

POR QUE AS LÍNGUAS SÃO DIFERENTES?

GIULIA BOSSAGLIA

Universidade Federal de Minas Gerais

No mundo existem hoje cerca de sete mil línguas vivas. Destas, pouco mais de vinte são faladas por cerca de 50% da população mundial — só para mencionar algumas: o chinês (mandarim), o inglês, o espanhol, o português, o híndi, o árabe. No entanto, quase três mil línguas se encontram em perigo de extinção, por serem faladas por muito poucas pessoas e por sofrerem a pressão das línguas dominantes no panorama mundial — línguas geralmente desconhecidas, sem a menor visibilidade ❶. O fato de apenas uma minoria das línguas efetivamente existentes ser mais “famosa” poderia dar aos leigos a impressão, errada, de que não há assim tantas diferenças entre as línguas do mundo. Isso porque, por exemplo, várias dessas línguas pertencem à mesma família e tendem a manifestar semelhanças, as mais evidentes pertencendo ao nível lexical (nas palavras em geral). Palavras como *father* em inglês, *padre* em espanhol, *pai* (e cf. *padre*) em português e *pitā* em híndi (todas línguas indo-europeias), entre muitas outras que se poderiam escolher como exemplo, são muito parecidas entre si e, ao mesmo tempo, muito diferentes de *fùqīn* em chinês (família sino-tibetana) e *alab* em árabe (família afro-asiática), todas significando “pai”.

Além disso, em geral, não há uma motivação lógica ou natural para determinado significado ser associado a uma sequência de sons, porque tal associação é majoritariamente arbitrária: consequência disso é que línguas distintas expressam com palavras distintas os mesmos significados — não há nada no significado de “pai” que determine que em inglês se diga *father*, ou em chinês *fùqīn*, ou em árabe *alab*.

No entanto, as diferenças entre as línguas podem ser observadas em níveis mais profundos e estruturais, não apenas nas palavras. Existem línguas como o pirahã (família mura; Brasil) ou o rotokas (família papua; Nova Guiné) que possuem apenas onze fonemas (sons) distintos, entre vogais e consoantes, e outras como o archi (família caucasiana; Daguestão) com noventa e um, ou o !xu (família khoisan; Namíbia, Botsuana) com 141 fonemas ❷. Para ter uma referência, em média, as línguas possuem aproximadamente 22-25 consoantes e 5-6 diferentes vogais ❸.

O português brasileiro, por exemplo, possui dezenove fonemas consonantais e dez vocálicos, entre os quais as vogais nasais, bastante raras nas línguas do mundo.

As línguas se distinguem pela maneira como formam suas palavras: algumas, como o vietnamita (família austro-asiática), o chinês, ou o

iorubá (família nigero-congolesa), possuem palavras em sua grande maioria monossilábicas e invariáveis, veiculando cada uma apenas um significado lexical ou gramatical:

vietnamita ④:

Tôi són nhà tắm.
1SG pintar casa banheiro
“Eu pinto o banheiro de casa.”

Outras, como a maior parte das línguas indo-europeias e o árabe, possuem palavras não muito grandes, mas flexionáveis, para veicular vários significados gramaticais diferentes ao mesmo tempo:

português:

Ceguei chegando.
Chegu-ei chega-ndo
chegar-1SG.IND.PRET.PF chegar-GER

Por outro lado, línguas como o turco (família altaica), o tâmil (família dravídica), ou o tagalog (família austro-asiática), possuem palavras às vezes muito longas, nas quais podem ser juntados, uns ao lado dos outros, muitos morfemas gramaticais diferentes:

turco (fonte: Kornfilt 2003 *apud*) ⑤:

Ben Hasan koşturdum.
ben Hasan-ı koş-tur-du-m
1SG Hasan-ACUS correr-CAUS-PRET-1SG
“Fiz o Hasan correr.”

Finalmente, há línguas em que as palavras são tão complexas quanto frases, podendo incluir simultaneamente os lexemas de sujeito, objeto direto, verbo e até outros elementos — como muitas das línguas nativas da América do Norte:

groenlandês ocidental (fonte: Fortescue 1984 *apud*) ⑥:

Paasinngilluinnarpara ilaajumasutit.
Paasi-nngil-luinnar-para ilaa-juma-sutit.
entender-não-completamente- vir-querer-você.PART
eu.SUJ.isso.OBJ.IND

“Não entendia completamente isso, (ou seja) que você queria vir.”

As línguas manifestam preferências distintas também no que diz respeito à ordem de sujeito, objeto e verbo nas frases declarativas: cerca

de 75% apresentam a ordem SOV, como o japonês (família japônica), o somali (família afro-asiática), o apinajé (família jê), ou svo, como o wolof (família nigero-congolesa), o indonésio (família austronésia), o biri (família Pama-Nyungan); cerca de 13% não manifestam preferência por uma ordem específica, podendo apresentar qualquer uma, como o nhanda (família Pama-Nyungan); e as ordens vso, vos, osv e ovs existem também, apesar de serem bem mais raras que as outras ⑦. Por exemplo, uma frase como *João comprou vários livros* seria construída como *João vários livros comprou* em japonês, língua SOV, mas como *Comprou vários livros João* em malgaxe (família austronésia).

Emerge destes poucos exemplos a existência de uma forte variação entre as línguas do mundo, em todos os níveis linguísticos. Resta entender de onde se originam todas essas diferenças, o que nem sempre é uma tarefa simples.

Como se viu pela diversidade de famílias linguísticas nos exemplos acima, um fator importante em determinar as diferenças entre as línguas é sua procedência genética, ou seja, a família à qual elas pertencem. As línguas de uma mesma família compartilham a mesma língua ancestral, da qual herdaram palavras e características específicas, distintas daquelas que línguas de outras famílias herdaram de suas línguas ancestrais. Isto explica de uma maneira simples por que entre o inglês, o português, o espanhol e o híndi as palavras são em geral mais parecidas, mas diferentes de palavras com o mesmo significado em línguas como o chinês e o árabe, pertencentes a outras famílias.

No entanto, a existência de relações genéticas entre as línguas não é necessariamente garantia de que elas não apresentem muitas diferenças. O híndi, por exemplo, se diferencia bastante do inglês, do português e do espanhol: possui ordem SOV na frase declarativa (*usne seb khaayaa* “ele uma maçã comeu”), e não svo; tem posposições (*mez par* “mesa em” = “na mesa”), e não preposições; é uma língua de caso, ou seja, os substantivos são flexionados de maneira distinta (e acompanhados de posposições diferentes) dependendo da função sintática que eles têm na frase: *lark-ā* “rapaz” é sujeito na frase intransitiva e, quando acompanhado pela posposição *ne*, também na frase transitiva. Já a forma *lark-e* é utilizada para objeto indireto (“ao rapaz”) e outros complementos, a depender também das posposições que a acompanham ⑧.

Todas essas diferenças têm a ver, em primeiro lugar, com o fato de que, dentro da família indo-europeia, a relação não é tão próxima entre os ramos genealógicos do híndi e do espanhol e português ou do inglês. O português e o espanhol são irmãos diretos, o inglês não, mas é fortemente latinizado, como se verá adiante. Portanto, como nas famílias biológicas, existem relações mais ou menos próximas entre os membros. A proximidade genética entre línguas se manifesta na presença de semelhanças mais evidentes, mas elas podem facilmente desaparecer, sobretudo quando entre as línguas ancestrais e as línguas mais recentes de uma família se passaram milhares de anos.

De fato, todas as línguas mudam ao longo do tempo, podendo adquirir características novas que originariamente não eram próprias de suas línguas ancestrais, ou perder traços que as caracterizavam: o sânscrito védico (II-I milênio a.C.), a língua ancestral mais próxima (comparada com o protoindo-europeu, o ancestral de toda a família) do moderno híndi, por exemplo, possuía preposições que se “perderam” ao longo de séculos, desenvolvendo o híndi um sistema de posposições ☹. O latim era uma língua de caso, mas as suas “filhas”, como o espanhol e o português, mantiveram oposição morfológica de caso apenas nas formas pronominais: *eu* vs. *me* vs. *mim*. O inglês também perdeu a marcação de caso nos substantivos, mantendo-a apenas nos pronomes (*he* “ele” sujeito vs. *him* objeto), enquanto o antigo inglês e, também, a língua ancestral de todas as línguas germânicas possuíam a marcação de caso nos substantivos. Entende-se, ainda, que o protoindo-europeu não possuía nenhuma marcação de caso, um traço inovador que as línguas descendentes mais antigas desenvolveram e as mais recentes perderam parcial ou completamente (em alguns casos, mantiveram). Mencionou-se que o chinês é uma língua com palavras invariáveis e formadas em sua maioria por apenas um morfema lexical ou gramatical, mas na protolíngua (ou seja, a língua ancestral) e nas línguas mais antigas da família sino-tibetana existiam sufixos e prefixos (em sua maioria derivacionais) que podiam ser acrescentados às palavras: protossino-tibetano **mək* “tinta” : **s-mək* “preto”, **kat* “ferir, machucar” : **fi-kat* “ser ferido, machucado” : **fi-kat-s* “ferida”; antigo chinês **tjək* “tecer” : **tjək-s* “objeto que foi tecido” ☹☹.

Por um lado, portanto, pertencer a famílias diferentes explica diferenças entre línguas; por outro, as mudanças que afetam qualquer

língua ao longo do tempo podem acentuar diferenças até entre línguas aparentadas.

Diversos fatores contribuem para que as línguas mudem ao longo do tempo. Alguns são propriamente linguísticos, como o fato de que na língua falada os sons são pronunciados com pequenas alterações, que ao longo do tempo podem mudar a forma das palavras e, muitas vezes, levar a mudanças mais profundas no sistema linguístico: a perda da marcação de caso no latim tardio e a consequente fixação de ordens específicas de palavras nas línguas românicas se deram devido à simples perda de [m] e [s] em final de palavra no latim popular falado 11. De fato, os sufixos *-us* e *-um*, de caso nominativo (sujeito) e acusativo (objeto direto), respectivamente, ao perderem na pronúncia as consoantes finais, tornando-se idênticos, perderam também a capacidade de distinguir essas duas funções sintáticas. Assim, uma simples frase como *Paulus amat Marium* “Paulo ama Mário” — cujo significado podia ser entendido perfeitamente independentemente da ordem em que se colocassem o sujeito e o objeto direto (*Marium amat Paulus*, *Marium Paulus amat*, *amat Paulus Marium* etc.) — se tornou de repente ambígua sem *-s* e *-m* finais, “obrigando” as línguas românicas a fixarem a ordem sujeito-verbo-objeto para contornar essa ambiguidade: *Paulo ama Mário* e *Mário ama Paulo*, possuem significados completamente diferentes em português.

No entanto, fatores extralinguísticos — destacadamente o contato linguístico — contribuem também para que as línguas mudem ao longo do tempo. O inglês é a língua germânica com o léxico mais latinizado, devido à dominação dos normandos que, a partir do século XI, introduziram o latim como a língua da hierarquia católica e o francês (anglo-normando é o termo que designa este dialeto do francês falado pelos normandos na Inglaterra) como a língua da nova classe dominante. O inglês continuava a ser usado nos demais contextos comunicativos, mas numa situação de contato com o latim e o francês. Assim, com o tempo, o inglês recebeu muitos empréstimos dessas línguas, tornando-se seu léxico muito mais latinizado do que o léxico das outras línguas germânicas: por exemplo, *peace* “paz” procede do anglo-normando *pes* (por sua vez derivado do latim *pāx*), que substituiu o antigo inglês *frið*, de significado igual, mas de origem germânica e não latina, como o alemão *Frieden*, holandês *vrede*, sueco *fred*, islandês *friður*, que continuam a raiz **friðu-* da protolíngua germânica 12. Mudanças no léxico

podem eventualmente afetar níveis mais profundos nas línguas: a partir de empréstimos, podem ser introduzidos novos fonemas (a oposição entre /f/ e /v/ não existia em inglês, mas foi introduzida através de empréstimos do normando), ou podem ser criados novos morfemas (como o inglês *-able* que entrou por empréstimo do normando e se tornou extremamente produtivo também com raízes germânicas: *readable* “legível”, *bearable* “suportável”). No inglês, em particular, a grande quantidade de morfemas de origem normanda (e, em parte, latina) levou a uma parcial alteração do sistema acentual das palavras, que nas línguas germânicas prevê a preferência para o acento na primeira sílaba ou no radical, enquanto no inglês se tornou mais complexo e misto, incluindo muitos vocábulos com o acento nas sílabas finais, como, por exemplo, os verbos *annóy* “irritar”, *obéy* “obedecer”, *oppréss* “oprimir” 13, ou duplas verbo-nome como *óbject* : *objéct* “objeto : objetar, opôr-se”, *súspect* : *suspéct* “suspeito : suspeitar” 14.

O contato linguístico não necessariamente está relacionado com o contato geográfico entre comunidades linguísticas. Ele pode ser apenas cultural. Contudo, línguas muito isoladas, menos expostas ao contato, mudarão a uma velocidade muito mais branda que outras. Algumas características das línguas do mundo têm distribuições geográficas específicas, confirmando que o contato geográfico entre línguas também contribui para explicar as diferenças entre elas. Por exemplo, todas as línguas que possuem ordens sintáticas em que o objeto precede o sujeito se encontram principalmente na América Central e na região amazônica, independentemente de suas filiações genéticas 15.