

## EDIÇÃO ELETRÔNICA DE DOCUMENTOS HISTÓRICOS DO SERTÃO USANDO O E-DICTOR

**Maiany Soares de Oliveira<sup>1</sup>; Mariana Fagundes de Oliveira**<sup>2</sup>

1. Bolsista FAPESB, Graduanda em Letras Vernáculas, Universidade Estadual de Feira de Santana, e-mail: [maianyso@gmail.com](mailto:maianyso@gmail.com)
2. Orientador, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: [marianafag@gmail.com](mailto:marianafag@gmail.com)

**PALAVRAS-CHAVE:** Português Brasileiro, Cartas Particulares, Edição eletrônica.

### INTRODUÇÃO

A Linguística Histórica *stricto sensu*, cujo objeto de estudo é a mudança linguística, depende do trabalho da Filologia, que é a “ciência do texto”. Não se pode fazer Linguística Histórica *stricto sensu*, segundo Mattos e Silva (2008, p. 14), sem a documentação remanescente do passado, criteriosamente editada, com a finalidade de estudos linguísticos.

Neste trabalho, apresenta-se a edição, em linguagem XML, usando o E-Dictor, desenvolvido por Kepler, Paixão de Souza e Faria (2007), de cartas particulares, da segunda metade do século XX, escritas por brasileiros falantes de PB popular ou semipopular, culto ou semiculto. O objetivo, com a realização desta edição, é constituir *corpus eletrônico* para estudo da história do PB.

Segundo Mattos e Silva (2004, p. 63), para recuperar uma história do PB, é preciso constituir um conjunto significativo de documentação representativa tanto do chamado português popular do Brasil como do português culto do Brasil. A segunda metade da década de 1990 inaugura, de acordo com Lobo (2009, p. 307), uma fase caracterizada pela laboriosa tarefa de construção de uma filologia de textos escritos no Brasil, ponto de partida, conforme a autora, incontornável para a descrição e análise das mudanças linguísticas que foram configurando o PB.

### MATERIAL, MÉTODOS OU METODOLOGIA

O material editado faz parte do Banco de dados do Projeto *Vozes do Sertão*, denominado *Documentos Históricos do Sertão* (DOHS), composto de documentos históricos do Semi-Árido baiano escritos em língua portuguesa, do século XVII ao século XX.

A documentação editada, uma documentação pessoal, de Valente, no Semiárido baiano, é constituída de 79 cartas, 25 cartões, 80 envelopes e 2 fotografias; a carta mais antiga é datada de 1980, e a mais recente, de 1993. A maior parte das 79 cartas são manuscritas em papel com pautas, havendo poucas cartas impressas e uma minoria sem envelope. Os remetentes – brasileiros, sobretudo baianos, semipopulares e semicultos –

são, a maioria deles, mulheres. As cartas são destinadas, sobretudo, a Adelmário Carneiro Araújo, natural de Valente-Bahia.

A edição, em linguagem XML, dos documentos é feita usando o E-dictor. O E-Dictor, um programa computacional, foi desenvolvido por Kepler, Paixão de Souza e Faria (2007), para facilitar a edição eletrônica, em linguagem XML, de documentos antigos. Esse programa amplia o alcance da linguagem XML e consegue unir a edição e a correção morfológica do texto em um mesmo ambiente, além de mapear as intervenções realizadas pelo editor no documento. A primeira versão dessa ferramenta surgiu para atender a uma demanda obtida durante a criação do Projeto *Corpus Anotado do Português TychoBrahe* (CTB/UNICAMP) e em atividades da equipe deste projeto com a equipe do *Projeto para a História da Língua Portuguesa* (PROHPOR/UFBA). Várias ferramentas foram testadas, como o *Multext*, *CLaRk*, e editores de texto, *Emacs*, *Kate*, *EditPlus* etc., mas elas não atendiam a necessidades específicas da edição; logo, foi planejada uma ferramenta de anotação específica para textos antigos. Esta viria a possibilitar a integração entre o sistema de edição e o sistema de correção da anotação morfológica. Além desses recursos, o E-Dictor permite que outros conteúdos sejam codificados, como: metadados, elementos do texto em geral, classes de palavras, níveis de edição filológica e comentários do editor, enriquecendo o material. Mesmo sendo desenvolvida a partir do CTB, o E-Dictor é flexível o suficiente para contribuir em outros contextos de construção de *corpora* de textos. O nosso objetivo é colaborar no transporte de documentos, que estão fac-similados em *Word*, para o XML, visando à composição do *corpus* eletrônico do Projeto CE-DOHS *Corpus Eletrônico de Documentos Históricos do Sertão* ([www.uefs.br/cedohs](http://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). Trata-se da edição em formato XML, via utilização da ferramenta computacional E-dictor (PAIXÃO DE SOUZA; KEPLER; FARIA, 2009), através do Termo Aditivo de Transferência de Tecnologia do *Corpus Histórico do Português Tycho Brahe* ([www.tycho.iel.unicamp.br](http://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.

## RESULTADOS E/OU DISCUSSÃO

A seguir, apresenta-se a edição, em linguagem XML, de duas das 79 cartas que foram editadas: a primeira, escrita em Cícero Dantas, interior da Bahia, em 25 de setembro de 1980, por Maria Lúcia B. Araújo, com Ensino Fundamental incompleto, falante de um português semipopular; a segunda, escrita em Salvador, no dia 26 de setembro de 1980, por Ângela Margarida Mesquita, estudante do curso de Contabilidade, falante de um português semiculto.

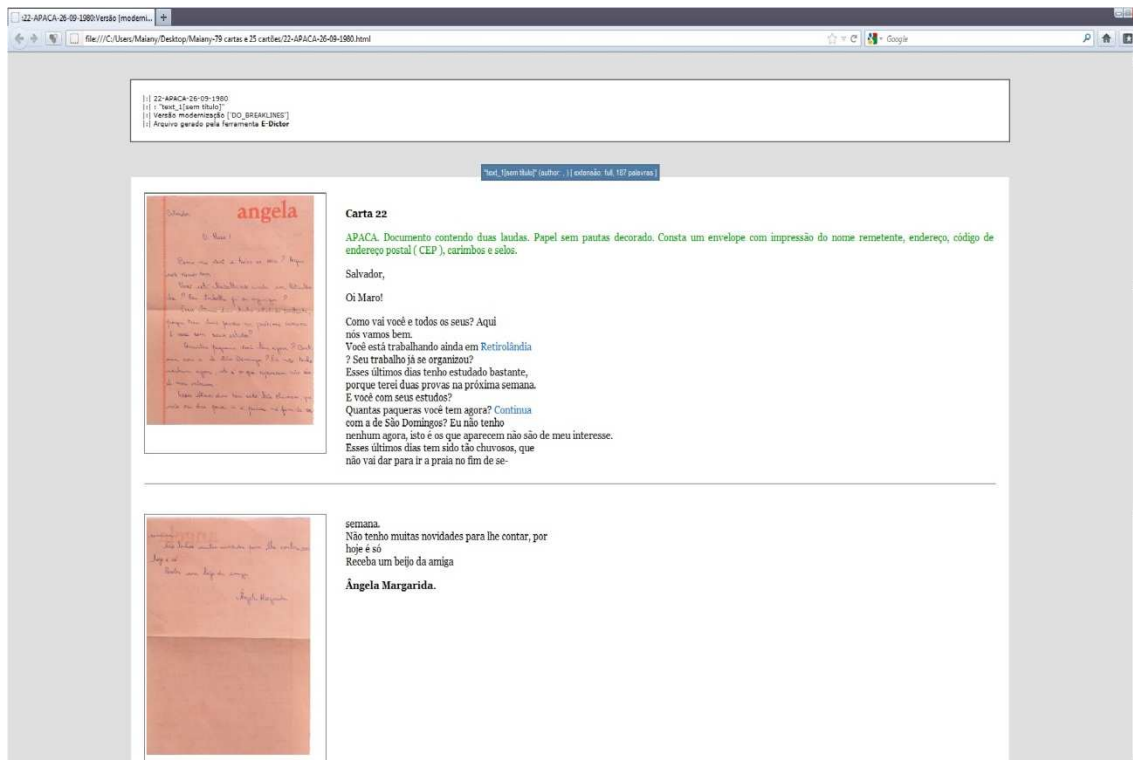


Figura 1: Edição Eletrônica da carta de Ângela Margarida Mesquita.

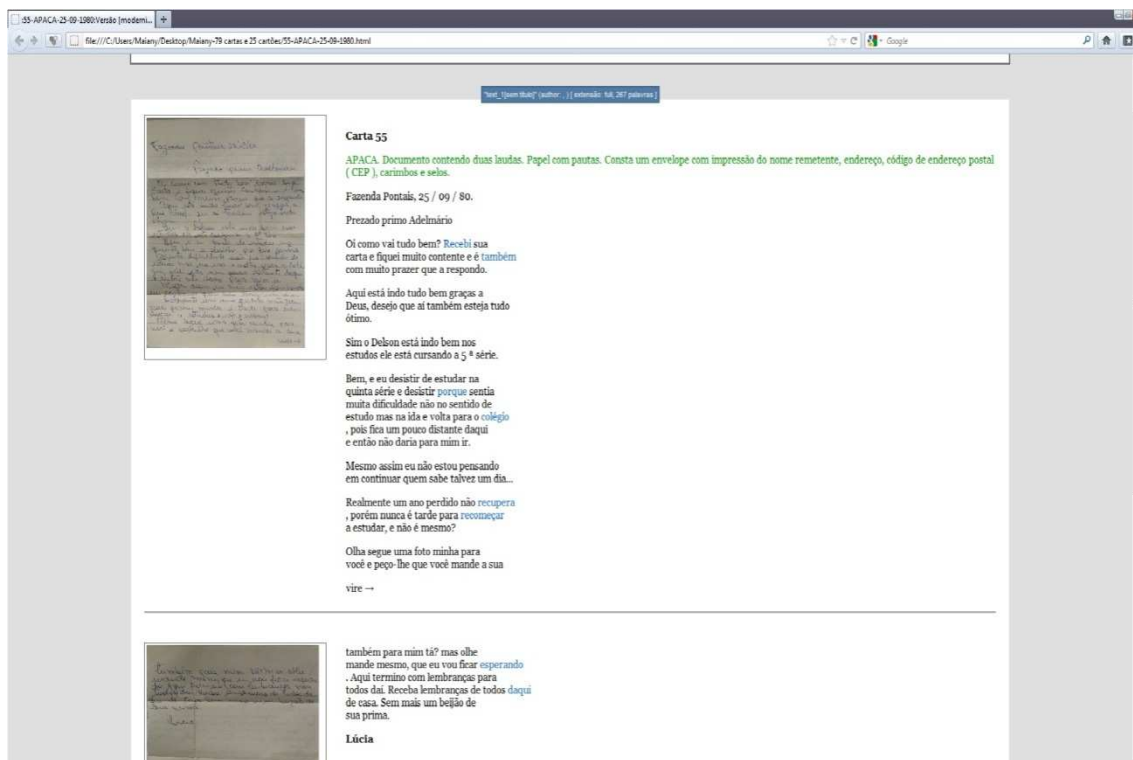


Figura 2: Edição Eletrônica da carta de Maria Lúcia.

## CONSIDERAÇÕES FINAIS

Espera-se que o trabalho realizado, servindo de *corpus* de estudo do PB a pesquisadores interessados, possa colaborar para um maior conhecimento da história da língua. As cartas encontram-se disponíveis no site do CE-DOHS: <http://www2.uefs.br/cedohs/apresenta.html>

## REFERÊNCIAS

CORPUS DOHS. Documentos Históricos do Sertão (disponível em <http://www2.uefs.br/cedohs/apresenta.html>) 2010.

LOBO, Tânia. Arquivos, acervos e a reconstrução histórica do português brasileiro. In: OLIVEIRA, MATTOS E SILVA, Rosa Virgínia. *Ensaio 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: 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 CORPUS ELETRÔNICO DE DOCUMENTOS HISTÓRICOS DO SERTÃO (disponível em <http://www2.uefs.br/cedohs/apresenta.html>), 2011.

PROJETO VOZES DO SERTÃO EM DADOS (disponível em: [www.uefs.br/nelp](http://www.uefs.br/nelp) ), 2011.