

PESQUISA COM CORPUS: CRITÉRIOS DE SELEÇÃO E FERRAMENTAS DE ANÁLISE

André Luiz Aguiar de Oliveira, Gessyca Mayara Rodrigues Alves, Renato Souza Lopes, Maria Claudete Lima

Pesquisas na Linguística apoiadas em modelos baseados no uso exigem a seleção de corpus representativo que sirva aos objetivos do estudo. Embora a representatividade de algo vasto como a própria língua seja polêmica, a linguística de corpus propõe critérios que ajudam a seleção de amostra significativa a depender da variável estudada, que, sendo o verbo, precisa de, no mínimo, 69 mil palavras (SARDINHA, 2000). Por outro lado, o manuseio de um corpus extenso toma tempo e sujeita-se a erros evitáveis por meio de ferramentas de coleta e análise. A presente pesquisa de base cognitivo-funcional, cujo objetivo é identificar os padrões construcionais de FICAR, analisa cartas pessoais dos séculos XIX e XX, advindas de dois acervos, por pressupor que a escrita pessoal se aproxima mais da fala. A 1ª. fase da pesquisa constituiu-se das etapas: seleção do corpus e coleta das ocorrências. A seleção das cartas seguiu essa metodologia: (1) contagem do nº de palavras de cada carta, de modo à amostra ser representativa; (2) distribuição equitativa de cada século para evitar enviesamento dos resultados; (3) seleção de cartas de cunho pessoal e de autores brasileiros; e (4) escolha de cartas em grafia modernizada para facilitar a identificação das frases. Assim, formou-se um corpus de 49 cartas do séc. XX e 102 do séc. XIX, somando 85.506 palavras. Para coleta dos dados por meio do AntConc, programa que analisa textos, fizeram-se dois procedimentos: inserção do corpus em .txt e busca de concordância com caractere-coringa para localizar todas as formas do lema FICAR. Como resultado, encontraram-se 183 ocorrências, 57,92% do séc. XX e 42,08% do séc. XIX, as quais serão depois analisadas no programa estatístico PSPP. A facilidade e a rapidez da coleta dos dados em um corpus extenso, possível graças ao AntConc, ilustra a relevância do uso de ferramentas na pesquisa linguística o que não dispensa, obviamente, o papel do pesquisador na seleção do corpus e análise das ocorrências coletadas.

Palavras-chave: modelos baseados no uso. AntConc. construções com ficar. linguística de corpus.