

REVEL NA ESCOLA: SEMÂNTICA LEXICAL

Mário A. Perini¹

Universidade Federal de Minas Gerais

Às vezes me perguntam: quantas palavras uma pessoa sabe? Essa é uma pergunta importante, principalmente para quem ensina línguas estrangeiras. Seria muito útil para quem planeja um curso de francês ou japonês ter uma estimativa de quantas palavras um nativo conhece; e quantas os alunos precisam aprender para usar a língua com certa facilidade. Também seria muito bom saber quantas palavras uma pessoa consegue aprender por semana. Essas informações seriam preciosas para quem está preparando um manual que incluía, entre outras coisas, um planejamento cuidadoso da introdução gradual de vocabulário.

À parte isso, a pergunta tem seu interesse próprio. Uma língua não é apenas composta de palavras (itens lexicais): inclui também regras gramaticais e um mundo de outros elementos que também precisam ser dominados. Mas as palavras são particularmente numerosas, e é notável como qualquer pessoa, instruída ou não, tem acesso a esse acervo imenso de informação com facilidade e rapidez. Assim, perguntar quantas palavras uma pessoa sabe tem também importância teórica, porque é parte do problema geral de o que é que uma pessoa tem em sua mente e que lhe permite usar a língua, falando e entendendo.

Mas antes de tentar responder essa pergunta, há outras que precisam ser consideradas. A primeira é simplesmente: o que é uma “palavra”?

Ora, alguém vai dizer, essa pergunta já está respondida: *todo mundo* sabe o que é uma palavra. Para que problematizar uma noção que é de conhecimento geral? Mas não é assim: a palavra *palavra* é usada de mais de uma maneira, e antes de podermos usá-la com clareza é necessário tirar do caminho algumas ambiguidades. Vou começar, portanto, mostrando como essa palavra é pouco precisa em seu uso cotidiano.

¹ Doutor em Letras pela University of Texas. Professor Voluntário na Faculdade de Letras da Universidade Federal de Minas Gerais – UFMG.

PALAVRA E LEXEMA

Vamos considerar a palavra *olho*. É muito claro que isso aí é uma palavra – mas será que *olhos* é a mesma palavra (só que no plural)? Ou será outra palavra?

Bom, há razões para responder das duas maneiras: é a mesma palavra, porque significa a mesma coisa (mas com o acréscimo de “mais de um”); e é outra palavra, porque se pronuncia diferentemente (*olhos* tem um *-s* final que *olho* não tem).² Ou seja, a pergunta faz sentido: *olho* e *olhos* é uma palavra ou duas? Antes de responder, vamos complicar um pouco mais a questão. Vamos considerar *olho* e *olheira*: é uma palavra, ou são duas? Aqui em geral as pessoas não hesitam, e dizem que são duas palavras. Por que é que não hesitamos no caso de *olho/olheira*, mas ficamos na dúvida com *olho/olhos*? Veremos que há boas razões para essa reação aparentemente incoerente.

A razão principal é que a relação entre *olho* e *olhos* é extremamente regular; ou seja, vale não apenas para esse par, mas para milhares de outros pares de elementos da língua. Formalmente, podemos citar muitíssimos pares que só diferem pela presença de um *-s* no final em oposição a nada: *olho/olhos*, *orelha/orelhas*, *gato/gatos*, *nuvem/nuvens* (no caso de *nuvens* há uma pequena alteração, puramente gráfica, de *m* para *n*, mas isso não tem importância). E, semanticamente, a relação é a mesma em todos os pares: a forma sem *-s* denota um objeto só, a forma com *-s* denota mais de um objeto (ou seja, singular e plural, respectivamente). Daí se tira uma consequência importante: não é preciso aprender e guardar permanentemente na memória cada caso individual; aprendemos uma regra geral (“faz-se o plural acrescentando um *-s* ao singular”), e estamos prontos.

Nada disso vale para o caso do par *olho/olheira*. Primeiro, muitas palavras não admitem esse sufixo *-eira*, mesmo se considerarmos palavras de significado próximo ao de *olho*: não temos **narizeira*, nem **palpebreira*, nem **ouvideira*.³ Depois, a relação semântica entre *olho* e *olheira*, alguma coisa como “pele escurecida em volta do olho”, só acontece nesse caso. Mesmo quando temos uma forma em *-eira*, a relação semântica é quase sempre diferente; existe *queixeira*, mas não tem nada a ver com a pele do queixo: a queixeira é aquela peça em cima do violino onde o músico apoia o queixo. E *cabeleira* é apenas o conjunto dos fios de cabelo de uma pessoa. Ou seja, ao aprendermos as palavras em *-eira*, não podemos

² Além da diferença de timbre das vogais tônicas: [o] e [ɔ].

³ O asterisco (*) é usado para marcar formas que não existem.

seguir uma regra geral; somos obrigados a aprender cada caso individualmente.⁴ Acho que é isso que as pessoas estão captando (sem saber) quando têm tanta certeza de que *olho* e *olheira* são duas “palavras” diferentes, e desconfiam que talvez *olho* e *olhos* sejam a mesma “palavra”.

O leitor deve ter notado que coloquei aspas em “palavra”. É que agora temos que enfrentar um detalhe terminológico, e vai ser necessário adquirir dois termos técnicos (peço paciência, e prometo manter a terminologia sob controle). Vamos admitir que uma **palavra** é uma forma única, independentemente de ser ou não relacionada com outras formas: assim, *olho*, *olhos* e *olheira* são três palavras distintas.

Mas a história não pode ficar aí; nós já sabemos que a relação entre *olho* e *olhos* é diferente da que existe entre *olho* e *olheira*. Por exemplo, se você sabe português e aprende a palavra *olho*, não precisa aprender separadamente a palavra *olhos*; mas a palavra *olheira* tem que ser aprendida separadamente. Vamos ver um exemplo em outra língua, para deixar isso mais claro: digamos que você já sabe bastante inglês, e aprende a palavra *bed*, que significa ‘cama’. Automaticamente, você já sabe que *beds* é o plural dessa palavra, e significa ‘camas’. Existe também a palavra *bedding*, que quer dizer ‘roupa de cama’. Essa palavra, entretanto, precisa ser aprendida independentemente, porque não há nenhuma regra que diga que acrescentando *-ing* temos uma palavra que significa ‘roupa usada em X’. Isso só funciona para *bedding*; de *table* ‘mesa’ não se pode tirar **tabling*, que seria ‘toalha de mesa’, e um pano que se coloca em cima da cadeira (*chair*) para protegê-la da poeira não se chama **chairing*. Isso já mostra como a relação *bed / beds* é diferente da relação *bed / bedding*.

Como exprimir isso em nossa linguagem gramatical? A solução mais comum é a seguinte: continuamos chamando **palavras** as formas individuais – assim, *olho*, *olhos* e *olheira* são três palavras distintas. E chamamos **lexema** um conjunto de palavras relacionadas de maneira regular e sistemática: assim, *olho* e *olhos* (assim como *bed* e *beds* em inglês) são duas palavras do mesmo lexema; mas *olheira* (e *bedding* em inglês) é uma palavra pertencente a outro lexema. Veremos que essa distinção tem importância quando formos responder as perguntas iniciais deste artigo.

A distinção entre palavra e lexema é muito importante para a análise dos verbos: temos as palavras *ajudar*, *ajudamos*, *ajudo*, *ajudando* e *ajudante* – um exame cuidadoso vai mostrar que as primeiras quatro pertencem a um lexema (que se chama “o verbo *ajudar*”), mas *ajudante* é outro lexema, apesar de ser de certa forma relacionado com o verbo *ajudar*.

⁴ Há casos intermediários, mas vamos esquecê-los para efeito da discussão do momento. Nada na língua é realmente simples.

Isso porque para todos os verbos temos formas formal e semanticamente paralelas às quatro primeiras (*correr, corremos, corro, correndo*), mas nem sempre uma forma correspondente a *ajudante* (que significa “pessoa que ajuda”): *corrente* existe, mas não significa “pessoa que corre”; e não temos em português **comente* (de *comer*), nem **vente* (de *ver*), nem muito menos **inte* (de *ir*); temos *estante*, mas essa palavra não tem relação semântica regular com o verbo *estar*.⁵

Daí se conclui que quando estamos aprendendo uma língua, o que aprendemos são lexemas novos, não palavras novas. Muitas palavras nós já conseguimos saber mesmo sem aprender – mesmo sem nunca ter visto um exemplo. Basta que a palavra em questão pertença a um lexema conhecido. Assim, você amanhã aprende uma palavra nova – digamos, *condilartro*. Se precisar usar o plural, não vai ter dificuldade nenhuma: aprender um singular é automaticamente aprender o plural. Assim temos que reformular a pergunta feita no início: quantos lexemas uma pessoa sabe?

É difícil chegar a uma resposta exata, por razões óbvias. Mas estudos recentes citados por Paul Nation⁶ sugerem que um falante educado do inglês conhece cerca de 20 mil lexemas. Podemos tomar esse número como ponto de referência.

O QUE É “CONHECER” UM LEXEMA?

Conhecer 20 mil lexemas é muita coisa, claro – principalmente quando lembramos que esses lexemas estão aí, à disposição, mesmo quando não são usados há muitos anos. Eu mesmo, outro dia, me surpreendi usando o verbo *pocar*, autêntica gíria capixaba que significa ‘quebrar’, e que eu certamente não usava desde a infância. Quando precisei, o lexema estava lá, bom criado às minhas ordens.

Isso em si já é espantoso; mas fica ainda mais espantoso quando consideramos a complexidade que é conhecer um lexema. Vamos pegar um verbo comum como *bater*, e tentar explicitar pelo menos uma parte dos conhecimentos que precisamos ter para usá-lo dentro das frases da língua.

⁵ Em termos técnicos, um lexema é um conjunto de palavras relacionadas por flexão; palavras relacionadas por derivação não formam um lexema.

⁶ Nation, I.S.P. (2001) *Learning vocabulary in another language*. Cambridge University Press. Nation é provavelmente a maior autoridade mundial em ensino e aprendizagem de vocabulário.

Primeiro, para saber esse verbo temos que conhecer sua forma: *bater*; e precisamos saber que é um verbo regular da segunda conjugação, ou seja, inclui as palavras *bato*, *batemos*, *bati*, *batendo* etc., e não **batei* nem **batando*.

Depois, temos que saber o que ele significa: e muitos verbos (e lexemas em geral) têm diversos significados, às vezes muito diferentes. *Bater* pode corresponder a ‘espancar’ (*o menino bateu no cachorro*); ou ‘remexer’ (*vou bater o omelete*); ou ‘aparecer’ (*ela bateu lá em casa às sete da manhã*); ou ‘tocar’ (*o sino bate de manhã cedo*); ou ‘mover rapidamente’ (*o beija-flor bate as asas muito depressa*); ou ‘insistir’ (*you vive batendo nesse assunto*); ou ‘chocar-se’ (*a porta bateu com força*); ou ‘ultrapassar’ (*a nadadora bateu o recorde olímpico*); ou ‘derrotar’ (*o Brasil bateu a Argentina no amistoso*)...

À primeira vista parece que essa multiplicidade de significados vai resultar em uma imensa confusão – como é que vamos saber o que é que nosso interlocutor quer dizer quando usa o verbo *bater*? Mas nós temos recursos para neutralizar esse perigo. Um deles é a valência do verbo – ou seja, as construções em que ele pode aparecer. Em certos casos, a construção nos diz qual a acepção do verbo que está valendo. Assim, se o complemento vier regido da preposição *em*, sabemos que o significado deve ser ‘espancar’: *o menino bateu no cachorro*, ou ‘derrotar’: *o Brasil bateu na Argentina*; ou então ‘insistir’: *you vive batendo nesse assunto*, ou ainda ‘chocar-se’: *o carro bateu no poste*. Mas não pode ser ‘ultrapassar’, porque nesse caso o complemento não pode ter preposição: *a nadadora bateu o recorde olímpico*. Na acepção de ‘derrotar’, a preposição é opcional: *o Brasil bateu (n)a Argentina*, mas se for ‘remexer’, a preposição não pode aparecer: *vou bater o omelete*, não *no omelete*. Quando não tem complemento nenhum, *bater* pode significar ‘soar’: *o sino acabou de bater*, mas não ‘derrotar’: acho estranho dizer *esse time bate muito* no sentido de que vence muitas partidas; seria aceitável no sentido de ‘os jogadores desse time são violentos’.

Essa variedade de acepções vem do fato de que o lexema *bater* tem conexões variadas com diversas estruturas presentes na memória, e que recebem o nome de **esquemas**⁷. Um desses esquemas é o conceito (não a palavra) de ESPANCAR, que inclui em sua estrutura um agente (o espancador) e um paciente (o espancado), e que é evocado quando ouvimos a frase *o menino bateu no cachorro*. Para processar essa frase, temos que efetuar muitas operações mentais, entre as quais: [a] evocar o esquema ESPANCAR; [b] associar o conceito de ‘menino’ com o agente da ação correspondente ao esquema; [c] associar o conceito de ‘cachorro’ com o paciente da mesma ação (agente, paciente, assim como outras relações:

⁷ Também se chamam **frames**.

instrumento, lugar, causa etc., são chamados **papéis temáticos**). Ou seja, as diversas acepções do verbo *bater* se correlacionam com diversas funções semânticas de seus complementos. Para pegar apenas o sujeito⁸, em uma frase como *a Paulinha bateu o omelete*, o sujeito (*a Paulinha*) é agente, ou seja, pratica a ação. Mas em *me bateu um cansaço*, o sujeito (*um cansaço*) não parece ser agente; é antes alguma coisa que “se movimenta”, em sentido figurado, e “vem” para cima de mim. Nesse caso *bater* evocou outro esquema, que tem esses elementos em sua estrutura.

Com alguns verbos as diferenças são muito grandes: *apanhar*, um verbo muito idiossincrático, tem sujeito agente quando seu complemento não tem preposição: *o gato apanhou o passarinho* (o gato praticou a ação, logo é agente); mas se o complemento final tiver a preposição *de*, ele é o agente, e o sujeito é paciente (sofre a ação): *o gato apanhou do passarinho*.

O caso de *apanhar* é excepcional, e quase sempre o sujeito é agente se o verbo exprime ação. Mas o verbo pode exprimir outras coisas, e nesse caso não há agente, e o sujeito será outra coisa. Por exemplo, em *aquela mulher loura é a professora de matemática* o sujeito e o outro complemento não “fazem” nada; a frase simplesmente explicita que são a mesma pessoa (como se disséssemos: “aquela mulher loura e a professora de matemática são uma e a mesma pessoa”)⁹. Isso está também na semântica do verbo, e é uma das coisas que distinguem *ser* de *apanhar* e *bater*. Ou seja, a semântica de um verbo precisa incluir se ele evoca um esquema correspondente a uma ação, ou um estado, ou um simples evento (*chover*). Isso vai determinar que papéis temáticos (agente, paciente, lugar, instrumento etc.) seus diversos complementos vão exprimir.

Mas uma coisa a semântica do verbo (seu esquema) não conta: qual complemento exprime qual papel temático. Isso tem que ser aprendido para cada verbo. Assim, em português dizemos *a gerente substituiu o relógio de ponto pela borboleta* – e isso significa que antes tínhamos um relógio de ponto, agora temos uma borboleta. Mas em inglês uma frase sintaticamente parecida seria:

inglês: *the manager substituted the time clock for the turnstyle*
português: *a gerente substituiu o relógio de ponto pela borboleta*

⁸ Alguns colegas podem ter objeções quanto a eu chamar o sujeito de “complemento”. Vou evitar essa discussão aqui, para manter as coisas (comparativamente) simples.

⁹ Ou, para os que gostam de fórmulas, “mulher loura” = “professora de matemática”. Note-se que *o menino viu o cachorro* **não** corresponde a “menino” = “cachorro”.

A frase inglesa significa que antes tínhamos uma borboleta (*turnstyle*), e agora temos um relógio de ponto (*time clock*) – exatamente a situação oposta nas duas línguas. Essa distribuição dos papéis temáticos precisa ser aprendida independentemente para o verbo *substituir* (português) e *substitute* (inglês): o esquema é o mesmo, mas sua representação sintática é diferente nas duas línguas.

Podemos agora fazer um cálculo aproximado (só aproximado, porque deixei de lado muita coisa, mas muita coisa mesmo) da quantidade de coisas que temos que conhecer para podermos usar um lexema – como *substituir*, *bater* etc. – corretamente. Alguém que ouve a frase *a menina bateu no cachorro* vai ter que computar coisas como as seguintes:

[a] a forma fonológica mostra que o verbo é *bater* (só esse lexema na língua inclui a forma *bateu*);

[b] está na forma *bateu*, e essa é a forma de pretérito perfeito do lexema *bater*, logo expressa um fato passado;

[c] o complemento tem a preposição *em*, e com essa forma de complemento *bater* evoca o esquema ESPANCAR, ou então CHOCAR-SE ... mas não SOAR (*o sino bateu*);

[d] o sujeito é *a menina*, e com o verbo *bater* (ao contrário de *apanhar*) o sujeito exprime o agente, logo foi a menina que praticou a ação (foi quem espancou ou se chocou, não quem foi espancado ou levou a trombada);

[e] o complemento não-sujeito é *no cachorro*, e com o verbo *bater* esse complemento exprime o paciente¹⁰ – logo foi o cachorro que foi espancado ou levou a trombada (não quem espancou ou se chocou);

Daí, o ouvinte já pode estabelecer que essa frase significa ‘a menina espancou o cachorro’, ou ‘a menina se chocou com o cachorro’. Como se vê, ainda há trabalho pela frente: o que é que a frase significa, exatamente? Isso vai depender de fatores não linguísticos – por exemplo, se eu estiver falando na presença de um cachorro machucado, o mais provável é que *bater* aí signifique ‘espancar’.

¹⁰ Mais precisamente, o não-agente.

Os pontos [a] a [e] acima mostram uma parte do que um usuário da língua precisa saber e fazer para entender frases que incluam o verbo *bater*. Depois, é só multiplicar isso por 20.000, e teremos uma ideia da massa de conhecimentos que qualquer falante de uma língua domina e usa com a maior facilidade.