

Físicos e linguista examinam a evolução do vocabulário de comunidades *on-line*

IGOR ZOLNERKEVIC

ILUSTRAÇÕES LARISSA RIBEIRO

Ninguém sabe quantas palavras nascem a todo momento. Os estudiosos da linguagem só têm certeza de que elas devem ser muitas e de que a imensa maioria é raramente usada, geralmente esquecida. Afinal, existem muito mais palavras do que um único ser humano conseguiria aprender ao longo da vida. Para se ter uma ideia, o serviço de busca Google registrou 13 milhões de palavras distintas em língua inglesa usadas pelo menos 200 vezes em páginas na internet até 2006, enquanto pesquisadores estimam que o tamanho do vocabulário de um adulto com bom nível educacional não ultrapassa 100 mil palavras.

O mistério da criação das palavras continua, mas um estudo publicado em maio na revista *PLoS ONE*, realizado pelos físicos brasileiros Eduardo Altmann e Adilson Motter em parceria com uma linguista norte-americana, ajuda a entender melhor como o vocabulário de uma comunidade evolui com o tempo. Ao analisarem estatisticamente milhares de palavras empregadas por quase 167 mil usuários de dois grupos de discussão na internet durante uma década, o trio de pesquisadores concluiu que as chances de uma palavra, velha ou nova, permanecer em uso no futuro não dependem tanto da frequência com que ela é usada atualmente, mas sim da variedade de assuntos em que é empregada e, mais importante ainda, do número de pessoas que a utilizam. Nas palavras do autor principal do estudo, Altmann, do Instituto Max Planck de Física de Sistemas Complexos, em Dresden, Alemanha, para manter a variedade de palavras em uso em uma comunidade, “é melhor muita gente falar pouco do que pouca gente falar muito”.

Esse não é o primeiro artigo sobre evolução do vocabulário assinado por Altmann e Motter, da Universidade Northwestern, em Evanston, no estado norte-americano de Illinois. A troca de mensagens entre milhões de pessoas por meios eletrônicos deixa vestígios na forma de bases de dados que cada vez mais os físicos estão se interessando em explorar, em busca de padrões que revelem a dinâmica social por trás da interação digital. “Físicos são muito bons em descobrir relações entre os mecanismos subjacentes e os padrões observados”, diz a outra autora do estudo, a linguista Janet Pierrehumbert, da Universidade Northwestern, sobre a colaboração. “Eles também são muito bons em fazer analogias entre um tipo de fenômeno e outro.”

Os pesquisadores escolheram analisar a atividade até 2008 de dois grupos de discussão da família de fóruns públicos Usenet, atualmente hospedados pelo Google, mas que já existiam em 1979, 10 anos antes da invenção das páginas *web*. Um dos fóruns estudados foi o *comp.os.linux.misc*, criado em 1993 para discutir o sistema operacional Linux, do qual participaram 128.903 pessoas, que iniciaram 140.517 tópicos de conversação. O outro foi o *rec.music.hip-hop*, grupo de discussão do gênero musical *hip-hop*, iniciado em 1995, em que 37.779 pessoas se engajaram ao menos uma vez em um dos 94.074 tópicos discutidos. O número total de palavras escritas pelos usuários de um desses grupos durante um intervalo de seis meses variava de quase 1 milhão a mais de 5 milhões.

Para quantificar como cada uma das palavras usadas nesses grupos era disseminada entre os usuários e os tópicos ao longo do tempo, não bastava simplesmente contar a cada seis meses o número de vezes que cada usuário usava a palavra e quantas vezes ela aparecia em cada tópico. A análise estatística precisou levar em conta o fato de que a atividade dos usuários e o tamanho das conversas desses grupos variavam muito. Alguns poucos usuários escreviam demais o tempo todo, enquanto muitos só contribuíam um pouquinho de vez em quando. Ao mesmo tempo, alguns poucos tópicos tinham mais de mil mensagens postadas, com a discussão durando mais de três anos, enquanto o tópico médio tinha cinco mensagens, durando cinco dias. No fim, conseguiram definir uma quantidade que mede o grau de disseminação de uma palavra entre usuários e conversas independentemente da frequência daquela palavra. Dessa maneira, conseguiram comparar a disseminação de palavras de ocorrência rara com a de palavras de uso frequente.

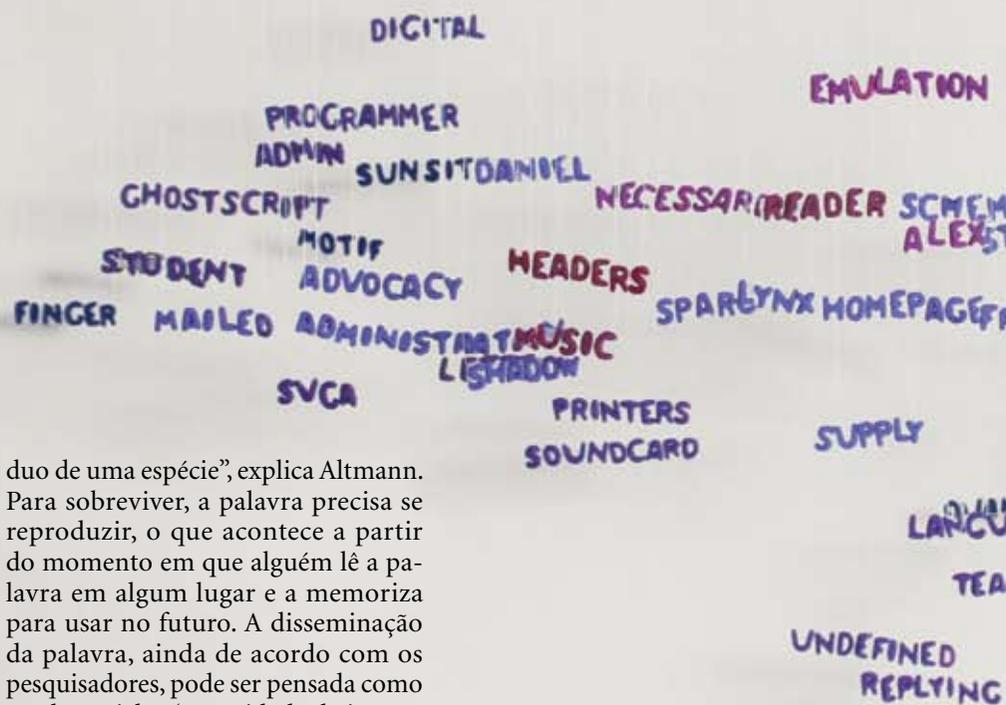
Presente e futuro - O passo seguinte foi comparar o número de vezes que cada palavra apareceu nas discussões e a medida de disseminação de cada uma delas em um período de seis meses com as mudanças na frequência de uso delas dois anos depois. Computando os números, os pesquisadores observaram que a frequência de uso de uma palavra num determinado momento informava pouco sobre a frequência com que seria empregada no futuro. Eles viram ainda que a quantidade de vezes que uma palavra seria mencionada dois anos mais tarde parecia ter uma estreita relação com a disseminação das palavras no passado. Concluíram, então, que a probabilidade do uso de uma palavra aumenta à medida que cresce o número de pessoas que a utilizam. Isso significa que, mesmo se uma palavra for muito utilizada hoje, ela corre o risco de cair em desuso daqui a alguns anos se o número de conversas e de tópicos em que é citada hoje for baixo.

Segundo os pesquisadores, a situação lembra muito a dos seres vivos lutando por sua sobrevivência. Cada palavra pode ser pensada como uma espécie biológica. “Cada uso da palavra pode ser comparado a um indiví-

VOCÁBULOS FLUTUANTES

De 1998 a 2000, palavras em inglês com alto grau de disseminação entre membros de um grupo de discussão sobre o sistema Linux cresceram em popularidade. No gráfico ao lado, elas estão representadas em cores que vão do vermelho ao amarelo. As palavras com baixa disseminação (do lilás ao preto) passaram a ser menos usadas. A variação na popularidade não dependeu da frequência de uso.

FONTE: EDUARDO ALTMANN



duo de uma espécie”, explica Altmann. Para sobreviver, a palavra precisa se reproduzir, o que acontece a partir do momento em que alguém lê a palavra em algum lugar e a memoriza para usar no futuro. A disseminação da palavra, ainda de acordo com os pesquisadores, pode ser pensada como sendo o nicho (capacidade de interação) da espécie no ambiente. Quanto mais estreito o nicho de uma espécie, mais risco ela corre de extinção. Por isso, uma explosão populacional não garante a sobrevivência da espécie, se o seu nicho for pequeno. “A palavra precisa estar distribuída entre certo número de usuários, caso contrário ela morre”, diz o físico.

Um dos resultados a que o grupo chegou – o fato de a frequência de uso da palavra no presente não influenciar a frequência de uso no futuro – contraria a conclusão de estudos recentes que analisaram a dinâmica das palavras em

períodos muito mais longos (séculos) e demonstraram a importância da frequência. No mais conhecido deles, publicado na revista *Nature* em 2007, um grupo liderado por Erez Lieberman, atualmente professor visitante no Google, mostrou que, em inglês, os verbos irregulares pouco usados tendem a se transformar em verbos regulares, enquanto apenas os mais adotados pela população mantêm a forma irregular. Isso explicaria por que o verbo irregular *to be*, o mais usado da língua, ainda é e deve permanecer irregular. Ao passo que o verbo irregular *to slink*,

FER

ASIEST MANUALS

STATES
EXCEPTION
FORWARD
ENTRIES
ODD
EXAMPLES
SHOT
RECONFIGURING
SUGGESTED
KILLED
SEARCHING
SAVED
CREATES
DETERMINE
INITIALIZE
EMPTY
STEPS
CUSTOM
REACH
PRINTED
FORMATTED
MATCHFIRE
AFFRAID
SOLVED

STANDARDS

EELY TASKS
LEARNED
GOD
LATE
HACK
APPLE
CUSTOMERS

M

MARKETING CHIPS

FREWARE

que significa caminhar sinuosamente e praticamente não é conhecido das pessoas, vem perdendo sua forma de passado irregular *slunk* em favor da variante regular *slinked*.

Janet acredita que esses estudos históricos estejam analisando casos muito específicos em que duas palavras competem por um mesmo nicho na linguagem. Ela explica que a maioria das palavras não está em competição umas com as outras, uma vez que sinônimos absolutos são extremamente raros. Por exemplo, as palavras *yes* e *yup* podem ambas significar “sim”,

mas o fato de a última ser mais coloidal que a primeira implica que cada uma é usada em situações diferentes e, portanto, cada uma tem seu nicho garantido. “Prevejo que esses fatores [disseminação entre usuários e tópicos] se mostrarão também muito importantes para explicar as flutuações de frequências em tempos históricos [da ordem de séculos]”, ela diz.

Nesse sentido, Altmann sugere que as medidas de disseminação das palavras desenvolvidas por eles sejam aplicadas em qualquer outra base de dados análoga, como a dos mais de 5 milhões de livros digitalizados pelo Google Books – alvo de estudo recente publicado na *Science* e encabeçado por Lieberman, que mensurou e comparou a frequência de diversas palavras-chave de interesse histórico e cultural (ver Pesquisa FAPESP nº 183). Nesse caso, os autores dos livros fariam o papel dos usuários e cada livro poderia ser pensado como uma entrada em certo tópico de discussão.

Outros dois resultados da análise dos grupos do Usenet intrigaram os pesquisadores. Um deles foi o fato de que a disseminação entre os usuários influencia as mudanças de frequências das palavras mais do que a disseminação entre os tópicos. O outro resultado é que as palavras em geral são mais ligadas a usuários do que a tópicos. Juntas, essas conclusões revelam que as idiosincrasias dos indivíduos ou de

subgrupos de indivíduos têm um papel central na manutenção do vocabulário da comunidade. “Quem ler as mensagens no grupo de *hip-hop*, por exemplo, vai perceber que as pessoas fazem um esforço para escrever de maneira diferente das demais para se posicionarem socialmente”, diz Altmann.

A disseminação das palavras não é o único fator determinante de seu sucesso. Altmann e seus colegas observaram que palavras ligadas a produtos comerciais como *wireless* e *Gnome* (plataforma de distribuição de Linux) ou a personalidades como Bush e o *rapper* norte-americano Eminem começavam sua vida nos grupos de discussão com um grau de disseminação muito baixo – o que, em princípio, as teria fadado ao esquecimento. Mas, nesses casos, forças externas aos grupos, como campanhas publicitárias e veiculação de notícias nos meios de comunicação, agiram para que essas palavras fossem incorporadas ao vocabulário deles.

Já as gírias e os jargões bem aceitos pelos grupos seguiram a tendência estatística das demais palavras, sugerindo que a aceitação deles dependia mais de fatores internos que externos. A linguista Eleonora Albano, da Universidade Estadual de Campinas, comenta que gírias e jargões são adotados por uma comunidade se contribuem para construir a identidade do grupo social.

Maria Helena Neves, linguista da Universidade Estadual Paulista e da Universidade Presbiteriana Mackenzie, considera interessantes os estudos quantitativos sobre conversações *on-line*, mas suspeita de que seus resultados não possam ser generalizados para a dinâmica da língua falada. “A amostra é restrita por causa do canal de expressão escolhido, do perfil dos usuários e do propósito da interação”, diz. Ela, aliás, desconfia sempre de generalizações. “Em linguagem não há receita pronta para nada, senão não existiriam a literatura e a poesia.” ■

» Ver gráfico original no site:
www.revistapesquisa.fapesp.br

Artigo científico

ALTMANN, E.G. *et al.* Niche as a determinant of word fate in online groups. **PLoS ONE**. mai. 2011.