

Resenha

BERBER SARDINHA, T. *Pesquisa em Lingüística de Corpus com WordSmith Tools*.
Campinas: Mercado de Letras, 2009, 272p.

RAPHAEL MARCO OLIVEIRA CARNEIRO
Graduando em Letras pela Universidade Federal de Uberlândia (UFU).
e-mail: raphael.olic@gmail.com.

Antes de partirmos para a descrição da obra em si, o maior expoente da Linguística de *Corpus* no Brasil, Tony Berber Sardinha, no clássico *Linguística de Corpus* (2004) diz que

a Linguística de Corpus ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados lingüísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador (BERBER SARDINHA, 2004, p. 3).

A Linguística de *Corpus*, assim concebida, pode beneficiar diversas áreas, não só em pesquisas, como no ensino também. Como exemplo, podemos citar outra produção de grande contribuição para a divulgação dos estudos que se utilizam de *corpora*, a saber, o livro *Corpora no Ensino de Línguas Estrangeiras*, organizado por Vander Viana e Stella E. O. Tagnin (2010). Nesse mesmo livro, várias pesquisas que se utilizaram da ferramenta *WordSmith Tools* são apresentadas, e por isso torna-se tão importante conhecer o funcionamento dessa ferramenta tão versátil. O livro em questão, *Pesquisa em Linguística de Corpus com WordSmith Tools*, de Tony Berber Sardinha torna-se, portanto, fundamental para o seu melhor entendimento.

O livro se propõe a traçar um panorama de possíveis pesquisas realizadas com o *WordSmith Tools*, bem como fazer uma explicação detalhada de seu funcionamento, para que possa facilitar o seu uso por aqueles que realizam pesquisas utilizando-se desse software. Basicamente, o *WordSmith Tools*, criado em 1996 por Mike Scott, da Universidade de Liverpool, Reino Unido, é um conjunto de ferramentas integradas utilizadas para análise lingüística. Ele permite fazer análises baseadas nas frequências e co-ocorrências de palavras em *corpora*. Desse modo, no primeiro capítulo o autor introduz o software em questão bem como alguns aspectos gerais referentes a ele.

No segundo capítulo, foca-se o uso da ferramenta em uma área bem promissora: o ensino de língua estrangeira. Como exemplo, o autor descreve o processo utiliza-

do na análise de um *corpus* composto por um texto em língua estrangeira, demonstrando como o programa *WordSmith Tools* pode ser utilizado no ensino de vocabulário por meio de colocações. A partir de exemplos retirados do *corpus* em estudo, também são levantadas questões relacionadas à importância do uso de textos autênticos no ensino de línguas.

Em sequência, o autor introduz alguns aspectos teóricos sobre a noção de gênero textual e as possibilidades de uso do *WordSmith Tools* na análise de gêneros. Em seguida, uma análise é feita sobre o gênero código de ética, revelando dados do *corpus* analisado e os passos metodológicos utilizados para a análise. Desse modo, por meio do programa em questão verificaram-se padrões centrais do gênero código de ética e como esses padrões normatizam a vida em sociedade.

No quarto capítulo, enfoca-se o uso do programa de análise lexical no estudo de metáforas conceituais. Inicialmente, é feita uma introdução sobre os estudos da metáfora na perspectiva cognitiva e do uso. O autor apresenta posteriormente, os recursos utilizados para a análise do *corpus* como as fontes de compilação do *corpus* e o tamanho do mesmo. Após as devidas análises, verifica-se a utilidade do *WordSmith Tools* na listagem de palavras-chave que contêm potencial metafórico, bem como a análise das mesmas a partir de linhas de concordâncias.

No capítulo seguinte, o autor apresenta os Universais de Tradução propostos por Baker (1993), a saber, simplificação, explicitação, normalização e estabilização. Por meio de um *corpus* paralelo composto por originais e traduções de 75 resumos de dissertações escritos em português e suas traduções para o inglês, retirados de teses de Linguística Aplicada defendidas no PPG em Linguística Aplicada e Estudos da Linguagem, da PUC-SP, o *corpus* é analisado de acordo com os três primeiros universais citados anteriormente. Em resumo, nesse tipo de análise, o *WordSmith Tools* foi mais adequado para os estudos de simplificação e explicitação e menos adequado na investigação de aspectos qualitativos como os da normalização.

Focando uma área ainda não muito conhecida no Brasil, o autor apresenta alguns aspectos introdutórios dos estudos realizados no âmbito da Linguística Forense. Trata-se de uma área que se preocupa com a análise da linguagem jurídica, a fim de fornecer evidências linguísticas em processos judiciais. Após os passos da análise exemplificada no capítulo, chega-se à conclusão de que os procedimentos descritos por meio do programa *WordSmith Tools* podem ser empregados em casos judiciais, visto que é possível desvendar casos em que se quer comprovar a autoria de determinado texto ou se essa autoria é colocada em dúvida.

Nos três capítulos seguintes, o autor apresenta uma vasta descrição de como utilizar cada ferramenta do programa *WordSmith Tools* versão 3. Sabe-se que existem versões atualizadas do programa¹, porém segundo o próprio autor, o conhecimento do funcionamento da terceira versão é suficiente para saber lidar com as versões posteriores. Basicamente, as ferramentas são definidas de acordo com suas respectivas funções:

¹ A versão 6.0 do *WordSmith Tools* (a atual), bem como as versões anteriores estão disponíveis para download no site www.lexically.net.

- WordList: produz listas de palavra contendo todas as palavras do arquivo ou arquivos selecionados, elencadas em conjunto com suas frequências absolutas e percentuais. Também compara listas, criando listas de consistência, onde se informa em quantas listas cada palavra aparece.
- Concord: realiza concordâncias, ou listagens de uma palavra específica (o “nódulo”, *node word* ou *search word*) juntamente com parte do texto onde ocorreu. Oferece também listas de colocados, isto é, palavras que ocorrem perto do nóculo.
- KeyWords: extrai palavras de uma lista cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras num outro corpus (de referência). Calcula também palavras-chave, que são chave em vários textos (BERBER SARDINHA, 2009, p. 9).

No capítulo dez, o autor também faz uma descrição dos utilitários presentes em cada ferramenta. São eles:

- File Manager: abre uma janela para gerenciamento de arquivos.
- Splitter: permite dividir um arquivo em vários menores.
- Text Converter: oferece várias funções para o pré-processamento de textos, como a substituição de palavras, partes de palavras ou partes de textos, simultaneamente num conjunto de arquivos, a renomeação em massa de arquivos, e a mudança de pasta (diretório) de arquivos que apresentem certas características.
- Viewer & Aligner: fornece meios para a visualização de textos e para o alinhamento (combinação) de dois textos num só (Ibid., p. 10).

Após as referências bibliográficas, o autor encerra o livro com uma lista de quatro *websites* que apresentam outros trabalhos, estudos e pesquisas que se valeram do *WordSmith Tools*.

A partir da obra assim apresentada, percebe-se que se trata de um volume ímpar, que além de descrever detalhadamente uma das ferramentas mais úteis e versáteis para análise linguística de *corpora*, discute os benefícios do uso do *WordSmith Tools* para várias áreas de estudo. Com objetividade e clareza, o autor demonstra pleno conhecimento e experiência na área trazendo de forma clara, simples e didática o que se tem feito e o que é possível fazer a partir dos estudos de *corpora*, embora as áreas não se restrinjam às discutidas pelo autor. Além disso, o leitor certamente encontrará na obra o embasamento teórico necessário que permitirá a devida compreensão dos temas contemplados ao longo da mesma.

Em resumo, além de demonstrar o incrível potencial do *WordSmith Tools* na investigação de questões de áreas da Linguística e da Linguística Aplicada, o livro apresenta múltiplas perspectivas para aqueles que pretendem iniciar, conhecer e desenvolver pesquisas nesse campo tão fascinante sobre os estudos da linguagem, a Linguística de *Corpus*.

Referências

BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Manole, 2004.

BERBER SARDINHA, T. *Pesquisa em Lingüística de Corpus com WordSmith Tools*. Campinas: Mercado de Letras, 2009.

TAGNIN, S. E. O.; VIANA, V. (org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010.

Resenha recebida em 27/07/2011

Aceito para publicação em 06/03/2012