CHEBANOV LAW AND VAKAR FORMULA IN MATHEMATICAL MODELS OF COMPLEX SYSTEMS

Josué-Antonio Nescolarde-Selva[*1], *Josep-Lluis Usó-Doménech*[1], *Miguel Lloret-Climent*[1] *and Lucía González-Franco*[2]

1Department of Applied Mathematics. University of Alicante. Alicante. Spain.
2Biodiversity Research Institute CIBIO, University of Alicante, Alicante, Spain

**Corresponding author: email: josue.selva@ua.es, Telephone. 0034680418381, Fax number. 0034 965 909 707*

ABSTRACT
Ecological models written in a mathematical language L(M) or model language, with a given style or methodology can be considered as a text. It is possible to apply statistical linguistic laws and the experimental results demonstrate that the behaviour of a mathematical model is the same of any literary text of any natural language. A text has the following characteristics: a) the variables, its transformed functions and parameters are the lexic units or LUN of ecological models; b) the syllables are constituted by a LUN, or a chain of them, separated by operating or ordering LUNs; c) the flow equations are words; c) the distribution of words (LUM and CLUN) according to their lengths is based on a Poisson distribution, the Chebanov's law. It is founded on Vakar`s formula, that is calculated likewise the linguistic entropy for L(M). We will apply these ideas over practical examples using MARIOLA model. In this paper it will be studied the problem of the lengths of the simple lexic units composed lexic units and words of text models, expressing these lengths in number of the primitive symbols, and syllables. The use of these linguistic laws renders it possible to indicate the degree of information given by an ecological model.

Keywords: Chebanov Low, Ecological Model, language, Mariola Model, Vakar formula

1. INTRODUCTION

The joint of linguistic enunciation submitted to the analysis is called text. Hence, the text is a linguistic behaviour sample that can be written or spoken. A text designates a written enunciation, it may be, long or short, ancient or new (Nescolarde-Selva and Usó-Doménech, 2013; Usó-Doménech and Nescolarde-Selva, 2012). The word STOP is so text as Hamlet. All studied linguistic material form a text, that it is collected of one or of several languages. With the naked eye, we have two types of text, the first is constituted of a simple designation of enunciation and the second must tolerate the existence of a series of conditions: have a written expression, be a connotative system, be closed and to possess logical order, temporary and spatial. All a series of elements: argument, style, syntax, etc. can act in the text as supports of an expressed ideological load indirectly through them. All text, also, possesses a references system to the Reality more or less rich. To this type of text is called "*literary text*".
A model is considered as a complex cognitive structure, at the same time it is expressed in a given language, which has been defined as L(M) (Usó-Domènech et al., 1997[b]; Usó-Domènech, P. Sastre-Vazquez and Mateu, 2001; Usó-Domènech, Sastre-Vazquez, 2002; Usó-Domènech, Vives Maciá and Mateu, 2006[a,b]; Villacampa & Usó-Domènech,

1999) and whose metalanguage is the formal mathematical language. In L(M) all the written records are texts.
A mathematical model will be a text if the following conditions are met (Nescolarde-Selva and Usó-Doménech, 2013; Sastre-Vazquez et al., 1999; Usó-Doménech and Nescolarde-Selva, 2012):

a) It must be an expression written in a formal language and the same text can have different levels of meaning or semantic levels.

b) It must be closed and any modification such as adding or removing any component (variable, flow equation or subsystem) converts it into a different text.

c) It must have, at least, three orders:

1) *Logical order*, where the logic relationship of sentences or the analysis of a proposition as an expression of a particular hypothesis is done. The hypothesis of the model will be given when the logical relationship is known.

2) *Temporary order*, also logical, since the time forms part of the logic structure of proposition.

3) *Spatial order,* which builds an unidimensional chain through the own restrictions of the model (Margalef, 1991).

Furthermore, a text possesses an own style labeled by the subjectivity of the modeller. The study of semantic structure of a text (ecological model) can lead us to interesting conclusions at the same that of outlining disaggregation in complex levels (semantic levels), in choosing its aggregation level as well as in interpreting the own model, not only in its globality (text), but also in its different parts (submodels, flow equations, etc.).
Classical text laws in the linguistic context, such as the range-frequency laws (Mandelbrot, 1954, 1961; Zipf, 1949; Vakar (in Marcus et al., 1966)), have been studied, in L(M) language models (Villacampa & Usó-Domènech, 1999) adapting them to the mathematical and ecological context, (Villacampa et al., 1999[a,b]; Sastre-Vazquez et al., 1999).
Zipf (1949) formulated a minimal effort principle in the natural language, which is not only applicable to the sounds of the speech, but also to other elements of the language, especially to the words. Such author observed that the product of the frequency of a word (or rather the number of times that is presented in a given text) and their ordinal position or rank (in the frequencies list: the most common word has rank 1, continues to that of rank or ordinal position 2, etc) is constant. This can be expressed through the formulation $f*r=$ constant (where f is the frequency and r is the degree of energy that the sign requires). Zipf interprets the symmetry of this law as the balance between two opposite forces: the speaker tends to repeat the same word as much as possible, that is to say, to use words as "thing" and "good", or pronouns and other substantive words instead of the exact word required by the context and the user needs the maximum clarity, with specify descriptions and the greater possible variety in the used words.

Between the two extremes of *"the same word for all the concepts* " and "*a special word for each one of the concepts*", it is established a balance expressed by the previous equation that represents the principle of the minimal effort. Thus each word has a certain probability and it is considerably more probable than the reader of a text find articles (which hardly influences the content of the text) and not the noun Patagonie, for example, that influences the content.
Kanding (1897) demonstrated that 15 less common words represent 25% of the total number of words of a text, that 66 most common extreme represent 50% of the text and 320 most common 72% of the total.
Thus, with a vocabulary of only 320 words, a person would be capable of understanding the three fourths of the words of any text. It is evident that this does not mean that the three fourths of the content is going to be comprehensible for him, since a considerable number of the most common words are empty of real content (articles, pronouns, etc), while some less common words of the text can occupy decisive positions, and one must understand them before that the text will interpreted. But on the other hand it has been demonstrated that it is possible to understand texts in foreign languages knowing only a very reduced vocabulary, conditioned on the fact that the vocabulary of the text will be quite basic.
In a general inspection of the methods and results of the statistics of the language, Guiraud (1954) summarizes the results in the following principles:

1) In any given text it will be found that a very small number of words constitutes the lion's share of the text.

2) In any given text, a very reduced and well chosen number of wordswill cover the great part of the text.

As an example consider the following: the 100 most common words will cover 60% of any text, the 1,000 most common words will cover 85%, and the 4,000 most common words 97.5%. All the other words therefore account for no more than 2.5% of the vocabulary in any given text.

The MARIOLA model (Usó-Domenech et al., 1995, 1997a) simulates certain shrub species found in a Mediterranean terrestrial ecosystem, and an indefinite number of models can be constructed for these same plants. The methodology used for this study permitted selection of the 'best model', i.e. the model that theoretically provided the most information. The best model can therefore be defined as the one that most closely reflects ecological reality (relationships and processes), thus enabling a better understanding of the ecosystem, with all the advantages that such an understanding provides. We are aware that it is impossible to achieve a 'perfect model', since according to Bonini's paradox (Bonini, Ch. P., 1963; Usó-Doménech, J. L., Nescolarde-Selva, J., & Lloret-Climent, M., 2014), a model as complex as the reality it simulates is identical to that same reality and thus becomes incomprehensible. In this paper, it will be studied the problem of the lengths of the simple lexic units composed lexic units and words of text models, expressing these lengths in number of the primitive symbols, and syllables respectively and it is one goal to prove empirically if the law stated by

Chebanov, which is fulfilled in the natural languages, also if it is satisfied in the formal language L(M).

2. THE MARIOLA MODEL

The MARIOLA model, so called for having taken as the base the mountainous terrestrial ecosystem of the Sierra de Mariola (Alicante, Spain), is a simulation of the behaviour and development of a typical bush ecosystem of the mediterranean area. In these shrub lands we find the representative bushes: *Bupleurus fruticescens* L., *U/ex parviflorus* Pourret, *Helychrysum stoechas* (L.) Moench, *Rosmarinus officinalis* L., *Lavandula latifolia* Medicus, *Sedum sediforme* (Jacq.) Pau, *Genista scorpius* (L.) OC. in Lam. and OC., *Marrubium vulgaris* L., *Thymus vulgaris* L, *Cistus albidus* L. They are common plants (Stübing et al.,1989) which play an important role in the shrub communities of the western Mediterranean region, specially during the first ten years after a fire. It is interspersed with areas of artificial reforestation of *Pinus halepensis.*

The MARIOLA can be characterized as flow lows. (1) It is a compartmental but not necessarilylinear model. (2) The input and output flows of each compartment or level are calculated by means of nonlinear regression equations. (3) The fauna is considered indirectly through a process of defoliation or destruction of the biomass by action of invertebrate predators and herbivore mammals. (4) Human action is not explicit. (5) The temporal unit for the measurements and simulation is one month for the reproductive submodel, the temporal resolution is one week. (6) The spatial extent is of 100 m^2. (7) The basic magnitude is biomass, with as unit grams of dry living material. (8) The model simulates the individual development of each bush species and the process of decomposition, in the space limited by the canopy of the plant. (9) The model does not take into consideration problems of competition. (l0) The disaggregation is intermediary, that is, it is not sufficiently disaggregated to study behaviour in the morpho or ecophysiological scales. (11) Processes of decomposition are considered as "black box"; that is, the existence of decomposers causing the decomposition is not taken into account. Nor are biochemical processes of degradation of cellulose and lignin considered. (12) The processes of decomposition of humus are referred to the 0 horizon of the soil. (13) In the actual state, the MARIOLA model has been validated with one shrub species, the *Cistus albidus.* Nothing impedes its validation in any other species, arbutus or herbaceous, provided that the equations of growth are known. (14) The model simulates the behaviour of the evolution of the plant biomass on short and medium terms and establishes an objective to observe the development in normal and limited (desertification) conditions. The MARIOLA model consists of the following submodels:

1. Submodel of growth:
   - growth,
   - defoliation,
   - destruction of the biomass.

2. Submodel on the decomposition of fallen biomass.
3. Submodel on reproduction:
   - formation of florist buds,
   - flowering,
   - fructification.

3. THE SEMI-FORMAL LANGUAGE L(M) AND ITS LEXIC UNITS

It has already been defined a model mathematical language L(M) or text-model language (Usó-Domènech et al., 1997[b]; Usó-Domènech, P. Sastre-Vazquez and Mateu, 2001; Usó-Domènech, Sastre-Vazquez, 2002; Usó-Domènech, Vives Maciá and Mateu, 2006[a,b]; Villacampa & Usó-Domènech, 1999) and it has been used in the particular methodology of complex systems modelling (Usó-Domènech et al., 1995, 1997[a]).
The language L (M) is a formal language regarding syntax and it is also natural with respect to recognizability Criteria (Villacampa et al., 1999[b]). Thus we can consider L (M) as a semi-formal language. In order to make linguistic analysis we first define basical linguistic units as (Sastre-Vazquez et al., 1999):

Primitive symbols: They are the sets of characters used to express variables and functions (VEVI, H, T, cos, tang, log, etc). Each variable x defines an associative field V_x, that it is formed by all possible transformed functions of the said variable, $\phi_x^n; n = 0,1,...$, being n the order of the transformed function. The zero-order transformed function of x is the same variable x.

Symbolic alphabet: It is built with primitive symbols

Textual alphabet: It is built with primitive symbols and numbers.

Simple lexic units (SLUN): They are the transformed functions as cos H, log VEVI, expT, etc. (Usó-Domènech et al., 1997[a]), the numbers and the variables (VEVI, T, H, etc.) and since they have a meaning.

Let x be one variable and let m be the number of primitive symbols of first order ϕ_x^1. Said number m is arbitrary, that is to say, it depends on the modeller. It is demostrated (Usó-Doménech, et. al, 1997[b]; Villacampa & Usó-Domènech, 1999) that the cardinal of the associative field of x, V_x is

$$V_x = \frac{\left(m^{n+1} - 1\right)}{m - 1} \quad (1)$$

Let α be the number of variables. As each variable has an associative field, and with the supposition that m is the same for each one, the cardinal N of the symbolic alphabet or total number of letters will be

$$N = \alpha \frac{m^{n+1} - 1}{m - 1} \quad (2)$$

On the other hand the linguistic units SLUN have meaning, which specifically lack the letters of an alphabet either. This meaning is given by the facts that correspond to behaviors of reality. For example, if expH is expressed by referring that atmospheric moisture is taking positive exponential behavior at some point in its trajectory,being any of the established criteria recognizability, who can "*choose*" such behavior among all possible. In addition, any SLUN has a signifier. The significance of SLUN is given by the observer. Of this duality signifier-significance[1] have been proposed principles of complementarity and uncertainty (Usó-Domènech et al, 1997[b]; 2000).

The operating lexic units or operator-LUN (op-LUN): They are the mathematical signs: +, -.

The ordenating lexic units or ordenating-LUN (or-LUN): They are the signs: =, <, >.

Special lexic unit: It is the sign d/dt and it defines the beginning a phrase (state equation).

Signs: They are (,), { ,}, [,], since they lack of meaning and they are the equivalent to the signs of ?, !, ; (,) in the natural languages.

Separating of lexic units (s-LUN): They are the signs * and /.

Composed lexic units (CLUM): They are the strings of SLUN separated by s-LUN.

Syllables or composed lexic units (CLUN): They are constituted by a SLUN, or a chain of them, separated by an op-LUN or a or-LUN.

[1] In any process, we can distinguish that it has a *signifier* as an inherent property, and having *significance* when it is related to the rest of the processes of the perceived Reality that the Subject considers as a system. The existence of information is independent of the fact that there is a Subject able to decode the message, to which the Subject is attempting to communicate. This objective information is termed signifier. The information in a message acquires meaning if a Subject decodes the message. This subjective information is termed significance. Therefore, the signifier is an ontic property, considering that the significance will be a system of meaning. The signifier is absolute and infinite, the significance is relative and finite. The signifier comes from Absolute Being and significance generates the relative being. The signifier is interpreted as the material or physical form of the sign and is something that can be caught (perception) by some of the traditional senses of the human being. The significance, on the other hand is a *mental construct*. In our approach, the signifier has a truth value equal to 1, that is to say, $v(S) = 1$, whereas the significance has as truth value a real positive number $v(s)$, between 0 and 1, with 0 corresponding to *absolute ignorance of the signifier* (therefore of the process) and 1 to *absolute understanding*, that is to say, $v(S) = v(s)$.

A-signifier (A-□) or the first order signifier is the signifier that is inherent to beings, processes or phenomena of the referring context. B-signifier (B-□), the second order signifier or connotation, is the signifier of significance s.

Words: They are flow equation (Forrester, 1961; Usó-Domènech et al., 1997[a]; Villacampa & Usó-Domènech, 1999). The symbols [ ] preceeding or by the symbols + and - are word separations.

Phrase: It is each ordinary differential equation (DOE) or state equation (Villacampa & Usó-Domènech, 1999).

Text: It is the concatenation of phrases.

Length of a text: It is understood by lengths (l_1, l_2, and l_3) of a text, the number of simple lexic units, composed lexic units and words that form it, take into account that each one of them must be counted as many times as appear in him.

In this paper it will be studied the problem of the lengths of the simple lexic units (SLUN), composed lexic units (CLUN) and words of text models, expressing these lengths in number of the primitive symbols, SLUN and syllables respectively and it is one goal to prove empirically if the law stated by Chebanov, which is fulfilled in the natural languages, also if it is satisfied in the formal language L(M).

4. THE CHEBANOV'S STATISTICAL LINGUISTIC LAW

We focus on studing the lenght of a text. For this we will used the criterion of Chebanov (1947), for Indo-European natural languages. According to said author, distribution of words according to the number of syllables follows a Poisson distribution.

The mean of the length of a word (or SLUN and CLUN), l_m is defined in terms of the following formula:

$$l_m = \sum nP(n) \quad (3)$$

where N: the number of syllables (or primitive symbols and SLUN), and P(n) is the appearance frequency of words (or SLUN and CLUN), with n syllables (or primitive symbols and SLUN). Then, between P(n) and l_m there exists the following relationship :

$$P(n) = \frac{(l_m - 1)^{n-1}}{(n-1)!} e^{-(l-1)} \qquad (4)$$

5. THE VAKAR'S FORMULA

Vakar (1966) used, for the Russian spoken language, the text of certain dramatic works admitting that these texts faithfully transcribed spoken language. The studies of that author succeeded in establishing the first 360 words of Russian common vocabulary, which allowed to evaluate limits on ranging vocabulary used in such a current conversation. Results of different linguistic backgrounds have come to show that Zipf's law is not always true. For pronouns, in the case of the Russian language, Vakar to deduce that the law is correct:

$$f_r = A \cdot 2^{-kr} \quad (5)$$

being f_r the frequency of occurrence of the letter occupying the rank r in the set of letters arranged according to the decreasing values of their frequency of occurrence; A and k denote two constants, which are characteristic of the proposed language, and number 2 only refers to the evaluation of information and is expressed in bits.

If N is the number of letters in an alphabet, we can write:

$$\sum_{r=1}^{N} f_r = 1 \quad (6)$$

Given (5), we must

$$\sum_{r=1}^{N} f_r = \sum_{r=1}^{N} A \cdot 2^{-rk} = A \sum_{r=1}^{N} 2^{-rk} = 1$$

As $\sum_{r=1}^{N} 2^{-rk}$ is an expression of the sum of a geometric progression, we will have:

$$A \frac{1-2^{-Nk}}{1-2^{-k}} 2^{-k} = 1 \quad (7)$$

For L(M), the number of charaters N has been defined according formula (2). Therefore

$$A \frac{1-2^{-Nk}}{1-2^{-k}} 2^{-k} = A \frac{1-2^{-\left(\alpha \frac{m^{n+1}-1}{m-1}\right)k}}{1-2^{-k}} 2^{-k} = 1 \quad (8)$$

In general, this relationship can be simplified having regard to the fact that while the constant k has a small value, the number of points N is sufficiently large so that the product

$$Nk = \left(\alpha \frac{m^{n+1}-1}{m-1}\right) >> 1 \quad (9)$$

Thus

$$A + 1 \approx 2^{k} \quad (10)$$

6. APPLICATION OF CHEBANOV LAW AND VAKAR FORMULA TO CASES

In order to illustrate the application of the statistical law of text we have used the MARIOLA model. We study the problem of the length of text considering:

1) Simple lexic units in function of number of primitive simbols.

2) Composed lexic units in function of number of simple lexic units.

3) Words in function of number of composed lexic units.

To prove if the distribution of LUN according to the number of primitive symbols (Sastre-Vazquez et al., 1999), follows a Poisson distribution, it was used a procedure goodness of fit based on chi squared distribution. The average parameter v = L - 1, of such Poisson distribution was estimated from the data as:

$$L = \sum n \cdot P(n) = \sum \left[n \left(\frac{f(n)}{\sum f(n)} \right) \right]$$

$$L = \left(\frac{1}{N} \right) \sum \left[n \cdot f(n) \right] = \left(\frac{1}{132} \right) \cdot 157 = 1.1893939 \qquad (11)$$

$$v = L - 1 = 0.1893939$$

Expected frequencies with F_i (n) = N p_i were calculated using a Poisson distribution with parameter 0.18, where N =132 is the total number of observations, and p_i is theoretical probability associated with class (i-1). Expected frequencies were obtained multiplying the respective probabilities by the size of the sample. The following hypothesis were tested.

H_0) $$P(n) = \frac{\left[0.18939^{(n-1)} \cdot e^{-0.18939} \right]}{(n-1)!} \qquad (12)$$

H_1) P(n) is not adjusted to Poisson with mean v = 0.18939.

The statistic was $\chi^2 = \Sigma \{ [f(n) - F(n)]^2 / F(n) \} = 34.78$ with k-p-1 = 1 degree of freedom, and since $\chi^2_{(1,\ 5\%)} = 3.84$, the hypothesis that appearance of LUN with n primitives symbols follows a Poisson distribution, is rejected.
When we carry out the same test for the lenght of CLN in function of its LUN, we obtain a similar result v=0.64 and $\chi^2 = \Sigma \{[f(n)-F(n)]^2/F(n)\} = 16{,}564$ which produces a rejetion of Poisson's distribution.
With respect the word distribution (N=23 flow equation) according to his lenght in based on the number of syllables (number of additions), we obtain $\chi^2 = \Sigma \{ [f(n) - F(n)]^2 / F(n) \} = 4.971$ with k-p-1 = 5 degrees of freedom , and since $\chi^2_{(5,\ 5\%)} = 11.07$, the hypothesis that appearance of words with n syllables follows a distribution

of Poisson, can not be rejected. Words (flow equations) have a mean length of v= 3,217 syllables.

7. CONCLUSIONS AND FUTHER REMARKS

The MARIOLA model, with all its limitations, has met the expectations placed on it. It has allowed us to calculate with sufficient credibility (relative errors no higher than 10% for the majority of variables) a series of variables that determine the development and behavior of bushes in Mediterranean ecosystems. It not only allows following such development at the ecosystem level, but can also be utilized for the monitoring of individual samples.

The MARIOLA model opens up problems that remain as yet unresolved. Mathematically, all flow equations are multiple nonlinear regressions. They are the input and outputs of ordinary differential equations. This is not a normal way of building a model. Specifically, the differential equations used in the MARIOLA model can be considered stochastic differential equations (SDEs). The Gauss method of integration was utilized. Other methods of integration (Korpelainen, 1989) are more accurate but also slower. Hence, the development of a mathematical theory of this kind of model is called for. This has already been done (Usó-Doménech et. al., 1997a).

It has been shown that for this particular signs system there exists an optimum distribution that allows a maximum of information to be transmitted with a minimum energy consumption, for the case of words (flow equations).

As occurs in natural languages, the most common words tend to be shorter while low-frequency words tend to have more syllables. In other words, the less "costly" signs of the code are combined with the most frequent concepts and conversely. However upon analyzing the lengths of the LUN and the CLUN this law is found not to apply.

A particularly simple alternative was suggested by Mandelbrot (1954). An essential notion is that spaces between words occur more or less randomly, with the result that short runs of letters (short words) will occur more frequently than long runs. There is also a greater variety of possible long words. Thus a relatively large probability must be assigned among a small number of short words, whereas only a small probability is available for the vast number of longer words. An author is considered, from this point of view, as a stochastic device for generating long sequences of letters, and consequently his words may be expected to follow the rank-frequency rule that is so commonly observed. The model predicts that the shortest words will be the most frequent, but this is not always the case. Since Mandelbrot shares Zipf's opinion that languages involve towards some maximally efficient state, and since the infrequent use of some short words is an inefficient way to communicate, he argues that letters (or phonemes) are not the proper units to measure the length of a word. "One must therefore make the weaker assumption that the structure of speech of words is influenced by some other coding, higher up in the receiving brain, considered as an optimal terminal information processing machine". Possibly, the length of the LUN and CLUN would have to be measured in function of the information they transmit.

Acknowledgments: The authors would like to thank Professor B. C. Patten (Institute of Ecology University of Georgia, Athens, Georgia. USA) and Professor Daniel Berend

(Department of Mathematics and Computer Science, University Ben-Gurion of the Negev, Israel) for their interest demonstrated in the development of the Text Theory.

REFERENCES

Bonini, Ch. P. 1963. *Simulation of information and decision systems in the firm,* Englewood Cliffs, N. J.: Prentice-Hall

Chebanov, S.G. 1947. O podchinenii rechevyj ukladov indo-evropeiskoi gruppy zakoma Puassona. *Doklady*, Akademia Nauk SSSR, 55, 2, 103-109. (In Russian).

Forrester, J., W. 1961. *Industrial Dynamics*. Massachusetts Institute of Technology Press, Cambridge, Massachusets.

Guiraud, P. 1954. *Les Caractères statístiques du vocabulaire, essai de méthodologie.* Paris. P.U.F.

Kanding, F.W. 1897. *Haudigkeitsworterbuch der deutschen sprache*. Berlin. (In German).

Korpelanien, E., 1989. Stokastien Differentiaali-Yhtälön Ratkaiseminen Runge-Kutta-Menetelmällä. Research Reports in Statistic (1). University of Joennsuu, Faculty of Social Sciences, Finland.

Mandelbrot, B. 1954. Structure formelle des textes et communication (deux études). *Word* 10, 1-27.

Mandelbrot, B. 1961. Word fequencies and Markovian models of discourse. Structure of Language and its Mathematical Aspects. In: Proccedings of Symposia in Applied Mathematics, 12. American Mathematical Society. Providenc, Rhode Island. 190-219.

Marcus, S., Nicolau, E. And Stati, S. 1966. *Introducere in lingvistica matematica*. Editura Stiintificä. Bucarest. (In Rumanian).

Margalef, R.1991. Networks in Ecology. *Theoretical studies of Ecosystems*. Cambridge University. Press. 41-57.

Nescolarde-Selva, J. and Usó-Doménech, J. L. 2013. Topological Structures of Complex Belief Systems (II): Textual Materialization. *Complexity*. Vol. 19, 2. 50-62

Sastre-Vazquez, P., J.L. Usó-Domènech, J. Mateu, P. Salvador and Y. Villacampa. 1999. Statistical Linguistic Laws in Ecological Models. *Cybernetic and Systems: an International Journal*. 30 (8). 697-724.

Stübing, G., Peris, J.B. and Costa, M., 1989. Los matorrales seriales termófilos valencianos. Phytoecoenologia, 17: 1-69.

Usó-Domenech, J.L., Villacampa, Y, Stubing, G., Karjalainen, T and Ramo, M. 1995 MARIOLA: A model for calculating the reponse of mediterranean bush ecosystem to climatic variations. *Ecological Modelling* , 80, 113-129.

Usó-Domenech, J.L., Mateu, J. and López, J.A.1997[a]. Mathematical and Statistical Formulation of an Ecological Model with Applications. *Ecological.Modelling*, 101,27-40.

Usó, J.L., Mateu, J., Villacampa, Y. and Salvador, P. 1997[b]. Adjustment versus meaning?. Towards a quantic theory of ecological models. In: Ecosystems and Sustainable Development (Edit J.L.Usó, C. Brebbia and H. Power). Serie Advances in Ecological Sciences. Vol I. Computational Mechanics Publications. Southampton-Boston. 375-393.

Usó-Domènech, J.L., Mateu, J., Villacampa, Y., and Sastre-Vazquez, P. 2000. A proposal of Uncertainty and Complementary Principles in Flow Equations of Ecological Models. *Cybernetic and Systems: an International Journal*. . 31(2), p 137-160.

Usó-Domènech, J.L., P. Sastre-Vazquez, J. Mateu. 2001. Syntax and First Entropic Approximation of L(MT): A Language for Ecological Modelling. *Kybernetes*. 30(9/10). 1304-1318.

Usó-Domènech, J.L., Sastre-Vazquez, P. 2002. Semantics of L(MT): A Language for Ecological Modelling. *Kybernetes* 31 (3/4), 561-576.

Usó-Domènech, J.L, Vives Maciá, F. and Mateu. J.. 2006[a]. Regular grammars of L(MT): a language for ecological systems modelling (I) –part I. *Kybernetes* 35 nº6, 837-850.

Usó-Domènech, J.L, Vives Maciá, F. and Mateu. J.. 2006[b]. Regular grammars of L(MT): a language for ecological systems modelling (II) –part II. *Kybernetes* 35 (9/10), 1137-1150.

Usó-Doménech, J. L. and Nescolarde-Selva, J. 2012. *Mathematic and Semiotic Theory of Ideological Systems.* Editorial LAP. Sarrebruck. Germany.

Usó-Doménech, J. L., Nescolarde-Selva, J., Lloret-Climent, M. 2014. Saint Mathew Law and Bonini Paradox in Textual Theory of Complex Models. 2 (4), pp. 89-93.

Vakar, N.P. 1966. *A Word Count of Spoken Russian: The Soviet Usage*. Ohio State University Press.

Villacampa, Y. and Usó-Domènech, J.L. 1999. Mathematical Models of Complex Structural systems. A Linguistic Vision. *Int. Journal of General Systems*. 28(1), 37-52.

Villacampa, Y., Usó-Domènech, J.L., Castro, M.A. and Sastre-Vazquez, P. 1999[a]. A Text Theory of Ecological Systems. *Cybernetic and Systems: an International Journal*. 30. 587-608.

Villacampa, Y., Usó-Domènech, J.L., Mateu, J. Vives, F. and Sastre, P. 1999[b]. Generative and Recognoscitive Grammars in Ecological Models. *Ecological Modelling*. 117, 315-332.

Zipf, G.K. 1949. *Human Behavior and the Principle of Leas Effort*. Cambridge, Massachussets. Adisson-Wessley.

ANEX A
TABLE 1
State and flow variables for MARIOLA model

a. State variables

Y	Description (unit)
BL	woody biomass (g)
BV	green biomass (g)
MOTS	total organic soil material (%)
NRO	organic material of animal origin on the ground (g)
RBL	litter of woody biomass on the ground (g)
RBV	litter of green biomass on the ground (g)

b. Flow variables

X	Description (unit)
ARRS	rate of loss of the organic soil material through dragging and washing (%)
CRBL	rate of production by growth of the woody biomass (g)
CRBV	rate of production by growth of the green biomass (g)
DBLAR	rate of destruction of the woody biomass through the action of arthropods (g)
DBLPL	rate of destruction of the woody biomass through the action of phytoplagues (g)
DBVFS	rate of destruction of the green biomass through the action of mammals (g)
DBVI	rate of destruction of the green biomass through the action of insects (g)
DBVPL	rate of destruction of the green biomass through the action of phytoplasgues (g)
DCBL	rate of catastrophic destruction of the woody biomass (g)
DCBV	rate of catastrophic destruction of the green biomass (g)
DF	rate of defoliation (g)
DMOTS	rate of decomposition of the total organic soil material (%)
DRBL	rate of decomposition of the litter of the woody biomass on the soil (g)
DRBV	rate of decomposition of the litter of the green biomass on the soil (g)
DRO	rate of decomposition of the detritus of an animal narure (g)
MOFD	rate of finely divided organic material (%)
PMOTS	rate of production of organic soil material (humus) (%)
PRO2	rate of production of organic detritus of animal origin (g)
VMN	rate of destruction of the woody biomass (g)

c. Exogenous variables

e	Description (unit)
H	environmental humidity (%)
IFAP	maximum intensity of precipitation (max.l/h)
PLU	precipitation(l)
POBHV	population of mammals (Oryctolagus cuniculus) (number of individuals)
T	environmental temperature (ºC)
VEVI	wind speed (km/h max)

d.-Auxiliary variables and parameters

a	Description (unit)
BT	total biomass (g)
CRO2	parameter of residual production of the rodents (g)
PORDT	the herbivore diet (%)

TABLE 2

Text for of MARIOLA model

First level

$$\frac{dBV}{dt} = CRBV - DF - DCBV - DBVFS - DBVI - DBVPL$$

$$\frac{dBL}{dt} = CRBL - VMN - DCBL - DBLAR - DBLPL$$

$$\frac{dRBV}{dt} = DF + DCBV - DRBV$$

$$\frac{dRBL}{dt} = VMN + DCBL - DRBL$$

$$\frac{dNRO}{dt} = PRO2 - DRO$$

$$\frac{dMOTS}{dt} = PMOTS + MOFD - DMOTS - ARRS$$

Second level

dBV/dt = $[a_1 T \cdot BV + b_1 H \cdot BV + c_1 \cdot BV + a_1 T \cdot BL + b_1 H \cdot BL + c_1 \cdot BL + d_1 PLU + f_1]$-

-$[a_3 BV + b_3 BV + c_3 BV^2 .T + d_3 BV.H + f_3 (\frac{1}{PLU}) + g_3]$-

-$[a_5 BV.IFAP + b_5 BV.VEVI + c_5 e^{(0.1IFAP)} + d_5]$-$[a_4 BV^2 + b_4 BV.POBHV + c_4]$- DBVI-1

dBL/dt =$[a_2 BV + a_2 BL + b_2 BV \cdot H + b_2 BL \cdot H + c_2 BV \cdot PLU + c_2 BL \cdot PLU + d_2]$--$[a_7 BL + b_7 BL.PLU + c_7]$-

$[a_8 \cos BL + b_8 BL.IFAP + c_8 BL.VEVI + d_8]$-DBLAR-$[a_9 BL.T + b_9 e^{(-0.1BL)} + c_9 \cos BL + d_9]$

dRBV/dt =

$$[a_3BV + b_3BV + c_3BV^2.T + d_3BV.H + f_3(\frac{1}{PLU}) + g_3] ++ [a_5BV.IFAP + b_5BV.VEVI + c_5e^{(0.1IFAP)} + d_5] -$$
$$- [a_{10}T^2 + b_{10}T.RBV + c_{10}H.RBV + d_{10}e^{(0.1H)} + f_{10}]$$

dRBL/dt = $[a_7BL + b_7BL.PLU + c_7] + [a_8 \cos BL + b_8BL.IFAP + c_8BL.VEVI + d_8] -$
$$- [a_{11}T^2 + b_{11}T.H + c_{11}e^{(0.1T)} + d_{11} \cos H + f_{11}]$$

dNRO/dt = $[POBHV.f20.\{f18.\log(a4BV^2 + b4BV.POBHV + c4)\}/100)] -$
$$[a_{12}NRO.T + b_{12}T^2 + c_{12} \cos NRO + d_{12}]$$

dMOTS/dt= $[\left(a_{13}T^2 + b_{13}T.H + c_{13}T.DBL + d_{13}T.DBV + f_{13} \cos T + g_{13}DRO + h_{13}\right)/100] +$
$$[(a_{14}T^2 + b_{14}T.H + c_{14}e^{(0.1T)} + d_{14}e^{(0.1H)} + f_{14})/100] -$$
$$[\left[MOTS(a_{15}MOTS + b_{15}T + c_{15}H + d_{15}PLU) + f_{15}T^2 + g_{15}\right]/100] -$$
$$[(a_{16}T^2 + b_{16}T.H + c_{16}T.PLU + d_{16}H^2 + f_{16})/100]$$

being a_i, b_i, c_i, d_i, f_i, g_i, h_i, i=1,2,...,16 parameters

Third level

dBV/dt = $a_1T \cdot BV + b_1H \cdot BV + c_1 \cdot BV + a_1T \cdot BL + b_1H \cdot BL + c_1 \cdot BL + d_1PLU + f_1$ --
$$\left[a_3BV + b_3BV + c_3BV^2.T + d_3BV.H + f_3(\frac{1}{PLU}) + g_3\right] - \left[a_5BV.IFAP + b_5BV.VEVI + c_5e^{(0.1IFAP)} + d_5\right] -$$
$$[a_4BV^2 + b_4BV.POBHV + c_4]$$ - DBVI- $\left[a_6BV^2 + b_6BV.T + c_6 \cos H + d_6\right]$

dBL/dt = $[a_2BT + b_2BT \cdot H + c_2BT \cdot PLU + c_2BL \cdot PLU + d_2]$ -- $[a_7BL + b_7BL.PLU + c_7]$ -
$\left[a_8 \cos BL + b_8BL.IFAP + c_8BL.VEVI + d_8\right]$ -DBLAR- $-\left[a_9BL.T + b_9e^{(-0.1BL)} + c_9 \cos BL + d_9\right]$

dRBV/dt =

$$[a_3BV + b_3BV + c_3BV^2.T + d_3BV.H + f_3(\frac{1}{PLU}) + g_3] ++ [a_5BV.IFAP + b_5BV.VEVI + c_5e^{(0.1IFAP)} + d_5] -$$
$$- [a_{10}T^2 + b_{10}T.RBV + c_{10}H.RBV + d_{10}e^{(0.1H)} + f_{10}]$$

dRBL/dt = $[a_7BL + b_7BL.PLU + c_7] + [a_8 \cos BL + b_8BL.IFAP + c_8BL.VEVI + d_8] -$
$$[a_{11}T^2 + b_{11}T.H + c_{11}e^{(0.1T)} + d_{11} \cos H + f_{11}]$$

dNRO/dt = $[POBHV.CRO2.PORDT/100)] - [a_{12}NRO.T + b_{12}T^2 + c_{12} \cos NRO + d_{12}]$

$$dMOTS/dt = [\left(a_{13}T^{2} + b_{13}T.H + c_{13}T.DBL + d_{13}T.DBV + f_{13}\cos T + g_{13}DRO + h_{13}\right)/100] +$$
$$[(a_{14}T^{2} + b_{14}T.H + c_{14}e^{(0.1T)} + d_{14}e^{(0.1H)} + f_{14})/100] -$$
$$[\left[MOTS(a_{15}MOTS + b_{15}T + c_{15}H + d_{15}PLU) + f_{15}T^{2} + g_{15}\right]/100] -$$
$$-[(a_{16}T^{2} + b_{16}T.H + c_{16}T.PLU + d_{16}H^{2} + f_{16})/100]$$

being $a_i, b_i, c_i, d_i, f_i, g_i, h_i$, i=1,2,...,16 parameters

In Table 3 are shown, for each considerate level, the corresponding frequencies from the different primitive symbols and LUN for the MARIOLA models.

TABLE 3
Frequencies in MARIOLA model.

First Level						Second Level						Thrid Level					
Primitive simbols			Lexic Units			Primitive simbols			Lexic Units			Primitive simbols			Lexic Units		
s	r	f	s	r	f	s	r	f	s	r	f	s	r	f	s	r	f
d/dt	1	6	DF	1	2	T	1	31	T	26	1	T	1	31	T	1	26
DF	2	2	DCBVDCB	2	2	BV	2	30	BV	25	2	H	2	24	BV	2	21
DCBV	3	2	LVMN	3	2	BL	3	23	BL	18	3	BV	3	21	H	3	19
DCBL	4	2	DBLPLDB	4	2	H	4	17	H	14	4	BL	4	16	BL	4	11
VMN	5	2	LARDBVFS	5	1	PLU	5	9	PLU	9	5	PLU	5	9	PLU	5	9
DBLPL	6	1	DBVPLDB	6	1	e	6	7	MOTS	5	6	e	6	7	BT	6	6
DBLAR	7	1	VIDbv/dt	7	1	cos	7	6	IFAP	4	7	cos	7	7	MOTS	7	5
DBVFS	8	1	DBL/dt	8	1	d/dt	8	6	VEVI	4	8	BT	8	6	IFAP	8	4
DBVPL	9	1	CRBL	9	1	IFAP	9	6	POBHV	3	9	d/dt	9	6	VEVI	9	4
DBVI	10	1	ARRS	10	1	MOTS	10	6	CosBL	3	10	IFAP	10	6	e01T	10	3
BV	11	1	CRBV	11	1	VEVI	11	4	e01H	2	11	MOTS	11	6	e01H	11	3
BL	12	1	MOFD	12	1	POBHV	12	4	e01T	2	12	VEVI	12	4	cosBL	12	3
CRBL	13	1	DRO	13	1	RBV	13	3	e01IFAP	2	13	RBV	13	3	POBHV	13	2
ARRS	14	1	DMOTS/dt	14	1	NRO	14	3	dMOTS/dt	1	14	NRO	14	3	e01IFAP	14	2
CRBV	15	1	PMOTS	15	1	RBV	15	3	dRVB/dt	1	15	POBHV	15	2	cosH	15	2
MOFD	16	1	DNRO/dt	16	1	DBLAR	16	1	dBV/dt	1	16	PORDT	16	1	dBV/dt	16	2
DRO	17	1	DRBVI	17	1	DRO	17	1	dBL/dt	1	17	DRO	17	1	RBV	17	2
MOTS	18	1	DRBL/dt	18	1	DBVI	18	1	cosNRO	1	18	DCBL	18	1	Drbv/dt	18	1
PMOTS	19	1	dRBV/dt	19	1	DCBL	19	1	cosT	1	19	DBVI	19	1	DMOTS/dt	19	1
NRO	20	1	PRO2	20	1	DCBV	20	1	log()	1	20	DBLAR	20	1	DNRO/dt	20	1
DRBVI	21	1	DRBL	21	1				cosH	1	21	DCBV	21	1	CosT	21	1
RBL	22	1	DMONTS	22	1				DBV	1	22	CRO	22	1	CosNRO	22	1
RBV	23	1		23	1				DBLAR	1	23				e-01BL	23	1
PRO	24	1		24	1				DRO	1	24				dBL/dt	24	1
DRBL	25	1							DNRO/dt	1	25				DRO	25	1
DMONTS	26	1							e-01BL	1	26				DCBL	26	1
									RBV	1	27				NRO	27	1
									DRBL/dt	1	28				PORDT	28	1
									DBL	1	29				DBVI	29	1
									DCBV	1	30				CRO2	39	1
									DBVI	1	31				DRBL/dt	31	1
									DCBL						DBLAR	32	1
															DCBV		