

CONSTITUIÇÃO DE UM CORPUS DO PORTUGUÊS FALADO NOS PAÍSES DE LÍNGUA OFICIAL PORTUGUESA (PALOP) E NO TIMOR LESTE

V Encontro de Iniciação Acadêmica

Pedro Guilherme da Silva Junior, Jhuliana Maria Silva de Souza, Lorena Rodrigues, Maria
Elias Soares

A linguística de corpus, de acordo com Sardinha (2004), vale-se da “abordagem que se ocupa da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente”, possibilitando consequentemente, a pesquisa da língua e da diversidade linguística. Ainda segundo Sardinha (2004), a linguística de corpus dedica-se ao estudo da linguagem através de evidências extraídas do computador, de modo que os dados devem encontrar-se em formato eletrônico. Com base nessa abordagem, o grupo de pesquisa PROFALA desenvolve o projeto de constituição de um corpus do português falado nos países africanos de língua oficial portuguesa (os PALOP): Angola, Cabo Verde, Guiné Bissau, Moçambique, São Tomé e Príncipe, e também na Ásia, no Timor-Leste. Assim, este trabalho tem como objetivo apresentar a metodologia utilizada pelo projeto, que trabalha na construção de um banco de dados do português falado nesses países, constituindo um acervo digital de áudios e transcrições ortográficas de entrevistas realizadas com informantes, de ambos os sexos, oriundos dos países citados, divididos em dois grupos: os que estão, há menos de seis meses ou há mais de seis meses no Brasil. Visto isso, apresentamos os resultados parciais, com foco nas transcrições das entrevistas, realizadas a partir de um questionário adaptado do Atlas linguístico do Brasil (ALIB), visando apreender informações sobre a língua, em diferentes níveis: fonético-fonológico, semântico-lexical e morfossintático. Atualmente, o projeto está com cerca de 83 áudios transcritos, dos 120 que já estão disponíveis no site do PROFALA. A formação do corpus permitirá que pesquisadores tenham acesso a uma gama de dados morfológicos, sintáticos, fonológicos e lexicais de diferentes variedades do português.

Palavras-chave: Língua portuguesa. PALOP. Corpus.