

**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**

**EDIÇÃO XML E ELABORAÇÃO DOS METADADOS DO DOCUMENTO E  
PROCESSAMENTO – DE CACHOEIRA A FEIRA DE SANTANA, BAHIA  
(1698-1881): DOCUMENTOS NOTARIAIS MANUSCRITOS**

**Kamila Moreira Souza<sup>1</sup>; Zenaide de Oliveira Novais Carneiro<sup>2</sup>;**

1. Bolsista PIBIC/CNPq, Graduanda em Letras Vernáculas, Universidade Estadual de Feira de Santana, e-mail:

[moreirakamila2@gmail.com](mailto:moreirakamila2@gmail.com)

2. Orientador, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail:

[zenaide.novais@gmail.com](mailto:zenaide.novais@gmail.com)

**PALAVRAS-CHAVE:** linguística; cedohs; edictor.

## **INTRODUÇÃO**

Os novos recursos que as atuais ferramentas tecnológicas trazem para o campo filológico, com as edições digitais, supõem uma verdadeira revolução, tanto em termos de armazenamento textual como em termos de elaboração de *corpora* com diferentes objetivos, como, por exemplo, quantificação de ocorrências, anotação sintática, extração automática das datações lexicais. E ainda, para além dessas possibilidades na análise de textos antigos, como afirmam Gonçalves e Banza (2013, p. 5),

Os acervos ou bibliotecas digitais apresentam-se ao investigador, e bem assim ao público em geral, como verdadeiras bibliotecas globais, possibilitando o acesso a fontes em qualquer lugar do mundo, sem sair do escritório, fontes que podem ser colecionadas num *pen-drive* ou partilhar livremente, com a simples mediação do *notebook* ou de um *tablet*.

O meio digital abre novas possibilidades de trabalho textual, e o projeto Corpus Eletrônico de Documentos Históricos do Sertão, ao qual se vincula este plano de pesquisa, colabora para a edição eletrônica de textos escritos em português; já disponibiliza diversos acervos, sobretudo de cartas manuscritas, organizando-as por grau de escolaridade e por grau de habilidade com a escrita; são 1084 cartas particulares (1808-2000), num total de 350.850 palavras, escritas por 422 remetentes (nascidos entre 1724 e 1980), extraída a maior parte de Carneiro *et al* (2011). Na segunda fase do projeto, que está em andamento, o número de documentos tem sido ampliado, tanto manuscritos como impressos. Entre os manuscritos, está de CARNEIRO & QUEIROZ (2011), foram os *De Cachoeira a Feira de Santana, Bahia (1698-1881)*: documentos notariais manuscritos. Nossa proposta aqui é elaborar os metadados (autor, conteúdo, destinatário, local etc.) do referido documento, bem como os metadados do processamento de suas edições semidiplomática e digital (primeira revisão, número de palavras etc.)

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

O material, como dito, é os *De Cachoeira a Feira de Santana, Bahia (1698-1881)*: documentos notariais manuscritos. Faz parte do projeto CE-DOHS.

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

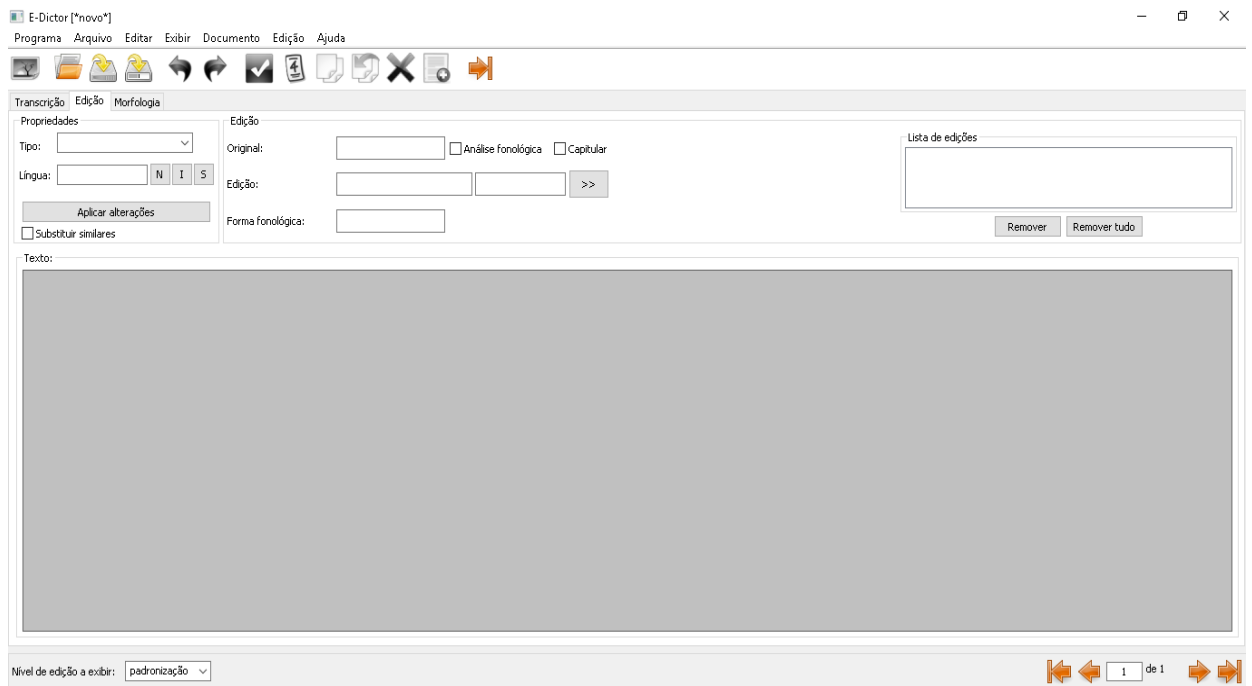
O eDicator, proposto por Paixão de Sousa, Kepler e Faria (2009), é um editor de arquivos *XML*, especialmente dedicado ao trabalho filológico e à análise linguística. A necessidade em realizar edições eletrônicas se baseia, principalmente, em realizar os estudos das mudanças linguísticas com o passar dos tempos. Surge, portanto,

“[...] a necessidade de uma ferramenta de anotação específica para textos antigos surgiu da nossa avaliação de que era preciso tornar o sistema concebido para a codificação do CTB mais amigável e confiável, ao mesmo tempo em que preservasse as vantagens de uma codificação especificamente voltada para edições filológicas”. (PAIXÃO DE SOUSA, KEPLER E FARIA, 2009, p.5)

À vista disso, o programa foi criado, entre outras funções, para auxiliar na conservação dos materiais escritos, garantindo a fidelidade com o arquivo original, além de contribuir para o processamento automático dos textos. A edição eletrônica, executada durante o período de vigência da bolsa, foi de testamentos, certidões com teor de testamento, inventários e declarações do arquivo Municipal de Cachoeira e do Centro de Pesquisa e Documentação de Feira de Santana da UEFS, totalizando 26 documentos oficiais, abrangendo o século XVII ao XIX.

Os documentos apresentam-se nos seus arquivos de origem em folhas avulsas escritas com pena de ave em papéis almaços soltos (ou amarrados por linhas no dorso esquerdo). O tipo de tinta utilizado varia, sendo as mais comuns, a de nós gaélica ou ainda as que têm ferro em sua base. O tipo de letra mais comum é a cursiva que se caracteriza pelo traçado livre e pessoal [...] E apresentam ainda muitas abreviaturas, sendo as mais utilizadas a do tipo contração ou síncope (supressão de letras intermediárias e preservação da letra inicial e das finais), além de algumas siglas. (CARNEIRO & ALMEIDA, 2002, p. 468)

No decorrer do processo de edição eletrônica dos arquivos, buscou-se, ao máximo, manter fielmente as informações dos documentos originais. Um conjunto de regras é adotado durante o procedimento, para que não tenham divergências em relação a outras edições. Cada documento é singular, então as etiquetas e formatações variam de acordo com as necessidades. Durante a edição é possível visualizar a palavra original e os processos realizados, na parte superior do programa, na lista de edições, os quais dependem do tipo de edição feita.



**Figura 1: Interface eDictor**

Na figura 1 é possível observar a interface do programa eDictor. Após a transcrição e gerar *XML*, as propriedades dos documentos são estabelecidas, seguindo para as padronizações ortográficas que são feitas por intermédio das etiquetas. Desta maneira, várias etiquetas foram utilizadas, como a padronização de “Deos”, “assignou”, “pesso”, entre outras variantes.

Já os metadados são divididos em original e processamento e preenchidos no próprio eDictor, na parte “Metadados CTRL+ALT+M”. As informações do original são equivalentes aos documentos: acervo, autor, código do autor, conteúdo, data de escrita, data de nascimento, escolaridade, etnia, grau de habilidade de escrita, local de escrita, local de nascimento, número da carta, para quem, referência, relação sexual, sexo, tipo documental. Os metadados de processamento contam com informações: 1ª revisão da edição XML, 2ª revisão da edição XML, edição semidiplomática, edição XML, revisão da edição semidiplomática, revisão final XML.

Os textos presentes *De Cachoeira a Feira de Santana, Bahia (1698-1881)*: documentos notariais manuscritos possuem fac-símiles, o que contribuiria para uma identificação mais precisa e consulta ao documento original. Entretanto, eles se encontram no Acervo do Monsenhor Renato Galvão, que devido à pandemia do Covid-19, tornou-se inviável o acesso por não possuir digitalização. Com a normalização dos expedientes, os arquivos serão organizados correspondendo à edição realizada e publicados no acervo do CE-DOHS. Ainda, é importante salientar que durante o período de distanciamento social, a pesquisa foi desenvolvida por meio de trabalhos remotos, utilizando como recurso reuniões através do Google Meet e o desenvolvimento das atividades regularmente em Home Office.

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

Os documentos notariais manuscritos oficiais, que compreendem o século XVII ao XIX, podem inferir em estudos linguísticos e extralinguísticos, podendo compor dados que expressam variantes históricas diversas do português brasileiro, ainda que

escassas. Sendo, desta forma, importante a edição dos arquivos para a disponibilização de dados a pesquisas em diversas áreas. Ampliando o banco de dados do CE-DOHS, agregando, junto aos *corpora* já existentes, gêneros textuais diversos e escritos em âmbitos formais, contribuindo, portanto, para “o melhor uso de maus dados” (LABOV, 1994, p. 11).

## REFERÊNCIAS

CARNEIRO, Zenaide de Oliveira Novais; ALMEIDA, Norma Lúcia Fernandes de; QUEIROZ, Rita de Cássia Ribeiro (Org.). **De Cachoeira a Feira de Santana, Bahia (1698-1881):** documentos notariais manuscritos. Feira de Santana: UEFS Editora, 2011. 1 CD-ROM : il.

CARNEIRO, Zenaide O. N. ; ALMEIDA, Norma L. F. . **Notícias sobre o banco de textos de Feira de Santana e Cachoeira.** In: Tânia Maria Alkimin. (Org.). Para a História do Português Brasileiro. São Paulo: Humanitas/Fapesp, 2002, v. 3, p. 465-488.

GONÇALVES, M. F.; BANZA, A. P. (org.). **Património textual e humanidades digitais:** da antiga à nova Filologia. Évora: CIDEHUS, 2013.

LABOV, W. **Principles of Linguistic Change.** Internal Factors. Cambridge, Massachusetts/Oxford: Blackwell Publishers. 1994. v. 1.

Paixão de Sousa, M. C.; Kepler, F. N.; Faria, P. **E-dictor:** Novas perspectivas na codificação e edição de corpora de textos históricos. In: VIII Encontro de Linguística de Corpus, 2009, Rio de Janeiro. Resumos, 2009.