Ecological Complexity 21 (2015) 27–33

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

Chebanov law and Vakar formula in mathematical models of complex systems

Josué-Antonio Nescolarde-Selva [a,*], Josep-Lluis Usó-Doménech [a], Miguel Lloret-Climent [a], Lucía González-Franco [b]

[a] Department of Applied Mathematics, University of Alicante, Alicante, Spain
[b] Biodiversity Research Institute CIBIO, University of Alicante, Alicante, Spain

ARTICLE INFO

Article history:
Received 3 June 2014
Received in revised form 20 November 2014
Accepted 22 November 2014
Available online 13 January 2015

Keywords:
Chebanov law
Ecological model
Language
MARIOLA model
Vakar formula

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; and (d) 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.

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., 1997b, 2001; Usó-Domènech and Sastre-Vazquez, 2002; Usó-Domènech et al., 2006a,b; Villacampa and 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:

* Corresponding author. Tel.: +34 680418381; fax: +34 965 909 707.
E-mail address: josue.selva@ua.es (J.-A. Nescolarde-Selva).

http://dx.doi.org/10.1016/j.ecocom.2014.11.003
1476-945X/

(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 labelled 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 and Usó-Domènech, 1999) adapting them to the mathematical and ecological context (Villacampa et al., 1999a,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 $f * r = \text{constant}$ (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 words will 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 1000 most common words will cover 85%, and the 4000 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-Domènech 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, 1963; Usó-Doménech et al., 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., *Ulex 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 necessarily linear 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. (10) 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