

SARDINHA, Tony Berber. 2004. *Lingüística de Corpus*. Barueri S.P: Editora Manole, xvi + 410 p.

Ao receber este pioneiríssimo volume de autoria de um dos mais competentes lingüistas aplicados brasileiros atuante na PUC-SP, recordei que em 1955 durante minha primeira ida à University of Michigan tomei conhecimento de uma obra de 1940, precursora do que muitos anos mais tarde viria a chamar-se *Lingüística de Corpus*. Refiro-me à *American English Grammar* do estruturalista Charles C. Fries de quem tive o privilégio de ser aluno-ouvinte em Ann Arbor. Quantas mudanças ocorreram de lá para cá: Fries adotou uma abordagem descritiva em que havia um lugarzinho ao sol para comentários de natureza quantitativa. Assim ao relatar as preferências de uso (a amostra era constituída por correspondência de militares americanos para seus familiares) aquele lingüista oferecia dados sobre variantes mais freqüentes. Assim ficou evidenciado que *Will* (sob forma contraída: *I'll*) ganhava da concorrente *Shall* desmistificando-se regras prescritivas até então encontradas em gramáticas.

Em 1999 com o aparecimento da *Longman Grammar of Spoken and Written English* de Douglas Biber et alii consolida-se uma Lingüística verdadeiramente quantitativa e Computadorizada. As fichas de dados datilografados usados por Fries deram lugar a corpora computadorizados armazenados em Bancos de um número crescente de línguas. Mas isso é outra história que por sinal está bem contada no oportuno livro de Sardinha. Aliás alternativamente pode-se começar a leitura de *Lingüística de Corpus* pelo capítulo 10 (o mais extenso: 46 páginas) intitulado *Lingüística de Corpus e Lingüística Aplicada: O Ensino de Língua Estrangeira*. Aos não-iniciados nesse produtivo e inspirador ramo da Lingüística o processamento do referido capítulo proporcionará uma visão do que se está realizando na área aplicativa. Assim os leitores conhecerão um pouco sobre *Corpora Pré-computador e Ensino de Línguas Usos de corpora no ensino*

de língua materna ou estrangeira, marcos importantes na história da Lingüística de Corpus Descrição da Linguagem de Aprendizizes (*learner corpora*), Concordâncias (Listagem de contextos em que itens lingüísticos ocorrem) no Ensino Aprendizado movido por dados (*Data driven learning*).

Este quase enciclopédico volume contém um Prefácio (3 p.), 11 capítulos de extensão Variável (de 14 a 46 páginas), Referências bibliográficas (20 p.), 5 Anexos e Índice Remissivo (6 p.). Os títulos de alguns capítulos podem dar uma idéia da abrangência e diversidade dos temas expostos em Português bem redigido e, onde necessário, também visual (figuras, reproduções de telas de computador, quadros): Visão geral da Lingüística de Corpus; Coleta armazenamento e pré-processamento de corpora; Um corpus de Português especializado na linguagem profissional; Frequência de palavras da língua portuguesa segundo o Banco de Português; Lingüística de Corpus e Tradução; Estudo da Variação com Lingüística de Corpus. Quase todos os capítulos seguem a mesma organização: Introdução diversas seções e Comentários finais. Nos utilíssimos Apêndices encontramos as 3.000 palavras mais freqüentes do Banco de Português (um dicionário inédito). Nove corpora na Internet (Português inglês alemão espanhol), 25 websites mencionados na obra Amostra de concordâncias do Banco de Português, Dendogramas para análise da padronização de *só*.

O Índice Remissivo mostra a riqueza conceitual-terminológica com que Sardinha trabalha e nomes de importantes centros de pesquisa na área. Poderia continuar a destacar qualidades neste primeiro livro em Lingüística de Corpus entre nós, mas esse privilégio e proveito cabe aos leitores das mais diversas áreas dentre as quais Letras Lingüística Tradução Comunicação Educação Psicologia Ciências da Computação Lexicografia e Terminologia. Em suma um livro que honra a Tradição Brasileira em Ciência da Linguagem e contribui significativamente para o desenvolvimento de um modo de fazer lingüística quantitativamente revelador e fascinante. Congratulações ao autor e à sua Editora pelo auspicioso lançamento.

Por/by: FRANCISCO GOMES DE MATOS  
(Letras/CAC/UFPE)  
E-mail: fcfgm@hotlink.com.br