

Capítulo 16

10 anos de Linguateca - depoimento

Violeta Quental

A Linguateca teve um papel pioneiro ao organizar e disponibilizar um importante acervo de recursos e ferramentas de análise computacional do português e de bibliografia correspondente. Ao longo desses últimos 10 anos, muitos foram os alunos de graduação e pós-graduação da PUC-Rio que, em algum momento de sua pesquisa, usaram os serviços da Linguateca, principalmente o projeto AC/DC (Santos e Sarmento, 2002) e o catálogo de publicações. A facilidade de uso da ferramenta de busca em corpora, o acesso gratuito, a variedade de recursos disponibilizados pelo portal foram e continuam sendo da maior importância para o ensino e para a pesquisa sobre o português. Para nós também, professores e pesquisadores, a Linguateca foi uma iniciativa extremamente útil e inspiradora – e como tal permanece, esperamos que por muito tempo ainda.

Nessa sessão de depoimentos, mais do que relatar as pesquisas feitas no âmbito da Universidade, pretendo comentar a experiência de elaboração do mini-dicionário *Caldas Aulete* (Geiger, 2004), em versão resumida dirigida ao público escolar. O público-alvo do dicionário definiu algumas das políticas de sua elaboração: a cobertura do léxico, de tamanho pequeno — cerca de 20.000 mil verbetes —, com um número de acepções para cada palavra limitado aos seus usos mais frequentes e atuais; a ordem de apresentação das acepções e das propriedades sintático-semânticas dos vocábulos escolhida a partir de pesquisa em corpora; exemplos e abonações breves e em linguagem “simples”.

Essa filosofia guiou a metodologia adotada pelos autores e a busca dos significados em uso, das regências verbais, dos exemplos, foi muito facilitada pelo uso dos recursos do projeto AC/DC, através de consulta por concordância e distribuição em corpora. Outros corpora foram também utilizados, especialmente o acervo literário digitalizado da editora, além da busca na rede, para busca de usos mais informais da língua, tentando evitar o viés fortemente jornalístico do CETEMPúblico (Rocha e Santos, 2000) e do NILC/São Carlos, principais recursos do AC/DC que utilizamos.

Essas escolhas de usar dados de corpus e de avaliar sua frequência já podem ser consideradas tradicionais na literatura lexicográfica. Para dicionários que se dirigem ao público escolar, aos aprendizes de uma língua, Rundell (1999) e Duran e Xatara (2006) propõe que se enfoquem os usos mais comuns, típicos e frequentes. Em nosso caso, para definir a frequência de uma determinada acepção de uma palavra, contávamos apenas com a leitura dos resultados de busca em concordância. Pode-se imaginar que, para palavras muito frequentes, essa leitura resulta em uma estimativa bastante imprecisa das acepções mais comuns. Tentamos evitar esse efeito, nos resultados do NILC/São Carlos - o mais usado por seu tamanho e por representar o português do Brasil - observando exemplos de cadernos variados do jornal: *Brasil*, *Cotidiano*, *Dinheiro*, *Ilustrada*, etc.

Já para a definição da regência verbal preferencial, por exemplo, a ferramenta oferece a busca por distribuição, que garante, em geral, resultados mais exatos. Logicamente, a teoria gramatical que fundamenta a anotação é relevante para a avaliação dos resultados.

Apresento a seguir um exemplo de busca da regência preferencial do verbo *pedir*: vê-se

que, no corpus NILC/São Carlos, a regência transitiva direta é muito mais frequente, se comparada à frequência da regência indireta.

Procura: [lema="pedir"] [func=«ACC.*"].
 Pedido de uma concordância em contexto
 Corpus: NILC/São Carlos anotado v. 4.5
 (3564 ocorrências.)

Procura: [lema="pedir"] [func=«DAT.*"].
 Pedido de uma concordância em contexto
 Corpus: NILC/São Carlos anotado v. 4.5
 (4 ocorrências.)

Observando os exemplos, podemos notar que aquilo que o corpus marca com a função dativa, no entanto, se restringe aos pronomes oblíquos. Vejamos:

Procura: [lema="pedir"] [func=«DAT.*"].
 Pedido de uma concordância em contexto
 Corpus: NILC/São Carlos anotado v. 4.5
 4 ocorrências.

Concordância

Procura: [lema="pedir"] [func=«DAT.*"].

par=118459: Um porteiro veio humildemente **pedir me** que me retirasse, oferecendo me com estúpida e revoltante aparência de benignidade a vil quantia, por que eu pagara o meu bilhete; resisti e furioso disse uma injúria ao mísero porteiro.

par=119045: – **Pedes me** uma segunda luneta mágica que te será fatal como a primeira.

par=119749: De súbito chegou se a mim um mancebo com o semblante abatido, e repassado de dor, e mal podendo falar, expôs me a sua situação que era das mais pungentes sem dúvida, e acabou, **pedindo me** o óbulo da minha caridade para enterrar o filhinho, o filho único, que deixara em casa morto no colo da consternada esposa.

par=119796: Ela tinha parado e olhava me provocadora, insolente, como a **pedir me** jantar...

Assim, o exemplo 1 a seguir, em que aparece o sintagma “ao papa”, não é apresentado se pedimos busca por complemento dativo, embora tenha o mesmo valor semântico de “lhe”.

1. *par=1805*: Henrique IV foi, então, **pedir perdão** ao papa, em Canossa (1077).

Este resultado deriva de classificação de Bick (2000), conforme podemos observar nos exemplos abaixo, de saídas do analisador PALAVRAS:

2.
 - pedir** [pedir] **V INF @IMV**
 - perdão** [perdão] **IN @<ACC**
 - a** [a] <sam-> **PRP @<PIV**
 - o** [o] <artd> <-sam> **DET M S @>N**
 - papa** [papa] **N M S @P<**
3.
 - pedir** [pedir] **V INF @IMV**
 - perdão** [perdão] **IN @<ACC**
 - a** [a] **PRP @<PIV**
 - ele** [ele] **PERS M 3S NOM/PIV @P<**
4.
 - pedir-** [pedir] **V INF @IMV**
 - lhe** [ele] **PERS M/F 3S DAT @<DAT**
 - perdão** [perdão] **IN @ADVL**

Para o dicionário, adotamos a classificação de regência indireta para sintagmas pronominais ou sintagmas nominais preposicionados com a mesma semântica de destinatário ou beneficiário. Para termos certeza, assim, dos resultados de frequência, teríamos de voltar à leitura dos trechos.

Outra questão interessante para a elaboração dos verbetes diz respeito à escolha de exemplos para ilustrar uma acepção. É tradição nos dicionários buscar abonações em textos literários. Nem sempre, para o aprendiz, esta é a melhor escolha, já que muitas das vezes a linguagem literária apresenta dificuldades a mais para quem desconhece o(s) significado(s) de uma palavra. Nem sempre também o exemplo do corpus é o mais esclarecedor.

Duran e Xatara (2006) discutem a questão a partir de 2 exemplos:

1. Faz mal molhar as plantas com sol quente. (Borba, 2002)
2. Molhou os pés no mar. (Houaiss, 2001)

O exemplo (1) foi extraído de corpus e o (2) elaborado por lexicógrafos. Embora não tenha nenhuma crítica a tais exemplos no contexto em que aparecem, se tivéssemos que escolher um para compor um dicionário de português para estrangeiros, o exemplo (2) seria mais adequado que o (1), uma vez que nele um dos complementos do verbo molhar (no mar) apresenta o traço (+ líquido), enquanto no exemplo (2) esse traço não aparece e a construção “molhar com sol quente” poderia gerar dúvidas no aprendiz que não conhece ainda o significado do verbo.

De modo geral, não utilizamos exemplos extraídos dos corpora por esse motivo e também pela necessidade de redigir exemplos curtos, uma restrição de formato de dicionário pequeno, mas nos inspiramos neles para construir nossos próprios exemplos.

Outra questão que surge sempre quando consultamos corpora não analisados sintática e semanticamente é a da homonímia. À época da confecção do dicionário, muitos enganos ainda apareciam nos resultados. A título de brincadeira, uso o exemplo que me foi fornecido pela Profa. Maria Carmelita Dias, de consulta para o verbo “gerar”, por seu lema, que retornou vários trechos sobre o ator Richard Gere. Hoje esta mesma consulta não traria esse resultado, mas ainda aparece como erro o nome de um restaurante famoso – o Gero.

Por fim, a questão mais sensível: a decisão de ir contra a visão diacrônica, talvez contestável em outro contexto que não o de um pequeno dicionário para uso escolar. Um exemplo de decisões que tomamos: o verbo “azeitar” que, nos dicionários mais conhecidos, têm como sua primeira acepção algo como “Temperar com azeite; pôr azeite em”. Esta acepção, nos corpora consultados, não apareceu uma única vez, mas é provavelmente a historicamente mais antiga. Seguindo a filosofia de obedecer ao uso, propusemos como primeira e única acepção a que surge nos corpora: “Passar óleo (em uma engrenagem, máquina etc.) ; LUBRIFICAR.”

Todas as decisões que a elaboração de um dicionário envolve são difíceis e sempre há um contra-exemplo para considerar. Sem dúvida, a contribuição dos recursos de consulta a corpora da Linguateca constituíram uma ajuda inestimável para a discussão lexicográfica.