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 © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.06.005

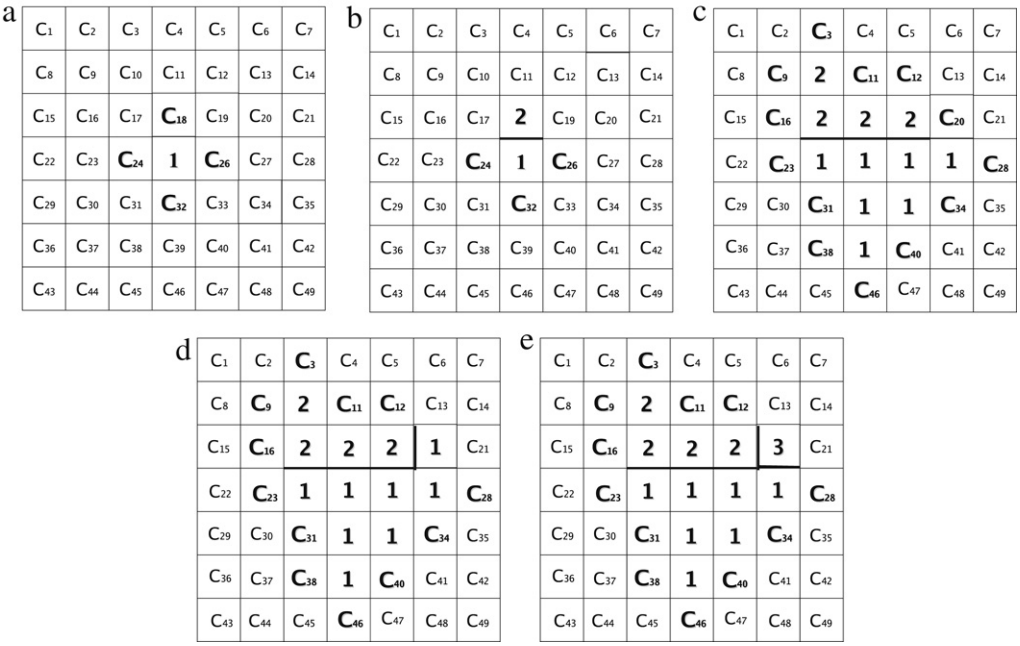

Fig. 1. In a lattice composed by $A = 7^2$ sites we show: (a) First site to be colonized by the ancestor language (labeled by number 1) and its four nearest neighbors (time $t = 1$); (b) The occurrence of a mutation giving rise to the language 2 (time $t = 2$); (c) Cluster whose sites were colonized 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 $C_i = \mathrm{e}^{-x_i}$, 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 $C_i = \mathrm{e}^{-x_i}$ 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

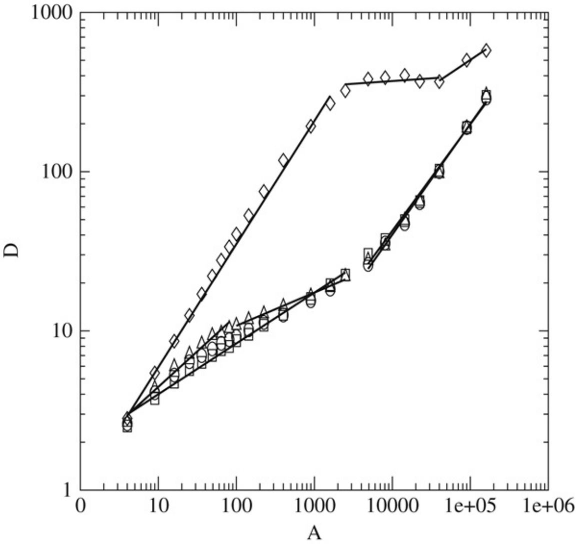

Fig. 2. Diversity D as a function of the area A for $\alpha = 0.7$ and $\lambda = 0(\square)$, 0.5 ($\circ$), 0.7 ($\triangle$) and 0.99 ($\diamond$). The exponent obtained for $\lambda = 0$ and $\lambda = 0.5$ is $z = 0.39 \pm 0.01$, while for $\lambda = 0.99$ we have two scaling regimes with exponents $z = 0.78 \pm 0.02$ and $z = 0.14 \pm 0.03$.

more 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. If a mutation occurs, we label the new language by the number 2 (Fig. 1b). The assumption of mutations mimics the situation of divergence of languages. The probability of occurrence of a mutation in the process of propagation of the language is $p = \frac{\alpha}{f}$, where α is a constant, and so the mutation probability is inversely proportional to the fitness of the language. This rule for the mutation probability was inspired by population genetics [23].

In the subsequent steps, we check what are the empty sites, which are on the boundary of the colonized cluster, and we choose one of those empty sites according to their capabilities (Fig. 1c). Those ones with higher capabilities have a higher likelihood to be occupied. We then choose the language to occupy the cell among their neighboring sites. The languages with higher fitness have a higher chance to colonize the cell (Fig. 1d). If a mutation occurs, the new language will be labeled by number 3 (Fig. 1e). This process will continue up to all sites are colonized. Next we present our simulation results.

3. Results and discussion

In Fig. 2 we show the diversity D as a function of the area A. The value of D is obtained by counting the total number of languages that were generated after the whole lattice was colonized. In the figure we have kept fixed $\alpha = 0.7$ and have considered four different values of the correlation parameter λ. We obtain each point by taking averages over 1000 independent simulations for $L < 300$ and over 500 for $L = 300, 400$. We notice from the figure the existence of two scaling regions over at least two decades, where $D \sim A^z$ for small and intermediate values of λ $(0 < \lambda < 0.5)$. The observed exponents were approximately the same in this range of correlation. We estimate the exponent $z = 0.32 \pm 0.01$ for the first region $(4 < A < 2500)$ and $z = 0.67 \pm 0.03$ for the second region $(2500 < A < 160{,}000)$. We notice from those results that the diversity increases faster with area in the region of large areas. The opposite behavior is observed when large values of λ are considered. The figure also shows the rising of another scaling region when λ is increased. For instance, for $\lambda = 0.7$ we find $z = 0.45 \pm 0.03$ (which is in good agreement with the exponent observed for the actual distribution of languages on Earth) for small areas $(4 < A < 81)$, $z = 0.21 \pm 0.01$ for intermediate areas $(81 < A < 2500)$ and $z = 0.70 \pm 0.03$ for large areas $(2500 < A < 160{,}000)$.

Fig. 3 displays the dependence of the exponent z on the correlation parameter λ for fixed $\alpha = 0.7$ in the first scaling regime (small values of A). We observe that z is approximately constant for small and intermediate values of λ, and start to grow when $\lambda > 0.5$, meaning a more quick increase of the diversity for large values of correlation. The same qualitative scenario is obtained for distinct values of α. The reasoning for the occurrence of such behavior is as follows. When the degree of correlation of the landscape is not high, the initial fitness of the first languages are very different. Thus, languages that possess high values of fitness will grow rapidly in the population and, consequently, the chance they give rise to new languages is very small since the mutation probability is inversely proportional to their fitness. When the landscape is highly correlated, the initial fitness of the languages which appear in the beginning of the dynamics is approximately the same. In this way, the fitness does not grow fast and so the diversity is higher than we find for low correlated landscapes.

In the following we have investigated the time evolution of the average area $\bar{A}(t)$ occupied by a typical language in such way to characterize the diffusion process. According to Ref. [1], the diffusion exponent d_w is given by $d_w = 2/(1 - z)$. Therefore d_w takes the standard Brownian value $d_w^B = 2$ when $z = 0$. In the interval $0 < z < 1$ we have $d_w > 2$ and the

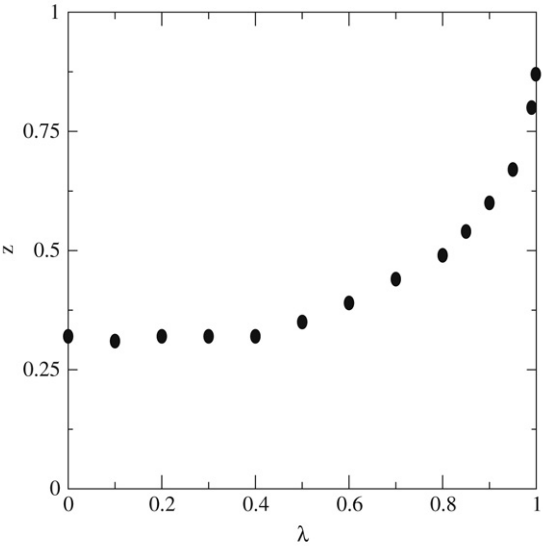

Fig. 3. Exponent z as a function of λ for $\alpha = 0.7$.

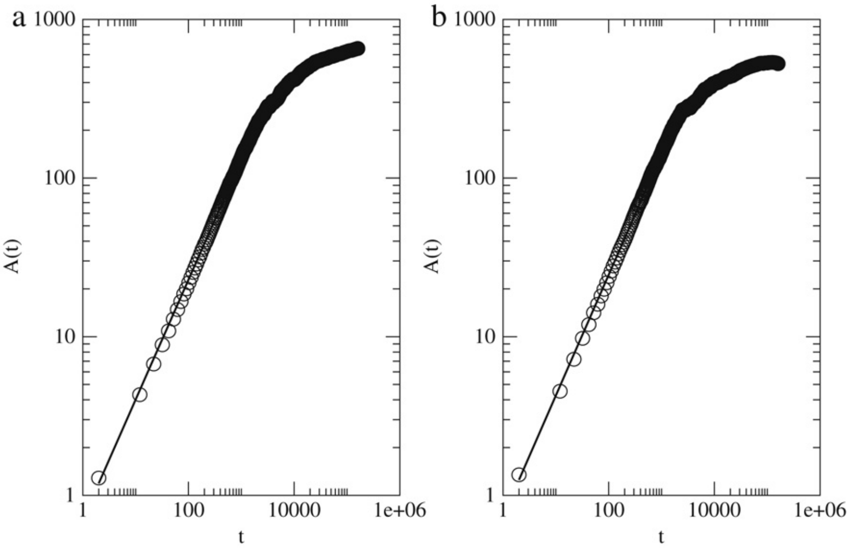

Fig. 4. Evolution of $\bar{A}$ for $\alpha = 0.7$ and (a) $\lambda = 0$ and (b) $\lambda = 0.99$.

diffusion is anomalous. The evolution of $\bar{A}(t)$ is displayed in Fig. 4 for 300 independent runs where $\alpha = 0.7$, $L = 400$ and $\lambda = 0$ and $\lambda = 0.99$. From the figure we notice the existence of two scaling regimes. The former for $A < 2,500$, where we have estimated $d_w = 2.63$ for the two values of λ, which characterizes an anomalous diffusion and the second scaling regime for $20{,}000 < A < 160{,}000$ where $d_w = 16.7$ for $\lambda = 0$ and $d_w = 18.18$ for $\lambda = 0.99$, which means a progressive hardness to promote diffusion with time.

In Fig. 5 we present the number of languages with population size greater than N, $n(> N)$ as a function of N. The data points are estimates over 350 independent runs for parameter values $L = 400$, $\alpha = 0.7$, $\lambda = 0$ and $\lambda = 0.99$. In order to calculate the population size, we have assumed that the population in a given site is its capability multiplied by a factor 100. The results are not sensitive to the particular choice of the multiplicative factor. We find one scaling regime $n(> N) \sim N^{-\tau}$, which takes place over three decades, where $\tau = 0.37 \pm 0.01$ for both values of λ. For large areas we do not observe a scaling regime.

4. Summary and conclusion

We have investigated the role of correlation among regions on the evolution of the linguistic diversity in a simple spatial model. We have seen that the relationship between diversity and area exhibits two scaling regimes along two decades for

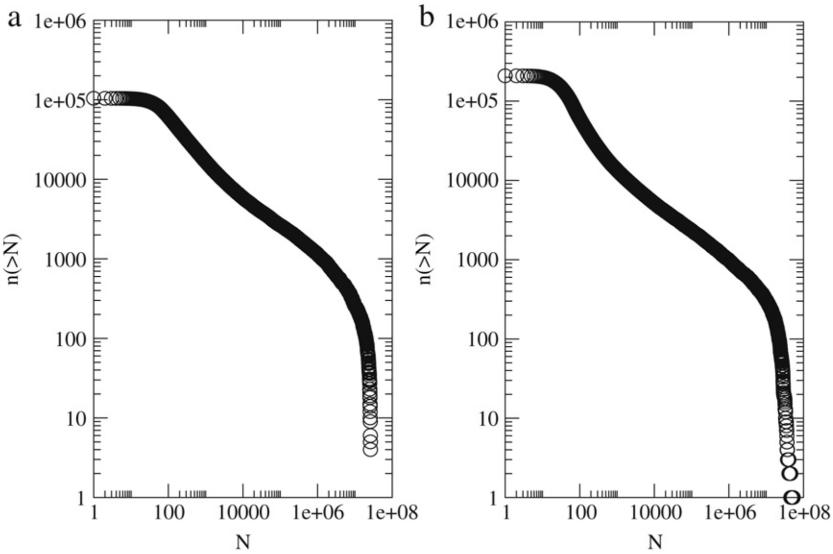

Fig. 5. Number of languages with population greater than N, $n(>N)$, as a function of N for $\alpha = 0.7$ and (a) $\lambda = 0$ and (b) $\lambda = 0.99$. $n(>N) \sim N^{-\tau}$ with $\tau = 0.37 \pm 0.01$ for $2{,}000 < A < 1{,}000{,}000$ and for both values of λ.

small and intermediate correlation. When high values of correlation are considered, we observe the occurrence of three distinct scaling regimes, and also a growth of the linguistic diversity on area when compared to the low correlated case. From those results, we can conclude that the linguistic diversity increases faster when the languages have approximately the same initial conditions to propagate than the case where the landscape is heterogeneous, and the facilities or difficulties faced by the languages depend on the particular site where they raised.

Acknowledgments

E. J. S. Silva is supported by Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE). V. M. de Oliveira is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and by the program PRONEX/MCT-CNPq-FACEPE.

References

[1] V.M. de Oliveira, M.A.F. Gomes, I.R. Tsang, Physica A 361 (2006) 361.
[2] V.M. de Oliveira, P.R.A Campos, M.A.F. Gomes, I.R. Tsang, Physica A 368 (2006) 257.
[3] D.M. Abrams, S.H. Strogatz, Nature 424 (2003) 900.
[4] M. Patriarca, T. Leppänen, Physica A 338 (2004) 296.
[5] J. Mira, A. Paredes, Europhys. Lett. 69 (2005) 1031.
[6] V. Schwämmle, Internat. J. Modern Phys. C 16 (2005) 1519.
[7] K. Kosmidis, J.M. Halley, P. Argyrakis, Physica A 353 (2005) 595.
[8] C. Schulze, D. Stauffer, Internat. J. Modern Phys. C 16 (2005) 781.
[9] A. Cangelosi, D. Parisi (Eds.), Simulating the Evolution of Language, Springer-Verlag, Berlin, 2002.
[10] J.P. Pinasco, L. Romanelli, Physica A 361 (2006) 355.
[11] C. Schulze, D. Stauffer, Physica A 379 (2) (2007) 661.
[12] D. Stauffer, Physica A 374 (2007) 835.
[13] D. Stauffer, Physica A 371 (2006) 719.
[14] P.M.C. de Oliveira, D. Stauffer, S. Wichmann, S.M. de Oliveira, Preprint: 0709.0868 on arXiv.org.
[15] D. Nettle, J. Anthropol. Archaeol. 17 (1998) 354.
[16] W.J. Sutherland, Nature 423 (2003) 276.
[17] M.A.F. Gomes, G.L. Vasconcelos, I.J. Tsang, I.R. Tsang, Physica A 271 (1999) 489.
[18] I. Gordo, P.R.A. Campos, Genetica 127 (2006) 217.
[19] L. Perfeito, I. Gordo, P.R.A. Campos, Eur. Phys. J. B 51 (2006) 301.
[20] O.F.F. Raposo, A.W.S. de Almeida, A.J.F. Souza, P.R.A. Campos, Eur. Phys. J. B 50 (2006) 653.
[21] E.A. Gonçalves, V.M. de Oliveira, A. Rosas, P.R.A. Campos, Eur. Phys. J. B 59 (2007) 127.
[22] C.O. Wilke, P.R.A Campos, J.F. Fontanari, J. Exp. Zool. 294 (3) (2002) 274.
[23] N.H. Barton, in: D. Otte, J.A. Endler (Eds.), Speciation and Its Consequences, Sinauer Associates, Sunderland, MA, 1989, p. 229.