



E0535

**ALGORITMOS DE ESTIMAÇÃO PARA CADEIAS DE MARKOV DE ALCANCE VARIÁVEL-
APLICAÇÕES A DETECÇÃO DO RITMO EM TEXTOS ESCRITOS**

Maíra Blumer Faretto (Bolsista PIBIC/CNPq e IC CNPq) e Profa. Dra. Nancy Lopes Garcia (Orientadora), Instituto de Matemática, Estatística e Computação Científica - IMECC, UNICAMP

Estudamos textos jornalísticos do século XX dos jornais Folha de São Paulo e O Público e textos históricos de escritores nascidos em Portugal entre o século XVI e XIX. Nestes textos marcamos todas as sílabas que são tônicas ou átonas. Usando Cadeias de Markov de Alcance Variável estimamos as árvores de contexto e as probabilidades de transição modelando cada texto. Este método nos permite discriminar completamente entre Português Europeu Moderno e Português Brasileiro. Para os textos clássicos, observamos diferenças em relação a ambas às línguas modernas, bem como uma maior variação nos padrões atestados. Neste projeto utilizaremos as chamadas Cadeias de Markov de Alcance Variável (VLMC, do inglês "Variable Length Markov Chain") como modelos para seqüências codificadas de sílabas em textos escritos. Nosso objetivo é comparar diversos algoritmos de estimação e/ou modificações no algoritmo de contexto de forma a verificar qual melhor classifica os textos modernos. E classificar os textos históricos para verificar quando houve a mudança e separação do Português Brasileiro do Português Europeu Moderno.

Detecção do ritmo - Alcance variável - Ritmos