

# CARTAS POPULARES DO SÉCULO XX: UMA PROPOSTA DE EDIÇÃO DE CARTAS NA LINGUAGEM XML PARA USO EM BANCOS ELETRÔNICOS

**Janaina de Oliveira Costa Mascarenhas<sup>1</sup>; Zenaide de Oliveira Novais Carneiro<sup>2</sup>**

1. Universidade Estadual de Feira de Santana, bolsista de iniciação científica FAPESB, Graduanda em Letras Vernáculas, e-mail: [janaina.mascarenhas@yahoo.com.br](mailto:janaina.mascarenhas@yahoo.com.br)
2. Universidade Estadual de Feira de Santana, orientadora, Departamento de Letras e Artes, e-mail: [zenaide.novais@gmail.com](mailto:zenaide.novais@gmail.com)

**PALAVRAS-CHAVE:** Linguística de *Corpus*, Português popular brasileiro, 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](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).

## MATERIAIS E MÉTODOS

Apresentar-se-á, neste trabalho, a edição eletrônica do acervo *Cartas de Sisal: Riachão do Jacuípe, Conceição do Coité e Ichu* (1906-2006). A amostra é constituída por 91 cartas pessoais, escritas ao longo do século XX por 43 sertanejos, oriundos da zona rural dos municípios de Riachão do Jacuípe, Conceição do Coité e Ichu, localizados na região sisaleira do semiárido baiano. São cartas trocadas entre familiares, compadres, namorados e amigos, principalmente para expressar saudades, obter notícias familiares e fazer pedidos. De modo geral, são textos próximos de uma escrita cotidiana, de caráter afetivo, em que há um significativo grau de intimidade entre os remetentes e destinatários. Compartilhando um contexto sociocultural semelhante, os remetentes são lavradores, trabalham com agricultura e criação de animais; possuem baixas condições financeiras e pouca escolarização. Um conjunto de características presentes nas cartas fornece algumas pistas para perceber que os seus autores são indivíduos pouco familiarizados com a escrita, e, por isso, a amostra revela-se representativa da variedade popular do português brasileiro. As cartas contidas neste acervo foram extraídas da edição fac-similar e semidiplomática de Santiago (2011).

Segundo Santiago (2013), a construção de *corpora* que apresentem dados representativos do português popular brasileiro tem especial relevância para a Linguística Histórica, tendo em vista a dificuldade de localização de documentos que sejam mais próximos de uma escrita cotidiana, produzidos por aqueles que não possuem maior domínio das habilidades de escrita. Ela também comenta a partir de Barbosa (2007) a importância dos estudos dos textos inábeis, já que são cartas que possuem essa característica, “[...] a inabilidade de reproduzir as soluções mais fonológicas de várias convenções gráficas torna os inábeis em escrita alfabética um grupo mais que desejado pela pesquisa histórica [...]” (BARBOSA, 2007, p. 484). Então, o trabalho desenvolvido, por Santiago (2011), no acervo *Cartas de Sisal: Riachão do Jacuípe*,

*Conceição do Coité e Ichu (1906-2006)*, é uma tentativa de contribuir para suprir a lacuna de dados percebida no século XX, demonstrando que traços de inabilidade identificados em séculos anteriores são aplicáveis a textos desse período. E o trabalho de linguística de *corpus* que realizei no projeto CEDOHS, também, no mesmo acervo, só tende a avançar nos estudos sobre a constituição do português brasileiro na vertente popular, já que é um trabalho de edição eletrônica, ou seja, de linguística computacional, pois o armazenamento de dados eletronicamente, só facilita e agiliza os estudos da língua, e a ferramenta utilizada para facilitar esse trabalho foi o E-Dictor.

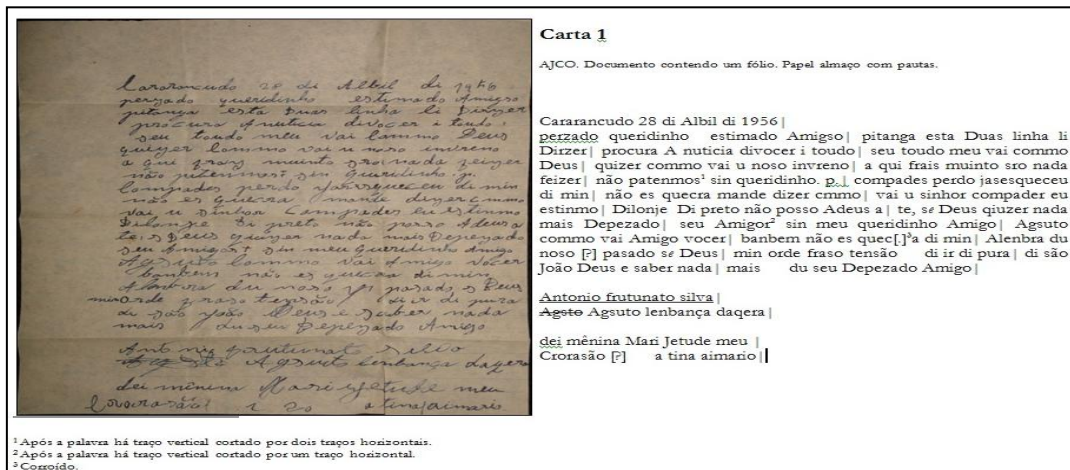


Figura1: Edição fac-similar e semidiplomática

Na edição em linguagem XML das cartas supracitadas, utiliza-se, fundamentalmente, a metodologia do projeto *Corpus* Histórico do Português Tycho Brahe, um *corpus* eletrônico anotado de textos em português, escritos por autores nascidos entre 1435 e 1845, desenvolvido desde 1998; no Tycho Brahe como no CEDOHS, utiliza-se o E-Dictor, pacote computacional desenvolvido por Kepler, Paixão de Souza e Faria (2009) possibilitado 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. Vejamos o exemplo na Figura 2:

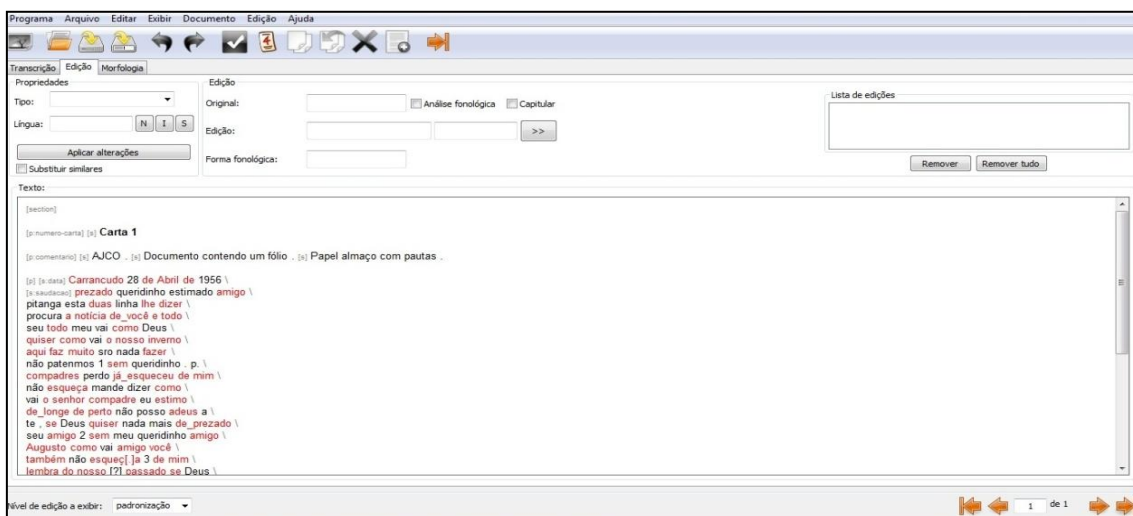


Figura2: Edição em linguagem XML

## RESULTADOS E/ OU DISCUSSÃO

Desenvolveu-se, neste trabalho, uma edição em linguagem XML, com o intuito de contribuir para a reconstituição da história da nossa língua. Essa edição foi realizada a partir da edição fac-similar e semidiplomática, feita por Santiago (2011) (cf. Figura 1), tendo como base os critérios filológicos do PHPB. A edição eletrônica do acervo de Sisal estará, em breve, disponível no site do CE-DOHS: <http://www2.uefs.br/cedohs/>. Assim, o resultado desse trabalho, feito em 91 cartas, 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. O usuário tem acesso também aos metadados e o léxico de edições. Uma amostra de alguns desses documentos podem ser visualizadas nas figuras 3, 4 e 5.

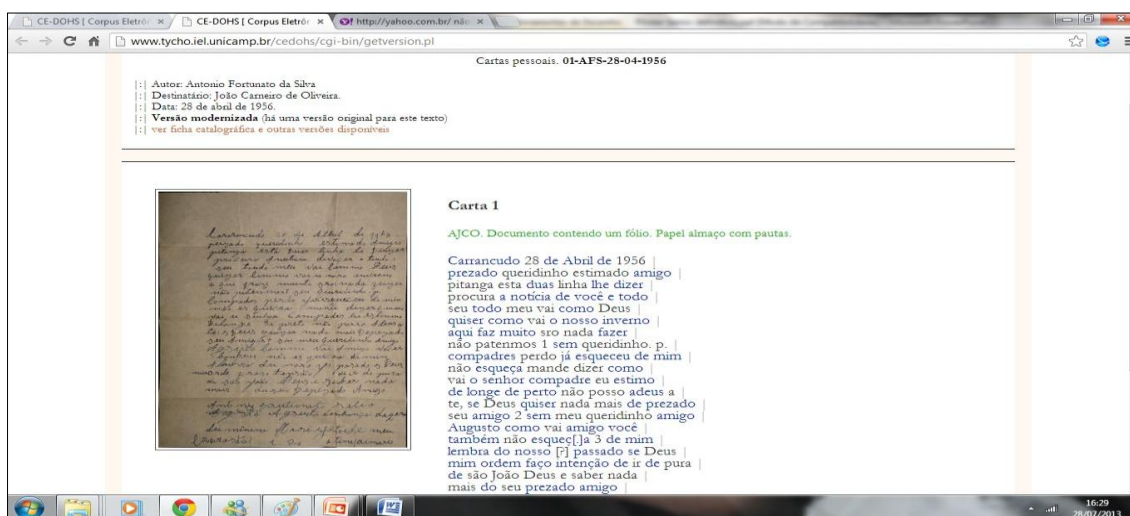


Figura 3: Edição XML modernizada



Figura 4. Metadados

Cartas pessoais: 01-AFS-28-04-1956

:| Autor: Antonio Fortunato da Silva  
 :| Destinatário: João Carneiro de Oliveira.  
 :| Data: 28 de abril de 1956.  
 :| **Lista dos itens editados** (por ordem alfabética)  
 :| ver ficha catalográfica e outras versões disponíveis

Item Original	Edições	Fonologia
A	pad: a,	
Adeus	pad: adeus,	
Agsuto	pad: Augusto,	
Agruto	pad: Augusto,	
Alhil	pad: Aheil,	
Alenbra	pad: iembra,	
Amigo	pad: amigo,	
Amigo	pad: amigo,	
Amigo	pad: amigo,	
Amigo	pad: amigo,	
Amigto	pad: amigo,	
a qui	jun: aqui,	
anbem	pad: também,	
Cararancudo	pad: Carrancudo,	
emmo	pad: como,	
como	pad: como,	
como	pad: como,	
como	pad: como,	
compader	pad: compadre,	

Figura 5. Léxicos das edições

## 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. E esse trabalho é bastante relevante, pois não há amostras suficientes que contribuam com a reconstrução da história da nossa língua, sobretudo quando se trata do Português Brasileiro em sua vertente popular. Portanto a produção desse trabalho é uma contribuição para a reconstituição histórica desse português na área de Linguística Histórica.

## REFERÊNCIAS

- PAIXÃO DE SOUZA M. C. *Mémoria do Texto* Revista Texto Digital Universidade Federal de Santa Catarina: 2006.
- 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 [www.uefs.br/cedohs/](http://www.uefs.br/cedohs/)), 2011.
- SANTIAGO, H. S. (Org.) . CD-ROM 1. (Vol. 3) *Cartas em Sisal: Riachão do Jacuípe, Conceição do Coité e Ichu (1906-2000): edição fac-similada*. In. SANTIAGO, H. S. (Org.); CARNEIRO, Z. O. N. (Org.); OLIVEIRA, K. (Org.). *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*. Feira de Santana: UEFS Editora, 2011.
- SANTIAGO, H. S. *Um estudo do português popular brasileiro em cartas pessoais de mãos “cândidas” do sertão baiano*. Dissertação de Mestrado. Feira de Santana: UEFS, 2012. Inédito
- SANTIAGO, H. S. Revista Pandora Brasil - Nº 52 Março de 2013. *Português brasileiro: algumas histórias* ISSN 2175-3318 [http://revistapandorabrasil.com/revista\\_pandora/portugues\\_brasileiro/huda.pdf](http://revistapandorabrasil.com/revista_pandora/portugues_brasileiro/huda.pdf). Acessado em: 17set2013.