.2017.7839754>
- [20] Deniz, F., Nunez-Elizalde, A.O., Huth, A.G. and Gallant, J.L. (2019) The Representation of Semantic Information across Human Cerebral Cortex during Listening versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience*, **39**, 7722-7736. <https://doi.org/10.1523/jneurosci.0675-19.2019>
- [21] Miller, G.A. (1994) The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, **101**, 343-352.
<https://doi.org/10.1037//0033-295x.101.2.343>
- [22] Crowder, R.G. (1993) Short-Term Memory: Where Do We Stand? *Memory & Cognition*, **21**, 142-145. <https://doi.org/10.3758/bf03202725>
- [23] Lisman, J.E. and Idiart, M.A.P. (1995) Storage of 7 ± 2 Short-Term Memories in Oscillatory Subcycles. *Science*, **267**, 1512-1515. <https://doi.org/10.1126/science.7878473>

- [24] Cowan, N. (2001) The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences*, **24**, 87-114. <https://doi.org/10.1017/s0140525x01003922>
- [25] Bacher, B.L. (2001) The Magical Number 4 = 7: Span Theory on Capacity Limitations. *Behavioral and Brain Sciences*, **24**, 116-117. <https://doi.org/10.1017/s0140525x01243921>
- [26] Saaty, T.L. and Ozdemir, M.S. (2003) Why the Magic Number Seven Plus or Minus Two. *Mathematical and Computer Modelling*, **38**, 233-244. [https://doi.org/10.1016/s0895-7177\(03\)90083-5](https://doi.org/10.1016/s0895-7177(03)90083-5)
- [27] Burgess, N. and Hitch, G.J. (2006) A Revised Model of Short-Term Memory and Long-Term Learning of Verbal Sequences. *Journal of Memory and Language*, **55**, 627-652. <https://doi.org/10.1016/j.jml.2006.08.005>
- [28] Richardson, J.T.E. (2007) Measures of Short-Term Memory: A Historical Review. *Cortex*, **43**, 635-650. [https://doi.org/10.1016/s0010-9452\(08\)70493-3](https://doi.org/10.1016/s0010-9452(08)70493-3)
- [29] Mathy, F. and Feldman, J. (2012) What's Magic about Magic Numbers? Chunking and Data Compression in Short-Term Memory. *Cognition*, **122**, 346-362. <https://doi.org/10.1016/j.cognition.2011.11.003>
- [30] Gignac, G.E. (2015) The Magical Numbers 7 and 4 Are Resistant to the Flynn Effect: No Evidence for Increases in Forward or Backward Recall across 85 Years of Data. *Intelligence*, **48**, 85-95. <https://doi.org/10.1016/j.intell.2014.11.001>
- [31] Trauzettel-Klosinski, S. and Dietz, K. (2012) Standardized Assessment of Reading Performance: The New International Reading Speed Texts Irest. *Investigative Ophthalmology & Visual Science*, **53**, 5452-5461. <https://doi.org/10.1167/iovs.11-8284>
- [32] Melton, A.W. (1963) Implications of Short-Term Memory for a General Theory of Memory. *Journal of Verbal Learning and Verbal Behavior*, **2**, 1-21. [https://doi.org/10.1016/s0022-5371\(63\)80063-8](https://doi.org/10.1016/s0022-5371(63)80063-8)
- [33] Atkinson, R.C. and Shiffrin, R.M. (1971) The Control of Short-Term Memory. *Scientific American*, **225**, 82-90. <https://doi.org/10.1038/scientificamerican0871-82>
- [34] Murdock, B.B. (1972) Short-Term Memory. In: *Psychology of Learning and Motivation*, Elsevier, 67-127. [https://doi.org/10.1016/s0079-7421\(08\)60440-5](https://doi.org/10.1016/s0079-7421(08)60440-5)
- [35] Baddeley, A.D., Thomson, N. and Buchanan, M. (1975) Word Length and the Structure of Short-Term Memory. *Journal of Verbal Learning and Verbal Behavior*, **14**, 575-589. [https://doi.org/10.1016/s0022-5371\(75\)80045-4](https://doi.org/10.1016/s0022-5371(75)80045-4)
- [36] Case, R., Kurland, D.M. and Goldberg, J. (1982) Operational Efficiency and the Growth of Short-Term Memory Span. *Journal of Experimental Child Psychology*, **33**, 386-404. [https://doi.org/10.1016/0022-0965\(82\)90054-6](https://doi.org/10.1016/0022-0965(82)90054-6)
- [37] Grondin, S. (2001) A Temporal Account of the Limited Processing Capacity. *Behavioral and Brain Sciences*, **24**, 122-123. <https://doi.org/10.1017/s0140525x01303928>
- [38] Pothos, E.M. and Juola, P. (2001) Linguistic Structure and Short-Term Memory. *Behavioral and Brain Sciences*, **24**, 138-139. <https://doi.org/10.1017/s0140525x01463928>
- [39] Conway, A.R.A., Cowan, N., Bunting, M.F., Theriault, D.J. and Minkoff, S.R.B. (2002) A Latent Variable Analysis of Working Memory Capacity, Short-Term Memory Capacity, Processing Speed, and General Fluid Intelligence. *Intelligence*, **30**, 163-183. [https://doi.org/10.1016/s0160-2896\(01\)00096-4](https://doi.org/10.1016/s0160-2896(01)00096-4)
- [40] Jonides, J., Lewis, R.L., Nee, D.E., Lustig, C.A., Berman, M.G. and Moore, K.S. (2008) The Mind and Brain of Short-Term Memory. *Annual Review of Psychology*, **59**, 193-224. <https://doi.org/10.1146/annurev.psych.59.103006.093615>

- [41] Barrouillet, P. and Camos, V. (2012) As Time Goes by: Temporal Constraints in Working Memory. *Current Directions in Psychological Science*, **21**, 413-419. <https://doi.org/10.1177/0963721412459513>
- [42] Potter, M.C. (2012) Conceptual Short Term Memory in Perception and Thought. *Frontiers in Psychology*, **3**, Article No. 113. <https://doi.org/10.3389/fpsyg.2012.00113>
- [43] Jones, G. and Macken, B. (2015) Questioning Short-Term Memory and Its Measurement: Why Digit Span Measures Long-Term Associative Learning. *Cognition*, **144**, 1-13. <https://doi.org/10.1016/j.cognition.2015.07.009>
- [44] Chekaf, M., Cowan, N. and Mathy, F. (2016) Chunk Formation in Immediate Memory and How It Relates to Data Compression. *Cognition*, **155**, 96-107. <https://doi.org/10.1016/j.cognition.2016.05.024>
- [45] Norris, D. (2017) Short-Term Memory and Long-Term Memory Are Still Different. *Psychological Bulletin*, **143**, 992-1009. <https://doi.org/10.1037/bul0000108>
- [46] Van Houdt, G., Mosquera, C. and Nápoles, G. (2020) A Review on the Long Short-Term Memory Model. *Artificial Intelligence Review*, **53**, 5929-5955. <https://doi.org/10.1007/s10462-020-09838-1>
- [47] Islam, M.A., Sarkar, A.K., Hossain, M.I., Ahmed, M.T. and Ferdous, A.H.M.I. (2023) Prediction of Attention and Short-Term Memory Loss by EEG Workload Estimation. *Journal of Biosciences and Medicines*, **11**, 304-318. <https://doi.org/10.4236/jbm.2023.114022>
- [48] Parkes, M.B. (2016) Pause and Effect: An Introduction to the History of Punctuation in the West. Routledge. <https://doi.org/10.4324/9781315247243>
- [49] Matricciani, E. and Caro, L.D. (2019) A Deep-Language Mathematical Analysis of Gospels, Acts and Revelation. *Religions*, **10**, Article No. 257. <https://doi.org/10.3390/rel10040257>
- [50] Battezzato, L. (2009) Techniques of Reading and Textual Layout in Ancient Greek Texts. *The Cambridge Classical Journal*, **55**, 1-23. <https://doi.org/10.1017/s1750270500000166>
- [51] Murphy, J.J. and Thaïss, C. (2020) A Short History of Writing Instruction: From Ancient Greece to the Modern United States. 4th Edition, Routledge. <https://doi.org/10.4324/9781003020899>
- [52] Spelman, H. (2019) Schools, Reading and Poetry in the Early Greek World. *The Cambridge Classical Journal*, **65**, 150-172. <https://doi.org/10.1017/s1750270519000046>
- [53] Fudin, R. (1989) Reading Efficiency and the Development of Left-to-Right Writing by the Ancient Greeks. *Perceptual and Motor Skills*, **69**, 1251-1258. <https://doi.org/10.2466/pms.1989.69.3f.1251>
- [54] Johnson, W.A. (2000) Toward a Sociology of Reading in Classical Antiquity. *American Journal of Philology*, **121**, 593-627. <https://doi.org/10.1353/ajp.2000.0053>
- [55] Johnson, W.A. (2010). Readers and Reading Culture in the High Roman Empire: A Study of Elite Communities. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195176407.001.0001>
- [56] Krauß, A., Leipziger, J. and Schücking-Jungblut, F. (2020) Material Aspects of Reading in Ancient and Medieval Cultures. Walter de Gruyter GmbH.
- [57] Vatri, A. (2012) The Physiology of Ancient Greek Reading. *The Classical Quarterly*, **62**, 633-647. <https://doi.org/10.1017/s0009838812000213>
- [58] Abramovitz, M. and Stegun, I.A. (1985) Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables. 9th Edition, Dover Publications.

- [59] Kirk, G.S. (2024) Homer. Encyclopedia Britannica.
<https://www.britannica.com/biography/Homer-Greek-poet>
- [60] Rosenzweig, M.R., Bennett, E.L., Colombo, P.J., Lee, D.W. and Serrano, P.A. (1993) Short-Term, Intermediate-Term, and Long-Term Memories. *Behavioural Brain Research*, **57**, 193-198. [https://doi.org/10.1016/0166-4328\(93\)90135-d](https://doi.org/10.1016/0166-4328(93)90135-d)
- [61] Kamiński, J. (2017) Intermediate-Term Memory as a Bridge between Working and Long-Term Memory. *The Journal of Neuroscience*, **37**, 5045-5047.
<https://doi.org/10.1523/jneurosci.0604-17.2017>

Appendix A. List of Mathematical Symbols and Meaning

Symbol	Definition
C_p	Characters per word
G_U	Universal readability index
I_M	Mismatch index
I_p	Word interval
M_F	Word intervals per sentence
P_F	Words per sentence
R	Noise-to-signal ratio
R_m	Regression noise-to-signal ratio
R_r	Correlation noise-to-signal ratio
S	Total number of sentences
W	Total number of words
n_C	Number of characters
n_W	Number of words
n_S	Number of sentences
n_I	Number of interpunctuations
n_{I_p}	Number of word intervals
γ	Signal-to-noise ratio
Γ	Signal-to-noise ratio (dB)
m_{jk}	Slope of regression line of text j versus text k
r_{jk}	Correlation coefficient between text j and text k

Appendix B. Linguistic Channels in Greek-1 Texts

Table A1. Greek-1. Correlation and slope of the regression lines between the indicated variables. Four digits are reported because some authors/texts differ only at the third/fourth digit.

Author	S-Channel		I-Channel		WI-Channel		C-Channel	
	Sentences vs words		Word Intervals vs Sentences		Words vs Interpunctuations		Characters vs Words	
	Correlation	Slope	Correlation	Slope	Correlation	Slope	Correlation	Slope
<i>Aeneas the Tactician</i>	0.9748	0.0448	0.9921	2.9668	0.9856	7.3856	0.9998	5.7334
<i>Aeschines</i>	0.8419	0.0363	0.8647	4.3939	0.9872	6.1227	0.9971	5.7281
<i>Aristides</i>	0.9795	0.0383	0.9858	3.5166	0.9980	7.2665	0.9989	5.4416
<i>Aristoteles</i>	0.9143	0.0352	0.9671	3.6124	0.9825	7.7066	0.9899	4.6854
<i>Demosthenes</i>	0.9899	0.0398	0.9874	3.7903	0.9988	6.5806	0.9999	5.0328
<i>Flavius Josephus</i>	0.9657	0.0315	0.9684	3.0637	0.9659	10.2912	0.9936	5.4927

Continued

<i>Herodotus</i>	0.9708	0.0377	0.9723	3.2090	0.9978	8.2387	0.9987	5.1998
<i>Pausanias</i>	0.9615	0.0354	0.9774	3.2647	0.9914	8.5999	0.9978	5.5776
<i>Plato</i>	0.9925	0.0644	0.9972	2.9594	0.9982	5.1887	0.9998	4.9659
<i>Plutarch</i>	0.9195	0.0371	0.9577	3.3539	0.9898	7.6165	0.9996	5.5026
<i>Polybius</i>	0.9971	0.0343	0.9885	3.2432	0.9949	8.9118	0.9997	5.9880
<i>Strabo</i>	0.8045	0.0334	0.8139	3.3826	0.9138	8.5624	0.9942	5.1707
<i>Thucydides</i>	0.6754	0.0290	0.6794	3.8304	0.8894	8.8060	0.9863	5.3551
<i>Xenophon</i>	0.9501	0.0425	0.9660	3.0978	0.9712	7.4113	0.9984	5.1957

Table A2. Greek-1. Average Γ , I-Channel. The author/text in the first row is the reference, *i.e.* the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j . For example, if *Aristides* is the input and *Demosthenes* is the output, then $\Gamma = 22.10$ dB, viceversa, if *Demosthenes* is the input and *Aristides* is the output, then $\Gamma = 22.76$. Cases with $\Gamma \geq 15$ dB are high lighted in colour: blue indicates not only that the number of sentences of the input and output texts are significantly very similar—for the same number of words—but also that the input author might have influenced the output author because he lived before; red indicates that the number of sentences of the input and output authors are very similar—for the same number of words, as in the blue cases—but no likely influence can be invocated because the input author lived after the output author. Largest Γ : *Flavius-Xenophon*, 36.67 dB; minimum Γ : *Plato-Thucydides*, -1.86 dB.

Author	<i>Aeneas</i>	<i>Aeschi</i>	<i>Aristi</i>	<i>Aristo</i>	<i>Demo</i>	<i>Flavius</i>	<i>Hero</i>	<i>Pausa</i>	<i>Plato</i>	<i>Plut</i>	<i>Poly</i>	<i>Strabo</i>	<i>Thuc</i>	<i>Xen</i>
<i>Aeneas</i>	∞	7.28	15.89	13.59	13.20	17.93	17.92	18.34	25.82	14.52	21.06	6.23	3.25	17.24
<i>Aeschines</i>	2.04	∞	5.53	7.98	6.49	4.54	5.18	5.09	1.23	7.12	3.88	9.82	8.37	4.91
<i>Aristides</i>	14.33	8.89	∞	20.88	22.76	15.08	18.35	20.85	13.19	17.17	21.28	5.93	2.97	15.31
<i>Aristoteles</i>	11.35	10.81	20.43	∞	19.57	14.93	17.86	18.62	10.03	21.35	15.71	7.72	4.40	15.59
<i>Demosthenes</i>	11.03	8.89	22.10	18.82	∞	11.57	13.86	15.25	10.43	14.00	15.45	4.89	2.20	11.81
<i>Flavius</i>	17.39	8.86	16.60	16.37	13.72	∞	26.41	22.89	14.48	20.55	19.17	8.84	5.16	36.81
<i>Herodotus</i>	16.78	9.20	19.42	18.92	15.56	25.96	∞	30.98	14.18	23.24	21.51	8.25	4.71	27.01
<i>Pausanias</i>	17.13	9.07	21.66	19.64	16.69	22.19	30.73	∞	14.66	21.79	24.15	7.59	4.23	22.58
<i>Plato</i>	25.86	6.71	15.03	12.56	12.81	15.07	15.45	16.09	∞	12.87	18.99	5.27	2.46	14.62
<i>Plutarch</i>	12.76	10.49	17.94	22.11	15.64	19.64	22.62	21.34	10.96	∞	16.49	9.43	5.51	21.01
<i>Polybius</i>	20.23	8.16	22.01	17.11	16.80	18.31	21.32	24.26	17.87	17.06	∞	6.26	3.21	18.10
<i>Strabo</i>	4.01	12.35	6.60	8.85	6.82	7.17	7.34	6.97	3.00	9.28	5.53	∞	13.28	7.55
<i>Thucydides</i>	-1.00	10.56	1.50	3.38	2.02	1.49	1.74	1.52	-1.86	3.26	0.38	11.41	∞	1.78
<i>Xenophon</i>	16.52	9.08	16.80	16.92	13.93	36.67	27.42	23.25	13.84	21.79	18.83	9.04	5.28	∞

Table A3. Greek-1. Average Γ , WI-Channel. The author/text in the first row is the reference, *i.e.* the channel input author/text Y_k ; the author/text in the first column is the dependent (output) author/text Y_j . For example, if *Aristides* is the input and *Demosthenes* is the output, then $\Gamma = 20.42$ dB, viceversa, if *Demosthenes* is the input and *Aristides* is the output, then $\Gamma = 19.54$. Cases with $\Gamma \geq 15$ dB are high lighted in colour: blue indicates not only that the number of sentences of the input and output texts are significantly very similar—for the same number of words—but also that the input author might have influenced the output author because he lived before; red indicates that the number of sentences of the input and output authors are very similar—for the same number of words, as in the blue cases—but no likely influence can be invocated because the input author lived after the output author. Largest Γ : *Plutarch-Aeneas*, 27.95 dB; minimum Γ : *Plato-Thucydides*, -0.19 dB.

Author	<i>Aeneas</i>	<i>Aeschi</i>	<i>Aristi</i>	<i>Aristo</i>	<i>Demo</i>	<i>Flavius</i>	<i>Hero</i>	<i>Pausa</i>	<i>Plato</i>	<i>Plut</i>	<i>Poly</i>	<i>Strabo</i>	<i>Thuc</i>	<i>Xen</i>
<i>Aeneas</i>	∞	13.70	19.17	26.96	14.74	10.75	17.12	16.77	6.91	27.95	14.87	11.76	10.19	23.02
<i>Aeschines</i>	15.33	∞	15.02	13.69	18.06	7.75	11.50	10.79	13.33	14.13	10.02	9.33	8.42	14.60
<i>Aristides</i>	19.45	13.17	∞	17.67	19.54	9.72	18.56	15.63	7.95	21.01	14.56	9.14	7.95	15.05
<i>Aristoteles</i>	26.54	11.67	16.74	∞	12.53	11.79	17.67	18.75	5.66	26.65	16.22	12.62	10.84	23.32
<i>Demosthenes</i>	16.27	16.99	20.42	14.50	∞	8.26	13.90	12.28	11.42	16.00	11.55	8.48	7.48	13.74
<i>Flavius</i>	7.66	3.07	5.94	9.12	3.64	∞	9.09	11.98	-0.54	8.26	12.29	11.15	10.35	8.19
<i>Herodotus</i>	15.72	8.68	17.47	16.64	11.94	11.88	∞	22.49	4.61	18.68	21.72	8.99	7.74	12.95
<i>Pausanias</i>	15.37	7.82	13.96	17.60	9.76	14.08	21.85	∞	3.50	17.74	26.83	10.56	9.12	13.77
<i>Plato</i>	10.25	15.15	10.87	9.42	13.49	5.91	8.63	7.98	∞	9.80	7.57	6.85	6.27	9.74
<i>Plutarch</i>	27.57	12.22	20.30	26.85	14.36	11.23	19.69	18.80	6.31	∞	16.50	11.23	9.70	19.63
<i>Polybius</i>	13.04	6.68	12.72	14.63	8.84	14.24	20.92	26.39	2.84	15.04	∞	9.27	8.02	11.58
<i>Strabo</i>	9.52	5.30	6.52	10.94	4.63	13.43	8.32	10.64	0.94	9.35	9.96	∞	24.21	11.73
<i>Thucydides</i>	7.50	3.86	4.89	8.71	3.16	12.59	6.60	8.71	-0.19	7.37	8.23	23.77	∞	9.36
<i>Xenophon</i>	22.96	12.69	14.71	23.88	11.96	11.05	14.54	15.52	5.99	20.08	13.77	13.74	11.84	∞

Table A4. Greek-1. Average Γ , C-Channel. The author/text in the first row is the reference, *i.e.* the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j . For example, if *Aristides* is the input and *Demosthenes* is the output, then $\Gamma = 21.83$ dB, viceversa, if *Demosthenes* is the input and *Aristides* is the output, then $\Gamma = 21.05$. Green color indicates very large Γ cases. Largest Γ : *Herodotus-Xenophon*, 44.99 dB; minimum Γ : *Aristotle-Polybius*, 9.99 dB.

Author	<i>Aeneas</i>	<i>Aeschi</i>	<i>Aristi</i>	<i>Aristo</i>	<i>Demo</i>	<i>Flavi</i>	<i>Hero</i>	<i>Pausa</i>	<i>Plato</i>	<i>Plut</i>	<i>Poly</i>	<i>Strabo</i>	<i>Thuc</i>	<i>Xen</i>
<i>Aeneas</i>	∞	24.99	24.34	11.39	17.12	19.42	19.32	25.15	16.22	27.37	27.38	16.70	15.28	19.09
<i>Aeschines</i>	25.01	∞	24.29	12.51	16.18	24.78	19.55	30.81	15.56	23.80	23.63	18.91	18.52	19.60
<i>Aristides</i>	24.89	24.85	∞	14.16	21.05	23.55	26.61	30.20	19.98	33.35	20.58	21.64	18.25	26.30
<i>Aristoteles</i>	13.62	14.43	15.84	∞	17.18	16.53	17.80	15.28	17.79	15.01	12.53	20.09	17.94	18.06
<i>Demosthenes</i>	18.25	17.52	21.83	16.08	∞	18.15	26.37	19.29	36.64	21.27	15.93	20.43	16.16	25.73
<i>Flavius</i>	20.10	25.30	23.39	15.11	16.96	∞	21.25	26.26	16.59	21.41	18.70	24.08	24.48	21.64
<i>Herodotus</i>	20.24	20.44	27.01	16.50	25.93	21.99	∞	23.19	24.85	24.57	17.48	24.82	18.74	44.97
<i>Pausanias</i>	25.57	31.07	29.91	13.52	18.22	26.01	22.56	∞	17.47	27.76	22.06	20.87	19.03	22.59
<i>Plato</i>	17.47	16.97	20.84	16.87	36.78	17.87	25.38	18.62	∞	20.19	15.35	20.60	16.23	24.98
<i>Plutarch</i>	27.74	24.35	33.18	13.14	20.48	21.38	24.01	27.98	19.29	∞	21.82	19.47	16.79	23.58

Continued

<i>Polybius</i>	27.00	23.03	19.71	9.99	14.42	17.56	16.21	21.29	13.73	21.08	∞	14.64	14.06	16.09
<i>Strabo</i>	17.98	19.88	22.34	19.14	19.98	24.61	24.92	21.68	19.96	20.34	16.23	∞	23.64	25.81
<i>Thucydides</i>	16.36	19.47	18.52	16.75	15.15	24.88	18.25	19.68	15.06	17.25	15.63	23.11	∞	18.63
<i>Xenophon</i>	20.05	20.48	26.72	16.79	25.27	22.37	44.99	23.21	24.43	24.17	17.38	25.73	19.14	∞

Appendix C. Linguistic Channels in Greek-2 Texts

Table A5. Greek-2. Correlation and slope of the regression lines between the indicated variables. Four digits of the correlation coefficient are reported because some authors/texts differ only at the third/fourth digit.

Author	S-Channel		I-Channel		WI-Channel		C-Channel	
	Sentences vs words		Word Intervals vs Sentences		Words vs Interpunctuations		Characters vs Words	
	Correlation	Slope	Correlation	Slope	Correlation	Slope	Correlation	Slope
<i>Aeschylus</i>	0.9150	0.0760	0.9106	2.2652	0.9019	5.5848	0.9947	5.2099
<i>Aesop</i>	0.9032	0.0545	0.9302	3.4236	0.9860	5.2809	0.9966	5.2351
<i>Euripides</i>	0.7416	0.0775	0.8521	2.3959	0.9673	5.1510	0.9943	4.9407
<i>Homer's Iliad</i>	0.9136	0.0343	0.9295	4.0631	0.9855	7.1000	0.9921	4.8988
<i>Homer's Odyssey</i>	0.9756	0.0412	0.9744	4.2355	0.9919	5.7158	0.9989	4.8945
<i>Pindarus</i>	0.9771	0.0455	0.9729	3.3394	0.9934	6.4488	0.9992	5.4343
<i>Sofocles</i>	0.8917	0.0744	0.9266	2.4612	0.9857	5.2563	0.9978	4.7420

Table A6. Greek-2. Average Γ , S-Channel. The author/text in the first row is the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j .

Author	<i>Aeschylus</i>	<i>Aesop</i>	<i>Euripides</i>	<i>Iliad</i>	<i>Odyssey</i>	<i>Pindarus</i>	<i>Sofocles</i>
<i>Aeschylus</i>	∞	8.04	9.75	-1.70	0.73	2.48	24.49
<i>Aesop</i>	10.95	∞	8.77	4.58	7.13	9.31	11.43
<i>Euripides</i>	9.41	4.43	∞	-3.29	-2.80	-1.61	10.86
<i>Iliad</i>	5.21	8.61	4.79	∞	12.54	10.71	5.36
<i>Odyssey</i>	6.56	10.52	5.09	10.08	∞	20.47	6.60
<i>Pindarus</i>	7.55	11.86	5.47	7.39	19.61	∞	7.54
<i>Sofocles</i>	24.83	8.71	11.56	-1.40	0.66	2.32	∞

Table A7. Greek-2. Average Γ , I-Channel. The author/text in the first row is the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j .

Author	<i>Aeschylus</i>	<i>Aesop</i>	<i>Euripides</i>	<i>Iliad</i>	<i>Odyssey</i>	<i>Pindarus</i>	<i>Sofocles</i>
<i>Aeschylus</i>	∞	9.37	17.69	7.07	6.42	9.17	21.12
<i>Aesop</i>	5.73	∞	6.06	16.06	12.88	16.52	8.15
<i>Euripides</i>	16.79	9.77	∞	7.47	6.48	8.68	15.67

Continued

<i>Iliad</i>	1.96	14.57	2.43	∞	16.39	11.06	3.73
<i>Odissey</i>	0.46	10.42	0.26	15.69	∞	11.42	2.25
<i>Pindarus</i>	5.12	16.95	4.38	13.37	13.49	∞	7.68
<i>Sofocles</i>	20.26	11.02	15.21	8.08	7.35	10.87	∞

Table A8. Greek-2. Average Γ , WI-Channel. The author/text in the first row is the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j .

Author	<i>Aeschylus</i>	<i>Aesop</i>	<i>Euripides</i>	<i>Iliad</i>	<i>Odissey</i>	<i>Pindarus</i>	<i>Sofocles</i>
<i>Aeschylus</i>	∞	10.21	12.95	10.21	9.79	9.71	10.21
<i>Aesop</i>	11.17	∞	20.46	11.83	21.45	14.60	46.00
<i>Euripides</i>	14.25	20.88	∞	11.01	16.30	12.72	21.12
<i>Iliad</i>	6.92	9.26	8.03	∞	12.11	18.56	9.10
<i>Odissey</i>	9.39	20.62	14.85	14.07	∞	18.85	20.12
<i>Pindarus</i>	7.40	12.75	10.21	19.60	17.79	∞	12.52
<i>Sofocles</i>	11.24	46.04	20.78	11.71	20.99	14.42	∞

Table A9. Greek-2. Average Γ , C-Channel. The author/text in the first row is the channel input Y_k ; the author/text in the first column is the channel dependent output Y_j .

Author	<i>Aeschylus</i>	<i>Aesop</i>	<i>Euripides</i>	<i>Iliad</i>	<i>Odissey</i>	<i>Pindarus</i>	<i>Sofocles</i>
<i>Aeschylus</i>	∞	33.56	25.25	23.35	21.12	22.71	19.45
<i>Aesop</i>	33.48	∞	23.75	21.64	22.01	25.20	19.53
<i>Euripides</i>	25.71	24.33	∞	33.58	24.25	19.23	24.51
<i>Iliad</i>	23.95	22.39	33.71	∞	22.04	18.04	23.12
<i>Odissey</i>	21.91	22.72	24.41	22.05	∞	20.04	28.43
<i>Pindarus</i>	22.09	24.69	18.13	16.77	19.13	∞	16.53
<i>Sofocles</i>	20.37	20.42	25.05	23.62	28.78	17.76	∞