



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2024**

EDIÇÃO MODERNIZADA DE AMOSTRAS DE FALA DA ZONA RURAL DE **FEIRA DE SANTANA**

Ana Gabriela Oliveira Martins Araújo¹; Mariana Fagundes de Oliveira Lacerda²

1. Bolsista – Modalidade Bolsa/FAPESB, Graduando em Letras Vernáculas, Universidade Estadual de Feira de Santana, e-mail: gabrielamartins2204@gmail.com
2. Orientador, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: marianafagundes@uefs.br

PALAVRAS-CHAVE: amostras de fala; edição modernizada; português brasileiro.

INTRODUÇÃO

O projeto A Língua Portuguesa no Semiárido Baiano – do Núcleo de Estudos de Língua Portuguesa (NELP), <<https://nelp.uefs.br/>>, da Universidade Estadual de Feira de Santana (UEFS) –, financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), começou a ser executado em 1998, tendo, como objetivo geral, contribuir para o estudo da sócio-história da língua portuguesa, na sua variedade falada, delineando as seguintes frentes de trabalho: (i) Formação de banco de dados com amostras de fala de comunidades rurais e urbanas do semiárido baiano; (ii) Realização de estudos sócio-históricos sobre a formação da língua portuguesa nas localidades pesquisadas; (iii) Realização de análises linguísticas, à luz de diferentes teorias linguísticas; (iv) Elaboração de materiais didáticos para serem utilizados nas escolas da região.

As amostras rurais, conforme Carneiro *et al* (2016), foram compostas em áreas passíveis de diversidade étnica, havendo indícios de que a população fosse formada por remanescentes de quilombo em Casinhas (Jeremoabo); por presença étnica de origem afro-brasileira em Barra/Bananal (Rio de Contas/Região da Chapada Diamantina), Piabas (Caém/Ancelino da Fonseca/Região do Piemonte da Diamantina) e Matinha (Feira de Santana/Região Paraguaçu); por mestiços de brancos de origem portuguesa em Lagoa do Inácio (Jeremoabo) e Mato Grosso (Rio de Contas). Isto implica a existência de variações linguísticas provenientes de contato linguístico entre línguas distintas, no processo de aquisição de língua.

O banco com gravações de fala do semiárido conta com mais de 90 horas de gravação em comunidades representativas da variedade falada, constituída nas regiões, cujos informantes tinham pouca ou nenhuma escolaridade.

Na presente pesquisa, são consideradas as amostras de fala da zona rural de Feira de Santana (região Paraguaçu) - Matinha (Almeida; Carneiro, 2008). Propõe-se a

realização da edição modernizada, com uso do eDictor (Kepler; Paixão de Sousa; Faria, 2009), desse conjunto de textos, a ser disponibilizada na página do Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS/NELP) <<http://www.uefs.br/cedohs/>>.

MATERIAL E MÉTODOS

O conjunto de textos de Matinha – com formação étnica marcante de matriz africana –, que faz parte do banco do NELP/ UEFS, reúne 12 entrevistas com falantes analfabetos ou pouco escolarizados, de ambos os sexos, divididos em três faixas etárias (Almeida; Carneiro, 2008). A metodologia de coleta foi feita nos moldes da sociolinguística quantitativa (Labov, 1972; 1982). Os inquéritos foram gravados na fase I do projeto A Língua Portuguesa no Semiárido Baiano, entre 1997 e 1999. Nessa fase, a intenção era contribuir para o conhecimento da realidade linguística brasileira e, de forma específica, para o estudo da língua falada em áreas do semiárido baiano. As comunidades foram escolhidas de forma que representassem o avanço da língua portuguesa na Bahia, a partir do século XVII, numa perspectiva sócio-histórica.



Figura 1: Coleção Amostras da Língua Falada no Semiárido Baiano (Almeida; Carneiro, 2008).
Fonte: CE-DOHS.

A edição modernizada será feita com o uso do eDictor, desenvolvido por Kepler, Paixão de Sousa e Faria (2009). Essa ferramenta combina um editor de XML e um etiquetador morfossintático, e permite a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em html), e de versões com anotação morfossintática (em texto simples e xml). Trata-se de um feliz conagraamento entre as mais novas tecnologias e a antiga Filologia.

RESULTADOS E/OU DISCUSSÃO

Nesta seção, apresentam-se, em síntese, os resultados alcançados com a execução da pesquisa.

Seguem figuras ilustrativas do passo a passo do processo de edição modernizada das amostras orais em questão, usando o eDíctor:

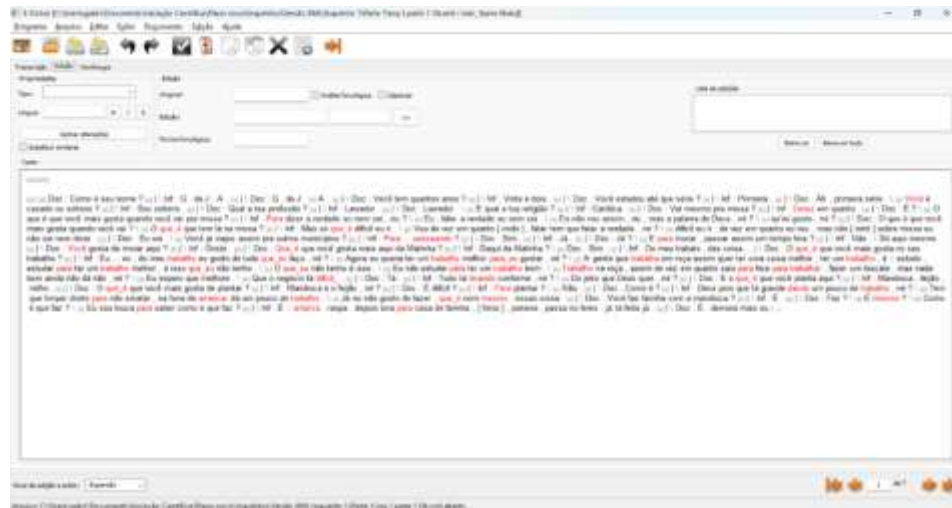


Figura 2 - Inquérito submetido à modernização.

Fonte: CE-DOHS.



Figura 3 - Realização da edição modernizada no nível da padronização.

Fonte: CE-DOHS.

Na Figura 3, está em destaque, a título de exemplificação, a palavra, originalmente realizada no *corpus* de fala, como “trabaio”. A intervenção em questão trata-se de modernização, por meio da etiqueta de padronização. A palavra é modernizada de “trabaio” para “trabalho”. A palavra padronizada aparece na *lista de edições*, na qual são indicadas todas as intervenções feitas na palavra.

CONSIDERAÇÕES FINAIS

Ao final do período de execução desta pesquisa, apresentou-se ao projeto Corpus Eletrônico do Documentos Históricos do Sertão (CE-DOHS) a edição modernizada das 12 entrevistas da zona rural de Feira de Santana, selecionadas como *corpus* de trabalho, para serem disponibilizadas na página online (http://www.uefs.br/cedohs/view/colecoes_documentais.html#A1823).

O CE-DOHS possui documentos de três categorias maiores: orais, manuscritos e impressos. A pesquisa aqui descrita traz resultados que se somam à agenda de constituição de *corpora* orais, em edição modernizada, em linguagem XML, com uso do eDictor.

O trabalho em questão contribui com a formação de banco de dados eletrônicos, em consonância com os princípios da Ciência Aberta, para estudo da história do português brasileiro.

REFERÊNCIAS

ALMEIDA, Norma Lucia Fernandes de; CARNEIRO, Zenaide de Oliveira Novais (Org.). **Coleção Amostras da Língua Falada no Semiárido Baiano**. Feira de Santana: UEFS Editora, 2008.

ALMEIDA, Norma Lucia Fernandes de; CARNEIRO, Zenaide de Oliveira Novais (Org.). **Variação linguística no semiárido baiano**. Feira de Santana: UEFS Editora, 2014.

CARNEIRO, Zenaide de Oliveira Novais; LACERDA, Mariana Fagundes de Oliveira. Corpus Eletrônico de Documentos Históricos do Sertão: etapa 1 (1750-2000). **Revista Binacional Brasil Argentina: Diálogo entre as Ciências**, v. 8, p. 205-221, 2019.

CARNEIRO, Z. O. N.; LACERDA, M. F. O. (Org). CE-DOHS - Corpus Eletrônico de Documentos Históricos do Sertão (2012-2025). Disponível em: <http://www.uefs.br/cedohs>. Acesso em: 15 jan 2024.

GONÇALVES, M. F.; BANZA, A. P. Fontes de metalinguísticas para a história do português clássico. In: GONÇALVES, M. F.; BANZA, A. P. **Patrimônio Textual e Humanidades Digitais: da antiga à nova filologia**. Évora: CIDEHUS, 2013. p. 73-112.

LACERDA, M. F. O; CARNEIRO, Z. O. N.; SANTIAGO, H. S. (Org). **Núcleo de Estudos de Língua Portuguesa da UEFS**. Disponível em: <https://nelp.uefs.br/>. Acesso em: 20 ago 2024.

MATTOS E SILVA, Rosa Virgínia. **Ensaio para uma sócio-história do português brasileiro**. São Paulo: Parábola Editorial, 2004, p. 91-108.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. E. dicator: Novas perspectivas na codificação e edição de corpora de textos históricos. In: **Anais do VIII Encontro de Linguística de Corpus**, realizado na UERJ, 13 a 14 de novembro de 2009. Rio de Janeiro, 2009. p. 69-105.

SANTIAGO, H. S.; LACERDA, M. F. O., BRITO, R. C.; CARNEIRO, Z. O. N. CEDOHS: um banco de dados sociolinguísticos para a história do português brasileiro. **LaborHistórico**, Rio de Janeiro, 7 (Especial): 311-329, 2021. DOI: <https://doi.org/10.24206/lh.v7iespec.41640> 2021. Acesso em: 15 jan. 2024.

SHEPHERD, Tania; SARDINHA, Tony Berber; PINTO, Marcia Veirano (Org.). **Caminhos da linguística de corpus**. Campinas: Mercado de Letras, 2010.