



**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

## **XXIV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2020**

### **O LIVRO DO GADO DO BREJO DO CAMPO SECO: EDIÇÃO EM LINGUAGEM XML USANDO O EDICTOR**

**Taine do Rosário<sup>1</sup> e Mariana Fagundes de Oliveira Lacerda<sup>2</sup>;**

1. Taine do Rosário, Graduanda em Letras com Língua Inglesa, Universidade Estadual de Feira de Santana, e-mail: taine.rosario99@gmail.com
2. Mariana Fagundes de Oliveira Lacerda, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: marianafagundes@uefs.br

**PALAVRAS-CHAVE:** Português brasileiro; Livro de Fazenda; Edição Digital.

#### **INTRODUÇÃO**

Este trabalho pretende apresentar os resultados alcançados com o desenvolvimento do plano de trabalho intitulado “O Livro do Gado do Brejo do Campo Seco: edição em XML usando o eDictor”. A construção de *corpora* para o estudo do português integra a grande agenda de reconstrução do português brasileiro, desenvolvida, em âmbito nacional, pelo Programa Para a História do Português Brasileiro (PHPB), juntamente com outros programas e projetos em rede, a exemplo do Programa Para a História do Português (PROHPOR) e do Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS). A edição digital do Livro do Gado, com a utilização da ferramenta eDictor, contribuiu diretamente com a referida agenda e, para além disso, trouxe discussões sobre possibilidades aprimoramento do software de edição. Os referenciais teóricos aqui usados são imprescindíveis na área da Língua Histórica e Linguística de *Corpus*, além daqueles específicos para o entendimento do material trabalhado, são eles os trabalhos de SANTOS FILHO (1956); MATTOS E SILVA (2001); PAIXÃO DE SOUSA, KEPLER e FARIA (2009); LACERDA e CARNEIRO (2016) e BRANDÃO (2019).

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

O corpus de trabalho trata-se do Livro do Gado do arquivo do Sobrado do Brejo, escrito por três gerações, no então povoado de Bom Jesus dos Meiras, atual Brumado, no interior da Bahia. O manuscrito traz registros representativos do tempo pretérito do português brasileiro, mais exatamente, o período setecentista. A referida documentação é de foro privado e importante para estudo da penetração e difusão da escrita no interior da Bahia. O Livro do Gado faz parte banco de textos do projeto CE-DOHS. A edição eletrônica foi

feita com o uso do eDictor, ferramenta desenvolvida por Kepler, Paixão de Sousa e Faria (2007), que permite a feitura de edições digitais de texto. O eDictor é voltado especialmente ao trabalho filológico e à análise linguística automática. 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). Tratou-se de um feliz conagração entre as mais novas tecnologias e a antiga Filologia.

### **RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)**

O Livro do Gado é um documento de foro privado que integra o Acervo do Sobrado do Brejo do Campo Seco, localizado no então povoado de Bom Jesus dos Meiras, atual município de Brumado, no interior da Bahia. O documento foi escrito por 3 gerações de indivíduos da família Pinheiro-Cangaçu, totalizando cerca de 150 anos, entre 1755 e 1832, de registros do cotidiano agro-pastoril da referida fazenda. O manuscrito, que foi preservado pela 4ª geração da família, encontra-se sobre os cuidados de Lycurgo Santos Neto, filho de Lycurgo Santos Filho, o qual concedeu anuência ao projeto CE-DOHS para que os estudos com o documento pudessem acontecer.

O Projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS) desde a sua criação, no ano de 2012, como um dos projetos integrantes do Núcleo de Estudos de Língua Portuguesa (NELP), do Departamento de Letras e Artes (DLA) da Universidade Estadual de Feira de Santana (UEFS), tem desenvolvido trabalhos na elaboração e disponibilização de uma vasta base documental para o estudo do Português Brasileiro (PB). O projeto, atualmente coordenado pela professora doutora Zenaide de Oliveira Novais Carneiro e co-coordenado pela professora doutora Mariana Fagundes de Oliveira Lacerda, nasce com o objetivo principal de compor um *corpus* que representasse, de maneira satisfatória, a realidade linguística de formação do Português Brasileiro (PB).

O CE-DOHS disponibiliza, na rede mundial de computadores, uma plataforma que, ao ser acessada, permite o acesso a dois conjuntos de documentos, sendo o primeiro composto por fontes escritas e orais, cujos os textos são representativos do período de 1823 a 2000, tendo sido produzidos por escreventes nascidos no Brasil, as amostras de fala desse mesmo conjunto foram gravadas nos anos 90 do século XX, também de indivíduos brasileiros, totalizando 1553 textos já editados e disponibilizados. O segundo conjunto retrocede ainda mais no tempo, e apresenta manuscritos que foram produzidos entre 1640 e 1822 por diversas populações nascidas no Brasil, estando esses ainda em processo de edição, para, posteriormente, serem liberados. O Livro do Gado do Brejo do Campo Seco faz parte deste segundo conjunto, e sua disponibilização é grande importância para o estudo da mudança linguística no português brasileiro.

Para que o documento seja disponibilizado, é necessário que se cumpram algumas etapas, a primeira delas dá origem à versão fac-similar. Os fac-símiles do Livro do Gado foram feitos com o uso do método Lapelinc, que trata-se de método fotográfico, cientificamente controlado, desenvolvido no Laboratório de Pesquisa em Linguística de Corpus da Universidade Estadual do Sudoeste da Bahia (UESB). A edição fac-similar, além de conservar os documentos, permite que a edição semidiplomática seja realizada.

Elaine Brandão Santos desenvolveu, como parte dos requisitos para obtenção do título de Mestre em Estudos Linguístico pelo Pós-graduação em Estudos Linguísticos (PPGEL) da Universidade Estadual de Feira de Santana (UEFS), as edições semidiplomáticas do *corpus* seguindo as Normas para transcrição de documentos Para a História do Português Brasileiro (MATTOS E SILVA, 2001).

A última etapa de preparação do documento, para que suas versões sejam disponibilizadas no banco de dados, diz respeito à edição digital, em linguagem XML, feita com o uso do eDictor, uma ferramenta que executa o que Gonçalves e Banza (2013, p. 4) chamam de “feliz congraçamento”, advindo da junção das inovações trazidas pela era das Humanidades Digitais com as técnicas e rigor já conhecidos da Filologia. Desenvolvido, no âmbito no projeto Corpus Histórico do Português Tycho Brahe (<http://www.tycho.iel.unicamp.br/~tycho/corpus/catalogo.html>), por Fábio Kepler, Maria Clara Paixão de Souza e Pablo Faria, o eDictor é um software gratuito especializado em edição filológica digital e anotação linguística.

Alguns desafios significativos foram encontrados no processo de edição, principalmente para a etiquetagem de elementos específicos do documento, visto que o conteúdo dos registros presentes no documento não seguem uma linearidade de escrita ou uma narrativa, além de conter símbolos e contagens, informações que ainda não são contempladas pelas funcionalidades do editor digital. Em função dessas dificuldades, os fólios se encontram parcialmente editados, pois há a necessidade de mudanças na estrutura computacional do software que não podem ser executadas por editores.

### **CONSIDERAÇÕES FINAIS (ou Conclusão)**

Apesar da não finalização do objetivo principal, tem-se, com a proposta deste plano de trabalho, a abertura de novos horizontes e possibilidades. As dificuldades encontradas servirão para o aprimoramento da ferramenta, sendo assim, futuros documentos que possam apresentar características semelhantes ao Livro do Gado serão facilmente editados.

### **REFERÊNCIAS**

- GONÇALVES, M. F.; BANZA, A. P. (org.). **Património textual e humanidades digitais: da antiga à nova Filologia**. Évora: CIDEHUS, 2013.
- LACERDA, M. F. O.; CARNEIRO, Z. O. N. **Edição filológica e digital do Livro do Gado e do Livro de Razão do Arquivo do Sobrado do Brejo (Bahia setecentista e oitocentista)**. In: Labor Histórico, Rio de Janeiro, 2, 2016. p. 151-163.
- Normas para transcrição de documentos manuscritos para a história do português do brasil**. In: MATTOS E SILVA, R. V. (Org.). Para a história do português brasileiro: primeiros estudos. V. II, tomos I e II. São Paulo: Humanitas/ FFCHL/USP:FAPESP, 2001.
- PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. **E-dictor: 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.

**PROJETO CORPUS ELETRÔNICO DE DOCUMENTOS HISTÓRICOS DO SERTÃO** (disponível em <http://www.uefs.br/cedohs/>), 2011.

SANTOS, E. B. **O Livro do Gado do Brejo do Campo Seco (Bahia)**: edição semidiplomática e descrição de índices grafo-fonéticos.

SANTOS FILHO, Lycurgo. **Uma comunidade rural do Brasil antigo**: aspectos da vida patriarcal no Sertão da Bahia, nos séculos XVIII e XIX. Salvador: UEFS: Fundação Pedro Calmon, 2012