



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 - 2022

Cartas de leitores e redatores da Bahia oitocentista: edição em linguagem XML com uso do eDictor

Ana Gabriela Oliveira Martins¹; Mariana Fagundes de Oliveira Lacerda²

1. Bolsista PIBIC/CNPq, Graduanda em Licenciatura em Letras Vernáculas, Universidade Estadual de Feira de Santana, e-mail: gabrielamartins2204@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; documentação impressa; edição XML.

INTRODUÇÃO

Este trabalho está vinculado ao projeto CE-DOHS: Corpus Eletrônico de Documentos Históricos do Sertão/DLA/UEFS, que 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, 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 para estudo linguístico. Como se vê, “Do feliz conagraamento 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 norte-americanos 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. Embora até hoje este *corpus* seja largamente utilizado, 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.”

Atendendo a uma das agendas do projeto CE-DOHS, aquela que diz respeito aos textos impressos, apresenta-se, aqui, a edição modernizada, em linguagem XML, usando o eDictor, de 80 cartas de leitores e redatores publicadas em jornais baianos, no século XIX, datadas entre 1830 e 1893. Essa documentação, em edição semidiplomática, realizada no âmbito do Projeto para a História do Português Brasileiro (PHPB), foi

publicada na obra intitulada *Críticas, queixumes e bajulações à Imprensa Brasileira do século XIX: Cartas de Leitores*, organizada por Barbosa e Lopes (2006).

MATERIAL E MÉTODOS

Trata-se de 80 cartas de leitores e redatores publicadas em jornais baianos, no século XIX, datadas entre 1830 e 1893. Essa documentação, em edição semidiplomática, realizada no âmbito do Projeto para a História do Português Brasileiro (PHPB), foi publicada na obra intitulada *Críticas, queixumes e bajulações à Imprensa Brasileira do século XIX: Cartas de Leitores*, organizada por Barbosa e Lopes (2006). A partir da edição semidiplomática das cartas, será feita a edição em linguagem XML, usando o eDictor, um editor de textos desenvolvido por Paixão de Souza, Kepler e Faria (2009). A linguagem XML permite a manipulação eletrônica do *corpus*, por meio de anotação e do uso de ferramentas de buscas de dados linguísticos.

Figura 1: Capa do livro *Críticas, Queixumes e Bajulações na Imprensa Brasileira do Século XIX: Cartas de Leitores e Redatores*.



RESULTADOS E DISCUSSÃO

O trabalho filológico tradicional enfrentou grandes limitações tecnológicas para realizar a edição e conservação dos escritos antigos, principalmente em razão dos altos custos, ocupação de muito espaço físico e alto índice de falhas nas edições em função das limitações humanas. Contudo, com os constantes avanços tecnológicos, sobretudo no âmbito das humanidades digitais, esse ofício foi ganhando suportes tecnológicos mais desenvolvidos, de modo que hoje esse trabalho tem sido computadorizado e conta com a rede mundial de computadores para o armazenamento de *Corpora*, busca automática de dados e conservação da originalidade do texto.

Concomitantemente aos avanços tecnológicos, programas foram desenvolvidos e aprimorados para facilitar a edição de documentos. Como resultado de tais aprimoramentos, surge o eDictor, programa criado por Paixão de Sousa, Kepler e Faria (2009), de edição eletrônica e registro morfológico feito para *corpora* em formato XML, que viabiliza uma melhora substancial nas edições do texto, cumprindo a premissa de conservar o texto sem perder sua originalidade.

Nesse cenário, o CE-DOHS, banco de textos do Núcleo de Estudos de Língua Portuguesa (NELP), realiza esse trabalho de edição de documentos antigos, através da aliança entre a filologia digital e a computação moderna. Reúne dois conjuntos de textos; a documentação impressa aqui apresentada integra o primeiro conjunto (XIX-XX). O segundo conjunto traz documentos do período XVI-XVIII.

No processo de edição modernizada dos textos (a partir da edição diplomática ou semidiplomática do documento), o CE-DOHS apenas faz intervenções na ortografia (modernização de baixo grau). Para ilustrar o processo de edição modernizada, em linguagem XML, com uso do eDictor, da coleção documental Cartas de Leitores e Redatores (XIX) – disponível em <<http://www.uefs.br/cedohs/view/coletaneas.html>> –, apresenta-se, na sequência, as intervenções, de natureza ortográfica, realizadas na carta 6, uma das cartas da coleção na qual se fizeram mais intervenções:

Figura 2: Print da carta 6, no eDictor.



Abaixo, o quadro com a descrição das intervenções ortográficas realizadas na carta 6 da coleção documental em questão:

Forma original	Níveis de edição	Forma modernizada
n'ella	Padronização	Nela
hum	Padronização	Um
avizo	Padronização	Aviso
meo	Padronização	Meu
ninguem	Padronização	Ninguém
commigo	Padronização	Comigo
negocio	Padronização	Negócio
vencer-se á	Junção/padronização	Vencer-se-á
tranzação	Padronização	Transação
inda	Expansão	Ainda
n'aquella	Padronização	Naquela
n'estes	Padronização	Nestes
effeitos	Padronização	Efeitos
tãobem	Segmentação/Padronização	Tão bem
vendeo	Padronização	Vendeu
Ouvedoria	Padronização	Ouvidoria
Civel	Padronização	Civil
cartorio	Padronização	Cartório
Tabellião	Padronização	Tabelião
vendí	Padronização	Vendi

recahir	Padronização	Recair
he	Padronização	É
calumnio	Padronização	Calúnia
calumnioso	Padronização	Calunioso
ella	Padronização	Ela

CONSIDERAÇÕES FINAIS

O CE-DOHS vem, ano após ano, superando as expectativas iniciais – que eram apenas de edições de documentos do sertão da Bahia – e galgando novos horizontes no que se refere ao armazenamento de dados para pesquisa linguística. O presente trabalho, que se propôs a editar 80 cartas de leitores e redatores publicadas em jornais baianos, no século XIX, datadas entre 1830 e 1893, são documentos inéditos, que fazem parte do rico banco de textos do NELP/UEFS.

O banco CE-DOHS, antes da realização da presente pesquisa, dispunha apenas de documentos impressos do século XX; a partir de agora, com os resultados alcançados aqui, dispõe também de documentos impressos do século XIX, ampliando o conjunto de materiais representativos do português brasileiro em oitocentos.

REFERÊNCIAS

- BARBOSA, A.; LOPES, C. *Críticas, queixumes e bajulações na Imprensa Brasileira do século XIX: cartas de leitores*. Rio de Janeiro: Ed. da UFRJ, 2006.
- CASTILHO, A. T. de (Org.). *Para a história do português brasileiro: primeiras idéias*. São Paulo: Humanitas Publicações / FFLCH/USP, 1998, v.1.
- CORPUS CE-DOHS. *Corpus Eletrônico de Documentos Históricos do Sertão* (FAPESB 5566/2010 - Consepe UEFS 202/2010). Coordenado por Zenaide de Oliveira Novais Carneiro e Mariana Fagundes de Oliveira Lacerda (UEFS). [Projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro (CNPq. 401433/2009-9 - Consepe UEFS 102/2009). (CNPq. Processo 401433/2009-9/Consepe: 102/2009) Disponível em: <<http://www5.uefs.br/cedohs/>>. Acesso em 21 jan. 2021.
- GONÇALVES, M. F.; BANZA, A. P. (org.). *Patrimônio textual e humanidades digitais: da antiga à nova Filologia*. Évora: CIDEHUS, 2013.
- MATTOS E SILVA, R. V. *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: 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.
- SHEPHERD, T. M. G.; SARDINHA, T. B.; PINTO, M. V. (Org.) *Caminhos da Linguística de Corpus*. Campinas: Mercado de Letras, 2012.