



SEGMENTAÇÃO E ANÁLISE ESTATÍSTICA EXPLORATÓRIA DE CORPORA DE REGISTROS ORAIS DE PORTUGUÊS EUROPEU MODERNO E PORTUGUÊS BRASILEIRO

Janaisa Martins Viscardi (Bolsista FAPESP) e Prof. Dr. Antonio Galves (Orientador), Instituto de Estudos da Linguagem - IEL, UNICAMP

Este projeto está inserido no projeto temático Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística (processo FAPESP 98/3382-0). O trabalho é feito a partir da hipótese de que é possível classificar as línguas de acordo com padrões rítmicos. O objetivo do projeto é preparar um corpus anotado, fazer uma análise estatística exploratória desses dados e, a partir dessa análise, verificar se há evidências de que o Português Brasileiro e o Português Europeu (doravante PB e PE, respectivamente) são línguas que pertencem a classes rítmicas distintas, com base na metodologia de Ramus et al (1999). O trabalho de Ramus é bastante importante por ser o primeiro a apresentar evidências acústicas sobre a existência das classes rítmicas. Até o presente momento, o trabalho consistiu na segmentação manual dos sinais acústicos em segmentos vocálicos e consonantais e na conseqüente análise estatística desses dados. No entanto, interessa ao projeto, a segmentação automática dos intervalos para que se possa analisar um grande número de dados em tempo menor. O programa que automatizará a segmentação dos corpora está em fase de desenvolvimento, tendo atualmente uma taxa de correção da ordem de 75%. Seu aperfeiçoamento depende de um trabalho de correção manual das saídas do programa e de identificação das características dos segmentos problemáticos. Participo ativamente desta etapa do projeto. Com base na análise estatística exploratória desse grande corpus será possível elaborar um gráfico que apresente a posição do PB e o PE em relação às demais línguas já estudadas e enquadradas no gráfico elaborado por Ramus et al.

Grupo Rítmico - Português Brasileiro - Português Europeu