

CORRESPONDÊNCIAS AMIGAS (SÉCULO XX): REVISÃO DA EDIÇÃO ELETRÔNICA E ESTUDO DE *TE* E *LHE* COMO CLÍTICOS ACUSATIVOS DE 2ª PESSOA

Emilia Carneiro dos Santos¹; Mariana Fagundes de Oliveira Lacerda²;

1. Bolsista PIBIC/CNPq, Graduando em Licenciatura em Letras vernáculas, Universidade Estadual de Feira de Santana, e-mail: emiliacarneiro.16@hotmail.com
2. Mariana Fagundes de Oliveira Lacerda, Departamento de Letras e Artes, Universidade Estadual de Feira de Santana, e-mail: marianafag@gmail.com

PALAVRAS-CHAVE: Banco de Dados. Edição Eletrônica. Função Acusativa.

INTRODUÇÃO

O projeto CE-DOHS, que conta com ótimos recursos tecnológicos, no universo das Humanidades Digitais, disponibiliza 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 no estudo linguístico. Como se vê, “Do feliz conagração 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.

Neste trabalho, realizamos a revisão da edição eletrônica, em linguagem XML, do acervo *Correspondências amigas*, de Valente-Bahia (1980-1993), revisão necessária, antes da próxima etapa, de anotação sintática, até que fique pronto para as pesquisas linguísticas por meio de buscas automáticas. O *corpus* em questão trata-se de cartas pessoais, de brasileiros (sobretudo baianos) de escolarização média, escritas no século XX, depositadas na cidade de Valente, localizada no interior da Bahia, em sua maioria tendo como principal destinatário o Senhor Adelmário Carneiro Araújo.

Além da revisão da edição XML, estudamos nesse *corpus* – colaborando com os estudos sobre a constituição histórica do português popular brasileiro, especialmente, do sertão baiano – o uso de *te* e *lhe* como clíticos acusativos de 2ª pessoa (LOPES &

MACHADO, 2005; ALMEIDA, 2009). Foram realizadas análises de dados em que essas formas se alternam para fazer referência à 2ª pessoa do discurso. Os exemplos seguintes ilustram esse recorte:

- a. Meus filhos **te** respeitam, Dora.
- b. As crianças **lhe** amam, professora.

MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)

A revisão da edição eletrônica do acervo *Correspondências Amigas*, do banco CE-DOHS, foi feita com o uso do eDictor, desenvolvido por Kepler, Paixão de Sousa e Faria (2007), um editor de textos especialmente voltado ao trabalho filológico e à análise linguística automática. Essa ferramenta combina um editor de XML e um etiquetador morfossintático, e permite a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em html), e de versões com anotação morfossintática (em texto simples e xml).

O *corpus* foi analisado com base nos princípios da sociolinguística. É constituído por 79 cartas, 25 cartões, 80 envelopes, havendo 6 cartas e 2 cartões no formato aerograma, destinados, em sua maioria, ao Senhor Adelmário Carneiro Araújo, nascido em Valente, em outubro de 1959. São 38 remetentes, 31 do sexo feminino e 7 do sexo masculino. Os remetentes são, a maior parte, nascidos no interior da Bahia, jovens com idade entre 20 e 30 anos, estudantes do Ensino Fundamental ou do Ensino Médio.

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

O eDictor é uma ferramenta computacional, com sistema de anotação de edição em Linguagem XML, que permite resguardar as informações filológicas fundamentais dos textos, ao mesmo tempo em que os torna aptos para o processamento automático, criado por Paixão de Sousa, Kepler e Faria (2007). A revisão final se faz necessário para melhor preparar o documento para a próxima etapa da edição: a anotação morfossintática. No exemplo abaixo, há uma das cartas em fase de revisão:

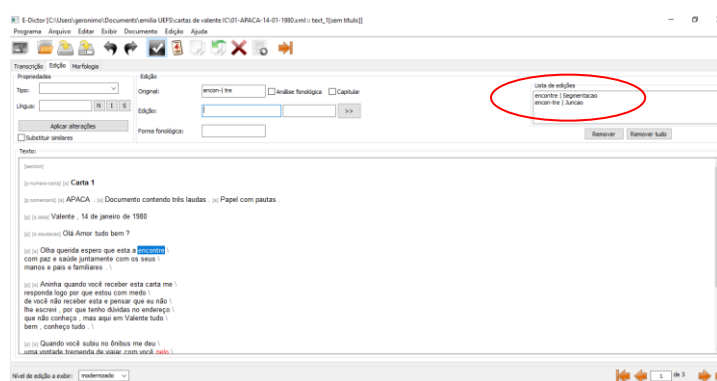


Figura 1: Processo de revisão do documento

Fonte: CE-DOHOS

O processo de revisão se dá seguindo os mesmos passos da edição, isto é, após inserir o documento na barra edição do e-Dictor, clica na palavra a ser editada e usa a etiqueta correspondente a alteração que ela precisa. As etiquetas são ferramentas que possibilitam ao editor fazer estas alterações. Por exemplo na palavra “encontre” da Figura 01 foram usadas duas etiquetas: a junção e a segmentação. Como se pode ver, cada modificação realizada fica visível, possibilitando ao editor observar quais foram as

alterações realizadas em cada palavra. Na revisão, observa-se estas alterações seguirem os padrões regidos pelo CE-DHOS. Pois, mesmo em edições modernizadas, que há uma maior intervenção no documento, não se pode padronizar concordância, crase etc., visto que são possíveis fenômenos de uma pesquisa linguística. Então no processo de revisão é preciso estar atento a todas essas questões.

Após a revisão do documento, faz-se a revisão dos metadados. Os metadados são as informações referentes ao documento. Eles são de dois tipos: original e de processamento. O primeiro diz respeito ao documento original, quem escreveu, localidade, datas etc. O segundo registra quem foi responsável pelas edições.

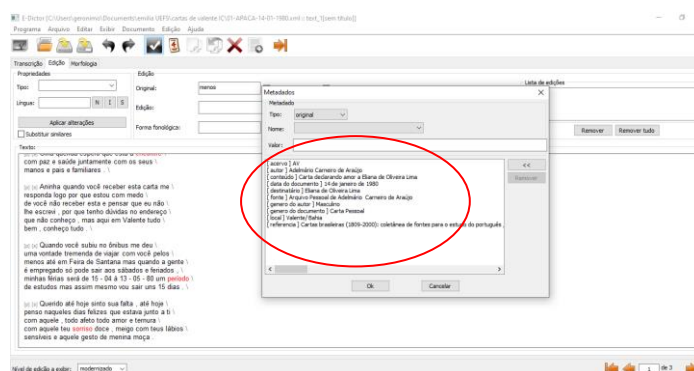


Figura 2: Inserção de metadados
Fonte: CE-DOHS

A figura 02 mostra a revisão dos metadados do tipo original. A revisão desses dados garante ao pesquisador maior confiabilidade no documento. Seja para estudos linguísticos como para outras áreas.

No que tange ao estudo da função dos clíticos, a Tabela 01 demonstra alguns resultados obtidos:

Tabela 01: Registro geral de ocorrências de *te* e *lhe* e suas respectivas funções

	Total de ocorrências	Acusativo	Dativo
Te	72	26	46
Lhe	45	12	33

Fonte: Elaboração do autor

Como se pode observar, os resultados demonstram que os remetentes preferiram, o clítico *te* ao *lhe*, para referir-se à segunda pessoa que é a forma prescrita pela Gramática Tradicional (GT). Foram registradas 72 ocorrências de *te* e 45 de *lhe*. Dessas 45 ocorrências de *lhe*, 12 ocorreram na função acusativa, contrariando a maioria das GT que classificam o clítico *lhe* como pronome oblíquo da terceira pessoa do discurso, o qual funciona como complemento, exercendo a função de clítico dativo, ou seja, Objeto Indireto (OI). Porém nesse estudo, as ocorrências referentes a este clítico aparecem sempre como referência a segunda pessoa do discurso, mesmo quando a função exercida por ele é de OI. Como pode-se observar nos exemplos abaixo:

- [...]trabalhe perto para **eu lhe encontrar** todos os dias. [...] (01-APACA-14-01-1980)
- [...]ela e legal me falou **que lhe conhece** e sua esposa. [...] (05-APACA-14-01-1990)

Estes resultados apontam o que outros estudos já confirmaram (LOPES & MACHADO, 2005; ALMEIDA, 2009). Isto é, é comum no Português Brasileiro o uso do clítico *lhe* com referência a segunda pessoa, inclusive, há gramáticos a exemplo de

Rocha Lima (2006), que já trazem esta possibilidade. Sobre a função acusativa, Almeida (2009) justifica esse fenômeno, propondo que o clítico de terceira pessoa está perdendo seu traço [+ dativo].

Sobre a ocorrência de *te*, seguem alguns exemplos encontrados no *corpus*:

- a. [...] **te achei** um cara legal mesmo. (48-APACA-25-08-1980)
- b. [...] Não sei como conseguir **te encontrar**[...] (01-APACA-14-01-1980)

CONSIDERAÇÕES FINAIS (ou Conclusão)

Os resultados obtidos a partir deste estudo foram:

- a) A Revisão da edição eletrônica do acervo Correspondências Amigas, e dos metadados. Os quais corroboram para os estudos linguísticos. Visto que, o *corpus* fica preparado para a próxima etapa, que é a anotação morfosintática. Colaborando com a constituição de *corpora* apropriados aos estudos linguísticos e contribuindo com as pesquisas sócio-históricas do português brasileiro.
- b) O levantamento do *corpus* de ocorrências *te* e *lhe* com função acusativa de 2ª pessoa, por meio de apêndice, bem como a sua quantificação e a descrição dos resultados, também contribuem com os estudos sobre a história do português popular brasileiro. Confirmando assim como outros estudos já demonstraram (ALMEIDA, 2009) a perda do traço [+ dativo] do pronome *lhe* e seu comum uso em referência a segunda pessoa do discurso.

REFERÊNCIAS

- ALMEIDA, G. de S. Quem te viu quem lhe vê: a expressão do objeto acusativo de referência à segunda pessoa na fala de Salvador. Dissertação (Mestrado em Linguística). Instituto de Letras.
- CARNEIRO, Zenaide de Oliveira Novais; SANTIAGO, Huda da Silva; OLIVEIRA, Klebson (Org.). Cartas Brasileiras (1809-2000): coletânea de fontes para o estudo do português. Feira de Santana, BA: UEFS Editora, 2011.
- LOPES, C. R. dos S.; MACHADO, A. C. M. Tradição e inovação: indícios do sincretismo entre a segunda e a terceira pessoas nas cartas dos avós. In: LOPES, C. R. S. (Org.) A Norma Brasileira em Construção. Fatos linguísticos em cartas pessoais do século 19. Rio de Janeiro: Faculdade de Letras, UFRJ, FAPERJ. 2005. p. 45-66.
- 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.
- SILVA, Rosa Virginia Mattos E. Caminhos da linguística histórica: ouvir o inaudível. São Paulo, SP: Parábola Ed., 2008.