Physica A 387 (2008) 5597–5601

Contents lists available at ScienceDirect

Physica A

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

Evolution of the linguistic diversity on correlated landscapes

E.J.S. Silva, Viviane M. de Oliveira *

Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, 52171-900, Recife-PE, Brazil

ARTICLE INFO

Article history:
Received 11 February 2008
Received in revised form 2 June 2008
Available online 12 June 2008

PACS:
87.23.Cc
89.75.-k
89.75.Da

Keywords:
Language size distribution
Linguistic diversity
Languages
Language competition

ABSTRACT

We have recently investigated the evolution of linguistic diversity by means of a simple spatial model that considers selective geographic colonization, linguistic anomalous diffusion and mutation. In the model, regions of the lattice are characterized by the amount of resources available to populations which are going to colonize the region. In that approach, the resources were ascribed in a randomly and uncorrelated way. Here, we extend the previous model and introduce a degree of correlation for the resource landscape. A change of the qualitative scenario is observed for high correlation, where the increase of the linguistic diversity on area is faster than for low correlated landscapes. For low correlated landscapes, the dependence of diversity on area shows two scaling regimes, while we observe the rising of another scaling region for high correlated landscapes.

1. Introduction

The evolution of linguistic diversity has been an issue of increasing interest in the last years [1–14]. Most of the investigations are focused in the rise, competition, extinction risk and death of languages. Some of these have demonstrated that the language diversity is affected by elements like geographic factors, climatic variability and economic features [15,16]. In particular, it was observed that the language diversity is high where the climate favors food production during the whole year, while it is low where the variability of the climate is great [15].

Gomes et al [17] analyzed approximately six and a half thousand languages on Earth and showed that the language diversity scales with area, according to a power law $D \sim A^z$, where $z = 0.41 \pm 0.03$. They also observed that the number of languages n spoken by a population of size larger than N, $n(> N)$, displays composite power laws: $n(> N)\, N^{-\tau}$, in which $\tau = 0.5$ for $5 \times 10^4 < N < 6 \times 10^6$ and $\tau = 1.0$ for $2 \times 10^7 < N < 1 \times 10^9$.

We have recently investigated the evolution of linguistic diversity by introducing a spatial computer simulation that considers a diffusive process which is able to generate and sustain the linguistic diversity [1]. The model borrows ideas well established in the population genetics literature [18–21]. It 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 can take place with probability $p = \alpha/f$, in which α is a constant. This model gives rise to scaling laws in close resemblance with those reported by Gomes et al [17]. In particular, we have found that the dependence of the linguistic diversity on the area displays two power law regimes described by critical exponents, which are dependent on the mutation probability. When $\alpha = 0.3$, we have found the exponent $z = 0.43 \pm 0.02$ for small areas and $z = 0.14 \pm 0.02$ for large areas, whereas for $\alpha = 0.73$, we have estimated the exponent $z = 0.88 \pm 0.01$ for small areas and $z = 0.35 \pm 0.03$ for large areas. We have also studied the dynamics of

* Corresponding author.
E-mail addresses: eduardo_ufrpe@yahoo.com.br (E.J.S. Silva), viviane@deinfo.ufrpe.br, vivoliveira@gmail.com (V.M. de Oliveira).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2008.06.005

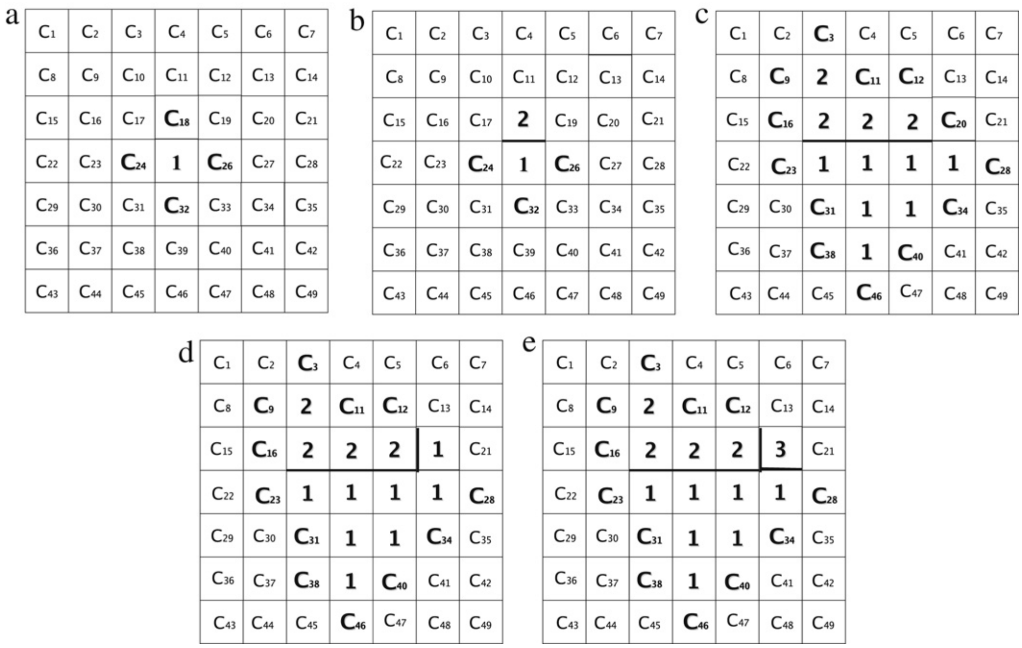

Fig. 1. In a lattice composed by $A = 7^2$ sites we show: (a) First site to be colonized by the ancestor language (labeled by number 1) and its four nearest neighbors (time $t = 1$); (b) The occurrence of a mutation giving rise to the language 2 (time $t = 2$); (c) Cluster whose sites were colonized by the populations speaking languages 1 and 2 and its boundary at time $t = 11$; (d) Increasing of the language labeled by the integer 1 (time $t = 12$); (e) The rising of a new language which is labeled by the integer 3 (time $t = 12$).

the linguistic diversity assuming that the fitness of each language is bounded by a given maximum (saturation) value which is randomly chosen from a uniform distribution [2]. This saturation term was introduced in order to mimic factors like the difficulty of learning the languages and economy that permit some languages to propagate more easily than others. In this case, we have found that the dependence of the linguistic diversity on the area displays just one power law regime with $z = 0.39 \pm 0.01$, which is in very good agreement with the actual distribution of languages on Earth.

In this work, we extend the aforementioned model by introducing a degree of correlation among the resources of the regions that can be occupied by the populations, in such way that we can vary the degree of heterogeneity of the landscape.

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. The model

Our model is defined on a two-dimensional lattice composed by $A = L \times L$ sites with periodic boundary conditions. Each lattice site s_i represents a region which can be occupied by a population speaking just one language. We ascribe to each site a given capability $C_i = e^{-x_i}$, that means the amount of resources available to the population which will colonize that place. It is expected that the population size in each cell is proportional to its capability. The value of x_i is estimated in the following way. First, we randomly choose one site s_i of the lattice and estimate x_i for this site from a normal distribution with mean zero and variance one. In the following, we estimate the values of x_j of its four nearest neighbors from a normal distribution with mean λx_i and variance $1 - \lambda^2$, where λ is the correlation parameter among the capabilities of the sites and is defined in the interval [0, 1]. To estimate the values of x_j for all other sites we take as reference the value of x_i obtained for one of their nearest neighbors. After ascribing the values of x_i to all sites, we take $C_i = e^{-x_i}$ as the capability of each site. When the correlation parameter λ is equal to 0, $\lambda = 0$, we have a completely uncorrelated landscape, whereas $\lambda = 1$ means that all sites have the same capability (maximum correlation). Intermediate degree of correlation is obtained for $0 < \lambda < 1$. This same procedure has been taken to study correlated fitness landscapes in population genetics [22].

In the first step of the dynamics, which comes after attributing the capabilities to all sites on the lattice, we randomly choose one site of the lattice to be colonized by a population that speaks the ancestor language, which is labeled as language one. To each language, we assign a fitness value f which is defined as the sum of the capabilities of the sites containing populations which speak that specific language. 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 first appears. It is introduced in order to take into account the factors that allow some languages to propagate more easily than others.

In the second step, one of the four nearest neighbors of this site will be chosen to be colonized, which occurs with likelihood proportional to its capability (see Fig. 1a). This means that regions containing a larger amount of resources are