

PROPOSTA DE SUPORTE COMPUTACIONAL À INSERÇÃO DE TRANSCRIÇÃO FONÉTICA EM LEMAS DE VERBETES DE DICIONÁRIOS DO PORTUGUÊS DO BRASIL

doutoranda: Vanessa Marquifável

Instituto de Biociências, Letras e Ciências

Exatas de São José do Rio Preto (IBILCE-UNESP)

linha de pesquisa: Tratamento do Léxico da Língua Geral

início da pesquisa: março de 2011

previsão de término: março de 2015

marquifavel@gmail.com

orientadora: Dra. Claudia Zavaglia

Instituto de Biociências, Letras e Ciências

Exatas de São José do Rio Preto (IBILCE-UNESP)

Departamento de Letras Modernas

linha de pesquisa: Tratamento do Léxico da Língua Geral

zavaglia@ibilce.unesp.br

Este resumo dedica-se à apresentação de um projeto de doutorado que tem como objetivo a construção de um ambiente *Web* gratuito com vistas à transcrição fonética automática em lemas de verbetes de dicionários do Português do Brasil. Numa atualidade motivada para o uso de várias aplicações tecnológicas ativadas pela fala, esse tipo de conversão de letras em som tem sido alvo de muitos estudos tanto em Ciência da Computação quanto de Engenharia Elétrica. Como bem cita Veiga *et al* (2011, p.39), apesar de ser já sólida e madura a investigação na área, ainda não se encontra totalmente esgotado, como se pode comprovar quer pelas taxas de erro publicadas nos artigos da área, quer pelos erros de conversão que persistem nos atuais sistemas existentes no mercado. Portanto, temos como meta em nosso projeto construir um ambiente computacional para transcrever foneticamente as unidades lexicais (ULs) que fazem parte da cabeça do verbete de dicionários mono ou plurilíngues, do Português do Brasil, e cuja disponibilização será gratuita, via *Web*. Como um de nossos objetivos gerais temos a intenção de mostrar de que forma um conhecimento fornecido pela Linguística Teórica pode ser integrado em sistemas automáticos de processamento de português brasileiro, nomeadamente na vertente da conversão letra-som, para produzir melhores resultados nessa tarefa. Em relação ao aparato teórico-metodológico, adotamos a Teoria da Variação ou Sociolinguística Variacionista, proposta por William Labov em 1966, cujo foco é a investigação da língua no contexto social, ou seja, enquanto falada por uma comunidade ou grupo social. Entre outras, foi escolhida devido ao fato de incorporar técnicas computacionais no tratamento de grandes conjuntos de dados, possibilitando ao linguista descrever, com base estatística, a variabilidade de um fenômeno, assim como os fatores que o influenciam. Tal aparato nos auxiliará a compreender melhor nosso objeto de estudo, ou seja, o comportamento fonético/fonológico das vogais e consoantes da variedade linguística praticada em São José do Rio Preto, principalmente das vogais mediais pretônicas <e> e <o>, uma vez que oferece uma metodologia quantitativa que permite extrair regularidades a partir de dados empíricos. A lacuna de conhecimento a ser preenchida com este nosso trabalho está relacionada a um dos problemas ainda com solução insatisfatória apresentado na literatura: problemas na transcrição fonética, ou seja, na conversão

letra-som de /e/, /o/ pretônicos. Esperamos, portanto, que a adição de uma maior quantidade de variáveis linguísticas de base sociovariacionista às regras de transcrição automática das vogais mediais pretônicas <e> e <o> poderão produzir resultados melhores que os citados na literatura.

Referências Bibliográficas

Veiga, A., Candeias, S., Perdigão, F. 2011. "Conversão de Grafemas para Fonemas em Português Europeu – Abordagem Híbrida com Modelos Probabilísticos e Regras Fonológicas". J. J. Almeida, A. Simões, X. Guinovart (eds.). *LinguaMÁTICA*, Revista para o Processamento Automático das Línguas Ibéricas, vol 3, nº2: 39-51. (ISSN: 1647-0818)