

As Línguas Indígenas e a Pré-História*

Denny Moore¹ e Luciana Storto²

¹*Museu Paraense Emílio Goeldi, Belém do Pará, PA, Brasil*

²*Museu Nacional, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil*

É possível, através da Lingüística, formular hipóteses sobre o passado, com base em dados de línguas atuais ou de línguas conhecidas através de estudos históricos. As hipóteses são elaboradas, primeiramente, sobre a pré-história das línguas, mas pode-se, também, chegar a conhecer como viviam os falantes das línguas do passado: sua cultura, adaptação ecológica, meio ambiente, localização e relações com outros grupos. A Lingüística pode ser entendida como uma subárea da Antropologia, como queria Sapir (1949), tendo muito a ganhar, ao cooperar com disciplinas afins como a Arqueologia, a Etnologia e a Antropologia Física, bem como com a História, a Etnobotânica e a Genética.

Não há, no entanto, necessariamente, uma correlação direta entre cultura, língua e genes. Por um lado, existem grupos humanos que são uma unidade cultural, uma unidade lingüística e uma população biológica. Por exemplo, na Austrália, em um estudo (White, 1997) com 39 'tribos' (unidades lingüísticas e culturais, sem organização política), 88% das tribos maiores (>população 500) tinham uma taxa de endogamia (casamento dentro do grupo) acima de 90%. Por outro lado, existem grupos nos quais as unidades culturais, lingüísticas e biológicas não coincidem. Naquele mesmo estudo na Austrália, as tribos menores tinham menos endogamia e membros das tribos do litoral central de Arnhem land (região no norte da Austrália), freqüentemente, casavam com pessoas de outras tribos.

Já na Amazônia Noroeste, temos a exogamia lingüística em um ambiente multilingüe (Sorensen, 1967). Ali, freqüentemente, uma pessoa pode ter avós que falam quatro línguas diferentes, que são mantidas distintas, sem mistura. Um exemplo extremo de independência de língua e cultura, por um lado, e de população biológica por outro, é o grupo dos kaduveu/mbayá. Nesta sociedade, segundo Öberg (1949:

*Agradecemos a Sérgio Meira pelos comentários úteis oferecidos.

55), o infanticídio era tão comum que famílias raramente tinham mais de uma criança, e a população era mantida pela captura de crianças de outros grupos. Em certas circunstâncias, uma língua pode ser gerada a partir de duas línguas, culturas e patrimônios genéticos, como acontece com várias línguas crioulas, nascidas em situações de contato colonial. Face aos exemplos mencionados, fica óbvio que a ligação entre a pré-história de uma língua e a pré-história dos falantes, em termos culturais, políticos ou biológicos, não deve ser pressuposta, mas é, a princípio, uma questão a ser investigada.

Há mais de 200 anos, lingüistas vêm estudando como as línguas mudam com o tempo e como recuperar informações sobre o seu passado. Neste artigo, explicaremos, brevemente, as idéias básicas da Lingüística Diacrônica (que estuda as mudanças das línguas com o tempo e procura reconstruir suas fases anteriores), bem como alguns métodos utilizados para relacionar os fatos lingüísticos com a pré-história dos falantes, indicando as limitações dos métodos e suas aplicações às línguas indígenas da América do Sul.

Mecanismos de Mudança Lingüística

Um dos mecanismos principais de mudança lingüística é a *mudança fonética*. Por exemplo, no português falado na maioria das regiões do Brasil, o som [l] (colchetes indicam transcrição fonética) em fim de sílaba tornou-se [w] (por exemplo, na palavra *hotel*), mas continua [l] em outras posições como, por exemplo, quando inicia uma sílaba, como em *hoteleiro*. Quando uma mudança fonética é completa em uma língua, todos os exemplos, no ambiente especificado, são mudados: todos os exemplos de [l] em final de sílaba, no português falado na maioria das regiões do Brasil (preservados ainda na escrita do português e na pronúncia de línguas aparentadas, como o espanhol), transformaram-se em [w], não somente alguns¹. As mudanças fonéticas seguem tendências universais, motivadas por princípios físicos ou fisiológicos. Assim, sons se transformam em sons parecidos em termos de articulação ou acústica, e diferem uns dos outros em poucos traços. Por exemplo, [l] e [w] são, na terminologia fonética, aproximantes sonoras, e esse tipo de mudança é natural e comum.

Mudanças fonéticas em línguas indígenas funcionam do mesmo modo. Por exemplo, algumas das ocorrências de [b] na língua gavião de Rondônia são o resultado da desnasalização de um *[m] que existia antigamente e ainda existe na língua aparentada, a suruí de Rondônia². Os dois sons, [m] e [b], são parecidos, pois ambos são consoantes bilabiais sonoras. A *regularidade* das mudanças fonéticas,

¹Existem, em certas condições, mudanças fonéticas irregulares, mas não nos ocuparemos de tais complexidades neste trabalho.

²Usa-se * para marcar formas não atestadas ou reconstruídas.

como veremos abaixo, nos dá a possibilidade de recuperar o som original, a partir do estudo dos resultados das mudanças.

Além do som, todos os outros aspectos de uma língua podem mudar com o tempo, inclusive a semântica das palavras. Por exemplo, uma só palavra antiga deu origem às formas abaixo, em três línguas descendentes:

- i 'água' em várias línguas tupi-guarani
- ii 'rio, chicha', na língua gavião de Rondônia
- e 'chuva', na língua karitiana

Outro mecanismo de mudança lingüística é a *criação analógica*, usada quando falantes modificam palavras para regularizar paradigmas. Por exemplo, a maneira de indicar plural em inglês antigo, através de *ablaut* ou alternância vocálica, existe, hoje, em poucas palavras como *foot ~ feet*, *goose ~ geese*. Outras palavras, em que o *ablaut* ocorria, regularizaram-se, como *book ~ books* ao invés de *book ~ *beek*. Na língua gavião de Rondônia, a vogal inicial de radicais está sendo reanalisada como um prefixo e regularizada como *a-* pelos falantes jovens, que tendem a falar *ákinì* 'ver' ao invés de *íkinì*.

Línguas também mudam por simples *perda*. Os marcadores de caso nominal do latim (nominativo, acusativo, etc.) não existem mais em português, a não ser na diferença de forma entre alguns elementos pronominais, como *eu* (1a. pessoa do caso reto) versus *me* (1a. pessoa do caso oblíquo). Pode-se levantar a hipótese de que a ordem de palavras, SVO (sujeito, verbo, objeto), do português, que é mais rígida do que a do latim, foi uma consequência da perda do sistema de marcação de caso, que era utilizado para identificar as relações gramaticais (sujeito, objeto, etc.).

Onde há contato com outras línguas, os *empréstimos* são uma fonte de possíveis mudanças. Contato esporádico resulta em algumas palavras emprestadas, freqüentemente adaptadas à pronúncia da língua que as recebe. Todavia, com muito contato e bilingüismo, o sistema de sons (Fonologia) ou o sistema de formação de palavras (Morfologia) podem ser modificados. Tantas palavras francesas com o som [•] (como em *janeiro*) entraram no inglês, que este som já faz parte do sistema fonológico inglês. No dialeto de nheengatu falado no Alto Rio Negro, como resultado de séculos de bilingüismo, palavras do português são livremente emprestadas e usadas com afixos de origem indígena, por exemplo, o vocábulo português 'caçar' pode ser marcado com um prefixo pessoal para derivar *a-kasai*, '(eu) caço', com o prefixo *a-* indicando primeira pessoa.

Aspectos Sóciopolíticos das Mudanças Lingüísticas

Cada geração aprende sua língua, modificando-a um pouco, mesmo nos casos de transmissão completa. Mudanças surgidas podem ser incorporadas ou

esquecidas. Se um subgrupo de falantes é separado do grupo, geograficamente ou por razões sóciopolíticas, as mudanças lingüísticas deixam de ser compartilhadas, criando divergências de fala entre os dois grupos. Assim, com a passagem do tempo, criam-se *dialetos* diferentes. Com mais tempo, a divergência é tanta que não há inteligibilidade entre as duas variantes lingüísticas, e pode-se falar, então, no aparecimento de duas *línguas* diferentes.

Às vezes, devido às condições políticas, uma língua pode ser parcialmente transmitida. Por exemplo, em situações coloniais, escravos africanos falantes de diferentes línguas, foram obrigados a se comunicar na língua de seus opressores europeus, sem oportunidade de aprendê-la perfeitamente. Uma língua simplificada, que não é a língua materna de ninguém, é chamada de *pidgin*. Quando o *pidgin* se torna a língua materna de uma geração, sua complexidade aumenta, e pode ser considerado *crioulo*, cujas propriedades podem ser bastante diferentes das propriedades das línguas envolvidas na sua criação. Pelas propriedades de uma língua, em comparação com outras línguas da sua família, um lingüista pode detectar casos que são o resultado de transmissão incompleta.

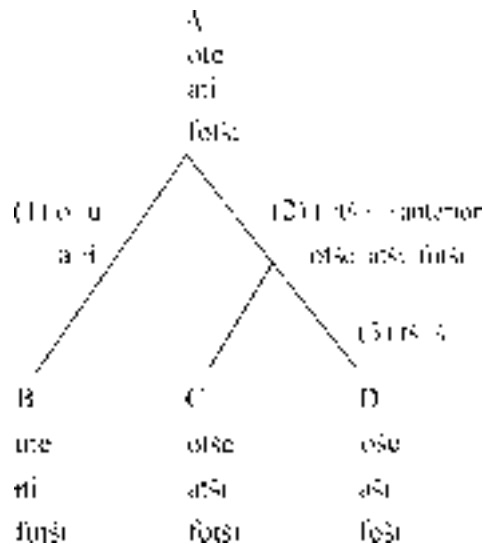
Um tipo de influência que uma língua pode sofrer durante sua transmissão é a dos *efeitos de substrato*, que são interferências da língua nativa dos aprendizes sobre a língua que está sendo aprendida. A língua geral amazônica, também conhecida como nheengatu, por exemplo, que foi aprendida por dezenas de milhares de índios que eram falantes nativos de outras línguas indígenas, foi alterada nesse processo, apesar de sempre ter tido uma comunidade de falantes nativos (quer dizer, nunca foi um *pidgin*). Similarmente, algumas línguas românicas tiveram, em sua origem, efeitos de substrato das línguas originais faladas nas regiões colonizadas pelos romanos. As mudanças de substrato são a versão diacrônica das influências que a língua nativa de um falante bilíngüe exerce na sua segunda língua.

Correspondências Fonéticas, Cognatos, Famílias e Troncos Lingüísticos

Vamos considerar um exemplo hipotético, bem simplificado, de mudança lingüística em um caso de transmissão completa. A Figura 1 representa, por meio de um diagrama arboriforme, as mudanças em três palavras de uma língua A durante sua transição para B, C e D. A comunidade de falantes de língua A se divide e uma parte acaba, após muito tempo, falando B, e a outra parte acaba falando C ou D. Uma das mudanças que ocorrem na transição de A para B é um aumento na altura das vogais *o* e *a*, que viram *u* e *i*, respectivamente. Esta mudança (1) transforma as palavras *ote*, *ati* e *fotši* em *ute*, *iti* e *futši*. Na outra parte da comunidade de fala, uma mudança (2) palataliza *t* em *tš* (como *tia*) antes de vogais anteriores, mudando *ote* e *ati* para *otše* e *atši*, sem afetar *fotši*. Subseqüentemente, esta parte da comunidade original se divide e uma outra mudança, de *tš* para *š* (como em *chuva*), muda *otše*,

atši e fotši em oše, aši e foši, respectivamente, em D.

Figura 1



Assim, no final, temos três conjuntos de *cognatos* (palavras descendentes da mesma palavra antiga, exibindo semelhança no sentido e na forma fonética). As palavras cognatas e as palavras antigas são mostradas na Tabela 1.

Tabela 1

A	B	C	D
ote	ute	otše	oše
ati	iti	atši	aši
fotši	futši	fotši	foši

Podemos notar que os cognatos exibem *correspondências de som*, por exemplo, *t* em B corresponde regularmente a *tš* em C e a *š* em D (*t*: *tš*: *š*). Por causa disso, mesmo que não existisse mais nenhum conhecimento sobre a língua A, seria possível levantar a hipótese da sua existência chamando-a de protolíngua. As palavras, na coluna A, seriam as protoformas. Se a protolíngua deixou muitas descendentes e as mudanças não foram tantas, é freqüentemente possível, através do *método comparativo*, *reconstruir* as protoformas a partir das correspondências de som nas formas cognatas das línguas filhas. Para ilustrar como se faz a reconstrução, vamos considerar duas correspondências nos cognatos, na Tabela 1.

○ som que deu origem à correspondência entre as consoantes no terceiro

conjunto de cognatos, $t\check{s} : t\check{s} : \check{s}$, pode ser, plausivelmente, reconstruído como $*t\check{s} = t\check{s} : t\check{s} : \check{s}$. Mas o que faria o lingüista se encontrasse os primeiros dois conjuntos que exibem a correspondência $t : t\check{s} : \check{s}$? Fiel ao princípio da regularidade da mudança fonética, que estipula que o mesmo som não deve ter duas correspondências no mesmo ambiente, o lingüista postula um outro proto-som para a outra correspondência: $*t = t : t\check{s} : \check{s}$, reconstrução que tem boa motivação fonética, uma vez que a mudança (2), na Figura 1 acima, $t > t\check{s}$, é natural e comum³, mas a reversa, $t\check{s} > t$, não é.

A reconstrução de protolínguas é bem mais complexa do que o exemplo simplificado dado, mas o exemplo é suficiente para introduzir os conceitos-chave e ilustrar, de modo geral, a lógica do processo. Utilizando o método comparativo, é possível, dentro de limites, reconstruir palavras em uma protolíngua som a som. É possível verificar a validade do método comparativo através de uma comparação entre os resultados e as formas escritas documentadas em línguas antigas, por exemplo, o latim.

Lingüistas dizem que as línguas A, B, C e D, na Figura 1, são relacionadas *geneticamente*, por descenderem de uma língua ancestral comum. A palavra *genético* é usada por lingüistas com um sentido diferente do que lhe dão os biólogos. Note-se, na Figura 1, que C e D têm uma relação genética entre si mais próxima do que com B, que é, geneticamente, mais distante. Essa *classificação interna* dos descendentes de A é representada pelo diagrama arboriforme. Relações genéticas entre línguas não devem ser confundidas com semelhanças *tipológicas*, que são somente características estruturais compartilhadas como, por exemplo, a ordem básica de palavras, a presença de vogais nasais, ou a complexidade morfológica. Semelhanças tipológicas não implicam em relações genéticas. Por exemplo, várias línguas aborígenes da Austrália têm sistemas de caso, mas esta propriedade tipológica não indica que são aparentadas geneticamente ao latim.

Línguas indígenas que descendem de uma língua mãe comum mostram cognatos e correspondências do tipo acima exemplificado hipoteticamente. Por exemplo, alguns cognatos entre quatro línguas da família tupari, do tronco tupi, estão na tabela abaixo. Apesar de os dados dessas línguas não mutuamente inteligíveis serem apenas preliminares, eles mostram uma correspondência $gw : \beta : kw : \beta$ que é bastante regular, bem como correspondências entre vogais (β é uma fricativa bilabial sonora, parecida com w).

Na Tabela 2, as formas na língua mãe das quatro línguas atuais, proto-tupari, foram reconstruídas, de forma preliminar, usando-se o método comparativo.

Tabela 2

Português	Proto-tupari	Ayuru	Makurap	Mekéns	Tupari
açaí	*gwit+i	gwiri	βirica	kwiri	βit+'i
machado	*gwi	—	βi	kwi	
distante	*gwetsok	gweeto	βetok	kwesop	

³A mudança é conhecida como palatalização, um tipo de assimilação de ponto de articulação, condicionada, no caso, pelas vogais que seguem a consoante t .

Segundo a terminologia usada no Brasil, as quatro línguas - ayuru, makurap, mekéns e tupari (com mais uma língua irmã, identificada somente em 1995) - constituem a *família lingüística* tupari, que deve ter uma profundidade temporal aproximadamente igual à da família lingüística latina, formada pelas línguas descendentes do latim: italiano, romeno, francês, português, etc. Nessa terminologia, *tronco lingüístico* é uma família de famílias lingüísticas. Por exemplo, a família tupari é uma das dez famílias (Rodrigues, 1984/85) do tronco lingüístico tupi; ou seja, a língua mãe de todas as línguas tupari, o proto-tupari, também tinha uma língua mãe, que é o proto-tupi. Comparando este exemplo com as línguas européias, temos a família lingüística latina (ou românica), que é uma das famílias do tronco lingüístico indo-europeu (juntamente com as famílias germânica, eslávica, índica, grega, celta, iraniana e báltica); a língua mãe das línguas românicas é o latim, que é uma das línguas filhas do proto-indo-europeu. Em outros países, usa-se o termo 'família lingüística' de modo mais geral, falando-se de 'famílias', 'subfamílias' ou 'ramos' de uma mesma família, etc.

Métodos Lingüísticos Ortodoxos de Fazer Inferências sobre o Passado

Distribuição de Agrupamentos Lingüísticos: Diversidade como Função da Antigüidade

Como vimos acima, a mudança lingüística produz diversidade com o passar do tempo. Dialectos da mesma língua têm uma profundidade temporal relativamente superficial; línguas irmãs, da mesma família lingüística, têm maior profundidade temporal, e línguas do mesmo tronco, mas de famílias diferentes, têm uma relação genética mais distante ainda. A *classificação interna* das descendentes indica a cronologia relativa da sua separação, como na Figura 1, na qual a distância genética entre B e C ou B e D é maior que a distância entre C e D.

Em geral, quando temos uma distribuição geográfica de dialectos ou de línguas geneticamente relacionadas, as regiões com pouca variação lingüística foram ocupadas mais recentemente que as regiões onde a diversidade é maior. Por exemplo, um lingüista do planeta Marte, após algumas pesquisas, poderia, facilmente, determinar que o ponto original da dispersão da língua portuguesa foi a Península Ibérica e não o Brasil, Macau ou Angola. As evidências são: 1) os dialectos mais divergentes da língua portuguesa se encontram em Portugal; 2) as línguas geneticamente mais próximas ao português (espanhol, italiano, outras línguas da família latina) se encontram na Europa, e 3) os dialectos mais diversos no Brasil encontram-se no litoral, fato consistente com a colonização oriunda da Europa.

O mesmo tipo de argumentação foi usado por Sapir (1949: 456-7) na sua hipótese de que as línguas atabascas originaram-se no norte de sua área de ocupação.

Havia três grupos geográficos das línguas atabascas: norte, oeste e sul. O grupo de línguas na região norte tinha subgrupos bem distintos, tão diferentes um do outro quanto entre eles e os grupos de línguas no oeste ou no sul. Os dois últimos grupos tinham pouca diversidade interna. Então, a região de maior profundidade temporal e de ocupação mais antiga deveria ser o norte. Além disso, o grupo de línguas atabascas constituía uma das três divisões do tronco na-dené, cujos outros membros, haida e tlingit, eram localizados ao norte.

No Brasil, inferências sobre origens de grupos de línguas estão sendo feitas na medida que o conhecimento das línguas indígenas e suas relações genéticas avança. Por exemplo, Rodrigues (1964) aponta o fato de que cinco das dez famílias lingüísticas do tronco tupi encontram-se em Rondônia: mondé, tupari, ramaráma, arikém e puruborá. É bem mais provável que as cinco famílias que estão hoje fora de Rondônia tenham saído daquela região do que as outras cinco tenham ido, independentemente, para lá (no caso de a terra original não ser Rondônia). Também, duas das famílias lingüísticas em Rondônia, mondé e tupari, têm uma profundidade temporal aproximadamente igual à da família latina, indicando que estão na região talvez há dois mil anos.

A hipótese de Rodrigues contradiz a posição de Lathrap e de Brochado (Lathrap, 1970; Brochado, 1984), que acreditam que o curso principal do Amazonas foi a terra de origem do tronco tupi, baseados na afirmação de Noble (1965) de que existe uma relação genética próxima entre os troncos tupi e aruák. O trabalho de Noble, porém, é duvidoso e Rodrigues (1985) apresenta indicações lingüísticas de uma relação genética entre os troncos tupi e karib, negando uma relação próxima entre aruák e tupi. Essas questões são fundamentais para um entendimento da pré-história da Amazônia.

A família karib, como foi observado por Rodrigues (1986), tem sua maior profundidade temporal na parte meridional da sua distribuição geográfica, onde se encontram as línguas kuikuru, txikão e bakairi, sugerindo que sua expansão possivelmente foi a partir do sul para o norte. Recentemente, Rodrigues (1999a) postulou, como hipótese, um tronco macro-jê composto de doze famílias: jê, kamakã, maxakali, krenak, puri, kariri, yaté, karajá, ofayé, boróro, guató e rikbaktsá. Destas, três são extintas: kamakã, puri e kariri. As relações postuladas são tão tênues e incertas que é difícil chegar a uma conclusão sobre a localização original dos falantes da proto-língua. Aikhenvald (1999) considera aberta a questão da região onde se falava o proto-aruaq.

Palavras e Coisas na Protolíngua e Protocultura

O vocabulário de uma língua é um inventário da cultura dos seus falantes. Assim, através do vocabulário de uma protolíngua, reconstruído pelo método comparativo, podemos ter uma “janela” na cultura dos falantes da protolíngua. Na Tabela 2, por exemplo, a existência de palavras para ‘açai’ e ‘machado’ no proto-tupari indica que os falantes habitavam uma região com palmeiras açai e que usavam machados. Uma parte do léxico de uma língua que é chave para um entendimento

da organização social dos seus falantes é a terminologia de parentesco. Sistemas terminológicos completos foram reconstruídos, por exemplo, no caso das línguas atabascas (Dyen & Aberle, 1974). Segundo Lehmann (1962: 208), é possível reconstruir palavras no proto-indo-europeu para ‘rebanho’, ‘vaca’, ‘cabra’, ‘porco’, ‘cachorro’, ‘cavalo’, ‘lobo’, ‘urso’, ‘ganso’, ‘pato’, ‘abelha’, ‘carvalho’, ‘faia’, ‘salgueiro’, e ‘grão’. Outros especialistas não concordam com algumas das palavras reconstruídas. Palavras para ‘arado’ e ‘jugo’, no proto-indo-europeu, indicam práticas de subsistência. A falta de palavra para ‘ferro’ ou itens lexicais ligados à sua produção sugere, para alguns estudiosos, que o proto-indo-europeu foi falado em um período anterior ao da utilização do ferro. Todavia, é sempre perigoso fazer inferências com base na ausência de palavras, pois estas podem, facilmente, ter se perdido com o tempo. Não há consenso sobre as correlações entre achados arqueológicos e culturas nas quais se falava o proto-indo-europeu. De fato, sem as evidências lingüísticas maciças que temos, ninguém teria imaginado a existência do tronco lingüístico indo-europeu ou a importância dos falantes das suas línguas.

No Brasil, a reconstrução do vocabulário de protolínguas está na “infância”, apesar da sua utilidade potencial em fornecer evidências para uma região em que pesquisas arqueológicas ainda são escassas. Todavia, existem algumas tentativas que são interessantes. Rodrigues (comunicação pessoal) achou evidências para a existência de palavras para ‘mandioca’, ‘batata doce’, ‘roça’, ‘pau para cavar na roça’ e ‘rede’ em proto-tupi, falado talvez entre 3.000 e 4.000 anos atrás.

Empréstimos

Quando algumas palavras em uma língua são parecidas, semântica e foneticamente, com palavras em uma segunda língua, há duas explicações possíveis, além do acaso. A primeira é que as duas línguas são relacionadas geneticamente, quer dizer, elas têm uma língua ancestral em comum, e as palavras parecidas são cognatas - retenções de palavras na língua ancestral comum. A segunda explicação é que uma (ou ambas) das línguas recebeu (receberam) as palavras em questão por empréstimo, o que indicaria um contato entre a língua doadora e a(s) língua(s) receptora(s). Determinar qual das duas possibilidades é a verdadeira, em um caso específico, é problema metodológico importante, que ocorre repetidamente em qualquer análise na Lingüística Diacrônica.

Um fato que ajuda a distinguir empréstimos de retenções é que empréstimos tendem a ocorrer em certos campos semânticos, refletindo a natureza do contato entre falantes. Por exemplo, após a Batalha de Hastings em 1066, a Inglaterra passou a ser governada pela França durante muito tempo, e naquele período, muitas palavras francesas entraram na língua inglesa nas esferas de governo (*government, parliament*), de negócios financeiros (*bank, finance, company*) e militares (*army, lieutenant, colonel*). Enquanto isto, o vocabulário básico (‘fogo’, ‘água’, partes do corpo, ‘dormir’, ‘correr’, etc.) continuou sendo da família germânica. Esta área do vocabulário tem uma forte tendência conservadora, como veremos abaixo, ao discutirmos a metodologia de Morris Swadesh.

Empréstimos também podem estar restritos a apenas algumas línguas da família, ao contrário das palavras cognatas. Por exemplo, os empréstimos de francês, mencionados acima, podem ser detectados pelo fato de que muitos ocorrem em formas parecidas nas línguas da família latina, mas raramente nas línguas da família germânica, das quais o inglês faz parte.

Empréstimos são, freqüentemente, marcados fonológica ou morfologicamente. Por exemplo, todas as palavras em inglês que têm o som [•] (*pleasure, garage*) são empréstimos, geralmente do francês, latim ou grego (Sapir, 1949: 448). Às vezes, é possível determinar a datação relativa desses empréstimos. Por exemplo, a língua inglesa adotou a palavra *banquet*, quando a consoante final *t* ainda era pronunciada na França, enquanto que a adoção da palavra *ballet* foi mais tardia, após a queda da consoante final, devido a uma mudança fonética na pronúncia francesa. Conseqüentemente, hoje em dia, em inglês, pronuncia-se um [t] final em *banquet*, mas não em *ballet*, apesar de a grafia ser a mesma em ambos os casos. Palavras emprestadas, antigamente, do português para a língua geral amazônica, por exemplo, *sorára* ‘soldado’ e *camishá* ‘camisa’ têm formas fonéticas mais divergentes do que as dos empréstimos atuais para o nheengatu, que são bem parecidos com a pronúncia portuguesa dos nossos dias.

Às vezes, há evidências fonológicas que indicam a direção do empréstimo. Comparando uma forma em português, *futebol*, com a forma correspondente em inglês, *football*, é evidente que inglês é a língua de origem da palavra (e do jogo que a palavra designa): uma vogal [e] teria sido inserida entre as duas consoantes oclusivas, se a palavra tivesse sido adotada pelo português (para adaptar a palavra ao padrão silábico português), mas não teria sido retirada, se a palavra tivesse sido adotada pelo inglês, cujo padrão silábico é mais amplo.

Esse exemplo ilustra um outro princípio que ajuda a determinar a direção de empréstimos: a língua na qual a palavra tem uma etimologia transparente é, provavelmente, a língua de origem. No caso, em inglês, a palavra *football* tem uma etimologia: *foot* ‘pé’ e *ball* ‘bola’, enquanto *futebol* não tem etimologia em português.

É mais difícil distinguir cognatos de empréstimos quando os empréstimos ocorrem entre línguas geneticamente relacionadas. Na evolução da língua francesa, o som [k] antes de [a], em latim, transformou-se regularmente em [š] em francês, mas continuou [k] em português:

caballus >	cheval, cavalo
capra >	chèvre, cabra

Porém, também houve empréstimos do latim - que continuou sendo uma língua associada a um alto nível de escolaridade - diretamente para o francês e o português, por exemplo, *canal, capital*, e, nestas palavras, o francês tem [k] ao invés de [š] (Fox, 1995: 66). Nesses casos, as palavras mais parecidas são provavelmente empréstimos, e as palavras mais diferentes (as com a correspondência francês [š] a português [k]) são provavelmente cognatas. Após serem detectados e estudados, os empréstimos podem indicar fatos importantes sobre o passado de sociedades sem

uma história escrita. No sul de Rondônia, a palavra para ‘milho’ é bem parecida em várias línguas da região, mesmo línguas de diferentes troncos lingüísticos: ayuru: *atiti*, makurap: *atiti*, mekéns: *asisi*, kanoé: *atiti*, djeoromitxi: *tšitši*, arikapú: *tšitši*, kwazá: *atšitši*. Provavelmente, o milho e a palavra para milho se espalharam na região em um período no passado. Com mais dados e estudo, deve ser possível traçar a rota do espalhamento.

Práticas magicoreligiosas podem ser emprestadas, junto com o vocabulário. O grupo gavião de Rondônia compartilha com o vizinho káro (arára de Rondônia) uma forma de xamanismo na qual espíritos aparecem no escuro, na frente de uma platéia. A palavra para o espírito é *olicicia* em gavião e *oricici* na língua do grupo káro, que ensinou o tipo de xamanismo ao grupo gavião.

Tipicamente, quando um povo entra em uma zona ecológica nova, ele aprende palavras para plantas e animais desconhecidos dos habitantes mais antigos na zona. Ver, por exemplo, os empréstimos das línguas tupi-guarani para português: *mutum*, *jaboti*, *tatu*, *açaí*, *muruci*, *mandioca*, etc. O mesmo fenômeno acontece com grupos indígenas. Dos empréstimos que Rodrigues (1985: 389-391) identifica como comum às famílias tupi-guarani e karib do Norte da Amazônia, várias são palavras para plantas e animais.

A língua karajá fornece evidência lingüística de um outro tipo sobre a localização original do grupo. Os karajás são divididos em quatro grupos, que falam dialetos diferentes. Três dos grupos moram no cerrado e o outro, os xambioá, moram em uma região mais amazônica. Segundo Eduardo Ribeiro (comunicação pessoal) o dialeto dos xambioá deriva novas palavras para novas plantas da Amazônia na base de palavras para espécies do cerrado, por exemplo, o termo para ‘açaí’, *kabirini*, é derivado do termo para ‘bacaba’, *kabiri*. Os quatro dialetos têm palavras parecidas para espécies do cerrado, e os mitos de origem, que são compartilhados pelos quatro grupos mencionam animais (seriema) e plantas (mangaba) típicas do cerrado. A inferência clara é que a localização original foi o cerrado e que os xambioá se deslocaram para Amazônia.

Nomes de lugares são, freqüentemente, adquiridos dos habitantes anteriores como, por exemplo, *Ipanema*, que quer dizer ‘lagoa fedorenta’ na língua tupi-guarani dos habitantes originais da região (Dietrich, 1999).

Métodos Lingüísticos Menos Ortodoxos de Fazer Inferências Sobre o Passado

Glotocronologia: Morris Swadesh

Morris Swadesh propôs o método chamado de glotocronologia ou léxico-estatística, em 1949 (Swadesh, 1951), baseando-se na pressuposição de que a parte básica do vocabulário das línguas é estável, mantendo-se na proporção de 80 a 85%

a cada 1000 anos. Da lista do “vocabulário básico”, excluem-se os artefatos e outros objetos culturalmente definidos e privilegiam-se aquelas palavras que são dificilmente emprestadas de outras línguas, como partes do corpo, termos de parentesco e elementos da natureza.

O cálculo léxico-estatístico permite estimar quando era falada a língua-mãe de duas línguas que são geneticamente relacionadas. A proporção de 80-85% foi calculada a partir da análise de textos de línguas modernas cuja relação genética estivesse comprovada, e cuja língua-mãe tivesse registros escritos, como é o caso das línguas românicas e do latim.

Na aplicação do seu método, Swadesh utilizava listas básicas, de 100 ou 200 palavras, que eram comparadas entre as línguas irmãs, cuja distância temporal deveria ser determinada, em busca de cognatos. Após determinar a porcentagem de cognatos descendentes de uma língua-mãe comum, seria possível calcular a antigüidade da língua-mãe, como se faz na datação com ^{14}C . No cálculo de manutenção do vocabulário básico de duas línguas, temos que considerar o fato de a retenção de palavras não significar que as mesmas palavras serão retidas em cada uma das duas línguas. Assim, o vocabulário comum entre as duas línguas, após um milênio, será menor que 80%, ou, mais precisamente, 64%.

O método da léxico-estatística foi muito criticado e, hoje, é considerado um método rápido, mas pouco confiável, para estimar a profundidade temporal de separação entre línguas irmãs. Um problema é a lista, que o próprio Swadesh revisou, várias vezes, para eliminar itens ligados à cultura. Um outro problema é o reconhecimento de cognatos, principalmente se a profundidade temporal é maior: dois lingüistas, trabalhando independentemente, nem sempre calculam a mesma porcentagem de cognatos compartilhados. O maior problema é a idéia de que a taxa de substituição lexical seja constante em todas as línguas e culturas do mundo. Testes, em várias regiões geográficas, mostraram resultados diferentes. Mesmo na Europa, a previsão do método da léxico-estatística pode estar seriamente errada: a estimativa para a separação da língua inglesa da língua alemã seria 954 anos, ou seja, por volta do século XI, enquanto a história indica que, de fato, a separação aconteceu a partir do século V d.C. (Fox, 1995: 286).

Comparação Multilateral: Joseph Greenberg

O método proposto por Greenberg, chamado de Comparação Multilateral, tem como objetivo produzir uma classificação genética, hipotética, de um grande número de línguas a partir da identificação de semelhanças entre um pequeno número de palavras comparadas em todas essas línguas. O próprio Greenberg esclarece que não se trata de um método que busca comprovar relações genéticas entre línguas, o que só pode ser feito através do método comparativo tradicional (Greenberg, 1987).

A versão da Comparação Multilateral que Greenberg aplicou ao analisar as línguas do continente americano, no seu controverso livro *Language in the Americas*

(1987), emprega pressupostos e instrumentos retirados da léxico-estatística, mais precisamente a taxa de retenção vocabular de 80-85% ao longo de um milênio, proposta por Swadesh, e a função de Joos (1964). Esta última é baseada na premissa de que o vocabulário básico tem uma distribuição estatística normal em termos da sua retenção e pode ser dividido em campos semânticos que apresentariam taxas de retenção variadas - algumas mais altas, outras mais baixas que a média de 80-85%. O resultado da aplicação do cálculo de Joos é que a taxa de perda vocabular decresce com o passar do tempo, pois algumas áreas vocabulares seriam bastante estáveis. A utilização que Greenberg faz da função de Joos é fundamental para justificar sua metodologia.

Todavia, a função de Joos é uma hipótese, não um fato comprovado. Vimos, na seção anterior, que a taxa de retenção média proposta por Swadesh já é duvidosa, e confirmar taxas de retenção diferentes de partes do vocabulário básico, em várias regiões do mundo, seria ainda bem mais difícil. Sem esse ajuste, após 10.000 anos, somente 1,1% (Greenberg, 1987: 343) dos cognatos do vocabulário básico seria retido em duas línguas irmãs - um item só, de uma lista de 100. Ele sustenta que a parte recuperável do vocabulário cognato, que é mantida em ao menos duas línguas, aumenta na medida que for maior o número de línguas comparadas. De fato, a comparação de muitas línguas aumenta bastante as chances de descobrir cognatos. Porém, depois de 10.000 anos, os números envolvidos são tão pequenos e as semelhanças entre cognatos tão tênues, devido às sucessivas mudanças fonéticas e semânticas, que sua identificação carece de significação estatística e não é replicável.

Os críticos de Greenberg argumentam que, ao aumentar o número de línguas, Greenberg corre o risco de aumentar, também, o número de possíveis coincidências acidentais. Ou seja, muitas das palavras que Greenberg considera comparáveis entre línguas podem, na verdade, não ser cognatos.

Ao propor sua classificação lingüística das Américas, que identificou três agrupamentos genéticos (ameríndio, na-dené e esquimó-aleuta), Greenberg foi atacado por Chafe (1987) e Goddard (1987) que criticaram severamente a qualidade dos dados utilizados por Greenberg, identificando inúmeros erros nas tabelas de *Language in the Americas* para línguas nas quais se especializam. Uma outra crítica séria à metodologia de Greenberg é a falta de rigor na identificação de cognatos. Ao contrário do método comparativo, a comparação multilateral não usa o critério de correspondências fonéticas para comprovar que duas palavras sejam cognatas. Além disso, o método aceita como cognatas palavras com grandes diferenças semânticas, como por exemplo, 'comer' e 'dente'.

A classificação lingüística das línguas nativas das Américas que Greenberg apresenta é, certamente, ousada. Ele afirma ter identificado três grandes agrupamentos genéticos, esquimó-aleuta, na-dené e ameríndio. Os primeiros dois agrupamentos já eram considerados prováveis, mas não o terceiro, ameríndio, que, segundo Greenberg, contém todas as outras línguas do hemisfério, agrupadas em somente onze subgrupos, dos quais sete estão localizados na América do Sul (*Macro-Gê, Macro-Panoan, Macro-Carib, Equatorial, Macro-Tucanoan, Andean e Chibchan-Paezan*). Segundo ele, o povo que falava a língua ancestral ao agrupamento ameríndio

entrou no Novo Mundo, talvez há 11.000-12.000 anos, antes dos povos que falavam línguas dos outros dois agrupamentos, que chegaram mais recentemente.

Como Greenberg chegou à estimativa de 11.000 a 12.000 anos de profundidade para o proto-ameríndio ao invés de, por exemplo, 20.000 ou 30.000 anos? Um método (Greenberg, 1987: 333) foi apelar para o registro arqueológico da época em que ele propôs sua classificação: ele identifica, assim, o proto-ameríndio com a cultura Clóvis, que teria entrado na América do Norte através do estreito de Bering há pelo menos 12.000 anos. O segundo método é puramente lingüístico (Greenberg, 1987: 335): ele aumenta a datação calculada por Swadesh (1958), que é de 7.000 anos (na base de uma lista de menos de 100 palavras), por razões no mínimo discutíveis, na proporção de 50%, para chegar a 10.000-11.000 anos. Claramente, o segundo método não comprova o primeiro, caracterizando-se, apenas, como um ajuste estatístico que tem como objetivo conformar os dados lingüísticos às datações arqueológicas, e, portanto, não se caracteriza como evidência independente para identificar a profundidade temporal da suposta protolíngua.

Os Nostraticistas

O nostrático é um macroagrupamento lingüístico, hipotetizado por um grupo de lingüistas, na sua maioria russos (por exemplo, Dolgopolskij, Illich-Svitych, Shevoroshkin), que inclui, na sua versão mais conservadora, os seguintes troncos: indo-europeu, cartveliano (georgiano e outras línguas relacionadas), urálico (línguas da região dos montes Urais) e altaico (composto de turco, mongol e tungus). É comum encontrarmos classificações de nostraticistas nas quais as famílias afro-asiáticas (em especial hamito-semítico) e dravídico (línguas do sul da Índia) estão incluídas. Às vezes, até mesmo o esquimó-aleuta, chukchi-camchatcano (noroeste da Ásia), níger-cordofaniano e sumério são considerados parte da macrofamília nostrática.

Os nostraticistas concebem seu trabalho como uma extensão do método comparativo, aplicado ao passado remoto. Eles usam protolínguas reconstruídas pertencentes a vários troncos lingüísticos e procuram reconstruir as relações genéticas a um nível mais profundo, em busca das línguas-mãe destas línguas. A opinião geral dos especialistas, nesses troncos lingüísticos, é que algumas das relações hipotetizadas pelos nostraticistas podem ser possíveis, mas que não há evidência suficiente para comprová-las, dado o caráter experimental e provisório das reconstruções originais, nas quais se baseiam as reconstruções mais profundas (até mesmo do proto-indo-europeu), e devido às mudanças semânticas duvidosas, exibidas nas palavras consideradas cognatas. As várias escolas de nostraticistas nem sempre concordam umas com as outras, e suas classificações são bastante diferentes da proposta por Greenberg. Por exemplo, o grupo eurasiático de Greenberg inclui japonês e coreano, que são, geralmente, excluídos do agrupamento nostrático, e exclui cartveliano, dravídico e hamito-semítico, que, como vimos, são freqüentemente considerados parte do nostrático.

Análise de Diversidade Tipológica e Genética: Johanna Nichols

A lingüista Johanna Nichols desenvolveu uma metodologia para investigar pré-história através da diversidade lingüística tipológica e genética das regiões do mundo, baseando-se na subdisciplina da Biologia conhecida como Genética de Populações. Sua preocupação primordial é com movimentos e expansões populacionais pré-históricas, sobre as quais ela afirma poder fazer inferências válidas, que vão além do que permite o método comparativo. Um tipo de diversidade com que ela trabalha é a diversidade genética (“genética” no sentido lingüístico do termo) de uma dada região: quantos troncos, famílias, ou grupos de descendência comum existem numa região. Um outro tipo de diversidade lingüística estudada por Nichols é a tipológica: quais parâmetros tipológicos, dentro dos conjuntos possíveis, estão presentes em uma região, e em que proporção. Os parâmetros tipológicos são selecionados após serem testados numa amostra global de línguas, a fim de se determinar quais deles são mais estáveis através dos anos. Para Nichols, por exemplo, o traço “marcação de núcleos” resiste muito a mudanças, enquanto a ordem relativa dos elementos da sentença (sujeito, verbo, objeto) pode ser, facilmente, influenciada por fatores regionais, ou seja, por características de outras línguas da região. Como os parâmetros tipológicos são gramaticais, ela não precisa lidar com as limitações impostas pelo desaparecimento inexorável das semelhanças entre os itens lexicais.

Equipada com essas formas de medir diversidade, Nichols investiga o que os vários padrões de diversidade têm a nos dizer sobre o passado. Dois conceitos são fundamentais para seu método: as zonas de espalhamento/dispersão (*spread zones*) e as zonas residuais (*residual zones*). As primeiras têm uma baixa diversidade tipológica e poucos troncos lingüísticos por unidade de área. São zonas como as estepes da Eurásia, onde grupos como o indo-europeu e o turco se expandiram, absorvendo ou expelindo outros grupos. Uma zona residual, em contraste, tem alta diversidade tipológica e um grande número de troncos lingüísticos acumulados. Um exemplo seria a região do Cáucaso. Para Nichols, o Velho Mundo tem uma diversidade tipológica menor e poucos troncos lingüísticos porque houve grandes expansões que reduziram o grau de diversidade naquela região do mundo. Na sua visão de pré-história remota (100.000 anos), sempre houve diversidade tipológica e um grande número de troncos lingüísticos. Tal visão se opõe, diametralmente, à dos nostraticistas, que pressupõem uma única linhagem genética que se desenvolveu ao longo de dezenas de milhares de anos. Nichols procura, às vezes, calibrar seu modelo com dados arqueológicos e com a ocorrência de eventos climáticos, ao contrário dos nostraticistas, que não distinguem agrupamentos que teriam uma profundidade temporal menor (e.g., entre semítico e indo-europeu, que são biológica e geograficamente próximos) de outros que exigiriam uma profundidade enorme (e.g., afro-asiático, indo-europeu e esquimó).

O que nos interessa, no entanto, é a visão que Nichols tem do Novo Mundo. Ela calcula que as migrações de novos troncos nas Américas, advindos da Ásia, teriam ocorrido com a frequência de um a cada “dois ou poucos milênios” (Nichols,

1990: 501), especialmente pelo fato de a diversidade de troncos lingüísticos ser baixa nas latitudes mais ao norte. Já que a introdução dos esquimó-aleutas produziu apenas um tronco, e a dos na-dené produziu apenas dois ou três, Nichols sustenta que a enorme diversidade lingüística do Novo Mundo (aproximadamente 140 troncos) não poderia ter surgido somente de uma língua ancestral, o Proto-Ameríndio, que, de acordo com a hipótese de Greenberg, era falado de 3.000 a 5.000 anos antes do proto-na-dené. Uma vez que as migrações de novos troncos lingüísticos para o Novo Mundo foram pouco freqüentes, Nichols conclui que a profundidade temporal da colonização teria sido “talvez 35.000 anos” (Nichols, 1990: 475). Este número é três vezes maior que a estimativa de Greenberg.

Com base em semelhanças tipológicas, Nichols chega a uma conclusão surpreendente: afirma que houve uma colonização a partir de regiões ao redor do Pacífico para o Novo Mundo há 35.000 anos (Nichols, 1992: 228), que está na origem das semelhanças tipológicas que ela afirma existir entre as Américas e a costa leste da Ásia. Essa argumentação é puramente lingüística, sem fundamentação arqueológica, e em conflito direto com as evidências biológicas. Ela comenta: “Neste movimento pelo Pacífico, as línguas viajaram mais longe e mais rapidamente que os traços biológicos: as afinidades lingüísticas do Novo Mundo chegaram tão longe quanto a Melanésia, enquanto as afinidades biológicas estenderam-se apenas ao norte da Ásia”. Frente a essa afirmação, fica difícil ver porque Nichols se preocupou em explicar o grau de diversidade lingüística do Novo Mundo na base de argumentos sobre migrações da Sibéria.

Discussão e Perspectivas

Neste artigo, apresentamos os métodos ortodoxos e não-ortodoxos que visam a investigar o passado através da lingüística. Alguns livros publicados, recentemente, também procuraram aplicar métodos lingüísticos e arqueológicos ao estudo da Pré-História de várias regiões (Blench & Spriggs, 1997; McConvell & Evans, 1997). Um outro exemplo de literatura interdisciplinar, que procura equacionar classificações genéticas com classificações lingüísticas, é o trabalho de Cavalli-Sforza. O assunto é complexo e extenso, mas vale a pena comentar, brevemente, seus objetivos e resultados, conforme foram apresentados em Cavalli-Sforza, 1991.

Cavalli-Sforza afirma: “Nossa terceira maior descoberta é que a distribuição de genes está correlacionada surpreendentemente bem com a distribuição de línguas. Concluímos que, em alguns casos, uma língua ou família de línguas pode servir para identificar uma população genética” (p. 108). Ele menciona o caso das línguas do grupo banto como um exemplo da família lingüística que fornece correlações biológicas bastante claras, permitindo vários *insights* sobre a Pré-História. Nenhum lingüista ficaria surpreso com tal afirmação, uma vez que a profundidade temporal daquela família lingüística é de aproximadamente 3.000 anos, e, portanto, está dentro dos limites dos métodos lingüísticos ortodoxos. O fato de que as línguas o grupo banto formam um agrupamento genético já era conhecido muito antes de Greenberg

classificar as línguas da África.

Já para os agrupamentos postulados com uma profundidade temporal bem maior, o caso é diferente. O autor (p. 109) afirma que estes “geraram resultados que também são consistentes com evidências genéticas completamente independentes”. Porém, um exame rápido dos quadros de agrupamentos lingüísticos, apresentados por Cavalli-Sforza, mostra as incongruências entre a classificação de Greenberg e aquelas de vários nostraticistas mencionados acima, isso sem mencionar os conflitos entre aquelas classificações e as conclusões de Nichols. Há, também, incongruências notáveis entre os agrupamentos biológicos e lingüísticos. Por exemplo, banto é bastante próximo, biologicamente, de nilótico, apesar de estes grupos serem totalmente independentes do ponto de vistas lingüístico, pertencendo a “superfamílias” (supertroncos, na terminologia que estamos utilizando) diferentes. A superfamília lingüística afro-asiática tem membros que são negros e outros que são caucasóides. Os subgrupos Indiano e Indiano Meridional são bastante próximos biologicamente, mas, de acordo com Greenberg, o subgrupo Indiano seria mais próximo de Esquimó do que do subgrupo Indiano Meridional de um ponto de vista lingüístico. A superfamília sino-tibetana tem um membro, o tibetano, que faz parte do agrupamento biológico do noroeste da Ásia, e um outro, chinês meridional, que já faz parte de um grupo biológico do sudoeste da Ásia, bastante distante do primeiro. Lingüisticamente, o subgrupo melanésio está agrupado com as línguas da Nova Guiné na superfamília indo-pacífica, mas, biologicamente, é um grupo mais próximo do subgrupo micronésio, da superfamília austronesiana.

As incongruências tendem a confirmar as incertezas da maior parte dos lingüistas no que diz respeito às classificações lingüísticas de grande profundidade temporal. É, no mínimo, duvidoso afirmar que as classificações lingüísticas e biológicas são completamente independentes, uma vez que a classificação biológica é diretamente influenciada por fatores geográficos, e os lingüistas tendem, consciente ou inconscientemente, a levar em conta fatores como geografia, teorias contemporâneas sobre o registro arqueológico e evidência biológica ao proporem hipóteses sobre relações genéticas distantes.

Portanto, não é possível afirmar que as classificações lingüísticas e as biológicas se confirmam. Existem outras questões não resolvidas, como os movimentos populacionais, utilizados como explicação para a distribuição de línguas e genes, cujas datas são freqüentemente discutíveis. Por exemplo, Cavalli-Sforza estima, com base na distribuição de genes, que houve uma primeira colonização do Novo Mundo há aproximadamente 30.000 anos, mais que o dobro da profundidade temporal que Greenberg postula. Há grandes debates na Arqueologia sobre o assunto. O mapa de migrações possíveis, apresentado por Cavalli-Sforza (p. 107), apresenta números referentes à “idade estimada do primeiro assentamento”, mas o autor afirma que o Velho Mundo Continental foi, provavelmente, colonizado por hominídeos há um milhão de anos. Há a possibilidade de que o último estágio da evolução humana tenha ocorrido em uma ampla frente geográfica, de forma que certas diferenças regionais atuais tenham resultado, pelo menos em parte, da continuidade de diferenças regionais antigas, e não apenas de processos como migrações e

bifurcações populacionais. Na verdade, de acordo com o quadro de distâncias genéticas (p. 108), o ramo africano tem menos diversidade interna do que o ramo asiático, apesar de os primeiros assentamentos, na África, terem ocorrido, segundo as datas dadas no artigo, antes que os da Ásia, o que deveria ter produzido maior diversidade com o passar do tempo. A pressuposição de alguns nostraticistas, de que *deve* ter havido uma língua, proto-sapiens, que deu origem a todas as línguas humanas atuais, não está, necessariamente, correta do ponto de vista da lógica. A situação pode ter sido mais parecida com a visão de Nichols. A profundidade temporal, postulada para a “Eva mitocondrial” de 150.000 a 200.000 anos (p. 106), é consistente com o desenvolvimento de uma diversidade lingüística enorme, além das possíveis mudanças lingüísticas que se seguiram, derivadas do contato com línguas diferentes de várias regiões, durante os movimentos populacionais.

Quais as implicações dessas questões para o estudo das línguas, culturas e populações indígenas do Brasil? Primeiramente, é importante concluir que não se deve dar credibilidade desmerecida aos métodos menos ortodoxos. Nem se deve supor que o problema das relações profundas entre troncos lingüísticos esteja resolvido. Por outro lado, métodos ortodoxos fornecem resultados muito mais ricos, e, se corretamente utilizados, podem ser aplicados, seguramente, a profundidades temporais de 5.000 a 6.000 anos. Os resultados interessantes obtidos para as famílias do banto e na-dené estão ao alcance desses métodos. Sem o auxílio da lingüística, teria sido muito difícil identificar e estudar estes agrupamentos. Podemos, sem sombra de dúvidas, combinar a Lingüística, a Arqueologia, e a Biologia Humana em busca de conclusões interessantes sobre o passado indígena no Brasil e no restante da América do Sul. As oportunidades para tanto são enormes, pois temos, aproximadamente, 90 troncos lingüísticos (definidos na maneira mais conservadora) na América do Sul, em contraste com 14 na África (Nichols, 1990: 479), e pouco trabalho interdisciplinar sobre eles.

Uma das razões pelas quais tal potencial não foi aproveitado é a falta generalizada de conhecimento que se tem, no Brasil, sobre a Lingüística Diacrônica e seus métodos; os programas de pós-graduação em Lingüística, no país, não costumam oferecer cursos nessa área. Além disso, a coleta de dados para serem usados em reconstrução de protolínguas é bem mais complicada do que, por exemplo, a coleta de amostras de sangue. O método comparativo exige uma ampla coleção de palavras que devem ser analisadas fonológica e morfológicamente, o que requer técnicas de trabalho de campo e capacidade analítica.

O estado atual da pesquisa diacrônica, na América do Sul, está resumido em Kaufman, 1990. Desde então, alguns dos estudos diacrônicos, que envolvem línguas indígenas brasileiras, incluem Dixon, 1999; Gildea, 1998; Jensen, 1999; Meira, 2000; Mello, 1992; Moore & Galucio, 1993; Noelli, 1998; Payne, 1991; Rodrigues, 1995, 1999a,b, 2000, e Rodrigues & Dietrich, 1997. Um guia útil sobre as línguas indígenas do Brasil é o de Rodrigues, 1986.

Um passo a ser dado para desenvolver o conhecimento da Pré-História indígena no Brasil seria iniciar um estudo interdisciplinar de um dos troncos principais do país, que combine dados lingüísticos, etnográficos, etno-históricos,

arqueológicos, e genéticos. Muitos desses dados já devem existir, e poderiam ser úteis, se organizados de acordo com os agrupamentos lingüísticos. O objetivo seria investigar o que pode ou não ser deduzido sobre o passado a partir dos dados, utilizando-se correlações ou a falta de correlações entre disciplinas, para entender os acontecimentos do passado e evitar cenários baseados em simplificações ou conjecturas. É importante que tal tipo de pesquisa se inicie o mais rápido possível, dado o perigo de extinção de várias línguas e culturas indígenas.

Referências Bibliográficas

- Aikhenvald A.Y. (1999). Arawak. In: Dixon R.M.W. & Aikhenvald A.Y. (Eds.), *The Amazonian Languages*. Cambridge University Press, Cambridge, pp. 65-106.
- Blench R. & Spriggs M. (Eds.) (1997). *Archaeology and Language I: Theoretical and Methodological Orientations*. Routledge, London & New York.
- Brochado J.P. (1984). An Ecological Model of the Spread of Pottery and Agriculture into Eastern South America. Tese doutoral, University of Illinois, Urbana-Champaign.
- Cavalli-Sforza L.L. (1991). Genes, Peoples and Languages. *Scientific American* 265: 72-78.
- Chafe W.L. (1987). Review of *Language in the Americas*, by J. Greenberg. *Current Anthropology* 28: 652-653.
- Dietrich W. (1999). A Importância do Tupi na Formação do Português do Brasil. In: Gartner E., Hundt C. & Schinberger A. (Eds.), *Estudos de História da Língua Portuguesa*. TFM, Frankfurt.
- Dixon R.M.W. (1999). Arawá. In: Dixon R.M.W. & Aikhenvald A.Y. (Eds.), *The Amazonian Languages*. Cambridge University Press, Cambridge, pp. 293-306.
- Dyen I. & Aberle D.F. (1974). *Lexical Reconstruction: The Case of the Proto-Athapaskan Kinship System*. Cambridge University Press, Cambridge.
- Fox A. (1995). *Linguistic Reconstruction; An Introduction to Theory and Method*. Oxford University Press, Oxford.
- Gildea S. (1998). *On Reconstructing Grammar: Comparative Cariban Morphosyntax*. Anthropological Linguistics Series. Oxford University Press, Oxford.
- Goddard I. (1987). Review of *Language in the Americas*, by J. Greenberg. *Current Anthropology* 28: 656-657.
- Greenberg J. (1987). *Language in the Americas*. Stanford University Press, Stanford.
- Jensen C. (1999). Tupi-Guarani. In: Dixon R.M.W. & Aikhenvald A.Y. (Eds.), *The Amazonian Languages*. Cambridge University Press, Cambridge, pp. 125-163.
- Joos M. (1964). Glottochronology with Retention-Rate Inhomogeneity. In: Lunt H.G. (Ed.), *Proceedings of the Ninth International Congress of Linguists*. Mouton, New York.
- Kaufman T. (1990). Language History in South America: What We Know and How to Know More. In: Payne D.L. (Ed.), *Amazonian Linguistics; Studies in Lowland South American Languages*. University of Texas Press, Austin, pp. 13-73.
- Lathrap D.W. (1970). *The Upper Amazon*. Praeger, New York.
- Lehmann W.P. (1962). *Historical Linguistics: an Introduction*. Holt, Rinehart, and Winston, New York.
- McConvell P. & Evans N. (1997). *Archaeology and Linguistics: Aboriginal Australia in Global Perspective*. Oxford University Press, Melbourne.
- Meira S. (2000). *A Reconstruction of Proto-Taranoan: Phonology and Morphology*. LINCOM Studies in Native American Linguistics. Vol. 30. LINCOM Europa, München.

- Mello A.A.S. (1992). Estudo Comparativo do Léxico da Família Lingüística Tupi-Guarani. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis.
- Moore D. & Galucio A.V. (1993). Reconstruction of Proto-Tupari Consonants and Vowels. In: *Survey of California and other Indian Languages, Report 8*, Proceedings of the Meeting of the Society for the Study of the Indigenous Languages of the Americas, July 2-4, 1993 and the Hokan Penutian Workshop, July 3, 1993, Columbus, Ohio, 119-138. Langdon M. (Ed.). Series ed. por Leanne Hinton. Department of Linguistics, Berkeley.
- Nichols J. (1990). Linguistic Diversity and the First Settlement of the New World. *Language* 66: 3.
- Nichols J. (1992). *Linguistic Diversity in Space and Time*. University of Chicago Press, Chicago.
- Noble G.K. (1965). *Proto-Arawakan and its Descendents*, Publication 38 of Indiana University Research Center in Anthropology, Folklore and Linguistics. (IJAL Vol. 31, No. 3, parte II.)
- Noelli F.S. (1998). The Tupi: Explaining origin and expansions in terms of Archeology and of Historical Linguistics. *Antiquity* 72: 648-663.
- Oberg K. (1949). *The Terena and Caduveu of Southern Mato Grosso*. Publication No. 9. Smithsonian Institution, Institute of Social Anthropology, Washington DC.
- Payne D.L. (1991). A Classification of Maipuran (Arawakan) Languages Based on Shared Lexical Retentions. In: Derbyshire D.C. & Pullum G.K. (Eds.), *Handbook of Amazonian Languages*. Vol. III. Mouton de Gruyter, Berlin, pp. 355-499.
- Rodrigues A. (1964). A Classificação do Tronco Lingüístico Tupi. *Revista de Antropologia* 12: 99-104.
- Rodrigues A. (1984/1985). Relações Internas na Família Lingüística Tupi-Guarani. *Revista de Antropologia* 27/28: 33-53.
- Rodrigues A. (1985). Evidence for Tupi-Carib Relationships. In: Klein H. & Stark L. (Eds.), *South American Indian Languages - Retrospect and Prospect*. University of Texas Press, Austin.
- Rodrigues A. (1986). *Línguas Brasileiras: Para o Conhecimento das Línguas Indígenas*. Edições Loyola, São Paulo.
- Rodrigues A. (1995). Glottalized Stops in Proto-Tupi. Comunicação apresentada no encontro de verão da Society for the Study of the Indigenous Languages of the Americas. University of New Mexico, Albuquerque.
- Rodrigues A. (1999a). Macro-Jê. In: Dixon R.M.W. & Aikhenvald A.Y. (Eds.), *The Amazonian Languages*. Cambridge University Press, Cambridge, pp. 165-206.
- Rodrigues A. (1999b). Tupi. In: Dixon R.M.W. & Aikhenvald A.Y. (Eds.), *The Amazonian Languages*. Cambridge University Press, Cambridge, pp. 105-124.
- Rodrigues A. (2000). Hipóteses sobre as Migrações dos Três Subconjuntos Meridionais da Família Tupi-Guarani. Comunicação apresentada no II Congresso Nacional da ABRALIN e XIV Instituto Lingüístico, Florianópolis.
- Rodrigues A. & Dietrich W. (1997). On the Linguistic Relationship between Mawé and Tupi-Guarani. *Diachronica* 14.2: 265-304.
- Sapir E. (1949). *Selected Writings of Edward Sapir in Language, Culture, and Personality*. In: Mandelbaum D.G. (Ed.). University of California Press, Berkeley e Los Angeles.
- Sorensen Jr. A.P. (1967). Multilingualism in the Northwest Amazon. *American Anthropologist* 69: 670-684.
- Swadesh M. (1951). Diffusional Cumulation and Archaic Residue as Historical Explanations. *Southwestern Journal of Anthropology* 7: 1-21.
- Swadesh M. (1958). Some New Glottochronological Dates for Amerindian Linguistics. *Word* 10: 306-332.
- White N. (1997). Genes, Languages and Landscapes. In: McConvell P. & Evans N. (Eds.), *Archeology and Linguistics; Aboriginal Australia in Global Perspective*. Oxford University Press, Oxford.