

Catherine Anne Lonngreen Sampaio - Construção e exploração de um corpus eletrônico de linguagem bilíngüe infantil

Orientadora: Profa. Dra. Paula Lenz Costa Lima

Esta dissertação descreve o processo de construção e exploração de um corpus eletrônico de linguagem bilíngüe infantil através do sistema de transcrição e análise do projeto do CHILDES (Child Language Data Exchange System), um sistema de troca de dados de linguagem infantil originalmente desenvolvido na Universidade de Carnegie nos Estados Unidos. A pesquisa enquadra-se na área da Linguística de Corpus, que pode ser caracterizada pela investigação de fenômenos lingüísticos baseada em exemplos reais de linguagem, armazenados em corpora eletrônicos, compilados de forma sistemática e segundo determinados critérios. O corpus descrito no presente trabalho é composto de transcrições da linguagem falada de duas crianças bilíngües que estão aprendendo português e inglês simultaneamente. Os dados foram coletados através de gravações durante um período de aproximadamente dois anos e transcritos seguindo as convenções estabelecidas pelo CHILDES. Fora as convenções mínimas, foram inseridos códigos específicos nas transcrições que permitissem a análise dos fenômenos a serem observados. As decisões tomadas nesse sentido são discutidas a cada passo. Uma vez construído o corpus, são realizadas várias análises, com ferramentas do próprio CHILDES, para ilustrar, de maneira geral, a potencialidade de um corpus transcrito de acordo com as convenções mínimas do sistema. Devido à inserção de códigos específicos para a separação das duas línguas, foi possível realizar análises mais específicas que permitissem a investigação do uso que as duas crianças fazem de mudança de código (Codeswitching), tanto de uma perspectiva estrutural quanto de uma perspectiva pragmática. No final do trabalho foram dadas algumas sugestões para a construção e exploração de corpora de diferentes naturezas, demonstrando o alcance potencial de uma investigação da linguagem fundamentada na Linguística de Corpus.