

Linguística: as línguas indígenas do Brasil

Aryon Dall'Igna Rodrigues

UNICAMP

1. Linguagem e língua. Linguagem é a capacidade inata que todo ser humano tem de adquirir e utilizar um sistema complexo de expressão e comunicação por meio de sons da fala; língua é cada um desses sistemas decorrentes da capacidade da linguagem. Nesse sentido, a linguagem é uma característica específica dos seres humanos, comum a todos eles e integrante da herança biológica da espécie; e não é, por conseguinte, um elemento cultural, que difira de uma sociedade para outra. Já as línguas, embora obedecendo a princípios gerais universais próprios da linguagem, diferem de sociedade para sociedade e, como os componentes das culturas, têm de ser aprendidas pelo indivíduo mediante interação com os demais membros da sociedade em que ele é criado.

A linguagem, como propriedade específica de toda a humanidade, que faculta a comunicação entre os indivíduos, pode ser considerada antecedente ou condicionadora da cultura, no sentido de que não é concebível o desenvolvimento cumulativo da cultura na ausência de um sistema linguístico de comunicação. Entretanto, é discutível se e em que medida a língua de uma sociedade particular condiciona a cultura dessa sociedade, e em que medida a cultura afeta a língua. É certo que o léxico, isto é, o vocabulário de uma língua reflete, enquanto nomenclatura, uma boa parte da cultura: as línguas têm nomes para os elementos das culturas a que estão normalmente vinculadas. Mas que aderência haverá das propriedades gramaticais de uma língua a propriedades culturais? Ou vice-versa, de propriedades culturais a propriedades gramaticais? A única generalização segura que se pode fazer a esse respeito é que toda língua é plenamente adequada às necessidades de expressão e comunicação da cultura a que está tradicionalmente associada. Nesse sentido, todas as línguas se equivalem, apesar de suas diferentes aparências.

2. Alteração e multiplicação das línguas. Todas as línguas se acham em constante, embora lenta e gradual, alteração. Os impulsos para essa alteração não se devem exclusivamente a fatores culturais, mas decorrem também de fatores inerentes seja a características fisiológicas e neurológicas dos indivíduos, seja à própria constituição interna desses complexos sistemas de comunicação. A permanente alteração das línguas faz com que aquilo que historicamente tem sido uma mesma língua, falada sem solução de continuidade por gerações sucessivas, seja hoje sensivelmente diferente do que foi tempos atrás. Essa alteração afeta não somente elementos do sistema lingüístico, mas também a estrutura desse sistema, por exemplo, as regras de associação e combinação dos elementos. Há, entretanto, um fator muito forte de coesão lingüística, que é o uso social da língua. Na medida em que cada indivíduo usa a língua como meio de interação com os demais indivíduos da comunidade a que pertence, ele constantemente participa sua experiência lingüística àqueles e coparticipa das experiências deles: inovações no seu uso da língua passam a integrar a experiência de seus interlocutores e podem propagar-se através destes aos demais membros da comunidade, e inovações em qualquer dos demais podem incorporar-se à experiência de cada indivíduo. Assim, embora alterando-se constantemente, a língua se mantém coesa, coerente e praticamente una.

Más a unidade de uma língua se rompe quando a sociedade que dela se serve divide-se em duas ou mais comunidades, de modo a interromper a cadeia de trocas permanentes de experiência. Nesse caso, as inovações surgidas numa comunidade não passam à outra comunidade, e vice-versa, e vêm a caracterizar diferenças entre a língua de uma e outra. Se a interrupção se mantém, daí por diante, no decurso do tempo, essas diferenças vão se acumulando na mesma proporção em que os elementos comuns vão diminuindo, e o que de início foi uma só língua passa a ser duas ou mais. Estas, primeiro tão semelhantes que permitem a intercompreensão quando eventualmente se encontram indivíduos das comunidades separadas, chegam a tornar-se, após séculos de separação, mutuamente ininteligíveis. Após alguns milênios de

separação e constante alteração divergente, a diferenciação pode ser tal, que a origem comum de duas ou mais línguas só se torna demonstrável mediante o emprego de métodos muito refinados de detecção de correspondências sistemáticas e de reconstrução de estágios pré-históricos dos sistemas linguísticos em questão.

3. Classificação genética das línguas. Línguas cuja origem comum pode ser demonstrada com recurso a métodos comparativos e reconstitutivos podem ser classificadas em grupos de diferentes ordens, segundo a maior ou menor proporção de elementos comuns que ainda conservam, ou seja, segundo a maior ou menor proximidade de sua comum origem. Línguas que têm tanto em comum, que são mutuamente inteligíveis, formam um grupo dialetal ou o que comumente se chama simplesmente de "língua": assim as línguas faladas majoritariamente nas cidades brasileiras e em boa parte de suas áreas rurais, bem como as que se falam em Portugal, constituem o grupo dialetal português ou a língua portuguesa (cerca de 500 anos de separação). Já as línguas entre as quais não há mais inteligibilidade mútua, mas cuja separação e diferenciação corresponde aproximadamente à de línguas como o português, o italiano, o francês (cerca de 2.000 anos de separação) constituem uma família linguística, neste caso a família linguística românica. Línguas ainda mais remotamente aparentadas, como o português, o alemão, o russo e o híndi (cerca de 5.000 anos de separação) formam um tronco linguístico, neste caso o tronco indo-europeu.

4. As línguas indígenas brasileiras. Falam-se no Brasil, ainda, de 120 a 150 línguas indígenas. Pode-se estimar que antes do descobrimento o número fosse bem maior, cerca do dobro do atual. A grande redução numérica é devida sobretudo à extinção dos povos indígenas decorrente da ocupação de seus territórios, durante os últimos 450 anos, pelos europeus e seus descendentes. As línguas sobreviventes situam-se quase todas a oeste de uma linha que se estende de São Luís do Maranhão até Chuí, passando por perto de Brasília: exceção constituem apenas as línguas iatê em Pernambuco, maxakalí em Minas Gerais, xoklêng em Santa Catarina e

alguns dialetos guarani de imigração recente na parte oriental dos estados de São Paulo, Rio de Janeiro e Espírito Santo.

Os trezentos anos do período colonial do Brasil legaram-nos documentos sobre apenas três línguas indígenas, uma das quais representada por três dialetos - o tupi antigo ou tupinambá dos séculos XVI e XVII, o guarani antigo do século XVII e a língua geral do Parará no século XVIII -, outra, a língua karirí ou kirirí, representada por dois dialetos - o kipeá e o dzubuká -, e mais a língua manau. Pode ter havido ainda algum documento de uma ou outra língua, mas que não chegou a nossos dias, como é certo que houve da língua maramonim.

É a partir da Abertura dos Portos, no início do século XIX, que começa a fazer-se uma documentação extensiva das línguas brasileiras. Essa documentação, devida sobretudo a naturalistas europeus (Martius, Spix, Pohl, Castelneau, Saint Hilaire, Wied-Neuwied, Natterer, Wallace, etc.), é feita quase exclusivamente na forma de amostras de vocabulário, mas atinge logo várias dezenas de línguas e vai estimular e permitir os primeiros ensaios de classificação linguística dos povos indígenas do Brasil, como o de von Martius, o de von den Steinen, o de Ehrenreich. Com von den Steinen, aliás, já no último quartel do século passado, inicia-se um esforço de documentação intensiva, caracterizada pela coleta de textos transcritos foneticamente e pela tentativa de análise da gramática. Von den Steinen procede assim com a língua bakairí das nascentes do Xingu, Capistrano de Abreu com o caxinauá do Acre, Koch-Grünberg com o taulipang de Roraima. É verdade que trabalho análogo já tinha sido feito pioneiramente, bem antes da vinda de von den Steinen, pelo General Couto de Magalhães, com seu excelente trabalho sobre a "língua tupi viva" ou nheengatu do baixo Amazonas. Nas cinco primeiras décadas deste século produzem-se documentos sobre muitas línguas, de diferentes naturezas e diversos níveis de qualidade técnica, por missionários e leigos, brasileiros e estrangeiros. Mas é só a partir do fim da década de 50 que passa a desenvolver-se o estudo científico sistemático de um número crescente de línguas. O fator decisivo foi sem dúvida a entrada em cena da organização linguístico-missionária Summer

Institute of Linguistics, interessada sobretudo na descrição técnica das línguas indígenas em todo o mundo. Hoje cerca de 40 línguas estão sendo descritas por pessoal daquela organização e umas 10 por pessoal independente, brasileiro ou estrangeiro, igualmente bem capacitado tecnicamente.

5. Classificação das línguas indígenas do Brasil. Com os dados acumulados até agora tem sido possível abordar em certa medida os problemas de classificação genética das línguas indígenas do Brasil. Reconhecem-se presentemente cerca de 25 famílias lingüísticas e umas vinte e tantas línguas não classificadas em famílias. Metade das famílias parecem incluir-se em dois troncos: sete no tronco tupi, seis no tronco macro-jê.

Das sete famílias que constituem o tronco tupi, quatro situam-se exclusivamente no Território de Rondônia, na área compreendida entre os rios Guaporé, Jiparanã e Pimenta Bueno; são elas as famílias arikém (línguas arikém, karitiana, kabixiana), tupari (línguas tupari, makurap, kepkiawat), ramarama (línguas ramarama, urumi, arara do Jiparanã), mondê (línguas mondê ou salamã, aruã, suruí do Jiparanã, cinta-larga). Na mesma região acha-se também a língua puruborã, não classificada ao nível de família.

No vale do Tapajós situa-se a família mundurukú (línguas mundurukú e kuruaya) e no vale do Xingu a família juruna (línguas juruna, xipaya e manitsawã). A sétima família, ao contrário das demais, não está restrita a uma região geográfica bem determinada, mas se encontra difundida amplamente, desde o litoral oriental até o extremo oeste do Brasil, de onde se estende ao Peru e à Bolívia: desde o litoral norte até o Rio Grande do Sul, estendendo-se até a Argentina e o Paraguai; no norte, penetra ainda na Guiana Francesa e se estendeu pelo Rio Negro e pelo Uaupês até alcançar a Colômbia. Este é o caso mais notável de expansão de uma família lingüística na América do Sul, embora tenha análogos nas famílias karib e aruak. A distribuição da família tupi-guarani é, no entanto, essencialmente sul-amazônica: sua penetração

ao norte do rio Amazonas se deu recentemente, a partir do século XVIII, seja no caso da língua oyampí (e dialetos: emeriõ, karipuna do Uaçã), que pelo vale do Jari foi chegar até o Oiapoque, na Guiana; seja no caso da língua geral do Amazonas ou nheengatú, que penetrou no vale do rio Negro na qualidade de língua da frente pioneira brasileira; em lugar do português, que só tardiamente a substituiu.

A família tupi-guarani compreende umas quinze línguas ou grupos dialetais: além das já mencionadas língua geral do Amazonas e oyampí, são exemplos clássicos dela o tupi antigo ou tupinambã, que nos séculos XVI e XVII se estendia pelo litoral, de São Paulo até o Maranhão; o guarani, que conhecemos na forma em que foi documentado nas reduções jesuíticas da Província de Guairá, isto é, da ampla região compreendida entre os rios Paranapanema e Iguazu, Paraná e Tibaji, no atual Estado do Paraná, bem como na forma dos dialetos modernos do sul de Mato Grosso, São Paulo, Paraná, Rio Grande do Sul. Algumas outras línguas tupi-guarani cujos nomes são mais conhecidos são o guajajara, o urubú (kaapõr), o parintintin, o kamayurã, o kayabí, o tapirapê, o mawé ou sataré, o awetí. Estas duas últimas se destacam linguisticamente das demais: enquanto a maioria das línguas da família tupi-guarani revelam extrema homogeneidade gramatical, que as separa nitidamente das outras famílias do tronco tupi, o mawé e o awetí apresentam desvios notáveis, revelando traços que os aproximam de uma ou outra dasquelas famílias.

Aparentada ao tronco tupi é, seguramente, a família karib, mas a natureza de seu parentesco com esse tronco ainda está sendo determinada. Esta família, ao contrário do tronco tupi, tem uma distribuição predominantemente norte-amazônica. Ao sul do Amazonas encontram-se hoje apenas dois sub-grupos desta família, ambos no Xingu: um constituído pelas línguas kalapalo, kuikúru, nahukwã, matipú e bakairí, e o outro pela língua txikão, estreitamente afim do parirí, do arara do Xingu e do apiakã do Tocantins, que já não são mais falados. As línguas karib

ao norte do Amazonas, que se estendem à Guiana, à Venezuela e à Colômbia, incluem, entre outras, no Parã, Amapá e Roraima, o galibí, o apalaí, o waiwai, o hixkaryana, o taulipang, o tiriô, o makuxí.

Outro tronco linguístico é o que se tem convencioneado chamar de macro-jê, o qual inclui seis famílias e umas cinco línguas não classificadas ao nível de família. A maior família linguística e a que melhor tem sobrevivido é a jê. Antigamente as línguas dessa família se estendiam do Maranhão e Parã oriental, passando por Goiás e São Paulo, até os três estados do Sul, alcançando também o nordeste da Argentina. Hoje verifica-se uma grande interrupção nessa distribuição, principalmente devido à extinção dos kayapô do sul, que viviam até o século passado no sul de Goiás, oeste de Minas e norte de São Paulo. Ao sul, nos estados de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul, ainda se fala a língua kaingang e, estreitamente aparentada a ela, a língua xokleng, esta na parte oriental de Santa Catarina. Ao norte, há o grupo dialetal timbira no Maranhão, Parã e norte de Goiás, que inclui os índios canelas, os gaviões do Tocantins, os apinayé, os kraô; o grupo dialetal kayapô, que se expandiu para oeste, no sul do Parã e norte de Mato Grosso, até além do rio Xingu, e em que se contam as línguas dos xikrí, gorotire, kubên-krã-kégn, txukahamãe; o grupo dialetal akwên, de que sobrevivem ainda as línguas xerente e xavante; e ainda a língua suyã, do Alto Xingu.

Outras famílias, cuja filiação ao tronco macro-jê tem sido proposta, são a kamakã, do sul da Bahia, cujas línguas já desapareceram todas; a maxakali, do nordeste de Minas e sul da Bahia, de que resta só a língua maxakali; a família coroadó, do leste de Minas e Estado do Rio de Janeiro, inteiramente desaparecida; a família kariri, que se estendia no passado da Bahia à Paraíba, mas cuja última língua foi substituída pelo português; a família bororo, que se estendia do centro de Mato Grosso para oeste até a Bolívia, e de que restam duas línguas, o

bororo oriental do rio São Lourenço e o umotina do rio Paraguai; e ainda as línguas botocudo (norte do Espírito Santo e nordeste de Minas), iatê ou fulniô (sul de Pernambuco), Karajá (Ilha do Bananal), ofayê (sul de Mato-Grosso) e guatô (rio Paraguai).

Independentes dos troncos tupi e macro-jê há ainda umas doze outras famílias lingüísticas em território brasileiro. A família aruak talvez integre um terceiro tronco, associada com a família arawá. As línguas desta última - kulina, yamadai, paumari, dani, etc. - situam-se na bacia do rio Purus, no sudoeste do Amazonas e no Acre, e se estendem ao território peruano. As línguas brasileiras da família aruak são cerca de trinta e se encontram desde o extremo norte, como o palikur no norte do Amapá ou o grupo dialetal conhecido como baniwa do Içana, no noroeste do Amazonas, até o sul de Mato Grosso e centro de São Paulo, como o terena. Algumas outras línguas da família aruak são o paresi do oeste de Mato Grosso, o grupo dialetal do Alto Xingu (waurá, mahinaku, yawalapití), o wapitxana do norte de Roraima.

Outras famílias são xirianá (ou yanonami) ao longo da fronteira com a Venezuela em Roraima e norte do Amazonas; tukano no noroeste do Amazonas; makú, na mesma região; katukina, no oeste do Amazonas; pano, no sudoeste do Amazonas, Acre e Rondônia (de onde se estende para o Peru e Bolívia); mura, na bacia do Madeira, no Amazonas e Rondônia; txapakura, em Rondônia (e Bolívia); nambikwara, no noroeste de Mato Grosso; guaikuru, de que sobrevive a língua kaduvêu, no sudoeste de Mato Grosso.

Línguas até agora não classificadas em nenhuma família e em nenhum tronco, as chamadas "línguas isoladas" ou isolados lingüísticos, são, entre outras, o mâku e o awakê, na fronteira de Roraima com a Venezuela; o tukuna, na fronteira do Amazonas com o Peru e a Colômbia; o jabuti, o arikapú, o kapixaná e o huarí, em Rondônia; o aripaktsá ou ribaktsá, no norte de Mato Grosso (para o qual foi proposta recentemente filiação ao tronco macro-jê); o irantxe ou menkú no oeste de Mato Grosso; o trumai no Alto Xingu.

6. Para dar uma ligeira idéia das semelhanças e dissemelhanças das línguas indígenas, a partir das quais devem ser elaboradas as hipóteses de relações genéticas que levam ao reconhecimento de famílias e troncos lingüísticos, vamos examinar umas dez palavras em algumas línguas:

	tupinambã	guarani	xetã	língua geral
1. minha mão	xê pô	txê pô	txi pô	se pũ
2. mão dele	ipõ	ipõ	ïpo	ipũ
3. me viu	xê repiãk	txê retxãg	txi rêtxa	se xipiã
4. foste	eresõ	erehõ	ereõ	resũ
5. caímos	oro'âr	oro'ã	oroã	iaãri
6. pedra	itã	itã	ïta	itã
7. sol	kwâr kwarasy	kwarahý	ñanetxãpe	kwarasi
8. mel	eïr	eï	ëi	ïra
9. arco	wyrapãr	wyrapã	warãpa	mirapãra
10. machado	jý	jý	ñeprãka	

Facilmente se percebe que essas quatro línguas são muito semelhantes entre si. O tupinambã e o guarani concordam em todos os itens, não só nos sons idênticos, mas também nos diferentes, que se correspondem de maneira sistemática: x = tx em 1 e 2, s = h em 4 e 7, r final = ausência de som em 5, 8 e 9 (o tupinambã tem duas formas para 7, uma breve outra longa). A língua geral se mostra mais semelhante ao tupinambã (s = s em 4 e 7, r = r em 5, 8 e 9, pi na segunda palavra de 3), ao passo que o xetã é mais próximo do guarani (tx = tx em 1 e 3, ausência de r em 5, 8 e 9, tx na segunda palavra de 3). Todas as quatro são línguas da família tupi-guarani. As diferenças entre elas são predominantemente fonológicas, isto é, dos sistemas de sons. Nos exemplos acima, só a língua geral apresenta um desvio gramatical, no número 5: as demais línguas distinguem duas formas para "nós" nos verbos, oro- "eu e ele(s)" e ja- "eu e você(s)", mas a língua geral não faz essa distinção e só tem a forma ia- com os dois significados. O xetã apresenta duas diferenças de vocabulário, as quais decorrem da substituição de expressões simples por expressões descritivas: no número 7, ñanetxãpe significa

"nossa luz" (em que *-txápe* corresponde ao guarani *-etsapê*) e, no número 10, *ñeprákã* significa "instrumento para colher" (correspondente ao guarani *jeporakáhã*), o que se explica por servir o machado principalmente para os xetã colherem frutas (cortando galhos) e mel e larvas de insetos (abrindo troncos). Agora, comparemos com as quatro línguas acima um segundo grupo de quatro:

	awetí	mundurukú	tupari	kipeã
1. minha mão	ipô	oby	opo	hiamysã
2. mão dele	ipô nãpô	iby	ipo	samysã
3. me viu	itúp	odjop	otoa	subí hidohó
4. foste	etô	edjy	etera	ewí
5. caímos	ozo'át	otxe'at	otekat	hidzidé
6. pedra	kytã	wita'a	wa'i	kró
7. sol	kwát	kaxi	kiakop	utxé
8. mel	ekýt	ewit	ewit	kentí
9. arco	yzapát	daryk	pen	seridzê
10. machado	ký	o'a	wi	bodzô

Ainda se reconhecem semelhanças, mas estas já não são tão óbvias, nem entre si, nem com o grupo precedente. O kipeã não mostra nenhuma semelhança com qualquer das outras sete línguas. O awetí mostra alguma coincidência com as primeiras quatro línguas, em particular com o tupinambã, principalmente se considerarmos que t corresponde a r em final de palavra nos itens 5, 7 e 9. No item 2 o awetí tem duas formas, nãpô empregada por homem, ipô usada por mulher; esta última é idêntica à que ocorre nas línguas do primeiro grupo, a segunda contém um elemento comum (pô "mão") e outro estranho àquele grupo (nã- "dele" na fala dos homens). Do mundurukú, o que mais imediatamente lembra o primeiro grupo são talvez as formas 5, 6 e 8, do tupari as formas 2 e 8. Há, entretanto, maior semelhança entre o mundurukú e o tupari nos marcadores de pessoa (o- "minha" em 1, i- "dele" em 2, o- "me" em 3, e- "tu" em 4, otxe- e ote- "nós" em 5) e na forma 8.

Com exceção do kipeã, essas línguas se filiam todas ao tronco tupi e representam três famílias: família tupari, família mundurukú e família tupi-guarani, esta abrangendo as quatro línguas do primeiro grupo e mais o awetí. Este é um

membro mais divergente dentro da família tupi-guarani, mas suas concordâncias com esta, na amostra acima, vão além das já assinaladas. Sistemáticamente o awetí apresenta ky onde nas outras línguas da família há ou houve jy ou ji: o número 10 é exemplo de ky = jy, enquanto que nos números 6 e 8 as formas das demais línguas tupi-guarani devem ter sido *jitá e *ejít, tendo o j desaparecido sistemáticamente quando diante de i (o mundurukú nos números 6 e 8 e o tupari nos números 8 e 10 têm wi nesse caso). Outra correspondência sistemática é a do t awetí ao s tupinambá e h guarani; o único caso aparente na amostra acima é o número 4: awetí -to = tupinambá -so = guarani -ho, todos significando "ir"; um outro caso teremos com o número 4 do awetí, em que -tup "ver" não corresponde às formas do mesmo significado em tupinambá e guarani, mas a -sub e -hu "visitar, achar". Considerando este último caso, verificamos que todas as dez formas do awetí encontram correspondência, senão total pelo menos parcial, nas demais línguas da família tupi-guarani. O mesmo não se dá com o mundurukú e o tupari: do primeiro as formas 7, 9 e 10 e do segundo as formas 4, 6 e 9 não encontram correspondentes em línguas da família tupi-guarani. Note-se que o número 7 do tupari, embora não corresponda ao 7 das demais línguas, significa literalmente "nosso (ki-) calor (-akop)", e seu último elemento coincide com o tupinambá -akúb e guarani -akú "quente". O uso de "nosso calor" para designar o sol é análogo ao xetá ñanetxápe, cujo significado literal é "nossa (ñane) luz (txápe)".

A língua kipeá, que na amostra não oferece nenhuma concordância com as demais, não pertence ao tronco tupi, mas ao tronco macro-jê.