Competition between two kinds of correlations in literary texts

S. S. Melnyk, O. V. Usatenko, and V. A. Yampol'skii*
A. Ya. Usikov Institute for Radiophysics and Electronics, Ukrainian Academy of Science, 12 Proskura Street, 61085 Kharkov, Ukraine

V. A. Golick
Academic Gymnasium No 45, 46-a Tobolskaya Street, Kharkov, Ukraine
(Received 6 February 2004; published 31 August 2005)

A theory of additive Markov chains with long-range memory is used to describe the correlation properties of coarse-grained literary texts. The complex structure of the correlations in the texts is revealed. Antipersistent correlations at small distances, $L \lesssim 300$, and persistent ones at $L \gtrsim 300$ define this nontrivial structure. For some concrete examples of literary texts, the memory functions are obtained and their power-law behavior at long distances is disclosed. This property is shown to be a cause of self-similarity of texts with respect to the decimation procedure.

DOI: 10.1103/PhysRevE.72.026140 PACS number(s): 02.50.Ga, 05.40.−a, 87.10.+e

I. INTRODUCTION

The problem of symbolic stochastic dynamic systems with long-range correlations has been under study for a long time in many areas of contemporary physics [1–6], biology [7–12], economics [8,13,14], etc. [8,15]. An important examples of such systems are literary texts [16–20].

One of the ways of gaining a correct insight into the nature of correlations in a symbolic system is to construct a mathematical object (for example, a correlated sequence of symbols) that possesses the same statistical properties as the initial dynamic system. There exist many algorithms for generating long-range correlated sequences: the inverse Fourier transformation [15,21], the expansion-modification Li method [22], the Voss procedure of consequent random additions [23], the correlated Levy walks [24], etc. [15]. Our belief is that, among the above-mentioned methods, the application of the *multistep Markov* chains is the most important one since it allows constructing random sequences with prescribed correlation properties. This was demonstrated in Ref. [25], where the Markov chains with the *stepwise memory function* (MF) had been studied. It was shown that there existed some symbolic dynamic systems (coarse-grained sequences of the Eukarya's DNA and dictionaries) with correlation properties that can be properly described by this model.

The multistep Markov chain is the sequence of symbols of some alphabet constructed using a conditional probability function, which determines the probability that some definite symbol of a sequence depending on N previous ones occurs. The additivity property of the Markov chain implies the *independent* influence of previous symbols on a generated one. The concept of additivity, originally introduced in paper [20], was later generalized for the case of binary *nonstationary* Markov chains [26]. Another generalization was based on the consideration of Markov sequences with a multivalued alphabet [27,28].

The efficient method of investigating the dynamic systems is to decompose the space of states into a finite number of parts labelled by definite symbols, which are naturally ordered according to the dynamics of the system. The most frequently used decomposition procedure rests on the introduction of two parts of the phase space. In other words, the approach presupposes mapping of two kinds of states onto two symbols, say 0 and 1. This procedure is often referred to as coarse graining. Thus, the problem reduces to investigating the statistical properties of binary sequences.

It might be thought that the coarse graining could result in losing, at least, the short-range memory in the sequence. The authors of Ref. [28] argued that the mapping of a given sequence onto a small-alphabet sequence did not necessarily imply that the long-range correlations existing in the initial text would be preserved. However, as was shown in Ref. [20], the statistical properties of coarse-grained texts depend, but not essentially, on the kind of mapping. This implies that only the small part of all possible kinds of mapping can destroy the initial correlations in the system. Below, we demonstrate that the coarse graining retains, although not completely, the correlations at all distances. This means that there is no point in coding every symbol (associating every part of the phase space of the system with its binary code) to analyze the correlation properties of the dynamic systems, as it is done, for example, in Ref. [18], but it is sufficient to use the coarse-graining procedure.

In the present work, we examine the coarse-grained literary texts as the additive Markov chains. We use the obtained in Ref. [29] equation that connects mutually-complementary characteristics of these sequences, the memory and correlation functions. Once the memory function of the original sequence is found from the analysis of the correlation function, we construct the corresponding Markov chain with the same statistical properties. This method for constructing the sequence of elements having a given correlation function seems to be very important for other applications, e.g., it can be employed to fabricate the effective filters of electrical or optical signals [30].

We show that the memory function of any coarse-grained literary text is characterized by a complex structure because

*Electronic address: yam@ire.kharkov.ua

of the competition between two kinds of correlations. One type of correlations is efficient at short distances, $L \lesssim 300$. The corresponding MF is negative, which reflects the *antipersistent* nature of such correlations. The other type of correlations with the positive memory function acts at long distances, $L \gtrsim 300$. The strength of these *persistent* correlations decreases as a power-law function. We demonstrate that the power-law decrease of the memory function results in the self-similarity phenomenon in the coarse-grained texts with respect to the decimation procedure.

The paper is organized as follows. In the next section, we introduce some general relations for the additive Markov chains and present an equation connecting the correlation and memory functions. Section III contains the application of the concept of additive Markov chains to literary works. In the Conclusion, we summarize the obtained results.

II. MATHEMATICAL MODEL

A. Markov processes

Consider a homogeneous binary sequence of symbols, $a_i=\{0,1\}$. To determine the *N-step Markov chain* we have to introduce the *conditional probability* $P(a_i|a_{i-N},a_{i-N+1},\ldots,a_{i-1})$ of occurring the definite symbol a_i (for example, $a_i=1$) after the N-word $T_{N,i}$, where $T_{N,i}$ stands for the sequence of symbols $a_{i-N},a_{i-N+1},\ldots,a_{i-1}$. Thus, it is necessary to define 2^N values of the P-function corresponding to each possible configuration of the symbols a_i in N-word $T_{N,i}$. Since we will apply our theory to the sequences with long memory lengths of the order of 10^6, some special restrictions upon the class of P-functions should be imposed. We consider the MF of the *additive* form,

$$P(a_i=1|T_{N,i}=\sum_{r=1}^{N} f(a_{i-r},r). \tag{1}$$

Here the function $f(a_{i-r},r)$ describes the additive contribution of the symbol a_{i-r} to the conditional probability that the symbol unity, $a_i=1$, occurs at the ith site. The homogeneity of the Markov chain is provided by independence of the conditional probability (1) of the index i. It is possible to regard Eq. (1) as the first term in expansion of conditional probability in the formal series, where each term corresponds to the additive (unary), binary, ternary, and so on functions up to the N-ary one.

Let us rewrite Eq. (1) in an equivalent form,

$$P(a_i=1|T_{N,i}=\bar{a}+\sum_{r=1}^{N} F(r)(a_{i-r}-\bar{a}), \tag{2}$$

with

$$\bar{a}=\frac{\sum_{r=1}^{N} f(0,r)}{\left\{1-\sum_{r=1}^{N}[f(1,r)-f(0,r)]\right\}}$$

and

$$F(r)=f(1,r)-f(0,r).$$

We refer to $F(r)$ as the *memory function* (MF). It describes the intensity of the influence of previous symbol $a_{i-r}(r=1,\ldots,N)$ upon a generated one, a_i. It can be shown that $\bar{a}$ coincides with the value of a_i averaged over the whole sequence. As far as we know, the concept of the memory function for multistep Markov chains was originally used in Refs. [20,25] where they were well-suited to describe the symbolic correlated systems.

The memory function $F(r)$ contains complete information about the correlation properties of the Markov chain. Nevertheless, the correlation function and other correlation moments are usually employed as the input characteristics describing the correlated random systems. However, the correlation function describes not only the direct interconnection of elements a_i and a_{i+r}, but also takes into account their indirect interaction via other intermediate elements. Our approach operates with the "origin" characteristics of the system, specifically with the memory function. This allows one to disclose the fundamental intrinsic properties of the system which provide the correlations between the elements.

A sequence of symbols in a Markov chain can be thought of as the sequence of states of some particle, which participates in a correlated Brownian motion. Thus, every L-word (a set of consequent symbols of the length L) can be considered as one of the realizations of the ensemble of correlated Brownian trajectories in the "time" interval L. The positive values of the MF result in persistent diffusion where previous displacements of the Brownian particle in some direction provoke its consequent displacement in the same direction. The negative values of the MF correspond to the antipersistent diffusion where the changes in the direction of motion are more probable. Another physical system, the Ising chain of spins with long-range interactions, could also be associated with the Markov sequence for which the positive values of the MF correspond to the attraction of spins whereas the negative ones conform to the repulsion.

Below we will make use of some more statistical characteristics of the random sequences. We consider the distribution $W_L(k)$ of the words of definite length L by the number k of unities in them, $k_i(L)=\Sigma_{l=1}^{L} a_{i+l}$, and the variance $D(L)$,

$$D(L)=\overline{(k-\bar{k})^2}, \tag{3}$$

where the average $\overline{g(k)}$ is defined as $\overline{g(k)}=\Sigma_{k=0}^{L} g(k)W_L(k)$. Another important value is the correlation function,

$$K(r)=\overline{a_i a_{i+r}}-\bar{a}^2, \quad K(0)=\bar{a}(1-\bar{a}). \tag{4}$$

By definition, the correlation function is even, $K(r)=K(|r|)$. It is related to the variance by the equation [20],

$$K(r)=\frac{1}{2}[D(r-1)-2D(r)+D(r+1)], \tag{5}$$

or

$$K(r)=\frac{1}{2}\frac{d^2D(r)}{dr^2} \tag{6}$$

in the continuous limit.

The memory function being applied in Refs. [20,25] is characterized by stepwise behavior and is defined by two parameters only: the memory depth N and the strength f of symbol's correlations. The value of f was assumed to be independent of the distance r between the symbols at $r<N$. This memory function was employed to describe the long-range persistent properties of the coarse-grained literary texts, specifically, the superlinear dependence of the variance $D(L)$. However, it does not provide the description of the antipersistent behavior of $D(L)$ at short distances, which was noted in Refs. [25]. Obviously, we need a more complex memory function for a detailed description of the both short-range and long-range properties of the coarse-grained texts.

B. Equation for the memory function

We suggest two methods for finding the memory function $F(r)$ of the Markov chain, a_i, that possesses the same correlation function as a given random sequence, b_i. The first one is based on the minimization of a "distance," dist, between the Markov chain generated by means of a sought-for MF and the initial sequence b_i. This distance is given as

$$\mathrm{dist} = \overline{[b_i - P(b_i = 1|T_{N,i})]^2} \tag{7}$$

with P function (2). Equating the variational derivative $\delta\,\mathrm{dist}/\delta F(r)$ to zero, we get the following relation between the memory function $F(r)$ and the correlation function $K(r)$:

$$K(r) = \sum_{r'=1}^{N} F(r')K(r-r'), \quad r \geqslant 1. \tag{8}$$

Equation (8) can also be derived by a straightforward calculation of expression $\overline{a_i a_{i+r}}$ in Eq. (4) using definition (2) of the memory function.

The second method establishes a relationship between the memory function $F(r)$ and the variance $D(L)$,

$$M(r,0) = \sum_{r'=1}^{N} F(r')M(r,r'), \quad r \geqslant 1,$$

$$M(r,r') = D(r-r') - \{D(-r') + r[D(-r'+1) - D(-r')]\}. \tag{9}$$

This relationship is obtained by the double integration of Eq. (8). Equations (5) and $D(-r)=D(r)$ are used here.

The function $K(r)$, being a second derivative of $D(r)$, is less convenient and robust in computer simulations. It is the exacting reason why we prefer to use the second method (9). This is our instrument for finding the memory function $F(r)$ of a sequence using the known variance $D(L)$. The robustness of the proposed method was demonstrated in Ref. [29].

III. LITERATURE TEXTS VIEWED AS THE MARKOV CHAINS

A. Variance and correlation function

Now we apply our method to investigating the correlation properties of the coarse-grained literary texts. At the outset of this subsection, we examine the variance $D(L)$ of the coarse-grained text of the King James Version of the Bible [31]. The result from the numerical simulation is represented by the solid line in Fig. 1. The straight dotted line describes the variance $D_0(L)=L\bar{a}(1-\bar{a})$, which corresponds to the noncorrelated biased Brownian diffusion. One of the typical coarse-graining procedure was used to map the letters of the text onto the symbols zero and unity, $[(a-m)\mapsto 0,(n-z)\mapsto 1]$. It is clearly seen that the diffusion is antipersistent at small distances, $L\lesssim 300$ (see inset), whereas it is persistent at long distances [32]. The deviation of the solid line from the dotted one testifies to the existence of the correlations in the text of the Bible. To confirm this statement we break down the original text into subsequences of a given length $L_0=3000$ and randomly shuffle them. The results from the calculation of the variance for the coarse-grained initial and shuffled texts of the Bible are given in Fig. 2. For $L \ll L_0$, the difference in $D(L)$ is negligible [33]. At $L \sim L_0$, the variance and correla-

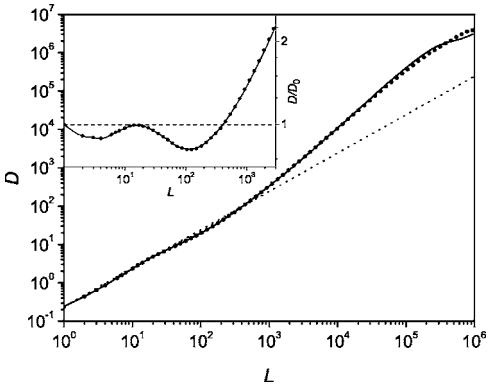

FIG. 1. The variance $D(L)$ for the coarse-grained text [letters $(a-m)\mapsto 0$, letters $(n-z)\mapsto 1$] of the Bible (solid line) and the Markov chain generated by means of the reconstructed memory function $F(r)$ (filled circles). The dotted straight line describes the noncorrelated Brownian diffusion, $D_0(L)=L\bar{a}(1-\bar{a})$. The inset demonstrates the antipersistent dependence of the dimensionless ratio $D(L)/D_0(L)$ upon L at short distances.

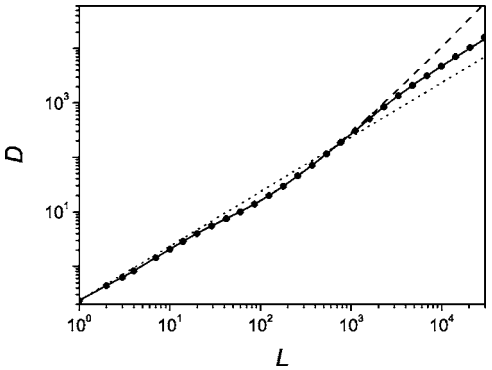

FIG. 2. The variance $D(L)$ for the coarse-grained text [(letters with even numbers in the alphabet) $\mapsto 1$, (ones with odd numbers) $\mapsto 0$] of the Bible (dashed line) and for the sequence obtained by shuffling the blocs of the length $L_0=3000$ (filled circles). The solid line represents the analytical results obtained from Eqs. (6) and (10). The dotted straight line describes the noncorrelated Brownian diffusion, $D_0(L)=L\bar{a}(1-\bar{a})$.

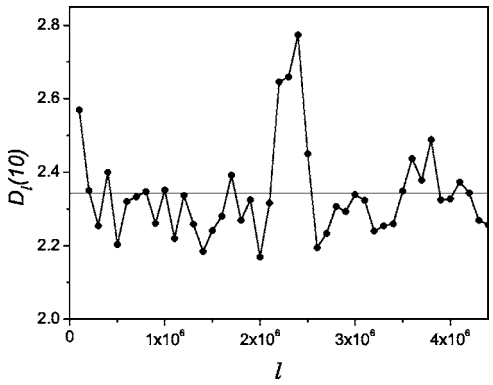

FIG. 3. The local variance $D_l(10)$ for the coarse-grained text of the Bible vs the distance l. The averaging interval is $L_0=10^5$.

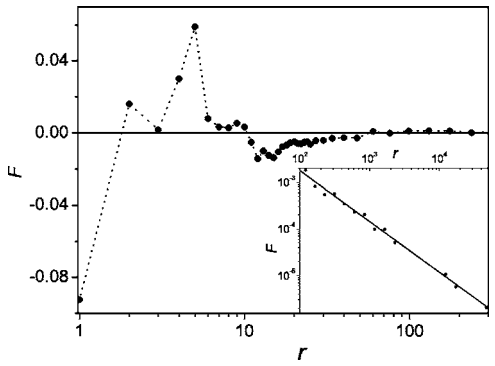

FIG. 4. The memory function $F(r)$ for the coarse-grained text of the Bible at short distances. The power-law decreasing portion of the $F(r)$ plot for the Bible is represented by the filled circles in the inset. The solid line is for the power-law fitting.

tion function of the shuffled sequence are less than the original ones. At $L \gg L_0$, the correlations in the shuffled text vanish. In this region, the variance $D(L)$ is a linear function, and the correlation function being the second derivative of the variance equals to zero.

It is easy to show that the correlation function of the shuffled sequence can be written as

$$K(r) = \begin{cases} K_0(r)\left(1 - \frac{r}{L_0}\right), & r < L_0, \\ 0, & r \geqslant L_0, \end{cases} \tag{10}$$

where $K_0(r)$ is the correlation function of the original nonshuffled sequence. The corresponding variance obtained by double numeric integration [see Eq. (6)] of the function $K(r)$ given by Eq. (10) is shown in Fig. 2 by the solid line.

Along with the global characteristic $D(L)$, it is interesting to study its local analogue,

$$D_l(L) = \langle (k - \langle k \rangle_{L_0})^2 \rangle_{L_0}, \tag{11}$$

where L_0 is the interval of local averaging and l is the coordinate of the left-hand border of this interval. An existence of a trend in the dependence $D_l(L)$ on l would be clearly indicative of nonstationarity of the stochastic process being studied. To verify the stationarity of the coarse-grained text of the Bible we perform the numerical simulation of the $D_l(L)$ dependence on l at different fixed values of L. As an example, the result of this simulation for $L=10$ is shown in Fig. 3. It is clearly seen that there exist regular fluctuations without a pronounced trend. The fluctuations result from the finiteness of the interval L_0 over which averaging is done. This fact allows us to make a conclusion about stationarity of the coarse-grained text of the Bible. The similar analysis of many other texts yielded the same result. It is expedient to study the global characteristics $D(L)$, Eq. (3), of the sequence instead of the local one, $D_l(L)$.

B. Memory function

According to Eqs. (8) and (9), the memory function can be restored using the variance or the correlation function. The MF thus obtained for the coarse-grained text of the Bible at $r<300$ is given in Fig. 4. At long distances, $r>300$, the memory function can be nicely approximated by the power function $F(r)=0.27r^{-1.1}$, which is shown by the solid line in the inset in Fig. 4. Note that the persistent part of the MF, $F(r>300) \leqslant 0.0008$, is much less than its typical magnitude 0.02 in the antipersistent region $r<40$.

It should be emphasized that the short-range part of the memory function at $r \lesssim 40$ and the $D(L)$ function at $L \lesssim 300$ are largely dependent on the method of coarse-graining. Nevertheless, the antipersistent correlations exist for practically all kinds of the coarse-graining procedure. An interesting feature is that the region $r \lesssim 40$ of negative antipersistent memory function provides far longer distances $L \sim 300$ of the antipersistent behavior of the variance $D(L)$.

To test the appropriateness of modeling the coarse-grained texts by the additive multi-step Markov chains, we have generated the sequence of symbols using conditional probability (2) with the reconstructed memory function $F(r)$ of the text of Bible. The variance $D(L)$ of this sequence is plotted by the filled circles in Fig. 1. The nice coincidence of this curve with the solid line describing the variance of the initial coarse-grained text of the Bible proves the robustness of our method of the MF reconstruction.

In order to prove the universal character in the power-law decrease of the memory function at long distances, we have made a comparison between the MFs of the coarse-grained texts for more than fifty different literary works. The texts are coarse-grained by mapping the letters from the first and second halves of the alphabet onto zero and unity, respectively. Subsequently, using Eq. (9), we first calculate the variances and then the memory functions. All the curves for the memory functions can be well fitted by the power-law functions $F(r)=cr^{-b}$. The results of the fitting for eight texts written or translated into Russian [34,35] are shown in Fig. 5. The exponents in all curves vary over the interval between $b_{min}=1.02$ for "War and Peace" by L. Tolstoy and $b_{max}=1.56$ for the Koran. Thus the constants c and b can be used for linguistic classification of different literary works. It is significant that the memory functions for the texts of the English- and Russian-worded Bible, as well as the texts of the Old and New Testaments are practically coincident.

The existence of two characteristic regions having different behavior of the memory function and, accordingly, of the

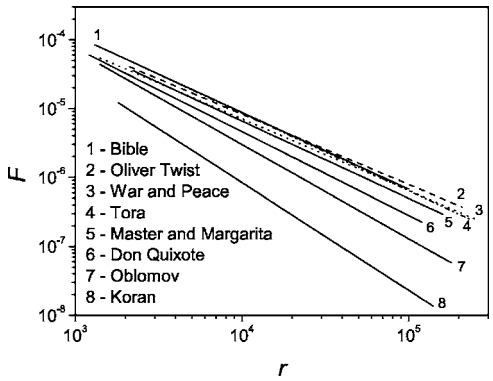

FIG. 5. The memory function at long distances for the coarse-grained texts of eight literary works: (1) The Bible (4.2×10^6 characters), (2) "Oliver Twist" by Charles Dickens (8.9×10^5 characters), (3) "War and Peace" by Leo Tolstoy (3.0×10^6 characters), (4) The Tora (6.7×10^5 characters), (5) "Master and Margarita" by Mikhail Bulgakov (8.0×10^5 characters), (6) "Don Quixote" by Miguel de Servantes (2.0×10^6 characters), (7) "Oblomov" by Ivan Goncharov (9.8×10^5 characters), and (8) The Koran (8.5×10^5 characters).

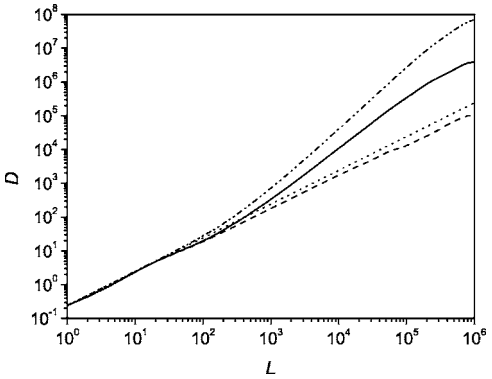

FIG. 6. The variance $D(L)$ for the coarse-grained text of the Bible [31] (the solid line), and for the sequences constructed using the persistent part of the MF (dash-dot-dotted line) and the antipersistent one (dashed line). The dotted line describes the noncorrelated Brownian diffusion, $D_0(L)=L\bar{a}(1-\bar{a})$.

persistent and antipersistent portions in the $D(L)$ dependence appears to be a prominent feature of all texts in any language. Note that the antipersistent portion of the memory function corresponds to the region where the grammatical rules are in use. Therefore, we call this kind of correlations the "grammatical" ones. The persistent correlations in a text at very long distances can be related to a general idea of the literary work. Thus, this kind of correlations is referred to as the "semantic" ones.

Two fundamentally different portions in the MF plots result from a peculiar competition between the two above-mentioned kinds of correlations. We would like to stress that both portions of the MF are equally important to provide an insight into the correlation properties of the literary texts. To support this statement we generate two special sequences. In both of them, only one kind of the memory function for the coarse-grained text of the Bible is taken into account, and the memory function in the other region is assumed to be zero. The variance $D(L)$ for these two sequences is given in Fig. 6. The lower (dashed) line represents the case where only the negative antipersistent portion, $r<40$, of the memory function is allowed for. The upper (dash-dot-dotted) curve is for the sequence that is generated by means of the long-range persistent memory, $F(r)=0.27r^{-1.1}$, $r>100$. It is evident that the generated sequence with the antipersistent memory function displays the subdiffusion only, whereas the sequence that corresponds to the persistent memory function is characterized by the superdiffusion behavior of the variance $D(L)$. The difference between the variances for two generated sequences and for the original coarse-grained text of the Bible, shown by the solid line in the same figure, corroborates our assumption about the significance of both kinds of the memory function.

C. Self-similarity of the coarse-grained texts

The power-law decrease (without characteristic scale) of the memory function at long distances leads to quite an essential property of *self-similarity* of the coarse-grained texts with respect to the *decimation* procedure discussed in Ref. [25]. This procedure implies the deterministic or random removal of some part of symbols from a sequence and is characterized by the decimation parameter $\lambda<1$ which represents the fraction of symbols kept in the chain. For example, under the random decimation each symbol is eliminated with probability $1-\lambda$. It can be shown that both of these procedures, deterministic and stochastic, are equivalent for a Markov chain. The sequence is self-similar if its variance $D(L)$ does not change after the decimation up to a definite value of L (which is dependent on the memory length of the original sequence and the decimation parameter). The model of the additive binary multistep Markov chain with the stepwise MF (which was discussed in Ref. [25]) offers the exact property of self-similarity at a length shorter than the memory length N. The coarse-grained literary texts are self-similar objects as well. It is indicated in Fig. 7 where three $D(L)$ curves are plotted for different values of the parameters

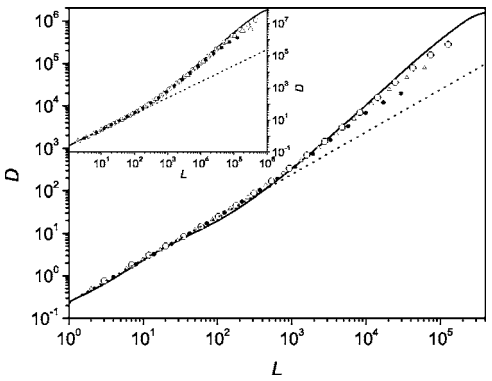

FIG. 7. The numerically calculated variance $D(L)$ for the coarse-grained text of the Bible [31] (solid line) and for the sequences obtained after their regular decimation. The circles, triangles, and dots are for the decimation parameters 2, 4, and 8, respectively. The dotted line describes the noncorrelated Brownian diffusion, $D_0(L)=L\bar{a}(1-\bar{a})$. The similar curves obtained for the sequence constructed by using the long-range part of Bible's memory function only are shown in the inset.

of the regular decimation. Note that the decimation procedure leads to a decrease in the effective memory length. As a result, the variance curves coincide up to the effective memory depth, which is proportional to the decimation parameter. A similar phenomenon occurs in the case of random decimation as well.

A question arises: what particular property of the memory function is crucial for the self-similarity of the coarse-grained literary texts? It is natural to assume that the persistent long-range scale-free portion of the memory function affords this property because the self-similarity is specifically manifest at long distances. To verify this supposition we carry out the decimation procedure with different λ for the Markov chain constructed by using the long-range part of the Bible memory function only and then plot the corresponding $D(L)$ dependence. The curves are shown in the inset in Fig. 7. The self-similarity for this sequence appears to be more pronounced than for the original coarse-grained text of the Bible. Moreover, the antipersistent part of the MF disappears very fast after the decimation. This is clearly revealed as a disappearance of the antipersistent sublinear portion of the $D(L)$ curves in Fig. 7. After the decimation, the solid line transforms into the entirely persistent superlinear curve, which goes above the curve $D_0=L\bar{a}(1-\bar{a})$. The conclusion about the invariance of the statistical properties of studied sequence with respect to the decimation procedure is an additional argument in favor of coarse-graining efficiency. The decimation can be thought of as additional coarse-graining of the initial random sequence.

IV. CONCLUSION

Thus we have demonstrated that the description of the literary works is suitable in terms of the Markov chains with complex memory functions. Actually, the memory function appears to be a convenient informative "visiting card" of any symbolic stochastic process. We have studied the coarse-grained literary texts and shown the complexity of their organization in contrast to a previously discussed simple power-law decrease of correlations. We have proved that the competition between the two kinds of correlations govern the statistical properties of the coarse-grained texts. The antipersistent correlations exist at short distances, $L \lesssim 300$, in the region where the grammatical rules are efficient. The other kind of correlations, the persistent one, plays the main role at long distances, $L \gtrsim 300$. It can be related to the general idea of a literary work. Therefore, the first kind of correlations may be referred to as the grammatical one, whereas the second kind may be called as semantic correlations. The additional investigations are necessary to clarify the nature of the correlations. For example, it would be of interest to elaborate the dynamical model, e.g., a cellular automaton that produces the sequence with the same enigmatic long-range behavior of the memory function as the one in the literary works.

If our understanding as to the nature of both kinds of correlations in the literary texts is correct, then several important questions would be of great interest, e.g., the following.

Does the lack of the antipersistent portion in the memory function [and in the $D(L)$ dependence [20]] in the DNA texts mean that the "grammatical rules" are nonexistent in the "DNA language"?

If we consider the variance $D(L)$ as a measure of information redundancy, can we explain the equality $D(L)_{|DNA} \simeq 10 \times D(L)_{|Text}$ resulting from the comparison between literary and DNA texts at $L \sim 3 \times 10^5$ [20] in the following way: the Nature is more careful about the conservation of the information stored in the DNA sequences than the Writer in his literary works?

We have studied the simplest examples of random sequences, the dichotomic one. However, our preliminary considerations suggest that the formulated concept of the additive Markov chains can by generalized to a larger class of random Markov processes with the finite or infinite number of states in the discrete or continuous "time." The suggested approach can be used for the analysis of other correlated systems in different fields of science.

[1] U. Balucani, M. H. Lee, and V. Tognetti, Phys. Rep. **373**, 409 (2003).
[2] I. M. Sokolov, Phys. Rev. Lett. **90**, 080601 (2003).
[3] A. Bunde, S. Havlin, E. Koscienly-Bunde, and H.-J. Schellenhuber, Physica A **302**, 255 (2001).
[4] H. N. Yang, Y.-P. Zhao, A. Chan, T.-M. Lu, and G. C. Wang, Phys. Rev. B **56**, 4224 (1997).
[5] S. N. Majumdar, A. J. Bray, S. J. Cornell, and C. Sire, Phys. Rev. Lett. **77**, 3704 (1996).
[6] S. Havlin, R. Blumberg Selinger, M. Schwartz, H. E. Stanley, and A. Bunde, Phys. Rev. Lett. **61**, 1438 (1988).
[7] R. F. Voss, Phys. Rev. Lett. **68**, 3805 (1992).
[8] H. E. Stanley *et al.*, Physica A **224**, 302 (1996).
[9] S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, M. E. Matsa, C.-K. Peng, M. Simons, and H. E. Stanley, Phys. Rev. E **51**, 5084 (1995).
[10] A. Provata and Y. Almirantis, Physica A **247**, 482 (1997).
[11] R. M. Yulmetyev, N. Emelyanova, P. Hänggi, F. Gafarov, and A. Prohorov, Physica A **316**, 671 (2002).
[12] B. Hao and J. Qi, Mod. Phys. Lett. A **17**, 1 (2003).
[13] R. N. Mantegna and H. E. Stanley, Nature (London) **376**, 46 (1995).
[14] Y. C. Zhang, Europhys. News **29**, 51 (1998).
[15] A. Czirok, R. N. Mantegna, S. Havlin, and H. E. Stanley, Phys. Rev. E **52**, 446 (1995).
[16] A. Schenkel, J. Zhang, and Y. C. Zhang, Fractals **1**, 47 (1993).
[17] I. Kanter and D. A. Kessler, Phys. Rev. Lett. **74**, 4559 (1995).
[18] P. Kokol and V. Podgorelec, Complexity International **7**, 1 (2000).
[19] W. Ebeling, A. Neiman, and T. Poschel, cond-mat/0204076.
[20] O. V. Usatenko, V. A. Yampol'skii, K. E. Kechedzhy, and S. S. Mel'nyk, Phys. Rev. E **68**, 061107 (2003).

[21] H. A. Makse, S. Havlin, M. Schwartz, and H. E. Stanley, Phys. Rev. E **53**, 5445 (1995).
[22] W. Li, Europhys. Lett. **10**, 395 (1989).
[23] R. F. Voss, in *Fundamental Algorithms in Computer Graphics*, edited by R. A. Earnshaw (Springer, Berlin, 1985) p. 805.
[24] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, Nature (London) **363**, 31 (1993).
[25] O. V. Usatenko and V. A. Yampol'skii, Phys. Rev. Lett. **90**, 110601 (2003).
[26] S. Hod and U. Keshet, Phys. Rev. E **70**, 015104(R) (2004).
[27] S. L. Narasimhan, J. A. Nathan, and K. P. N. Murthy, Europhys. Lett. **69**, 22 (2005).
[28] S. L. Narasimhan, J. A. Nathan, P. S. R. Krishna, and K. P. N. Murthy, cond-mat/0409053.
[29] S. S. Melnyk, O. V. Usatenko, and V. A. Yampol'skii, physics/0412169.
[30] F. M. Izrailev, A. A. Krokhin, and S. E. Ulloa, Phys. Rev. B **63**, 041102(R) (2001).
[31] The Old Testament of the King James Version of the Bible, http://www.writersbbs.com/bible/.
[32] At length L of the order of the full text length $\mathcal{M} \sim 10^6$, we note the violation of the monotonous growth in the variance $D(L)$. It is a manifestation of the border effect. Obviously, the number $k(L)$ of unities in any L word tends to $\bar{k}$ and the variance $D(L)$ tends to zero [see definition Eq. (3)] at $L \to \mathcal{M}$.
[33] Note that the antipersistent regions in Figs. 1 and 2 differ from each other; this distinction results from the difference in the coarse-graining procedures used in these two cases.
[34] Russian Synodal LiO 31/7/91, http: //www.lib.ru/ hristian/bibliya/nowyj_zawet.txt.
[35] http://www.lib.ru.