



**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

## **XXVII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA – 2023**

CE-DOHS/UEFS: EDIÇÃO MODERNIZADA DE AMOSTRAS DE FALA DE  
BARRA DOS NEGROS/BANANAL

**Ana Gabriela Oliveira Martins Araújo**<sup>1</sup>; Mariana Fagundes de Oliveira Lacerda<sup>2</sup>

1. Bolsista PIBIC/CNPq/UEFS, Graduanda em Licenciatura em Letras Vernáculas, e-mail:  
[gabrielamartins2204@gmail.com](mailto:gabrielamartins2204@gmail.com)
2. Orientadora, Departamento Letras e Artes, Universidade Estadual de Feira de Santana, e-mail:  
[marianafagundes@uefs.br](mailto: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://nelpuefs.wordpress.com/>>, 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, tendo 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.

Neste plano de trabalho, são consideradas as amostras de fala de Barra dos Negros/Bananal (CARNEIRO; ALMEIDA, 2008). Propõe-se a realização da edição modernizada, com uso do eDictor (KEPLER; PAIXÃODESOUSA; FARIA, 2009), desse conjunto de textos, a ser disponibilizada na página do CE-DOHS, < <http://www.uefs.br/cedohs/>>.

**Figura 1:** Página inicial do site CE-DOHS



**Fonte:** <<http://www.uefs.br/cedohs/>>

## **MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)**

O conjunto de textos de Barra dos Negros/Bananal – 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 (CARNEIRO; ALMEIDA, 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.

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 conagraçamento entre as mais novas tecnologias e a antiga Filologia.

**Figura 2:** Logomarca da ferramenta eDictor.

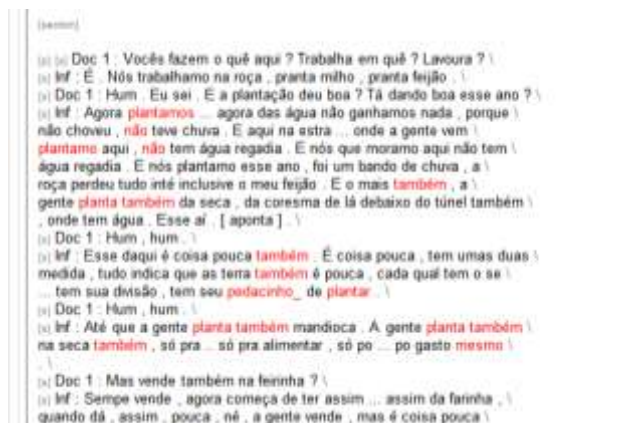


**Fonte:** <<https://humanidadesdigitais.org/edictor/>>

## RESULTADOS E/OU DISCUSSÃO

A seguir, ilustra-se o processo de edição modernizada do *corpus* em questão, apresentando parte do inquérito 5 :

**Figura 2:** Inquérito 5



(barr) \

(D) Doc 1 : Vocês fazem o quê aqui ? Trabalha em quê ? Lavoura ? \

(I) Irf : E . Nós trabalhamos na roça , pranta milho , pranta feijão . \

(D) Doc 1 : Hum . Eu sei , E a plantação deu boa ? Tá dando boa esse ano ? \

(I) Irf : Agora **plantamos** ... agora das água não ganhamos nada , porque \

não choveu , **não** teve chuva , E aqui na estrada ... onde a gente vem \

**plantamo** aqui , **não** tem água regadia . E nós que moramos aqui não tem \

água regadia . E nós plantamos esse ano , fui um bando de chura , a \

roça perdeu tudo até inclusive o meu feijão . E o mais **também** , a \

gente **planta também** da seca , da coresma de lá debaixo do túnel também \

, onde tem água . Esse aí , [ aponta ] . \

(D) Doc 1 : Hum , hum . \

(I) Irf : Esse daqui é coisa pouca **também** . É coisa pouca , tem umas duas \

medida , tudo indica que as tem **também** é pouca , cada qual tem o se \

... tem sua divisão , tem seu **pedacinho** de plantar . \

(D) Doc 1 : Hum , hum . \

(I) Irf : Até que a gente **planta também** mandioca . A gente **planta também** \

na seca **também** , só pra ... só pra alimentar , só po ... po gasto **mesmo** \

... \

(D) Doc 1 : Mas vende também na feirinha ? \

(I) Irf : Sempre vende , agora começa de ter assim ... assim da farinha , \

quando dá , assim , pouca , né , a gente vende , mas é coisa pouca \

**Quadro 1:** Intervenções no inquérito 5.

Forma Original	Níveis de edição	Forma modernizada
Prantamos	Padronização	Plantamos
Num	Padronização	Não
Tombém	Padronização	Também
Pranta	Padronização	Planta
Pedacim	Padronização	Pedacinho
Mehmo	Padronização	Mesmo
Sábo	Padronização	Sábado
Trabaia	Padronização	Trabalha
Sempe	Padronização	Sempre
Choveno	Padronização	Chovendo

## CONSIDERAÇÕES FINAIS

Ao final do período de execução do plano de trabalho, apresentou-se, no âmbito do projeto Corpus Eletrônico do Documentos Históricos do Sertão (CE-DOHS), a edição modernizada das 12 entrevistas de Barra dos Negros/Bananal, que fazem parte do banco de textos do NELP/UEFS. A edição modernizada, com o léxico de edições, será disponibilizada no site, em <[http://www5.uefs.br/cedohs/view/colecoes\\_documentais.html#A1823](http://www5.uefs.br/cedohs/view/colecoes_documentais.html#A1823)>.

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

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.

## REFERÊNCIAS

- 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 às Ciências, v. 8, p. 205-221, 2019.
- CARNEIRO, Zenaide de O. N.; ALMEIDA, Norma Lucia F. de (Org.). Amostras da língua falada na zona rural de Rio de Contas. 2. ed. Feira de Santana: Editora da UEFS/Edufba, 2008. v. 1.
- CARNEIRO, Z. O. N.; LACERDA, M. F. O. (Org). CE-DOHS - Corpus Eletrônico de Documentos Históricos do Sertão (2012-2025). URL: <http://www.uefs.br/cedohs>. Accessed in: Acesso em: dia, mês e ano.
- 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). NELP: Núcleo de Estudos de Língua Portuguesa da UEFS. Disponível em: <https://nelp.uefs.br/>. Acesso em: Acesso em: 20 ago 2023.
- MATTOS E SILVA, Rosa Virgínia. Ensaio para uma sócio-história do português brasileiro. São Paulo: Parábola Editorial, 2004, p. 91-108.
- PAIXÃODESOUSA, M. C.; KEPLER, F. N.; FARIA, P. Edictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: Anais do VIII Encontro de Linguística 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. 2022.
- SHEPHERD, Tania; SARDINHA, Tony Berber; PINTO, Marcia Veirano (Org.). Caminhos da linguística de corpus. Campinas: Mercado de Letras, 2010.