

MODELAGEM ESTOCÁSTICA DOS PADRÕES PROSÓDICOS E DA MUDANÇA SINTÁTICA NO PORTUGUÊS<sup>1</sup>: A.P. Alencar<sup>2</sup>, A. Galves (orientador)  
Departamento de Estatística - IME/USP

Na primeira parte do trabalho, o principal objetivo é identificar os processos estocásticos associados aos contornos acentuais no Português Brasileiro (PB) e Europeu (PE). Para isso, consideramos as distâncias entre elementos acentuados em amostras de transcrições fonéticas codificadas no Instituto de Estudos da Linguagem (IEL) na UNICAMP. A sequência destas distâncias foi modelada como uma cadeia de Markov de alta ordem. Utilizamos os critérios: Akaike, de Bayes e de mínima entropia, para a escolha da ordem da cadeia e estimamos as probabilidades de transição pelo método da máxima verossimilhança. Este trabalho resultou na elaboração do programa PROSPAT, em linguagem C, cujo objetivo é estimar as probabilidades de transição entre as distâncias e os coeficientes AIC e BIC. Os resultados deste trabalho estão no artigo "Markovian modeling of the stress contours of Brazilian and European Portuguese" aceito pela REBRAPE.

A segunda parte do trabalho consiste na construção de um corpus acústico anotado de língua falada de PB e PE. Os dados foram gravados e analisados utilizando-se o programa WinPitch desenvolvido por Philippe Martin da Universidade de Toronto. Este programa fornece gráficos de pressão, intensidade, frequência fundamental e espectrograma em tempo real. A partir destes gráficos, o trabalho de anotação e análise dos dados é realizado.

Além disso, está sendo construído o corpus Tycho Brahe, que é um corpus anotado sintaticamente do Português histórico. Para a implantação do corpus realizamos a adaptação de rotinas em linguagem Perl que permitem a busca de expressões regulares no corpus.

<sup>1</sup> Projeto financiado pela FAPESP e PRONEX; <sup>2</sup> Bolsista da Pró-Reitoria da USP.