

O Método LAPELINC na Construção do Corpus Dovic: Topografia de Documentos Oitocentistas de Vitória da Conquista.¹

Thais Reis Silva ², Cristiane dos Santos Namiuti ³

RESUMO

No presente trabalho apresenta-se os resultados do subprojeto de pesquisa “Topografia de documentos do Livro de Notas E15 (1881-1882), E16 (1883-1884), E21 (1880-1890) do Corpus DOVIC (Documentos Oitocentistas de Vitória da Conquista e Região): uma contribuição para ampliação das fontes de investigação do português no Brasil oitocentista”, vinculado ao projeto temático “Do português pré-clássico às variantes modernas: contribuições para o estudo da sintaxe e interfaces”, tal projeto está inserido no campo da Linguística de Corpus e tem como objetivo contribuir para a construção do corpus Dovic (Documentos Oitocentistas de Vitória da Conquista) através da topografia das cartas de alforria dos Livros de Notas E 15, E 16 e E 21 (1880-1890), justificando-se pela necessidade de preservar digitalmente estes documentos essenciais para a memória local e nacional, dado seu valor como fonte de informações sobre o período escravocrata brasileiro. A metodologia aplicada é o método LAPELINC (Namiuti; Santos, 2021), que assegura a fidedignidade na transferência do texto físico para o meio digital por meio de etapas controladas. A pesquisa fundamenta-se teoricamente nos trabalhos desenvolvidos pelo laboratório Lapelinc, como os de Namiuti e Santos (2021; 2019), Namiuti e Costa (2021) e Santos e Brito (2021; 2014), além de estudos de outros pesquisadores da área, como Souza (2006) e Sardinha (2000). Dessa forma, esta pesquisa visa não apenas garantir a preservação da memória contida nesses documentos, mas também disponibilizar um corpus digital confiável que sirva de base para futuras investigações linguísticas e históricas a partir dos dados compilados.

PALAVRAS-CHAVE: Linguística de Corpus, Corpus Dovic, método LAPELINC, topografia

ABSTRACT

The LAPELINC Method in the Construction of the Dovic Corpus: Topography of Nineteenth-Century Documents from Vitória da Conquista.

In the present work, the results of the sub-project "Topography of documents from the Livro de Notas E15 (1881-1882), E16 (1883-1884), E21 (1880-1890) of the DOVIC Corpus (Documentos Oitocentistas de Vitória da Conquista e Região): a contribution to expanding the sources for the investigation of Portuguese in 19th-century Brazil" are presented. This sub-project is linked to the thematic project "From pre-classical Portuguese to modern variants: contributions to the study of syntax and interfaces," and is situated within the field of Corpus Linguistics. Its objective is to contribute to the construction of the DOVIC corpus (Documentos

1 Pesquisa financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), por meio de bolsa de fomento à discente pesquisadora

2 Discente do curso de Graduação de Letras Modernas da Universidade Estadual do Sudoeste da Bahia (UESB). E-mail: thaisreis2806@gmail.com

3 Professora do Departamento de Estudos Linguísticos e Literários (DELL) e do Programa de Pós-Graduação em Linguística (PPGLin), ambos da Universidade Estadual do Sudoeste da Bahia (UESB).

E-mail: cristianenamiuti@uesb.edu.br

Oitocentistas de Vitória da Conquista) through the topographic transcription of the letters of manumission from the Livros de Notas E 15, E 16, and E 21 (1880-1890). This work is justified by the need to digitally preserve essential documents on Brazilian slavery. The applied methodology is the LAPELINC method (Namiuti; Santos, 2021), which ensures the reliability of the transfer from the physical text to the digital medium through controlled steps. The research is theoretically grounded in the work developed by the Lapelinc laboratory, such as Namiuti and Santos (2021; 2019), Namiuti and Costa (2021), and Santos and Brito (2021; 2014), in addition to studies by other researchers in the field, such as Souza (2006) and Sardinha (2000). Thus, this research aims to preserve these documents and provide a reliable digital corpus for future studies.

KEYWORDS: Corpus Linguistics; Dovic Corpus; Digital Preservation; LAPELINC Method; Textual Topography

INTRODUÇÃO

A preservação de fontes documentais históricas, fundamentais para as Humanidades, esbarra na inevitável degradação material de seus suportes físicos. Como alternativa, a constituição de corpus digitais tem se mostrado uma solução promissora. No entanto, essa migração para o meio digital transfere o desafio crítico da preservação física para o da garantia de fidedignidade dos documentos. Para assegurar esse rigor metodológico, surge o método LAPELINC (Namiuti; Santos, 2021), um método controlado que articula técnicas de fotografia, linguística, palográfico e computação. Seu objetivo final é a criação de corpus digitais anotados que preservem a integridade informacional e do documento original, assegurando confiabilidade para pesquisa

Diante disso, o presente trabalho tem como objetivo demonstrar a aplicação e a contribuição dessa metodologia para a construção do corpus Dovic (Documentos Oitocentistas de Vitória da Conquista), que possui documentos importantes do período escravocrata do Brasil. Busca-se, ainda, evidenciar o potencial do método para viabilizar e otimizar a formação de corpora em geral, bem como as pesquisas linguísticas e filológicas subsequentes.

A fundamentação teórica deste estudo ancora-se, principalmente, nos trabalhos de Namiuti e Santos (2021; 2019), Namiuti e Costa (2021) e Santos e Brito (2021; 2014), os quais detalham o funcionamento e as potencialidades do método LAPELINC. Complementam essa base os estudos de Souza (2006) e Sardinha (2000), que discutem os fundamentos e propósitos da Linguística de Corpus

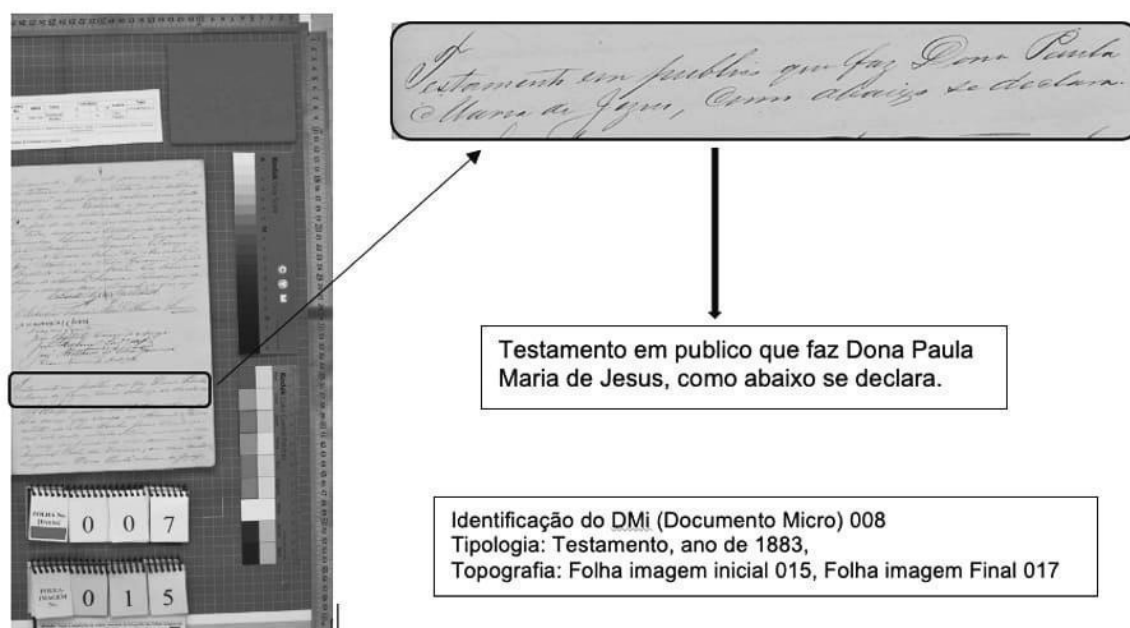
Na seção a seguir, apresentaremos os passos metodológicos adotados para a pesquisa.

MATERIAIS E MÉTODOS

O método LAPELINC (Namiuti; Santos, 2021) estrutura-se em três etapas principais: a transposição, que consiste na captura digital do documento físico por meio de fotografia cientificamente controlada; a transcrição, que aplica os rigorosos princípios da paleografia para a fiel conversão do texto antigo; e a compilação, que resulta na anotação digital do documento em formato XML (*Extensible Markup Language*). Nesta pesquisa, enfatizou-se atividades das etapas de transposição e transcrição, mais especificamente a Topografia dos Documentos Micros (DMis) contidos nos Livros de Notas (Santos; Namiuti, 2019).

A figura 1 ilustra os processos e os objetos envolvidos na topografia: à esquerda a Folha-Imagem 015 do Livro de notas E16, folha 07 retro do livro físico, com elementos de controle da transposição material feita através da fotografia cientificamente controlada, e à direita a possibilidade de ampliação para a leitura e identificação do início de um documento, no caso um testamento.

FIGURA 1: TOPOGRAFIA DO LIVRO E16: IDENTIFICAÇÃO DO DMI 008



Fonte: Elaborada pelas autoras

O percurso metodológico desta pesquisa consistiu na realização da topografia dos DMis a partir das folhas-imagem do Livro de Notas E16. Após a execução da topografia, procedeu-se à tabulação de dados referentes tanto ao processo de análise quanto às características intrínsecas dos documentos. Foram registrados métricas relacionadas ao próprio método, como o tempo despendido na topografia de cada documento, bem como

atributos dos livros analisados, tais como: o número total de documentos, a extensão lexical dos cabeçalhos e a frequência de cada tipologia documental identificada no corpus.

RESULTADOS E DISCUSSÃO

Com a leitura das folhas-imagem capturadas segundo o método LAPELINC, e sua possibilidade de ampliação para a leitura, realizamos a identificação e transcrição dos cabeçalhos e fechos dos documentos contidos.

As folhas imagens do livro Notas E16 foram imprescindíveis para uma boa e facilitada topografia. Isso ocorre, pois conforme (Santos; Brito, 2021) a fotografia cientificamente controlada do método Lapelinc, não só pode preservar o documento físico e facilitar o seu acesso para os pesquisadores, como também pode ser usado para adicionar mais praticidade ao trabalho dos formadores de corpora eletrônicos, uma vez que as ferramentas de edição fotográficas podem ser usadas para facilitar uma melhor visualização daqueles documentos que possam estar apagados pelo o tempo.

O levantamento topográfico realizado gerou um conjunto de dados que pode subsidiar a formulação de questões de pesquisa. Entre as informações coletadas, destaca-se a distribuição quantitativa de documentos por tipologia, conforme ilustrado no gráfico abaixo.

FIGURA 2: Distribuição tipológica dos documentos do livro E16



Fonte: Elaborado pelas autoras

A análise do gráfico suscita a seguinte questão investigativa: que fatores históricos, sociais e administrativos explicam a predominância de procuções no Livro de Notas E16 (1883–1884) de Vitória da Conquista, em comparação com os demais tipos documentais registrados no mesmo período? Essa indagação provavelmente surgiria a um pesquisador ao examinar o corpus aqui descrito. A partir do material compilado, ele dispõe de uma base

empírica sólida para explorar tal problema, o que, segundo Sardinha (2000), constitui justamente a finalidade dos recursos oferecidos pela construção de corpora digitais.

CONCLUSÕES/CONSIDERAÇÕES

A análise dos dados demonstraram que o método LAPELINC possui grande potencial para a elaboração de corpora digitalmente anotados. Ele estabelece um conjunto de etapas sistemáticas e controladas, entre as quais se destaca a fotografia científica, que atua como ponte entre o documento físico (DF) e o documento digital (DD) (Namiuti; Santos, 2021). Esse procedimento metodológico viabiliza tanto a preservação dos documentos quanto oferece maior praticidade para seu manuseio e análise por pesquisadores.

A análise dos dados obtidos por meio do processo de topografia descrito neste trabalho, que combina o uso de fotografias cientificamente controladas (Santos; Brito, 2021) com as regras paleográficas para transcrição e análise das informações, permitiu identificar diversas oportunidades de investigação acadêmica. Um exemplo disso é a pergunta de pesquisa gerada através do estudo do recorte da tipologia documental presente nos livros de notas E16, discutido neste trabalho, o que reforça o valor científico do *corpus* construído

Em síntese, através do método LAPELINC, contribuímos, de forma cientificamente controlada, para a construção do corpus DOVIC (*Documentos Oitocentistas de Vitória da Conquista*), assegurando fidelidade aos documentos originais, aspecto fundamental para o desenvolvimento de pesquisas científicas baseadas nesse material.

REFERÊNCIAS BIBLIOGRÁFICAS

1. NAMIUTI, Cristiane; SANTOS, Jorge Viana. **Novos desafios para antigas fontes: a experiência DOVIC na nova linguística histórica.** In: Humanidades digitais e o mundo lusófono. Organização Ricardo M. Pimenta, Daniel Alves. Rio de Janeiro: Editora FGV, 2021, págs. 69-89.
2. NAMIUTI, Cristiane; COSTA, Aline. **Reflexões sobre anotação sintática e ferramentas de busca - Uso da linguagem XML para anotação sintática no corpus digital DOViC.** In: Letras e Letras, v. 30, n. 2, p. 82-103, 2014.
3. PAIXÃO DE SOUSA, Maria Clara. **Memórias do Texto.** Revista Texto Digital, v. 2, n. 1, 2006.
4. SANTINHA, Tony. **Linguística de corpus: histórico e problemática.** In RETROSPECTIVA/RETROSPECTIVE D.E.L.T.A., Vol. 16, N.º 2, 2000(323-367).
5. SANTOS, Jorge; BRITO, Giovani. **A fotografia digital na transposição de documentos manuscritos históricos.** In: Humanidades digitais e o mundo lusófono.

Organização Ricardo M. Pimenta, Daniel Alves. Rio de Janeiro: Editora FGV, 2021, págs. 103-115.

6. SANTOS, J. V.; BRITO, G. S. **Fotografia técnica de documentos para formação de corpora digitais eletrônicos**: o método desenvolvido no Lapelinc. *Letras & Letras*, v. 30, n. 2, p. 421-430, 2014.
7. SANTOS, Jorge Viana; NAMIUTI Cristiane. **DOViC (Documentos Oitocentistas de Vitória da Conquista-BA e Região)**. Corpus eletrônico, UESB, Vitória da Conquista, 2014

