Increase of Complexity from Classical Greek to Latin Poetry

R. Mansilla
E. Bush

Center for Interdisciplinary Research in Science and the Humanities,
National Autonomous University of Mexico

In this paper a method is developed to analyze the increase of complexity from classical Greek poetry to classical Latin poetry by mapping large samples of poems onto a symbolic time series. This mapping setup intends to characterize the regular succession of rhythms, that is, the patterns of stressed and unstressed syllables in a verse. Using techniques from information theory; more precisely, the mutual information function, it is shown how the rhythmical patterns in Greek poetry evolve to more complex behavior in Latin poetry. Some interesting results are reported.

1. Introduction

The human language is a system of objects, rules, and relations of remarkable complexity. The purest expression of abstract and symbolic human imagery is without doubt poetry. Without taking blank verse into account, poetry in English and Spanish is normally characterized by a regular succession of rhythm, that is, a pattern of stressed and unstressed syllables in a verse. Classical Greek and Latin poetry, on the other hand, not only has as a distinctive feature a chain of stressed and unstressed syllables, but also one of length, in other words, a pattern of "long" and "short" quantities of vowel length.

The study of symbolic sequences has a long history. Gatlin's work [1] is probably the cornerstone for this kind of study. However, another pertinent publication which has exerted influence in the community is [2].

The discovery of the DNA structure has been influential in the study of symbolic sequences, and has revolutionized our thoughts regarding biological evolution. The studies of long-range correlation in DNA [3–17] have been a fruitful line of research in recent years. A symbolic sequence is said to have long-range correlations if its power spectrum scales as $1/d^c$, where d is the distance between symbols in the sequence and $c \approx 1$.

Success reached in the study of regularities of the DNA molecule has stimulated the study of written languages and music with similar techniques [18–22]. A recent paper in this direction is [23].

In the present paper, borrowing tools from information theory that have been used before in the DNA long-range correlation studies, we shall demonstrate that a particular type of rhythm in Greek and Latin poetry, the hexameter, can be analyzed with techniques similar to those reported above. We show that although both Greek and Latin poets used the hexameter the former did it in a more strict sense, while the Latin poets tended to "break the rules," creating (in reference to rhythm) more complex pieces of poetry. The evolving patterns of mutual and partial information functions, when applied to symbolic sequences mapped from real verses, allow us to support our claims. The structure of the paper is as follows. In section 2 we develop a brief overview of the ancient poets and their works analyzed here, as well as a description of the methodology used to map the verses into a symbolic sequence. In section 3 we develop the main concepts and results related to information theory used in the paper. In section 4 we discuss our results and section 5 concludes the paper.

2. An overview on ancient poetry

We shall give a synoptic overview of the ancient poets analyzed here, as well as a description of the methodology used to map the verses into a symbolic sequence.

2.1 Classical Greek poetry

The hexameter, so named on account of it being a verse of six meters, or feet, was first used by Homer for the *Iliad* and the *Odyssey*,[1] both epic poems that date back as far as the eighth century B.C. The former deals with the wrath of Achilles in the last year of the war against the Trojans, and the latter with the last year of the homecoming, travels, and wanderings of Odysseus (Ulysses) 10 years after the sack of Troy. Linguistically, the poem is packed with formulae, a string of words repeated in many places throughout both poems and with epithets, for example, "the grey-eyed Hera." Both poetical devices are used in order to meet the criteria of rhythm within a verse.

But the hexameter was not only limited to the singing of heroes and their deeds and travails in antiquity, it was also used in the genre of didactic and bucolic poetry. The first poet to use the hexameter for pedagogical and philosophical purposes was Hesiod, a poet whose worldview was radically different from Homer's. His *Works and Days* is

---

[1]There is a considerable amount of bibliography on whether Homer actually existed, and when and where the *Iliad* and the *Odyssey* were composed. For the purposes of the present paper, we have deliberately set the so-called "Homeric question" aside and have focused, instead, on considering the two extant poems as independent wholes. In a future paper we will discuss our result about this topic.

didactic in nature given that his purpose in composing it was to teach his brother how to work a piece of land and which days were particularly propitious for agriculture. Hesiod's other poem, the *Theogony*, was an attempt at explaining the origins of the universe and the genealogy of the major and minor deities in ancient Greece. This systematization was also an attempt at giving the world, albeit in a mythical and religious context, a logical explanation.

Regarding the use of the hexameter for the genre of bucolic poetry, Theocritus is the model. His *Idylls* are characterized mainly by the description of bucolic situations in the countryside. Naturally, the urban element proper to the Hellenistic period of Greek history (323 B.C. to 30 B.C.), is absent from his poetry, with the exception of *Idyll* 15, and the emphasis is on the simple lifestyle of farmers. Theocritus is analyzed in this paper because Vergil, whom we will discuss in the section on classical Latin poetry, based some of his poems on the Greek poet from Syracuse.

It must be noted here that Homer and Hesiod, in their respective use of the hexameter, composed in the Ionian-epic dialect, an "artificial" dialect that was never actually spoken in the Greek-speaking areas of the Mediterranean, reserved only for poetry. It must also be remembered that using this dialect had religious and ritualistic purposes given that the epic was normally inspired by Calliope, the Muse of epic poetry. Theocritus, on the other hand, composed in the Doric dialect of ancient Greek, which was spoken in Syracuse. We should also remember that the nature of poetic language, in general, is to distance itself from everyday colloquial speech, hence the use of rhythm.

2.2 Latin poetry

With regards to classical Latin poetry, the poets analyzed here are Lucretius, Vergil, and Ovid. We decided that these three poets represent the best models of the Latin hexameter, since they emulated the Greek models mentioned above.

Lucretius, whose work *On the Nature of Things*, set out to explain in hexameters the ordering of the universe and man's relation to it using the Epicurean philosophy as his base, especially the concept of atoms as the force that constitutes every type of being. His purpose was, like Hesiod's, didactic and philosophical, but what differentiates them is that the Roman approached the world from a more rational perspective, whereas Hesiod was more emotional.

Vergil was especially proficient in the use of the hexameter for his poetical works, the *Aeneid* and *Georgics*. The *Aeneid* relates the wanderings of Aeneas, son of Venus and a survivor of the destruction of Troy, and his quest to found a new homeland. It is considered by many philologists to have its model in the contents of the *Odyssey*, but it is

also evident that Apollonious also exercised a good amount of influence on the Roman, since, like *Argonautica*, there is an absence of formulae and epithets in the Latin work. Vergil's objective was propagandistic, for the poem is filled with anachronistic references to the "greatness" of Rome. The *Georgics*, on the other hand, were didactic and, like Hesiod's *Works and Days*, dealt with descriptions of and advice on agriculture.

Ovid, Vergil's contemporary, was another Latin poet who found that the hexameter fit his poetical purposes. His best-known work is *Metamorphoses*, a compendium of myths of gods, semigods, and heroes who were transformed, or suffered a metamorphosis, from one state of being to another. The poem is a wealth of information on mythology in the ancient world.

▍ 2.3 The hexameter

As was noted at the beginning of this section, the hexameter in classical Greek and Latin poetry is rhythmical not only because of the regular succession of stressed and unstressed syllables, but also because each syllable can have a "long" or a "short" quantity. The rhythm of the hexameter is TA-ta-ta or TA-ta, depending on the combinations in each foot. But before we go any further, we shall look at the technical aspect of this type of verse.

The hexameter, as the etymology of the word indicates, is a verse of six meters, or feet, whose basic feature is the dactyl, that is, one long followed by two short syllables. Graphically, the verse looks like this:

__ uu | __ uu | __ uu | __ uu | __ uu | __ u

where __ indicates a long syllable, and u a short one. The dactyl __ uu is thus represented. But there is another feature of the hexameter, the spondee, which is represented thus __ __, that is, a long followed by another long. The vertical bar separates each foot from the other.

In the first four feet there can be either dactyls or spondees. The last foot also allows substitution of the short syllable for the long. The verse, then, is represented as follows with the possible substitutions:

__ u̲u̲ | __ u̲u̲ | __ u̲u̲ | __ u̲u̲ | __ uu | __ u.

Notice that the fifth foot does not allow any substitution. This, generally speaking, occurs in the Greek and Latin poets. However, when the dactyl of the fifth foot is substituted with a spondee, the fourth foot without exception is a dactyl. Each foot has a princeps and a biceps, the former being their first half, always long, and the biceps being the second half, either two shorts or one long. The princeps is, rhythmically speaking, characterized by the arsis, or the up beat, and the biceps by the thesis, or the down beat, resulting in the rhythm, as was seen above, of TA-ta-ta or TA-ta.

Another essential feature of the hexameter is the caesura, or pause. There is widespread disagreement on how the caesura is really used: as a rhythmical pause, allowing a poet to, as it were, "catch his breath," or as a semantic marker, that is, a device used to call attention to a certain word or turn of phrase within the verse. Either way, Greek and Latin poets followed certain rules regarding the placement of the caesura.

Normally, the caesura is placed at a word-end within the foot, and it can also be placed so as to coincide with a word-end and a foot-end. The placement depends on the foot, that is, the position is normally after the princeps or between the biceps when there is no substitution. In the first case, the caesura is "masculine," and in the second it is "feminine." The most common caesurae are known as trimimeral, penthemimieral, and hephthemimieral, respectively, which is the same as saying they fall after the third, fifth, and seventh princeps. There is also the bucolic caesura, which coincides with the end of the fourth foot. An example of an analysis of a verse, also called scansion, is taken from the first verse of Book I of the *Aeneid*:

> *Arma uirumque cano, Troiae qui primus ab oris*

and it is analyzed thus

> __ uu | __ uu | __ ⇑ __ | __ uu | __ uu | __ u.

The double-bar arrow indicates a caesura, which in this case, is penthemimeral. It is placed at a word-end, *cano*, and coincides with the comma. The rhythm of the verse is read as follows:

> *ARma uirUMque caNO, TroiAE qui PRImus ab Oris*   (1)

or

> TA-ta-ta TA-ta-ta TA ⇑ ta TA-ta-ta TA-ta-ta TA-ta.   (2)

Generally, if the penthemimeral appears, there are no other caesurae. If there is a trimimeral, it is usually corresponded by the hephthemimeral or the bucolic caesura. These rules are not airtight, since there are instances in which punctuation (commas, periods, etc.) plays an important role and can be found in the first foot, thus indicating a caesura.

The possible combinations within a foot are

> __ uu, __ __, __ ⇑ uu, __ u ⇑ u, __ ⇑ __, __ uu ⇑, __ __ ⇑, __ u.

In this paper, all the elements of the hexameter are mapped into symbolic time series using a trinary system where a long syllable is translated as 0, a short as 1, and the caesura as 2. Verses (1) and (2) analyzed above are thus 0110110200101101. In order to have representative samples of each poet, we decided that the first 100 verses of each Greek and Latin poet were sufficient. However, two things should be noted: first, we did

not take the grammatical and syntactic aspects into account; second, further study is forthcoming in which the complete works of the Greek and Latin poets mentioned here will be analyzed so as to arrive at more definite and sounder conclusions (see footnote 1).

At the risk of sounding repetitive, the poets and the first 100 verses of their works analyzed here are: Homer, Books I of the *Iliad* and the *Odyssey*; Hesiod, *Works and Days*; Theocritus, Book I of *Idylls*. The Latin poets and their works are: Vergil, Books I of the *Aeneid* and the *Georgic*; Ovid, Book I of *Metamorphoses*; and Lucretius, Book I of *On the Nature of Things*. The texts used are from the Oxford editions.

3. The mutual information function

The concept of a "mutual information function" can be found for the first time in the epic 1948 paper by Shannon [24]. His results were generalized to abstract alphabets by several Soviet mathematicians, culminating in the work of Dobrushin [25]. Initially they were used to measure the difference between the average uncertainty in the input of an information channel before and after the output is received. More recently the method has been used to study some properties of regular languages and cellular automata [26], DNA long-range correlation [1, 6–8, 15, 16] and some properties of strange attractors [27]. A thorough discussion of the properties of this function can be found in [28–30].

Let us denote by $A = \{\alpha_1, \ldots, \alpha_n\}$ an alphabet and $s = (\ldots, a_0, a_1, \ldots)$ an infinite string with $a_i \in A$, $i \in Z$, where Z represents the set of all integer numbers. The mutual information function of the string s is defined as:

$$M(d, s) = \sum_{\alpha, \beta \in A} P_{\alpha, \beta}(d, s) \ln\left[\frac{P_{\alpha, \beta}(d, s)}{P_\alpha(s) P_\beta(s)}\right] \tag{3}$$

where $P_{\alpha, \beta}(d, s)$ is the joint probability of having the symbol α followed d sites away by the symbol β on the string s and $P_\alpha(s)$ is the density of the symbol α in the string s.

We will also denote the partial (α, β) information function as:

$$M_{\alpha, \beta}^P(d, s) = P_{\alpha, \beta}(d, s) \ln\left[\frac{P_{\alpha, \beta}(d, s)}{P_\alpha(s) P_\beta(s)}\right]. \tag{4}$$

Notice that the function defined in equation (3) represents the sum of the contribution of functions defined in equation (4) for every pair $(\alpha, \beta) \in A^2$. When no confusion is possible we drop the symbol s from equations (3) and (4), resulting in the following:

$$M(d) = \sum_{\alpha, \beta \in A} P_{\alpha, \beta}(d) \ln\left[\frac{P_{\alpha, \beta}(d)}{P_\alpha P_\beta}\right] \tag{5}$$

$$M^P_{\alpha,\beta}(d) = P_{\alpha,\beta}(d) \ln \left[\frac{P_{\alpha,\beta}}{P_\alpha P_\beta} \right]. \tag{6}$$

In our work we consider the alphabet $A = \{0, 1, 2\}$ and obviously our sequences are not infinite, but large enough to allow stable statistical estimations of $P_{\alpha,\beta}(d)$ and P_α.

As we show below, equations (5) and (6) often have periodic behavior. Therefore, we briefly describe here a method widely used to study those kinds of behavior. Fourier spectra [31] are used in time series analysis, because the visual representation in the frequency domain can more easily reveal patterns that are harder to discern in the primary data, for example, intricate periodical behavior. Here we use a Fourier transform of mutual and partial information functions to detect some periodical behavior of those functions when applied to the symbolic strings mapped from the verses. From this point forward, we denote the power spectra of mutual (partial) information functions to the product of a Fourier transform of those functions by its complex conjugate:

$$\hat{S}(k) = \theta \left| \sum_{d=1}^{L} M(d) e^{-2\pi(k/L)d} \right|^2$$

where θ is a constant related with the sample frequency and L is the number of data available for $M(d)$.

4. Discussion

We begin our analysis with the behavior of the mutual information function related to each verse. It drastically changes from Greek to Latin poets. In Figure 1(a) the mutual information function of Homer's *Iliad* and in Figure 1(b) that of Vergil's *Aeneid* is shown. They represent
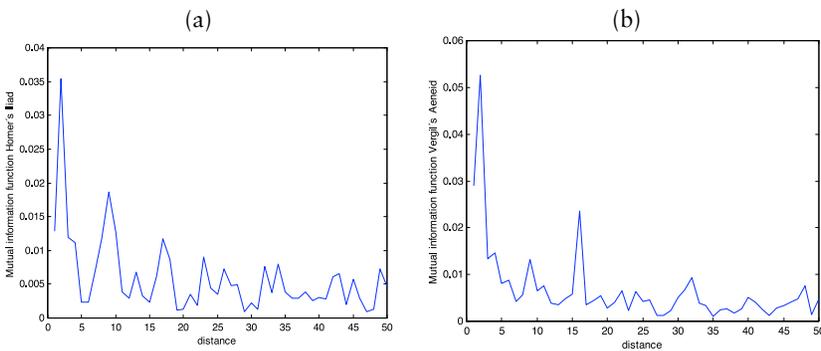

Figure 1. Mutual information functions of Homer's *Iliad* (a) and Vergil's *Aeneids* (b).

the average behavior for Greek and Latin poems respectively. The peak at distance $d = 2$ is more pronounced in Vergil than in Homer. It is related to the long syllable common to the dactyl and the spondee and the substitution between them. The peak at distance $d = 9$ in Homer's *Iliad* almost disappears in Vergil's *Aeneid*. Therefore, some structures inside the verses were relaxed in Latin poetry with reference to the Greek. Below, we study these facts more accurately. Recall that the mutual information function of equation (5) is the sum of the partial contributions from equation (6). Finally, the peak at distance $d = 17$ in Homer's *Iliad* moves to $d = 16$ in Vergil's *Aeneid* and is more pronounced there. As we show below, this fact is related to the use and position of caesurae.

We will now take a closer look at the structure of the verses using partial information functions. A remarkable property of the functions $M_{0,1}^P(d)$ and $M_{1,1}^P(d)$ is that their graphs are a mirror image of each other with respect to the horizontal axis. Due to the substitution of two short syllables for a long syllable in the biceps position, when at certain distances there is an abundance of zeroes there should be a shortage of ones, which is the result of correlations and anticorrelations. An important consequence of this fact is that $M_{0,1}^P(d) + M_{1,1}^P(d) = 0$ and hence they contribute nothing to the mutual information function. In Figure 2(a) the $M_{0,1}^P(d)$ and $M_{1,1}^P(d)$ functions of Homer's *Iliad* are shown. In Figures 2(b) and 2(c) the $M_{0,1}^P(d)$ and $M_{1,1}^P(d)$ functions for the Greek and Latin poets, respectively, are shown. Therefore peaks at distances $d = 2, 9$ in mutual information functions of every poem are contributed by the remainder of the partial information functions.

The relationship between the princeps of the feet and caesurae is also different in Greek and Latin poetry. In Figures 3(a) and 3(b) the $M_{0,2}^P(d)$ functions for Greek and Latin poems, respectively, are shown. Notice the peak at distance $d = 9$ in both sets of plots. The height is larger in Greek poems than in Latin ones. The peak at distance $d = 13$ almost disappears in Latin poems. These facts suggest that the use of more than one caesura in each verse is more frequent in Latin than in Greek poems. Besides, the rhythmical structure in Greek poems encompasses several verses as the peaks at distances $d = 26, 44$ show, which almost disappears in Latin poems.

The distance between two consecutive caesurae also yields some insights about how more complex Latin poetry is. In Figure 4 $M_{2,2}^P(d)$ for Homer's *Iliad*, Homer's *Odyssey*, and Lucretius' *Nature of Things* is shown. Notice the remarkable periodic behavior of $M_{2,2}^P(d)$ for *Iliad* that maintains a high correlation between caesurae at a distance of 170 symbols. This feature disappears in Lucretius' verse. The periodic structure persists however in Latin poets, although it is weakened. In Figure 5 the averaged power spectrum of Greek and Latin poems is shown. The
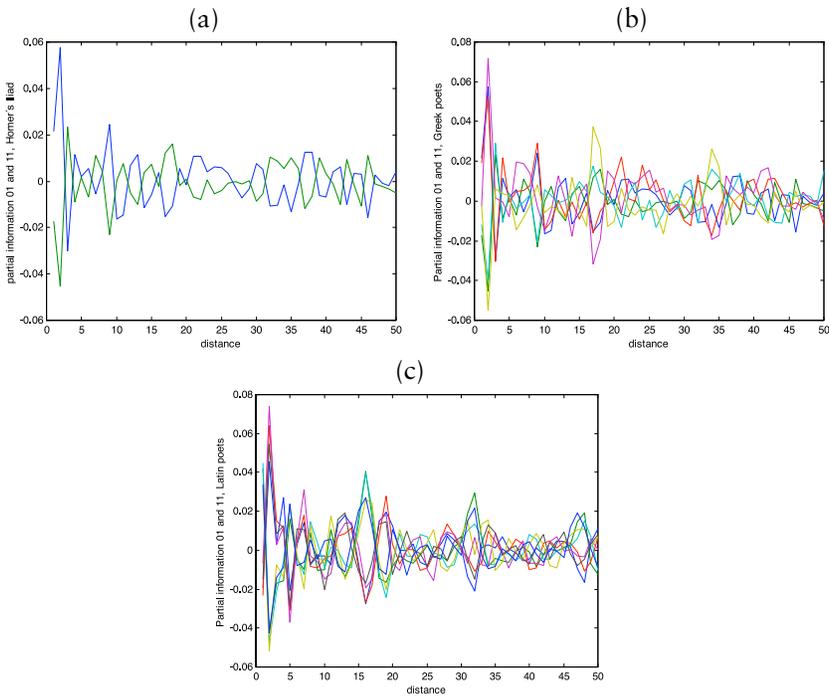
Figure 2. Partial information functions $M_{0,1}^P(d)$ and $M_{1,1}^P(d)$ for Homer's *Iliad* (a), all the Greek poems (b), and all the Latin poems (c). Notice the mirror behavior of the partial information functions (see the text).
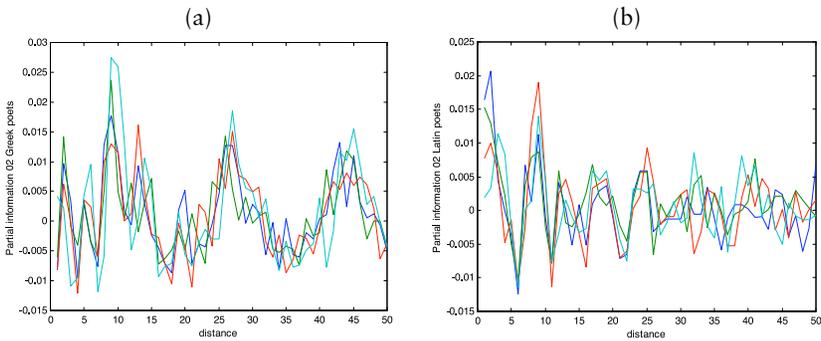

Figure 3. Partial information functions $M_{0,2}^P(d)$ for Greek poems (a) and Latin poems (b).
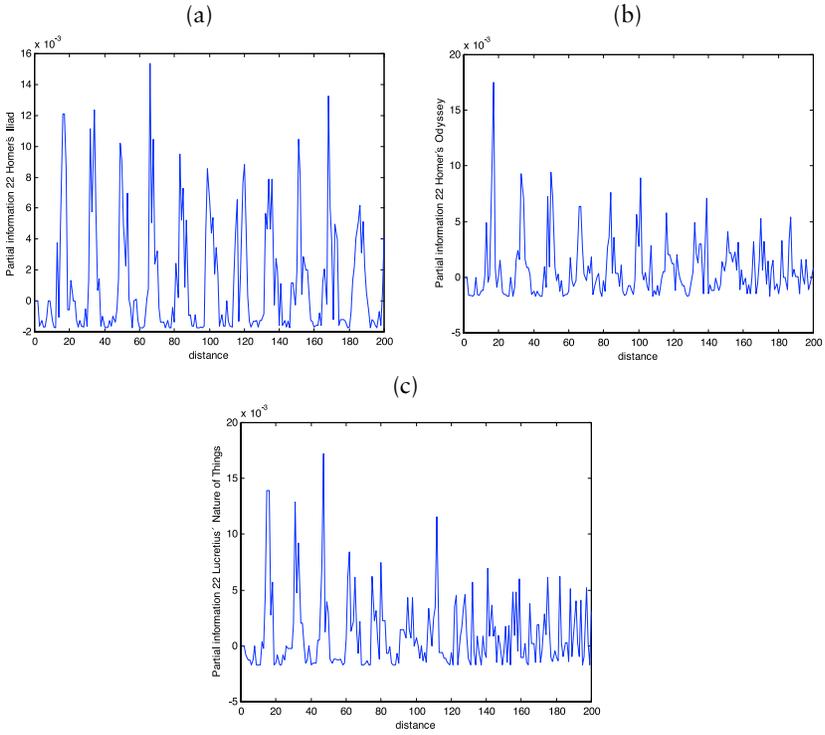
Figure 4. Partial information functions $M_{2,2}^{P}(d)$ for Homer's *Iliad* (a), Homer's *Odyssey* (b), and Lucretius' *Nature of Things* (c).
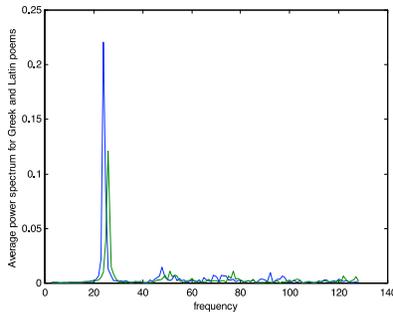

Figure 5. Averaged power spectrum for Greek and Latin poems.

first large peak almost coincides in position, but is smaller in Latin poems, reflecting a lower influence of the corresponding harmonic. Our explanation of this fact is that rhapsodes in Homeric times had to memorize large pieces of verses, because no written verses existed at the time. Therefore, a strict rhythm facilitated memorization. Meanwhile, in Lucretius' time declaimers also memorized verses, but composition could be done by writing them, which allows more complicated structures in the poetry.

With reference to the *Odyssey*, it proved to be very flashy how the same person could write two epic poems in the same period with that striking difference in rhythm. The behavior of *Odyssey* is closer to Lucretius' *Nature of Things* than to Homer's *Iliad*. Our results probably could throw light on the question of if the *Iliad* and *Odyssey* are by the same author.

5. Conclusions

This paper shows that an increase in the complexity of rhythmical structure arises in the use of hexameter from Greek classic poets to Latin ones. The method allows discrimination between Greek and Latin poems using the differences in the use of dactyl and spondee, and in the position and use of caesurae. The results also open the possibility of deciding on the Homeric question.

References

[1] L. L. Gatlin, *Information Theory and the Living System* (Columbia University Press, 1972).

[2] H. Herzel, "Complexity of Symbol Sequences," *Sys. Anal. Mod. Sim.*, **5** (1988) 435–444.

[3] D. Borsnitk, *et al.*, "Analysis of Apparent $1/f^{\alpha}$ Spectrum in DNA Sequences," *Europhysics Letters*, **23** (1993) 389–394.

[4] A. L. Buldyrev, *et al.*, "Long-range Correlation Properties of Coding and Noncoding DNA Sequences: GenBank Analysis," *Physical Review E*, **51** (1995) 5084–5091.

[5] S. Karlin and V. Brendel, "Patchiness and Correlations in DNA Sequences," *Science*, **259** (1993) 677–680.

[6] W. Li, "Generating Nontrivial Long-range Correlation and $1/f$ Spectra by Replication and Mutation," *International Journal of Bifurcation and Chaos*, **2** (1992) 137–154.

[7] W. Li and K. Kaneko, "Long-range Correlation and Partial $1/f^{\alpha}$ Spectrum in a Noncoding DNA Sequence," *Europhysics Letters*, **17** (1992) 655–660.

[8]   P. Miramontes, "Cellular Automata, Genetic Algorithms, and DNA evolution," (Ph. D. thesis UNAM, 1992).

[9]   C. K. Peng, *et al.*, "Long-range Correlation in Nucleotide Sequences," *Nature*, **356** (1992) 168–170.

[10] S. Nee, "Uncorrelated DNA Walks," *Nature*, **357** (1992) 450.

[11] V. V. Prabhu and J. M. Claverie, "Correlations in Intronless DNA," *Nature*, **359** (1992) 782.

[12] W. Li and K. Kaneko, "DNA Correlations," *Nature*, **360** (1992) 635–636.

[13] C. A. Chatzidimitrou-Dreismann and D. Larhamar, "Long-range Correlation in DNA," *Nature*, **361** (1993) 212–213.

[14] C. K. Peng, *et al.*, "Finite Size Effects on Long-range Correlations: Implications for Analyzing DNA Sequences," *Physical Review E*, **47** (1993) 3730–3733.

[15] R. Mansilla and R. Mateo-Reig, "On Mathematical Modeling of Intronic Sectors of the DNA Molecule," *International Journal of Bifurcation and Chaos*, **5** (1995) 1235–1241.

[16] R. Voss, "Evolution of Long-range Correlations and $1/f$ Noise in DNA Base Sequences," *Physical Review Letters*, **68** (1992) 3805–3808.

[17] R. Mansilla and G. Cocho, "Multiscaling in Expansion Modification Systems: An Explanation for Long Range Correlation in DNA," *Complex Systems*, **12** (2000) 207–240.

[18] R. Voss, and J. Clarke, "$1/f$ Noise in Music and Speech," *Nature*, **258** (1975) 317–318.

[19] Yu. Klimontovich and J. Boon, "Natural Flicker Noise ($1/f$ Noise) in Music," *Europhysics Letters*, **3** (1987) 395–399.

[20] A. Schenkel, J. Zhang, and Y. Ch. Cheng, *Fractals*, **1** (1993) 47–52.

[21] Y. Shi, "Correlations of Pitches in Music," *Fractals*, **4** (1996) 547–553.

[22] T. Saiki, *et al.*, "Fluctuations of Character Centroid Intervals in Laterally Written Japanese Sentences," *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, **E82A**(3) (1999) 520–526.

[23] M. A. Montemurro and P. A. Pury, "Long-range Fractal Correlations in Literary Corpora," cond-mat/0201139 (to appear in *Fractals*).

[24] C. Shannon, "A Mathematical Theory of Communication," *Bell Systems Technologies Journal*, **27** (1948) 379–423.

[25] R. Dobrushin, "General Formulation of Shannon's Main Theorem in Information Theory," *Uspieji Matematicheskix Nauk*, **14** (1959) 1–104. English translation in *American Mathematical Society Transactions*, **33** (1959) 323–438.

[26] W. Li, "Power Spectra of Regular Languages and Cellular Automata," *Complex Systems*, **1** (1987) 107–113.

[27] A. Fraser and H. Swinney, "Independent Coordinates for Strange Attractors from Mutual Information," *Physics Review A*, **33** (1986) 1134–1151.

[28] R. Blahut, *Principles and Practice of Information Theory* (Addison-Wesley, 1987).

[29] W. Li, "Mutual Information Function versus Correlation Function," *Journal of Statistical Physics*, **60** (1990) 823–831.

[30] A. Renyi, *The Probability Theory* (Pergamon Press, 1970).

[31] D. B. Percival and A. T. Walden, *Spectral Analysis for Physics Applications* (Cambridge University Press, 1993).