

## Transposição material do papel ao texto digital: Testamentos do Corpus DOViC

Marina Martins Pinchemel Amorim<sup>1</sup>, Cristiane Namiuti Temponi<sup>2</sup>, Jorge Viana Santos<sup>3</sup>

1. Estudante de IC da Universidade Estadual do Sudoeste da Bahia – UESB; \*[marinapinchemel@gmail.com](mailto:marinapinchemel@gmail.com)

2. Professora do Departamento de Estudos Linguísticos e Literários e do Programa de Pós Graduação em Linguística – UESB, Vitória da Conquista/BA;

3. Professor do Departamento de Estudos Linguísticos e Literários e do Programa de Pós Graduação em Linguística – UESB, Vitória da Conquista/BA.

Palavras Chave: *Humanidades digitais, Corpora anotados, Manuscritos antigos.*

### Introdução

De acordo com Paixão de Sousa (2006), a questão central que se coloca para o trabalho com textos antigos como fundamentos para estudos linguísticos no meio eletrônico é a busca por uma abordagem global do texto, em termos conceituais e tecnológicos, que se reflita numa integração entre diferentes planos de análise. Este trabalho se enquadra na proposta delineada em Namiuti, Santos e Leite (2011) para o desenvolvimento da pesquisa linguística baseada em *corpora*, problematizando a necessidade de se colocar na posição de um pesquisador formador de *corpora* (PFC) e não apenas de um pesquisador pragmático. Ele está vinculado ao projeto temático “Sintaxe diacrônica em *corpus* eletrônico: do português pré-clássico às variantes modernas” e visa trabalhar especificamente com testamentos catalogados no *Corpus DOViC*, contribuindo para a construção deste, garantindo o controle das edições, uma vez que tal controle permite ao pesquisador analisar a complexidade do documento histórico ao se tornar documento digital.

### Resultados e Discussão

A transposição material é realizada por meio de etapas. Nosso trabalho parte do resultado da etapa de processamento de imagem, chamada de “Fotografia” e da primeira fase de transcrição de documentos. O acesso aos testamentos guardados no computador servidor do Laboratório de Pesquisa em Linguística de Corpus (LAPELINC), já digitais em forma de imagem e texto transcrito, se dá remotamente através de PC terminal. Nosso trabalho inicia com a etapa de transformação do documento imagem em documento texto, que possui mais duas etapas. A primeira delas é a Transcrição, que envolve a leitura e interpretação do documento/imagem em dois momentos: a transcrição paleográfica em tarefa individual e em seguida a revisão do documento transcrito. Foram realizadas as revisões de 26 testamentos selecionados para compor este tipo de documento no *corpus* Beta DOViC, datados do século XIX e arquivados em 3 livros de notas guardados no fórum da cidade baiana de Vitória da Conquista. Acerca dessas revisões foi elaborado um Diário de Revisão, contendo os dados de identificação do testamento, data da revisão, tipo e alteração feita e o tempo gasto por folha-imagem. Com a análise, vimos que o tempo gasto para a revisão de uma folha-imagem em comparação ao texto digitado é em média 19 minutos e que a maioria das alterações nas transcrições foi de substituição de letra, devido à dificuldade ao interpretar a grafia do tabelião que redige o testamento. Esses dados contribuem para estudos futuros de estatística e para a gerência de atividades planejadas para a equipe do LAPELINC. Passada essa etapa, segue a Edição em camadas, que é bastante complexa, pois envolve a inserção de codificação de formatos e metadados. Esta tarefa exige estudo e treinamento de uma ferramenta de edição específica para a construção de grandes *corpora* anotados para a pesquisa linguística – o eDictor – que permite edição eletrônica em XML de textos antigos para fins de análise linguística automática. Todo o material preparado segue para o processo de Anotação Linguística, a etapa que conclui a transposição. O documento em formato texto está pronto para ser submetido a uma ferramenta acoplada ao e-Dictor, responsável por anotar morfologicamente as palavras do texto.

### Conclusões

A preservação do patrimônio histórico e linguístico de um povo, como os testamentos, é importante para o estudo das mais variadas áreas, como a História, Paleologia, Economia, Filologia, Direito e, finalmente, a Linguística, em suas várias ramificações. A relevância desse trabalho contempla vários segmentos de pesquisa, pois se dispõe a preparar um material a ser lançado na *web* e que poderá ser consultado por qualquer pessoa ao redor do mundo. Este projeto reafirma a importância da Linguística de *Corpus* e dos Estudos Diacrônicos para a preservação da memória (linguística e cultural), apresentando os desafios metodológicos impostos pela Linguística aliada às vertentes tecnológica (Fotografia) e computacional (sistemas de gerenciamento de informação, banco de dados, edição e anotação eletrônicas de textos), bem como alguns desenvolvimentos metodológicos envolvendo essas duas tendências combinadas para o trabalho com textos antigos.

### Agradecimentos

Agradecemos à Universidade Estadual do Sudoeste da Bahia (UESB), ao Laboratório de Pesquisa em Linguística de Corpus (LAPELINC) e à Fundação de Amparo à Pesquisa do Estado da Bahia (Fapesb).

NAMIUTI, C.; SANTOS, J. V.; LEITE, C. M. B. *Propostas e Desafios dos Novos Meios das Antigas Fontes: A Preservação da Memória pela Linguística de Corpus*. In: X Colóquio Nacional e II Colóquio Internacional do Museu Pedagógico UESB, 2011, Vitória da Conquista. Anais do X Colóquio Nacional e II Colóquio Internacional do Museu Pedagógico UESB. Vitória da Conquista: UESB, 2011. v. 1. p. 1-11.

PAIXÃO DE SOUSA, M.C.; KEPLER, F. N.; FÁRIA, P.P.F. E-Dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: SHEPHERD, T.; SARDINHA, T. B.; PINTO, M. V. (Orgs). *Caminhos da linguística de corpus*. Campinas: Mercado de Letras, 2010.

Cristiane NAMIUTI-TEMPONI, Jorge Viana SANTOS, Aline Silva COSTA. *New challenges for ancient sources: DOViC experience in the new Historical Linguistics*. Congresso de Humanidades Digitais em Portugal: construir postes e quebrar barreiras na era digital. 8-9 de Outubro, 2015 — Lisboa/Portugal.