

Lincoln Almir Amarante Ribeiro
(*Universidade Federal de Minas Gerais*)

Uma proposta de método quantitativo aplicado à análise comparativa das línguas Pano e Tacana

ABSTRACT

In this paper, we consider a quantitative binary method of comparison of languages based in a statistical analysis of word lists. The method presents reasonable solutions to problems discussed in previous works on the subject. As an application of the method, we perform a comparative analysis of languages of the Pano and Tacana families, considering the items defined as proto-forms. The application of the method straightens the hypothesis about genetic relationship among languages of these two families.

KEYWORDS Historical Linguistics; Quantitative Method; Pano-Tacana Stock.

RESUMO

Este artigo apresenta um método quantitativo binário para a comparação histórica de línguas, baseado em uma análise léxico-estatística. O método mostra uma solução viável para os problemas que foram encontrados em trabalhos de características semelhantes realizados previamente. Como aplicação ilustrativa do método, realizou-se uma análise comparativa de línguas das famílias Pano e Tacana, considerando listas de palavras definidas como Proto-Formas. A aplicação do método resultou num fortalecimento da hipótese de relacionamento genético entre as línguas dessas duas famílias

PALAVRAS-CHAVE Linguística Histórica; Métodos Quantitativos; Tronco Pano-Tacana.

0. PRELIMINARES

Embora pautada somente em listas de palavras, muitas vezes, de línguas que têm entre si apenas a coincidência da localização geográfica, uma provável relação de parentesco entre as línguas faladas pelos povos Pano e Tacana foi sugerida por vários pesquisadores. Dentre estes, podemos citar Groeteken (1907) e Créqui-Montfort & Rivet (1921), os quais notaram semelhanças gramaticais entre as duas famílias; Schüller (1933) conclui que, embora essa relação não seja muito intensa, certamente há uma relação entre os dois grupos de línguas.

Rivet & Loukotka (1952) foram além classificando, ainda, as línguas do grupo Tacana como membros da família Arawak, tal como afirmam:

“... relacionamos com a família Arawak um grupo o de línguas classificadas há bastante tempo com o nome de família Tacana. Se estas línguas têm uma gramática que apresenta grandes analogias morfológicas com a família Pano, seu vocabulário é em grande parte Arawak”. (p.1294).

Nessa mesma perspectiva, Mason (1950) afirmou, ainda que de forma equivocada, como ele mesmo reconsiderou posteriormente (Mason, 1963), que as línguas Tacana tinham uma possível afinidade com o grupo Arawak.

Outros autores como Greenberg (1956) e Swadesh (1959) também usaram as expressões Tacana-Pano e Pano-Tacana em suas classificações, bem como Villarejo (1959), em um trabalho puramente intuitivo (já que as listas de léxico disponíveis não poderiam induzi-lo a tal hipótese), também deduziu que “os huarayos faziam parte da grande tribo nahua”. Aliás, talvez seja essa intuição que tenha influenciado o trabalho de Key (1968), em que postulou a existência de um tronco Pano-Tacana, embora com um grau de parentesco bastante afastado entre as respectivas famílias. Entretanto, esse estudo, ainda que muito criticado por Girard (1970) e Chavarria (1973), constitui-se atualmente como a principal fonte para uma hipótese praticamente definida como verdadeira (Suarez, 1973; Wistrand-Robinson, 1989).

Outra perspectiva de parentesco foi apontada ainda por Loos (1973) ao relacionar os grupos Pano e Tacana ao Moseten:

“La familia lingüística Pano consiste de unos 30 ó más idiomas esparcidos entre el Amazonas y Bolivia en ambos los lados entre el Perú y Brasil. Pertenecen al grupo mayor Pano-Tacana y posiblemente Pano-Tacana-Moseten, pero más allá no está claro con qué tronco está emparentado”.

Finalmente, em um trabalho mais recente, Lanes (2000) utiliza-se do método glotocronológico e da léxico-estatística para postular que a família Pano trata-se, na realidade, de um tronco. Isso se deveria ao afastamento da língua Matsés dos demais idiomas Pano considerados por esse autor. Contudo, nesse estudo não há menção às línguas Tacana.

Como vemos, muitas hipóteses foram levantadas, mas ainda não contamos com conclusões concretas sobre o tema. Em todo o caso, como previram d’Ans et al, (1973), a problemática Pano-Tacana haveria de converter-se em um dos focos de interesse primordial da lingüística comparativa amazônica. E os conhecimentos nesse campo acabariam contribuindo para a elaboração de classificações lingüísticas realmente consistentes em comparação com as mencionadas. Corroborando essa afirmação, apresentamos neste artigo uma proposta de método quantitativo de comparação, a qual visa contribuir para a resolução da problemática que envolve a classificação das línguas Pano e Tacana como pertencentes a um mesmo agrupamento genético lingüístico.

1. A DESCRIÇÃO DO MÉTODO

Nosso método de “prova” (ou não) de parentesco entre duas línguas baseia-se no conceito de *redução ao absurdo* muito utilizado na demonstração de teoremas na Matemática e na Lógica. Para adaptarmos esse conceito à análise lingüístico-comparativa recomendamos atenção a alguns procedimentos, os quais passamos a descrever na seqüência.

Dado um par de línguas quaisquer e supondo, a priori, que elas não sejam geneticamente relacionadas, inicialmente faremos um cálculo da chance de existência de pares de cognatos, os quais, se encontrados, supostamente deverão ser resultados de empréstimos ou simplesmente do acaso. Esse cálculo será feito com base nas propriedades fonológicas e fonotáticas das línguas em análise. Como resultados do cálculo, encontraremos, então, as inúmeras probabilidades, **p**, de se achar em uma lista de, por exemplo, **n** palavras com significados idênticos das duas línguas, mais de um número **r** de cognatos. Se a probabilidade de encontrarmos, digamos, mais de **r** (= 10) cognatos em uma lista de **n** (= 200) palavras for muito pequena, suponhamos **p** (=1 em 1000), e se encontramos em uma situação real mais de 10 cognatos, então, a hipótese inicial de que as línguas são geneticamente não relacionadas, é falsa. Isso nos leva a concluir, por *redução ao absurdo*, que as línguas são na realidade geneticamente relacionadas.

Para a aplicação do método devemos sistematizá-lo com base nas seguintes etapas: 1) determinação do par de línguas que serão comparadas; 2) estabelecimento do inventário de fonemas dessas línguas; 3) levantamento dos tipos silábicos encontrados nas línguas; 4) Definição do número de ocorrências de cada fonema em suas diversas posições dentro de um vocábulo; 5) escolha do critério de equivalência entre fonemas visando estabelecer os cognatos nas línguas, ou seja, o que deverá ser considerado como equivalente em termos de fonemas nessas línguas.

Todas essas etapas serão muito importantes; contudo, as línguas a serem comparadas e a escolha do critério de equivalência dos fonemas serão fundamentais para a confiabilidade nos resultados finais.

Com respeito às línguas, ressaltamos que o método aqui apresentado é em princípio binário, isto é, ele compara duas línguas de famílias supostamente diferentes¹ a fim de mostrar uma possível relação genética entre elas. Nesse caso, sugerimos que se os trabalhos de reconstrução dessas duas famílias já estiverem adiantados, no sentido de que já existam formas “proto” reconstruídas, essas proto-formas sejam as escolhidas para o trabalho de comparação. Caso não seja essa a situação, recomendamos um certo cuidado na definição das línguas. Preferencialmente, sugerimos a escolha de línguas que tenham se separado de sua respectiva proto-língua mais recentemente. Para detectar isso, parece-nos conveniente utilizar um método também quantitativo para determinar o tempo de separação das línguas envolvidas. O método glotocronológico de Swadesh (1955), ao nosso ver, demonstra ser ideal nesse caso, já que esse método bem como aquele que aqui apresentamos são quantitativos e baseados em listas de vocábulos.

Devemos ressaltar que essa etapa do trabalho, às vezes, costuma ser difícil, tendo em vista que com o passar do tempo, os vocabulários das línguas envolvidas podem sofrer tantas mudanças que a probabilidade de se encontrar cognatos, caso as línguas sejam geneticamente aparentadas, fica abaixo daquela de se encontrar cognatos ao acaso. Isso, sem dúvida pode obscurecer ou até mesmo tornar impossível a determinação do parentesco por meio do método (Warnow, 1997).

Outro detalhe para o qual devemos estar atentos, diz respeito ao critério de uniformidade relativa das línguas envolvidas em termos cronológicos. Ou seja, as línguas,

¹ Embora possa ser também aplicável a línguas de uma mesma família.

pelo motivo exposto acima, devem ter se separado de suas respectivas proto-línguas em tempos equivalentes. Nesses termos, ou seja, devemos escolher línguas que sejam “mais representativas”, em termos de vocabulário.

Com relação à escolha do critério de equivalência dos cognatos, um problema que não consideraremos por ser muito “livre” e, portanto, subjetivo, é permitir a chamada *Tolerância Semântica* (“semantic shift”), isto é, a possibilidade de podermos considerar, por exemplo, ***badi** ‘lua’ do Proto-Tacana como cognato de ***bari** ‘sol’ do Proto-Pano. Com uma pequena modificação em nossos cálculos, entretanto, isso poderia ser feito, porém a nossa segurança em relação ao resultado ficaria abalada, pois com isso, poderíamos encontrar um número de formas cognatas maior, o que é interessante para métodos de reconstrução, mas não naqueles do tipo de comparação quantitativa. De acordo com o que afirmamos, então, ao necessitarmos de uma tabela de equivalência de fonemas, por exemplo, admitiremos uma forma como ***eya** ‘galho de árvore’ do Proto-Tacana como cognato de ***poya** ‘braço’ do Amahuaca (Pano) ou **pojami** ‘braço’ do Proto-Pano. Ou seja, caso as palavras pertençam a línguas diferentes e tenham o mesmo significado, nesse caso, então, admitiremos a *Tolerância Fonética* (“phonetic shift”) em nosso método.

2. O PROTO-PANO E O PROTO-TACANA: A APLICAÇÃO DO MÉTODO

Para fins deste artigo, utilizaremos as propostas de Shell (1975) para o Proto-Pano, e de Key (1968) para o Proto-Tacana. Justificamos nossa escolha pelas proto-formas por serem elas, a nosso ver, o que mais nos aproximaria do que estamos denominando vocabulário “médio” das línguas dos dois grupos em questão, ou seja, o vocabulário que encerra contribuições de todas as línguas das respectivas famílias. Aliás, com isso também podemos responder ao seguinte questionamento: sendo as línguas “proto” hipotéticas, isso não poderia afetar nossa hipótese de trabalho? Haja vista que de certa forma cada proto-lexema, pela sua construção, deve conter características “médias” de todas as línguas de suas respectivas famílias, certamente, suas características proto não trariam problemas à análise.

Quanto aos inventários fonológicos dessas línguas, segundo Shell (1975), no Proto-Pano encontramos 16 consoantes: *p, *t, *k^w, *k, *ʔ, *ts, *tʃ, *r, *m, *n, *s, *ʃ, *ʂ, *β, *y e *w; 4 vogais orais: *a, *i, *ĩ, *o; e 4 vogais nasais: *ã, *ĩ̃, *ĩ̃̃, *õ. Já os fonemas do Proto-Tacana, de acordo com Key (1968), são as consoantes: *p, *t, *k, *b, *d, *ts, *ć, *tʃ, *s, *ś, *ʃ, *x, *m, *n, *w, *r, *í, *y, *ʔ; e as vogais orais: *i, *o, *e, *a.²

Considerando que as vogais nasais não existem no Tacana, resolvemos fazer, ainda, uma análise da frequência desse tipo de vogais ou daquelas nasalizadas no Pano. O resultado é mostrado abaixo:

² Os símbolos ć, í, ś referem-se a segmentos sem valor fonético sendo somente representativos de conjuntos de reflexos.

| Vogais | a | i | ɨ | o |
|---------------|----------|----------|----------|----------|
| Oral | 37% | 29% | 18% | 12% |
| Nasal | 2% | 2% | 0,5% | ~0% |

Tabela 1: Freqüência das vogais nasais ou nasalizadas no Pano.

Como vemos os dados na Tabela 1 nos fazem concluir que as vogais nasais não desempenham um papel importante no léxico das línguas Pano. Elas só figuram realmente em marcações morfológicas nas chamadas formas longas das palavras. Devido a essas circunstâncias, então, nos parece normal que não haja no Proto-Tacana vogais nasais ou nasalizadas.

Para a obtenção de informações sobre a estrutura dos vocábulos em termos de número de sílabas possível, efetuamos uma exaustiva contagem de sílabas das palavras das duas famílias lingüísticas, com base em uma lista de 498 palavras do Proto-Pano apresentadas em Shell (1975) e, por outro lado, de uma lista, segundo Key (1968), de 204 palavras do Proto-Tacana. O resultado é dado na Tabela 2, seguinte:

| | Proto-Pano | Proto-Tacana |
|-----------------------|-------------------|---------------------|
| Monossílabos | 9% | 7% |
| Dissílabos | 72% | 54% |
| Trissílabos | 18% | 33% |
| Quadrissílabos | 1% | 6% |

Tabela 2: Distribuição do número de sílabas no Proto-Pano e no Proto-Tacana.

O quadro, acima, revela-nos que não existem em ambas as línguas palavras com mais de quatro sílabas, a não ser os casos de compostos, os quais não consideraremos neste estudo. Para os propósitos da aplicação do método que ora apresentamos, sugerimos a partir dos dados na Tabela I, uma proto-estrutura padrão em termos de quantidade máxima de sílabas para todos os proto-vocábulos dessas famílias, a saber: CVCVCVCV, a qual, será fundamental na análise dos dados aqui proposta. Por outro lado, o padrão silábico do Proto-Pano é (C)V(C), enquanto no Proto-Tacana a estrutura silábica padrão é (C)V. Ou seja, existem sílabas travadas no Proto-Pano, bem como registramos a ocorrência em ambas as proto-famílias de várias palavras iniciadas em vogais. De posse dessas informações, sugerimos, além dos inventários fonológicos expostos anteriormente, mais dois segmentos para cada uma das duas línguas: a consoante “vazia” \emptyset e a vogal “vazia” \emptyset^3 . Nosso intuito com isso é estabelecer uma certa regularidade das formas lingüísticas a serem comparadas. A introdução destes fonemas *virtuais* nas formas e a hipótese de *redução ao absurdo* constituem o núcleo do nosso método e o distingue dos procedimentos usuais apresentados em Rosenfelder (2002) e Ringe (1992). Tomemos um exemplo do Proto-Pano, a forma

³ Ressaltamos que o símbolo representado pela letra grega \emptyset não deve ser confundido com aquele apresentado no Alfabeto Fonético Internacional (IPA). Por convenção, utilizaremos esse símbolo em correspondência à idéia de vazio ou zero. Com isso, estaremos apenas uniformizando o processo de contagem do número de sílabas e de segmentos fonológicos na análise das palavras de duas línguas.

como ***risisti**, ‘fio’ que em princípio é composta por três sílabas, CV.CVC.CV e, como vemos, tem uma sílaba travada medial. Considerando a quantidade possível de sílabas no Proto-Pano e supondo que a proto-estrutura silábica a rigor seja CV e não CVC, então, deduziremos que a sílaba travada em ***risisti** seja o resultado de um processo de degeneração devido à queda do núcleo silábico da consoante em coda da sílaba anterior. Nesse sentido, quando estivermos computando o número de sílabas de ***risisti**, ao invés de três sílabas, registraremos quatro, pois no lugar onde esperaríamos um núcleo silábico, será introduzida a vogal vazia ϕ . Assim, a forma apresentada para contagem será ***risis ϕ ti** e a estrutura da palavra quanto ao número de sílabas será CVCVCVCV.

Essa regularização deverá ser aplicada na contagem de todas as palavras das listas. Por exemplo, se temos a palavra ***?awini** ‘esposa’ e postulamos que a proto-estrutura da palavra teria quatro sílabas, então, as consoantes e vogais vazias deverão complementar os espaços de uma possível sílaba perdida. Nesse caso, a forma ***?awini** poderia ser *** $\phi\phi$?awini** ou ***?awni $\phi\phi$** . Entretanto, optaremos sempre pela forma de “completar” ϕ à esquerda por uniformidade e, assim, eliminaremos a possibilidade de contagem dupla. Como poderemos observar, com isso, conseguiremos o intento de uniformizar as palavras das duas línguas levando-se em conta todos os padrões silábicos que elas apresentam. Notemos que ao proceder dessa maneira estamos tão somente criando lexemas virtuais para fins de contagem.⁴ Este procedimento é um dos pontos centrais do método e evita várias críticas tais como as que são apresentadas em Poser (2002), Rosenfelder (2002), entre outros. Ou seja, não enfrentaremos problemas ao contarmos os lexemas de ambas as línguas como tendo a forma apresentada neste trabalho.

Na comparação da língua Proto-Pano com a Proto-Tacana levaremos em conta, ainda, algumas equivalências que nos permitirão ter maior flexibilidade na determinação dos cognatos. A tabela abaixo mostra as equivalências que serão utilizadas:

| Proto-Pano | Proto-Tacana |
|----------------|--------------|
| p | p,b |
| t | t,d |
| k | k,x |
| k ^w | k,x |
| ʃ | ʃ, s,sʃ,ts |
| ʂ | ʃ, s,sʃ,ts |
| m | m,n,b |
| n | m,n,d |
| r | r, r̄ |
| s | ʃ, s,sʃ,ts |
| β | b,p |

⁴ Um processo de preenchimento de posições em segmentos por meio de fonemas vazios também é utilizado pela Fonologia de Governo (Carr, 1994). Entretanto, naquele modelo teórico o preenchimento se dá somente nos casos de posições nucleares e tem implicações fonológicas. Isso não ocorre em nosso método, pois, como já enfatizamos, a introdução dos fonemas vazios tem somente a função de facilitar a comparação das palavras de duas línguas ao considerá-las com um mesmo número de sílabas e padrão silábico.

| | |
|----|--------|
| w | w,u |
| tʃ | ć,tʃ,t |
| ts | ć,tʃ,t |
| y | y,i |
| a | a |
| i | i,e |
| u | o |
| ɨ | i,e |
| ã | a,e |
| ĩ | i,e,a |
| o | o |
| ĩ | a,i,e |

Tabela 3: Equivalência entre fonemas do Proto-Pano e do Proto-Tacana.

Considerando essas equivalências, fizemos uma contagem da frequência dos fonemas das duas línguas proto e usando somente lexemas com a forma máxima de sílabas CV, igual a CVCVCVCV nas duas línguas e, ainda, com o artifício de inserir consoantes e vogais vazias. Na Tabela 4, abaixo temos a estatística para o Proto-Pano.

| PANO | | | | | | | | |
|----------------|-----------------|-----------------------|---------------------------|----------------------|--------------------------|-----------------------|---------------------------|---------------|
| Fonema | Posição Inicial | Primeira Vogal medial | Primeira consoante medial | Segunda vogal medial | Segunda consoante medial | Terceira vogal medial | Terceira consoante medial | Posição final |
| p | 0.002 | | 0.020 | | 0.064 | | 0.062 | |
| t | 0.002 | | 0.014 | | 0.060 | | 0.084 | |
| k | | | 0.002 | | 0.068 | | 0.112 | |
| β | 0.006 | | 0.050 | | 0.064 | | 0.050 | |
| k ^w | | | | | 0.020 | | 0.006 | |
| m | 0.008 | | 0.002 | | 0.076 | | 0.060 | |
| n | 0.004 | | 0.018 | | 0.074 | | 0.177 | |
| r | 0.006 | | 0.014 | | 0.020 | | 0.042 | |
| s | | | 0.016 | | 0.038 | | 0.040 | |
| ʃ | | | 0.004 | | 0.034 | | 0.024 | |
| ʂ | 0.004 | | 0.018 | | 0.034 | | 0.076 | |
| w | 0.010 | | 0.006 | | 0.042 | | 0.042 | |
| y | 0.004 | | 0.010 | | 0.032 | | 0.018 | |
| ʔ | 0.008 | | 0.038 | | 0.102 | | 0.054 | |
| ts | | | | | 0.014 | | 0.026 | |
| tʃ | 0.002 | | | | 0.020 | | 0.028 | |
| a | | 0.022 | | 0.080 | | 0.337 | | 0.275 |
| i | | 0.020 | | 0.056 | | 0.165 | | 0.213 |
| ĩ | | 0.006 | | 0.042 | | 0.120 | | 0.139 |

| | | | | | | | | |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| o | | 0.010 | | 0.042 | | 0.197 | | 0.212 |
| ã | | 0.002 | | 0.008 | | 0.010 | | 0.026 |
| ĩ | | 0.002 | | 0.002 | | 0.018 | | 0.014 |
| ĩ | | | | | | 0.006 | | 0.004 |
| õ | | | | 0.004 | | 0.010 | | 0.006 |
| Fonema vazio | 0.944 | 0.938 | 0.750 | 0.766 | 0.336 | 0.137 | 0.137 | 0.111 |

Tabela 4: Frequência de fonemas em cada posição dos vocábulos do Proto-Pano.

Já na Tabela 5, a seguir, observamos as estatísticas para o Proto-Tacana:

| TACANA | | | | | | | | |
|---------------|-----------------|-----------------------|---------------------------|----------------------|--------------------------|-----------------------|---------------------------|---------------|
| Fonema | Posição Inicial | Primeira Vogal medial | Primeira consoante medial | Segunda vogal medial | Segunda consoante medial | Terceira vogal medial | Terceira consoante medial | Posição final |
| p | | | 0.019 | | 0.058 | | | |
| t | 0.005 | | 0.010 | | 0.077 | | 0.078 | |
| k | 0.005 | | 0.010 | | 0.053 | | 0.059 | |
| b | 0.005 | | 0.014 | | 0.016 | | 0.039 | |
| m | | | 0.019 | | 0.097 | | 0.054 | |
| n | | | 0.005 | | 0.048 | | 0.132 | |
| ř | | | | | | | 0.005 | |
| ř | | | | | 0.029 | | 0.078 | |
| r | | | 0.005 | | 0.039 | | 0.054 | |
| s | | | 0.014 | | 0.034 | | 0.034 | |
| ʃ | | | | | 0.010 | | 0.020 | |
| ś | | | 0.014 | | 0.025 | | 0.059 | |
| w | | | 0.010 | | 0.024 | | 0.059 | |
| y | | | 0.005 | | 0.019 | | 0.010 | |
| ts | | | 0.005 | | | | 0.049 | |
| ć | | | | | 0.034 | | 0.034 | |
| tʃ | | | | | 0.010 | | 0.015 | |
| x | | | 0.010 | | 0.054 | | 0.059 | |
| a | | 0.005 | | 0.082 | | 0.382 | | 0.366 |
| e | | 0.024 | | 0.135 | | 0.210 | | 0.161 |
| i | | | | 0.053 | | 0.196 | | 0.308 |
| o | | | | 0.043 | | 0.161 | | 0.165 |
| Fonema vazio | 0.985 | 0.971 | 0.855 | 0.687 | 0.339 | 0.051 | 0.084 | 0.000 |

Tabela 5: Frequência de fonemas em cada posição dos vocábulos do Proto-Tacana.

Utilizando-nos agora dos dados dessas duas tabelas e daqueles expressos na Tabela 3 de equivalência de fonemas, podemos calcular as chances de coincidências entre as duas línguas de uma maneira simples e direta. Primeiro, precisamos calcular as probabilidades de coincidência de todos os fonemas em todas as posições e, em seguida, somá-las. Não vamos apresentar todos os cálculos aqui, pois eles, embora muito demorados, são simples. Para ilustrar como isso é feito, vamos calcular a probabilidade de se ter uma coincidência entre um fonema do Proto-Pano, por exemplo, /β/, e outro que figura na tabela de equivalência na posição de segunda consoante medial. A frequência de /β/ ocorrer é de 0.050. Pela tabela de equivalência /β/ pode ser substituído por /b/ e /p/ do Proto-Tacana. Na tabela de frequência dos fonemas do Tacana, encontramos para esses fonemas respectivamente as frequências 0.014 e 0.019. Assim, a probabilidade de termos uma coincidência na ocorrência da segunda consoante das palavras dessas duas línguas é:

$$0.050 * (0.014 + 0.019) = 0.00165 (0,165\%).$$

Na expressão, acima, obtivemos o valor 0,165% como o produto da frequência de se encontrar o fonema /β/ em uma determinada posição (naquela ocupada pela segunda consoante) em um lexema Pano, multiplicado pela soma das frequências de se encontrar /p/ (0,014) e /b/ (0,019) na mesma posição num lexema Tacana. Calculando-se dessa maneira as probabilidades de coincidência de ocorrência de todos os fonemas em suas respectivas posições e somando-se essas probabilidades obteremos facilmente, os resultados mostrados na Tabela 6, abaixo:

| | |
|---------------------------|-------|
| Posição inicial | 0,930 |
| Primeira vogal medial | 0,911 |
| Primeira consoante medial | 0,646 |
| Segunda vogal medial | 0,554 |
| Segunda consoante medial | 0,179 |
| Terceira vogal medial | 0,298 |
| Terceira vogal medial | 0,125 |
| Posição final | 0,291 |

Tabela 6: Probabilidades de ocorrências de coincidências em todas as posições nos vocábulos.

Com base nesses dados, concluímos que a chance de obtermos uma coincidência de lexemas na forma CVCVCVCV nas línguas Proto-Pano e Proto-Tacana é o produto das probabilidades acima, ou seja, 0,000588323 ou 0,06% o que equivale a 1 chance em 1700. Utilizando a lista das palavras do Proto-Tacana encontradas em Key (1968) em número de 204 e utilizando uma distribuição binomial, podemos calcular as probabilidades de se encontrar r=1,2,... cognatos na lista de n=204 lexemas com a probabilidade constante p=

0,000588323 de se encontrar uma coincidência, o que é dado pela fórmula: $P(r) = C_r^n \cdot p^r \cdot (1-p)^{n-r}$, onde C_r^n é o coeficiente binomial. De posse dos valores numéricos, não será difícil encontrarmos os seguintes resultados:

$$P(0) = 0,886873 \text{ (88,7\%)}$$

$$P(1) = 0,106503 \text{ (10,7\%)}$$

$$P(2) = 0,006364 \text{ (0,64\%)}$$

$$P(3) = 0,000252 \text{ (0,03\%)}$$

$$P(4) = 0,000007 \sim (0\%)$$

A fim de calcularmos as probabilidades de encontrarmos numa lista de n vocábulos, mais de r cognatos, poderemos usar a seguinte relação:

$$P(r \geq k+1) = 1 - P(r < k) - P(r = k) \quad k=0,1,2\dots$$

Aplicando então a fórmula acima, poderemos calcular os dados da seguinte tabela:

| r | Extamente r coincidências | r ou mais coincidências |
|----------|----------------------------------|--------------------------------|
| 0 | 0,886873 | |
| 1 | 0,106503 | 0,113127 |
| 2 | 0,006364 | 0,006624 |
| 3 | 0,000252 | 0,000260 |
| 4 | 0,000007 | 0,000008 |

Tabela 7: Probabilidade de se encontrar r ou mais coincidências na lista de 204 vocábulos.

Na lista de 204 palavras do Proto-Tacana, encontramos os seguintes cognatos com o Proto-Pano:

| Proto-Tacana | Proto-Pano | Glossa |
|---------------------|---------------------|---------------|
| *etaka | *tak ^w a | fígado |
| *ebiti | *βitʃi | pele |
| *eyana | *ana | língua |
| *ena | *ini | rio |
| *baka | *βaki | menino |
| *ami | *imi | sangue |
| *ti | *tʃiʔi | fogo |
| *ewane | *ʔawini | esposa |
| *tsoke | *ʃ oki | tucano |

| | | |
|--------|-------|---------|
| *a | *ʔa | fazer |
| *ma | *ma | não |
| *maxe | *maki | piranha |
| *etsao | *ʃao | osso |

Tabela 8: Cognatos Proto-Tacana e Proto-Pano.

Como encontramos um número de cognatos superior ao determinado pela probabilidade de ocorrência de coincidências por chance, conforme estabelecido anteriormente, somos conduzidos a reconhecer que a hipótese inicial de que as duas famílias através de seus representantes não estavam relacionadas é falsa. Portanto, por *redução ao absurdo* devemos concluir que há uma relação genética entre o grupo Pano e o grupo Tacana. Notemos que os cognatos que apresentamos são todos “básicos”, pois são nomes de parentesco, partes do corpo humano, animais, entre outros, ou seja, de vocábulos que de uma maneira geral não são empréstimos de outras línguas. Isso nos dá uma segurança adicional em nossa análise.

3. CONCLUSÃO

Neste estudo, buscamos, através de algumas hipóteses simplificadoras de trabalho, apresentar cálculos das chances de coincidência de vocábulos, ou seja, a probabilidade de encontrarmos cognatos em duas línguas que supostamente não eram relacionadas lingüisticamente. Nessa hipótese de trabalho, incluímos vários aspectos importantes que sanaram falhas que foram exaustivamente criticadas em estudos anteriores, como, por exemplo, não utilizar as propriedades fonêmicas e fonotáticas das línguas analisadas e admitir *tolerância semântica* excessiva. (McMahon, 2002; Ringe, 1992). Procuramos ainda tornar, na aplicação do método com propriedade, as condições para se encontrar cognatos as mais restritivas possíveis, ou seja, o único tipo de tolerância que admitimos foi aquele que considera como equivalentes fonemas que possuam pontos ou modos de articulação muito próximos ou aqueles que a literatura da lingüística comparativa considera como prováveis e ainda aqueles que são reflexos de uma língua na outra. Com isso, garantimos o rigor na exigência da contagem de cognatos, o que é muito importante para línguas que supostamente estejam separadas com um “time depth” muito grande e que, portanto, oferecem poucas chances de encontrarmos “cognatos perfeitos”, ou seja, aqueles que não têm desvio algum, quer seja fonêmico quer semântico.

Acreditamos que nosso método é de fácil aplicação, embora, reconheçamos que, às vezes, muito trabalhoso e, ao contrário do que pode parecer à primeira vista não exige conhecimentos de matemática acima das noções básicas da estatística tradicional clássica. Acreditamos que embora os princípios operacionais do método sejam conhecidos há muito tempo, pois se trata de aplicações matemáticas simples dentro da léxico-estatística, a hipótese *de redução ao absurdo* seja nova e introduz uma facilidade insuperável na comparação de línguas: a existência de listas muito exaustivas de vocábulos para a reconstrução no que agora, acreditamos, não é necessária para a determinação do parentesco de duas línguas quaisquer.

Enfim, esse é o “espírito” de nosso método quantitativo de comparação de línguas. Acreditamos que sua principal virtude é a filosofia de trabalho por meio da qual não será mais preciso especular qual a porcentagem de cognatos que fazem com que duas línguas sejam consideradas geneticamente relacionadas. Com isso não precisaremos recorrer a listas pré-elaboradas e sim a qualquer lista com as palavras escolhidas ao acaso e não necessariamente extensas, a não ser que nosso objetivo seja fazer a reconstrução do proto-idioma.

Nosso próximo passo será aplicar o método à grande maioria das famílias de línguas da América e assim tratar de verificar, via método quantitativo aqui apresentado, o parentesco destas línguas e ainda tentar resolver os problemas controversos levantados por Greenberg (1990), quando da sua classificação das línguas da América.

REFERÊNCIAS BIBLIOGRÁFICAS

- CARR, P. (1994). *Phonology*. London: The MacMillan Press LTD.
- CHAVARRIA M.M.C. (1973). *Esbozo fonológico del Ese-éxa o “Huarayo” (Tacana)* Tesis de Bachiller en Lingüística. Lima: Universidad Nacional Mayor de San Marcos.
- CRÉQUI-MONTFORT, G. de & RIVET P. (1921). La famille linguistique Takana. *Journal de la Société des Americanistes*, Part I, vol.13: 281-301
- d’ANS, A.; CHAVARRIA M.C.; GUILLÉN F.N. & SOLIS F.G. (1973). *Problemas de Clasificación de lenguas No-Andinas en el Sur-este Peruano*. Lima: Universidad Nacional Mayor de San Marcos.
- GIRARD, V. (1970). Resenha de KEY (1968). *International Journal of American Linguistics* 35(1):73-78.
- GREENBERG, J.H. (1956). The General classification of Central and South American Languages. In Anthony Wallace (ed.). *Men and cultures: Selected Papers of the Fifth International Congress of Anthropological and Ethnological Sciences*, p. 791-94. Philadelphia: university of Pennsylvania Press.
- _____. (1990). The American Indian Language Controversy. *The Review of Archaeology*. Reimpresso in: *California Linguistic Newsletter*, 22 (3): 16-18
- GROETEKEN, A.P. (1907). Bishop Armentia O.F.M. und die Erforschung des Rio Madre de Dios. *Anthropos* 2: 730-734.
- KEY, M.R. (1968). *Comparative Tacanan Phonology with Caviñena Phonology and notes on Pano-Tacanan Relationship*. Janua Linguarum. Series Práctica, 50. La Haya-Paris: Mouton.
- LANES E.J. (2000). *Mudança Fonológica em línguas da Família Pano*. Dissertação de Mestrado em Lingüística. Rio de Janeiro: UFRJ.
- LOOS, E.E. (1973). *Estudios Pano II*. Lima: Instituto Lingüístico de Verano.
- MASON, J.A. (1963). The Languages of South American Indians. In: Julian H. Steward (ed.). *Handbook of South American Indians*. Vol. 6, p. 157-317. New York: Cooper Square Publishers, Inc.
- McMAHON, A. (2002)-. Snap! What count as a match. In *Second Conference on the Archeology and Linguistics of Australia*. Canberra.

- POSER J.P. (2002). Sir Tomas Young and Statistical evidence of Historical Relationship. *Historiographia Linguistica* 29(1/2): 262-268.
- RINGE, D.A. (1992). On calculating the factor of chance. *Transactions of the American Philosophical Society* 82, part 1:1-110.
- RIVET, P. & LOUKOTKA, È. (1952). Langues de l'Amérique du Sud et des Antilles. In: *Les Langues du Monde*, Nouvelle édition, p. 1099-1169. Paris: Champion.
- ROSENFELDER, M. (2002). *How likely are chances resemblances between languages?* Disponível em <<http://www.zompist.com/chance.htm>>
- SCHULLER, R. (1933). The Language of Tacana Indians (Bolivia). *Anthropos*: 28: 99-116, 463-484.
- SHELL, O.A. (1975). *Estudios Pano III: Las Lenguas Pano y su reconstrucción*. Lima: Instituto Lingüístico de Verano.
- SUAREZ, J.A. (1973). Macro-Pano-Tacanan. *International Journal of American Linguistics* 39:137-154.
- SWADESH, M. (1959). Mapas de Clasificación lingüística de México y las Américas. *Cuadernos del Instituto de Historia. Serie Antropológica* 8. México: UNAM.
- VILLAREJO, R.P.A. (1959). *La Selva y el hombre. Estudio antropocosmológico del aborigen amazónico*. Lima: Editorial Ausonia.
- WARNOW, T. (1997). Mathematical Approaches to comparative linguistics. USA: *Proceedings of the Nautical Academy of Sciences* 9: 6585-6590.
- WISTRAND-ROBINSON, L. (1989). Uto-Aztekan affinities with Panoan of Peru I: Correspondences. In Mary R. Key (ed.) *Studies in South American Indian Languages*, p. 243-276. Philadelphia: University of Pennsylvania Press.

Recebido: 30/01/2003

Aceito: 17/12/2003