

Determinação da Taxa de Relaxação de Vocábulos em um Processo de Relaxação Exponencial nas Línguas Indígenas da Família Pano

Lincoln Almir Amarante RIBEIRO
Universidade Federal de Minas Gerais

Resumo: No presente artigo, apresentamos um estudo da evolução do vocabulário de línguas indígenas da família Pano. Para tanto, adotamos um modelo de relaxação exponencial com taxas de relaxação diferentes para cada um dos significados de palavras comparadas entre pares (aleatórios) de línguas dessa família, ou seja, calculamos a contribuição individual de cada significado na estimativa do tempo de separação das línguas. Na oportunidade, discutimos também a conexão deste procedimento com o método glotocronológico de Swadesh (1950).

Palavras-chave: Lingüística Histórica; Léxico-Estatística; Línguas Indígenas; Línguas Pano.

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 argue the connection of this procedure with the glottochronological method of Swadesh (1950).

Key words: Historical Linguistics; Lexicostatistics; Indigeneous languages; Pano languages.

Resumen: En el actual artículo, estudiamos la evolución del vocabulario de las lenguas de la familia Pano, adoptando un modelo de relajación exponencial con diferentes tasas de relajación para cada uno de los significados de palabras entre pares aleatorios de lenguas comparadas,

o sea, calculamos la contribución individual de cada significado en la estimación del tiempo de separación de los idiomas. En la ocasión discutimos la conexión de este procedimiento con el método glotocronológico de Swadesh (1950).

Palabras clave: Lingüística Histórica; Léxico-estadística; Lenguas Indígenas; Lenguas Pano.

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), o tempo decorrido antes de uma eventual falha em um sistema mecânico, o intervalo entre acidentes de trânsito, entre outros.

A propriedade característica de uma distribuição exponencial consiste no fato de que a persistência das propriedades de um evento em um determinado tempo independe do seu comportamento histórico. Baseando-se nessa hipótese foi que os lingüistas Swadesh (1950, 1951, 1952, 1955), Lees (1953) e Dyen (1962, 1965) propuseram uma teoria léxico-estatística. Para Swadesh, o precursor da teoria, a retenção de palavras no comportamento histórico de uma língua não depende da história de retenção de tais palavras. Além disso, Dyen (1962, 1965), ao estudar línguas malaio-polinésias, demonstrou 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.

Tendo em mente tais pressupostos, analisaremos, neste artigo, a evolução do vocabulário de 20 línguas indígenas pertencentes à família Pano que, de acordo com Erikson (1998), possui uma grande homogeneidade territorial, lingüística e cultural.

Para lingüistas como Shell (1975), a família Pano é uma das "mais conhecidas" da América do Sul, com cerca de três dezenas de línguas faladas por cerca de 30 mil pessoas, as quais habitam 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. Como é comum ocorrer entre as populações indígenas, as etnias Pano ocupam sempre áreas às

margens de rios. Assim, do Norte para o Sul estão localizados em terras que vão do Rio Amazonas ao alto Madeira e Beni e do Oeste para Leste, estão na região que vai do Ucayali até as cabeceiras do Javari, Juruá e Purus.

Antes de darmos início à nossa análise, apresentaremos, na seqüência, algumas preliminares acerca do modelo de relaxação exponencial, da teoria léxico-estatística e, ainda, do método glotocronológico.

1 Preliminares

A teoria léxico-estatística apresenta algumas hipóteses simplificadoras para o processo de mudança lingüística que podem, em certos casos, comprometer a aplicabilidade prática do método, mas certamente as suas limitações não são os aspectos estocásticos da teoria. Tais hipóteses são:

- 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 demonstrou Sankoff (1969), a formalização dessas hipóteses nos conduz diretamente ao modelo de relaxação exponencial. Para representar tal formalização, consideremos um conjunto de m pares de línguas comparadas entre si através de uma lista de n significados, sendo os pares indexados pela variável inteira $i = 1, 2 \dots m$, enquanto 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, 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. Convencionemos, ainda, que se $\delta(i, j) = 1$, o i-ésimo par tem uma representação cognata. Caso contrário, $\delta(i, j) = 0$. $\delta(i, j)$ é a Função Delta de Kroenecker. Dessa forma, podemos escrever:

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

A conexão com a teoria de Swadesh, na forma como foi apresentada por Lees (1953), é direta. Nessa teoria, $P(i, j)$ é dada por:

$$(3) P(i, j) = r(j)^{2t(i)}$$

em que $r(j)$ é a constante de retenção de cognatos para o j-ésimo par da lista de significados e $t(i)$ o tempo de separação do par. Na teoria de Swadesh, há a hipótese de que a taxa de retenção de cognatos seja a mesma para todos os significados da lista léxico-estatística, de modo que a expressão (3) se torna:

$$(4) P(i, j) = r^{2t(i)}$$

Tomando o logaritmo em qualquer base da expressão em (4) e explicitando o valor de $t(i)$, encontraremos imediatamente a fórmula de Lees (1953), ou seja,

$$(5) t(i) = \log P(i,j) / 2 \log r$$

para o tempo de separação das línguas do par em função do número de cognatos da lista. Combinando as fórmulas (1) e (3), encontraremos a relação entre as constantes de relaxação e de retenção, a saber:

$$(6) 1/\tau(j) = -2 \log r(j)$$

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). Entretanto, esta tem recebido muitas críticas, sendo, inclusive, confundida na literatura 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. Por exemplo, Bergsland & Vogt (1962) apresentaram quatro casos em que o tempo de separação de línguas não é tão preciso (sendo três muito baixos e um muito alto) e rejeitaram a teoria de Swadesh baseados nesse fato. Na mesma linha de argumentação, Fodor (1961) também rejeitou a léxico-estatística. Chretien (1962) calculou uma quantidade enorme de funções binomiais ordinárias para provar que t e, portanto, $P(i, j)$ são variáveis aleatórias e, por isso, não seriam adequadas para uma formulação matemática aceitável da teoria de Swadesh.

Curiosamente, esse desconhecimento básico da natureza de estimativas estatísticas é uma característica não só dos muitos críticos da léxico-estatística como também de alguns daqueles que utilizam o método. Ora, mas não é justamente o caráter aleatório de $P(i, j)$ que justifica o uso de um modelo estocástico para o parâmetro de retenção? Claramente, a situação é semelhante ao modelo do comportamento de um gás ideal. Um gás consiste em um volume finito com um grande número de partículas realizando um movimento rápido, em que cada uma delas é descrita de maneira determinística pelas leis de Newton. Assim, aplicando essas leis, podemos conhecer em cada tempo a posição e a velocidade precisas de cada partícula. Do ponto de vista prático, no entanto, isso seria impossível devido ao grande número de partículas presentes no gás. O que é possível (e nisso consiste o trabalho de Maxwell, Boltzmann e Einstein), seria considerar a natureza de cada

partícula como um processo aleatório e construir o modelo. É justamente a complexidade e a grande dificuldade de se fazer previsões acerca do comportamento individual das partículas que possibilitam um processo estocástico para os gases. Em situação análoga, o grande número de palavras e a aleatoriedade de cognatos tornam viável uma teoria léxico-estatística da evolução das línguas.

Outro ponto que merece destaque nas críticas feitas ao método léxico-estatístico é a 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) 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) 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?

Quanto à determinação da quantidade de itens de que deve ser constituída uma lista de palavras, um procedimento presente na metodologia das ciências pode nos ser bastante útil nesse sentido. Quando duas tendências contrárias: a) uma de aumentar o número de palavras na lista para atender às exigências estatísticas e b) outra de diminuir, já que poucas palavras podem ser consideradas úteis para a lista no sentido de serem difíceis de sofrer mudanças, então, deve haver um ponto de equilíbrio entre essas duas tendências que favoreça o surgimento de uma lista que atenda aos dois requisitos de forma simultânea. Naturalmente, a lista será constituída pelo vocabulário básico das línguas. Por esse motivo, devemos evitar usar critérios tais como elaborar listas com números “redondos” de itens, tais como 100, 200, 500, entre outros.

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. Nesse caso, altas taxas de retenção sinalizam para um maior grau de relacionamento e, inversamente, baixas taxas de retenção nos levam a um menor grau de relacionamento. Duas estimativas para a constante de retenção são conhecidas: a) uma feita por Lees (1953), comparando a evolução de 13 línguas para listas de 215 palavras, obtendo um valor de $r = 0,81\%$; b) a outra, feita por Swadesh (1950), comparando o Inglês antigo com o moderno a partir de uma lista de 100 palavras, obtendo a taxa de 86%. Antes de prosseguir, notemos que a probabilidade de um conjunto de determinações é dada por:

$$(7) P = \prod_{j: \delta(i,j)=1} P(i,j) \cdot \prod_{j: \delta(i,j)=0} Q(i,j), \text{ em que}$$

$$(8) Q(i,j) = 1 - \exp(-t(i)/\hat{o}(j))$$

é a probabilidade de que a representação lexical do j -ésimo significado do i -ésimo par não seja cognata. A expressão (7), acima, é escrita sob a hipótese de que todos os cognatos são estatisticamente independentes uns dos outros.

Em vista desses apontamentos preliminares, passemos imediatamente à nossa análise propriamente dita.

2 A Análise

Como já o dissemos em nossa introdução, para estabelecer nossas estimativas, utilizamos línguas da família etnolingüística Pano, mais especificamente consideramos um conjunto de 20 idiomas, ou seja, $m = 20$. Essas línguas são conhecidas pelos seguintes nomes: Arara, Amahuaca, Capanahua, Chacobo, Cashibo, Isconahua, Katukina, Kaxarari, Kaxinawá, Marubo, Matis, Matsés, Pacahuara, Panobo (ou Wuariapano), Poyanawa, Shanenawa, Sharanahua, Shipibo, Yaminahua e Yawanawa.

A lista de palavras utilizada em nossa empreitada contém 100 itens e foi organizada por Swadesh (1950). Os itens serão mantidos em língua inglesa visando resguardar o significado original das palavras escolhidas pelo autor da lista. Isto porque entendemos que, sendo o

método baseado em comparações léxicas, é necessário que os significados das palavras não mudem ou apresentem-se dúbios. Afinal, como temos visto, são recorrentes os estudos em que inadvertidamente os pesquisadores que utilizam a lista de Swadesh em seus trabalhos de campo, ao “traduzirem-na” para sua língua materna se esquecem de que erros de tradução podem se propagar com facilidade. É comum, por exemplo, o pesquisador traduzir a palavra ash como “cinza” (pó, resíduos da combustão de certas substâncias), mas ao coletar o dado correspondente a tal significado na língua pesquisada, ocorrer de seu informante fornecer-lhe outro significado, no caso, aquele relativo à “tonalidade” (cor). Diante disso, parece-nos fundamental manter a lista de itens para comparação na língua de origem de Swadesh.

Em geral, o método para se fazer o ajuste dos 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. Sem dúvida, essa é uma tarefa árdua, o que nos levou a adotar uma solução aproximada, a qual é apresentada em Dyen (1962, 1965) para as línguas da família Malaio-Polinésia. A solução aproximada é obtida através de uma análise da variância da transformação $\log \log$ da proporção dos pares de cognatos de palavras nos grupos de significados. Em concordância com Dyen (1962, 1965), denominaremos P (Produtividade) de um significado, a proporção de pares de cognatos encontrados para esse significado em todos os pares de línguas considerados.

É preciso notar vários pontos que diminuem a chance de cometermos enganos em nossa análise, ao contarmos cognatos. Em primeiro lugar, o que chamamos de “cognato” em nosso trabalho é, na realidade, um “cognato aparente”. Isso porque estamos considerando itens que apresentam semelhança fonológica e não há como, embora tentemos evitar esse procedimento, considerar como cognatos verdadeiros, palavras semelhantes por chance, empréstimos, resultados de universais ou de patologias verbais (exemplo: tabus), entre outros. É preciso ressaltar, entretanto, que sempre que possível recorreremos à reconstrução efetuada por Shell (1975), do Proto-Pano, já que sua comparação histórica é a única forma quase segura de garantirmos uma relação genética.

Nessas perspectivas, em primeiro lugar, apresentamos na Tabela 1, abaixo, as produtividades em ordem decrescente de log log dos 100 itens propostos por Swadesh (1950) para as 20 línguas Pano em análise. Por razões de espaço, não apresentaremos aqui as diversas listas de palavras dessas línguas, pois isso implicaria um aumento de cerca de 40 páginas ao presente texto. Contudo, cópias dessas listas poderão ser obtidas mediante consulta ao autor que, oportunamente, pretende disponibiliza-las na internet.

Tabela 1 – Produtividade, log log e tempos de relaxação para os significados em 20 línguas da família Pano

ORDEM	SIGNIFICADO	P	log(-log P)	1/
1	foot	1,0000	-	0.000
2	moon	1,0000	-	0.000
3	neck	1,0000	-	0.000
4	rain	1,0000	-	0.000
5	sleep	1,0000	-	0.000
6	tongue	1,0000	-	0.000
7	you	1,0000	-	0.000
8	white	0,9053	-3,033	0,054
9	blood	0,9000	-2,250	0,057
10	come	0,9000	-2,250	0,057
11	hand	0,9000	-2,250	0,057
12	liver	0,9000	-2,250	0,057
13	meat	0,9000	-2,250	0,057
14	we	0,9000	-2,250	0,057
15	hair	0,8211	-1,624	0,107
16	knee	0,8105	-1,560	0,114
17	bone	0,8053	-1,530	0,118
18	eat	0,8053	-1,530	0,118
19	eye	0,8053	-1,530	0,118
20	feather	0,8053	-1,530	0,118
21	sand	0,8053	-1,530	0,118
22	smoke	0,8053	-1,530	0,118
23	tail	0,8053	-1,530	0,118
24	who	0,8053	-1,530	0,118
25	i	0,7316	-1,163	0,170
26	breast	0,7211	-1,118	0,178
27	night	0,7211	-1,118	0,178
28	cold	0,7158	-1,096	0,182
29	earth	0,7158	-1,096	0,182
30	fat	0,7158	-1,096	0,182

31	leaf	0,7158	-1,096	0,182
32	louse	0,7158	-1,096	0,182
33	tooth	0,7158	-,1096	0,182
34	egg	0,6474	-0,833	0,236
35	fire	0,6368	-0,796	0,245
36	name	0,6368	-0,796	0,245
37	road	0,6368	-0,976	0,245
38	heart	0,6316	-0,778	0,250
39	nose	0,6316	-0,778	0,250
40	sun	0,6316	-0,778	0,250
41	two	0,6316	-0,778	0,250
42	one	0,5684	-0,571	0,307
43	ashes	0,5632	-0,555	0,312
44	fingernail	0,5632	-0,555	0,312
45	fly	0,5632	-0,555	0,315
46	what	0,5608	-0,548	0,317
47	give	0,5579	-0,539	0,317
48	head	0,5579	-0,539	0,317
49	skin	0,5579	-0,539	0,322
50	not	0,5526	-0,522	0,351
51	bark	0,5242	-0,437	0,382
52	horn	0,4947	-0,351	0,377
53	new	0,4895	-0,336	0,388
54	hear	0,4842	-0,336	0,390
55	root	0,4842	-0,336	0,390
56	tree	0,4842	-0,336	0,390
57	lie	0,4789	-0,301	0,400
58	this	0,4789	-0,301	0,400
59	ear	0,4632	-0,262	0,418
60	star	0,4579	-0,247	0,424
61	black	0,4316	-0,174	0,457
62	sit	0,4263	-0,159	0,463
63	man	0,4211	-0,145	0,470
64	woman	0,4211	-0,145	0,470
65	round	0,4105	-0,116	0,484
66	warm	0,3684	-0,001	0,543
67	seed	0,3211	0,128	0,617
68	person	0,3053	0,171	0,645
69	cloud	0,3000	0,186	0,654
70	drink	0,3000	0,186	0,654
71	kill	0,3000	0,186	0,654
72	yellow	0,2947	0,200	0,664
73	red	0,2789	0,244	0,694
74	bird	0,2632	0,289	0,725
75	belly	0,2579	0,304	0,736

76	see	0,2579	0,304	0,771
77	mouth	0,2421	0,349	0,783
78	green	0,2368	0,365	0,783
79	know	0,2368	0,365	0,783
80	water	0,2368	0,365	0,783
81	say	0,2263	0,396	0,807
82	stone	0,2263	0,396	0,807
83	swim	0,2263	0,396	0,807
84	many	0,2211	0,412	0,820
85	that	0,2063	0,456	0,858
86	walk	0,1947	0,492	0,889
87	long	0,1925	0,499	0,895
88	mountain	0,1895	0,509	0,903
89	big	0,1684	0,577	0,968
90	dog	0,1684	0,577	0,968
91	die	0,1632	0,594	0,985
92	stand	0,1632	0,594	1,985
93	full	0,1368	0,688	1,081
94	fish	0,1316	0,707	1,102
95	bite	0,1211	0,747	1,147
96	dry	0,1211	0,747	1,147
97	burn	0,1158	0,7682	1,171
98	good	0,1053	0,811	1,223
99	all	0,0947	0,857	1,281
100	small	0,0688	0,985	1,454

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 desses significados. Assim, na Classe 1, por exemplo, estão os significados que aparecem representados por vocábulos com formas iguais ($P = 1,0$) ou análogas ($0,9 < P < 1,0$) nas 20 línguas analisadas; por outro lado, na Classe 5 estão os significados que apresentam uma maior diversidade formal nessas línguas. Na Tabela 2, abaixo, representamos esse agrupamento:

Tabela 2 – Produtividade das classes

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

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, a seguir:

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

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

Em seguida, procedemos a uma análise da variância. Da relação em (1), tomando o logaritmo de ambos os membros, temos:

$$(9) \quad \log P(i, j) = - t(i)/\tau(j)$$

Se $P(i, j) < 1$, então, podemos ainda tomar o logaritmo de ambos os membros da relação acima, obtendo:

$$(10) \quad \log(-\log P(i,j)) = \log t(i)/\tau(j), \text{ ou}$$

utilizando-se da aditividade da função logarítmica:

$$(11) \quad \log(-\log P(i,j)) = \log t(i) - \log \tau(j)$$

Vamos definir $X(i, j)$ por:

$$(12) \quad X(i,j) = \log(-\log P(i,j)). \text{ Então,}$$

$$(13) \quad X(i,j) = \log t(i) - \log \tau(j)$$

Esta é a expressão básica que nos permite calcular $t(i)$ e $\tau(j)$. Seja $\hat{t}(i)$ e $\hat{\tau}(j)$ respectivamente os tempos de separação e de relaxação estimados em nossos cálculos. Em analogia à equação (1), podemos introduzir as correções $\varepsilon(i,j)$:

$$(14) \quad X(i,j) = \log \hat{t}(i) - \log \hat{\tau}(j) + \varepsilon(i,j)$$

Os erros $\varepsilon(i,j)$ serão pequenos se o modelo se ajustar aos dados. Uma maneira razoável de se obter estimativas destas grandezas é através de médias. Como os erros $\varepsilon(i, j)$ são aleatórios, suas médias devem ser tomadas como nulas, isto é:

$$(15) \quad \langle \varepsilon(i,j) \rangle_i = \langle \varepsilon(i,j) \rangle_j = 0$$

O subscripto indica que a média está sendo tomada em relação ao índice especificado. Tomando-se as médias em relação a i e j na equação (14) e tendo-se em conta (15), teremos:

$$(16) \quad \langle X(i,j) \rangle_i = \log \hat{t}(i) - \langle \log \hat{\tau}(j) \rangle_i \quad \text{e}$$

$$(16) \quad \langle X(i,j) \rangle_j = \langle \log \hat{t}(i) \rangle_j - \log \hat{\tau}(j),$$

onde $\langle \log \hat{\tau}(j) \rangle_i$ e $\langle \log \hat{t}(i) \rangle_j$ são constantes a determinar.

Para se determinar essas constantes, observemos que os t e τ são medidos na mesma unidade de tempo, embora sem dados históricos, não possamos atribuir qualquer unidade física para essas grandezas. Podemos, no entanto, padronizar nossas medidas definindo como unidades dessas grandezas, arbitrariamente, como sendo o valor da constante de relaxação da quinta classe:

$$(17) \hat{\tau}_5 = 1,000 \text{ ou seja } \log \hat{\tau}_5 = 0.$$

Assim, fazendo-se $j = 5$ na equação (17), encontramos:

$$(18) \langle X(i,j) \rangle_{j=5} = \langle \log t(i) \rangle_j - \log \hat{\tau}_5 = \langle \log t(i) \rangle_j$$

Levando-se na equação (17) encontramos:

$$(19) \log \hat{\tau}(j) = \langle X(i,j) \rangle_{j=5} - \langle X(i,j) \rangle_j$$

Como podemos notar, esta expressão permite calcular os tempos de relaxação através das médias de $X(i,j)$.

Para obtermos os dados necessários para os cálculos dos tempos de relaxação de cada significado e dos tempos de separação dos pares de línguas, necessitamos recorrer aos dados. Estes são apresentados nas Tabelas 4 e 5, na seqüência.

Na Tabela 4, apresentamos para cada par de línguas comparado a porcentagem de cognatos encontrada para cada uma das classes de significados:

Tabela 4 – Porcentagem de cognatos para as cinco classes nos 10 pares de línguas

PAR DE LÍNGUAS	TOTAL %	CLASSE DE SIGNIFICADOS j = 1, 2 ... 5				
		1	2	3	4	5
1	0,630	0,857	1,000	0,611	0,700	0,241
2	0,670	1,000	0,895	0,833	0,650	0,276
3	0,450	0,786	0,684	0,556	0,350	0,138
4	0,270	0,571	0,368	0,167	0,300	0,103
5	0,600	0,929	0,947	0,611	0,700	0,138
6	0,550	1,000	0,789	0,611	0,300	0,310
7	0,360	0,714	0,632	0,500	0,200	0,034
8	0,370	0,643	0,526	0,611	0,250	0,069
9	0,640	0,929	0,947	0,778	0,550	0,276
10	0,570	0,929	0,789	0,556	0,400	0,379

Já na Tabela 5 apresentamos o cálculo log log de cada uma das porcentagens de cognatos para cada um dos pares de línguas comparados bem como cada uma das classes de significados que foram mostradas na Tabela 4.

Tabela 5 – Transformações log log das porcentagens de cognatos das cinco classes de significados.

PAR DE LÍNGUAS	SIGNIFICADOS j = 1, 2, ... 5					TOTAL
	1	2	3	4	5	
1	-1,869	0,000	-0,708	-1,031	0,353	-3,255
2	0,000	-2,199	-1,699	-0,842	0,253	-4,487
3	-1,423	-0,968	-0,533	-0,049	0,683	-1,893
4	-0,571	0,000	-0,582	0,186	0,821	1,018
5	-2,610	-2,910	-0,708	-1,031	0,683	-6,576
6	0,000	-1,440	-0,710	0,186	0,158	-1,806
7	-1,089	-0,779	-0,367	0,476	1,218	-0,541
8	-0,817	-0,442	-0,707	0,327	0,983	-0,656
9	-2,610	-2,910	-1,382	-0,514	0,253	-7,163
10	-2,610	-1,440	-0,533	-0,090	-0,030	-4,703
TOTAL	-13,60	-13,09	-7,93	-2,38	5,38	-30,06

Com os dados da Tabela 5, acima, e o uso da fórmula (19) podemos facilmente encontrar os valores dos tempos de relaxação para cada uma das classes de significados. Os resultados são mostrados na Tabela 6, abaixo:

Tabela 6 – Constantes de relaxação

CLASSE DE SIGNIFICADOS j = 1, 2 ... 5	MÉDIAS <X(i,j)> _j	log $\hat{t}(j)$	$\hat{t}(j)$
1	-1,360	1,898	6,673
2	-1,309	1,847	6341
3	-0,793	1,331	3,785
4	-0,238	0,866	2,377
5	0,538	0,000	1,000
GRANDE MÉDIA	-0,632	1,188	

Da Tabela 6, obtemos $\langle \log \hat{t}(j) \rangle_i = 1,188$, de modo que podemos obter a seguinte relação em (20):

$$(20) \log t(i) = \langle X(i,j) \rangle_i + \langle \log \hat{t}(j) \rangle_i = \langle X(i,j) \rangle_i + 1,188$$

Essa relação nos permite encontrar os tempos de separação para qualquer par de línguas, como nos mostram os resultados, na Tabela 7, abaixo:

Tabela 7 – Tempos de separação dos diversos pares de línguas

PAR DE LÍNGUAS i = 1, 2 ...10	MÉDIA <X(i,j)> _i	log $t(i)$	$t(i)$
1	-0,651	0,537	1,711
2	-0,897	0,291	1,338
3	-0,379	0,809	2,246
4	-0,204	1,392	4,022
5	-1,315	-0,127	0,881
6	-0,361	0,827	2,286
7	-0,108	1,080	2,945
8	-1,131	1,057	2,880
9	-1,433	-0,245	0,783
10	-0,941	0,247	1,280

Estamos, agora, em condições de verificar se os dados das línguas Pano se ajustam ao modelo estocástico exponencial proposto. Para tanto, verificamos que $\log(-\log P(i,j))$ em função de $\log \tau(j)$, apresenta como gráfico uma reta conforme visto na por meio do método de regressão linear, desprezando os pontos da Classe 1, já que a equação em (11) só vale para $P(i,j) > 1$. Nesse caso, a equação é:

$$(21) \quad \log(-\log P(i,j)) = 0,601 - 0,991 \log \tau(j),$$

Além disso, verificamos que a média dos tempos de relaxação é (desprezando-se a primeira classe para a qual o modelo não é aplicável) de $\langle \hat{\tau}(j) \rangle_i = 3,376$. Aplicando esse valor na expressão em (6), encontramos para a taxa de retenção média das línguas Pano o valor $\langle r(j) \rangle_i = r = 0,8623$. Esse valor é muito próximo daquele encontrado por Swadesh (1950) para a lista de 100 palavras.

Com a equação em (19), podemos expressar os cálculos dos tempos de relaxação apresentados na Tabela 1. A figura abaixo é um gráfico dessa expressão:

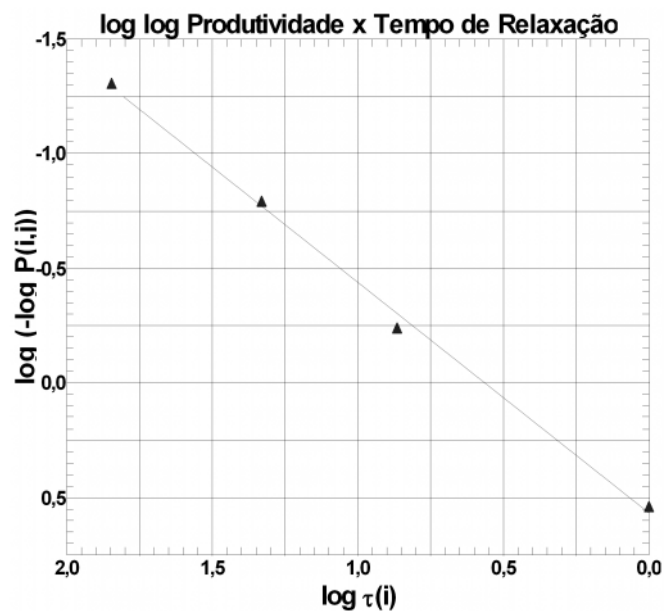


Figura 1 – log log da Produtividade em função do tempo de relação

Os pequenos triângulos, na figura, são pontos experimentais obtidos com os 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). De forma análoga, isso ocorreu em nosso estudo sobre as línguas da família Pano. A Figura 1 fala por si só. Dessa forma, a conclusão de que a variação histórica do léxico de uma língua, pelo menos no contexto de inserção em uma família geneticamente bem definida, como acreditamos ser o caso da família de línguas Pano, se ajusta a um modelo de relaxação exponencial com uma taxa de relaxação que varia de vocábulo para vocábulo. A partir de um ajuste cuidadoso dos dados disponíveis das línguas da família 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. Também é possível calcular, através do ajuste, os tempos de relaxação de cada um dos vocábulos 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. Por meio do ajuste dos dados disponíveis das línguas de uma determinada família a uma particular função de relaxação, procuramos mostrar 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, observando, por exemplo, que a taxa de retenção média para uma

lista de 100 palavras, como aquela proposta por Swadesh (1950), aproxima-se dos 86% obtidos por Swadesh (1955) e quando Dyen (1967) usou listas de 200 palavras a taxa média de retenção ficou próxima do resultado obtido por Lees (1953). A partir das datas relativas, e utilizando-se a taxa de retenção média obtida, poderíamos, por exemplo, fixar $P = (0,86)^2$. Com tal procedimento, obteríamos para $\hat{t}_5 = 1,6$ ao invés do valor unitário apresentado em (18). Dessa forma, a unidade de tempo, ou seja, a unidade em que mensuramos t e \hat{t} seria definida como a constante de tempo médio da Classe 5, estimada em cerca de 1600 anos. Desse modo, é possível construir um diagrama de árvore das línguas da família e, portanto, sua classificação interna.

Para o futuro, estamos propondo outros problemas como o de tentar fazer ajustes para o maior número possível de famílias de línguas conhecidas e o de tentar obter definitivamente uma taxa média de retenção que possa ser usada para línguas para as quais não temos nenhum registro histórico. Além disso, seria interessante também construir uma lista básica de vocábulos que possa ser realmente confiável. Isto pode ser feito, acreditamos, examinando-se as produtividades de cognatos em várias famílias de línguas distantes e adotando-se algum critério de "score", ou seja, uma espécie de "placar", em que significados com produtividade média superior a tal "placar" fizessem parte da lista de palavras. Para tanto, naturalmente, é necessário esquecer o número de elementos da lista.

Referências Bibliográficas

BERGSLAND, K.; VOGT, H. On the validity of glottochronology. *Current Anthropology*, n. 3, v. 2, p. 115-153, 1962.

BLOOMFIELD, L. *Language*. Chicago: University of Chicago Press, 1933.

CHRÉTIEN, C. D. The mathematical models of glottochronology. *Language*, n. 18, v. 1, p. 11-37, 1962.

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

_____. A lexicostatistical classification of the Austronesian languages. Indiana University Publications in Anthropology and Linguistics, Memoir 19, 1965.

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

ERIKSON, P. Uma singular pluralidade: a etno-história Pano. In: CUNHA, M. C. et al. (Org.). *História dos Índios no Brasil*. São Paulo: Shwarcz, 1998.

FODOR, I. The validity of glottochronology on the basis of the Slavonic languages. *Studia Slavica* 7, p. 295-346, 1961.

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

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

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

SHELL, O. A. - *Estudios Panos III: las lenguas Pano y su reconstrucción*. Lima: ILV SLP, n. 12, 1975.

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

_____. Diffusional cumulation and archaic residue as historical explanations. *Southwestern Journal of Anthropology*, n. 7, p. 1-21, 1951.

_____. Lexico-statistical dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96, p. 452-463, 1952.

_____. Toward greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21, p. 121-137, 1955.