Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Physica A 368 (2006) 257–261

PHYSICA A

www.elsevier.com/locate/physa

Bounded fitness landscapes and the evolution of the linguistic diversity

Viviane M. de Oliveira[a,*], Paulo R.A. Campos[b], M.A.F. Gomes[a], I.R. Tsang[c]

[a]*Departamento de Física, Universidade Federal de Pernambuco, 50670-901 Recife, PE, Brazil*
[b]*Departamento de Física e Matemática, Universidade Federal Rural de Pernambuco, 52171-900 Dois Irmãos, Recife, PE, Brazil*
[c]*Centro de Informática, Universidade Federal de Pernambuco, 50670-901 Recife, PE, Brazil*

Received 27 October 2005; received in revised form 16 November 2005
Available online 28 February 2006

Abstract

We have recently introduced a simple spatial computer simulation model to study the evolution of the linguistic diversity. The model considers processes of selective geographic colonization, linguistic anomalous diffusion and mutation. In the approach, we ascribe to each language a fitness function which depends on the number of people that speak that language. Here, we extend the aforementioned model to examine the role of saturation of the fitness on the language dynamics. We found that the dependence of the linguistic diversity on the area after colonization displays a power law regime with a nontrivial exponent in very good agreement with the measured exponent associated with the actual distribution of languages on the Earth.

Keywords: Linguistic diversity; Languages; Language competition

1. Introduction

The research in language dynamics has raised an increasing interest of the complex systems community in the last years [1]. Most of the researchers focus their investigations on issues like rise, competition, extinction risk and death of languages [2–12]. Furthermore, recent advances in archeology, genetics and linguistics have provided relevant contributions to a better comprehension of the linguistic diversification [13,14]. Some investigations have demonstrated that distinct causes have greatly affected the evolution of the linguistic diversity. Among the main elements are geographic factors, economic features, complexity of the language, to cite just a few. For instance, Sutherland [2] has shown that beside country area, forest area and maximum altitude contribute to increase diversity, whereas the diversity decreases for a larger latitude. According to Bellwood [15,16] and Renfrew [17,18] the occurrence of agricultural expansion was responsible for the massive population replacements initiated about 10,000 years ago and caused the disappearance of many of the Old World languages.

---

[*] Corresponding author.
E-mail address: viviane@df.ufpe.br (V.M. de Oliveira).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.11.058

In a recent work, we investigated the evolution of the linguistic diversity by introducing a spatial computer simulation model that considers a diffusive process which is able to generate and sustain the diversity [1]. The model describes the occupation of a given area by populations speaking several languages. To each language was assigned a fitness value f which is proportional to the number of sites colonized by populations that speak that language. In the process of colonization, language mutation or differentiation and language substitution can take place, which affords the linguistic diversity. This simple model gives rise to scaling laws in close resemblance with those reported in Ref. [19].

In the current contribution, we study the dynamics of the linguistic diversity but now we assume that the fitness of each language is bounded by a given maximum (saturation) value which is randomly chosen from a uniform distribution. The saturation hypothesis mimics factors like the difficulty/ease of learning the languages and economy that permit some languages to propagate more easily than others.

The paper is organized as follows. In Section 2 we introduce the model. In Section 3 we discuss the results. And finally, in Section 4 we present the conclusions.

2. Model

Our model is defined on a two-dimensional lattice of linear size L, and composed of $A = L \times L$ sites with periodic boundary conditions. Each lattice site s_i represents a given region, which can be occupied by a single population speaking just one language. We ascribe to each site a given capability C_i, whose value we estimate from a uniform distribution, defined in the interval 0–1. The capability means the amount of resources available to the population which will colonize that place. It is implicit that the population size in each cell s_i is proportional to its capability C_i.

In the first step of the dynamics, we randomly choose one site of the lattice to be colonized by a single population that speaks the ancestor language, while all other sites are empty. Each language is labeled by an integer number. As soon as a new language arises, it is labeled by the next higher integer. To each language, we assign a fitness value f, which is calculated as the sum of the capabilities of the sites which speak that specific language. Thus, the initial fitness of the ancestor language is the capability of the initial site. But now differently from Ref. [1], the fitness cannot exceed an integer value γ_k which we have chosen to be in the range 1–2000. This saturation term γ_k is randomly chosen when the language k appears.

In the second step, one of the four nearest neighbors of the site containing the ancestor language will be chosen to be colonized. It will be selected with probability proportional to its capability. We assume that regions containing larger amount of resources are most likely to be colonized faster than poor regions. The referred site is then occupied by a population speaking the ancestor language or a mutant version of it. Mutations are the mechanisms responsible for generating diversity, and together with the natural selection maintains diversity on the system. The probability of occurrence of a mutation in the process of propagation is $p = \alpha/f$, where α is a constant, and so the mutation probability is inversely proportional to the fitness of the language. The form of the mutation probability p is inspired by population genetics, where the most adapted organisms are less likely to mutate than poorly adapted organisms [20]. The probability of producing reverse mutations is zero, that is, the language generated by a mutation is always different of the previous ones.

In the subsequent steps, we check the empty sites which are located on the boundary of the colonized cluster, and we then choose one of those empty sites according to their capabilities. Again, those sites with higher capabilities enjoy of a greater likelihood to be occupied. After that, we choose the language to be incorporated in the chosen cell among those languages occupying the neighboring sites. Languages with higher fitness have higher chance to expand. The process continues while there are empty sites in the network. After completion, we count the total number of languages D.

3. Results and discussion

In Fig. 1, we show the diversity D as a function of the area A (total number of sites in the lattice) for mutation parameter $\alpha = 0.3$ and saturation values defined in the interval 1–2000. The points are averages over 100 independent simulations when $L < 400$ and over 20 simulations when $L = 500$. We observe that the curve presents just one scaling region which extends over five decades. The exponent $z = 0.39 \pm 0.01$ is in quite