

APLICAÇÃO DA EDIÇÃO XML EM CARTAS OITOCENTISTA DO INTERIOR DA BAHIA

Janaina de Oliveira Costa Mascarenhas¹; Zenaide de Oliveira Novais Carneiro²

1. Janaína Oliveira Costa Mascarenhas, bolsista de iniciação científica FAPESB, Graduanda em Letras Vernáculas, Universidade Estadual de Feira de Santana e-mail: janaina.mascarenhas@yahoo.com.br
2. Zenaide de Oliveira Novais Carneiro, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: zenaide.novais@gmail.com

PALAVRAS-CHAVE: Linguística de *Corpus*; Português Brasileiro; Século XIX. Edição XML.

INTRODUÇÃO

O objetivo deste trabalho é fazer edições em linguagem XML de cartas pessoais do século XIX extraídas do Banco de Dados (DOHS) do Projeto *Vozes do Sertão em Dados: história, povos e formação do português brasileiro* (CNPq. Processo 401433/2009-9/Consepe: 102/2009) editadas em versão semi-diplomáticas para estudo do português brasileiro. A edição XML é feita para uso em bases eletrônicas na forma como desenvolvida no Projeto CE-DOHS - *Corpus Eletrônico de Documentos Históricos do Sertão* (www.uefs.br/cedohs), (FAPESB, Processo 5566/2010/Consepe:202/2010), coordenado por Zenaide de Oliveira Novais Carneiro e Mariana Fagundes de Oliveira, sediado no Núcleo de Estudos de Língua Portuguesa (NELP), na Universidade Estadual de Feira de Santana (UEFS).

MATERIAIS E MÉTODOS

As cartas utilizadas neste trabalho foram editadas em versão semi-diplomática na tese de doutoramento de Carneiro (2005), conforme visualizada na Figura 1.

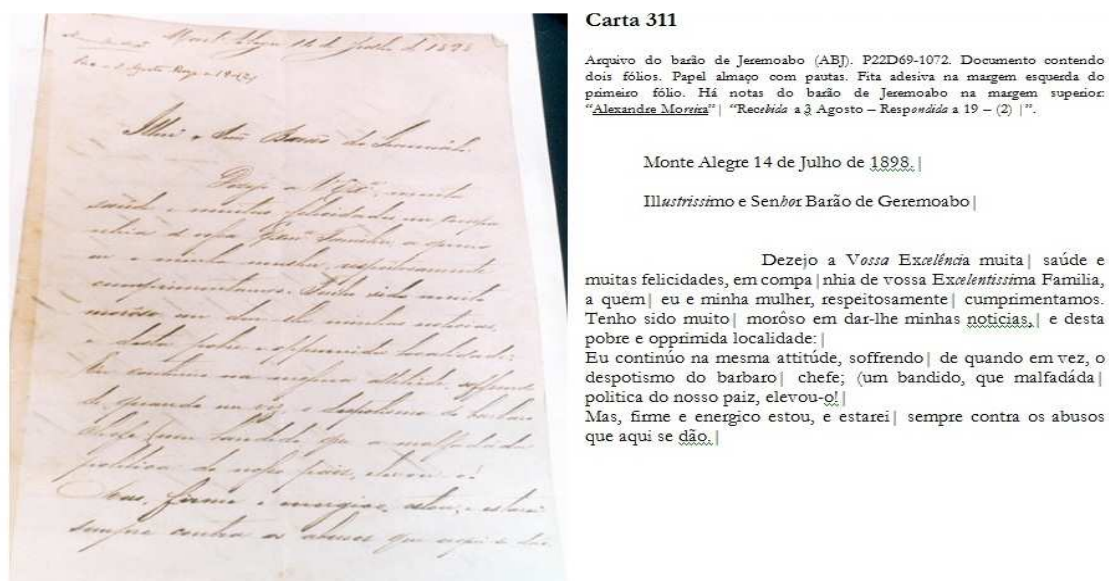


Figura1: Edição fac-similada semi-diplomática

As cartas datadas de 1880 a 1903 são oriundas do interior da Bahia e foram escritas por 43 sertanejos para o Barão de Jeremoabo e fazem parte do Fundo ABJ, depositadas no Centro de Documentação da Fundação Clemente Mariani. Trata-se de cartas manuscritas que abordam temas diversos, como, por exemplo, política, seca e finanças.

Neste trabalho, editamos em versão XML 95 cartas de um total de 190. O método empregado nessa tarefa científica é o da Linguística de Corpus, em versão eletrônica a partir do uso da linguagem XML. A edição XML é feita através do uso da ferramenta computacional E-dictor (PAIXÃO DE SOUZA; KEPLER; FARIA, 2009) possibilitado através do Termo Aditivo de Transferência de Tecnologia do *Corpus Histórico do Português Tycho Brahe* (www.tycho.iel.unicamp.br), sediado no Instituto de Estudos da Linguagem (IEL) da Universidade Estadual de Campinas/UNICAMP, coordenado por Charlotte Marie Chambelland Galves.

Vejamos o exemplo na Figura 2.

RESULTADOS E/ OU DISCUSSÃO

A partir da edição semi-diplomática (cf. figura 1), foi feita a edição em versão XML com uso do E-dictor, como exemplificado na Figura 2.

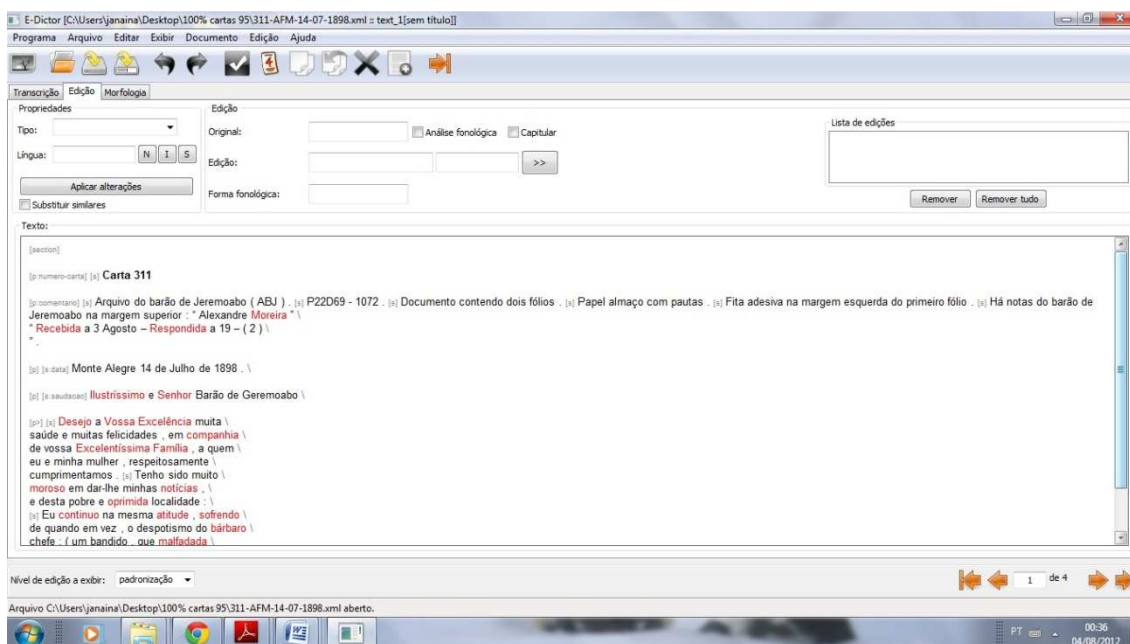


Figura2: Edição em linguagem XML

Esse trabalho feito em 95 cartas. O resultado pode ser visto em <http://www.tycho.iel.unicamp.br/cedohs/corpora/catalog-CDM.html>. Nessa página é possível gerar automaticamente edições em vários formatos, semi-diplomática, em versão XML, edição modernizada como na Figura 3 e técnica. O usuário tem acesso também aos metadados e o léxico de edições.

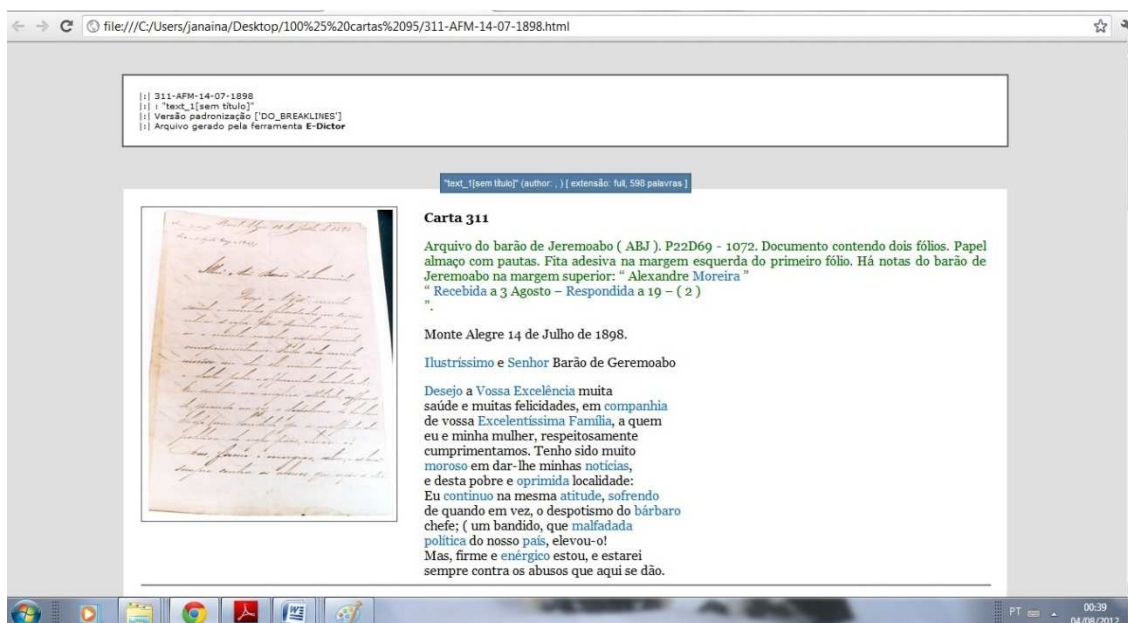


Figura 3: Edição fac-similada em linguagem XML

CONSIDERAÇÕES FINAIS

O acervo editado está disponível no site do projeto em (<http://www.uefs.br/cedohs/>) e é de fácil acesso aos usuários e pesquisadores. Cabe salientar que essa edição será utilizada em uma segunda fase para anotação morfológica e sintática permitindo a busca automática para proveito dos estudiosos da área.

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://www2.uefs.br/cedohs/>), 2011.

CORPUS DOHS. Documentos Históricos do Sertão (disponível em <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.

MATTOS e SILVA, Rosa Virgínia. *O português são dois: novas fronteiras, velhos problemas*. São Paulo: Parábola, 2004

PAIXÃO DE SOUZA, M.C., KEPLER, F.N. & FARIA, P. (a sair) *"E-Dictor: novas perspectivas na codificação e edição de corpora de textos históricos"*. In: 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.

PAIXÃO DE SOUSA, M.C. “*Memórias do Texto*”. Revista Texto Digital. Universidade Federal de Santa Catarina: 2006.

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