

## O MÉTODO LAPELINC: ALIMENTAÇÃO DE DOCUMENTOS DO CORPUS DOViC NO APLICATIVO WEBSINC

Amanda Moreno Fonsêca de Andrade  
(UESB/CNPq)

Patrick Pereira Campos Brito  
(UESB)

Jorge Viana Santos  
(UESB/CNPq)

Cristiane Namiuti Temponi  
(UESB/CNPq)

### RESUMO

O Método Lapelinc tem como proposta a construção de um *corpus* digital que parte de documentos notariais manuscritos guardados em arquivos da região do Sudoeste Baiano: o *Corpus* DOViC (Corpus de Documentos Oitocentistas de Vitória da Conquista) (SANTOS; NAMIUTI, 2016). O método possui três etapas: *Transposição*, *Transcrição Paleográfica* e *Compilação de Corpora*. No presente trabalho, enfatizamos algumas subetapas das fases do método realizadas por pesquisadores do Lapelinc com a finalidade de alimentar o aplicativo *WebSinc*<sup>1</sup> com dados de documentos presentes no *corpus* citado.

**PALAVRAS-CHAVE:** Método Lapelinc; Corpus Dovic; *WebSinc*

### INTRODUÇÃO

O Método Lapelinc foi desenvolvido por pesquisadores da Universidade Estadual do Sudoeste da Bahia. O método iniciou em 2009 e continua a ser desenvolvido no Lapelinc (Laboratório de Pesquisa em Linguística de Corpus-UESB) e possui três etapas, a

---

<sup>1</sup> O *WebSinc* (Namiuti, Santos e Costa 2015 e Costa e Namiuti 2015) é um aplicativo web para o trabalho de registro, armazenamento, disponibilização, visão e busca de

**IX SEMINÁRIO DE PESQUISA E ESTUDOS LINGÜÍSTICOS**  
**21 e 22 de setembro de 2017**

---

saber: *Transposição, Transcrição Paleográfica e Compilação de Corpora*. Após os documentos serem trabalhados nestas três etapas o material digital permanecerá fidedigno ao original, podendo ser acessado remotamente via *Web*. As subetapas integrantes de cada uma das 3 etapas citadas do Método serão apresentadas na seção metodologia e, na seção seguinte, serão relatadas descritivamente com seus devidos resultados as que foram executadas pelos pesquisadores Amanda M. F. Andrade e Patrick P. C. Brito.

## **MATERIAL E MÉTODOS**

O Método Lapelinc consiste na transposição do Documento Físico (DF) para o Documento Digital Imagem (DDI) que, após transcrito, origina o Documento Digital Texto (DDT). O método possui 3 etapas: 1) *Transposição*; 2) *Transcrição Paleográfica* e 3) *Compilação de Corpora*. Concernente à primeira, *Transposição*, temos as seguintes subetapas: 1.1) Controle: etapa da captura de informações da fonte; 1.2) Captura fotográfica da imagem do original: fotografa-se o DF (Documento Físico) para se formar o DDI (Documento Digital Imagem). Nesta etapa está presente a Mesa Cartesiana<sup>2</sup> que é um equipamento que possibilita o registro visual das características físicas dos documentos; 1.3) Edição fotográfica que consiste em preparar versões do DDI controladas para diferentes fins; 1.4) Criação de imagens de uso co-indexadas à imagem-original; 1.5) Análise descritiva/ topográfica que se divide em duas etapas, sendo: a) Localizar o início e o final de Documentos Micro nas folhas-Imagem do Documento Macro; b) identificar, a partir dos cabeçalhos dos Documentos Micro, a tipologia do documento. Na segunda etapa, *Transcrição Paleográfica*, é estabelecida uma relação entre a Fotografia e a Paleografia, que se subdivide em: 2.1) Leitura e Decodificação dos DDIs; e 2.2) Transcrição Paleográfica. Segundo Namiuti e Santos (2016), a etapa de *Transcrição* envolve as soluções técnicas para a edição

---

<sup>2</sup> Mesa Cartesiana foi idealizada e concebida para ser um equipamento que possibilitasse, junto com o registro do documento, o registro visual cientificamente controlado de suas características físicas

**IX SEMINÁRIO DE PESQUISA E ESTUDOS LINGÜÍSTICOS**  
**21 e 22 de setembro de 2017**

---

especializada de textos antigos em meio eletrônico. Por fim, para que a terceira etapa, *Compilação de Corpora*, seja iniciada, é necessário que o documento tenha ganhado o formato digital de texto simples (TXT) após ter passado pelas outras etapas. Desta forma, para dar continuidade ao trabalho, nesta etapa são essenciais algumas ferramentas computacionais, a exemplo do: *eDictor* (PAIXÃO DE SOUSA, KEPLER E FARIA, 2010) que é um editor de marcação extensível; e do *WebSinc* (NAMIUTI; SANTOS e COSTA, 2015 e COSTA e NAMIUTI, 2015).

## RESULTADOS E DISCUSSÃO

Levando em conta a metodologia adotada nas atividades propostas pelo Método, obtivemos os seguintes resultados das atividades circunscritas das etapas 1 e 3 realizadas pelos pesquisadores Amanda M. F. Andrade e Patrick P. C. Brito envolvidos na alimentação de dados do *WebSinc*. Na primeira etapa, *Transposição*, as atividades desenvolvidas foram: primeiramente, edição fotográfica, que consiste no recorte da imagem, que, neste caso, foram recortadas as imagens dos livros E3, E11, E1, E5, E15 e E16, totalizando 6 livros editados. Abaixo pode-se observar a imagem antes do corte e pós o corte, nas figuras 1 e 2:

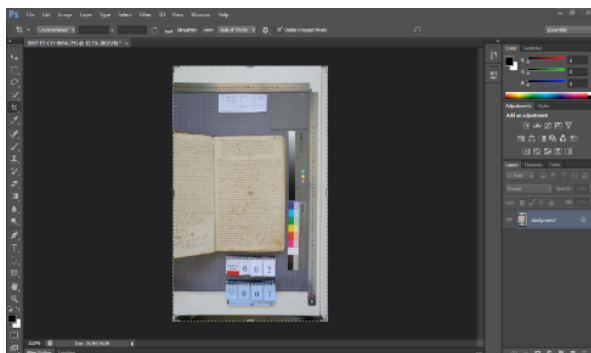


Figura 1: folha-imagem 007-E5-0004.JPG pertencente ao livro E5 C11 antes do corte.

**IX SEMINÁRIO DE PESQUISA E ESTUDOS LINGÜÍSTICOS  
21 e 22 de setembro de 2017**

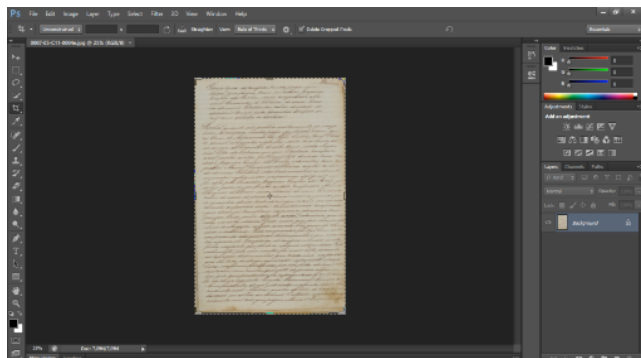


Figura 2: folha-imagem 007-05-0004e.JPG pertencente ao livro E5 C11 após o corte.

E, em seguida, concernente à terceira etapa, foi feito: a) Alimentação de dados dos documentos macro no *WebSinc* (Namiuti, Santos e Costa 2015 e Costa e Namiuti 2015) a catalogação de 10 livros da coleta 11 (C11), a saber os livros: E1, E3, E5, E6, E11, E13, E14, E15, E16 e E21, totalizando 10 livros cadastrados no aplicativo, conforme a figura 3 abaixo:

Cadastro de Documento Macro				
<a href="#">+ Novo</a>				<a href="#">Início</a>
Código	Título do documento	Tipo de documento	Vinculados ao documento	Ações
E14	Livro de Notas 14 dos annos de 1882 a 1883	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E16	Livro de Notas 16 dos annos de 1883 a 1884	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E1	Livro de Notas 1 dos annos de 1841 a 1848	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E5	Livro de Notas 5 dos annos de 1862 a 1866	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E3	Livro de Notas dos annos de 1849 a 1858	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E6	Livro de Notas dos annos de 1869 a 1870	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E11	Livro de Notas dos annos de 1877 a 1880	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E13	Livro de Notas dos annos de 1880 a 1881	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E21	Livro de Notas dos annos de 1880 a 1890	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	
E15	Livro de Notas dos annos de 1881 a 1882	Livro de notas	<a href="#">Imagens</a> <a href="#">Manuscritos</a>	

Figura 3: Página do Websinc onde os livros do *Corpus Dovic* foram cadastrados.

Fonte: *WebSinc*

**IX SEMINÁRIO DE PESQUISA E ESTUDOS LINGÜÍSTICOS**  
**21 e 22 de setembro de 2017**

---

b) Anexação das imagens dos Documentos Macros cadastrados anteriormente; c) Cadastro dos dados e cabeçalhos dos Documentos Micros presente nos Documentos Macros cadastrados.

## **CONCLUSÃO**

Os materiais históricos e/ou raros precisam ser trabalhados de maneira cuidadosa, pois além de serem antigos, são frágeis. Assim, é necessário cuidado ao manuseá-los. Tendo isto em vista, o Método Lapelinc utiliza a Fotografia como ferramenta no processo de transposição do documento garantindo a sua fidedignidade. O presente trabalho contribui na construção do aplicativo WebSinc, afim de, futuramente, auxiliar pesquisas tanto na área da Linguística, quanto no Direito, por exemplo.

## **REFERÊNCIAS**

BRITO, Giovane Santos. **“Do texto ao documento digital: transposição fotográfica de documentos manuscritos históricos para formação de corpora linguísticos eletrônicos”**. (Dissertação) Mestrado em Linguística – Programa de Pós Graduação Linguística da Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2015. Orientador: Jorge Viana Santos; Co-orientadora: Cristiane Namiuti Temponi.

COSTA, Aline Silva; NAMIUTI-TEMPONI, Cristiane. **WebSinC: Buscas online em corpora sintaticamente anotados**. “E-Book do Congresso de Humanidades Digitais em Portugal: Construir pontes e quebrar barreiras na era digital – 2015”. Lisboa: Universidade Nova de Lisboa, 2016 (no prelo).

FARIA, P.; KEPLER, F. N.; PAIXÃO DE SOUZA, M. C. An Integrated Tool for Annotating Historical Corpora , In: Fourth Linguistic Annotation Workshop, LAW IV, **48th Annual Meeting of the ACL**, 2010, Uppsala. Proceedings of the Fourth Linguistic Annotation Workshop, 2010. p. 217-221.

NAMIUTI, Cristiane; SANTOS, Jorge e COSTA, Aline. **“New challenges for ancient sources: an important dialogue between**

**IX SEMINÁRIO DE PESQUISA E ESTUDOS LINGÜÍSTICOS  
21 e 22 de setembro de 2017**

---

**Computer Science and new Historical Linguistics.”** Trabalho apresentado no Workshop: The New Historical Linguistics and the World of Annotated Corpora, Campinas, Brasil, Março 9-13, 2015.

NAMIUTI, Cristiane; SANTOS, Jorge Viana. **Novos desafios para antigas fontes: a experiência DOViC na nova linguística histórica.** E-Book do Congresso de Humanidades Digitais em Portugal. 2015. Lisboa: Universidade Nova de Lisboa, 2016 (prelo).