

*ATAS DO II ENCONTRO NACIONAL DO  
GRUPO DE ESTUDOS DE LINGUAGEM  
DO CENTRO-OESTE:  
INTEGRAÇÃO LINGÜÍSTICA, ÉTNICA E SOCIAL*

*Denize Elena Garcia da Silva  
(Organizadora)*

*Brasília*

*2004*

Componentes da Diretoria do Grupo de Estudos de Linguagem do Centro-Oeste  
GELCO

Presidente

Denize Elena Garcia da Silva (UnB)

Vice-Presidente

Maria Zaira Turchi (UFG)

Primeira Secretária

Gláucia Muniz Proença Lara (UFMS)

Segunda Secretária

Hilda Orquídea Hartman Lontra (UnB)

Primeiro Tesoureiro

Manoel Mourivaldo de Almeida (UFMT)

Segunda Tesoureira

Maria Raquel Galán (ULBRA/TO)

E56	<p>Encontro nacional do grupo de estudos de linguagem do Centro-Oeste: integração linguística, étnica e social (2. 2003 : Goiânia)</p> <p>Atas do II encontro nacional do grupo de estudos de linguagem do Centro-Oeste: integração linguística, étnica e social / Denize Elena Garcia da Silva / (organizadora). — Brasília : Oficina Editorial do Instituto de Letras da UnB, 2004. 3v.</p> <p>1. Linguística-Centro-Oeste. 2. Linguística-conferência. 3. Linguística aplicada. 4. Literatura. I. Silva, Denize Elena Garcia da. II. Título.</p> <p>CDU 801(817)(061.3)</p>
-----	--

Endereço para correspondência:

Grupo de Estudos de Linguagem do Centro-Oeste – **GELCO**

UnB – IL – LIV

Campus Universitário Darcy Ribeiro – ICC Norte, subsolo, módulo 20

CEP 70910-900 – Brasília – DF

## **TAXA DE RELAXAÇÃO DE VOCÁBULOS EM UM MÉTODO LEXICO-ESTATÍSTICO APLICADO A LÍNGUAS DA FAMÍLIA PANO**

*Lincoln Almir Amarante Ribeiro (UFMG)*

*Abstract: In the present paper we study the evolution of the Pano language family vocabulary, using an exponential relaxation model with different relaxation taxes for each word meaning in the comparison between any random pair of languages, or either, we calculate the individual contribution of each meaning in the estimate of the separation time of the languages. In the chance we discuss the connection of this procedure with the glottochronological method of Swadesh (1950).*

*Keywords: Historical Linguistics; Léxicostatistics; Indigeneous languages; Pano languages.*

### **Introdução**

Muitos eventos na natureza são simulados por um modelo de relaxação exponencial, dentre os quais podemos citar o tempo de desintegração radioativa de átomos (processo usado na datação por Carbono 14) e as mutações do DNA mitocondrial (processo utilizado nas datações de mutações genéticas em populações humanas).

A relaxação exponencial é caracterizada pelo fato de que a persistência das propriedades de um evento em um determinado tempo independe do seu comportamento histórico. Baseando-se nisso, Swadesh (1950:162), Lees (1953:115) e Dyen (1962:42) propuseram uma teoria léxico-estatística. Para Swadesh, a retenção léxica no comportamento de uma língua não depende de sua história passada. Dyen (op. cit.), ao estudar línguas malaio-polinésias mostrou que tal independência entre a probabilidade de persistência de um evento e sua história é uma condição necessária e suficiente para a validade do modelo de relaxação exponencial.

No presente artigo, pretendemos analisar a evolução do vocabulário de 20 línguas indígenas pertencentes à família Pano, a qual, segundo Shell (1975:15), é uma das “mais conhecidas” da América do Sul, com cerca de 30 línguas faladas por mais ou menos 30 mil pessoas habitantes de uma área que tem a forma aproximada de um quadrilátero limitado pelos paralelos 3° S e 14° S e pelos meridianos 72° W e 64° W na região amazônica peruana, boliviana e brasileira. Do Norte para o Sul, os Pano estão localizados em terras que vão do rio Amazonas ao alto Madeira e Beni; já do Oeste para Leste, estão na região que vai do Ucayali até as cabeceiras do Javari, Juruá e Purus.

### **1. Preliminares**

A teoria léxico-estatística é baseada nas seguintes hipóteses simplificadoras para o processo de mudança lingüística:

- 1) Existe um conjunto básico e universal de significados não culturais a partir dos quais é possível construir uma lista conveniente;
- 2) em cada língua natural, é possível, em um dado instante de tempo, encontrar uma única representação léxica correspondente a cada um desses significados;
- 3) em pequenos “intervalos de tempo”, a representação de qualquer significado tem um risco constante e pequeno de ser substituído por uma outra representação (não cognata) e

4) a substituição ou não da representação de um significado ocorre independentemente de qualquer outro significado ou dos períodos de tempo em que ela ocorre.

Conforme Sankoff (1969), a formalização dessas hipóteses nos conduz diretamente ao modelo de relaxação exponencial. Seja  $m$  um conjunto de pares de línguas comparadas entre si através de uma lista de  $n$  significados, sendo os pares indexados pela variável inteira  $i = 1, 2 \dots m$  e os significados, pela variável inteira  $j = 1, 2 \dots n$ . De acordo com Sankoff (op. cit.), a probabilidade de que o  $i$ -ésimo par de línguas tenha um par de cognatos para o  $j$ -ésimo significado é dada pela expressão em (1), abaixo:

$$(1) P(i,j) = \exp - t(i)/\tau(j)$$

Em (1), admitimos que cada significado da lista tem uma constante de relaxação  $\tau(j)$  que mede a retenção do vocábulo correspondente à única representação léxica daquele significado. Diz-se que a representação léxica de um significado da lista é retida para um par de línguas se, no instante de sua comparação, tal representação é cognata<sup>1</sup>, ou seja, apresenta semelhança fonológica entre si. Consideremos também que para cada par de línguas exista uma SEPARAÇÃO TEMPORAL  $t(i)$  igual ao dobro do tempo decorrido, desde que as línguas em questão tenham se separado da língua antecedente em sua árvore genealógica.<sup>2</sup> Convencionemos, ainda, que se  $\delta(i, j) = 1$ , o  $i$ -ésimo par tem uma representação cognata. Caso contrário,  $\delta(i, j) = 0$ . Dessa forma, podemos escrever:

$$(2) P(i, j) = \sum_{i,j} \delta(i, j)/N$$

A hipótese de que a constante de retenção independe do par de línguas e do tempo em que ela é medida é o ponto principal da teoria de Swadesh (1950:163). Contudo, esta tem sido muito criticada e, até mesmo foi confundida com a hipótese estocástica de que a evolução do léxico de uma língua se dá através de um processo de relaxação exponencial. Na realidade, tais hipóteses são independentes e é, justamente o caráter aleatório de  $P(i, j)$  que justifica o uso de um modelo estocástico para o parâmetro de retenção. O grande número de palavras e a aleatoriedade de cognatos tornam viável uma teoria léxico-estatística da evolução das línguas.

Outras críticas ao método léxico-estatístico estão ligadas à afirmação de que não existe uma lista básica de vocábulos. Naturalmente, o desejável seria a existência de uma lista: a) cujas palavras fossem resistentes a mudanças e b) que fosse suficientemente grande para os procedimentos estatísticos. Não há dúvidas de que a primeira restrição é atendida, embora não saibamos o número de glossas que atendem a esta característica. Meillet (1958:84) observa que todas as palavras não apresentam a mesma facilidade de serem importadas e, ainda, que palavras comuns são mais difíceis de serem substituídas. A esse respeito Bloomfield (1933:298) também afirmou que “a semelhança aparece especialmente em aspectos que são comuns na conversa cotidiana”. Ora, mas o que seriam palavras “comuns” ou da “conversa cotidiana”? Não seriam sérias candidatas ao vocabulário básico?

Para determinar a evolução do vocabulário das línguas da família Pano, usaremos o modelo exponencial, permitindo que as taxas de retenção sejam diferentes para os vários significados de um par qualquer de línguas dessa família. Isto é, calcularemos a contribuição individual de cada significado na estimativa do tempo de separação de línguas relacionadas, ou seja, seu grau de relacionamento.

---

<sup>1</sup> Esse método só pode ser aplicado em línguas para as quais já se demonstrou que são pertencentes a uma mesma família através do método histórico-comparativo. Assim, pode-se afirmar que itens da lista que apresentam a mesma forma o mesmo significado sejam realmente cognatos.

<sup>2</sup> O modelo de árvore está pressuposto para a família em questão.

## 2. A análise

Para estabelecer as estimativas, utilizamos a família etnolingüística Pano, considerando um conjunto de 20 línguas: Arara, Amahuaca, Capanahua, Chacobo, Cashibo, Isconahua, Katukina, Kaxarari, Kaxinawá, Marubo, Matis, Matsés, Pacahuara, Panobo (ou Wuariapano), Poyanawa, Shanenawa, Sharanahua, Shipibo, Yaminahua e Yawanawa.

Em geral, o método que ajusta os dados à equação em (1), chamado método de máxima verossimilhança, consiste em achar o máximo da função em (1). Para encontrar os pontos estacionários de  $P(i, j)$ , igualam-se suas derivadas parciais em relação aos parâmetros a zero e resolve-se o sistema de  $n + m$  equações simultâneas não lineares. Essa tarefa árdua levou-nos a adotar uma solução aproximada, a qual é apresentada em Dyen (1967:163) para as línguas da família Malaio-Polinésia. A solução aproximada é obtida através da análise da variância da transformação  $\log \log$  da proporção dos pares de cognatos de palavras nos grupos de significados. Em concordância com Dyen (1967:161), denominaremos **P** (Produtividade) de um significado, a proporção de pares de cognatos encontrados para esse significado em todos os pares de línguas considerados.

Nessas perspectivas, em primeiro lugar, apresentamos na Tabela 1, a seguir, as produtividades em ordem decrescente de  $\log \log$  dos 100 itens propostos por Swadesh (1950:166) para as 20 línguas Pano em análise.<sup>1</sup>

ORDEM	SIGNIFICADO	P	$\log(-\log P)$	$1/\tau$
1	<i>foot</i>	1,0000	-	0,000
9	<i>blood</i>	0,9000	-2,250	0,057
17	<i>bone</i>	0,8053	-1,530	0,118
33	<i>tooth</i>	0,7158	-1,096	0,182
41	<i>two</i>	0,6316	-0,778	0,250
51	<i>bark</i>	0,5242	-0,437	0,382
65	<i>round</i>	0,4105	-0,116	0,484
69	<i>cloud</i>	0,3000	0,186	0,654
85	<i>that</i>	0,2063	0,456	0,858
98	<i>good</i>	0,1053	0,811	1,223
99	<i>all</i>	0,0947	0,857	1,281
100	<i>small</i>	0,0688	0,985	1,454

Tabela 1: Produtividade,  $\log \log$  e tempos de relaxação para os significados em 20 línguas da família Pano.

A constante de relaxação é uma função  $\tau = \varphi(P)$  monotonicamente crescente da produtividade. Para determinarmos tal função, agrupamos os significados em cinco classes formadas a partir da produtividade dos significados. Assim, na Classe 1 estão os significados representados por vocábulos com formas iguais ( $P = 1,0$ ) ou análogas ( $0,9 \leq P < 1,0$ ) nas 20 línguas analisadas; por outro lado, na Classe 5 estão os significados que apresentam maior diversidade formal nas línguas. Na Tabela 2, na seqüência, representamos esse agrupamento:

CLASSE	NÚMERO DE ITENS	DESCRIÇÃO DOS ITENS	PRODUTIVIDADE
1	14	1-14	0,900-1,000
2	19	15-33	0,700-0,899
3	18	34-52	0,500-0,699
4	20	52-71	0,300-0,499
5	29	72-100	0,000-0,299

Tabela 2: Produtividade das classes.

<sup>1</sup> Por questão de espaço, não apresentaremos aqui as diversas listas de palavras das línguas, pois isso implicaria um acréscimo de cerca de 40 páginas a este texto. Contudo, cópias das listas poderão ser obtidas junto ao autor que, aliás, deverá disponibilizá-las na internet em breve. Já a Tabela 1 completa, ou seja, com 100 glossas encontra-se em Amarante Ribeiro (2003).

Nosso estudo prosseguiu com um sorteio (visando manter a aleatoriedade da escolha) entre as línguas, do qual resultou a construção de 10 pares para comparação. Enumeramos esses pares, também de forma aleatória, conforme apresentado na Tabela 3, abaixo:

ORDEM	PAR DE LÍNGUAS
1	Isconahua x Panobo
2	Sharanahua x Kaxinawa
3	Kaxarari x Yawanawa
4	Chacobo x Matsés
5	Marubo x Arara
6	Poyanawa x Yaminahua
7	Amahuaca x Cashibo
8	Matis x Capanhua
9	Shanenawa x Katukina
10	Shipibo x Pacahuara

Tabela 3: Pares (para comparação) de línguas da família Pano.

Em seguida, procedemos a uma análise da variância<sup>1</sup> que mostra que os dados das línguas se ajustam ao modelo estocástico exponencial proposto.<sup>2</sup> Para tanto, verificamos que desprezando os pontos da Classe 1, já que só se pode tomar o logaritmo de números para os quais  $P(i,j) > 1$ , a expressão de  $\log(-\log P(i,j))$  em função de  $\log \tau(j)$ , apresenta como gráfico uma reta cuja equação é:

$$(3) \quad \log(-\log P(i,j)) = 0,601 - 0,991 \log \tau(j),$$

Alem disso, a média dos tempos de relaxação é (desprezando-se a primeira classe para a qual o modelo não se aplica) de  $\langle \tau_{\downarrow}(j) \rangle_i = 3,376$ , o que dará a taxa de retenção média das línguas Pano o valor  $\langle r(j) \rangle_i = r = 0,8623$ , valor este muito próximo do encontrado por Swadesh (1950) para a lista de 100 palavras. Assim, com a equação em (3), podemos calcular os tempos de relaxação expressos na Tabela 1. A figura abaixo é um gráfico dessa expressão:

<sup>1</sup> Para ver os procedimentos em detalhes, consultar Amarante Ribeiro (2003).

<sup>2</sup> Análises e cálculos matemáticos feitos a partir de aqui são inteiramente paralelos ao estudo de Dien (1962).

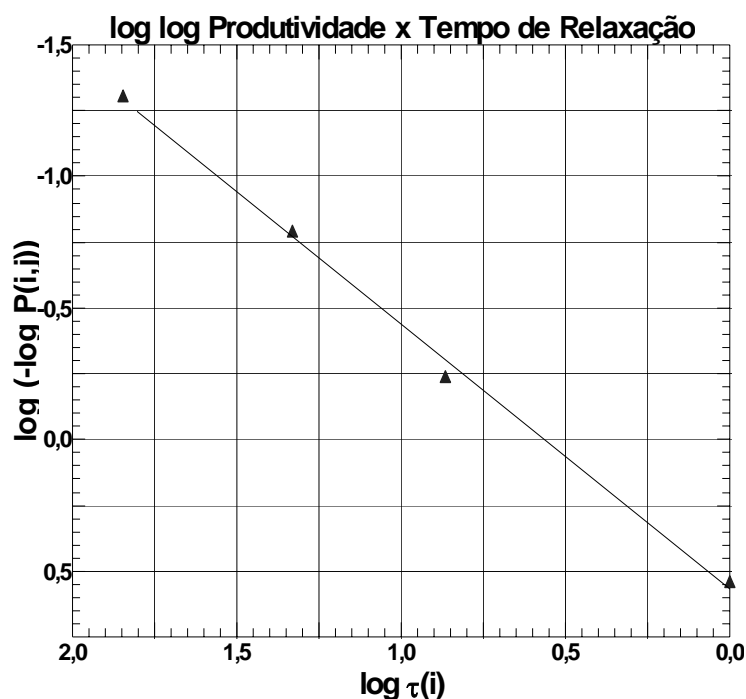


Gráfico 1: *log log* da Produtividade em função do tempo de relação.

Os pequenos triângulos na Figura I são pontos experimentais obtidos a partir dos dados das quatro classes das línguas Pano. A reta representa os pontos teóricos. Parece-nos que o modelo de relaxação exponencial se aplica muito bem à mudança histórica do léxico de línguas, pois foi aplicado com sucesso na análise da família Malaio-Polinésia por Dyen (1967:63) e no presente estudo para uma família de línguas completamente diferente genético e tipologicamente. O Gráfico 1 A partir de um ajuste cuidadoso dos dados disponíveis das línguas a uma particular função de relaxação, é possível calcular os tempos de separação dos diversos pares de línguas que compõem a família e assim descrever sua evolução histórica. No entanto, tal ajuste permite determinar tempos de separação relativos e carece de outros argumentos históricos para obtermos datas absolutas. Contudo, a partir dessas datas relativas é possível construir um diagrama de árvore das línguas da família e, portanto, sua classificação interna. É possível ainda calcular os tempos de relaxação de cada vocábulo da lista e assim tentar obter um tempo de relaxação médio válido para toda a lista.

### Conclusão

Nosso intuito neste artigo foi mostrar que a variação do léxico de uma língua segue um modelo de relaxação exponencial com uma taxa de relaxação que varia de vocábulo para vocábulo. Mostramos que é possível calcular os tempos de separação dos diversos pares de línguas que compõem essa família e, assim, descrever sua evolução histórica.

No entanto, como pudemos observar, esse ajuste somente permite determinar tempos de separação relativos. Para obtermos datas absolutas, seria necessário recorrer a outros argumentos históricos. Desse modo, é possível construir um diagrama de árvore das línguas da família e, portanto, sua classificação interna.

**Referências bibliográficas**

BLOOMFIELD, L. (1933). *Language*. Chicago: University of Chicago Press. p. 298-299.

DYEN, I. (1962). The lexicostatistical classification of the Malayo-Polynesian languages. *Language*. n. 38, p. 38-46.

DYEN, I.; JAMES A.T. & COLE J. W. L. (1967). Language divergence and estimated word retention rate. *Language*. n. 43, v. 1, p. 150-171.

LEES, R. B. (1953). The basis of glottochronology. *Language*. n. 29, p. 113-127.

MEILLET A. (1958). *Linguistique Historique et Linguistique Générale*, tome 1. Paris: Librairie Ancienne Honoré Champion Editeur, p.84.

SANKOFF, D. (1969). Simulation of word-meaning stochastic processes. *International Conference on Computational Linguistics*. Preprint, n. 49. Coling.

SHELL, O. A. (1975). *Estudios Panos III: Las Lenguas Pano y su Reconstrucción*. 1 ed. Lima: ILV SLP, n. 12.

SWADESH, M. (1950). Salish International relationships. *International Journal of American Linguistics*. n.16, p. 157-167.