

A Meia-Vida das Palavras

No princípio era o verbo. Um verbo irregular. Mas o tempo passou e ele tornou-se regular. Como pode a matemática ajudar a entender a língua nossa de cada dia? Um projecto inovador mostra em detalhe como a linguagem evolui – do irregular ao regular.

Verbos irregulares são um inferno. Tanto para as crianças quanto para os estudantes de uma língua estrangeira. Mesmo adultos fluentes e bem educados por vezes tropeçam em conjugações exóticas. Se o próprio nome faz-nos supor que tais estruturas são mais excepção do que regra, porque, então, os verbos mais comuns (ser, estar, ter, haver) são irregulares?

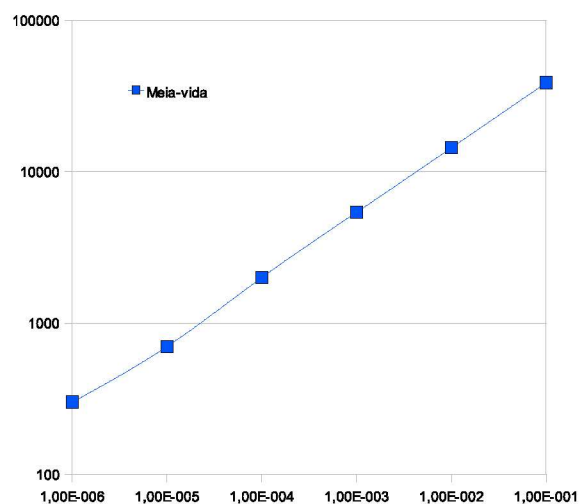
Esta questão tem intrigado os linguistas há muito tempo. Recentemente, um grupo de matemáticos e biólogos de Harvard resolveu tirar a questão a limpo com técnicas muito distantes das que são normalmente usadas no estudo da língua [1]. Tal equipa, liderada por Martin Nowak, do Programa de Dinâmica Evolutiva, tem aplicado modelos matemáticos simples para tudo aquilo que evolua: genes, pessoas, países e, porque não?, línguas.

Assim, classificaram tantos verbos irregulares em inglês quanto foram capazes de encontrar na literatura clássica e recente: desde *Beowulf* (século VII a IX) até a literatura contemporânea, passando pelos *Contos da Cantuária* (*Canterbury tales*), de 1200. No inglês arcaico do primeiro livro, havia 177 verbos irregulares. Já no medieval, apenas 145. Hoje resistem 98. Portanto, não há dúvidas, as línguas evoluem – gostem ou não os amantes da excepção!

Como se dá esta evolução? O que faz um verbo resistir a sua regularização e outros não? Em inglês, é bom lembrar, a regra básica é que a forma passada de um verbo faz-se pela adição de “-ed” em seu final. Raríssimas vezes um verbo regular torna-se irregular (como “sneak” - esgueirar-se, em uma tradução livre –

cujo passado modificou-se de “sneaked” para “snuck”). O caminho comum é o inverso, e é disto que trata o artigo de Liberman.

Para quantificar esta mudança, os autores tomaram emprestado um conceito da física nuclear: a “meia-vida”. A “meia-vida” de uma substância radioactiva é o tempo necessário para que a sua actividade diminua pela metade. Assim, os verbos foram divididos em seis grupos de acordo com o sua frequência de uso. No primeiro estavam os dois únicos verbos cujo uso é superior a um em dez (10^{-1}): “be” e “have”. Em seguida vinham os que têm uso



Relação entre a meia-vida dos verbos e a sua frequência de uso, em escala logarítmica. Os quatro primeiros numeros são calculados; os dois últimos estimados.

entre IOⁿ e IOⁿ. Este grupo inclui "come", "do", "find", "get", "give", "go", "know", "say", "see", "take" e "think". Os dez verbos mais comuns em inglês são todos irregulares e estão num destes dois grupos. A divisão continua, em mais quatro grupos marcados por potências sucessivas de 10. O grupo a que cada verbo pertence é definido pela sua frequência actual de uso, pois não há disponibilidade confiável de dados equivalentes para épocas passadas.

Os dois primeiros grupos (os de uso mais frequentes) não tiveram nenhuma regularização desde a época de *Beowulf*. No terceiro grupo, 10% dos verbos deixaram de ser irregulares. No quarto e no quinto grupos 43% e 72% dos verbos foram regularizados pelo uso, respectivamente. Finalmente, na categoria dos infrequentes, 91% dos verbos irregulares em inglês antigo já não o são.

O tamanho de cada grupo diminui no tempo, tornando possível associar uma meia-vida para os verbos de cada categoria. Para os verbos cujo uso está entre 10% e IOⁿ, a meia-vida foi estimada em 300 anos. Desta forma, "wreak" (vingar) mudou da forma passada "wrekan" para a regular "wreaked". Um exemplo na categoria acima, o tempo característico para a regularização obtido foi de 700 anos, tempo suficiente para o passado de "mourn" (lamentar) mudar de "mournen" para "mourned" mas insuficiente para "drove" (passado de dirigir) tornar-se "drived". Subimos uma categoria e a meia-vida sobe para 2000 anos. Mais um pouco, e temos 5400 anos, fazendo com que verbos regularizados como "help" se tornem, neste grupo, mais uma excepção do que uma regra. Finalmente, nas duas categorias de topo, encontramos tempos característicos de 14.400 e 38.000 anos, fazendo com que as mudanças nestes dois grupos sejam tão raras que provavelmente nunca ocorrerão. Podemos até dizer que antes de "be" e "have" virarem regulares a própria língua inglesa já terá deixado de existir.

Se não houve nenhuma regularização nestas duas categorias mais altas, como é possível saber a vida média de seus componentes? Usa-se um processo chamado "extrapolação". A partir das quatro categorias onde esta conta é possível chega-se a uma lei surpreendente pela sua simplicidade: a meia-vida da forma irregular de um verbo é proporcional ao inverso da raiz quadrada da sua frequência de uso. Isto permite-nos obter, por extrapolação dos resultados, a frequência de ocorrência de

regularizações em grupos de palavras onde ainda não aconteceu nenhuma regularização. E, sem surpresa, o tempo característico é muito maior do que o tempo de observação desta investigação.

A idéia de que as palavras menos usadas são mais rapidamente regularizadas é antiga. Aquilo que mais usamos está mais forte na nossa memória; além disto, alterar uma forma que usamos muito pode prejudicar a comunicação - portanto somos mais conservadores exactamente nestas palavras.

Outro estudo, publicado no mesmo número da *Nature* e que compara a velocidade de alteração das palavras em Inglês, Espanhol, Grego e Russo reforça esta ideia: o que usamos menos muda mais rápido [2]. É por este motivo que quando se compara duas línguas, as primeiras palavras a ser estudadas são *um, dois, noite, morrer* etc, palavras cuja má comunicação pode ser fatal e que, por serem das que menos se alteram, melhor permitem estabelecer relações entre as varias línguas.

E no português? Temos muito mais riqueza de conjugação, cada pessoa exigindo uma terminação verbal específica. Muitos verbos (aceitar, eleger, expressar, imprimir, fritar, matar, prender, suspender etc.) têm dois participios - um irregular, geralmente derivado directamente do latim, e um regular, frequentemente a forma mais recente, mostrando uma tendência próxima àquela descrita acima.

Uma curiosa consequência deste resultado é poder fazer uma previsão para o futuro. De acordo com Lieberman e colegas, o próximo verbo a ser regularizado é "wed" (casar). Assim, estamos a assistir o fim da época dos "newly-wed" (recém-casados) e entrando na época dos "newly-wedded". Uma rápida pesquisa na internet mostra que a nova forma, regular, já está muito corrente.

Referencias

[1] - Lieberman E, JB Michel, J Jackson, T Tang, MA Nowak (2007). "Quantifying the evolutionary dynamics of language" *Nature* 449:713-716

[2] - Pagel M, Atkinson AD, Meade A (2007). "Frequency of word-use predicts rates of lexical evolution throughout Indo-European history" *Nature* 449:717-720