

RESENHA: *CORPUS LINGUISTICS*, DE TONY MCENERY E ANDREW HARDY

por Guilherme Fromm* (UFU)** e Hadinei Ribeiro Batista (UFRJ)***

McEnery, T.; Hardie, A. (2012) *Corpus linguistics*. Cambridge: University Press.

Esta resenha busca inserir-se no hall de áreas do conhecimento que estabelecem relações estreitas com as Tecnologias da Informação e Comunicação (TICs), principalmente no que tange ao processamento textual e construção de bancos de dados/*corpora* a partir da exploração de ferramentas tecnológicas. A Linguística de *Corpus* (LC), área da linguagem que se dedica à coleta e ao tratamento tecnológico de um conjunto vasto de usos da língua, mostra-se como uma interlocutora em potencial para as TICs.

O livro, editado em 2012, constituído de 9 capítulos, volta-se para a LC do ponto de vista da teoria e da prática. Cada capítulo é dividido em seções e composto de um resumo final com indicações de leituras, atividades e questões para discussão. O livro possui ainda um glossário ao final e a tese principal defendida na obra, a contribuição da LC para diferentes estudos que exploram a linguagem, é sempre acompanhada de exemplos, inclusive de vários estudos desenvolvidos pelo próprio autor.

Os autores definem LC como o estudo de dados da língua em grande escala, viabilizado por ferramentas computacionais capazes de tratar e analisar extensas coleções de textos escritos/transcritos. Para se ter uma visão geral, o capítulo I destina-se a introduzir o leitor à LC, em que se busca definir os diferentes tipos de *corpora* (forma e composição), ou seja, questões relacionadas à etiquetagem e à constituição textual dos *corpora* no sentido de abranger uma ou várias línguas. O capítulo II trata das anotações dos *corpora* para a análise computacional, bem como avalia e apresenta um overview de ferramentas/concordanceadores e de procedimentos estatísticos que podem ser aplicados ao conjunto de dados. No capítulo III, são discutidas algumas questões relacionadas à ética e à legalidade na construção de *corpora*, apontando os desafios práticos da LC quanto à distribuição de dados e ao tratamento ético das pessoas envolvidas no *corpus data*. O capítulo IV destina-se a fazer um overview da tradição da linguística de *corpus* do inglês, considerando as contribuições de pesquisadores de universidades como College London, Lancaster, Birmingham, Catholique de Loivan e Nottingham; além, é claro, dos desenvolvidos nos USA e da Universidade do Norte do Arizona. O capítulo mostra o percurso,

* guifromm@uol.com.br

** Universidade Federal de Uberlândia

*** Universidade Federal do Rio de Janeiro

desde a década de 60, na construção de *corpora* do inglês capazes de explorar as variedades de usos da língua (falada e escrita) bem como possibilitar a construção de modelos de coleta que viabilizassem estudos comparativos tanto sincrônicos quanto diacrônicos, além de servirem de base para a confecção de gramáticas do inglês. O capítulo V centra-se no diálogo da LC com estudos sobre variação e mudança linguística. Primeiramente, tratou-se de apontar os benefícios de alguns *corpora* para abordagens sincrônicas e diacrônicas e, em seguida, foram explorados, de modo geral, os estudos da variação textual/registros, em especial os estudos de Biber.

Julgamos importante interromper aqui a visão geral da obra ora apresentada para discutir, de forma mais acurada, a argumentação desenvolvida no capítulo V. Especificamente quanto à seção 5.5, em que se discute a relação entre *corpora* e sociolinguística variacionista, uma das dificuldades apontadas pelos autores sobre a aplicação da LC a essa corrente teórica é o interesse da sociolinguística na variação da pronúncia e a falta de transcrição fonética em *corpora* de fala extensos. Outro ponto que os autores consideram primordial é o interesse dos sociolinguistas em fatores sociais relacionados aos falantes da amostra. Os *corpora*, considerando a metodologia da LC, centram-se na linguagem dos registros e não na linguagem de um grupo particular de pessoas. Como ressaltam os autores, a tendência atual é a de que vários *corpora*, como o CANCODE, BNC, NECTE e outros, sejam desenvolvidos de forma a atender melhor as necessidades da corrente teórica em tese, considerando a preocupação na transcrição de conversações e na inserção de fatores externos ou sociais. Embora os autores apontem o desencontro entre essas duas abordagens, é indiscutível o ponto de convergência entre a Sociolinguística e a LC no que tange ao interesse pela coleta e tratamento de dados da língua em uso.

Assinala-se, ainda, que a limitação quanto à transcrição fonética está sendo superada, embora os estudos sociolinguísticos variacionistas não se restrinjam a esse nível de análise. Outra preocupação que merece considerável atenção é o mapeamento de fatores sociais (ou extralinguísticos). Os *corpora* citados apresentam um controle bastante amplo de traços identitários dos falantes, tradicional nas análises sociolinguísticas, que se mostram, no entanto, pouco eficientes na explicação da variação e mudança linguística de estudos de base quantitativa. Nessa convergência, sugere-se um refinamento dos indicadores sociais, bem como um levantamento de dados a partir de comunicação virtual, seja através de sites ou de aplicativos de celulares, uma vez que esse viés da interação tornou-se a ‘febre’ do momento.

Voltando a atenção para o capítulo VI, este explora os estudos (em especial ‘Firthianos’ e ‘neoFirthianos’) sobre os ‘colocados’, ou seja, a ideia de que o significado da palavra é denotado a partir de sua combinação com outras com as quais co-ocorre. Tal discussão levou ao debate sobre considerar a LC como teoria ou como metodologia. A noção de *corpus-based* e *corpus-driven*, várias vezes discutida na obra, é retomada com essa intenção, uma vez que os dados de *corpora* podem contrariar conceitos preestabelecidos. O Capítulo VII aborda a relação da LC com outras áreas da linguística, tanto estrutural quanto cognitiva. A argumentação principal baseia-se em demonstrar que a LC, por prover uma grande quantidade de dados, mostra-se eficaz em explicar determinados fenômenos dificilmente capturados por um conjunto restrito de dados que normalmente baseiam

estudos sobre noções mais cognitivas como metáforas e de estrutura sintática. No capítulo VIII, o foco é na convergência de estudos funcionalistas, psicolinguísticos e LC. A intenção foi desenvolver uma argumentação em favor do link desses dois campos teóricos com a Linguística de *Corpus* com o objetivo de buscar melhores esclarecimentos para a natureza da linguagem humana.

Para além da discussão sobre a independência teórica da LC, vários estudos mostram que sua metodologia auxilia sobremaneira as investigações dessas áreas na medida em que fornece grande quantidade de dados que permite mapear, com maior clareza, o funcionamento de estruturas sintáticas, além de dar suporte ao processamento cognitivo da linguagem. De fato, mesmo que estudos centrados na psicolinguística partam da elaboração de enunciados fictícios para resgatar o processamento cognitivo do falante, os padrões linguísticos mais ou menos recorrentes no uso espontâneo da língua podem ser fornecidos pela LC como parâmetros eficazes para a metodologia empregada pela corrente teórica.

No capítulo IX e último, é retomada a relação da LC com outras correntes linguísticas bem como em relação a outras áreas das ciências sociais, que notadamente exploram a linguagem em seus estudos, e argumentou-se, novamente, em favor da expansão da LC no sentido de responder a novas questões, de estabelecer novas integrações metodológicas bem como de gerar novas disciplinas acadêmicas, dada sua diversidade na exploração de *corpus data*.

Por fim, o livro vem ao encontro das necessidades atuais inovadoras, dada a observação de que o abismo entre a Sociolinguística e a LC está cada vez mais estreito. Cabe assinalar, ainda, que o livro alerta para os problemas já sobejamente conhecidos dos *corpora* tradicionais. Note-se que, vencidos os impasses ora discutidos, a metodologia da LC pode se mostrar de fundamental importância na construção e tratamento de banco de dados altamente eficazes na explicação de fenômenos varáveis, assim como tem se prestado a vários outros campos solidamente reconhecidos no âmbito acadêmico, com exceção, é claro, daqueles que se baseiam em dados intrinsecamente introspectivos.

Recebido em: 30/08/2015

Aceito em: 08/09/2015