

## EDIÇÃO CONTROLADA DE DOCUMENTOS DO SERTÃO BAIANO DO SÉCULO XX (2ª METADE)

**Priscila Starline Estrela Tuy Batista<sup>1</sup>; Igor Leal Souza<sup>2</sup>; Zenaide de Oliveira Novais Carneiro<sup>3</sup>**

1. Bolsista PROBIC, Graduando em Licenciatura em Letras Vernáculas, Universidade Estadual de Feira de Santana, e-mail: [priscilatuy@gmail.com](mailto:priscilatuy@gmail.com)
2. Bolsista FAPESB, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: [igorengcomp@gmail.com](mailto:igorengcomp@gmail.com)
3. Orientadora, 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 Computacional, Edição Eletrônica, Português Brasileiro.

### INTRODUÇÃO

Nesta pesquisa, usamos a 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) ([www.uefs.br/nelp](http://www.uefs.br/nelp)), coordenado por Zenaide de Oliveira Novais Carneiro, banco DOHS ([www.uefs.br/dohs](http://www.uefs.br/dohs)), especificamente em um subconjunto de textos da 2ª metade do século XX, o APJCPV – Acervo Particular Dr. João da Costa Pinto Victorio e a convertemos em formato eletrônico no âmbito 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). Estamos desenvolvendo o trabalho de edição em linguagem XML, utilizando a ferramenta computacional E-Dictor (PAIXÃO DE SOUZA; KEPLER; FARIA, 2009). Essa ferramenta amplia o alcance da linguagem XML e consegue unir a edição do texto e a correção em um mesmo ambiente, além de mapear as intervenções realizadas. A linguagem XML possibilita a codificação completa dos textos, tanto no cabeçalho, com fins de informações, catálogo e busca, quanto na estrutura dos textos, como paragrafação e paginação.

### METODOLOGIA

Nossa metodologia baseia-se naquela utilizada pelo *Corpus Histórico do Português Tycho Brahe*, composto por um *corpus* eletrônico anotado de textos em português, escritos por autores nascidos entre 1435 e 1845, desenvolvido desde 1998 em <http://Tycho.iel.unicamp.br/~tycho/corpus/>, sítio eletrônico no qual estão definidos os modelos e as ferramentas que estão subsidiando o projeto CE-DOHS, um *corpus* voltado a um banco de dados inédito de textos do sertão baiano, 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

Temos como resultado a edição em XML (Figura 1) de 30 das 102 cartas do acervo citado, o qual, em uma primeira versão, encontra-se uma edição Semi-Diplomática com o Fac-Símile (Figura 2). Esse é um trabalho minucioso, exigindo que o editor esteja sempre atento. Ao término de cada edição, antes de ser destinado à alimentação do banco de dados, o documento é revisado pela orientadora, para garantir que não haja erros na edição. O XML disponibiliza as versões Diplomática e Semi-Diplomática do documento, além da versão técnica, da ficha com metadados e do léxico de edições.

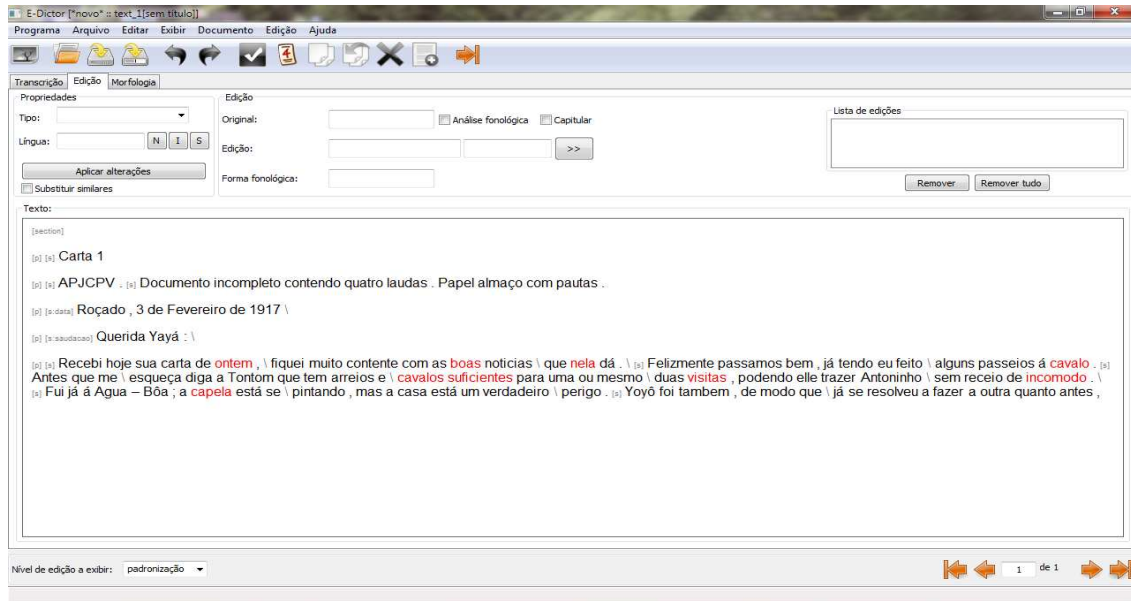


Figura 1: Modelo de edição em XML utilizando o E-Dictor.

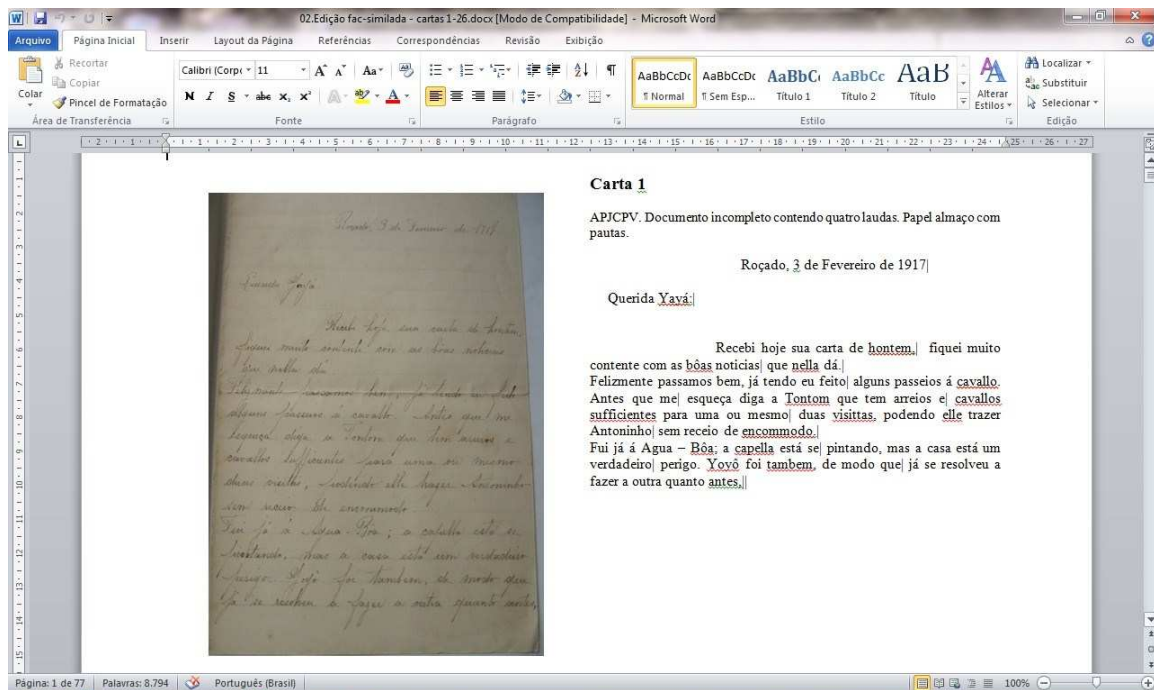


Figura 2: Modelo da edição Semi-Diplomática fac-similada.

Em paralelo ao processo de edição do APJCPV, finalizamos a revisão para publicação do Acervo *PUBLICA-SE EM FEIRA DE SANTANA: Dos anúncios e das cartas de leitores e redatores no “O Progresso” e no “Folha do Norte” (1901-2006)*. Esse acervo, também resultado de bolsistas anteriores, compõe o *corpus* do banco de dados DOHS, futuramente será editado em linguagem XML e disponibilizado no site do CE-DOHS. Houve, ainda, a colaboração na revisão do acervo *Correspondências amigas: o acervo de Valente, Bahia (1980-1993)*, Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português / Zenaide de Novais Carneiro (Org.).

Também como resultado do trabalho e dos estudos, realizamos a apresentação de painel intitulado “Edição fac-similada de cartas do Arquivo Particular de Dr. João da Costa Pinto Victória: versão XML”, no Castilho - II Congresso Internacional de Linguística Histórica, ocorrido de 07 à 10 de fevereiro de 2012, na Universidade de São Paulo (USP). Além disso, o trabalho foi apresentado na modalidade de

comunicação, no *I Congresso Internacional de Estudos Filológicos*, realizado de 29 de julho a 02 de agosto de 2012, na Universidade Federal da Bahia (UFBA).

### CONSIDERAÇÕES FINAIS

O objetivo desta nova edição, feita em linguagem XML, é disponibilizar este material para pesquisadores interessados em estudar a história do Português Brasileiro (PB). Visamos à composição de um banco eletrônico de dados, o CE-DOHS – Corpus Eletrônico de Documentos Históricos do Sertão (<http://www2.uefs.br/cedohs/>), a partir do corpus do DOHS – Documentos Históricos do Sertão. De forma mais geral, o nosso trabalho colabora com a composição de *corpora* anotados de longo alcance de busca em parceria com o Projeto Corpus Histórico do Português Tycho Brahe, além de outros parceiros.

### REFERÊNCIAS

CARNEIRO, Z. & C. GALVES (2010) “**Variação e Gramática: Colocação de clíticos na história do português brasileiro**”, a sair em Revista de Estudos da Linguagem, UFMG.

DUARTE, Maria Eugênia Lamoglia & Callou, Dinah (org.). Para a História do Português Brasileiro – **Notícias de corpora e outros estudos – Vol. IV**. Faculdade de Letras da UFRJ/FAPERJ, Rio de Janeiro, 2002.

GALVES, C. (2010). **Periodização e competição de gramáticas: o caso do português médio**, a sair em LOBO, Tânia; CARNEIRO, Zenaide; RIBEIRO, Silvana; SOLEDADE, Juliana; ALMEIDA, Ariadne. (Orgs.) Coletânea de estudos em homenagem a Rosa Virgínia Mattos e Silva. Salvador: EDUFBA. (no prelo)

PAIXÃO DE SOUSA, M. C. & KEPLER, F. (2007). **E-Dictor: Uma ferramenta integrada para a anotação de edição e classe de palavras**. Comunicação ao VI Encontro de Lingüística de Corpus. USP. São Paulo, 7 de setembro.

PAIXÃO DE SOUSA, M.C. (2005a). **Memórias do Texto**. Participação na mesa-redonda “Bibliotecas e bancos de dados digitais de literatura”, II Simpósio Nacional de Literatura e Informática. Universidade Federal de Santa Catarina (UFSC), Florianópolis, outubro de 2005.

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

PAIXÃO DE SOUSA, M. C. & KEPLER, F. (2007). **E-Dictor: Uma ferramenta integrada para a anotação de edição e classe de palavras**. Comunicação ao VI Encontro de Linguística de Corpus. USP. São Paulo, 7 de setembro.

PAIXÃO DE SOUSA, M. C. (2007b). **Linguística de Corpus e História da Língua Portuguesa: Propostas, Resultados e Desafios**, Coordenação de Mesa Redonda no V Congresso Internacional da Associação Brasileira de Linguística – ABRALIN. Belo Horizonte, 2 de março de 2007.

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.