



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

XXVI SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2022

UM ESTUDO PILOTO DE USO DE FERRAMENTAS COMPUTACIONAIS EM CORPORA ELETRÔNICOS NO ÂMBITO DA PARCERIA CE-DOHS E CORPUS TYCHO BRAHE, VIA CONVÊNIO UEFS-UNICAMP DE TRANSFERÊNCIA DE TECNOLOGIAS

Bruno Lopes de Almeida¹; Zenaide de Oliveira Novais Carneiro²

1. Bolsista PIBIC/CNPq, Graduando em Letras com Língua Inglesa, Universidade Estadual de Feira de Santana, e-mail: brunouefs@outlook.com
2. Orientadora, Departamento de Letras e Artes (DLA), Universidade Estadual de Feira de Santana, e-mail: zenaide.novais@gmail.com

PALAVRAS-CHAVE: Edição eletrônica; Linguística computacional; Linguística de corpus; Plataforma Tycho Brahe.

INTRODUÇÃO

A discussão se divide em seções, respectivamente enumeradas, que discutem e elucidam partes fundamentais ao desenvolvimento do plano de trabalho: A ferramenta de edição eletrônica eDictor, a apresentação dos documentos, a Plataforma Tycho Brahe e os resultados das testagens e aprimoramentos realizados e/ou sugeridos.

MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)

O material são as cartas de Jozé da Silva Lisboa. A edição eletrônica foi realizada utilizando as ferramentas da Plataforma Tycho Brahe, que absorvem os conceitos de edição trazidos pelo eDictor como ferramenta de edição filológica e incorpora elementos multimídia como integração de áudio e traduções.

RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)

1 O Edictor

Diante do avanço tecnológica e do êxodo científico para o mundo virtual, fez-se necessário que os modelos antigos de se fazer filologia fossem aprimorados. Nessa perspectiva, o eDictor, desenvolvido por Maria Clara Paixão de Sousa, Fábio Kepler e Pablo Faria (2007; 2010), surge como uma ferramenta filológica, agora digital, capaz de

proporcionar a edição eletrônica e modernizada, a partir da linguagem computacional XML, de textos e documentos que, após suas edições, são disponibilizados para análises e estudos linguísticos. As cartas de Jozé da Silva Lisboa, utilizadas na realização desse plano de trabalho, foram editadas com o eDictor e publicadas no site do CE-DOHS.

2 As cartas de Jozé da Silva Lisboa

As cartas de Jozé da Silva Lisboa integram Coleção documental composta da edição fac-similar e semidiplomática de 208 cartas, extraídas de Carneiro (2005). Essas cartas, dirigidas a diversos destinatários, estão depositadas no Instituto Geográfico e Histórico da Bahia (IGHBA), Arquivo Público do Estado da Bahia (APEBA) e Centro de Estudos Feirenses (CENEF), da Universidade Estadual de Feira de Santana (UEFS). Ao todo, do referido remetente, são cinco cartas, todas destinadas a Manoel Ignacio da Cunha e Menezes (visconde do Rio Vermelho). Os documentos, com suas devidas edições, estão disponíveis no acervo do Projeto CE-DOHS.

3. A Plataforma Tycho Brahe

A Plataforma Tycho Brahe é uma plataforma computacional que se configura como um sistema online para o gerenciamento e a manipulação de *corpora* linguísticos com ferramentas integradas para etiquetamento, parsing, busca e edição de árvores sintáticas. A plataforma tem por objetivo facilitar e centralizar os processos de criação de *corpus* anotado sintaticamente por meio do diálogo entre as tecnologias modernas e ciências como as humanidades digitais (Gonçalves e Banza, 2013), linguística de *corpus* e linguística computacional.

Concebida no ano de 2013 e organizada a partir de ferramentas computacionais já existentes como o eDictor (edição de textos) e Corpus Draw (estruturas sintáticas), a Plataforma Tycho Brahe sofreu diversas adaptações para que pudesse prover serviços para diferentes tipos de *corpora*. Atualmente a plataforma está em sua segunda fase de desenvolvimento, tendo como desenvolvedores Luiz Henrique Lima Veronesi, estudante do curso de doutorado do Instituto de Estudos da Linguagem (IEL) da Universidade Estadual de Campinas (UNICAMP) e a professora Charlotte Marie Chambelland Galves (IEL/UNICAMP/CNPq). É nessa segunda fase que está inserida a integração do Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS/UEFS) na plataforma, legalmente assegurada pelo Convênio UEFS-UNICAMP (2019-2023) de transferência de tecnologias. Como a Plataforma Tycho Brahe ainda está sendo desenvolvida, o presente

plano de trabalho se apresentou como um piloto que tentou, de maneira inédita, utilizar os recursos da plataforma. Para isso, como referido anteriormente, foram utilizadas cinco cartas escritas por Jozé da Silva Lisboa. É importante ressaltar que, para além das edições modernizadas dos documentos através das ferramentas da Plataforma Tycho Brahe, o presente plano auxiliou com questões de funcionalidade, layout e forneceu relatórios que proporcionaram o melhoramento e o aperfeiçoamento da plataforma.

4 RESULTADOS: TESTAGENS, APRIMORAMENTOS E SUGESTÕES

Como informado, a Plataforma Tycho Brahe ainda está em desenvolvimento. Muitas alterações precisam ser feitas a todo momento e adaptações de listas, layouts, abas, tags etc. acontecem com frequência. Nesse contexto, durante a vigência da bolsa, fiz indicações que pudessem contribuir nesse sentido, com a melhor fluidez e consistência da plataforma. Partindo para a testagem do novo eDictor e das diversas possibilidades e interfaces de edição compiladas, aprimoradas e desenvolvidas na Plataforma Tycho Brahe, fiquei responsável por testar e observar as ferramentas que competem à edição gráfica do texto dos documentos. Basicamente, a transposição do eDictor e suas funcionalidades para a Plataforma Tycho Brahe. Minha sugestão, muito funcional para esse processo e adotada na plataforma, foi a criação de uma aba específica para editores. Uma aba em que pudéssemos adicionar documentos, editá-los e classificá-los, de maneira online ou offline, sem nenhuma interferência direta na plataforma e em seus bancos de dados. Diferente de criar pastas, drives e fazer os armazenamentos em nossos computadores, tudo agora se faz disponível e arquivado na própria plataforma. A plataforma oferece as mesmas funcionalidades do eDictor em versão software, mas com um layout mais intuitivo, funcional e dinâmico. Um fator interessante é poder ver a versão fac-similar com zoom, ao lado da aba de transcrição do texto. Isso possibilita que uma edição filológica seja feita diretamente na plataforma, sem o intermédio do docx (word). As maiores mudanças aconteceram na aba de edição. Apesar de utilizar os mesmos conceitos e organização do eDictor em software, a plataforma traz uma abordagem completamente diferente e integrada no que se refere à maneira e às possibilidades de manuseio do texto. Antes, por exemplo, a edição era feita palavra por palavra. Atualmente, é possível editar um período inteiro numa mesma aba, otimizando o tempo. Além disso, não é mais necessário aplicar as etiquetas hierarquicamente, selecionando uma por uma. A edição acontece de maneira simultânea, através de caixas de texto. Uma vez editadas, as palavras agora ganham uma cor azul, que outrora era vermelha (no

eDictor software). Outra novidade interessante é o preenchimento dos metadados, que se dá através de caixas de texto muito simples e podem ser adicionados no momento em que se faz a inserção do documento na plataforma. Durante a vigência da bolsa, as cinco cartas de José da Silva Lisboa foram utilizadas e submetidas a esses processos de edição para testar as ferramentas e o funcionamento da plataforma. No percurso, muitos bugs, botões e ferramentas não funcionais foram encontrados, mas a plataforma continua em desenvolvimento e os aprimoramentos também.

CONSIDERAÇÕES FINAIS (ou Conclusão)

Creio que os resultados foram satisfatórios e que cumpri bem com a proposta de testagem das ferramentas e acompanhamento do desenvolvimento da plataforma.

REFERÊNCIAS

- CASTRO, I. **Para uma história do português clássico**. In: DUARTE, I.; LEIRIA, I. (Org.). Actas do Congresso Internacional sobre o Português. Lisboa: Associação Portuguesa de Linguística, 1996. v. 2. p. 135-150.
- CONVÊNIO UEFS-UNICAMP DE TRANSFERÊNCIA DE TECNOLOGIAS (disponível em <http://www5.uefs.br/cedohs/view/convenios.html>), 2019-2023.
- CARNEIRO, Z. de O. N. **Cartas Brasileiras (1808-1904): um estudo lingüístico-filológico**. Tese de Doutorado, Unicamp – Campinas, SP: 2005, volume 2, tomo 1.
- GALVES, C. **The Tycho Brahe Corpus Project: (big) annotated data for a new historical syntax**. T-AP Workshop, Washington DC, 28/29 Janeiro 2015.
- GALVES, C.; ANDRADE, A. L. de; FARIA, P. **Corpus Histórico do Português Tycho Brahe** (disponível em <http://www.tycho.iel.unicamp.br/~tycho/corpus/>), 2017.
- SANDALO, F.; SENA, T. A.; VERONESI, L. H. L. **Annotating Polysynthetic Language: From Portuguese to Kadiwéu**. In Cadernos de Estudos Linguísticos – (59.3), Campinas, pp. 629-645. Set/Dez. 2017.
- GONÇALVES, M. F.; BANZA, A. P. (org.). **Património textual e humanidades digitais: da antiga à nova Filologia**. Évora: CIDEHUS, 2013.
- MACHADO FILHO, A.V.L.; MATTOS SILVA, R.V. (Org.). **O Português Quinhentista: Estudos Lingüísticos**. Salvador: EDUFBA, 2002.
- MOREAU, F. E. **Arquitetura Militar em Salvador da Bahia Séculos XVI a XVIII**. Tese (Doutorado em Arquitetura) – Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo. São Paulo, p. 373. 2011.
- 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.
- PROJETO CORPUS ELETRÔNICO DE DOCUMENTOS HISTÓRICOS DO SERTÃO (disponível em <http://www.uefs.br/cedohs/>), 2011.
- SANTOS, P. V. P. dos. **Trabalhar, Defender e Viver em Salvador no Século XVI**. Dissertação (Mestrado em História) – Faculdade de Filosofia e Ciências Humanas, Universidade Federal da Bahia. Bahia, p. 136. 2004.