



**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**  
**Elaboração dos metadados do Livro do Gado do**  
**Sobrado do Brejo (séculos XVIII e XIX): uma documentação escrita**  
**por três mãos**

**Aracele Santos de Lima<sup>1</sup> e Mariana Fagundes de Oliveira Lacerda<sup>2</sup>**

1. Aracele Santos de Lima, Graduanda em Letras com Língua Inglesa, Universidade Estadual de Feira de Santana, e-mail: aracele.santosdelima@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**

Apresenta-se aqui a elaboração dos metadados do Livro do Gado do Brejo do Campo Seco (séculos XVIII e XIX). Os metadados reúnem informações diversas, sobre dados do documento: autor, conteúdo, destinatário, local, data do documento, referência, fonte, gênero do autor, gênero do documento; e dados do processamento: revisão final da edição XML, primeira revisão da edição XML, edição semidiplomática, revisão da edição semidiplomática, edição XML, número de palavras.

O Livro do Gado referido pertence ao projeto CE-DOHS: Corpus Eletrônico de Documentos Históricos do Sertão – coordenado pelas professoras doutoras Mariana Fagundes de Oliveira Lacerda e Zenaide de Oliveira Novais Carneiro –, do Núcleo de Estudos de Língua Portuguesa (NELP), da UEFS.

O CE-DOHS conta com ótimos recursos tecnológicos, no universo das Humanidades Digitais, disponibilizando edições semidiplomáticas, em PDF, e, por meio do estabelecimento de redes com projetos que desenvolvem a Linguística de Corpus – como o projeto Corpus Histórico do Português Tycho Brahe (<http://www.tycho.iel.unicamp.br/~tycho/corpus/>), coordenado por Charlotte Galves (UNICAMP) –, também edições em linguagem XML, usando o eDictor.

O eDictor é um programa computacional, desenvolvido por Kepler, Paixão de Souza e Faria (2007), para facilitar a edição eletrônica de textos antigos, que, anotada sintaticamente, permite a busca automática de dados no estudo linguístico. Como se vê,

Do feliz conagraçamento entre as mais recentes tecnologias e a antiga Filologia, surgiu um novo universo de possibilidades para a preservação, disponibilização e análise de textos antigos, universo em que é possível oferecer ao leitor mais de uma edição do mesmo texto, permitindo que tenha ao seu dispor o texto editado, em diferentes versões, e o seu original. (GONÇALVES; BANZA, 2013, p. 4)

Segundo Shepherd *et al.* (2012, p. 11), “A ideia de coligir coleções de textos naturais com o objetivo de os submeter à análise linguística remonta ao trabalho dos estruturalistas norteamericanos da década de 1950, tais como Harris (1951) e Fries (1952). Com o Brown Corpus (Francis e Kucera, 1954), surgiria o primeiro corpus

eletrônico compilado para este fim. Na altura, praticamente não existiam textos escritos em formato digital, os computadores eram máquinas enormes e caras, que ocupavam salas inteiras, e os programas informáticos demoravam horas e até dias a correr.

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

O documento em questão se trata de um documento de foro privado, o Livro do Gado do Brejo do Campo Seco, escrito por três gerações, do último quartel do século XVIII ao terceiro quartel do século XIX, no sertão baiano, de grande importância para a história da penetração e difusão da língua escrita na Bahia. Faz parte do banco de textos do projeto CE-DOHS, tendo sido editado, em versão fac-similar e semidiplomática, por Brandão (2019), no âmbito do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL), da UEFS, sob a orientação da professora doutora Mariana Fagundes de Oliveira Lacerda.

A elaboração dos metadados foi feita com base nos pressupostos da Linguística de Corpus e nos critérios do Projeto CE-DOHS e do projeto Tycho Brahe. Trazem dados biográficos (nome, sexo, data de nascimento, local de nascimento, etnia, escolaridade, grau de habilidade de escrita, profissão) e dados do documento (número do documento, data de escrita, local de escrita, para quem, tipo documental, fonte).

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

Entendemos que, para se estudar uma língua, faz-se necessário, também, conhecer a história e os costumes da sociedade. Portanto, seguindo Petrucci (2003), consideram-se no CE-DOHS as seguintes questões: a saber: *Cuándo? Dónde? Quién? Para qué?* São questões importantes, voltadas para a compreensão das condições externas da história de uma língua. Abarcam-se aspectos econômicos e demográficos da região em questão, bem como o perfil sociocultural dos escreventes do Livro do Gado, além de sua finalidade.

Na pesquisa aqui descrita, preenchemos, no banco CE-DOHS, a partir da descrição que Brandão (2019) apresenta, levando em consideração as questões de Petrucci (2003), listadas acima, os metadados do documento; foram também preenchidos os metadados do processamento do documento (revisão final da edição XML, primeira revisão da edição XML, edição semidiplomática, revisão da edição semidiplomática, edição XML, número de palavras).

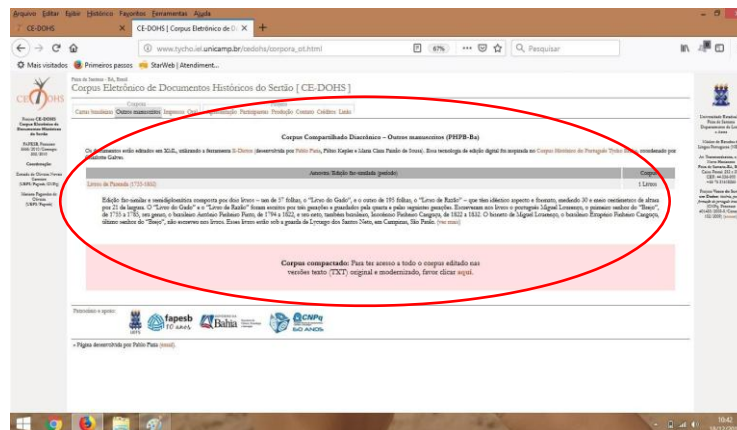
Segundo Brandão (2019, p. 127-128), a partir da investigação dessas questões de Petrucci,

foi possível realizar uma caracterização confiável do *corpus*, desvendando tanto o perfil sociocultural de quem produziu o Livro do Gado quanto a estrutura social da localidade pertencente ao referido manuscrito. Como resultado, além de fornecer dados sobre os aspectos demográficos, econômicos e culturais da região, foram elaboradas fichas com a descrição do perfil biográfico de cada escrevente e uma árvore genealógica de suas famílias, que abrange mais de 300 anos de descendência dos Pinheiros Canguçu, informação que possibilitará estudos na perspectiva gerativa e, também, sociolinguística.

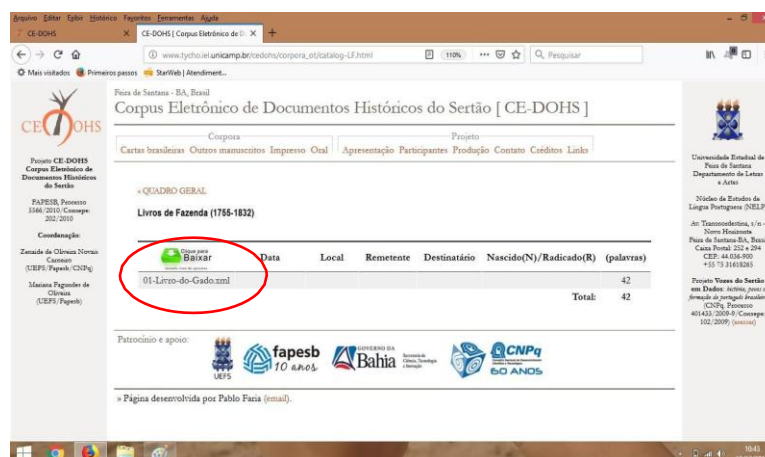
**Figura 1:** Página inicial do site CE-DOHS.



**Figura 2:** Página apresentação do Livro de Fazenda.



**Figura 3:** Página de acesso ao Livro do Gado



A atualização da página *online* dos metadados depende da finalização da edição XML do Livro do Gado, o que se dará assim que a ferramenta eDICTOR seja aperfeiçoada, tendo em vista as especificidades do referido livro manuscrito. A colega de projeto Taine do Rosário está responsável por essa edição modernizada do documento, já em fase de finalização. No SEMIC, serão apresentados os resultados finais.

## CONSIDERAÇÕES FINAIS

O banco CE-DOHS veio somar-se aos *corpora* eletrônicos constituídos fundamentalmente para análises linguísticas, dos quais o Brown Corpus é o primeiro representante. Trata-se de um trabalho valioso essa formação de banco de dados nas plataformas digitais, para os estudos linguísticos de maneira geral, especialmente, no que diz respeito ao CE-DOHS – considerando as perguntas sócio-históricas que embasaram sua constituição –, para os estudos da formação do português brasileiro, na área da Linguística histórica.

## REFERÊNCIAS

BACELAR DO NASCIMENTO, M. F. O lugar do corpus na investigação linguística. Disponível em: <http://www.clul.ul.pt/equipa/berlim-2000-nascimento.pdf>. Acesso em: 20 abr. 2004.

CASTILHO, A. T. de (Org.). Para a história do português brasileiro: primeiras idéias. São Paulo: Humanitas Publicações / FFLCH/USP, 1998, v.1. 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.

MATTOS E SILVA, R. V. Para a história do português culto e popular brasileiro: sugestões para uma pauta de pesquisa. In: ALKMIM, Tânia M. Para a história do português brasileiro: novos estudos. São Paulo: Humanitas/FFCHL/USP: FAPESP, 2002, v. 2, p. 443-464.

MATTOS E SILVA, R. V. Ensaios para uma sócio-história do português brasileiro. São Paulo: Parábola, 2004.

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.

SANTOS, E. B. *O Livro do Gado do Brejo do Campo Seco (Bahia): edição semidiplomática e descrição de índices grafo-fonéticos*. Dissertação (Mestrado em Linguística). Feira de Santana: UEFS, 2019.

SHEPHERD, T. M. G.; SARDINHA, T. B.; PINTO, M. V. (Org.) Caminhos da Linguística de Corpus. Campinas: Mercado de Letras, 2012.