

EDUARDO JOSÉ DE SOUZA SILVA

**MODELO ESTOCÁSTICO PARA EVOLUÇÃO DA DIVERSIDADE LINGÜÍSTICA
EM RELEVOS CORRELACIONADOS**

RECIFE-PE – FEVEREIRO/2008.



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**MODELO ESTOCÁSTICO PARA EVOLUÇÃO DA DIVERSIDADE LINGÜÍSTICA
EM RELEVOS CORRELACIONADOS**

Dissertação apresentada ao
Programa de Pós-Graduação em
Biometria e Estatística Aplicada como
exigência parcial à obtenção do título de
Mestre.

**Área de Concentração: Modelagem
Estatística Computacional (com ênfase
nas áreas agrárias, biológicas e
humanas)**

Orientador(a): Profa. Dra. Viviane Moraes de Oliveira

Co-orientador(a): Prof. Dr. Paulo Roberto de Araújo Campos

RECIFE-PE – FEVEREIRO/2008.

FICHA CATALOGRÁFICA

S586m Silva, Eduardo José de Souza
Modelo estocástico para evolução da diversidade lingüística em relevos correlacionados / Eduardo José de Souza Silva. -- 2008.
52 f. : il.

Orientadora : Viviane Moraes de Oliveira
Dissertação (Mestrado em Biometria) -- Universidade Federal Rural de Pernambuco. Departamento de Estatística e Informática.
Inclui bibliografia.

CDD 574.018 2

1. Biometria
2. Diversidade lingüística
3. Línguas
4. Competição
- I. Oliveira, Viviane Moraes de
- II. Título

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**MODELO ESTOCÁSTICO PARA EVOLUÇÃO DA DIVERSIDADE
LINGÜÍSTICA EM RELEVOS CORRELACIONADOS**

EDUARDO JOSÉ DE SOUZA SILVA

Dissertação julgada adequada para
obtenção do título de mestre em
Biometria e Estatística Aplicada,
defendida e aprovada por
unanimidade em 22/02/2008 pela
Comissão Examinadora.

Orientador:



Prof(a). Dr(a). Viviane Moraes de Oliveira
Universidade Federal Rural de Pernambuco

Banca Examinadora:



Prof(a). Dr(a). Tatijana Stosic
Universidade Federal Rural de Pernambuco



Prof(a). Dr(a). Eufrazio de Souza Santos
Universidade Federal Rural de Pernambuco



Prof(a). Dr(a). Emande Barbosa da Costa
Universidade Federal Rural de Pernambuco

DEDICATÓRIA

Esta dissertação é toda dedicada em memória de Cleide Maria de Souza Silva. Eternas saudades do esposo, filhas, neto, irmãos, sobrinhos, parentes, amigos e do seu filho Eduardo José de Souza Silva.

"Porque a inclinação da carne é a morte, mas a inclinação do espírito é vida e paz."(ROM 8,6)

"... O mundo não é só de lutas, matérias e bombas atômicas, é também de sentimentos, música, amor, poesia e revelação."Cleide Maria de Souza Silva

AGRADECIMENTOS

O amor de Deus é como o oceano, você pode ver o começo, mas não o fim.

Aos padres capuchinhos da Basílica da Penha frei João e Jânio e a todos os padres da Basílica pelas palavras de amor, apoio e conforto.

Meu querido pai José Dionísio da Silva e minha querida mãe(em memória) Cleide Maria de Souza Silva pela educação, por não ter deixado eu desistir nos momentos mais difíceis da dissertação e pelo incentivo.

Minhas irmãs Simone Maria de Souza Silva e Karina Maria de Sousa Silva pelo apoio e incentivo. Meus tios Clodoaldo, Cláudio, Geraldo, Clóvis e minhas tias Célia, Maria Tereza e Albinha. Meus primos André Geraldo, Silvia Analha e Karol Souza, Clodoaldo Junior pelos incentivos e apoios durante minha vida. A minha namorada Juliana Lins pelo apoio, incentivo e principalmente pela paciência.

As professoras Doutoras do departamento de Estatística da Universidade Federal de Pernambuco Maria Cristina, Audrey Cysneiros e Jacira Guiro pelas cartas de indicações.

Ao coordenador do curso de pós-graduação professor Dr. Eufrázio pela confiança.

A minha orientadora Dra. Viviane Moraes de Oliveira e ao meu co-orientador Dr. Paulo Roberto de Araújo Campos pelo incentivo, orientação e paciência.

Aos professores doutores do departamneto de informática e estatística Gauss Cordeiro, Bork Stosic, Tatijana Stosic e Laélia Campos, CNPQ e FACEPE.

Aos professores mestres Edilson Gonçalves do departamento de Física da Universidade Federal Rural de Pernambuco, Yara Leal do departamento de Educação da Universidade Católica de Pernambuco.

Ao professor Dr. Pedro Hugo do departamento de Física da Universidade Rural de Pernambuco.

Aos amigos Christian Seraffin, Marcelo Silva, Josete Cavalcanti, Ana clara, Suly, Reynaldo, Manuelle.

Aos amigos de laboratório Pedro Neto, Carlos Augusto , Osmar Francisco, Raphael Fonseca e Yuri Romma.

Aos amigos de turma Angela, Rosângela, Esdras, Adriano, Iran, Frank, Luiz José, Luiz Henrique, Janilson, Moacy, Domingos e Lucas.

Aos funcionários do DEINFO em especial Dona Zuleide França.

Resumo

As origens das línguas têm sido tema de investigação e grande interesse desde a Antigüidade, e os avanços recentes na arqueologia, genética e lingüística têm sido importantes para a obtenção de uma melhor compreensão da diversificação lingüística. Entretanto, não há um consenso universal com respeito à evolução desta diversidade. Algumas similaridades entre grupos distintos de línguas sugerem que elas devam ter um ancestral comum. Neste trabalho, estudamos a evolução da diversidade lingüística por meio de uma simulação computacional simples que considera os mecanismos de difusão fundamentais que geram e mantêm esta diversidade. O nosso modelo é usado para descrever a ocupação de uma dada área por populações que falam diversas línguas. No processo de colonização das regiões podem ocorrer mutações e dessa forma a diversidade lingüística é aumentada. No modelo, regiões da rede são caracterizadas pela quantidade de recursos disponíveis às populações que vão colonizar a região. A cada língua é atribuído um fitness f que é definido como a soma dos recursos dos sítios contendo populações que falam aquela língua específica. Estudamos os casos em que os valores dos fitness são ilimitados e limitados por um valor máximo (saturação). Nosso modelo dá origem a leis de escala em boa concordância com os observados na distribuição de línguas da Terra. Uma mudança abrupta no cenário qualitativo apenas é observada no limite de correlação muito alta, onde o crescimento da diversidade lingüística com a área é mais rápido do que para relevos com baixa correlação. No caso em que os fitness das línguas são limitados encontramos que para relevos com baixa correlação a dependência da diversidade com a área apresenta um regime de escala que está em boa concordância com a distribuição de línguas na Terra.

Abstract

The origins of the languages have been an issue of investigation and broad interest since ancient times, and recent advances in archeology, genetics and linguistics have been important to a better comprehension of the linguistic diversification. However, there is not a universal consensus concerning the evolution of this diversity. Some similarities among distinct groups of languages suggest that they must have a common ancestor. Here we study the evolution of the linguistic diversity by means of a spatial computer simulation which considers the underlying diffusion mechanisms that generate and sustain this diversity. The model is used to describe the occupation of a given area by populations speaking various languages. In the process of colonization of regions, language mutation or differentiation can take place, and so increase the linguistic diversity. In the model, regions of the lattice are characterized by the amount of resources available to populations which are going to colonize the region. To each language, we assign a fitness value f which is defined as the sum of the resources of the sites containing populations which speak that specific language. We study both the cases where the values of fitness are unbounded and bounded by a maximum value (saturation). We introduce 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. Our model gives origin to scaling laws in close resemblance with the ones observed to the actual distribution of languages on Earth. An abrupt change of the qualitative scenario is only observed in the limit of very high correlation, where the increasing of the linguistic diversity on area is faster than for low correlated landscapes. In the case where the fitness of the languages are bounded we found that for low correlated landscapes, the dependence of diversity on area shows a scaling regime which is in good agreement with the actual distribution of languages on Earth.

Lista de Figuras

1	Diversidade média de línguas D como função da área A (Gomes et al,1999). . .	p. 10
2	Frequência de línguas como função do tamanho população(Sutherland,2003). . .	p. 11
3	Em uma rede composta por $A = 5^2$ sítios apresentamos: (a) O primeiro sítio a ser colonizado pela língua ancestral, L_1 e seus quatro vizinhos mais próximos;(b) A escolha de um dos quatro vizinhos mais próximos que será colonizado pela língua L_1 ou uma língua mutante;(c) A ocorrência de uma mutação dando origem à língua L_2 ;(d) Aglomerado cujos sítios foram colonizados por populações falando as línguas L_1 e L_2 ;(e) A escolha de um sítio na fronteira de colonização que poderá ser colonizado pelas línguas L_1, L_2 ou uma mutante destas.	p. 20
4	Diversidade D como função da área A para $\alpha = 0.15$ (a), $\alpha = 0.3$ (b), $\alpha = 0.7$ (c), $\alpha = 0.9$ (d) e $\lambda = 0(\bigcirc)$, $\lambda = 0.5(\triangle)$, $\lambda = 0.93(\square)$ e $\lambda = 0.99(\diamond)$	p. 22
5	Expoente z como função de λ para (a) $\alpha = 0,15$ e (b) $\alpha = 0.7$	p. 23
6	Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.3$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$	p. 24
7	Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.7$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$	p. 25
8	Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.9$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$	p. 26
9	Número de línguas com população maior do que N como função de N para $\alpha = 0.5$, $\lambda = 0$ (a), $\lambda = 0.5$ (b) e $\lambda = 0.99$ (c)	p. 27
10	Número de línguas com população maior do que N como função de N para $\alpha = 0.9$, $\lambda = 0$ (a), $\lambda = 0.5$ (b) e $\lambda = 0.99$ (c)	p. 28
11	Evolução da $\bar{A}(t)$ para $\alpha = 0.3$, $\lambda = 0$ (a) e $\lambda = 0.99$ (b).	p. 34
12	Evolução da $\bar{A}(t)$ para $\alpha = 0.9$, $\lambda = 0$ (a) e $\lambda = 0.99$ (b).	p. 34
13	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.3$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 35

14	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.7$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 35
15	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.9$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 35
16	Diversidade D como função da área A para $\alpha = 0.7$ e $\lambda = 0(\square)$, $\lambda = 0.5(\circ)$, $\lambda = 0.93(\triangle)$ e $\lambda = 0.99(\diamond)$	p. 37
17	Expoente z como função de λ para $\alpha = 0.7$	p. 38
18	Número de línguas com população maior do que N , $n(> N)$, como função de N para $\alpha = 0.7$ e $\lambda = 0$ com $\tau=0,52\pm0,01$	p. 39
19	Número de línguas n como função do tamanho da população N para $\alpha = 0.3$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 40
20	Número de línguas n como função do tamanho da população N para $\alpha = 0.7$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 41
21	Número de línguas n como função do tamanho da população N para $\alpha = 0.9$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 42
22	Evolução de $A(t)$ para $\alpha = 0.7$ e $\lambda = 0$	p. 43
23	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.3$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 43
24	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.7$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 44
25	Diversidade $D(t)(\square)$ e área média $A(t)(\bigcirc)$ como função do tempo t para $\alpha = 0.9$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$	p. 44

Sumário

DEDICATÓRIA	p. i
AGRADECIMENTOS	p. ii
1 Introdução	p. 8
2 Modelos	p. 13
2.1 Equações Diferenciais	p. 13
2.2 Modelos baseados em indivíduos	p. 16
3 Evolução da diversidade lingüística em relevos correlacionados	p. 18
3.1 Modelo	p. 18
3.2 O processo de difusão	p. 30
3.2.1 Passeio aleatório unidimensional	p. 30
4 Análise do caso com saturação do fitness	p. 36
4.1 Introdução	p. 36
4.2 Resultados	p. 36
5 Conclusões	p. 45
Referências	p. 46

1 Introdução

As origens das línguas têm sido assunto de investigação e grande interesse desde os tempos remotos. Recentes avanços na arqueologia, genética e lingüística têm sido importantes para uma melhor compreensão dessa diversidade. Contudo, não existe um consenso a respeito da evolução da diversidade de línguas(Renfrew,1994).

Os arqueólogos têm identificado vários marcos no comportamento humano, mas ainda não há um consenso sobre quais deles implicam em capacidade de linguagem(Holden,2004). O aparecimento da primeira ferramenta de pedra, datada de 2,4 milhões de anos, para alguns pesquisadores pode indicar facilidade de comunicação. Outro possível ponto inicial se deu há 2 milhões de anos quando o cérebro dos hominídeos começou a desenvolver-se, inclusive em áreas primárias associadas ao processamento da linguagem. Estudos com esqueletos de hominídeos revelaram que há cerca de 300.000 anos nossos antepassados tornaram-se mais ou menos modernos anatomicamente, e que possuíam uma laringe localizada na parte superior da traquéia(Holden, 2004). Essa característica aumentou o intervalo de sons que os humanos podiam produzir.

Outros marcos possíveis vieram dos estudos em genética. Como um exemplo, pesquisadores do Instituto Max Planck de Antropologia Evolucionária relataram que o 'gene da língua' FOXP2 pode ter sofrido sua última mutação há cerca de 100.000 mil anos(Enard et al,2002), o que fixa as bases para uma nova fluência lingüística.

Os primeiros hominídeos surgiram na África, há cerca de 4 ou 5 milhões de anos. O Homo Erectus se desenvolveu neste continente e se dispersou pela Ásia e Europa. De acordo com a *Out of Africa Hypothesis*, a passagem do Homo Erectus para o Homo Sapiens teria sido restrita à África, e nossos ancestrais saíram da África nos últimos 200.000 anos substituindo todos os povos nativos encontrados(Renfrew,1994). Uma teoria multiregionalista alternativa diz que o processo de transição do Homo Erectus para o Homo Sapiens não ficou restrito à África, incluindo também a Ásia e a Europa. As evidências genéticas favorecem a primeira hipótese(Nie,1995). O Homo Sapiens alcançou nossa forma atual, o Homo Sapiens Sapiens, há cerca de 100.000 anos. Este se dispersou pela Terra falando uma ou várias línguas. Há 40.000 anos atrás nossa espécie colonizou o Levante, a Europa, a Ásia, Nova Guiné e Austrália. Entre 37.000 e 16.000 anos atrás os asiáticos cruzaram o Estreito de Bering dando início à colonização das Américas.

Existem quatro processos pelos quais uma língua passa a ser falada em uma determinada região: (a) colonização de uma região desocupada; (b) a convergência, que ocorre quando línguas contemporâneas influenciam umas às outras através do empréstimo de palavras, frases e formas gramaticais; (c) divergência, que ocorre quando novas línguas surgem a partir de uma; (d) substituição lingüística, onde uma língua é trocada por outra.

Há mais de duzentos anos, os lingüístas concordam que algumas línguas possuem semelhanças em vocabulário, gramática, formação de palavras e no uso de sons e que, por esta razão, devem ter um ancestral comum(Renfrew,1994). Esses grupos com características em comum foram denominados de famílias lingüísticas. A classificação mais famosa deste tipo foi empreendida em 1786 por Sir Willian Jones, juiz britânico da Suprema Corte de Calcutá, que observou relações entre Gótico, Grego, Latin, Persa e Sânscrito, do ponto de vista de palavras comuns e de características gramaticais. Hoje essa família lingüística é conhecida como Indo-Européia.

Em 1963, o lingüista americano Joshep H. Greenberg da Universidade de Stanford deu um passo significativo oferecendo uma visão unificada das línguas e classificando os idiomas da África em quatro macrofamílias dominantes: a Afro-Asiática, a Khoisana, a Negro-Kordofaniana e a Nilo-Sahareana. Ele propôs esta classificação utilizando o método de análise multilateral que consiste em examinar várias palavras simultaneamente em muitos idiomas(Refrew,1994). Ele também aplicou este procedimento para as línguas das Américas e identificou três famílias: a Esquimó-Aleuta, Na-Dene e Ameríndia.

Pelo intenso intercâmbio lingüístico entre diversos povos e comunidades, as línguas se modificam dando origem a novos dialetos e até outras formas de escrita. Os idiomas em todo o planeta estão sendo submetidos a uma rápida mudança por causa da demografia mundial, das novas tecnologias e da internacionalização da comunicação, podendo muitas delas desaparecerem do mapa, caso essas tendências atuais se matenham. A demografia mundial é uma das causas da crise das línguas. A população do planeta cresceu rapidamente durante o século vinte e populações das nações subdesenvolvidas migraram para os países desenvolvidos gerando grandes centros urbanos. Estes grandes centros urbanos são regiões onde idiomas misturam-se e mudam rapidamente transformando dialetos nativos antigos em novos dialetos(Graddol,2004).

O futuro das línguas é um tema de interesse e preocupação(Diamond,1997). Estima-se que ao menos 50% das línguas existentes podem estar extintas no próximo século(Sampat, 2001; Graddol, 2004). Enquanto uma centena de línguas são faladas por cerca de 90% da população mundial, a maioria das línguas está presente em pequenas regiões. A perda de diversidade lingüística é motivo de preocupação não apenas dos lingüístas, porque línguas fornecem uma maneira importante de entender melhor o passado de nossa espécie. Como algumas línguas possuem um vocabulário muito elaborado para descrever o mundo, sua perda também implica em perda de conhecimento ecológico.

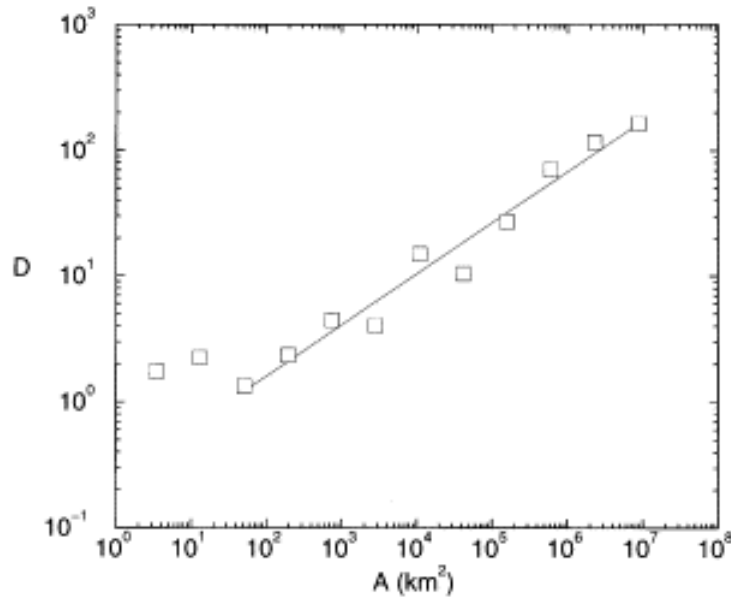


Figura 1.1: Diversidade média de línguas D como função da área A (Gomes et al,1999).

Nos últimos anos tem havido um crescente interesse da comunidade de sistemas complexos pelo estudo de alguns aspectos interdisciplinares da dinâmica de línguas, em particular em temas como competição entre línguas, risco de extinção e morte de línguas (Abrams e Strogatz, 2003; Patriarca e Leppänen, 2004; Schulze e Stauffer, 2004; Mira e Paredes, 2005; Schwämmle, 2005; Kosmidis et al, 2005).

Recentemente Gomes e colaboradores (Gomes et al, 1999) investigaram a distribuição das aproximadamente 6700 línguas vivas na Terra. Eles observaram a existência de algumas relações de escala entre a diversidade de línguas e a área dos países, assim como com suas populações. Dividindo os 228 países em 12 grupos de acordo com suas áreas e calculando a diversidade média D de línguas em cada grupo, eles obtiveram que a diversidade lingüística aumenta com a área A de acordo com uma lei de potência, $D \sim A^z$, onde, $z = 0.41 \pm 0.03$ (Figura 1.1). Eles também observaram uma dependência em lei de potência entre a diversidade e a população média N de cada grupo. Neste caso, $D \sim N^v$, com $v = 0.50 \pm 0.04$. Ainda neste estudo eles observaram que o número de línguas com população maior que N , $n(> N)$, como função de N apresenta dois regimes em lei de potência: $n(> N) \sim N^{-\tau}$ com $\tau = 0.5$ para $5 \times 10^4 < N < 6 \times 10^6$ e $\tau = 1$ para $2 \times 10^7 < N < 1 \times 10^9$, sendo cada uma destas leis de potência válidas para duas décadas. Eles propõem que o fato de o expoente τ diferir para línguas intermediárias e grandes pode ser visto como uma evidência de que essas duas classes de línguas possuem dinâmicas de crescimento distintas.

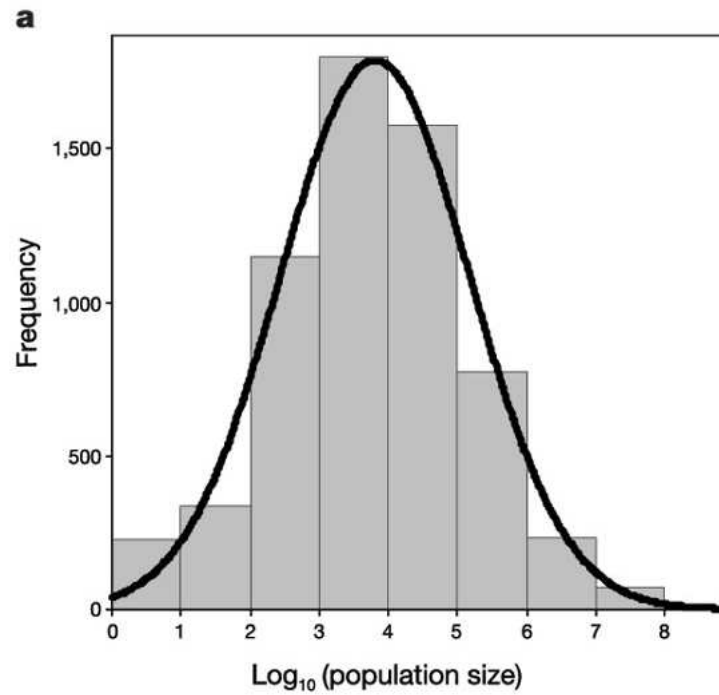


Figura 1.2: Frequência de línguas como função do tamanho população(Sutherland,2003).

Em outra investigação sobre a distribuição de línguas na Terra, Sutherland(Sutherland,2003) observou que a frequência de línguas com um dado tamanho de população segue aproximadamente uma distribuição lognormal(Figura 1.2). Embora, o teste de normalidade de Kolmogorov-Smirnov rejeite a hipótese de normalidade mostrando que os dados não seguem uma distribuição normal.

No capítulo 2 descrevemos alguns modelos utilizados no estudo da evolução da diversidade lingüística. Alguns desses modelos são baseados em equações diferenciais e outros baseados em indivíduos. Dentre esses, citamos o modelo introduzido por de Oliveira e colaboradores (de Oliveira et al, 2006), que consiste em um modelo espacial simples que inclui processos de colonização geográfica seletiva, difusão e mutação.

Nos capítulos 3 e 4 extendemos o modelo de Oliveira e colaboradores, introduzindo uma correlação espacial no relevo. Temos como objetivo estudar os efeitos da variação da correlação no relevo sobre a evolução da diversidade lingüística. Comparamos nossos resultados com aqueles obtidos por Gomes (Gomes et al, 1999) e Sutherland (Sutherland,2003) para a distribuição atual de línguas no planeta Terra.

No capítulo 5 apresentamos nossas conclusões e perspectivas.

2 Modelos

Nos últimos anos, um grande número de pesquisadores tem investigado as origens da linguagem e a evolução da diversidade de línguas. A maior parte dos modelos desenvolvidos têm como foco o surgimento, a competição e os riscos de extinção. Neste capítulo descreveremos brevemente alguns modelos utilizados no estudo da evolução da diversidade lingüística. Apresentaremos, inicialmente, os modelos baseados em equações diferenciais, onde assume-se uma população constante e observa-se a evolução das concentrações de indivíduos que falam uma determinada língua. Em seguida, apresentaremos os modelos baseados em indivíduos (agent-based models) (ABM).

2.1 Equações Diferenciais

A simulação computacional mais simples para a evolução de línguas foi desenvolvida por Abrams e Strogatz (Abrams, Strogatz, 2003). Eles consideraram um sistema de duas línguas, X e Y , que competem em uma população altamente conectada, sem estrutura espacial, onde cada indivíduo fala apenas uma língua. A atratividade de um língua cresce com o seu número de falantes e com seu status s , um parâmetro que reflete as oportunidades sociais ou econômicas proporcionadas a seus falantes.

O modelo para a mudança de língua é

$$\frac{dx}{dt} = yP_{YX}(x, s) - xP_{XY}(x, s), \quad (2.1)$$

onde x e $y = 1 - x$ são as frações de indivíduos que falam as línguas X e Y , respectivamente. $P_{yx}(x, s)$ é a probabilidade, por unidade de tempo, que um indivíduo que fala a língua Y passe a falar a língua X . Eles assumiram funções da forma $P_{YX}(x, s) = cx^\alpha s$ e $P_{XY}(x, s) = c(1 - x)^\alpha(1 - s)$.

Eles observaram que a Equação (2.1) possui três pontos fixos, onde apenas as soluções $x = 0$ e $x = 1$ são estáveis. Portanto, as duas línguas não podem coexistir de maneira estável. O modelo descreve bem o declínio de línguas como o, Galês, o Gaélico Escocês, e o Quechua.

O modelo de Abrams e Strogatz não contempla sociedades bilíngües, como é o caso da Galícia

no noroeste da Espanha, onde a maioria da população fala a língua galega e a castelhana. Mira e Paredes(Mira e Paredes,2005) generalizaram o modelo de Abrams e Strogatz considerando que uma fração da população é bilíngüe.

Neste modelo, X , Y e W são subconjuntos da população que falam apenas a língua X , apenas a língua Y e ambas as línguas, respectivamente. x , y e w são as frações da população que pertencem a esses grupos, de forma que $x + y + w = 1$.

A variação da fração de indivíduos que falam a língua X no tempo é dada por

$$\frac{dx}{dt} = yP_{yx} + bP_{bx} - x(P_{xy} + P_{xb}) \quad (2.2)$$

onde $P_{\alpha\beta}$ é a fração do grupo α que é transferida para grupo β por unidade de tempo. As frações do grupo X que são transferidas para os grupos Y e W , por unidade de tempo, são dadas por

$$P_{xb} = cks_y(1 - x)^\alpha \quad (2.3)$$

e

$$P_{xy} = c(1 - k)s_y(1 - x)^\alpha, \quad (2.4)$$

onde s_Y é o status de Y ($0 \leq s_Y \leq 1$; $s_X = 1 - s_Y$), c e α são constantes e o parâmetro k ($0 \leq k \leq 1$) reflete a facilidade de se tornar bilíngüe, ou seja, a similaridade entre as duas línguas. $k = 0$ representa a situação onde a conversação entre duas pessoas que falam línguas diferentes é impossível e reduz o modelo àquele de Abrams e Strogatz. Quando $k = 1$ temos $X = Y$. As equações para P_{yw} e P_{yx} são similares às Equações (2.3) e (2.4). Eles também assumiram que $P_{wx} = P_{yx}$ e $P_{wy} = P_{xy}$.

Eles observaram que para cada valor de s_x existe um valor de k , k_{min} , tal que a língua com menor status é extinta para todo $k < k_{min}$. Em outras palavras, o bilingüismo é possível quando há uma certa similaridade entre as línguas.

Pinasco e Romanelli(Pinasco,Romanelli,2006) propuseram um modelo pertencente à família de modelos Lotka-Volterra para explicar a coexistência de duas línguas. O modelo mais simples dessa família de modelos consiste em um par de equações diferenciais, não lineares e de primeira ordem, e que são freqüentemente utilizadas para descrever dinâmicas de interação entre duas espécies, onde uma delas atua como presa e a outra como predadora. Este conjunto de equações foi proposto independentemente por Alfred J. Lotka em 1925 e Vito Volterra em 1926. A forma usual dessas equações é a seguinte:

$$\frac{dy}{dt} = r_1y - \alpha_1xy \quad (2.5)$$

$$\frac{dx}{dt} = -r_2x + \alpha_2xy \quad (2.6)$$

onde x é o número de indivíduos do predador e y é o número de indivíduos da presa. A população da presa cresce exponencialmente com uma taxa r_1 na ausência do predador enquanto que a população do predador decresce exponencialmente com uma taxa r_2 na ausência da presa. α_1 e α_2 são as taxas de decrescimento da população de presas e crescimento da população de predadores, respectivamente, devido ao encontro entre as duas espécies. No modelo de Pinasco e Romanelli, duas línguas, x e y , competem e possuem populações $X(t)$ e $Y(t)$ no tempo t . Por simplicidade, eles assumiram que x é a única língua atrativa, e definiram um coeficiente c que é a taxa de conversão da língua y para a língua x . Este coeficiente representa o status de x , isto é, retrata as oportunidades econômicas ou sociais permitidas aos que falam x .

Para modelar a competição entre as duas línguas eles utilizaram um sistema de equações diferenciais dado por

$$\frac{dx}{dt} = cxy + \alpha_x x \left(1 - \frac{x}{S_x}\right) \quad (2.7)$$

$$\frac{dy}{dt} = -cxy + \alpha_y y \left(1 - \frac{y}{S_y}\right). \quad (2.8)$$

A taxa de conversão de y em x é proporcional a xy . Os termos $\alpha_x x \left(1 - \frac{x}{S_x}\right)$ e $\alpha_y y \left(1 - \frac{y}{S_y}\right)$ são as taxas de crescimento, α_x e α_y são parâmetros positivos e incluem as taxas de natalidade e mortalidade, S_x e S_y são os termos de saturação das populações.

Os pontos fixos do sistema são

$$\frac{dx}{dt} = 0 \Leftrightarrow Y = \frac{\alpha_x X}{cS_x} - \frac{\alpha_X}{c} \quad \text{ou} \quad x = 0, \quad (2.9)$$

$$\frac{dY}{dt} = 0 \Leftrightarrow Y = -\frac{cS_y X}{\alpha_y} + S_y \quad \text{ou} \quad y = 0 \quad (2.10)$$

onde temos os pontos de equilíbrio $(0, S_y)$ e $(S_x, 0)$ correspondendo à morte de uma língua e $(0,0)$ correspondendo à extinção de ambas as populações. Existe outro ponto de equilíbrio onde, $(x_e, y_e) = \left(\frac{\alpha_y S_x (cS_y + \alpha_x)}{c^2 S_x S_y + \alpha_x \alpha_y}, \frac{\alpha_x S_y (\alpha_y - cS_x)}{c^2 S_x S_y + \alpha_x \alpha_y}\right)$, se a condição $S_x < \frac{\alpha_y}{c}$ for satisfeita. Portanto, a coexistência é possível.

Nowak e colaboradores (Nowak, 2001; Nowak, 2002) apresentaram uma dinâmica evolucionária para o aprendizado de línguas ou gramáticas por crianças. Eles utilizaram um conjunto de equações

$$\frac{dx_i}{dt} = \sum_{j=1}^n f_j Q_{ji} x_j - \phi x_i \quad i = 1, \dots, n \quad (2.11)$$

onde x_i é a fração de indivíduos que falam a língua i . $f_j = \sum_{i=1}^n x_i F_{ji}$ é o fitness da língua j , que é determinado pela probabilidade F_{ji} de que um indivíduo que fala a língua j seja entendido por pessoas que falam a língua i . $\phi = \sum_j f_j x_j$ é o fitness médio da população. Q_{ji} é a probabilidade de que crianças que possuem pais que falam a língua j passem a falar a língua i . O tamanho da população é mantido constante $\sum_i x_i = 1$.

Eles observaram que para taxa de mutação igual a zero, existem n soluções de equilíbrio estável da forma $x_i = 1$ e $x_j = 0$ (para todo $i \neq j$). Essas soluções correspondem a situações onde todos os indivíduos de uma população adotaram a mesma língua. Para taxas de mutação elevadas, a única solução estável é aquela em que todas as línguas ocorrem com frequências aproximadamente iguais.

2.2 Modelos baseados em indivíduos

Stauffer e colaboradores (Stauffer et al, 2004) introduziram um modelo onde as línguas são descritas por seqüências de l bits, onde cada bit pode receber o valor 0 ou 1. O tamanho da seqüência é fixado para cada simulação.

Eles iniciam com um indivíduo que fala a língua zero (todos os bits são iguais a zero). A cada iteração $t = 1, 2, \dots$, cada indivíduo vivo se reproduz dando origem a um filho. O filho fala a mesma língua do pai ou, com probabilidade p um de seus l bits muda de valor. Cada indivíduo pode morrer com probabilidade $\frac{N(t)}{K}$, onde $N(t)$ é a população total no início do passo de tempo t . K é o termo de saturação. Dessa forma a população é duplicada a princípio em cada iteração até ela alcançar um plateau em $\frac{K}{2}$.

Eles também assumem que um indivíduo que fala a língua i , de tamanho $S_i = x_i N(t)$ troca sua língua pela de uma pessoa selecionada aleatoriamente na população com probabilidade $(1 - x_i)^2$. Ou seja, quanto menor for o número de pessoas que falam uma língua, maior a probabilidade de troca por outra.

Variando a taxa de mutação p eles observaram dois regimes. Para p pequeno, uma língua domina e é falada por quase todas as pessoas na população. Quando p é grande a distribuição de tamanhos das línguas é aproximadamente lognormal.

O primeiro modelo a simular populações humanas se espalhando sobre um continente onde nenhum ser humano vivia antes foi o de de Oliveira e colaboradores (de Oliveira et al, 2006). A dinâmica proposta inclui processos de colonização geográfica seletiva, difusão, mutação e interação entre populações que ocupam regiões diferentes. O modelo apresentado por de Oliveira e colaboradores é definido sobre uma rede de duas dimensões, $A = L \times L$ sítios, sendo que cada sítio S_i representa uma região que pode ser ocupada por uma população falando uma língua. A cada sítio S_i é atribuída uma capacidade C_i que é estimada de uma distribuição uniforme de intervalo $[0, 1]$. A capacidade mede a soma de recursos disponíveis para a população colonizar a região.

O primeiro passo da dinâmica é escolher aleatoriamente um sítio da rede para ser colonizado por uma população falando a primeira língua. A cada língua é atribuído um fitness f como sendo soma das capacidades dos sítios contendo populações que falam uma língua específica. No passo

seguinte um dos quatros vizinhos mais próximos do sítio deverá ser escolhido com probabilidade proporcional as capacidades. O sítio escolhido é então ocupado por uma população falando a primeira língua ou uma mutação desta. A probabilidade de ocorrer uma mutação no processo de propagação é dado por $p = \frac{\alpha}{f}$. A etapa seguinte consiste em verificar quais sítios vazios estão na fronteira de colonização e eleger um destes sítios de acordo com as capacidades elegendo uma língua para ocupar esta região entre os vizinhos mais próximos. A simulação pára quando todos os sítios da rede tiverem sido colonizados.

Eles analisaram a diversidade D como função da área A (total de números de sítios na rede) para dois valores distintos de α obtendo duas escalas distintas, onde $D \sim A^z$. Quando $\alpha = 0.3$ estimou-se o expoente $z = 0,43 \pm 0,04$ para $4 < A < 1000$ e $z = 0,14 \pm 0,02$ para $1000 < A < 490000$. Para $\alpha = 0.73$ estimou-se $z = 0,88 \pm 0,01$ para $4 < A < 1000$ e $z = 0,35 \pm 0,03$ para $1000 < A < 490000$. Para ambos os valores de α os expoentes estão de acordo com a distribuição de línguas no planeta apresentado por Gomes e colaboradores (Gomes et al, 1999). Para áreas pequenas e intermediárias a diversidade de línguas cresce mais rapidamente, quando comparado com áreas grandes.

Eles também estudaram a distribuição de línguas, ou seja, a função de distribuição acumulada correspondendo ao número de línguas com população maior do que N , ($n > N$), como uma função de N . Para obter os gráficos de distribuição acumulada, de Oliveira et al assumiram que cada sítio da rede contribui com uma pessoa da população. Em analogia com a distribuição de línguas no planeta (Gomes et al, 1999), eles encontraram o regime de escala $n(> N) \sim N^{-\tau}$ onde $\tau = 0.36 \pm 0.01$, ao longo de quase três décadas em N .

3 Evolução da diversidade lingüística em relevos correlacionados

Como vimos anteriormente, a evolução da diversidade lingüística tem despertado grande interesse na comunidade científica e muitos modelos têm sido desenvolvidos com o intuito de entender melhor os mecanismos presentes nesta evolução. Neste capítulo, apresentamos uma extensão do modelo proposto por de Oliveira et al (de Oliveira et al, 2006), introduzindo um grau de correlação entre os recursos das regiões que podem ser ocupadas pelas populações, de tal forma que podemos variar o grau de heterogeneidade do relevo. Nosso objetivo é investigar o papel da correlação entre as regiões sobre a evolução da diversidade lingüística.

3.1 Modelo

O nosso modelo é definido sobre uma rede de duas dimensões, composta por $A = L \times L$ sítios, onde cada sítio S_i representa uma região que pode ser ocupada por uma população falando uma língua. Atribuímos a cada sítio S_i uma capacidade C_i , onde $C_i = e^{-x_i}$, que representa a quantidade de recursos disponíveis para a população que ocupará aquele sítio. O valor de x_i é estimado da seguinte forma: primeiro, escolhemos um sítio aleatoriamente e obtemos x_j de uma distribuição normal padrão, ou seja, $x_j \sim N(0, 1)$. Na etapa seguinte, obtemos os valores de x_i para os quatro vizinhos mais próximos desse sítio, a partir de uma distribuição Gaussiana de média $\mu = \lambda x_j$ e variância $\sigma^2 = 1 - \lambda^2$, ou seja,

$$P(x_i|x_j) = \frac{1}{\sqrt{2\pi(1-\lambda^2)}} \exp \left[-\frac{1}{2} \frac{(x_i - \lambda x_j)^2}{1 - \lambda^2} \right], \quad (3.1)$$

onde λ é o parâmetro de correlação entre as capacidades dos sítios e é definido no intervalo $[0, 1]$. O processo de obtenção dos x_i 's continua tomando como referência os valores obtidos para os vizinhos mais próximos. Após atribuir os valores de x_i a todos os sítios, tomamos $C_i = e^{-x_i}$ como a capacidade de cada sítio S_i . Quando o parâmetro de correlação λ é igual a zero, temos um relevo completamente descorrelacionado, enquanto que λ igual a um significa que todos os sítios possuem a mesma capacidade. Um grau intermediário de correlação é obtido quando λ varia entre zero e um ($0 < \lambda < 1$).

No início da dinâmica(Figura 3.1a) , escolhemos aleatoriamente um sítio da rede para ser colonizado por uma população que fala a primeira língua, a qual denominamos de língua L_1 . A cada língua atribuímos um fitness(adaptabilidade) f que é definido como a soma das capacidades dos sítios contendo populações que falam aquela língua específica. Desta maneira, o fitness inicial da primeira língua é a capacidade do sítio inicial. No segundo passo(Figura 3.1b), um dos quatros vizinhos mais próximos do sítio é escolhido com probabilidade proporcional à sua capacidade. Isto significa que as regiões que contêm uma grande quantidade de recursos possuem uma maior probabilidade de serem colonizadas primeiro. O sítio escolhido é então ocupado por uma população falando a primeira língua ou uma versão mutante desta. Se uma mutação ocorrer, a nova língua será denominada de língua L_2 (Figura 3.1c). A probabilidade de ocorrer uma mutação no processo de propagação da língua é $p = \frac{\alpha}{f}$, onde α é uma constante($0 < \alpha < 1$). Esta regra para probabilidade de mutação foi inspirada pela genética de populações. Observou-se que pequenas populações são mais vulneráveis à deriva genética, e que a taxa de deriva é inversamente proporcional ao tamanho da população. Deriva genética é um mecanismo de evolução que muda as características das espécies no tempo.

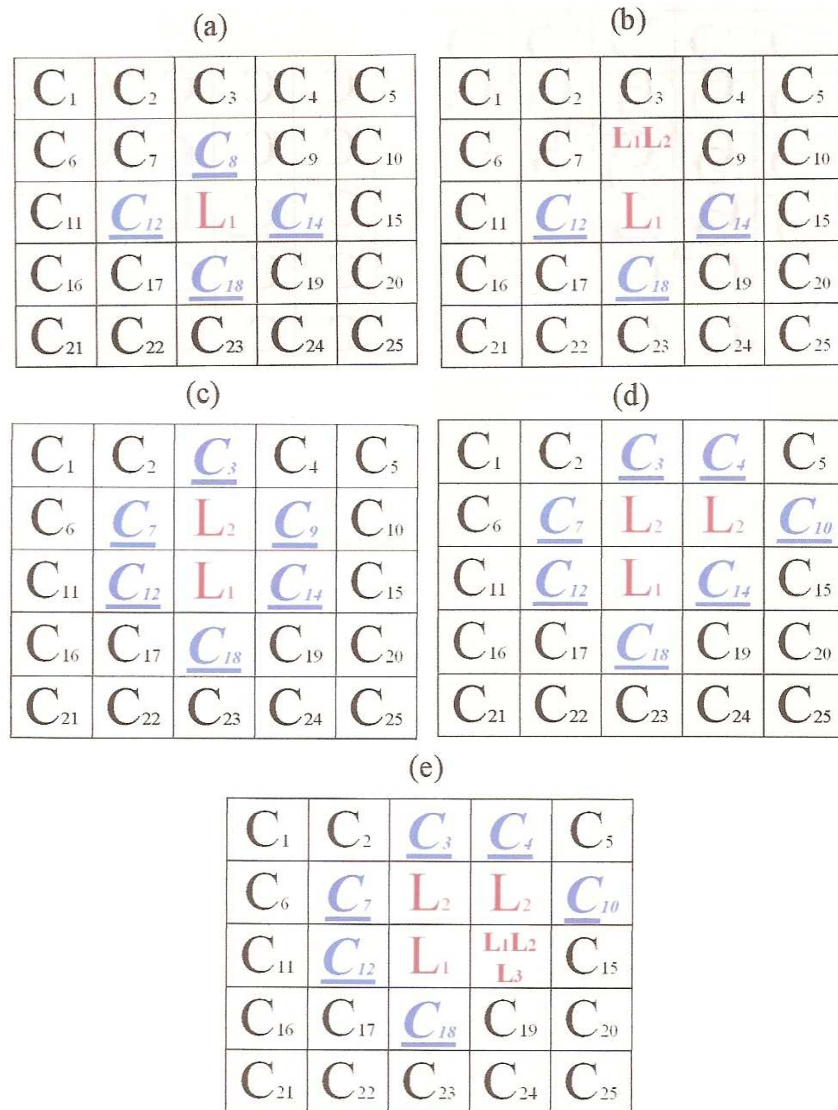


Figura 3.1: Em uma rede composta por $A = 5^2$ sítios apresentamos: (a) O primeiro sítio a ser colonizado pela língua ancestral, L_1 e seus quatro vizinhos mais próximos; (b) A escolha de um dos quatro vizinhos mais próximos que será colonizado pela língua L_1 ou uma língua mutante; (c) A ocorrência de uma mutação dando origem à língua L_2 ; (d) Aglomerado cujos sítios foram colonizados por populações falando as línguas L_1 e L_2 ; (e) A escolha de um sítio na fronteira de colonização que poderá ser colonizado pelas línguas L_1, L_2 ou uma mutante destas.

Nos passos seguintes verificamos quais sítios vazios estão na fronteira de colonização (Figura 3.1d). Escolhemos, então, um destes sítios de acordo com suas capacidades (Figura 3.1e). O sítio de maior capacidade tem maior probabilidade de ser colonizado primeiro. Sorteamos então uma língua para ocupar esse sítio. A língua com maior fitness em sua vizinhança tem maior chance de ocupar o sítio. O processo continua até que todos os sítios da rede sejam colonizados.

Na Figura 3.2 apresentamos a diversidade D como função da área A para $\alpha = 0.15$, $\alpha = 0.3$, $\alpha = 0.7$ e $\alpha = 0.9$ e quatro valores distintos de λ , a saber, $\lambda = 0$, $\lambda = 0.5$, $\lambda = 0.93$ e $\lambda = 0.99$. A diversidade é obtida contando-se o número de línguas na rede após todos os sítios terem sido colonizados. Obtemos cada ponto tomando médias sobre 1000 configurações para $L < 300$ e sobre 500 para $L = 300$ e 400.

Observamos que para um valor fixo de λ , a diversidade cresce mais rápido com a área quando o valor de α é aumentado. Isto ocorre porque a probabilidade de mutação é proporcional a α .

Quando fixamos o valor de α , notamos que a diversidade cresce mais rápido com a área à medida que o relevo torna-se correlacionado. A razão para a ocorrência deste comportamento é a seguinte: quando a correlação do relevo é baixa, os fitness iniciais das línguas que surgem são muito diferentes. Dessa forma, línguas que adquirem um fitness inicial alto crescerão rapidamente e, como a probabilidade de mutação é inversamente proporcional ao fitness, a chance de surgimento de novas línguas se torna muito pequena. Quando o relevo é altamente correlacionado, o fitness inicial das línguas que surgem é aproximadamente o mesmo. Assim, as línguas iniciais competem em igualdade de condições no início do processo de colonização e como seus fitness ainda não são altos, a diversidade é mais alta do que no caso em que o relevo é pouco correlacionado.

Com exceção do caso em que $\alpha = 0.15$, observamos duas regiões de escala em lei de potência quando o relevo é descorrelacionado ($\lambda = 0$) ou pouco correlacionado ($\lambda = 0.5$). Curiosamente, os expoentes obtidos em cada região são aproximadamente os mesmos para $\lambda = 0$ e $\lambda = 0.5$. Com o aumento da correlação do relevo, passamos a observar apenas uma região de escala para áreas pequenas e um decrescimento da diversidade com a área para áreas grandes. Como as capacidades dos sítios em um relevo altamente correlacionado possuem valores muito próximos, o fitness de uma língua é aproximadamente duplicado quando sua população coloniza dois sítios, é triplicado quando coloniza três sítios, e assim por diante. Conseqüentemente, para áreas grandes temos línguas com fitness muito grandes, resultando em uma taxa de mutação próxima de zero.

Com o objetivo de solucionar este problema, no próximo capítulo estudaremos esse modelo introduzindo um termo de saturação ao fitness, de forma a tornar o modelo mais realístico.

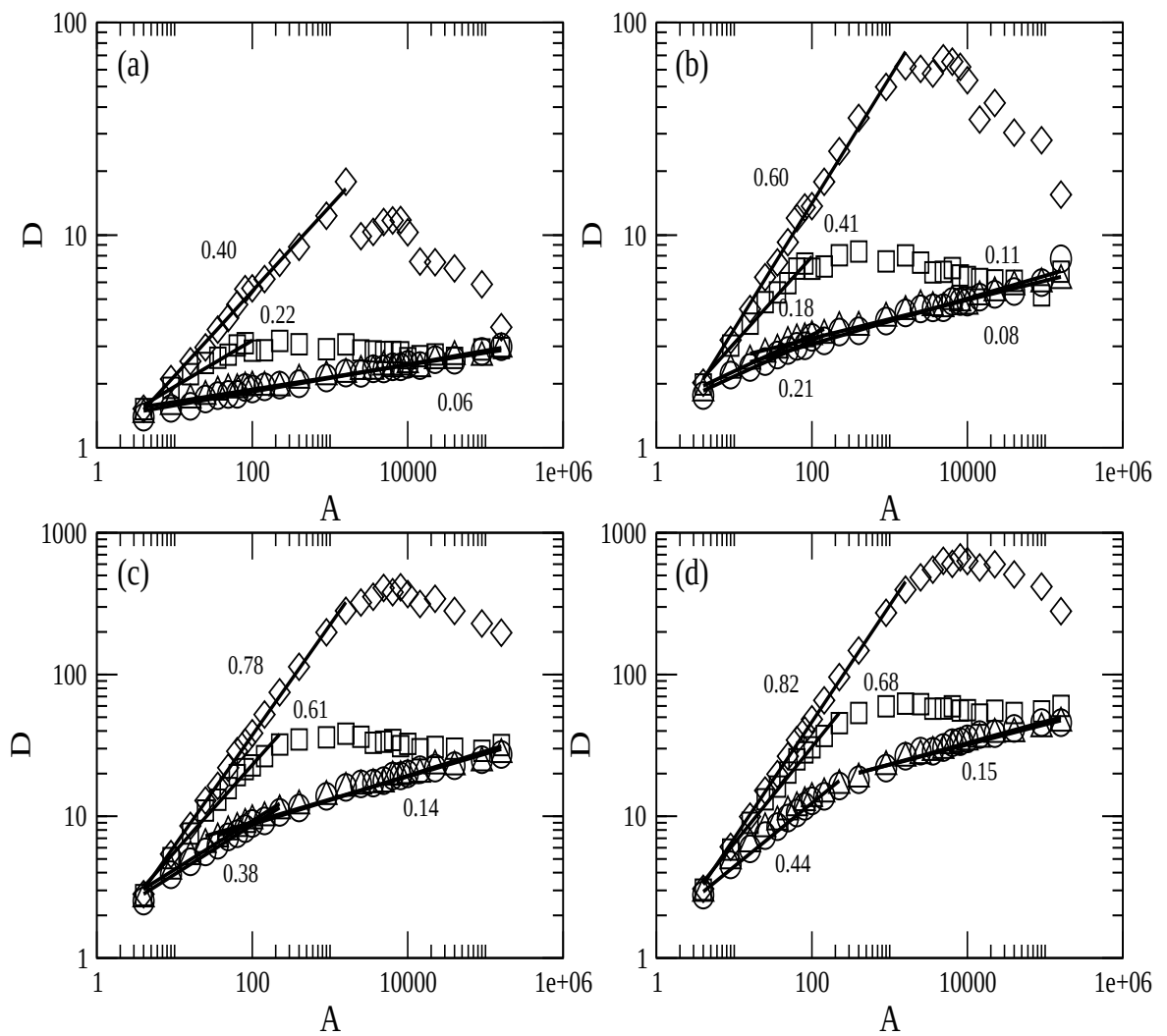


Figura 3.2: Diversidade D como função da área A para $\alpha = 0.15$ (a), $\alpha = 0.3$ (b), $\alpha = 0.7$ (c), $\alpha = 0.9$ (d) e $\lambda = 0(\bigcirc)$, $\lambda = 0.5(\triangle)$, $\lambda = 0.93(\square)$ e $\lambda = 0.99(\diamond)$

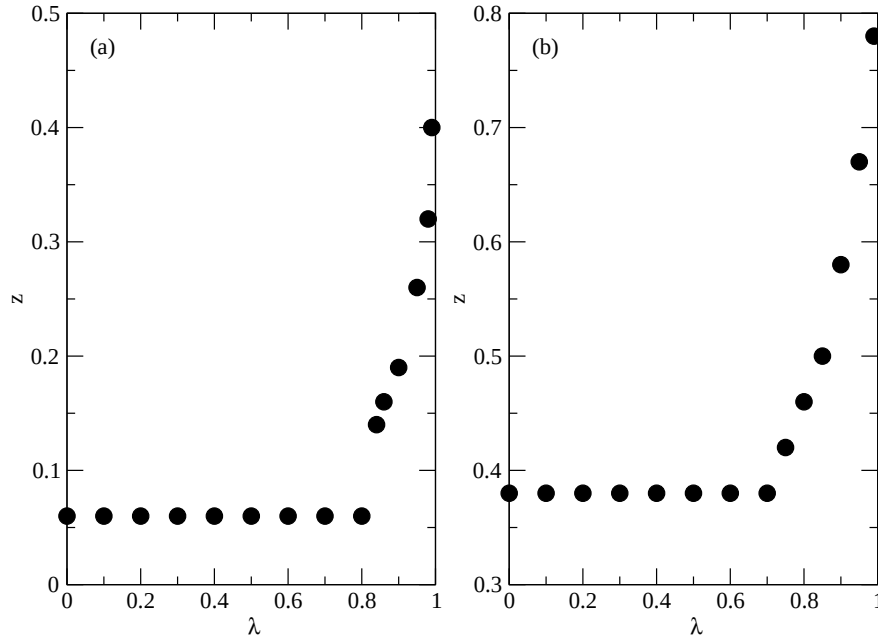


Figura 3.3: Expoente z como função de λ para (a) $\alpha = 0,15$ e (b) $\alpha = 0,7$.

Na Figura 3.3 mostramos a dependência do expoente z em relação ao parâmetro de correlação λ para $\alpha = 0.15$ e $\alpha = 0.7$. Observamos que z é constante para valores pequenos e intermediários de λ , mas cresce rapidamente quando $\lambda > 0.7$ e $\lambda > 0.8$, significando que a diversidade aumenta mais rapidamente para valores altos do parâmetro de correlação. Este mesmo comportamento qualitativo é observado para outros valores de α .

Na Figura 3.4 apresentamos o número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.3$ e três valores distintos de λ . Para calcular o tamanho da população assumimos que a população num dado sítio é igual à sua capacidade multiplicada por 100. Os gráficos são obtidos utilizando-se 400 configurações com $L = 400$. Observamos uma distribuição aproximadamente lognormal para populações pequenas e intermediárias. Também observamos um pico para grandes populações. Como o número de línguas com populações muito grandes na Terra é pequeno quando comparado ao número total de línguas (lembramos que em torno de 100 línguas são faladas por 90% da população da Terra), este resultado não apresenta boa concordância com o obtido empiricamente (Sutherland, 2003). Quando a correlação do relevo aumenta, este pico torna-se menos pronunciado. Mostramos nas Figuras 3.5 e 3.6 que ele tende a diminuir quando aumentamos a probabilidade de mutação. Como a população é constante, se o número de línguas aumenta, a consequência é um decréscimo no tamanho das populações que falam estas línguas.

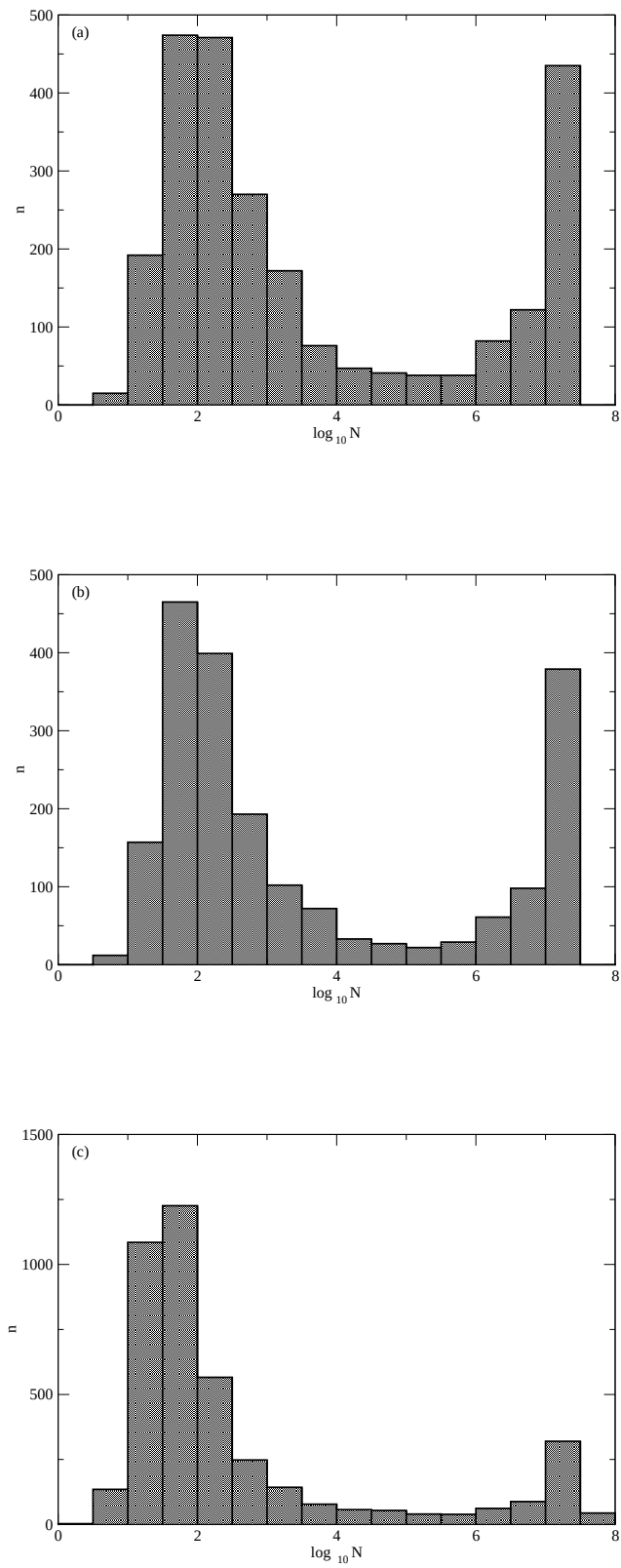


Figura 3.4: Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.3$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$.

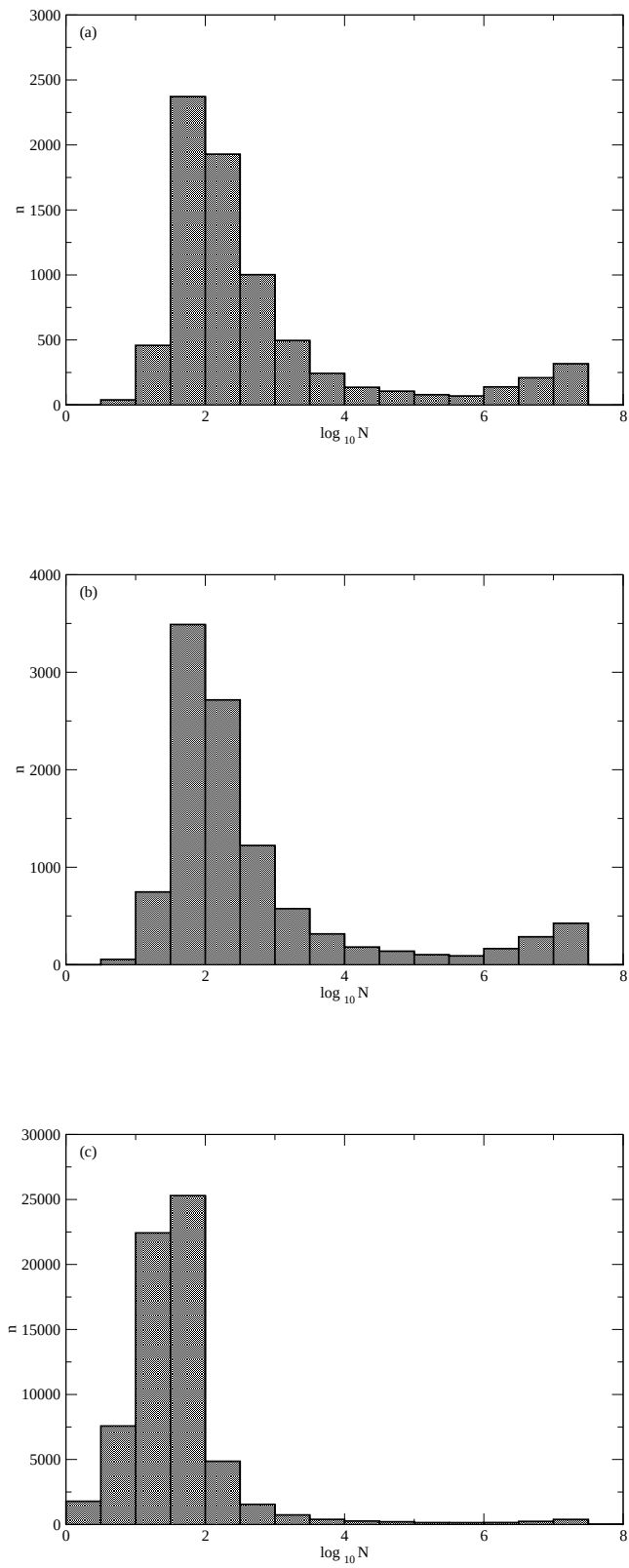


Figura 3.5: Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.7$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$.

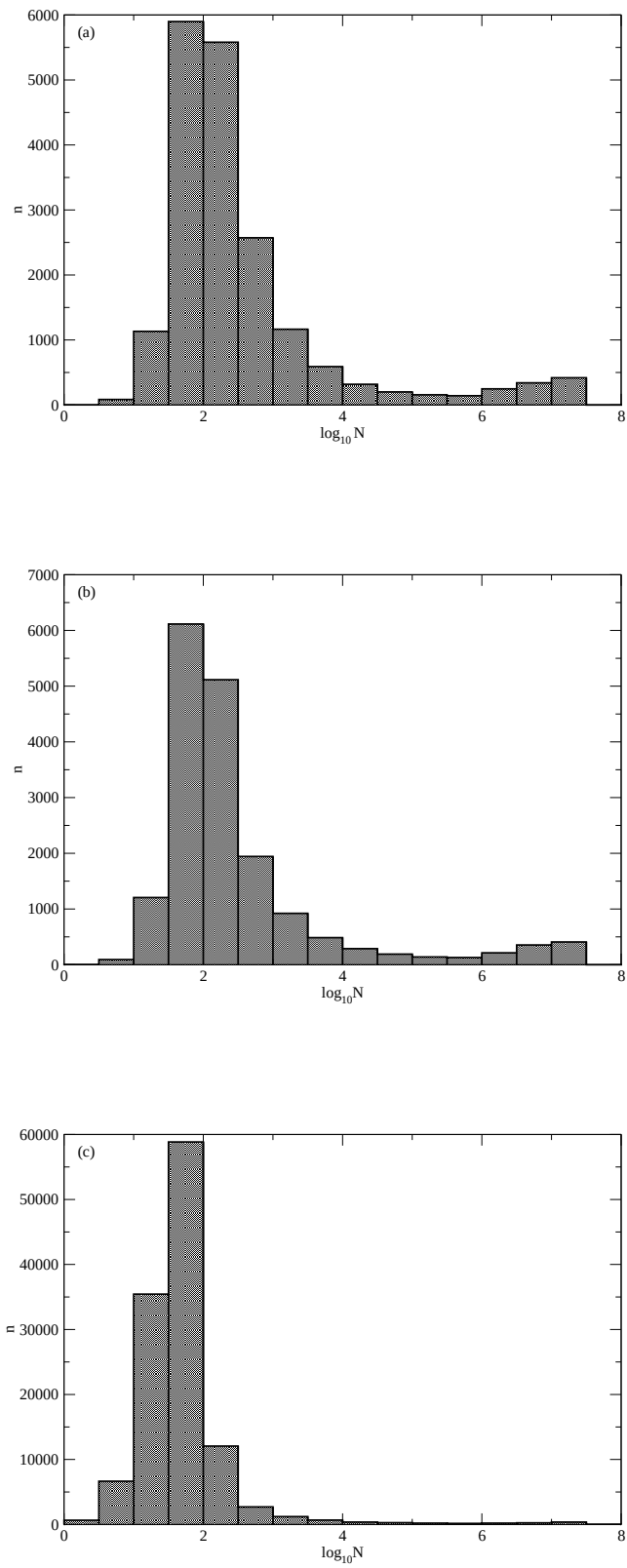


Figura 3.6: Número de línguas n como função do logaritmo do tamanho da população N para $\alpha = 0.9$ e (a) $\lambda = 0$, (b) $\lambda = 0.5$ e (c) $\lambda = 0.99$.

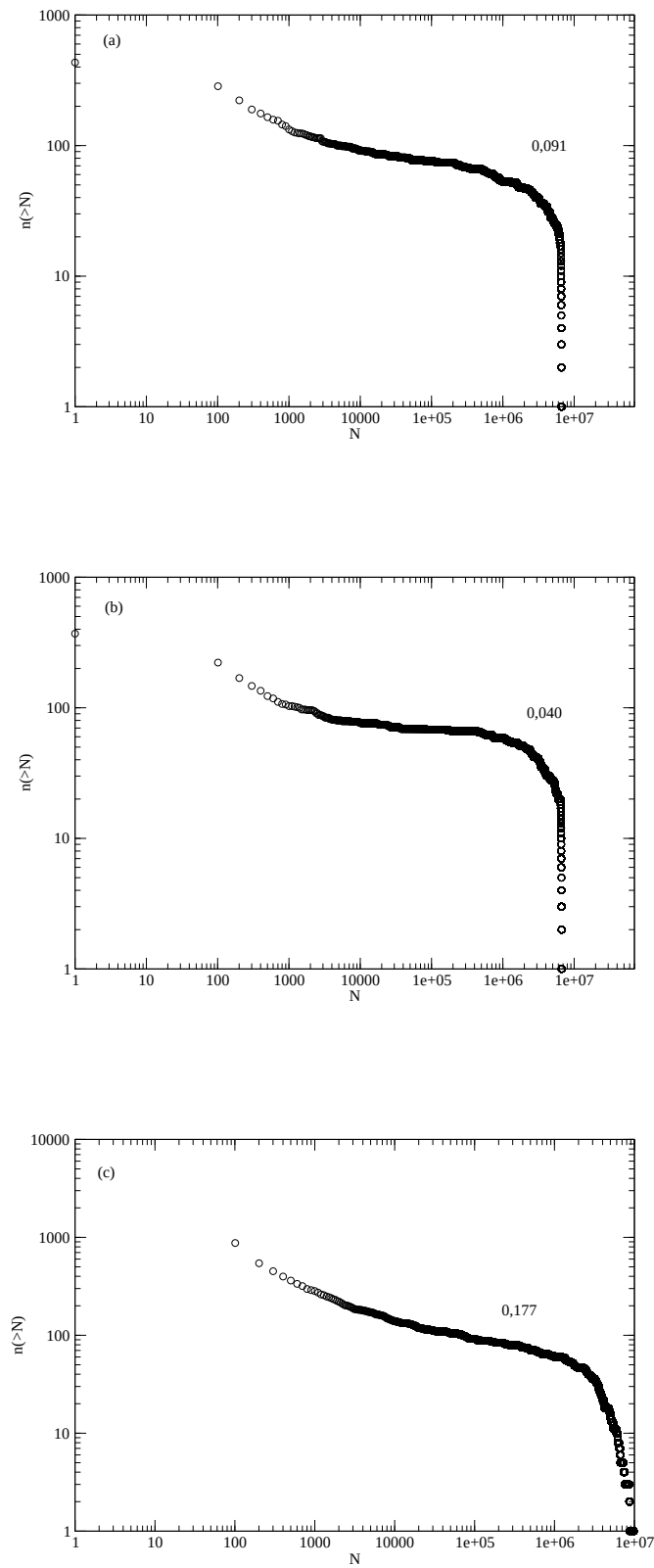


Figura 3.7: Número de línguas com população maior do que N como função de N para $\alpha = 0.5$, $\lambda = 0$ (a), $\lambda = 0.5$ (b) e $\lambda = 0.99$ (c)

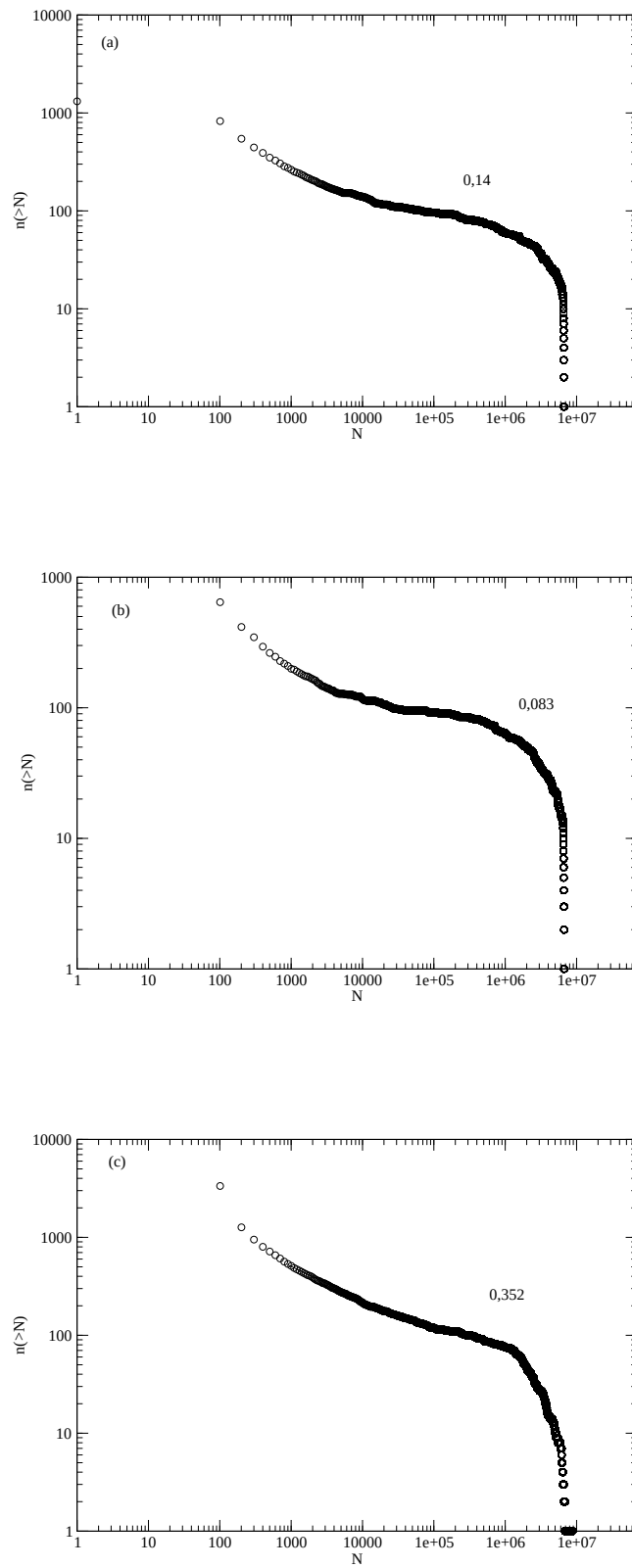


Figura 3.8: Número de línguas com população maior do que N como função de N para $\alpha = 0.9$, $\lambda = 0$ (a), $\lambda = 0.5$ (b) e $\lambda = 0.99$ (c)

Investigamos também o número de línguas com população maior do que N , $n(> N)$, como função de N (Figuras 3.7 e 3.8). As curvas foram estimadas a partir de 40 simulações com $L = 200$ para $\alpha = 0.5$ e $\alpha = 0.9$ para regiões sem correlação ($\lambda = 0$), regiões com correlação intermediária ($\lambda = 0.5$) e regiões altamente correlacionadas ($\alpha = 0.99$). Observando os gráficos verificamos para valores intermediários de N um comportamento em lei de potência dado por $n(> N) \sim N^{-\tau}$. Para $\alpha = 0.9$ e $\lambda = 0.99$ encontramos que τ está em boa concordância com o expoente obtido para a distribuição de línguas no planeta (Gomes et al, 1999) para populações intermediárias. Cada lei de potência prolonga-se por duas ou mais décadas.

3.2 O processo de difusão

A difusão é um fenômeno comum na natureza. Em um processo de difusão num conjunto de elementos que se movem - energia, movimento linear, átomos, moléculas, etc - cada elemento realiza uma trajetória aleatória. Como resultado desse movimento individual altamente irregular, o conjunto se difunde (Pedron, dos Santos Mendes, 2005).

A união entre a formulação estocástica de passeio aleatório e o conceito físico de difusão foi feita através do teorema central do limite, que afirma que a variável aleatória z definida por

$$z = \frac{1}{\sqrt{Nb}} \left(\sum_{j=1}^N \xi_j - Na \right) \quad (3.2)$$

possui distribuição gaussiana

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (3.3)$$

no limite $N \rightarrow \infty$, onde as variáveis aleatórias ξ_j , $j = 1, \dots, N$ são independentes e com mesma distribuição. Para que o teorema central do limite seja válido, basta existir a média a e a variância b (Tome, de Oliveira, 2001).

3.2.1 Passeio aleatório unidimensional

Considere uma partícula se movendo ao longo de uma reta, partindo da origem. A cada intervalo de tempo τ , ela salta uma distância h para direita com probabilidade p e uma distância h para esquerda com probabilidade q . A fim de descrever o movimento da partícula, introduzimos variáveis aleatórias independentes $y_1, y_2, \dots$ que toma valores $+1$ ou -1 de acordo com o salto. A variável y_j indica se no j -ésimo instante a partícula deve saltar para a direita ou para esquerda, assim, ela toma o valor $+1$ com probabilidade p e -1 com probabilidade q . A posição da partícula no instante $t = n\tau$ será $x = hm$ onde $m = y_1 + y_2 + \dots + y_n$ (Tome, de Oliveira, 2001).

A média e a variância de y_j são dadas por

$$a = E(y_j) = (-1) * q + (1) * p = p - q \quad (3.4)$$

$$b = E(y_j^2) - E(y_j)^2 = (1)^2 p + (-1)^2 q - (p - q)^2$$

$$b = 1 - (p - q)^2 = 1 - [p^2 - 2pq + q^2]$$

$$b = 1 - [p(1 - q) - 2pq + q(1 - p)]$$

$$b = 1 - [p - pq - 2pq + q - pq]$$

$$b = 1 - [1 - 4pq]$$

$$b = 4pq \quad (3.5)$$

A função geradora dos momentos é uma ferramenta muito útil. Contudo, tem a desvantagem de que a integral que a define pode, nem sempre ser finita e, portanto, nem sempre existirá. Para contornar esse tipo de problema usaremos uma nova transformada denominada de função característica. Uma das maiores utilidades da função característica é a forma relativamente simples que esta função toma para certas leis de distribuição. Além disso, a partir da função característica é extremamente simples deduzir os momentos de qualquer ordem.

Seja $g(k)$ a função característica da variável y_j dada por

$$g(k) = E(e^{iky_j}) = pe^{ik} + qe^{-ik} \quad (3.6)$$

Para obter a probabilidade $P_n(m)$ de a partícula estar a m passos da origem após n intervalos de tempo, determinamos primeiro a correspondente função característica. Dado que as variáveis têm a mesma distribuição de probabilidade $G(k) = [g(k)]^N$ assim,

$$G_n(k) = [g(k)]^n = (pe^{ik} + qe^{-ik})^n. \quad (3.7)$$

Tomando a expansão binomial $a = pe^{ik}$ e $b = qe^{-ik}$

$$\begin{aligned} (a + b)^n &= \sum_{l=0}^n \binom{n}{l} a^l b^{n-l} \\ G_n(k) &= \sum_{l=0}^n \binom{n}{l} p^l q^{n-l} e^{lik} e^{-nik} e^{lik} \\ G_n(k) &= \sum_{l=0}^n \binom{n}{l} p^l q^{n-l} e^{2lik} e^{-nik} \\ G_n(k) &= \sum_{l=0}^n \binom{n}{l} p^l q^{n-l} e^{ik(2l-n)} \end{aligned} \quad (3.8)$$

e comparando com a definição de $G_n(k)$, dada por

$$G_n(k) = \sum_{m=-n}^n P_n(m) e^{ikm}, \quad (3.9)$$

onde $P_n(m)$ é a distribuição de probabilidade de m . Fazendo a mudança de variável $m = 2l - n$ obtemos

$$P_n(m) = \frac{n!}{\left(\frac{n+m}{2}\right)! \left(\frac{n-m}{2}\right)!} p^{(n+m)/2} q^{(n-m)/2}. \quad (3.10)$$

Portanto a média e a variância de m são

$$E(m) = na = n(p - q) \quad (3.11)$$

$$E(m^2) - E(m)^2 = nb = 4npq. \quad (3.12)$$

Para obter a distribuição de probabilidade para $n \gg 1$, basta utilizar o teorema central do limite, já que as variáveis $y_1, y_2, \dots$ são independentes. Obtemos

$$P_n(m) = \frac{1}{\sqrt{2\pi nb}} \exp\left\{-\frac{(m - na)^2}{2nb}\right\}, \quad (3.13)$$

pois $P_n(m)dm = f(z)dz$ e $dm = \sqrt{nb}dz$.

A posição da partícula no instante $t = n\tau$ será $x = hm$ onde, $m = \frac{x}{h}$ e $n = \frac{t}{\tau}$ logo, a densidade de probabilidade é $f(x) = \frac{P_n(m)}{h}$ da variável x no instante t .

Considerando as relações $c = \frac{ha}{\tau}$, $D = \frac{h^2b}{\tau}$, $b = \frac{\tau D}{h^2}$ e m na equação (3.13) obtemos:

$$f(x) = \frac{h}{\sqrt{2\pi Dt}} \exp\left(-\frac{\left(\frac{x}{h} - \frac{ta}{\tau}\right)^2}{\left(\frac{2tb}{\tau}\right)}\right)$$

$$f(x) = \frac{h}{\sqrt{2\pi Dt}} \exp\left(-\frac{h^2\left(\frac{x}{h} - \frac{ta}{\tau}\right)^2}{h^2\left(\frac{2tb}{\tau}\right)}\right)$$

$$f(x) = \frac{h}{\sqrt{2\pi Dt}} \exp\left(-\frac{\left(x - \frac{tah}{\tau}\right)^2}{\left(\frac{2tbh^2}{\tau}\right)}\right)$$

$$f(x) = \frac{h}{\sqrt{2\pi Dt}} \exp\left(-\frac{(x - ct)^2}{(2Dt)}\right)$$

Comparando com a distribuição normal obtemos os seguintes resultados:

$$E(x) = ct \quad (3.14)$$

$$E(x^2) - E(x)^2 = Dt \quad (3.15)$$

onde c é a velocidade média da partícula e D é o coeficiente de difusão.

Matematicamente, os efeitos da difusão são medidos por um aumento no tempo da variância σ^2 do conjunto de partículas aleatórias em movimento,

$$\sigma^2(t) = E(x^2(t)) - E(x(t))^2 = Dt. \quad (3.16)$$

Quando $E(x(t)) = 0$ a variância é dada por $\sigma^2(t) = E(x^2(t))$, o deslocamento médio quadrático.

A teoria de passeios aleatórios têm sido aplicada em muitas áreas das ciências. Num sistema Euclidiano, o deslocamento médio quadrático de uma trajetória aleatória, $E(x^2(t))$, é proporcional ao tempo t , $E(x^2(t)) \propto t$. A dependência linear no crescimento temporal do deslocamento médio quadrático é uma característica do movimento browniano e, portanto, da difusão usual. É uma consequência do teorema central do limite e da natureza estocástica do processo. Em contraste, a difusão anômala, em geral, tem como característica o crescimento não linear no decorrer do tempo, com $\langle x^2(t) \rangle \propto t^{2/d_w}$, onde d_w é o expoente de difusão (Havlin, Ben-Avraham, 2002).

Com o objetivo de caracterizar o processo difusivo em nosso modelo, investigamos a evolução no tempo da área média $\bar{A}(t)$ ocupada pelas línguas. Cada passo de tempo representa o processo de colonização de um sítio por uma língua. A área média no tempo t é dada por:

$$\bar{A}(t) = \frac{\sum_{i=1}^{D(t)} A_i(t)}{D(t)} \quad (3.17)$$

onde $A_i(t)$ é a área ocupada por uma língua i no tempo t e $D(t)$ é a diversidade de línguas no tempo t . No final do processo de ocupação $t = t_f = L^2$ logo, $\bar{A}(t) \sim L^{2(1-z)}$, desde que a área total é igual a L^2 e $D \sim A^z$. Desta maneira um expoente de difusão d_w pode ser introduzido usando a relação $\bar{A}(t_f) \sim t_f^{(2/d_w)}$ (Havlin, Ben-Avraham, 2002).

Seja as seguintes relações:

$$t_f \sim L^2 \quad (3.18)$$

e

$$\bar{A}(t_f) \sim L^{2(1-z)} \sim t_f^{(1-z)}. \quad (3.19)$$

Pode-se concluir que $\frac{2}{d_w} = 1 - z$ logo, $d_w = \frac{2}{1-z}$. Desta forma d_w assume o valor Browniano padrão $d_w = 2$ se $z = 0$. Para $0 < z < 1$ temos uma difusão anômala e $d_w > 2$, indicando uma progressiva dificuldade para as línguas difundirem-se ou propagarem-se.

Mostramos a evolução no tempo da área média para dois valores distintos de α para regiões sem correlação ($\lambda = 0$) e regiões altamente correlacionadas ($\lambda = 0.99$). Para $\alpha = 0.3$ (Figura 3.9) os expoentes são praticamente iguais e $d_w = 2.16$ e $d_w = 2.14$. Neste caso, os valores dos expoentes de difusão se aproximam de um valor Browniano padrão. Para $\alpha = 0.9$ (Figura 3.10) obtivemos os expoentes $d_w = 2.30$ e $d_w = 2.34$, o que sugere uma maior dificuldade de propagação das línguas em relação ao caso anterior.

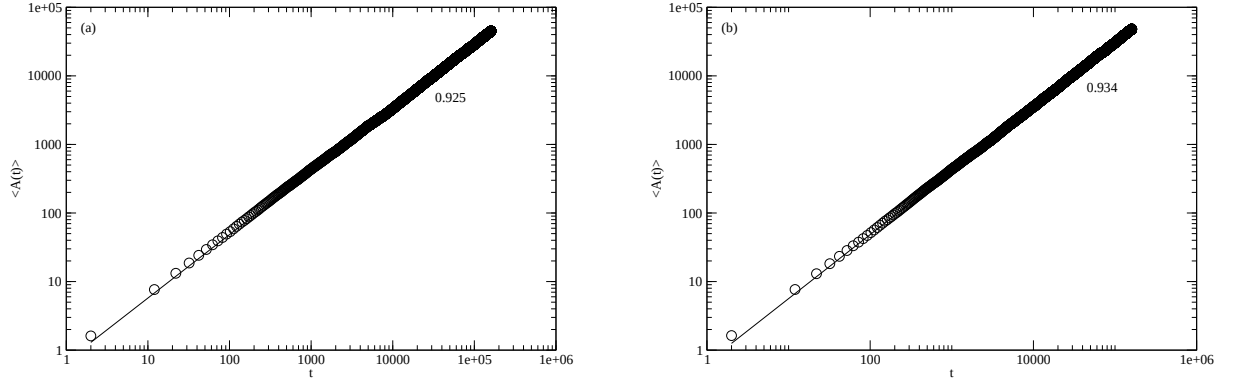


Figura 3.9: Evolução da $\overline{A}(t)$ para $\alpha = 0.3$, $\lambda = 0$ (a) e $\lambda = 0.99$ (b).

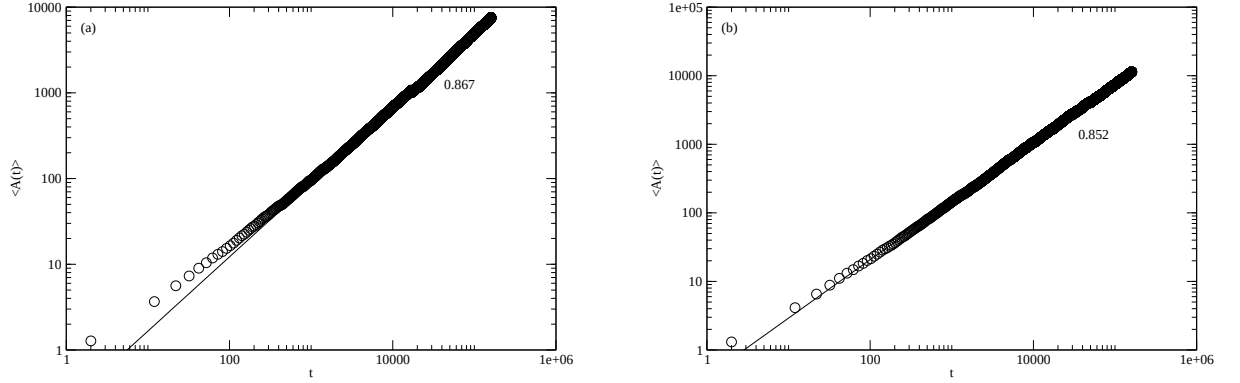


Figura 3.10: Evolução da $\overline{A}(t)$ para $\alpha = 0.9$, $\lambda = 0$ (a) e $\lambda = 0.99$ (b).

Observamos o número de línguas $D(t)$ e a área média $A(t)$ como função do tempo t para uma simples realização da dinâmica com $L = 400$. Comparamos o comportamento de $A(t)$ e de $D(t)$ para três valores distintos de α para relevos sem correlação ($\lambda = 0$) e relevos altamente correlacionados ($\lambda = 0.99$) (Figuras 3.11, 3.12 e 3.13).

Observamos que quando o relevo é muito correlacionado a diversidade cresce mais rápido no início do processo de colonização quando comparamos ao caso onde o relevo é descorrelacionado. Quando $\alpha = 0.3$, observamos longos períodos onde não há surgimento de novas línguas. Esses períodos se tornam mais curtos com o aumento de α .

A área média tem seu valor diminuído a cada surgimento de uma nova língua. Esse efeito pode ser observado mais claramente para valores de α pequenos, desde que a taxa de surgimento de novas línguas é pequena.

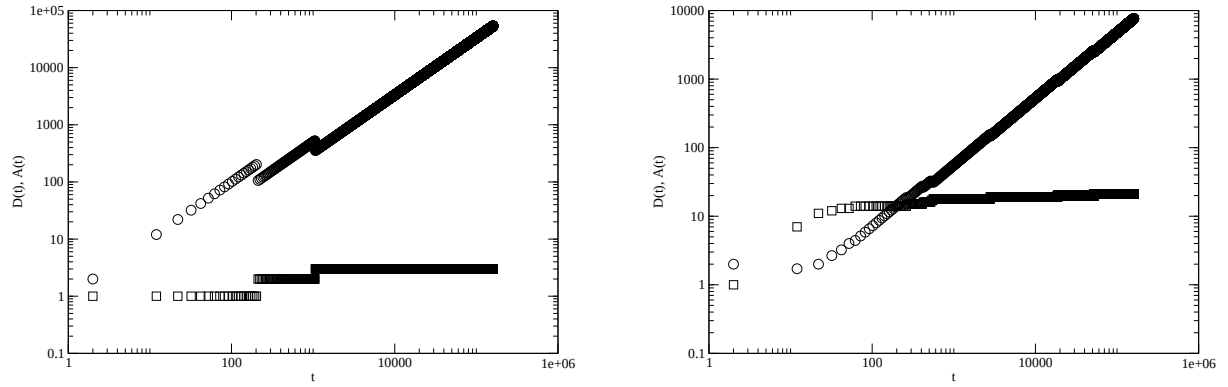


Figura 3.11: Diversidade $D(t)(\square)$ e área média $A(t)(\circ)$ como função do tempo t para $\alpha = 0.3$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

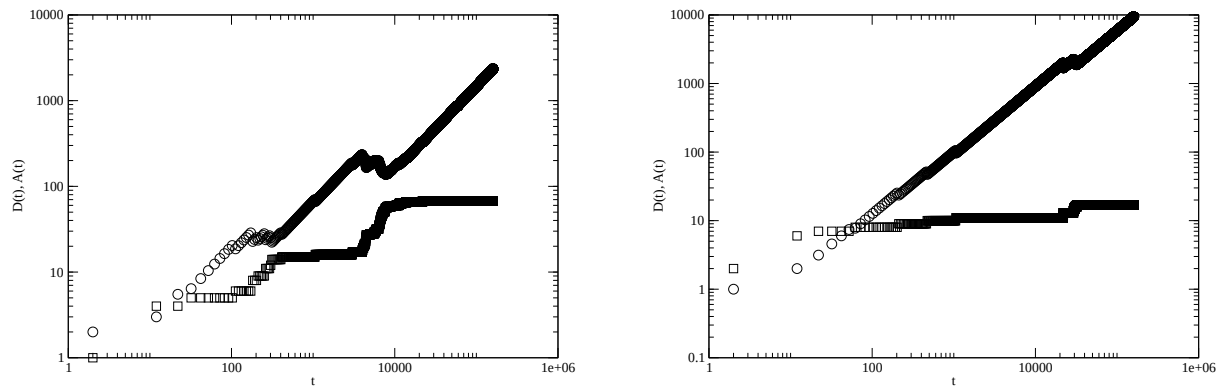


Figura 3.12: Diversidade $D(t)(\square)$ e área média $A(t)(\circ)$ como função do tempo t para $\alpha = 0.7$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

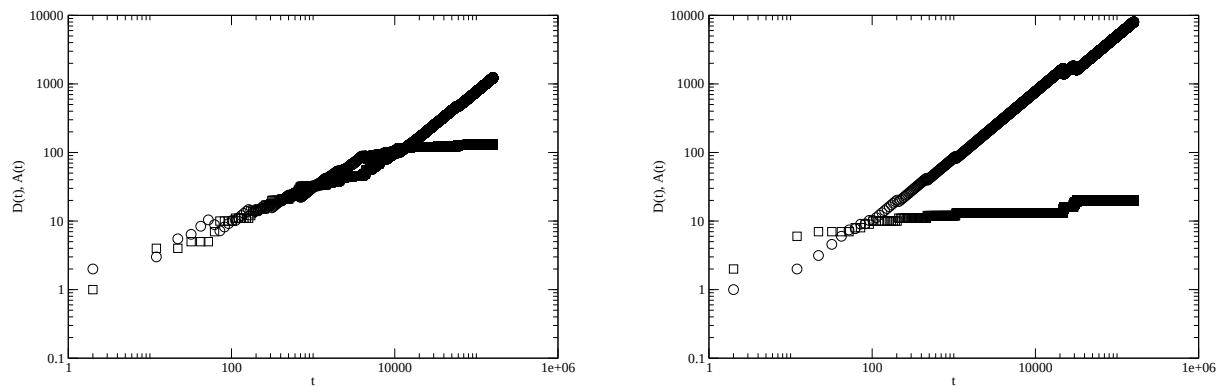


Figura 3.13: Diversidade $D(t)(\square)$ e area media $A(t)(\circ)$ como função do tempo t para $\alpha = 0.9$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

4 Análise do caso com saturação do fitness

4.1 Introdução

Algumas investigações demonstraram que a evolução da diversidade de línguas é afetada por alguns fatores. Entre os principais, temos os fatores geográficos, características econômicas das diferentes populações e a complexidade da língua. Como exemplo, Sutherland (Sutherland,2003) mostrou que a diversidade cresce com a área em florestas e com a altitude, enquanto decresce com a latitude. De acordo com Renfrew (Renfrew,1994), a ocorrência da expansão da agricultura foi responsável pela substituição de muitas populações iniciada a cerca de 10.000 anos, que resultou no desaparecimento de muitas línguas.

Neste capítulo estudamos a dinâmica da diversidade de línguas assumindo que o fitness de cada língua é limitado por um valor máximo (saturação), que é escolhido aleatoriamente de uma distribuição uniforme no intervalo $[1,2000]$. A hipótese de saturação simula fatores como a dificuldade/facilidade de aprendizado de uma língua, economia, entre outros fatores que permitem que algumas línguas se propaguem mais facilmente que outras.

4.2 Resultados

Na Figura 4.1 apresentamos a diversidade D como função da área A . Mantivemos $\alpha = 0.7$ e consideramos quatro valores de λ . Obtivemos cada ponto tomando médias sobre 1000 configurações para $L < 300$ e sobre 500 para $L = 300$ e 400 . Notamos a existência de uma região de escala sobre quase cinco décadas onde $D \sim A^z$ com $z = 0.39 \pm 0.01$ para valores pequenos e intermediários de λ , diferentemente do caso onde não consideramos saturação do fitness, que apresentou dois regimes de escala para os mesmos valores de α e λ . Este resultado está em muito boa concordância com o obtido para a distribuição de línguas na Terra, onde se obteve $z = 0.41 \pm 0.03$ (Gomes et al,1999). Para correlação alta ($\lambda = 0.99$), notamos a existência de dois regimes de escala: o primeiro na região $4 < A < 1.600$ com expoente $z = 0.78 \pm 0.02$ e o segundo para áreas grandes ($1600 < A < 160.000$) que fornece um expoente muito pequeno,

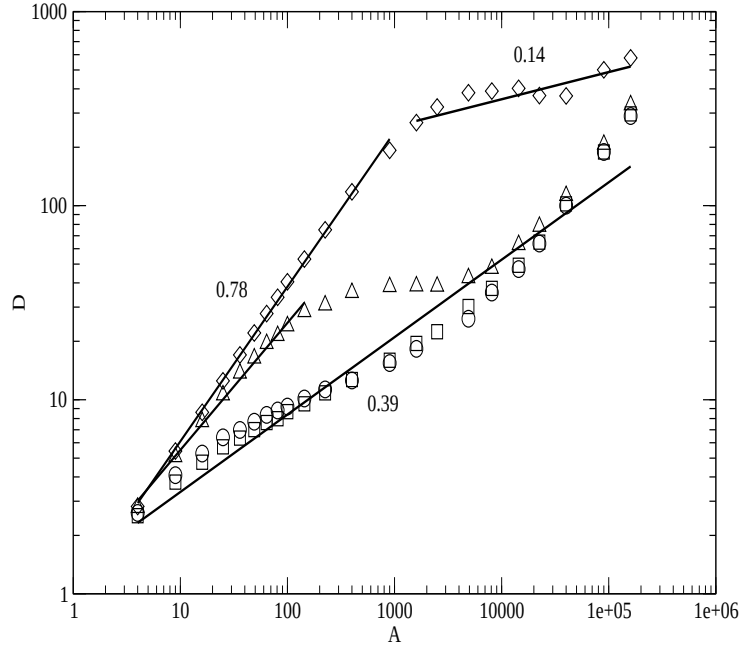


Figura 4.1: Diversidade D como função da área A para $\alpha = 0.7$ e $\lambda = 0(\square)$, $\lambda = 0.5(\circ)$, $\lambda = 0.93(\triangle)$ e $\lambda = 0.99(\diamond)$

$z = 0.14 \pm 0.03$. Também observamos uma tendência de oscilação com relação ao comportamento de escala. Esta tendência também é observada na distribuição real de línguas(Gomes et al,1999).

Por meio da curva onde $\lambda = 0.93$, podemos verificar como se dá a transição do caso onde observamos apenas um regime de escala para o caso onde há duas regiões de escala.

Comparando os resultados com aqueles obtidos para o caso onde o valor do fitness é ilimitado, observamos que com a introdução da saturação obtemos um crescimento mais rápido da diversidade com a área para áreas grandes.

Mostramos na Figura 4.2 a dependência do expoente z em relação ao parâmetro de correlação λ para $\alpha = 0.7$. Observamos que z é aproximadamente constante para valores pequenos e intermediários de λ , mas cresce rapidamente quando $\lambda > 0.8$, significando que a diversidade aumenta rapidamente para valores altos do parâmetro de correlação. O mesmo cenário qualitativo é obtido para outros valores de α .

Investigamos também o efeito do termo de saturação sobre o número de línguas com população maior do que N , $n(> N)$ como função de N para 10 simulações com $L = 400$, $\alpha = 0.7$ e $\lambda = 0$ (Figura 4.3). Para calcular o tamanho da população, assumimos que a população num dado sítio é sua capacidade multiplicada por um fator 100. Encontramos uma lei de potência

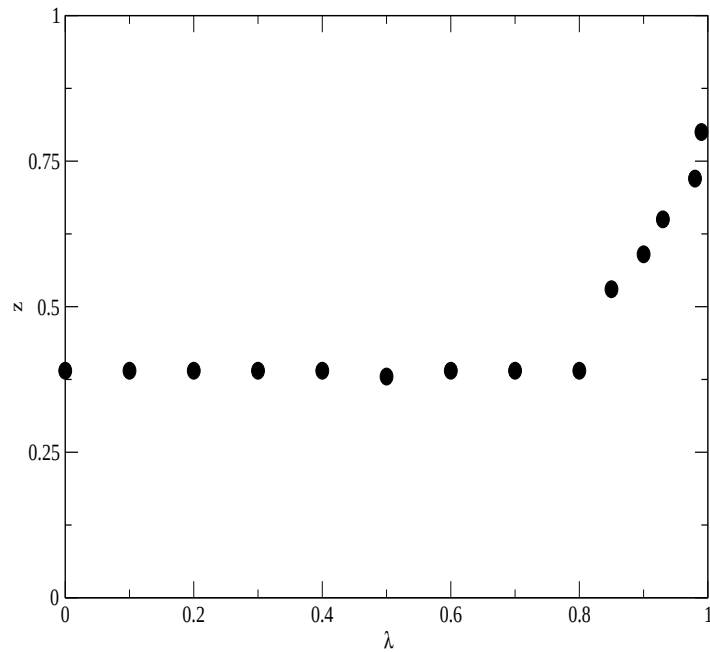


Figura 4.2: Expoente z como função de λ para $\alpha = 0.7$

$n(> N) \sim N^{-\tau}$, onde $\tau=0,52\pm0,01$ ao longo de três décadas. Este resultado esta em boa concordância com a distribuição de línguas no planeta para áreas intermediárias(Gomes et al). Para áreas grandes não obtemos lei de potência.

Observamos o número de línguas n como função do logaritmo do tamanho da população N para três valores distintos de α , a saber, $\alpha = 0.3$ (Figura 4.4), 0.7 (Figura 4.5) e 0.9 (Figura 4.6) para relevos sem correlação($\lambda = 0$) e relevos altamente correlacionados($\lambda = 0.99$). Fizemos um teste de Kolmogorov-Smirnov com o intuito de verificar a normalidade dos dados. Observamos que para os níveis de significância de 1% e 5% rejeitamos a hipótese de normalidade($p - valor < 2.2e^{-16}$). Este cenário qualitativo é observado para outros valores de α .

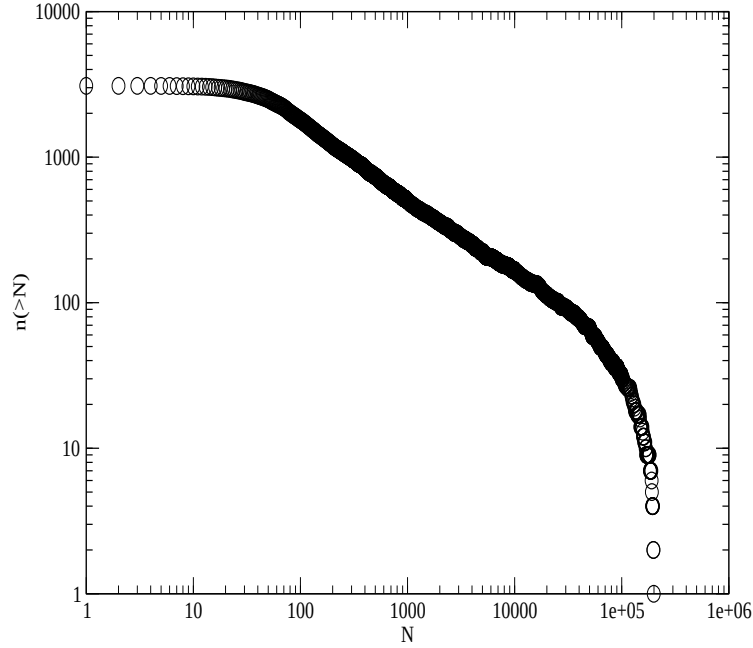


Figura 4.3: Número de línguas com população maior do que N , $n(> N)$, como função de N para $\alpha = 0.7$ e $\lambda = 0$ com $\tau=0,52\pm0,01$.

Observamos o efeito da saturação sobre a evolução no tempo da área média $\bar{A}(t)$ ocupada pelas línguas que é apresentado na Figura 4.7 para 300 configurações com $L = 400$, $\alpha = 0.7$ e $\lambda = 0$. Encontramos duas regiões de escala. A primeira para $A < 2.500$ com $d_w = 2.63$, que caracteriza uma difusão anômala e a segunda para $20.000 < A < 160.000$ com $d_w = 16.7$ verificando que existe uma progressiva dificuldade para as línguas se difundirem no tempo. Esse mesmo comportamento qualitativo é observado para outros valores de α e λ .

Na Figuras 4.8, 4.9 e 4.10 apresentamos o número de línguas $D(t)$ e a área média $A(t)$ como função do tempo t para uma simples realização da dinâmica com $L = 400$, $\alpha = 0.3, 0.7$ e 0.9 e $\lambda = 0$ e 0.99 . Quando comparamos esses resultados com os obtidos para o caso sem saturação observamos que os intervalos de tempo em que a diversidade se mantém constante diminuem consideravelmente, ou seja, a geração de diversidade é observada durante todo o processo de colonização da rede. Este comportamento é observado tanto no caso em que a correlação é nula ($\lambda = 0$), quanto no caso em que o relevo é muito correlacionado ($\lambda = 0.99$).

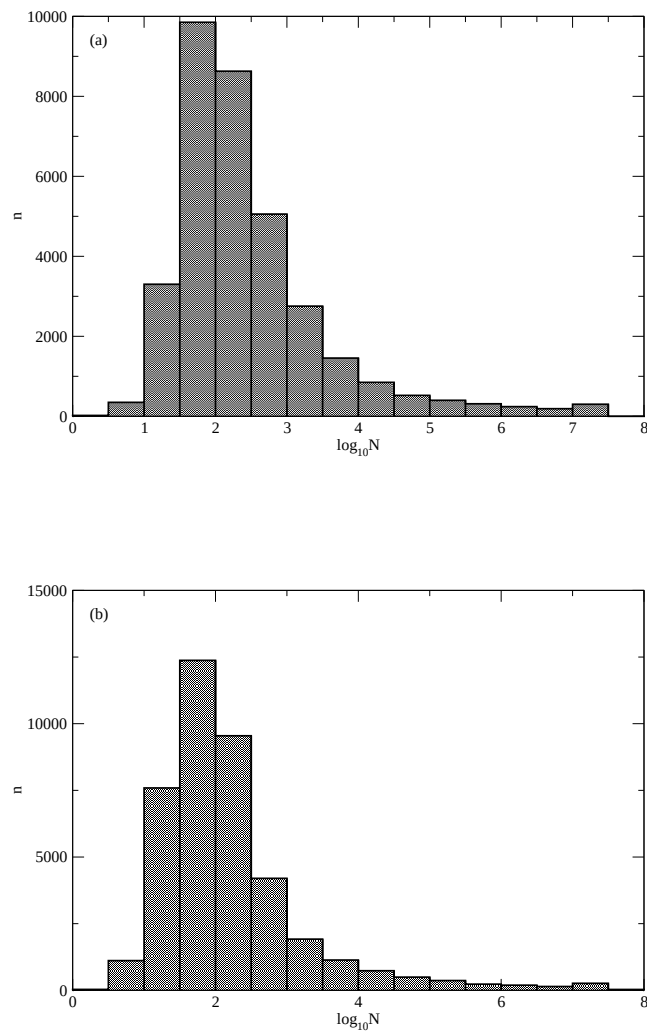


Figura 4.4: Número de línguas n como função do tamanho da população N para $\alpha = 0.3$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

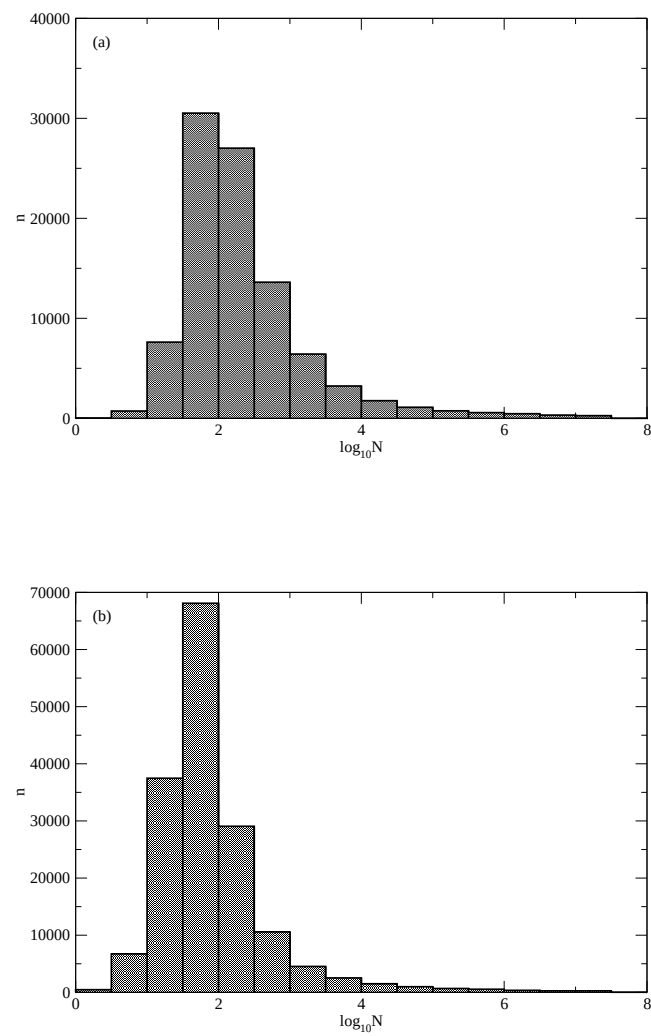


Figura 4.5: Número de línguas n como função do tamanho da população N para $\alpha = 0.7$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

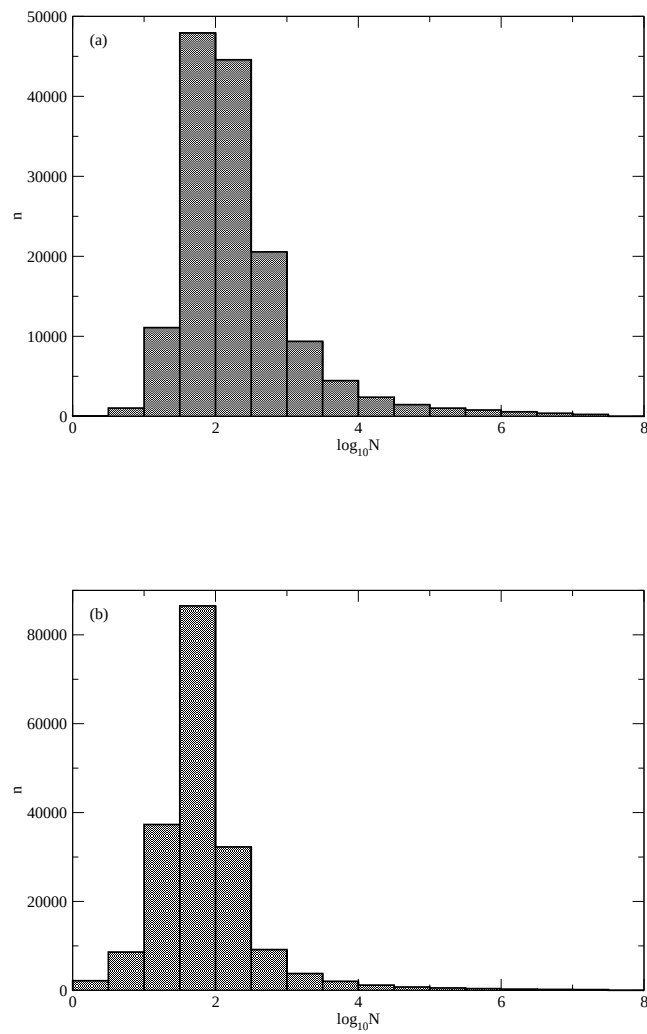


Figura 4.6: Número de línguas n como função do tamanho da população N para $\alpha = 0.9$, (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

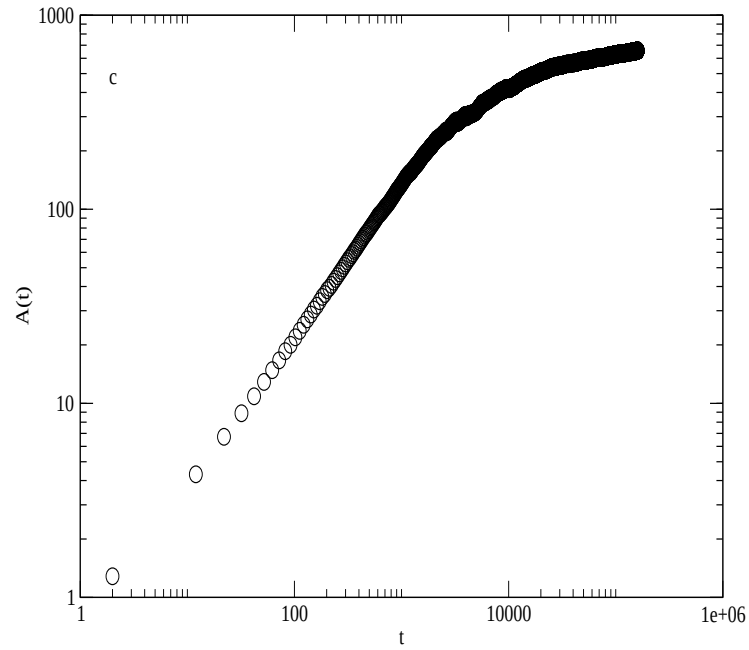


Figura 4.7: Evolução de $A(t)$ para $\alpha = 0.7$ e $\lambda = 0$.

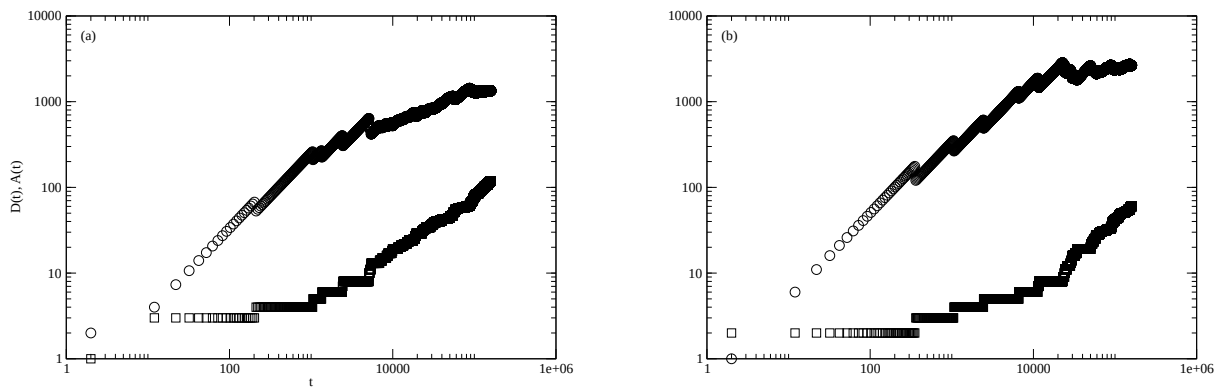


Figura 4.8: Diversidade $D(t)$ ($\square$) e área média $A(t)$ ($\circ$) como função do tempo t para $\alpha = 0.3$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

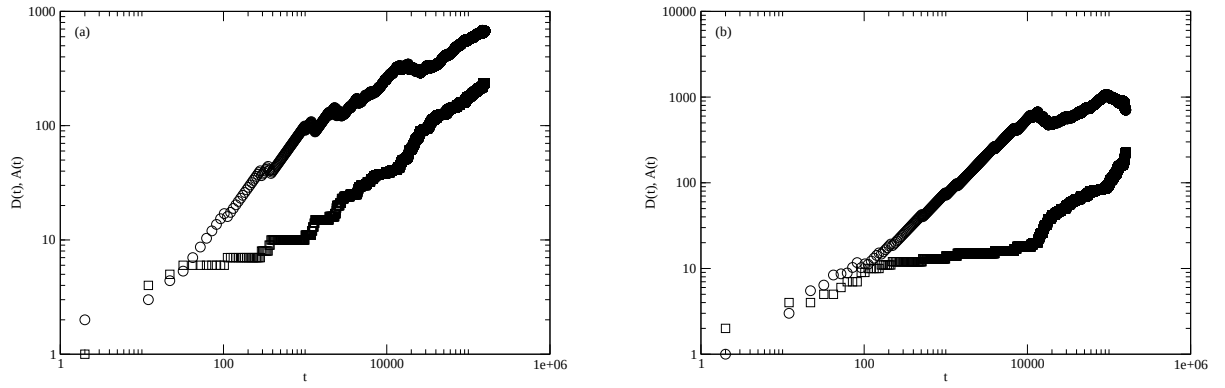


Figura 4.9: Diversidade $D(t)$ ($\square$) e área média $A(t)$ ($\circ$) como função do tempo t para $\alpha = 0.7$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

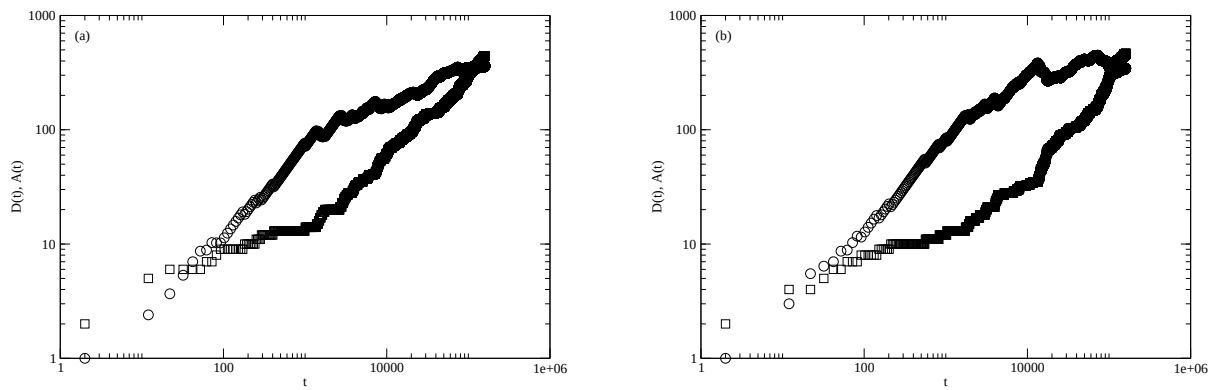


Figura 4.10: Diversidade $D(t)$ ($\square$) e área média $A(t)$ ($\circ$) como função do tempo t para $\alpha = 0.9$ e (a) $\lambda = 0$ e (b) $\lambda = 0.99$.

5 Conclusões

As origens e a evolução das línguas e sua diversidade têm, nos últimos anos, sido tema de grande interesse da comunidade científica que estuda sistemas complexos. As descobertas nos campos da lingüística, genética e arqueologia tem sido importantes para um melhor entendimento de como se deu esta diversificação.

Vários modelos foram introduzidos para o estudo de processos como a competição entre línguas e seus riscos de extinção. Dentre esses modelos, citamos o de de Oliveira e colaboradores (de Oliveira et al,2006) que consiste em um modelo espacial que considera os mecanismos de difusão que geram e mantêm a diversidade de línguas. O modelo é usado para descrever a ocupação de uma dada área por populações que falam diversas línguas.

Nesta dissertação, extendemos o modelo acima introduzindo um grau de correlação entre os recursos das regiões que podem ser ocupadas pelas populações, de forma que podemos observar como se dá o processo de evolução da diversidade lingüística quando variamos o grau de heterogeneidade do relevo. Observamos que para relevos altamente correlacionados a diversidade cresce mais rapidamente como função da área quando comparamos com o caso onde a correlação do relevo é baixa.

Estudamos os casos onde os fitness das línguas são limitados e ilimitados. No caso em que os fitness das línguas são limitados encontramos que para relevos com baixa correlação a dependência da diversidade com a área apresenta um regime de escala que está em boa concordância com a distribuição de línguas na Terra.

Através de uma investigação do processo difusivo, foi possível calcular o expoente de difusão e concluir que a difusão das línguas no nosso modelo é anômala.

Uma seqüência natural do nosso trabalho seria introduzir uma dinâmica de interação entre as populações e também considerar relevos fragmentados, ou seja, considerar a existência de regiões que não podem ser colonizadas, tornando assim o modelo mais realístico.

Referências

- Abrams, D. M., Strogatz, S. H., (2003) *Modelling the dynamics of language death*. Nature **424**, 900.
- Diamond, J. M., (1997) *Linguistics The language steamrollers*. Nature **389**, 544.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., Paabo, S., (2002) *Molecular evolution of FOXP2, a gene involved in speech and language*. Nature **418**, 869.
- Gomes, M.A.F., Vasconcelos, G. L., I.J., Tsang, I.J., Tsang, I.R., (1999) *Scaling relations for diversity of languages*. Physica A **271**, 489.
- Graddol, D., (2004) *The future of language*. Science **303**, 1329-1331.
- Havlin, S., Ben-Avraham, D., (2002) *Diffusion in disordered media*. Advances in Physics **51**, 187.
- Holden, C., (2004) *The origin of speech*. Science **303**, 1316-1319.
- Kosmidis, K., Halley, J. M., Argyrakis, P., (2005) *Language evolution and population dynamics in a system of two interacting species*. Physica A **353**, 595-612.
- Mira, J., Paredes, A., (2005) *Interlinguistic similarity and language death dynamics*. Europhysics Letters, **69(6)**, 1031-1034.
- Nettle, D., (1999) *Linguistic diversity of the Americas can be reconciled with a recent colonization*. Proceeding of the National Academy of Sciences of the United State of America **96**, 3325-3329.
- Nie, M., (1995) *Genetic support for the Out-of-Africa theory of human evolution*. Proceedings of the national academy of sciences **92**, 6720.
- Nowak, Martin A., Komarova, N. L., Niyogi, P., (2001) *Evolution of universal grammar*. Nature **291**, 114.
- Nowak, Martin A., Komarova, N. L., Niyogi, P., (2002) *Computational and evolutionary aspects of language*. Nature **417**, 611-617.
- de Oliveira, V.M., Gomes, M.A.F., Tsang, I.R., (2006) *Theoretical model for the evolution of the linguistic diversity*. Physica A **361**, 361-370.
- de Oliveira, V.M., Campos, P.R.A., Gomes, M.A.F., Tsang, I.R., (2006) *Bounded fitness landscapes and the evolution of the linguistic diversity*. Physica A **368**, 257-261.
- Patriarca, M., Leppänen, T., (2004) *Modelling language competition*. Physica A **38**, 296-299.
- Pedron, I.T., dos Santos Mendes, R., (2005) *Difusão anômala e equações generalizadas de difusão*. Revista Brasileira do Ensino de Física **27**, 251-258.
- Pinasco, J.P., Romanelli, L. (2006) *Coexistence of languages is possible*. Physica A **361**, 355-

360.

Renfrew, C., (1994) *World linguistic diversity*. Scientific American **270**, 104.

Sampat, P., (2001) *Last words*. World Watch **14**, 34.

Schulze, C., Stauffer D. (2005) *Monte Carlo simulation of the rise and the fall of languages*. International Journal of Modern Physics C **16(5)**, 781-787.

Schwämmle, V., (2005) *Simulation of competition of languages with an ageing sexual population*. International Journal of Modern Physics C **16(10)**, 1519-1526.

Stauffer, D., Schulze, C., (2005) *Microscopic and macroscopic simulation of competition between languages*. Physics of Life Reviews **2**, 89-116.

Sutherland, W. J., (2003) *Parallel extinction risk and global distribution of languages and species*. Nature **423**, 276-279.

Tomé, T., de Oliveira, M.J., (2001) *Dinâmica estocástica e Irreversibilidade* Editora da Universidade de São Paulo, São Paulo.