The pre-meaning of state sequences

Dr. C. S. Chassapis[1,2]

***Abstract* — A generalized information measure, called degree of pre-meaning, is discussed here, with primary objective to understand the most about state sequences. Computational experiments over selected long archetypal strings have been made and analytical expressions are presented. Learning, optimal alphabets and fusion of states have been also discussed. Laws of learning have been unveiled.**

***Index terms* — entropy, information, texts, strings, learning, alphabets.**

I. INTRODUCTION

A generalized information measure, called *degree of pre-meaning*, is discussed here, with primary objective to understand the most about state sequences. Our approach adheres to the framework set by previous classic works, such as [1, 2, 3]. We name *state sequence* any linear chain (string) of states (for example: English or formal texts, time-series, DNA-sequences, etc.), *1-state* the element (atom) of which the state-sequence is made of, *s-state* (or *s-word*) a combination of s consecutive 1-states. These 1-states may represent numbers, words, or other concepts. Also, importantly, any s-state can be considered a 1-state in a redesigned analysis. We use the term *state space* or *alphabet* for the set of 1-states. When we discuss not an actualized state-sequence but all potential ones given a state space, we use the term *syntactic space*. Also, inversely, an actualized syntactic space *is* a given sequence.

II. DEFINITION OF THE DEGREE OF PRE-MEANING

We recognize that the meaning of a state increases with the number of states, N, of the space that we know that this state belongs to. Let $m_1(N) := \log_2(N)$ be the *per-state minimal meaning*, or *m-meaning*, of a state-space of N 1-states. Also let $M_1(N) := N\, m_1(N)$ be the *total m-meaning*. In general, for s-states, $m_s(N) := s \log_2(N)$, and $M_s(N) := N^s\, s \log_2(N)$. Since $M_s(N)$ is the total number of bits needed to articulate and store *a priori* all s-words, it can also be interpreted as a volume of bits. The *total minimal meaning of all at-most K-state long combinations of the 1-states*, $M_*(N,K)$, is the sum of all $M_s(N)$ up to s=K. The quantity K can be, for example, an appropriately defined limit of a detection- or cognitive- system. Following analogous reasoning and using entropies as in [1, 2, 3], a definition of the *total information* can be $I_1(N, L) := NH(1, L)$, where H(s, L) is the entropy from the s-words of an actualized state-sequence of length L. This can be interpreted as the bit-volume needed to describe and store an actualized state-space. For the total information related to the s-words, we have $I_s(N, L) := N^s\, H(s, L)$. Adding all $I_s(N, L)$ we obtain the *total information of all at-most K-state long combinations of the 1-states*, $I_*(N,K,L)$. We define the *degree of pre-meaning* to be the *ratio* of two measures of information (a ratio of bit volumes):

$$\mu(N,K,L) := M_*(N,K) / I_*(N,K,L) \tag{1}$$

By forming this ratio we give to μ a pure sense free from units. μ can be also seen as the inverse of

---

[1] Algosystems S.A., 206 Sygrou Ave, 17672 Kallithea, Greece (preferred mail address), Email: chassapis@algosystems.gr.

[2] Research Associate, CAG, Division of Applied Technologies, Demokritos NRC 15310 Aghia Paraskevi, Athens, Greece.

the percentage of the M_* bit-volume actualized, or, more loosely speaking, as the *ratio of "definition" to "surprise"*. Obviously $\mu(N,K) \geq 1$.

Even if storage capacities and bit volumes are not enough to attack what we have aggressively named pre-meaning, understanding and memory are closely related, and, as E. Alvarez-Lacalle et al. say in [5]: *"Understanding texts requires memory: the reader has to keep in mind enough words to create meaning. This calls for a relation between the memory of the reader and the structure of the text".* We believe that μ captures a lot of the interesting aspects of understanding, and this is what we will try to show in what follows. Also, another interesting aspect of our approach is that one can define extrema of the degree of pre-meaning, even some kind of an average degree of pre-meaning related to a specific syntactic space, investigating therefore the capacity of a particular space as a carrier of meaning. This can even be connected to cultural ways of representing "reality".

III. COMPUTATIONAL EXPERIMENTS

In the next pages, while discussing various phenomena we will corroborate our claims using computational experiments over several archetypal strings, which are defined as follows:

SYMDYN: Aperiodic symbolic-dynamics-oriented sequence so constructed to have very-long-range correlations. The sequence is build by what we may call the "rabbit law". We start by the binary 0. Then we iterate several times the productions (rules): 0→1, 1→10. The resulted binary text of bits is deflated by block-grouping three bits to form a sequence of integers. The number of consecutive bits in the group is so selected (experimentally) to produce only 4 different integer codes. The autocorrelation function (a.c.f.) of such a state sequence shows very strong harmonic-like behavior with very slow amplitude decrease. Here L = 600 000 000.

PERIODIC: Perfect saw-tooth-like periodic sequence, with 4 different numerical codes recorded. The a.c.f of such a state sequence is a slowly decreasing periodic function. Here L = 220 000 000.

DNA: Homo sapiens Chromosome 1 DNA sequence. The sequence contains fuzzy regions named N (in addition to the traditional A, C, T, G letters) and few artifacts like, carriage return, and space. The raw "fasta"-formatted sequence contains 7 different codes (characters). Here L = 222 827 847 and N = 4 because we have cut all artifacts and concatenated the rest in one large string.

TEXTS: The texts have been copied from Project Gutenberg web site (http://www.gutenberg.org) and have been cleared from any translators' comments, punctuation marks, redundant spaces, or any other metadata like line numbers. Only 63 characters are recognized, (normal and capital English letters, numbers and space).

- *Whitman:* Whitman, Walt, *Poems by Walt Whitman*, EText-No. 8388. Release Date 2005-06-01. L = 275801, N = 63. WRD: L = 51314, N = 8328.
- *Tolstoy:* Tolstoy, Leo, *War and Peace*, EText-No. 2600. Release Date 2001-04-01. L = 3026626, N = 63. WRD: L = 572593, N = 19106.
- *Byron:* George Gordon Byron, *The Works of Lord Byron Vol. 4*, EText-No. 20158. Release Date 2006-12-22. L = 478105, N = 63. WRD: L = 94442, N = 11390.
- *Kant:* Kant, Immanuel, *The Critique of Pure Reason*, EText-No. 4280. Release Date 2003-07-01. L = 1212043, N = 63. WRD: L = 209756, N = 7176.

The word analysis (WRD) happens as follows: all words (i.e. strings of characters between spaces or other punctuation marks, without the punctuation marks) are treated as symbols and are put in the symbol table for a normal pre-meaning analysis. Obviously this way N is greatly increased and L decreased, but universally μ is also increased. As can be seen in figure 5, μ can clearly differentiate poetry from text in both variations of the analysis, i.e. character based or word based. More details can be discerned inside the families of curves.

The probabilities of the s-words are calculated by the programs for s-words that may overlap.

For any periodic sequence we have $\mu = K \; N^K/(N^K\text{-}1) \; - \; 1/(N\text{-}1)$. This is also observed by our programs to a remarkable accuracy (see figure 1). This theoretical result also serves as a numerical crosscheck.

Computational analysis of the SYMDYN sequence that is also gradually shuffled is displayed in figure 2. Shuffling is the process of randomly selecting pairs of positions in the sequence (until the required percentage if satisfied) and then exchanging the characters in these positions. This process does not alter $\mu(1)$. Obviously the percentage of the altered positions is smaller than the shuffling percentage due to the possibility of exchanging identical characters, and the (smaller) probability by chains of replacements to return to the initial position the same character that was initially there. In figure 3 the same degradation of μ is investigated for DNA. Probably there are biological reasons for the distance between curve K=16 and the rest. This distance signals a resistance to destruction of the 16-words of the DNA. Figure 4 shows a comparison between the full chromosome 1 sequence (circles) with three other, smaller sequences (10 million characters approximately each), taken at different starting points within the long one. The clear differentiation of the smaller sub-sequences for small words (small K) clearly shows a region of what can be called different biological “meaning”.

μ can also be used in fundamental studies of the mechanism of construction of meaning, and in cryptanalysis studies, where the μ of a text is measured and compared to supposed similar ones to unveil the possibility of a hidden message. A variety of equally important themes is explored next.

IV. FUZINESS

An analysis based on μ has advantages, compared for example to simple block-entropy analyses, because it can directly capture state fusion and de-fusion phenomena and this ability is discussed later and put in use to determine optimal alphabets. Since the m-meaning of a state is synonymous to the unambiguous determination of it among the set of states that it belongs to, it seems that, the fuzzier the determination of a state becomes, the less degree of pre-meaning is conveyed, because, due to fusion of states, M_* is lowered. Things, nevertheless, are not so simple, because, in many cases, fuzziness, by emphasizing the important aspects of a normality or a symmetry, can effectively increase μ because I_* is lowered. This phenomenon is captured by μ.

V. DISORDER

μ as is defined in (1) has an interesting alternative interpretation. We can write (1) in the form

$$\mu(N,K) = 1 / \langle D(N,K) \rangle \qquad (2)$$

the inverse of an appropriately averaged disorder. This average is weighted by the percentage of the m-meaning related only to s-states, $w_s(N,K) := M_s(N) / M_*(N,K)$, and:

$$\langle D(N,K) \rangle := \sum_{s=1}^{K} w_s(N,K)\, D_s(N),$$

with $D_s(N)$ the *s-disorder*, our generalization of the simple disorder defined in [4] that is related to our s-states

$$D_s(N) := -\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\cdots\sum_{i_s=1}^{N} p(i_1, i_2, ..., i_s) \log_{N^s}(p(i_1, i_2, ..., i_s)) = \frac{H(s)}{s\log_2 N}.$$

We omit L, but since we are using s-word occurrence counting to approximate probabilities, L is implied on all D and H quantities. Maybe a better approach could have been to use the s-th order *empirical entropy* of Manzini [6], often used in discussions of compression algorithms, which applies to finite texts, accepting that the classical s- block entropy that we are now using can only, strictly speaking, be defined for infinite sources. The empirical entropy coincides with the statistical estimation of the entropy taking the text as a finite sample of the infinite source, and is not making any assumptions on the distribution of the text internals. This approach has advantages and we plan to use it in the future, but for now that we are trying to understand the potential of the degree of pre-meaning, it is better to link it with well understood notions. The empirical entropy is calculated using the occurrences of substrings consisting of concatenated selected ends of the s-words and the meaning of the process is not easy to grasp.

VI. INFORMATION PROCESSING

To understand the mechanics of μ we study few information processing activities. Two important macroscopic "information processing" parameters of our approach are N and K, (a third one, equally important, is L), and at this point, these questions arise naturally: Q1) Are there any relations that μ satisfies in relation to N and K? Q2) Is it possible to gradually determine μ? Q3) Which might be the relations that μ satisfies with respect to special probabilistic substructures? We already know that $\mu = 1$ for the random case. Can we have in particular μ = constant with respect to N or K, more generally? After some algebraic manipulation the following relation is derived:

$$\langle D(N,K) \rangle = \big(1-w_K(N,K)\big) \langle D(N,K-1) \rangle + w_K(N,K)\, D_K(N),\ K = 2,3, ...,\ \&\ \langle D(N,1) \rangle = D_1(N). \quad (3)$$

This answers Q2. From (3) we answer Q3: put $\langle D(N,K) \rangle = \langle D(N,K-1) \rangle$ and observer that, *if* $D_K(N)\,\mu(N,K-1) = 1$ then, *μ remains unchanged* after introducing one-more-state-long words of the 1-states. Eq. (3) also gives

$$\frac{\mu(N,K) - \mu(N,K-1)}{\mu(N,K)} = w_K(N,K) \big(1 - D_K(N)\, \mu(N,K-1)\big). \quad (4)$$

This means, more generally, that, whether we are going to see a relative increase or a decrease to μ after the introduction of one-more-state-long combinations of the 1-states, is controlled by $(D_K(N)\,\mu(N,K-1))$ being smaller or larger than 1. Equation (4) is also half answer to Q1. The other half of Q1 is answered soon.

VII. LEARNING

We investigate when an enlargement of the state-space does not contribute to an increase of μ. To start, we observe that, even if there is a non-zero probability of appearance of one new state, the probabilities of manifestation of this state in conjunction with previous states (to form s-words), is, at the first stages of the process of construction of meaning, zero (this is the situation where a cognitive system may or may not produce any higher level association, at time t, while being fed with (random or not) s-states for all s). To explore this situation we put $p(i_1=N+1) > 0$ & $p(i_1,i_2, ..., i_r =N+1, ..., i_s) \approx 0$, for all r, s, with r, s = 1, 2, ..., K. We assume that $N \gtrsim K$ (true, for example, for most words of the majority of occidental languages, interpreting K as a maximum word-length). Then we can use the approximation $w_s(N+1,K) \approx w_s(N,K)$, for $N \gtrsim K$. Also, for N>>1, we have $\log_{(N+1)^s}(p) \approx \log_{N^s}(p)$. These approximations, lead to $D_1(N+1) \neq D_1(N)$ and $D_s(N+1) \approx D_s(N)$ for s = 2, 3, ..., K. So, here $\langle D(N+1,K) \rangle \approx \langle D(N,K) \rangle + w_1(N,K) \big(D_1(N+1) - D_1(N)\big)$, thus, we find:

$$\frac{\mu(N+1,K) - \mu(N,K)}{\mu(N+1,K)} = - w_1(N,K)\, \mu(N,K) \big(D_1(N+1) - D_1(N)\big) \quad (5)$$

Equation (5) is an important result of this work because we see that, *only if the 1-disorder clearly drops with N*:

$$\Delta_N D_1(N) := \Big(D_1(N+1) - D_1(N) \Big) < 0 \qquad (6)$$

and if, at the same time, the established degree of pre-meaning is of significant value:

$$w_1(N,K)\ \mu(N,K) > 1 \qquad (7)$$

we can see a significant relative increase of μ with N at the first stages of creation of meaning. We call equations (6), (7) "*laws of learning*". These laws do not say what can be learned but how to learn efficiently.

One can generalize our "laws of learning" to more complex set-ups, like for example, in a theory of scientific understanding. We will not pierce in such a theory here, but we will note the similarities of our laws of learning to the basic ingredients of an outline of a theory of scientific understanding presented in [7]. To quote from these authors: *"The basic theory of scientific understanding exploits three main ideas: First, that to understand a phenomenon P, a given agent has to be able to fit P into its cognitive background corpus C. Second, that to fit P into C is to connect P with parts of C, such as the unification of C increases. Third, that the cognitive changes involved in unification can be treated as sequences of shifts of phenomena in C".* We keep for our purposes the key importance of the cognitive background corpus *C*, that we can obviously link to μ, and that the changes in *C* must increase the unification of *C* and must be treatable within the context of *C*, and we take the liberty to link that to a lowering of a disorder in *C*.

VIII. EXCISION OF A SMALL STATE-SUBSPACE

Let us investigate what happens when a small subset, S, of the state-space A *is cut* from the rest. To keep algebraic manipulation to reasonable margins, we fix K=2 (this choice is not compromising generality because the 1-states that combine to form the 2-state can be of arbitrary "internal" complexity). Let us assume that the set A has N elements, and that the set S, (subset of A) has n elements. Then, we put p(i,j) = 0 if i XOR j $\in$S. So,

$$D_2^{(A)}(N) = \frac{\log_2(N-n)}{\log_2(N)} D_2^{(A-S)}(N-n) + \frac{\log_2(n)}{\log_2(N)} D_2^{(S)}(n)$$

$$D_1^{(A)}(N) = \frac{\log_2(N-n)}{\log_2(N)} D_1^{(A-S)}(N-n) + \frac{\log_2(n)}{\log_2(N)} D_1^{(S)}(n)$$

Now,

$$\langle D^{(A)}(N,2)\rangle = w_1(N,2)\, D_1^{(A)}(N) + w_2(N,2)\, D_2^{(A)}(N)$$

$$\langle D^{(S)}(n,2)\rangle = w_1(n,2)\, D_1^{(S)}(n) + w_2(n,2)\, D_2^{(S)}(n)$$

$$\langle D^{(A-S)}(N-n,2)\rangle = w_1(N-n,2)\, D_1^{(A-S)}(N-n) + w_2(N-n,2)\, D_2^{(A-S)}(N-n)$$

From these, we get after some algebra

$$\frac{\log_2(N)}{w_2(N,2)}\langle D^{(A)}(N,2)\rangle = \frac{\log_2(N-n)}{w_2(N-n,2)}\langle D^{(A-S)}(N-n,2)\rangle + \frac{\log_2(n)}{w_2(n,2)}\langle D^{(S)}(n,2)\rangle +$$

$$- \frac{n}{2N}\frac{\log_2(N-n)}{N-n} D_1^{(A-S)}(N-n) - \frac{N-n}{2N}\frac{\log_2(n)}{n} D_1^{(S)}(n) \qquad (8)$$

To understand eq. (8) we will study a special case with 1 << n << N. Then eq. (8) becomes

$$\frac{\log_2(N)}{w_2(N,2)}\langle D^{(A)}(N,2)\rangle \approx \frac{\log_2(N-n)}{w_2(N-n,2)}\langle D^{(A-S)}(N-n,2)\rangle + \frac{\log_2(n)}{w_2(n,2)}\langle D^{(S)}(n,2)\rangle - \frac{\log_2(n)}{2n} D_1^{(S)}(n)$$

or

$$\langle D^{(A)}(N,2)\rangle \approx \frac{\log_2(N-n)}{\log_2(N)}\frac{w_2(N,2)}{w_2(N-n,2)}\langle D^{(A-S)}(N-n,2)\rangle + \frac{\log_2(n)}{\log_2(N)} w_2(N,2)\, D_2^{(S)}(n),$$

or, since $\frac{\log_2(N-n)}{\log_2(N)} \frac{w_2(N,2)}{w_2(N-n,2)} = 1 + O\left(\frac{n}{N\log_2(N)}\right)$, and $w_2(N,2) = \frac{2N}{1+2N} = 1 + O\left(\frac{1}{2N}\right)$,

$$\langle D^{(A)}(N,2) \rangle \approx \langle D^{(A-S)}(N-n,2)\rangle + \frac{\log_2(n)}{\log_2(N)} D_2^{(S)}(n).$$

So

$$\frac{\mu^{(A)}(N,2) - \mu^{(A-S)}(N-n,2)}{\mu^{(A)}(N,2)} \approx - \mu^{(A-S)}(N-n,2)) \frac{\log_2(n)}{\log_2(N)} D_2^{(S)}(n) \qquad (9)$$

The r.h.s. of (9) is always negative. This is a remarkable result. It says that $\mu^{(A)}(N,2) < \mu^{(A-S)}(N-n,2)$, or that *in the first stages of construction of meaning, when K is small, once a space of a large alphabet is altered by an appropriate excision of a subspace of adequate extent, then the degree of pre-meaning rises!* This reminds to us a familiar learning methodology where one simplifies a conceptual space by ignoring some concepts at start.

IX. OPTIMAL ALPHABETS

The probabilities of the s-states do not tell whether a state is a member of the alphabet or is an s-state. So given a state-sequence of large length, L, without knowing the alphabet and N, and having measured the probabilities p_a, a=1, 2, ..., Q, of *all* actualized s-states with s=1, 2, ..., K, with 1 « K « L, then we may (temporarily) assume that these Q states constitute the alphabet and:

$$\mu(Q,1) = \frac{1}{\langle D(Q,1) \rangle} = \frac{1}{w_1(Q,1)D_1(Q)} = \frac{1}{D_1(Q)}. \qquad (10).$$

Obviously if we know the "true" N-member alphabet, then the same probabilities of the states, (with a *renaming* of the states) give, as we know from equation (1), μ(N,K), where the probabilities of those states that never appear a-posteriori are zero (but because of the alphabet they could have a-priori appeared), since in general

$$Q \leq \sum_{s=1}^{K} N^S \qquad (11)$$

Based on the fact that in (10) and (1) we have the *same* states and that (11) also holds, we show the theorem*:*

Always μ(N,K) ≥ μ(Q,1), except for N=2 at K=2, where, the inequality may fail. (12)

[*Proof:* $\mu(N,K) \geq \mu(Q,1) \Leftrightarrow \sum_{s=1}^{K} w_s(N,K)\, D_s(N) \leq D_1(Q) \Leftrightarrow$

$$-\sum_{s=1}^{K} \frac{w_s(N,K) \log_2 Q}{s \log_2 N} \sum_{i_1=1}^{N} \dots \sum_{i_s=1}^{N} p(i_1,..., i_s) \log_2\Big(p(i_1, ..., i_s)\Big) \leq -\sum_a p(a) \log_2(p(a)) =$$

$$-\sum_{s=1}^{K} \sum_{i_1=1}^{N} \dots \sum_{i_s=1}^{N} p(i_1,..., i_s) \log_2\Big(p(i_1, ..., i_s)\Big).$$

In the last equation, we just renamed the states under summation. The states do not change. Now,

$$\frac{w_s(N,K) \log_2 Q}{s \log_2 N} = \frac{s N^S \log_2 Q}{\sum_{r=1}^{K} r N^r \; s \log_2 N} \leq \frac{N^K \log_2\left(\sum_{r=1}^{K} N^r\right)}{\sum_{r=1}^{K} r N^r \; \log_2 N} = \frac{N^K \log_2\left(\frac{N^{K+1} - N}{N-1}\right)(N-1)^2}{\Big(N^{K+2} K - N^{K+1} K - N^{K+1} + N\Big) \log_2 N}$$

which is always ≤ 1 except on the N=K=2 case where it jumps to $\frac{2}{5}\frac{\ln(6)}{\ln(2)} \approx 1.034$ if Q=6. So inequality (12) is true under the stated conditions.

Please note that the anomaly at N=K=2, does no essential harm to the usability of the theorem, because we will usually have Q>>1 and we will use the theorem to visit large N cases, smaller or much smaller than Q of course, while seeking for the "best" alphabet.**]**

Eq. (12) provides, first, a means for computational discovery of the unknown alphabet, and, second, reveals to us *the power of the proper naming of things*. There are two ways to perform the search for an optimum alphabet in a given text. Either start from a very fine stratification homogeneously *fuse states*, compute μ, compare and reiterate, or, start from a very coarse-grained sequence and iteratively homogeneously *defuse states* and compute μ. Always in a way to maximize μ. We present an algorithm based on a systematic fusion of states, starting from a very fine stratification.

▼**Algorithm:**

0) Given an initial encoded text of 1-state-length equal to L, ...
1) Stratify extensively to a given N_0 (i.e. use all available codes)
2) Calculate $\mu(N_0, K)$ for several K's
3) Store the set of all $\mu_0(N_0, K)$
4) Select randomly a 2-state
5) Perform a fusion of this 2-state to an 1-state that is decided at random to be the first or the second 1-state, but once decided then is homogeneously applied to the whole text.
6) Calculate $\mu_1(N_1, K)$ for several K's
7) Store the set of all $\mu_1(N_1, K)$
8) If $\mu_1(N_1, K) \gtrsim \mu_0(N_0, K)$ for several corresponding K's, then

    8a) Set $\mu_0(N_0, K) = \mu_1(N_1, K)$ for all K's in the calculation

    8b) Set $N_0 = N_1$

   Endif

9) Go to step 3, UNTIL satisfied. ▲

What resides in the symbol-table at end, is the optimum alphabet. A 1-state of this symbol-table may consist of one or blocks of the original 1-states. We can calculate the transition matrix p(i|j) for these 1-states and use this p(i|j), for time-series prediction of a future 1-state.

X. REFERENCES

[1] C. E. Shannon, Bell Syst. Techn. J., 27 (1948) 379.

[2] C. E. Shannon, Bell Syst. Techn. J., 27 (1948) 623.

[3] A. I. Khinchin, "*Mathematical Foundations of Information Theory*", Dover, N.Y., 1957.

[4] P. T. Landsberg, "*Thermodynamics and Statistical Mechanics*", Dover, N.Y., 1990. (Page 366).

[5] E. Alvarez-Lacalle, B. Dorow, J.-P. Eckmann, E. Moses, *A quantitative analysis of concepts and semantic structure in written language: Long range correlations in dynamics of texts*, (Dated: May 29, 2006), arXiv:physics/0510276.

[6] G. Manzini, *An analysis of the Burrows-Wheeler Transform*, J. Assoc. Comp. Mach. 48, 3, (2001) pp 407-430.

[7] G. Schurz, K.Lambert, Synthèse, 101 (1994), pp 65-120.

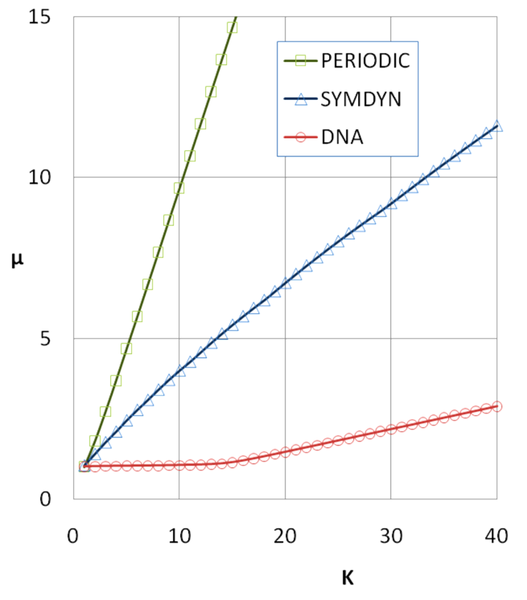

Figure 1: μ versus k. Here N=4 for all sequences.

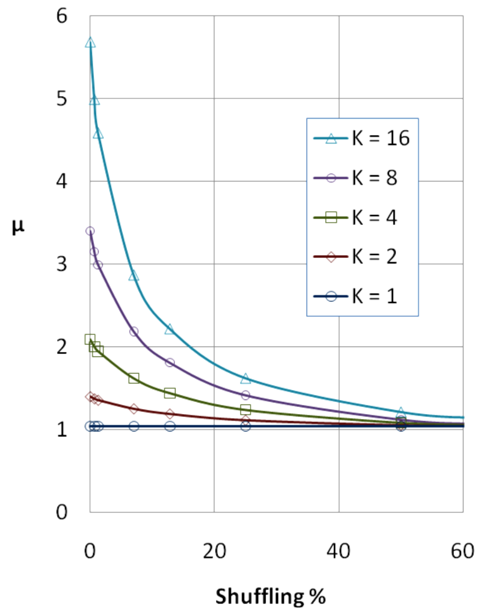

Figure 2: μ versus the percentage of shuffling for SYMDYN.

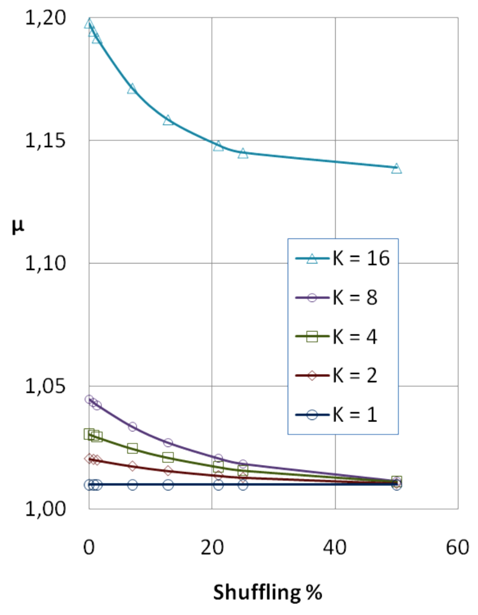

Figure 3: μ versus the percentage of shuffling for DNA

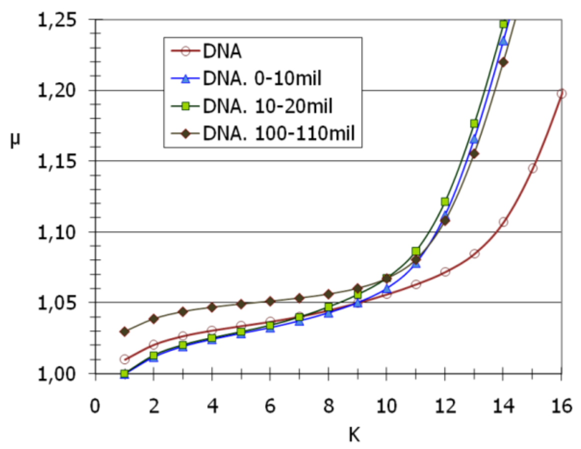

Figure 4: μ versus K for DNA for various sub-sequences starting at various positions inside the long sequence

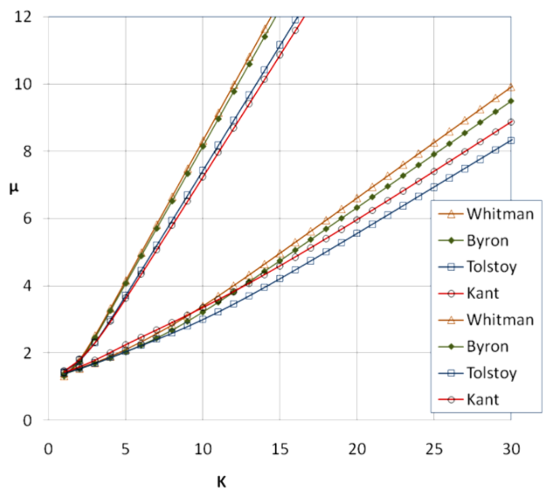

Figure 5: μ versus k. Here N=63 for all sequences. The upper family of curves comes from WRD analysis.