

---

# DESCRIPTIVE STATISTICS, ENTROPY AND INTERNET

**Professor habil. Gheorghe Săvoiu, Ph.D.**

## **Abstract**

*An investigation of the evolution of language in general and of that archived on the Internet is not an isolating (mono)disciplinary problem of anthropological, etymological, linguistic, historical, physical, statistical, cybernetic or computer science type, but certainly requires an inter-, trans-, and multidisciplinary approach, including much more sciences in relation to even those mentioned before. This article, due to the small dimensions characteristic of the specific genre and the strictly statistical topics of the review in which it is published, remains an interdisciplinary one, an interdisciplinary way of thinking of the II<sup>nd</sup> level, starting from the introductory approach of the research, and also from some defining landmarks of language, Internet and entropy, developing the entropy from thermodynamics and placing it universally in the words and networks defined as www. The paper evolves structurally with the help of a survey theory in the direction of an investigation that infers, estimates and provides some approximate parameters and with a certain level of the error limit of Internet communication and especially the evolution of entropy in the universe of networks. As natural as possible, the investigation gives rise to descriptive statistics, matrices and modeling, along with comments and final remarks whose real usefulness can be assessed by the readers of the article in its entirety.*

**Key words:** *language, Internet, survey, random sample, average representativeness error, limit error, descriptive statistics, entropy, thermodynamic entropy, statistical entropy, informational entropy, economic entropy, correlation matrix, econometric model.*

**JEL classification:** C10, C13

## **Introduction**

Language, originally understood as exclusively oral transmission of information, has ancestral echoes in the evolution of the human condition and interpersonal communication, defined as the defining formative condition of man and his extended society under the name of population. Proto-language as the ancestor of language in the vision of modern anthropology seems to be related to Homo Ergaster, the hominid who lived in Africa almost two million years ago (1.4-1.9 million years ago), first resorted to symbolic thinking and willing to act figuratively or imaginatively to project his thoughts in the long

---

run (Carrizo, 2018). The scientific argument for such a long evolution from protolanguage to modern language is given by the size of Homo Ergaster's skull, which already reaching 850 cm<sup>3</sup> allowed the appearance of the speech center with the Broca and Wernicke areas able to support a form of real linguistic or symbolic communication. . A moderate theory focused on genetic evidence, respectively on the relatively recent discovery of a language-related gene and whose spread lasted between 10 and 20 thousand years, places the emergence of language as the engine of human evolution almost 200 thousand years ago. However, the theory cannot specify qualitative aspects related to the fact if the language was a primitive one of mimesis type (imitative) or a more evolved one (personalized). The maximum reductionist variant links language and communication not with the appearance of Homo Sapiens Sapiens (about 50 thousand years ago) but with its maturation through the development of a complex social structure of human communities, implicitly to intra-community communication.

Modern language has become more than “*a method of human communication, spoken or written, consisting in the use of words in a structured and conventional way*” (Oxford English Dictionary, 2020), constantly multiplying in increasingly varied scientific conceptualizations from psychological ones as *a non-verbal method of expression* to computer ones as *a system of symbols and rules for writing programs or algorithms*, detailing both linguistically and geographically as *a communication system used by a particular country or community*, or structurally as *phraseologies and vocabularies specific to certain professions, fields or social groups*, etc.

Internet is a relatively recent term introduced into general and scientific language. This concept can be replaced in modern language in a limited way, with the acronym WorldWideWeb (www or W3), and its meaning is related to direct access to information through search engines, but especially for storing or archiving information (Săvoiu, Cudanov, 2020). The Internet simultaneously means both “an initiative to access and retrieve extensive information or hypermedia” (Greene, 2000) and an offer that “aims to provide general information through quick access to a large universe of documents” (Mowbray, 2017). Although Nikola Tesla, in the early twentieth century, was actually the first scientist to believe that the “global wireless system” was feasible, imagining what this virtual world would look like, and after another six decades, Licklider became the one who expanded the idea of an intergalactic computer network, the modern Internet was officially born only in 1990, when it was practically invented by its creative designers Tim Berners Lee and Robert Cailliau. However, this would not have been possible if in 1972, ARPAnet (Advanced Research Projects Agency Network - created

---

and funded by the US Department of Defense) had not invented e-mail, and computer scientists Robert Kahn and Vint Cerf had not developed plans for the first Internet protocol. The Internet, suggestively described as “organized chaos”, was used first by physicists and later by social work professionals, especially those in academia and professional associations, from universities to economic companies, from scientific associations, to organizations dedicated to art and culture, etc. The real history of the Internet connects in a unique way fields of communication with linguistic fields, statistical fields with computer fields, academic sectors, with military sectors, engineering actions with online social interactions, educational theory with research practice.

The current use of Internet has long exceeded the intentions of its inventors, the meaning of the concept extending from the technical source „available through a telecommunications system that includes at least one computer” (Manovich, 2001) to broader meanings of connection, access, service with information or even at the maximum extent of availability of communication and information. Finally, all these multiple meanings are reduced to change and adaptation with an emphasis on language, along with communication and information, because the Internet facilitates and resizes better the volume and importance of information, but also the information system, language and even communication (Săvoiu, Cudanov, 2020).

The association between Internet and any other thermodynamic system naturally induces the need to identify statistical dimensions related to maximum, minimum, average, modal or median values, standard dispersion and deviation, asymmetry and vaulting, normality or evolutionary abnormality of language, etc. The spread and change of Internet language has at least an interesting investigative dynamic, which has hypothetical chances to identify anything other than heterogeneity and abnormality in the potentially detectable network of billions of words, in a continuous dilation caused by the increase in the number of web pages etc. The variable number of words has a size and a dynamic only seemingly predetermined and a dominant eclectic meaning being dependent on many factors such as language, quantified word dimensions, synonymy and homonymy, spatialization by regionalization along with timing given by the depth of the archaisms, as well as other causes not at all semantic or grammatical, but rather related to abbreviation and acronymization, etc.

The words’ volume of Internet could be analysed and interpreted both as flows and as statistical stocks following at least four types of distinct evaluations, focused on: i) maximum and minimum values, medians and averages of the approximations; ii) values of the statistical scattering or dispersion of the same approximations; iii) associations or major correlations of the values specific to the descriptive statistics of the approximations; iv)

---

estimates and modelling with statistical investigation tools, recognized and appropriate as the statistical survey and the econometric model validated partially or in full, associating statistical methods to investigate the variable pattern of a web page. Thus, the size of Internet starts from the appearance and evolution of a prototypical or standardizable meaning of the form and content of the web page and finally reaches their network. As a result, the research in this article statistically investigates the size of the Internet as a source, reservoir, corpus, system or statistical set using a sample of words taken randomly in Romanian and, finally, tries to specify or estimate the parameters of the www network, with an acceptable error level. In parallel, the paper follows the evolution of the mentions of a randomly word sample, using the Google Search engine considered by the author as the most important in terms of access dimensions. The chronology of the investigation starts from the hypothesis that a research can be considered irrelevant for short time intervals, and for this reason we try in parallel to estimate the entropy of the sample of selected words, starting from the simple hypothesis of its growth in any system exposed to the passage of time. A few final remarks close the article and this brief introductory presentation (reminder) reshapes linguistic reasoning with the help of statistical thinking, which together evolves in the direction of the interdisciplinarity of linguistic statistics.

### **Methodology, data, results and discussions**

#### **• About the statistical entropy or the entropy of statistically described systems and isolated assemblies**

In any more or less successful attempt to define or conceptualize, again and again, “the subject or the definite is equal to the predicate or definer” by definition (Săvoiu, 2009). Univocity as an unchanged meaning in time and space of the term thus becomes the desideratum of a precise definition in the case of entropy, and ambiguity and imprecision are expressions of imbalance and inequality of content materialized in the sentence or statement of any definition. “In essence, the necessary and sufficient criteria for the definition are related to the presence of synonymy or equivalence of meanings, but also to the absence of ambiguity. The well-defined subject or term finally becomes a linguistic synonym, a logical and non-contradictory equivalent or unambiguous, in relation to the predicate or definer” (Săvoiu, 2009). The simplest of the definitions of entropy, while retaining the specific precision of physics and a location in the www network or an Internet placement, is that of “measuring the degree of unavailability of a www system to produce orders” or thinking closer to the language of “the measure of the degree of disorder or chaos of language and the Internet as isolated systems”. The more orderly a

---

system is able to produce order, the lower its entropy, the limit being given in time by the fact that the system reaches its final threshold when it can no longer produce order and generates exclusively chaos, which records the maximum entropy of his. The major problem of such a definition is that the definer (lat. *definiens*) is meant to facilitate knowledge and specify the meaning of entropy, although the term entropy naturally offers consistently different meanings (Săvoiu, 2009). The scientific definition seen as an ensemble finally identifies a convention of the use of language and configures the concept of entropy as a measure of the lack of order and this makes its meaning more difficult to understand. The classics tried to define the essence of the phenomenon, by similarities (proximate gender) or by difference, but not by absence. Modern scientific theory and its language have given up virtually any restriction of definitions. For this reason, this article proposes that the entropy assimilated to the language and the Internet system to be defined linguistically and informally by disorder or chaos of the language system placed on www, using statistical (probabilistic) determinations. As authors or readers, senders and respondents of messages, people announce, initiate and respond to a written dialogue or simply communicate on the Internet, distribute information with words, permanently expanding the number of words used or the volume of common language. Information flows naturally from those who know more, to those who know less, as in thermodynamics and heat circulates from the hot body which thus loses it through contact with the colder body which heats up. The more intensely people are placed in information communication processes, structured in senders and receivers of messages, appreciate and comment, modify and redistribute messages (texts) on Internet, the more each has in the end more decrypted data for further discussion, amplifying information received in personal memory and the Internet, forming new communication groups. Through permanent or intermittent communication, the information available on the Internet thus redefined as a multidisciplinary universe of words or scientific language or as a multiverse of universes of (mono)disciplinary languages, develops in an ascending and exclusively continuous manner and thus has an increasing chance to become more tangled, more disordered, more chaotic, etc., confirming over time the increase in entropy, both in the case of language and network (www). Entropy was conceived and can be redefined as a quantitative reflection of the evolution of the degree of asymmetry or irreversibility of any information process. Entropy acquires an increasingly precise identity, reflecting an arrow of time and characterizing the behavior of any communication system, including the Internet. This multiplicative and dissipative dynamic of words and language, only in one sense, respectively in the sense of an expansion existing anywhere in the network and continuously,

---

describes an evolutionary thermodynamic inexorability: even the Internet goes through a natural process of maturation and finally aging. This requires adapting logically and to the Internet all the principles of thermodynamics (Săvoiu, 2021):

i) the principle of “equivalence between the forms of energy exchange between the Internet system, defined by the internal informational energy of one’s own language as a function of state and the language existing in the surrounding world” (the principle of non-existence of a type I *perpetuum mobile* of species);

ii) the principle that “spontaneous energy transformations take place from the highest potential to the lowest potential” (the principle of non-existence of a type II *perpetuum mobile*), in which case the entropy of the Internet as a state quantity is additive in subsystems or (sub)structured networks; greater than zero for irreversible transformations (the entropy of the Internet cannot tend to zero or even cannot decrease outside an energy transfer from outside the Internet system and the Arnold Sommerfeld alternative is practically impossible);

iii) the principle according to which “when the temperature (information) tends to absolute zero, the entropy of a system (Internet) becomes zero” or of the system hypothesis with unchangeable entropy known as the alternative Walther Nernst and Max Plank.

In the Internet case, the second principle of thermodynamics contains the highest applicability of all three basic principles, constituting by logical symmetry the most important principle for this article. Thus, the statement that in any isolated (closed) system, the thermodynamic entropy of the system grows, it transforms into the www network in the entropy of the Internet certainly grows according to the dynamics of the volume of language as Josiah Willard Gibbs (1902) pointed out: “*In any transformation of an isolated physical system, its entropy always increases thermodynamic equilibrium states. Entropy remains constant only in the limit case of reversible transformations*”. The effective measurement of the entropy of systems began with Rudolf Julius Emanuel Clausius, who extended the laws of thermodynamics to an infinite universe, concluding that “*the energy of the world remains constant, the entropy of the world tends to the maximum*”, to a moment of equilibrium, defined as “*thermal death*”. The concept of **absolute entropy** ( $S_{abs}$ ) is characterized by the value zero, according to the third law of thermodynamics, actually defining entropy at absolute zero, while the variation of absolute entropy is denoted by  $\Delta S$  and describes the distance between any  $S$  and  $S_{abs}$ . Ludwig Eduard Boltzmann relativized Clausian-type *thermal death* and issued the “*fluctuation hypothesis*”, establishing the statistical nature of entropy (1780).

---

**Statistical entropy** is an essential notion for the macro state of a macroscopic system, in relation to which we note the microstates that define a statistical set consisting of a large number of identical systems, each with the same probabilities  $w_i$  to be in microstate  $I$ , and the condition their standardization. Ludwig Eduard Boltzmann's **statistical entropy** is thus a measure of the number of possible microscopic states of an equilibrium system, in accordance with its macroscopic properties. In a system described by multiple variables, where a certain number of configurations ( $\Omega$ ) is identified, hypothetically just as statistically probable, the entropy ( $S$ ) is given by the natural logarithm of the number of configurations, multiplied by the Boltzmann constant ( $1,3806504 \times 10^{-23} \text{ J / K}$  or  $8.617343 \times 10^{-5} \text{ eV K}^{-1}$ ) :

$$S = -k_B \times \ln \Omega \quad (1)$$

For the entropy of a system (language or Internet) to decrease, energy from outside the system should be transferred to it. The first to treat entropy and information similarly was Léon Brillouin, followed by Josiah Willard Gibbs, in whom entropy becomes the maximum amount of information in the system:

$$S = -k_B \sum_{i=1}^n p_i \times \ln (p_i) \quad (2)$$

where  $k_B$  is the same Boltzmann constant,  $p_i$  is the probability that an energy  $E_i$  will occur during system fluctuations (the operator  $\Sigma$  is quantified for all possible states of the system)

**Informational entropy** is a measure of the *degree of uniformity or diversity of probability distribution* and was expounded by Claude Elwood Shannon and Warren Weaver in *The Mathematical Theory of Communication* (1949), transformed into a conditioned entropy of two events  $X$  and  $Y$  taking values  $x_i$  and  $y_j$ , when it loses not only its initial thermodynamic significance, but also the historical burden of the units of measurement (J/K), being redefined by the relation:

$$S(X|Y) = - \sum_{i=1}^n p(x_i|y_j) \times \log \frac{p(x_i|y_j)}{p(y_j)} \quad (3)$$

where  $p(x_i | y_j)$  is the conditional probability for  $X = x_i$  and  $Y = y_j$ . According to the Shannon relation, with the notation of entropy slightly modified from the original formula from  $H(X)$  to  $S$ , this can also be determined in the case of a discrete variable  $X$  with the values  $\{x_1, \dots, x_n\}$  and with the function of probability  $p(x_i)$ :

$$S = H(X) = - \sum_{i=1}^n p(x_i) \times \log_b p(x_i) \quad (4)$$

where  $b$  is a basis for logarithms, real and super unitary of type  $e$ , 2, 10 etc. (for  $b = 2$  the unit of measurement of information becoming the famous bit).

---

**Informational entropy** practically indicates the volume of information existing in a word / message in bits (minimum length) or in bits per symbol, constituting the optimal limit of a lossless compression applicable to a communicated/message (Shannon entropy in bits per symbol  $\times$  number of symbols from word/message). **Economic entropy** is described as defining for any economic process. This type of entropy “*dominates all material transformations in an economy, developing in an irreversible way in the direction of resource depletion*” (Georgescu-Roegen, 1971), constituting the theoretical (bio) economic variant of entropy able to describe the measure of continuous and irremediable degradation of a potential of existing resources (the appearance of disorganization or chaos).

$$S = \sum_{i=1}^n w_i \times \ln(f_i), \text{ cu } f_i = N_i/N \quad (5)$$

where  $w_i$  = weighting coefficient (aggregation).

A simplified variant of calculus in Octav Onicescu’s logic becomes informational energy:

$$E_i = \sum_{i=1}^n (N_i/N)^2 = \sum_{i=1}^n (f_i)^2, \text{ with } f_i = N_i/N \quad (6)$$

where a receiver expects a word/message  $N_i$  from  $N$  possible words/messages, with obviously different probability of occurrence (Săvoiu, 2013). Finally, the information entropy of Claude Elwood Shannon as a measure of the uncertainty associated with a random variable therefore determines the amount of information needed to represent all states of the variable was selected in this article as the most appropriate solution in investigating the language and Internet.

• **A representative sample for sizing the active Internet language**

This paper’s autor takes into account an obviously approximate value of the probable limit error ( $\Delta_{mean}$ ) and finally select an acceptable level of error corresponding to a significance threshold  $\alpha = 0.01$  and for a standard deviation of 2091593 mentions. Thus, the probable limit error ( $\Delta_{mean}$ ) of Google Internet searches (for  $\alpha = 0.01$  and  $z = 2.58$  is estimated as follows:

$$\Delta_{mean} = z \times \sigma_{mean} = z \times \sqrt{\frac{\sigma^2}{n-1}} = 525000 \text{ mentions} \quad (7)$$

and the volume of a randomly extracted sample repeated becomes equal, finally, with  $n = 106$  words:

$$n = [z \times \sigma]^2 : [\Delta_{mean}]^2 = \{[z \times \sigma] : [\Delta_{mean}]\}^2 = \{[2.58 \times 2091593] : [525000]\}^2 = (10.278686)^2 = 105.65 \text{ words} \quad (8)$$



The determinations for a simple non-repeated random survey (N-RRS) do not bring significant changes, because the finite population correction is very small in terms of dimensions in real conditions ( $n = 106$  words in the sample and  $N = 65920$  words in the explanatory dictionary - DEX, 2012), imposing changes only at the level of the fourth decimal:

$\sqrt{[(1 - \frac{n}{N})]} = 0,9991957$ . This fact does not influence the volume of the N-RRS type sample (for the significance threshold  $\alpha = 0.01$  and  $z = 2.58$ ):

$$n = \left[ \frac{z^2 \sigma^2}{(\Delta_{ma})^2 + \frac{z^2 \sigma^2}{N}} \right] = \{(2,58^2 \times 2091593^2) : [(525000)^2 + (2,58^2 \times 2091593^2) : (65290)]\} = 105,5 = 106 \text{ words (9)}$$

Ludwig Eduard Boltzmann's statistical entropy leads to a reasoning in which equilibrium corresponds to a physically dynamic internal condition, but statistically stabilized in the way particles (words in this article) occupy their positions in the available volume (language and Internet), without this meaning configurations of particles left unchanged in time (mentioning words with similar results). The occupation of space in time was most likely that for which the configurations of particles (words) were more distributed and the internal energy of the isolated system (language or the Internet) has become practically the sum of the kinetic energies of the constituents in conditions of chaotic motion (real informational communication, from those who know to those who do not know).

For real systems, in quantum mechanics and later in quantum physics, the number of particles is colossal ( $10^{23}$ ). If in a single drop of water there are about 10211021 molecules, and in the Romanian dictionary 65920 words and in the English language 600,000 words, extended by the investigations of Harvard University and Google already since 2010 to 1022000 words, then estimates of words placed on the Internet in Romanian of about 1 trillion, in English of almost 15 trillion and in the total language of the Internet of about 30 trillion words (Săvoiu, 2021), confirm the investigative similarity in the article according to which language in modern communication will soon approach the dimensions of a statistical ensemble specific to statistical physics ( $10^{23}$ ).

The analogy is justified also starting from the reasoning of Ludwig Eduard Boltzmann, in which the balance is nothing but the stationary probability of the distribution of microstars (mentions of words investigated with Google search), where chaos increases naturally (entropy) and agitated particles (words) occupy time all the positions (calculated probabilities of

the words) that are allowed to them. In the Boltzmannian ergodic hypothesis, the statistical distribution of microstars or the internal probabilistic structure depends only on the number of microstars that the system can include, which justifies taking a sample ( $n = 106$ ) in this article and quantifying the partial entropy. As the order of a physical ensemble always evolves towards disorder, so the communication focused on words and language, benefits from the assumed hypothesis of the existence of entropy, estimated quantitatively or inferred (Table no. 1) with the capitalization of natural logarithm ( $\ln(x_i)$  or  $\log_e(x_i)$ , where  $x_i$  represents the frequency or number of mentions, following Internet searches of the sampled word based on  $e$ . The mathematical value of the transcendental and irrational number  $e$  can be approximated by 20 decimal places at  $e = 2,71828182845904523536$ , although it has in reality an infinite number of decimals and quantifies a constant of the mathematician John Napier (1550-1617), who describes the exponential function with the property of having the value of the derivative  $F(x) = e^x$  always equal to 1, at the point  $x = 0$ . The inverse of this functions was translated by the same John Napier into natural logarithms ( $e$ -base logarithms) provides a natural fit for the calculation of entropy. In the author's opinion, the maximum abstracted character of these logarithms gives them a natural adequacy to the entropy calculation in this paper (Table no. 1).

### The random extracted 106 Romanian words on Internet and the entropy calculus

Tabel no. 1

| Page | Column 1 | Column 2 | Random extracted word | Number of letters | Google mentions on Internet |            | Determined values * per word in          |  |
|------|----------|----------|-----------------------|-------------------|-----------------------------|------------|--|--|
|      |          |          |                       |                   | 21.01.2021                  | 15.03.2021 | 21.01.2021<br>$P(x_i) \times \ln P(x_i)$ | 15.03.2021<br>$P(x_i) \times \ln P(x_i)$ |
| 50   | 4        |          | antispumant           | 11                | 13400                       | 15800      | -2.05465E-05                             | -2.37418E-05                             |
| 372  |          | 39       | exponent              | 8                 | 23700000                    | 24900000   | -0.016040637                             | -0.016594925                             |
| 488  | 50       |          | iepuroi               | 7                 | 33700                       | 33500      | -4.81131E-05                             | -4.74794E-05                             |
| 1109 |          | 60       | tâmpănă               | 7                 | 3                           | 3          | -6.57817E-09                             | -7.42759E-09                             |
| 251  |          | 44       | criticism             | 9                 | 237000000                   | 225000000  | -0.097901496                             | -0.093704606                             |
| 92   |          | 49       | bavură                | 6                 | 36800                       | 41300      | -5.21685E-05                             | -5.75525E-05                             |
| 1038 |          | 43       | spectacol             | 9                 | 5560000                     | 5390000    | -0.004686447                             | -0.004529025                             |
| 18   | 73       |          | afet                  | 4                 | 12200000                    | 13100000   | -0.00918511                              | -0.009686205                             |
| 672  | 20       |          | monosepal             | 9                 | 3890                        | 4200       | -6.51502E-06                             | -6.94311E-06                             |
| 759  | 11       |          | ovreică               | 7                 | 4490                        | 4670       | -7.4473E-06                              | -7.66382E-06                             |
| 277  | 12       |          | decât                 | 5                 | 53700000                    | 55200000   | -0.031314354                             | -0.031797899                             |
| 118  | 46       |          | bregmă                | 6                 | 481                         | 497        | -9.20948E-07                             | -9.4207E-07                              |
| 487  |          | 8        | idiotism              | 8                 | 141000                      | 120000     | -0.00017819                              | -0.000152686                             |
| 160  |          | 35       | căruțaș               | 7                 | 58700                       | 44200      | -8.0074E-05                              | -6.1253E-05                              |
| 861  |          | 50       | prăvălie              | 8                 | 142000                      | 153000     | -0.00017934                              | -0.000190453                             |
| 559  |          | 1        | învălu                | 7                 | 94500                       | 207000     | -0.000123757                             | -0.000250565                             |
| 1195 | 63       |          | veșteji               | 7                 | 36400                       | 37600      | -5.16472E-05                             | -5.27972E-05                             |
| 503  | 70       |          | individualizare       | 15                | 119000                      | 123000     | -0.0001527                               | -0.000156158                             |

|      |    |    |                 |    |            |            |              |              |
|------|----|----|-----------------|----|------------|------------|--------------|--------------|
| 676  |    | 37 | moșneguț        | 8  | 10300      | 13600      | -1.61032E-05 | -2.06676E-05 |
| 172  | 53 |    | cerdac          | 6  | 129000     | 120000     | -0.00016434  | -0.000152686 |
| 870  |    | 14 | prestigiu       | 9  | 826000     | 1180000    | -0.000876619 | -0.001195084 |
| 38   | 55 |    | amplexiune      | 10 | 526        | 692        | -1.00086E-06 | -1.28568E-06 |
| 541  | 3  |    | încleștare      | 10 | 55800      | 55600      | -7.64425E-05 | -7.56024E-05 |
| 964  | 13 |    | rotocol         | 7  | 525000     | 528000     | -0.000584425 | -0.000582973 |
| 29   | 64 |    | alicuantă       | 9  | 700        | 760        | -1.31037E-06 | -1.40393E-06 |
| 433  |    | 59 | geobotanic      | 10 | 1850000    | 497000     | -0.001792512 | -0.00055216  |
| 98   | 18 |    | băzai           | 5  | 769000     | 784000     | -0.000822425 | -0.000830427 |
| 928  |    | 10 | receptioner     | 11 | 526000     | 693000     | -0.000585424 | -0.000743749 |
| 489  | 40 |    | iezuitic        | 8  | 2170       | 2130       | -3.77914E-06 | -3.6854E-06  |
| 263  |    | 63 | curpen          | 6  | 293000     | 263000     | -0.000345738 | -0.000311199 |
| 976  | 37 |    | sarcosporidioză | 15 | 577        | 580        | -1.09277E-06 | -1.08922E-06 |
| 1159 | 27 |    | țacănit         | 7  | 68600      | 65900      | -9.23552E-05 | -8.83358E-05 |
| 1052 |    | 42 | stigmă          | 6  | 56900000   | 56200000   | -0.032803155 | -0.032259353 |
| 684  |    | 47 | muslim          | 6  | 54000000   | 57500000   | -0.17213221  | -0.178194773 |
| 329  |    | 57 | drag            | 4  | 78400000   | 81700000   | -0.216430252 | -0.220597924 |
| 399  | 21 |    | fixism          | 6  | 25800      | 24700      | -3.76231E-05 | -3.58621E-05 |
| 497  | 7  |    | imunotransfuzie | 15 | 281        | 303        | -5.55492E-07 | -5.9137E-07  |
| 193  |    | 61 | clasicism       | 9  | 76800      | 78700      | -0.000102401 | -0.000103907 |
| 1005 | 69 |    | senar           | 5  | 14700000   | 13000000   | -0.010753442 | -0.009623578 |
| 262  |    | 76 | curea           | 5  | 8940000    | 10700000   | -0.007049081 | -0.008157555 |
| 608  |    | 72 | machetist       | 9  | 19400      | 18200      | -2.89264E-05 | -2.70559E-05 |
| 1065 | 23 |    | subînțelege     | 11 | 238000     | 195000     | -0.000286506 | -0.000237362 |
| 1043 | 39 |    | sprânceană      | 10 | 3460000    | 2490000    | -0.003104364 | -0.002310649 |
| 378  | 74 |    | faclă           | 5  | 382000     | 774000     | -0.000439151 | -0.000820963 |
| 779  |    | 25 | pas             | 3  | 2990000000 | 2920000000 | -0.36698091  | -0.366031729 |
| 40   | 2  |    | analitic        | 8  | 1240000    | 1100000    | -0.001258289 | -0.001122832 |
| 935  |    | 59 | regal           | 5  | 230000000  | 235000000  | -0.095799698 | -0.096708665 |
| 897  | 29 |    | pumnișor        | 8  | 2020       | 1890       | -3.53559E-06 | -3.2958E-06  |
| 1095 |    | 20 | știre           | 5  | 5760000    | 5556000    | -0.00483171  | -0.031964258 |
| 1082 |    | 9  | sfadă           | 5  | 114000     | 129000     | -0.000146845 | -0.000163078 |
| 714  | 16 |    | neurolimfă      | 10 | 387        | 171        | -7.501E-07   | -3.44853E-07 |
| 579  | 66 |    | laterit         | 7  | 638000     | 677000     | -0.000695971 | -0.000728373 |
| 798  |    | 41 | penetrație      | 10 | 2970000    | 3640000    | -0.002716676 | -0.003220848 |
| 735  |    | 28 | odolean         | 7  | 129000     | 139000     | -0.00016434  | -0.000174541 |
| 23   |    | 15 | ahtiat          | 6  | 32300      | 36800      | -4.6272E-05  | -5.17638E-05 |
| 442  | 75 |    | glorie          | 6  | 8700000    | 8920000    | -0.006886961 | -0.006984828 |
| 1068 |    | 38 | succesor        | 8  | 499000     | 564000     | -0.000558385 | -0.000618496 |
| 983  | 45 |    | sângerare       | 9  | 495000     | 583000     | -0.000554365 | -0.000637138 |
| 225  | 36 |    | constituție     | 11 | 1290000    | 1310000    | -0.001303185 | -0.001311197 |
| 353  |    | 67 | endometrită     | 11 | 34600      | 52900      | -4.92938E-05 | -7.22301E-05 |
| 145  |    | 56 | cardanic        | 8  | 407000     | 458000     | -0.000464937 | -0.000513083 |
| 663  | 53 |    | moale           | 5  | 7850000    | 9050000    | -0.006024946 | -0.007071753 |
| 155  |    | 33 | căciulă         | 7  | 2650000    | 2800000    | -0.002458572 | -0.002561007 |
| 298  |    | 49 | desfăta         | 7  | 120000     | 131000     | -0.000153868 | -0.000165377 |
| 479  | 63 |    | homeoterm       | 9  | 13000      | 13300      | -1.99785E-05 | -2.02454E-05 |
| 829  | 32 |    | platelaj        | 8  | 3530       | 3200       | -5.95174E-06 | -5.38882E-06 |
| 845  |    | 40 | politicesc      | 10 | 4020       | 3940       | -6.71777E-06 | -6.5419E-06  |
| 216  | 6  |    | concentra       | 9  | 79700000   | 39900000   | -0.042871346 | -0.02445522  |
| 598  | 12 |    | loc             | 3  | 258000000  | 313000000  | -0.104067462 | -0.118619035 |
| 485  | 76 |    | iaz             | 3  | 3620000    | 3300000    | -0.003229175 | -0.002956751 |

|                   |    |                                    |                  |                    |                    |                          |                          |              |
|-------------------|----|------------------------------------|------------------|--------------------|--------------------|--------------------------|--------------------------|--------------|
| 252               |    | 61                                 | cromocistoscopie | 16                 | 4500               | 4690                     | -7.46213E-06             | -7.69437E-06 |
| 25                |    | 28                                 | albie            | 5                  | 3570000            | 3300000                  | -0.00319026              | -0.002956751 |
| 636               |    | 72                                 | meditație        | 9                  | 5690000            | 5580000                  | -0.00478096              | -0.00466672  |
| 923               |    | 57                                 | râuri            | 5                  | 1730000            | 1420000                  | -0.001689529             | -0.001408294 |
| 702               | 17 |                                    | nefresc          | 8                  | 190000             | 147000                   | -0.000233625             | -0.000183652 |
| 273               |    | 9                                  | dârmoz           | 6                  | 9480000            | 9620000                  | -0.007411182             | -0.007450423 |
| 1000              |    | 4                                  | seimen           | 6                  | 648000             | 393000                   | -0.000705724             | -0.000447097 |
| 637               | 75 |                                    | megafon          | 7                  | 15300000           | 17000000                 | -0.011122251             | -0.012066736 |
| 962               | 7  |                                    | rontgen          | 7                  | 8240000            | 7210000                  | -0.006574092             | -0.005820082 |
| 1193              |    | 73                                 | vernis           | 6                  | 52500000           | 48200000                 | -0.030750491             | -0.028507874 |
| 237               | 67 |                                    | corijent         | 8                  | 7710               | 6610                     | -1.23103E-05             | -1.05867E-05 |
| 836               |    | 21                                 | pluti            | 5                  | 960000             | 842000                   | -0.0010023               | -0.000885037 |
| 639               | 74 |                                    | melodramă        | 9                  | 19400000           | 21300000                 | -0.013575167             | -0.014573424 |
| 1099              | 60 |                                    | tachina          | 7                  | 560000             | 572000                   | -0.000619247             | -0.000626354 |
| 907               |    | 1                                  | radioastronomie  | 15                 | 696000             | 646000                   | -0.000752304             | -0.00069846  |
| 21                |    | 34                                 | agoniza          | 7                  | 3370000            | 4490000                  | -0.003033788             | -0.003865949 |
| 10                | 3  |                                    | act              | 3                  | 2820000000         | 2880000000               | -0.365022909             | -0.365529208 |
| 916               | 10 |                                    | răgușit          | 7                  | 93500              | 134000                   | -0.000122561             | -0.00016882  |
| 325               |    | 2                                  | documentație     | 12                 | 8280000            | 8680000                  | -0.006601412             | -0.006823782 |
| 1024              | 41 |                                    | sistem           | 6                  | 372000000          | 324000000                | -0.134458975             | -0.121516765 |
| 1066              | 5  |                                    | subordonat       | 10                 | 360000             | 422000                   | -0.000416306             | -0.000476677 |
| 1113              | 35 |                                    | teleancheta      | 11                 | 133                | 87                       | -2.7363E-07              | -1.82129E-07 |
| 871               |    | 56                                 | preursi          | 7                  | 996                | 1940                     | -1.82402E-06             | -3.37724E-06 |
| 979               | 37 |                                    | sălbatic         | 8                  | 3030000            | 3080000                  | -0.002764617             | -0.002783768 |
| 373               |    | 70                                 | extenuant        | 9                  | 108000             | 144000                   | -0.000139785             | -0.000180241 |
| 949               | 66 |                                    | retrimitere      | 11                 | 153000             | 128000                   | -0.000191925             | -0.000161927 |
| 346               | 36 |                                    | electricitate    | 13                 | 4350000            | 4770000                  | -0.003788835             | -0.004074261 |
| 99                | 11 |                                    | belcanto         | 8                  | 5540000            | 5780000                  | -0.004671876             | -0.00481087  |
| 1218              |    | 55                                 | zărghală         | 9                  | 959                | 1040                     | -1.7595E-06              | -1.88412E-06 |
| 641               |    | 30                                 | mercur           | 6                  | 4330000            | 4500000                  | -0.0037737               | -0.003873422 |
| 864               |    | 27                                 | precedat         | 10                 | 9850               | 10400                    | -1.12909E-05             | -1.61215E-05 |
| 653               |    | 23                                 | mijloc           | 6                  | 8280000            | 9080000                  | -0.006601412             | -0.007091783 |
| 772               | 51 |                                    | parafrazare      | 11                 | 27200              | 11600                    | -3.95006E-05             | -1.78378E-05 |
| 572               | 44 |                                    | kilometru        | 9                  | 2680000            | 2780000                  | -0.00248295              | -0.002544978 |
| 1107              |    | 18                                 | tălgerei         | 8                  | 16900              | 15900                    | -2.54642E-05             | -2.38807E-05 |
| 14                | 43 |                                    | adjunct          | 7                  | 38200000           | 40300000                 | -0.023757872             | -0.024654729 |
| Total = 106 words |    | Total = 837 (7,9 letters/<br>word) |                  | 8730721923         | 8805003403         | -1.890196788             | -1.909348493             |              |
|                   |    |                                    |                  | Mean =<br>82365301 | Mean =<br>83066070 | Entropy =<br>1.890196788 | Entropy =<br>1.909348493 |              |

\* Note: The determination of the probability value  $p(x_i)$  is one of a frequency based nature  $[p(x_i) = f(x_i) : \sum f(x_i)]$

The entropy of the words of the sample, evaluated in time, on 21.01.2021 and 15.03.2021, respectively, evolves naturally ascending, from **1.890196788** to **1.909348493** through the determinations of informational type  $[S = H(X) = - \sum_{i=1}^n p(x_i) \times \log_b p(x_i)]$ , where  $b = e$  and not taking into account the Boltzmann constant in the phenomena of linguistic statistics]. The emphasis of this research is placed on the evolution or dynamics of entropy quantification processes and not on the effective level of entropy. Entropy thus

attests to the passage of time in the seemingly isolated system of the Internet, in terms of mentioning words in the language of network communications. Thus, entropy is redefined and constitutes in linguistic statistics the degree of disorder of communication in the universe, which increases continuously and naturally, as shown by the actual determinations. Modern communication multiplied through the language or network of the Internet, describes a system that evolves irreversibly and follows the arrow of time, ie the value of entropy always increases, through a continuous flow of information from those who know to those who do not know (sometimes the latter do not know in the so plastic way expressed by Socrates, according to which the more they know the more they realize how much they do not know or that in fact what they do not know grows much faster than what they know).

The descriptive statistics and econometric analysis identifies in parallel some other important aspects, related to the association of the dimensions with the mentions of words on Internet (table no.2) and especially matrix of correlation with high correlative intensities between the variables used (table no. 3 and 5) or some models with differentiated dynamics and intensities between those two samples or mentions extracted with Google Search (table no. 4, and 6-8)

The initial data series (SER02 and SER03) described in table no. 2 are characterized by an accentuated distributional abnormality of type, accompanied by a high heterogeneity and a very pronounced asymmetry, while by logarithm of the mentions the whole database is normalized, homogenized and symmetrized (SER04 and SER05):

### Descriptive statistics for the first three basic series and those derived by logarithm

*Table no. 2*

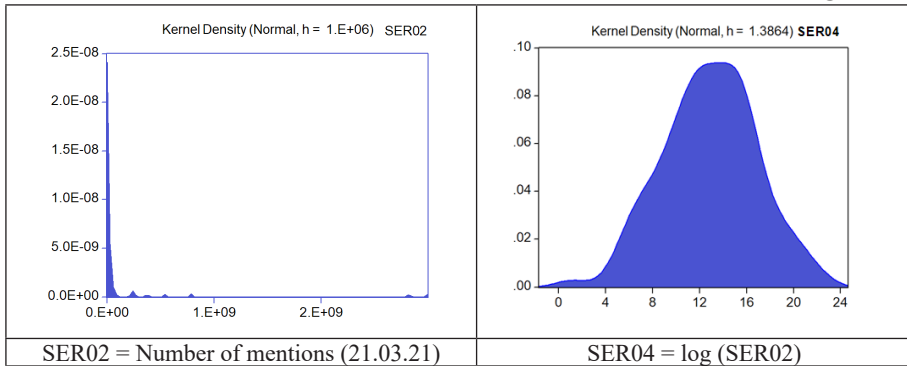
| Sample: 1 106 | Word's dimension | Number of mentions |          |            |            |
|---------------|------------------|--------------------|----------|------------|------------|
|               |                  | SER02              | SER03    | SER05      | SER06      |
| Indicators    | SER01            | 21.03.21           | 15.03.21 | log(SER02) | log(SER03) |
| Mean          | 7.896226         | 82365301           | 83066070 | 12.85743   | 12.88265   |
| Median        | 8.000000         | 497000.0           | 512500.0 | 13.11634   | 13.14660   |
| Maximum       | 16.00000         | 2.99E+09           | 2.92E+09 | 21.81854   | 21.79485   |
| Minimum       | 3.000000         | 3.000000           | 3.000000 | 1.098612   | 1.098612   |
| Std. Dev.     | 2.658067         | 4.07E+08           | 4.07E+08 | 3.959881   | 3.983315   |
| Skewness      | 0.772488         | 6.451298           | 6.405496 | -0.138605  | -0.156136  |
| Kurtosis      | 3.892969         | 44.72987           | 44.16034 | 2.907912   | 2.908925   |
| Jarque-Bera   | 14.06419         | 8426.377           | 8207.468 | 0.376857   | 0.467321   |
| Probability   | 0.000883         | 0.000000           | 0.000000 | 0.828260   | 0.791630   |
| Sum           | 837.0000         | 8.73E+09           | 8.81E+09 | 1362.888   | 1365.560   |
| Sum Sq. Dev.  | 741.8585         | 1.74E+19           | 1.74E+19 | 1646.469   | 1666.014   |

*Source: Realized by the author based on software package EViews*

Two Kernel-type distributions are clearly shown in fig. no. 1, describing the process of statistical normalization caused by the use of the exponential function and the derived specific logarithm, as can be seen from the confrontation of two graphs showing the initial series (SER02) and the same logarithmic series (SER04):

**Graphic confrontation of the distribution of mentions and logarithms of these values**

*Fig. no. 1*



Source: Realized by the author based on software package EViews

Potential or anticipated associations are identified by means of a correlation matrix (table no. 3) and describe a reverse statistical connection that must be tested statistically, between the dimensions and mentions of words in Internet communication and language. The ascent of the intensity of the correlation is noticeable, which increases with the entropy as the arrow of time.

**Correlation matrix**

*Table no. 3*

|       | SER01           | SER02     | SER03     | SER04     | SER05           |
|-------|-----------------|-----------|-----------|-----------|-----------------|
| SER01 | 1.000000        | -0.311137 | -0.315902 | 0.364028  | <u>0.373932</u> |
| SER02 | -0.311137       | 1.000000  | 0.999415  | -0.938801 | -0.936403       |
| SER03 | -0.315902       | 0.999415  | 1.000000  | -0.940829 | -0.940017       |
| SER04 | 0.364028        | -0.938801 | -0.940829 | 1.000000  | 0.997897        |
| SER05 | <u>0.373932</u> | -0.936403 | -0.940017 | 0.997897  | 1.000000        |

Source: Realized by the author based on software package EViews

Test F- statistic (Fisher – Snedecor) validates the correlation in both moments of time, with accents of increased intensity simultaneously with the growth of entropy’s level, using two econometric models similarly describing

the variation of the mentions in the logarithmic series in direct relation to the letter dimensions of the words. These models are described in table no. 4, where, on the left are modelled the data from 21.03.21 and on the right those from 15.03.21.

**Econometric confrontation in time of the models of the mentions according to the dimensions of the words in the Internet**

*Table no. 4*

| Dependent Variable: SER04 Method: Least Squares |             |                       |             |           |
|---|-------------|-----------------------|-------------|-----------|
| Sample: 1 106 Included observations: 106        |             |                       |             |           |
| Variable  | Coefficient | Std. Error            | t-Statistic | Prob.     |
| C   | -0.081595   | 0.016871              | -4.836329   | 0.0000    |
| SER01   | 0.008075    | 0.002026              | 3.985851    | 0.0001    |
| R-squared                                       | 0.132517    | Mean dependent var    |             | -0.017832 |
| Adjusted R-squared                              | 0.124175    | S.D. dependent var    |             | 0.058963  |
| S.E. of regression                              | 0.055181    | Akaike info criterion |             | -2.937723 |
| Sum squared resid                               | 0.316669    | Schwarz criterion     |             | -2.887469 |
| Log likelihood                                  | 157.6993    | F-statistic           |             | 15.88701  |
| Durbin-Watson stat                              | 2.094178    | Prob(F-statistic)     |             | 0.000125  |

| Dependent Variable: SER05 Method: Least Squares |             |                       |             |           |
|---|-------------|-----------------------|-------------|-----------|
| Sample: 1 106 Included observations: 106        |             |                       |             |           |
| Variable  | Coefficient | Std. Error            | t-Statistic | Prob.     |
| C   | -0.083688   | 0.016846              | -4.967945   | 0.0000    |
| SER01   | 0.008317    | 0.002023              | 4.111649    | 0.0001    |
| R-squared                                       | 0.139825    | Mean dependent var    |             | -0.018013 |
| Adjusted R-squared                              | 0.131554    | S.D. dependent var    |             | 0.059123  |
| S.E. of regression                              | 0.055097    | Akaike info criterion |             | -2.940764 |
| Sum squared resid                               | 0.315708    | Schwarz criterion     |             | -2.890510 |
| Log likelihood                                  | 157.8605    | F-statistic           |             | 16.90566  |
| Durbin-Watson stat                              | 2.109717    | Prob(F-statistic)     |             | 0.000002  |

*Source: Realized by the author based on software package EViews*

In the evaluation of entropy in the second stage on 15.03.21, three additional data series are generated by partial calculation (Annex no.1) representing the frequency determined frequency [SER06= $p(x_i)$ ], the logarithm of this probability [SER07= $\log_b p(x_i)$ , where  $b = e$ ] and the entropy (SER08), detailed per word. From all of these series, the logarithmic series of mentions (R = -0.447) and the entropy's series (R = -0.374) are correlated much more intensely with the number of letters, as can be seen from table no. 5:

## Correlation matrix of the data from Annex 1

*Table no. 5*

|       | SER01     | SER03     | SER06     | SER07     | SER08     |
|-------|-----------|-----------|-----------|-----------|-----------|
| SER01 | 1.000000  | -0.315902 | -0.315902 | -0.447054 | 0.373932  |
| SER02 | -0.315902 | 1.000000  | 1.000000  | 0.419067  | -0.940017 |
| SER06 | -0.315902 | 1.000000  | 1.000000  | 0.419067  | -0.940017 |
| SER07 | -0.447054 | 0.419067  | 0.419067  | 1.000000  | -0.552637 |
| SER08 | 0.373932  | -0.940017 | -0.940017 | -0.552637 | 1.000000  |

*Source: Realized by the author based on software package EViews*

Test F-statistic validates in the data series from 15.03.21 the maximum intensity correlation described in the econometric model of the statements (SER03) in relation to the data from their logarithmic series (SER07) in relation to the letter dimensions of the words (SER01), exposed model in table no. 6:

### Econometric model of the word mentions explained by the logarithmic series of probabilities of these mentions and to the dimensions of the words

*Table no. 6*

| Dependent Variable:SER03 Method: Least Squares Sample:1 106 Included observations: 106 |             |                       |             |          |  |
|--|-------------|-----------------------|-------------|----------|--|
| Variable   | Coefficient | Std. Error            | t-Statistic | Prob.    |  |
| C  | 6.33E+08    | 1.22E+08              | 5.172132    | 0.0000   |  |
| SER07  | 35500626    | 10096203              | 3.516235    | 0.0007   |  |
| SER01  | -24615515   | 15129929              | -1.626942   | 0.1068   |  |
| R-squared  | 0.196272    | Mean dependent var    |             | 83066070 |  |
| Adjusted R-squared   | 0.180665    | S.D. dependent var    |             | 4.07E+08 |  |
| S.E. of regression   | 3.69E+08    | Akaike info criterion |             | 42.31633 |  |
| Sum squared resid  | 1.40E+19    | Schwarz criterion     |             | 42.39171 |  |
| Log likelihood   | -2239.766   | F-statistic           |             | 12.57638 |  |
| Durbin-Watson stat   | 2.177465    | Prob(F-statistic)     |             | 0.000013 |  |

*Source: Realized by the author based on software package EViews*

Finally, the article offers in table no.7 a solution for calculating the entropy, a specific solution to linguistic statistics, starting from its econometric model (SER08) in relation to the number of mentions on the Internet of the selected words (SER03), frequencies of the mentions of the words in the specific language of the Internet (SER07):



**Econometric model of entropy explained by the number of mentions of the selected words and the natural logarithms of the frequency probabilities of the mentions**

*Table no. 7*

| Dependent Variable:SER08 Method: Least Squares Sample:1 106 Included observations: 106 |             |                       |             |           |
|--|-------------|-----------------------|-------------|-----------|
| Variable   | Coefficient | Std. Error            | t-Statistic | Prob.     |
| C  | -0.036269   | 0.005185              | -6.994869   | 0.0000    |
| SER03  | -1.25E-10   | 4.62E-12              | -27.03143   | 0.0000    |
| SER07  | -0.002857   | 0.000472              | -6.055791   | 0.0000    |
| R-squared  | 0.914186    | Mean dependent var    |             | -0.018013 |
| Adjusted R-squared   | 0.912520    | S.D. dependent var    |             | 0.059123  |
| S.E. of regression   | 0.017487    | Akaike info criterion |             | -5.226849 |
| Sum squared resid  | 0.031496    | Schwarz criterion     |             | -5.151469 |
| Log likelihood   | 280.0230    | F-statistic           |             | 548.6349  |
| Durbin-Watson stat   | 1.450300    | Prob(F-statistic)     |             | 0.000000  |

*Source: Realized by the author based on software package EViews*

An extension of the model with the inclusion of the number of letters in the word (inclusion in the model of SER01) is susceptible to (multi) collinearity as can be seen from table no. 8:

**Econometric model of entropy explained by the number of letters in the words, the number of mentions and the natural logarithms of the frequency probabilities of the mentions**

*Table no. 8*

| Dependent Variable: SER08 Method: Least Squares Sample: 1 106 Included observations: 106 |             |                       |             |           |
|--|-------------|-----------------------|-------------|-----------|
| Variable   | Coefficient | Std. Error            | t-Statistic | Prob.     |
| C  | -0.038808   | 0.006535              | -5.938500   | 0.0000    |
| SER01  | 0.000468    | 0.000729              | 0.641458    | 0.5227    |
| SER03  | -1.24E-10   | 4.69E-12              | -26.51275   | 0.0000    |
| SER07  | -0.002738   | 0.000508              | -5.386895   | 0.0000    |
| R-squared  | 0.914531    | Mean dependent var    |             | -0.018013 |
| Adjusted R-squared   | 0.912017    | S.D. dependent var    |             | 0.059123  |
| S.E. of regression   | 0.017537    | Akaike info criterion |             | -5.212007 |
| Sum squared resid  | 0.031370    | Schwarz criterion     |             | -5.111500 |
| Log likelihood   | 280.2364    | F-statistic           |             | 363.8039  |
| Durbin-Watson stat   | 1.470832    | Prob(F-statistic)     |             | 0.000000  |

*Source: Realized by the author based on software package EViews*

**Conclusions**

There are notable exceptions also to the meaningless terminological ambiguity of the language present on the Internet, but confronted temporally, and the most eloquent example is that of the word *recession* (recesiune) which gave rise to the famous *Rindex*, as a signal in the emergence of crises or recessions. Descriptive statistics, matrices of correlation and econometric modelling together reveal an opportunity to estimate the entropy and to

---

verify its permanent growth in the communication system and in the Internet language.

In this paper, entropy also marks the inexorable passage of time in the space of statistical linguistics and generalizes the aging of all philological systems, generated by national languages (theoretically isolated), including communication through language and Internet. Entropy is also here an arrow of time according to statistical landmarks to quantify the course or route of the arrow in the linguistic space... In order to ensure an evolutionary continuity of language and Internet, the well-known „*zero principle of thermodynamics*” will certainly be observed, applied in particular cases of economics, sociology and demography. According to this principle, “*any particular statistical set or system - described by the sciences that ensures the second order interdisciplinarity of statistical physics, respectively the distinct case of econophysics, sociophysics and demographics - located in external conditions invariable in time will reach, after a sufficient time long, a necessary and expected state of equilibrium, and simultaneously one of transitivity of this equilibrium*”, from one generation to another (Săvoiu, 2021). As one can see in this paper, this truth is practically validated in statistical linguistics, in the destructured Internet communication at the level of national languages and scientific languages.

#### References

1. A short history of the Web | CERN. 2020. *Where the Web was born. 2020*. Available online at: <https://home.cern/science/computing/shot-history-web>. Accessed on the 27th of January 2021.
2. Anderson, R.H., Bikosn, T. K., Law, S. A. & Mitchell, B.M. 1995. *Universal Access to E-Mail: Feasibility and Societal Implications*. Santa Monica: Rand Corporation.
3. Bellamy, D. 1987. *Innovative applications of computer technology in social work*. Paper presented at the Conference of the Canadian Association of Schools of Social Work, Learned Societies Meeting, Hamilton, Ontario.
4. Carrizo, P. G. 2018. *Cultura generală în 365 de zile*, București: Editura Lider Internațional.
5. Georgescu-Roegen, N. 1971. *The Entropy Law and the Economic Process*, Cambridge: Harvard University Press.
6. Gibbs, J.W. 1902. *Elementary Principles in Statistical Mechanics*, Dover: Yale Univ. Press.
7. Greene, R., 2000. *Web work: a history of internet art*, Art Forum International, pp.162–190.
8. Manovich, L. 2001. *The Language of New Media*. Cambridge MIT: Cambridge, Press.
9. Marson, S. M. 1997. A Selective History of Internet Technology and Social Work. *Computers in Human Services*, vol. 14(2), pp. 35–49.
10. Mowbray, S. 2017. *Did Switzerland Invent the Internet?* Available online at: <https://theculturetrip.com/europe/switzerland/articles/did-switzerland-invent-theinternet/>, Accessed on the 26th of January, 2021.

- 
11. Săvoiu, G. 2009. *Statistica: mod de gândire și metode*, București: Editura Universitară.
  12. Săvoiu, G., Cudanov, M. 2020. New Typology of Online Scientific Research and Multi-disciplinarity, *ESMSJ*, vol 9(2), pp. 4-8. Available online at: [https://www.esmsj.upit.ro/No20\\_2020.html](https://www.esmsj.upit.ro/No20_2020.html) Accessed on the 23th of January 2021.
  13. Săvoiu, G. 2013. *Situații statistice financiar-contabile și sisteme de indicatori statistici derivați*, București: Editura universitară [*Financial and accountancy statistical situations and systems of derived statistical indicators*, Bucharest: Universitara Publishers].
  14. Săvoiu, G. 2021. *Metode statistice și interdisciplinaritate*, București: Editura Universitară.
  15. Shannon, C.E. și Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.
  16. The birth of the Web | CERN 2020. *The World Wide Web was invented by British scientist Tim Berners Lee in 1989 while working at CERN*. Available online at: <https://home.cern/science/computing/birth-web>. Accessed on the 25th of January 2021.
  17. Language. 2020. Oxford English Dictionary, Language, Available on line at: <https://www.lexico.com/definition/language>, Accessed on the 12<sup>th</sup> of February 2021.

**Detailed calculation of the entropy of the randomly extracted 106 words' sample from DEX (2012), according to Internet mentions, quantified with Google Search on 15.03.2021**

| Pag. | Col. 1 | Col. 2 | Random extracted word | Number of letters | Google mentions on Internet | Determinant values* per word in 15.03.2021 |              |                             |
|------|--------|--------|-----------------------|-------------------|-----------------------------|--|--------------|-----------------------------|
|      |        |        |                       |                   | 15.03.2021                  | $P(x_i)$                                   | $\ln P(x_i)$ | $P(x_i) \times \log P(x_i)$ |
| 50   | 4      |        | antisumpant           | 11                | 15800                       | 1.79443429E-06                             | -13.23082075 | -2.37418E-05                |
| 372  |        | 39     | exponent              | 8                 | 24900000                    | 0.002827937578                             | -5.868207604 | -0.016594925                |
| 488  | 50     |        | ieपुरoi               | 7                 | 33500                       | 3.80465498E-06                             | -12.47928525 | -4.74794E-05                |
| 1109 |        | 60     | tămpănă               | 7                 | 3                           | 3.40715371E-10                             | -21.79997368 | -7.42759E-09                |
| 251  |        | 44     | criticism             | 9                 | 22500000                    | 2.55536528E-02                             | -3.666975005 | -0.093704606                |
| 92   |        | 49     | bavură                | 6                 | 41300                       | 4.69051494E-06                             | -12.26996819 | -5.75525E-05                |
| 1038 |        | 43     | spectacol             | 9                 | 5390000                     | 6.12151950E-04                             | -7.398530022 | -0.004529025                |
| 18   | 73     |        | afet                  | 4                 | 13100000                    | 1.48779045E-03                             | -6.510463177 | -0.009686205                |
| 672  | 20     |        | monosepal             | 9                 | 4200                        | 4.77001519E-07                             | -14.55574616 | -6.94311E-06                |
| 759  | 11     |        | ovreică               | 7                 | 4670                        | 5.30380261E-07                             | -14.44967161 | -7.66382E-06                |
| 277  | 12     |        | decăt                 | 5                 | 55200000                    | 6.26916282E-03                             | -5.072112454 | -0.031797899                |
| 118  | 46     |        | bregmă                | 6                 | 497                         | 5.64451798E-08                             | -16.68999594 | -9.4207E-07                 |
| 487  |        | 8      | idiotism              | 8                 | 120000                      | 1.36286148E-05                             | -11.20333894 | -0.000152686                |
| 160  |        | 35     | cărușă                | 7                 | 44200                       | 5.01987313E-06                             | -12.2021059  | -6.1253E-05                 |
| 861  |        | 50     | prăvălie              | 8                 | 153000                      | 1.73764839E-05                             | -10.96039276 | -0.000190453                |
| 559  |        | 1      | învălu                | 7                 | 207000                      | 2.35093606E-05                             | -10.65811189 | -0.000250565                |
| 1195 | 63     |        | veșteji               | 7                 | 37600                       | 4.27029932E-06                             | -12.36382664 | -5.27972E-05                |
| 503  | 70     |        | individualizare       | 15                | 123000                      | 1.39693302E-05                             | -11.17864633 | -0.000156158                |
| 676  |        | 37     | moșneguț              | 8                 | 13600                       | 1.54457635E-06                             | -13.38076089 | -2.06676E-05                |
| 172  | 53     |        | cerdac                | 6                 | 120000                      | 1.36286148E-05                             | -11.20333894 | -0.000152686                |
| 870  |        | 14     | prestigiu             | 9                 | 1180000                     | 1.34014713E-04                             | -8.917560969 | -0.001195084                |
| 38   | 55     |        | amplexiune            | 10                | 692                         | 7.85916789E-08                             | -16.35900001 | -1.28568E-06                |
| 541  | 3      |        | înceleștare           | 10                | 55600                       | 6.31459154E-06                             | -11.97264749 | -7.56024E-05                |
| 964  | 13     |        | rotocol               | 7                 | 528000                      | 5.99659053E-05                             | -9.721734403 | -0.000582973                |
| 29   | 64     |        | alicuantă             | 9                 | 760                         | 8.63145606E-08                             | -16.26526753 | -1.40393E-06                |
| 433  |        | 59     | geobotanic            | 10                | 497000                      | 5.64451798E-05                             | -9.78224066  | -0.00055216                 |
| 98   | 18     |        | băzai                 | 5                 | 784000                      | 8.90402836E-05                             | -9.326421666 | -0.000830427                |
| 928  |        | 10     | recepționer           | 11                | 693000                      | 7.87052507E-05                             | -9.449800687 | -0.000743749                |
| 489  | 40     |        | iezuitic              | 8                 | 2130                        | 2.41907913E-07                             | -15.23470871 | -3.6854E-06                 |
| 263  |        | 63     | curpen                | 6                 | 263000                      | 2.98693808E-05                             | -10.41867665 | -0.000311199                |
| 976  | 37     |        | sarcosporidioză       | 15                | 580                         | 6.58716384E-08                             | -16.53555786 | -1.08922E-06                |
| 1159 | 27     |        | țăcănit               | 7                 | 65900                       | 7.48438098E-06                             | -11.80269224 | -8.83358E-05                |
| 1052 |        | 42     | stigmă                | 6                 | 56200000                    | 6.38273461E-03                             | -5.05415865  | -0.032259353                |
| 684  |        | 47     | muslim                | 6                 | 57500000                    | 6.53037794E-02                             | -2.728705367 | -0.178194773                |
| 329  |        | 57     | drag                  | 4                 | 81700000                    | 9.27881527E-02                             | -2.377436312 | -0.220597924                |
| 399  | 21     |        | fixism                | 6                 | 24700                       | 2.80522322E-06                             | -12.78402744 | -3.58621E-05                |
| 497  | 7      |        | imunotransfuzie       | 15                | 303                         | 3.44122525E-08                             | -17.18485316 | -5.9137E-07                 |
| 193  |        | 61     | clasicism             | 9                 | 78700                       | 8.93809990E-06                             | -11.62518753 | -0.000103907                |
| 1005 | 69     |        | senar                 | 5                 | 13000000                    | 1.47643327E-03                             | -6.51812605  | -0.009623578                |
| 262  |        | 76     | curea                 | 5                 | 10700000                    | 1.21521816E-03                             | -6.712831666 | -0.008157555                |
| 608  |        | 72     | machetist             | 9                 | 18200                       | 2.06700658E-06                             | -13.08940909 | -2.70559E-05                |
| 1065 | 23     |        | subînțelege           | 11                | 195000                      | 2.21464991E-05                             | -10.71783113 | -0.000237362                |
| 1043 | 39     |        | sprânceană            | 10                | 2490000                     | 2.82793758E-04                             | -8.170792697 | -0.002310649                |
| 378  | 74     |        | faclă                 | 5                 | 774000                      | 8.79045657E-05                             | -9.339258813 | -0.000820963                |
| 779  |        | 25     | pas                   | 3                 | 292000000                   | 3.31629628E-01                             | -1.103736512 | -0.366031729                |
| 40   | 2      |        | analitic              | 8                 | 1100000                     | 1.24928969E-04                             | -8.987765228 | -0.001122832                |
| 935  |        | 59     | regal                 | 5                 | 235000000                   | 2.66893707E-02                             | -3.623489893 | -0.096708665                |
| 897  | 29     |        | pumnișor              | 8                 | 1890                        | 2.14650684E-07                             | -15.35425386 | -3.2958E-06                 |
| 1095 |        | 20     | știre                 | 5                 | 55560000                    | 6.31004867E-03                             | -5.065611889 | -0.031964258                |
| 1082 |        | 9      | sfadă                 | 5                 | 129000                      | 1.46507609E-05                             | -11.13101828 | -0.000163078                |
| 714  | 16     |        | neurolimfă            | 10                | 171                         | 1.94207761E-08                             | -17.75692241 | -3.44853E-07                |

|                        |    |                                   |                  |    |            |                |              |              |
|------------------------|----|-----------------------------------|------------------|----|------------|----------------|--------------|--------------|
| 579                    | 66 |                                   | laterit          | 7  | 677000     | 7.68881020E-05 | -9.473159413 | -0.000728373 |
| 798                    |    | 41                                | penetrație       | 10 | 3640000    | 4.13401317E-04 | -7.791091726 | -0.003220848 |
| 735                    |    | 28                                | odolean          | 7  | 139000     | 1.57864789E-05 | -11.05635675 | -0.000174541 |
| 23                     |    | 15                                | ahliat           | 6  | 36800      | 4.17944188E-06 | -12.38533284 | -5.17638E-05 |
| 442                    | 75 |                                   | glorie           | 6  | 8920000    | 1.01306037E-03 | -6.894779461 | -0.006984828 |
| 1068                   |    | 38                                | succesor         | 8  | 564000     | 6.40544897E-05 | -9.655776435 | -0.000618496 |
| 983                    | 45 |                                   | săngerare        | 9  | 583000     | 6.62123537E-05 | -9.62264635  | -0.000637138 |
| 225                    | 36 |                                   | constituție      | 11 | 1310000    | 1.48779045E-04 | -8.81304827  | -0.001311197 |
| 353                    |    | 67                                | endometrită      | 11 | 52900      | 6.00794771E-06 | -12.02242735 | -7.22301E-05 |
| 145                    |    | 56                                | cardanic         | 8  | 458000     | 5.20158800E-05 | -9.863961502 | -0.000513083 |
| 663                    | 53 |                                   | moale            | 5  | 9050000    | 1.02782470E-03 | -6.88031065  | -0.007071753 |
| 155                    |    | 33                                | căciulă          | 7  | 2800000    | 3.18001013E-04 | -8.05345599  | -0.002561007 |
| 298                    |    | 49                                | desfăta          | 7  | 131000     | 1.48779045E-05 | -11.11563336 | -0.000165377 |
| 479                    | 63 |                                   | homeoterm        | 9  | 13300      | 1.51050481E-06 | -13.40306665 | -2.02454E-05 |
| 829                    | 32 |                                   | platelaj         | 8  | 3200       | 3.63429729E-07 | -14.82767988 | -5.38882E-06 |
| 845                    |    | 40                                | politicesc       | 10 | 3940       | 4.47472854E-07 | -14.61964996 | -6.5419E-06  |
| 216                    | 6  |                                   | concentra        | 9  | 39900000   | 4.53151443E-03 | -5.396699083 | -0.02445522  |
| 598                    | 12 |                                   | loc              | 3  | 31300000   | 3.55479704E-02 | -3.336872217 | -0.118619035 |
| 485                    | 76 |                                   | iaz              | 3  | 3300000    | 3.74786908E-04 | -7.889152939 | -0.002956751 |
| 252                    |    | 61                                | cromocistoscopie | 16 | 4690       | 5.32651697E-07 | -14.4453981  | -7.69437E-06 |
| 25                     |    | 28                                | albie            | 5  | 3300000    | 3.74786908E-04 | -7.889152939 | -0.002956751 |
| 636                    |    | 72                                | meditație        | 9  | 5580000    | 6.33730590E-04 | -7.363886631 | -0.00466672  |
| 923                    |    | 57                                | răuri            | 5  | 1420000    | 1.61271942E-04 | -8.732418536 | -0.001408294 |
| 702                    | 17 |                                   | nefiresc         | 8  | 147000     | 1.66950532E-05 | -11.0003981  | -0.000183652 |
| 273                    |    | 9                                 | dărmoz           | 6  | 9620000    | 1.09256062E-03 | -6.819231143 | -0.007450423 |
| 1000                   |    | 4                                 | seimen           | 6  | 393000     | 4.46337136E-05 | -10.01702107 | -0.000447097 |
| 637                    | 75 |                                   | megafon          | 7  | 17000000   | 1.93072044E-03 | -6.249862063 | -0.012066736 |
| 962                    | 7  |                                   | rontgen          | 7  | 7210000    | 8.18852608E-04 | -7.107606456 | -0.005820082 |
| 1193                   |    | 73                                | verniz           | 6  | 48200000   | 5.47416029E-03 | -5.207716386 | -0.028507874 |
| 237                    | 67 |                                   | corijent         | 8  | 6610       | 7.50709534E-07 | -14.10224703 | -1.05867E-05 |
| 836                    |    | 21                                | pluti            | 5  | 842000     | 9.56274474E-05 | -9.255050672 | -0.000885037 |
| 639                    | 74 |                                   | melodramă        | 9  | 21300000   | 2.41907913E-03 | -6.024368335 | -0.014573424 |
| 1099                   | 60 |                                   | tachina          | 7  | 572000     | 6.49630641E-05 | -9.641691695 | -0.000626354 |
| 907                    |    | 1                                 | radioastronomie  | 15 | 646000     | 7.33673765E-05 | -9.520031183 | -0.00069846  |
| 21                     |    | 34                                | agoniza          | 7  | 4490000    | 5.09937338E-04 | -7.581222706 | -0.003865949 |
| 10                     | 3  |                                   | act              | 3  | 288000000  | 3.27086756E-01 | -1.117529834 | -0.365529208 |
| 916                    | 10 |                                   | răgușit          | 7  | 134000     | 1.52186199E-05 | -11.09299089 | -0.00016882  |
| 325                    |    | 2                                 | documentație     | 12 | 8680000    | 9.85803140E-04 | -6.922053879 | -0.006823782 |
| 1024                   | 41 |                                   | sistem           | 6  | 324000000  | 3.67972601E-02 | -3.302331892 | -0.121516765 |
| 1066                   | 5  |                                   | subordonat       | 10 | 422000     | 4.79272955E-05 | -9.945825372 | -0.000476677 |
| 1113                   | 35 |                                   | teleancheta      | 11 | 87         | 9.88074576E-09 | -18.43267785 | -1.82129E-07 |
| 871                    |    | 56                                | preursi          | 7  | 1940       | 2.20329273E-07 | -15.32814271 | -3.37724E-06 |
| 979                    | 37 |                                   | sălbatic         | 8  | 3080000    | 3.49801114E-04 | -7.95814581  | -0.002783768 |
| 373                    |    | 70                                | extenuant        | 9  | 144000     | 1.63543378E-05 | -11.02101739 | -0.000180241 |
| 949                    | 66 |                                   | retrimitere      | 11 | 128000     | 1.45371892E-05 | -11.13880042 | -0.000161927 |
| 346                    | 36 |                                   | electricitate    | 13 | 4770000    | 5.41737440E-04 | -7.520729102 | -0.004074261 |
| 99                     | 11 |                                   | belcanto         | 8  | 5780000    | 6.56444948E-04 | -7.328671725 | -0.00481087  |
| 1218                   |    | 55                                | zărgheală        | 9  | 1040       | 1.18114662E-07 | -15.95160997 | -1.88412E-06 |
| 641                    |    | 30                                | mercur           | 6  | 4500000    | 5.11073056E-04 | -7.578998011 | -0.003873422 |
| 864                    |    | 27                                | predecedat       | 10 | 10400      | 1.18114662E-06 | -13.64902488 | -1.61215E-05 |
| 653                    |    | 23                                | mijloc           | 6  | 9080000    | 1.03123186E-03 | -6.877001215 | -0.007091783 |
| 772                    | 51 |                                   | parafrazare      | 11 | 11600      | 1.31743277E-06 | -13.53982559 | -1.78378E-05 |
| 572                    | 44 |                                   | kilometru        | 9  | 2780000    | 3.15729577E-04 | -8.06062448  | -0.002544978 |
| 1107                   |    | 18                                | tălgerei         | 8  | 15900      | 1.80579147E-06 | -13.22451158 | -2.38807E-05 |
| 14                     | 43 |                                   | adjunct          | 7  | 40300000   | 4.57694315E-03 | -5.386723938 | -0.024654729 |
| Total = 106<br>cuvinte |    | Total= 837 (7,9litere/<br>cuvânt) |                  |    | 8805003403 | 1.00000000E+00 | -            | -1.909348493 |

\* Note: The determination of the probability value  $p(x_i)$  is one of a frequency based nature  $[p(x_i) = f(x_i) / \sum f(x_j)]$