

IDENTIFICAÇÃO DE FENÔMENO DE TRANSFERÊNCIA GRAFO-FÔNICO-FONOLÓGICA PB-ILE UTILIZANDO REDES NEURAIS ARTIFICIAIS

Atos Apollo Silva Borges (bolsista ICV/UFPI); Prof. Dra. Aratuzza Rodrigues Silva Rocha (Co-Orientadora, Programa de Pós-Graduação em Linguística Aplicada – UFC); Prof. Dr. Fábio Rocha Barbosa (Departamento de Engenharia Elétrica – UFPI)
atosborges00@gmail.com

Introdução

Os fenômenos de transferência entre línguas cometidos por falantes não nativos já são conhecidos e bem documentados na literatura. Contudo, o reconhecimento e a classificação dos mesmos são feitos, principalmente, através de transcrições, um processo lento e trabalhoso. A identificação rápida desse tipo de fenômeno seria de grande valia para sistemas de nivelamento de alunos de língua estrangeira, podendo ser usados em escolas de idiomas, sistemas de ensino à distância, aprendizes autodidatas, aplicativos de língua estrangeira ou até mesmo por pesquisadores linguistas. No presente trabalho, foi realizada uma verificação das possibilidades de técnicas e ferramentas que podem ser utilizadas na classificação automática desses processos de transferência, em especial do fenômeno de transferência grafo-fônico-fonológica de Sonorização de <s> (SOS), guiado sob uma ótica do paradigma conexionista da linguagem.

Metodologia

Seleção de palavras

A seleção de palavras da língua inglesa para a composição dos dados de treino foi feita a partir da literatura pesquisada (Rocha, 2012). Baseando-se na metodologia de outros trabalhos (Zimmer, 2006), foram escolhidas palavras que sejam capazes de provocar os processos de transferência em humanos.

Tabela 1: Processo investigado na pesquisa.

Palavra exemplo	Pronúncia nativa	Pronúncia com processo
case	[keɪs]	[keɪz]

Tabela 2: Distribuição de palavras selecionadas

Alta frequência	Baixa frequência	Cognatas	Não-cognatas	Logatomas	Total
61	49	69	41	15	125

Coleta das produções

Para gerar os fenômenos foi utilizado sistema de TTS do software livre Google Tradutor® em duas opções de idioma: Português Brasileiro e Inglês.

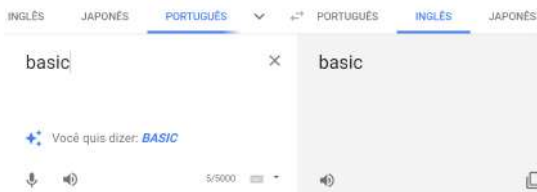


Figura 1: Método para a simulação de fenômeno no Google Tradutor.

Extração de Descritores

Foram utilizados dois descritores de sinal (Frequência formante – FF e Coeficiente Mel-Cepstrais – MFCC) obtidos através do software especializado em tratamento de áudio Praat versão 6.0.20.

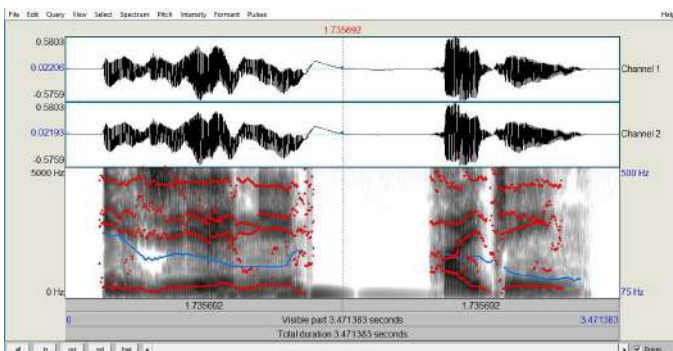


Figura 2: Exemplo de tela do Praat onde os áudios foram tratados

Desenvolvimento e treinamento da RNA

Para realizar a identificação de fenômenos nas produções foi desenvolvida uma Rede Neural Artificial (RNA) do tipo *Multilayer Perceptron* com algoritmo de treinamento Levenberg-Marquardt. Para esse desenvolvimento utilizou-se o software Matlab®, já que este apresenta funções auxiliares na construção de Redes Neurais Artificiais.

Resultados e discussão

Resultado da simulação

Tabela 2: Porcentagem de ocorrência de fenômeno de transferência em cada categoria de palavras.

Alta frequência (%)	Baixa frequência (%)	Cognatas (%)	Não cognatas (%)	Logatomas (%)	Total (%)
72,13	95,92	88,41	73,17	93,34	84,00

Resultado da RNA com Frequências Formantes

Tabela 3: Porcentagem de acerto para a RNA com FF após 100 testes.

Acerto da RNA para cada divisão de dados (Treino/Validação/Testes)(%)								
20/15/65			50/15/35			70/15/15		
Melhor Rede (%)	Média (%)	DP	Melhor Rede (%)	Média (%)	DP	Melhor Rede (%)	Média (%)	DP
100	89,68	0,0464	97,37	89,84	0,0472	100	90,66	0,0412

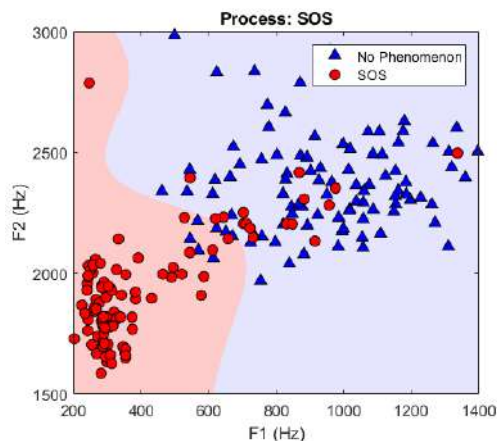


Figura 3: Fronteira de decisão da RNA

Resultado da RNA com MFCC

Tabela 4: Porcentagem de acerto para RNA com MFCC após 100 testes.

Acerto da RNA para cada divisão de dados (Treino/Validação/Testes)(%)								
20/15/65			50/15/35			70/15/15		
Melhor Rede (%)	Média (%)	DP	Melhor Rede (%)	Média (%)	DP	Melhor Rede (%)	Média (%)	DP
100	99,63	0,0118	100	99,74	0,0088	100	99,61	0,0132

Conclusão

A partir dos resultados obtidos na pesquisa é possível afirmar que as hipóteses iniciais foram todas confirmadas. A investigação dos processos revelou tanto novas informações sobre a simulação desses fenômenos como mostrou técnicas necessárias para sua identificação. Portanto, a partir das técnicas desenvolvidas, as possibilidades de implementação de um sistema de identificação automático em plataformas *online* ou *mobile* são viáveis e representariam um avanço tanto para o aprendizado de línguas como para a própria pesquisa da área. O algoritmo de TTS do Google Tradutor® também se mostrou uma ferramenta importante na compreensão dos processos de transferência entre línguas, reforçando ainda mais o paradigma conexionista da linguagem.

Agradecimento

Agradeço à UFPI pelo apoio, assim como aos colegas do Grupo de Pesquisa da Universidade Federal do Piauí na qual faço parte.

Referências

- ROCHA, A. R. S. Os efeitos da instrução explícita em fonologia na produção e percepção de consoantes da língua inglesa. Universidade Estadual do Ceará. Fortaleza. 2012.
ZIMMER, M. C.; ALVEZ, U. K. A produção de aspectos fonético-fonológicos da segunda língua: instrução explícita e conexonismo. *Revista Linguagem & Ensino*, Pelotas, v. 9, n. 2, p. 101-143, 2006.