

Proposta de Formação e Anotação Padrão de *Corpus* Oral para Estudo Fonético Acústico de Vogais do Português Brasileiro

Regina CRUZ (UFPA/CNPq)
Socorro CAMPOS (UFPA)
Raquel COSTA (UFPA)
Carlos CAVALCANTE (Aluno de Mestrado CML/UFPA)

RESUMO

O presente trabalho tem como objetivo principal apresentar como estão sendo organizados, tratados e anotados os *corpora* formados para o estudo acústico do sistema vocálico do português falado na Amazônia Paraense. Trata-se de um estudo vinculado ao Projeto Nacional PROBRAVO e tem como objetivo principal analisar acusticamente o sistema vocálico átono do Português Brasileiro falado na Amazônia Paraense. O projeto já conta com *corpora* das quatro variedades locais - Belém (CRUZ 2011; REIS, 2011; REIS *et al* 2011), Cametá (COSTA, em andamento), Mocajuba (CAMPOS, em andamento) e Bragança (CAVALCANTE, em andamento) -, formados previamente para um projeto anterior executado em nível de estágio pos-doutoral (CRUZ, 2011), quando se estabeleceu a metodologia aqui aplicada. Portanto, o projeto dispõe de um material para análise acústica gravado com 108 informantes, sendo 18 nativos da capital paraense, 36 da cidade de Cametá (18 da zona urbana e 18 da zona rural), 36 da cidade de Mocajuba (18 da zona urbana e 18 da zona rural) e 18 da cidade de Bragança, todos estratificados socialmente, como descrito acima, em sexo, escolaridade e faixa etária.

Cada informante participou de dois protocolos diferentes de coleta de dados: a) um teste de projeção de imagens e b) a leitura em voz alta de um texto.

Para a formação do *corpus* final, foram utilizados na coleta de dados 74 (setenta e quatro) vocábulos, contendo os mesmos contextos selecionados pelo programa Varbrul como favorecedores para explicar o fenômeno de variação das vogais alvos em suas descrições variacionistas prévias (RODRIGUES & ARAÚJO 2007, DIAS *et al* 2007, CAMPOS 2008, CRUZ *et al* 2008, CASSIQUE *et al* 2009, SOUSA 2010).

Dentre estes 74 (setenta e quatro), utilizaram-se no texto instrumento de coleta de fala lida 51 (cinquenta e um) vocábulos. Para formar o *corpus* do teste de imagens, utilizamos quase todos os vocábulos presentes no texto instrumento de coleta de fala lida, retirando-se apenas quatro: 'querida', 'sobrinho', 'senhoras' e 'escravo'; e acrescentamos 23 (vinte e três) novos vocábulos.

Aos 108 (cento e oito) sujeitos foi solicitado primeiro que dissessem o que viam logo depois que cada imagem era projetada. A sequência das 70 (setenta) imagens foi projetada duas vezes a

cada participante da pesquisa. A nenhum deles foi revelado de imediato os verdadeiros objetivos da pesquisa. Eles tomaram conhecimento dos reais objetivos da pesquisa, apenas quando lhes foi solicitado a assinatura no termo de consentimento esclarecido do projeto.

Uma vez aplicado o teste de imagens, foi solicitada a leitura do texto acima mencionado. Foi dado a cada informante um tempo de 10 minutos para familiarização do mesmo com o tema do texto – futebol – de modo que a leitura fosse a mais natural possível.

Uma vez as gravações obtidas em trabalho de campo, o tratamento realizado consiste em seis etapas bem estabelecidas. Primeiramente as gravações originais são segmentadas em 6 (seis) níveis no programa PRAAT: enunciado, indicação da palavra alvo, duração da palavra, sílaba, transcrição fonológica da vogal, transcrição fonética da vogal. Em seguida procede-se à extração do vocábulo alvo e seu isolamento em um arquivo individual, preservando seu sinal sonoro e sua transcrição no TextGrid. Para tal utiliza-se um recurso do próprio PRAAT (Extract Selected Sound e Extract Selected TextGrid). Passa-se então a codificação do vocábulo contendo as informações sociais do locutor, do tipo de situação de fala gravado (teste de imagens ou fala lida) e o próprio vocábulo em si. O código criado é utilizado para nomear o novo arquivo criado.

De posse dos vocábulos em arquivos isolados, procede-se ao levantamento e registro das variantes de vogais médias pretônicas identificadas no *corpus* por locutor em um formulário controle padrão. São então tomadas medidas de F1, F2, F0 e da duração das vogais médias alvo na posição pretônica na parte central das vogais. A tomada de medidas acústicas é feita também com o auxílio do programa PRAAT. São também realizadas medidas das vogais tônicas destes mesmos 74 (setenta e quatro) vocábulos com o objetivo de apresentação de um espaço vocálico de referência. O tratamento estatístico compreende tomada de média, desvio padrão e teste T dos valores obtidos.

Análises dos dados da variedade de Belém (PA) de caráter ainda preliminar (CRUZ 2011, REIS 2011, REIS *et al* 2011) tem demonstrado que a variedade do português falada na Amazônia Paraense possui no nível acústico 4 (quatro) variantes para as vogais médias pretônicas: a) alteamento como em fut[i]bol e c[u]madre; b) manutenção como em fut[e]bol e c[o]madre; c) abaixamento como em fut[E]bol e c[O]madre; d) mas há também casos de enfraquecimento das vogais pretônicas ou mesmo de total apagamento das mesma como em fut[Ø]bol e c[Ø]madre. As análises acústicas têm confirmado os resultados das descrições sociolinguísticas de que a manutenção das médias é a variante de maior ocorrência no português falado na Amazônia Paraense, seguida do alteamento das médias pretônicas e a variante de abaixamento é a de menor ocorrência nos *corpora*. Com relação ao nível de escolaridade, os falantes de baixa escolaridade possuem comportamento linguístico no que diz respeito à realização das vogais médias pretônicas distinto dos falantes de alta escolaridade (níveis médio e superior) independentemente do sexo.

O outro trabalho semelhante de que se tem conhecimento, apesar de mais voltado para as características acústicas das vogais tônicas, utilizou dados artificiais e não existentes no sistema português como um todo (ESCUADERO *ET AL* 2009). Portanto o presente estudo prioriza uma análise acústica de dados naturais do português.

Com relação ao PROBRAVO, o presente estudo pretende contribuir diretamente na implementação de uma metodologia adequada para a análise acústica das vogais médias pretônicas de todas as quinze variedades estudadas pela equipe nacional.

Referências Bibliográficas

CAMPOS, Benedita M. do S. Pinto. *Alteamento vocálico em posição pretônica no português falado no Município de Mocajuba-Pará*. 2008. Dissertação (Curso de Mestrado em Letras) – Universidade Federal do Pará, Belém: UFPA.

CASSIQUE, Orlando *et al.* “Variação das Vogais Médias Pré-tônicas no português falado em Breves (PA)”, in D. Hora (org.), *Vogais no ponto mais oriental das américas*. João Pessoa (PB): Ideia, 2009, pp. 163-184.

COSTA, Raquel Maria da Silva. *Descrição sociolinguística das vogais médias postônicas não-finais /o/ e /e/ no português falado no município de Cametá-PA*. Belém – Pará: UFPA, 2010. (Dissertação de Mestrado).

CAVALCANTE, Carlos. *O Sistema Vocálico do português Falado em Bragança (PA): Análise Experimental*. Belém: UFPA/ILC/CML, **em andamento**. (Projeto de Dissertação de Mestrado).

CRUZ, Regina. *Projeto de Pesquisa “Brazilian Amazon Portuguese vowel system: acoustic analysis”* (processo BEX1754/10-6) Brasília: CAPES/ Fulbright, New York (USA): New York University, 2011. (Relatório Técnico-científico Aprovado).

CRUZ, Regina *et al.* “As Vogais Médias Pretônicas no Português Falado nas Ilhas de Belém (PA)”, in Maria do Socorro Silva de ARAGÃO (Org.). *Estudos em fonética e fonologia no Brasil*. João Pessoa: GT-Fonética e Fonologia / ANPOLL, 2008.

DIAS, Marcelo Pires; **CASSIQUE**, Orlando & **CRUZ**, Regina. “O alteamento das vogais pré-tônicas no português falado na área rural do município de Breves (PA): uma abordagem variacionista”. *Revista Virtual de Estudos da Linguagem (REVEL)*. Porto Alegre, n. 9, vol. 5, jul. 2007. (Disponível em: http://www.revel.inf.br/site2007/_pdf/9/artigos/).

ESCUADERO, Paola *et al.* “A cross-dialect acoustic description of vowels: Brazilian and European Portuguese”, *Journal of Acoustical Society of America*, volume 126, número 3, Setembro 2009, pp. 1379-1393.

REIS, Érica. Plano de Trabalho *O Sistema Vocálico do Português falado em Belém (PA): análise experimental*. Belém: UFPA/ILC/FALE, 2011. (Relatório de Iniciação Científica).

REIS, Érica *et al.* *O Sistema vocálico do Português falado em Belém (PA): análise qualitativa, comunicação oral* apresentada na IV Jornada do Curso de Letras do Campus Universitário de Cametá, Cametá (PA), 14 a 16 de setembro de 2011.

RODRIGUES, Doriedson do Socorro & **ARAÚJO**, Marivana dos Prazeres. “As vogais médias pretônicas / e / e / o / no português falado no município de Cametá/PA – a harmonização vocálica numa abordagem variacionista”, *Cadernos de Pesquisa em Linguística, Variação no Português Brasileiro*, Leda Bisol & Cláudia Brescancini (orgas.), volume 3, Porto Alegre, novembro de 2007, pp. 104-126.

SOUSA, Josivane do Carmo Campos. *A Variação das Vogais Médias Pretônicas no Português Falado na Área Urbana do Município de Belém/PA*. 2010. Dissertação (Curso de Mestrado em Letras) – Universidade Federal do Pará, Belém: UFPA.

Palavras-chave: *corpus* oral, vogais médias pretônicas, análise acústica, português brasileiro.

Trabalho para Comunicação Oral