

## EDIÇÃO EM LINGUAGEM XML DE PARTE DO ACERVO CARTAS PARA O BARÃO DE JEREMOABO

Sueli Meireles Conceição<sup>1</sup>; Zenaide de Oliveira Novais Carneiro<sup>2</sup>

**PALAVRAS-CHAVE:** Português Brasileiro, Linguística de *Corpus*, Edição XML.

### INTRODUÇÃO

O presente trabalho é resultado da pesquisa desenvolvida no projeto *Corpus Eletrônico de Documentos Históricos do Sertão* (CE-DOHS) FAPESB, Processo 5566/2010/Consepe:202/2010), coordenado por Zenaide de Oliveira Novais Carneiro e Mariana Fagundes de Oliveira, que trabalha em parceria com o projeto *Corpus Histórico do Português Tycho Brahe*, ([www.tycho.iel.unicamp.br](http://www.tycho.iel.unicamp.br)), do Instituto de Estudos da Linguagem (IEL), da Unicamp, coordenado por Charlotte Galves, através do Termo Aditivo de Transferência de Tecnologia.

Na pesquisa, aplicamos a linguagem XML na edição das cartas para uso em bancos eletrônicos e na elaboração de *metadados* e de informações, com base em Carneiro (2005), especificamente, do *acervo Carta para Cícero Dantas – Barão de Jeremoabo* (ABJ) que atualmente faz parte da base de dados do *Projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro* (CNPq. Processo 401433/2009-9/Consepe: 102/2009), do Núcleo de Estudos da Língua Portuguesa (NELP).

A proposta é disponibilizar para a rede mundial de computadores o acervo do *Corpus DOHS* ([www.uefs.br/dohs](http://www.uefs.br/dohs)) para estudos linguísticos voltados ao português brasileiro. A edição em linguagem XML possibilita o controle e o mapeamento das intervenções realizadas nos documentos, de forma que podemos editá-los e prepará-los para buscas automáticas de análise linguística e, ao mesmo tempo, garantir a recuperabilidade das formas originais.

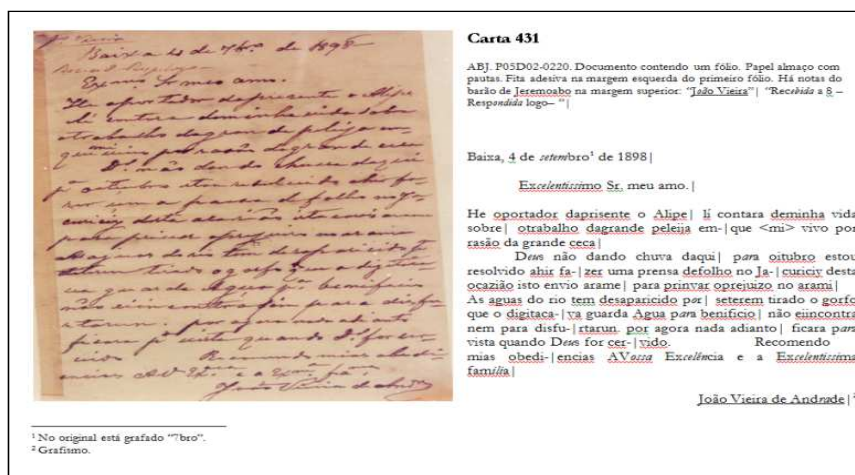
### MATERIAL E MÉTODO

O material editado pertence ao *acervo* intitulado *Cartas para Cícero Dantas Martins (Barão de Jeremoabo)*, constituído por 190 cartas, escritas entre 1880-1903, oriundas do interior da Bahia. A documentação teve sua edição fac-similada semi-diplomática desenvolvida por Carneiro (2005) (Figura 1).

---

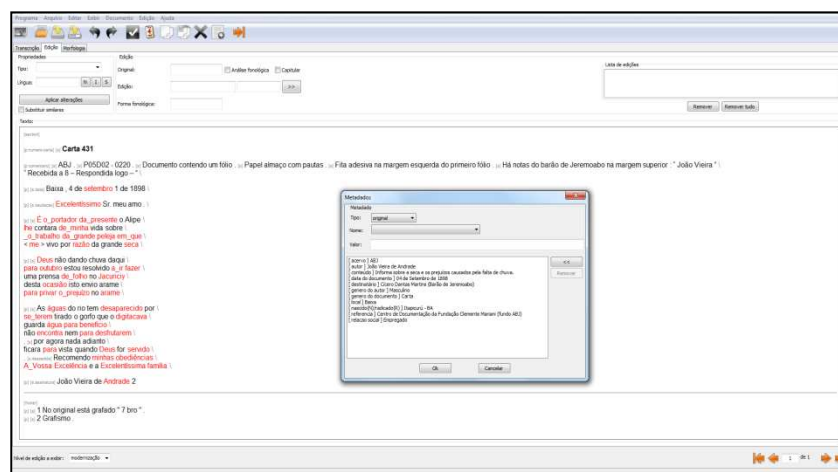
<sup>1</sup> Bolsista PIBIC/FAPESB, Graduanda em Licenciatura em Letras Vernáculas, Universidade Estadual de Feira de Santana (UEFS), bolsista IC/ FAPESB, [suelimeireles.c@gmail.com](mailto:suelimeireles.c@gmail.com).

<sup>2</sup> Professora da Universidade Estadual de Feira de Santana (UEFS) e Pós-Doutora pela Universidade Estadual de Campinas (UNICAMP), [zenaide.novais@gmail.com](mailto:zenaide.novais@gmail.com).



**Figura 1.** Edição fac-similada semi-diplomática

A metodologia que adotamos para edição e elaboração dos *metadados* dos autores/redatores do *corpus* DOHS ([www.uefs.br/dohs](http://www.uefs.br/dohs)) baseia-se na utilizada pelo Corpus Histórico do Português Tycho Brahe. A edição de uma carta em linguagem XML pode ser vista na (Figura 2) feita através do uso da ferramenta E-Dictor (PAIXÃO DE SOUZA; KEPLER; FARIA, 2009).

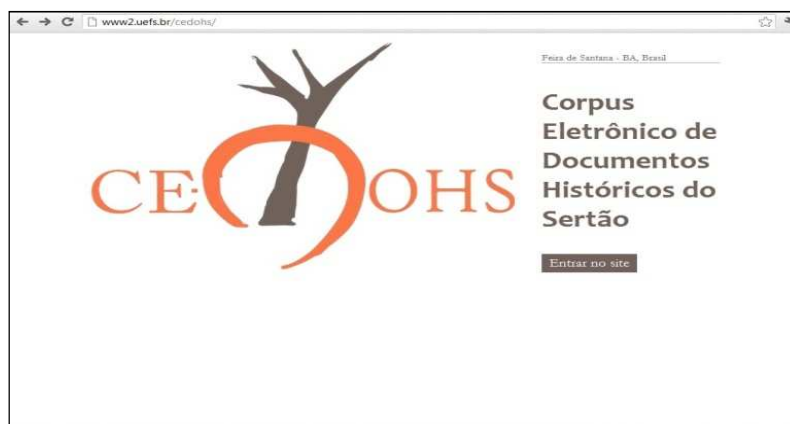


**Figura 2.** Edição e elaboração de metadados em linguagem XML

Os *metadados* têm a função de trazer informações como: nome do acervo, remetente, destinatário, conteúdo, data do documento, local de origem, gênero do documento, entre outros. Para preencher esses dados é preciso utilizar o E-Dictor, pois ele oferece diversas ferramentas, e entre elas, a ferramenta *documento*. Nessa aba encontram-se algumas opções, inclusive, a opção *metadados*. Ao selecionar *metadados*, e depois o a opção *original*, a ficha do autor/remetente é gerada e está pronta para ser preenchida. A ficha eletrônica facilita a localização do documento e o levantamento de dados para a construção do perfil biográfico.

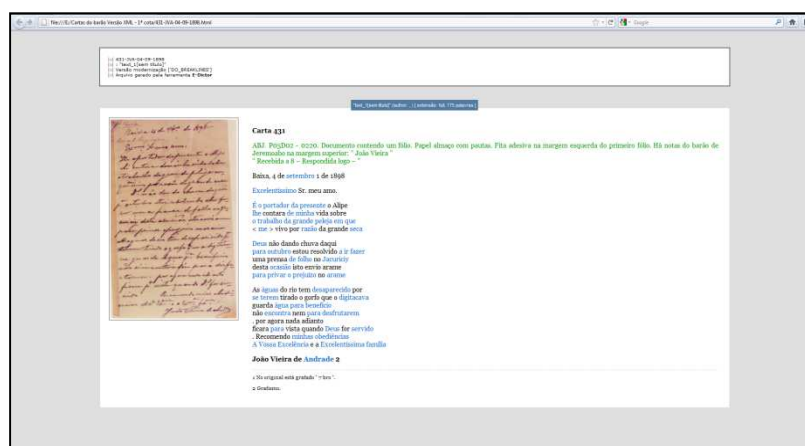
## RESULTADOS E DISCUSSÕES

Como resultados desse trabalho, temos a edição em linguagem XML e dos dados dos autores/remetentes de cartas que compõem o *Acervo Cartas para Cícero Dantas Martins, Barão de Jeremoabo (1880-1903)*, disponibilizado no site do CE-DOHS (<http://www.uefs.br/cedohs>) (Figura 3), já podendo ser consultada em <http://www.tycho.iel.unicamp.br/cedohs/corpora/catalog-CDM.html>.



**Figura 3.** Site do projeto Corpus Eletrônico de Documentos Histórico do Sertão

No site, são encontrados todos os materiais nos quais foram aplicados a edição em linguagem XML, além das informações sobre os documentos editados, contidas na ficha catalográfica. A edição eletrônica dos textos permite gerar diferentes versões de edição, como a versão *modernizada* (Figura 4), *texto simples*, *léxico de edições*, entre outras.



**Figura 4.** Modelo edição XML na *versão modernizada*

## CONSIDERAÇÕES FINAIS

O objetivo desta nova edição, feita em linguagem XML, é disponibilizar este material para pesquisadores interessados em estudar a história do PB. Com base nesse tipo de edição é também possível fornecer base para aplicação de anotação morfológica e sintática de forma automática para busca por pesquisadores na área da Linguística Histórica.

## REFERÊNCIAS

CARNEIRO, Zenaide. *Cartas Brasileiras: um estudo lingüístico-filológico*. Tese de Doutorado, Campinas: Unicamp, 2005.

CE-DOHS – DOCUMENTOS HISTÓRICOS DO SERTÃO EM DADOS (disponível em <http://www.uefs.br/cedohs>).

CORPUS DOHS. Documentos Históricos do Sertão (disponível em [www.uefs.br/dohs](http://www.uefs.br/dohs)), 2010.

MATTOS E SILVA, Rosa Virgínia. (2002). 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, v. 2, p. 443-464.

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. (a sair em: Shepherd, T., Berber Sardinha, T. e Veirano Pinto, M. (2009) (Org.). *Linguística de Corpus: Sínteses e Avanços*. Anais do VIII Encontro de Linguística de Corpus, realizado na UERJ, 13 a 14 de novembro de 2009. Rio de Janeiro, RJ, 2009. p. 69-105.

PROJETO VOZES DO SERTÃO EM DADOS (disponível em [www.uefs.br/nelp](http://www.uefs.br/nelp)), 2010.

SANTIAGO, Huda. *Cartas populares*, (no prelo).