

TOPOGRAFIA DE DOCUMENTOS DOS LIVROS DE NOTAS E1 (1841-1848),
E2 (1841-1855), E5 (1852-1866) DO CORPUS DOVIC (DOCUMENTOS
OITOCENTISTAS DE VITÓRIA DA CONQUISTA E REGIÃO)

Emilly Silva Moreira Viana¹, Cristiane Namiuti²

RESUMO

Esse resumo expandido apresenta o resultado da pesquisa do projeto de iniciação científica “TOPOGRAFIA DE DOCUMENTOS DOS LIVROS DE NOTAS E1 (1841-1848), E2 (1841-1855), E5 (1852-1866) DO CORPUS DOVIC (DOCUMENTOS OITOCENTISTAS DE VITÓRIA DA CONQUISTA E REGIÃO): UMA CONTRIBUIÇÃO PARA AMPLIAÇÃO DAS FONTES DE INVESTIGAÇÃO DO PORTUGUÊS NO BRASIL OITOCENTISTA”, que se vincula a um projeto maior denominado “Do português pré-clássico às variantes modernas: Contribuições para o estudo da sintaxe e interface”, que vem sendo desenvolvido no Lapelinc (Laboratório de Pesquisa em Linguística de Corpus), laboratório criado e coordenado por Cristiane Namiuti (PPGLin/UESB) e Jorge Viana (PPGLin/UESB). Pensando na importância da transposição de um Documento Físico para o meio digital, o método em questão tem se desenvolvido no Laboratório de Pesquisa de Linguística de Corpus (LAPELINC) da UESB, para construção de *corpora* digitais anotados cientificamente controlados (Santos e Namiuti 2016) e estabelece práticas para a transposição material do documento, originalmente em papel, para o suporte digital. Assim, dando continuidade à vertente linguística do projeto temático, mais especificamente à construção de corpus para estudos linguísticos. Este projeto de iniciação científica tem como objetivo colaborar na construção do corpus DOVIC (Corpus de Documentos Oitocentistas de Vitória da Conquista e Região), tencionando tanto viabilizar pesquisas científicas sobre os atributos linguísticos presentes nos documentos, quanto preservar a memória e a história da época e da região.

PALAVRAS-CHAVE: Linguística de Corpus; DOVIC; Sintaxe

TOPOGRAPHY OF DOCUMENTS FROM NOTEBOOKS E1 (1841-1848), E2 (1841-1855), E5 (1852-1866) OF THE CORPUS DOVIC (NINETEENTH CENTURY DOCUMENTS FROM VITÓRIA DA CONQUISTA AND REGION)

ABSTRACT

This expanded summary presents the results of the research conducted as part of the scientific initiation project “TOPOGRAPHY OF DOCUMENTS FROM THE NOTEBOOKS E1 (1841-1848), E2 (1841-1855), AND E5 (1852-1866) OF THE DOVIC CORPUS

¹ Estudante da Universidade Estadual do Sudoeste da Bahia, emillyviana194@gmail.com

² 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, cristianenamiuti@gmail.com

(NINETEENTH-CENTURY DOCUMENTS FROM VITÓRIA DA CONQUISTA AND REGION): A CONTRIBUTION TO THE EXPANSION OF RESEARCH SOURCES ON EIGHTEENTH-CENTURY PORTUGUESE IN BRAZIL“ which is linked to a larger project called “FROM PRECLASSICAL PORTUGUESE TO MODERN VARIANTS: CONTRIBUTIONS TO THE STUDY OF SYNTAX AND INTERFACE” which is being developed at Lapelinc (Corpus Linguistics Research Laboratory), a laboratory created and coordinated by Cristiane Namiuti (PPGLin/UESB) and Jorge Viana (PPGLin/UESB). Considering the importance of transposing a physical document to the digital medium, the method in question has been developed and applied to build scientifically controlled annotated digital corpora and establishes controlled practices for the material transposition of the document, originally on paper, to digital media. Thus, continuing the linguistic aspect of the thematic project, more specifically the construction of corpora for linguistic studies, this scientific initiation project aims to collaborate in the construction of the DOViC corpus (Corpus of Eighteenth-Century Documents from Vitória da Conquista and Region), with the intention of enabling scientific research on the linguistic attributes present in the documents and preserving the memory and history of the period and the region.

KEYWORDS: Corpus Linguistics; DOVic; Syntax.

INTRODUÇÃO

A Linguística de Corpus é o campo de pesquisa, cujos estudos são realizados através da análise de grandes conjuntos de textos, denominados *corpora*, para identificar características de uso da língua nestes contextos. Em vista disso, ela se ancora em ferramentas de coleta e análise desses dados, a fim de auxiliar os processos de construções teóricas acerca de léxico, gramática e outras vertentes linguísticas da língua escrita e falada.

As fontes documentais antigas são essenciais para muitos estudos na área das humanidades, que avançam nas discussões atuais recorrendo ao passado. No entanto, a limitação destes documentos em formato físico em função da relação de submissão entre eles, o tempo e o espaço é um grande problema neste processo.

A partir disso, os pesquisadores em linguística de *corpus* vêm buscando criar e renovar meios para a compilação de *corpora* atuais, através da união dos conhecimentos acerca da Linguística, Computação e Filologia. Com o objetivo de compreender a importância das fontes documentais antigas para diversos estudos na área das humanidades, esta pesquisa vem sendo realizada a partir da leitura do material teórico, realização de reuniões e tabulação de dados, pretendendo analisar e transcrever documentos antigos e entender como são gerados os dados a partir disso. Considerando essa conjuntura, focalizamos no auxílio da construção do corpus DOVic (Corpus de Documentos Oitocentistas de Vitória da Conquista), através da topografia de documentos dos Livros de Notas E 1 (1841-1848), E 2 (1841-1855), E 5 (1852-1866). Estes documentos são importantes, não somente para a memória conquistense, como

também para todo o país, devido às informações que eles possuem sobre o período escravocrata do Brasil.

MATERIAIS E MÉTODOS

Como materiais e métodos para a realização da pesquisa em questão foram realizadas leituras bibliográficas de autores como Namiuti e Santos (2021; 2019), Santos e Brito (2021; 2014), Sardinha (2000) e Souza (2006), bem como a discussão desse referencial teórico; a participação como ouvinte de eventos científicos, a fim de compreender pormenorizadamente o mundo da pesquisa; topografia e transcrição dos cabeçalhos dos documentos do livro de nota E1 (1841-1848); reuniões como ferramentas de inspeção do trabalho realizado; e a tabulação dos dados levantados a partir das transcrições realizadas, conforme exemplo na Figura 1.

O método LAPELINC (Namiuti; Santos, 2021) estrutura-se em três etapas cruciais: (i) a transposição, que consiste na captura digital do documento físico por meio de fotografia cientificamente controlada; (ii) a transcrição, que materializa as minuciosas técnicas da paleografia para a digitalização fiel do texto; (iii) e a compilação, que consiste na anotação digital do documento em linguagem XML (Extensible Markup Language).

O percurso metodológico desta pesquisa constituiu na realização da topografia dos documentos do Livro de Notas E1, a partir das folhas imagens, reunindo as seguintes informações: (1) Número sequencial do documento localizado no livro de notas - Documento Micro (DMi) segundo Santos e Namiuti (2019); (2) Transcrição do cabeçalho; (3) Identificação da Tipologia do DMi; (4) Ano; (5) Folha Imagem Inicial; (6) Folha Imagem final.

Para a tabulação dos dados ligados ao próprio processo metodológico e às características dos documentos, registrou-se especificidades acerca do próprio método, como o tempo utilizado na topografia dos documentos, além de questões do livro topografado, como o número total de documento, a quantidade de palavras por cabeçalho e a frequência tipológica dos documentos que compõem o livro.

RESULTADOS E DISCUSSÃO

A partir das atividades realizadas no projeto de Iniciação Científica, elaboramos uma planilha de controle de dados na qual pontuamos e registramos todas as atividades relativas ao processo de topografia, conforme demonstra a Figura 1.

FIGURA 1: RECORTE DA PLANILHA DE TOPOGRAFIA DO LIVRO DE NOTAS E1

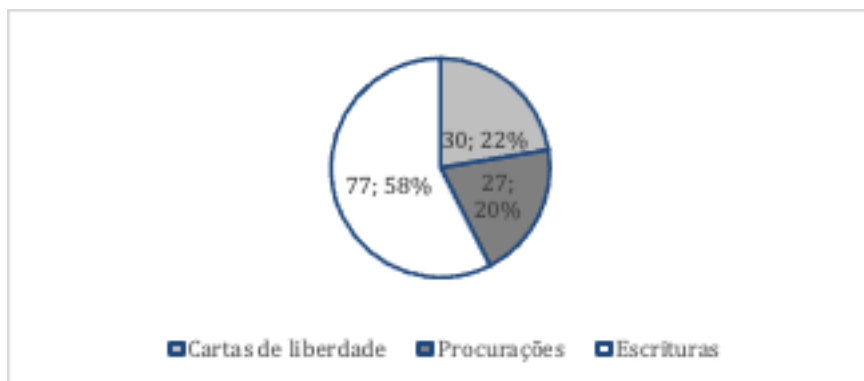
Topografia de documentos micro (DMi) no Livro de Notas – Documento Macro (DMa) E1

DMi	Cabeçalho	Tipo	Fecho	Ano	Folha- imagem inicial	Folha- imagem final
00	Termo de Abertura Este livro ha de servir para nele se lançar os papeis de Notas vai todo por mim numerado e rubricado com a rubrica que uso.	Termo de Abertura	Neste assinatura de João Dias de Miranda fis. 101-verso - assinatura de Luiz Fernandes de Oliveira fis. 107- assinatura de Maria Clemência do Amor Divino e de seu marido João de Oliveira Freitas fis. 132- ass. de [inint.] Fernandes Ribeiro 141 - ass. de José Nunes [inint.] de Jacinto Fernandes Ribeiro.	1980	002	002
01	Procuração bastante que faz Liberto Hansalves de Figueiredo na melhor forma e direito para o que abaixo se declara.	Procuração	Imperio aos desesete dias do me de fevereiro do ditto anno neste Fazenda da Soledade termo diz e declara que nam tem effeito esta escripta, de que dou fé/ (assinatura)	1841	007	007
02	Procuração bastante que faz Liberto Hans Alves de Figueiredo para o que a baxo se declara.	Procuração	tambem assignarão com o [inint.] depois de lido este perante todos, que sam reconhecidos de mim pelos proprios, de que tracto, e dou fé. (assinaturas)	1841	007	010
3	Carta de Liberdade do Cabra de nome João por Anthonomazia Gato, [inint.] pelo seo Senhor Ignacio Gonçalves Viana, e appres entada por Lodovico Gonçalves Chaves, como abaixo se declara.	Carta de Liberdade	Nada mais se continha essa ditta carta de liberdade que o contheudo aqui descripto [inint.] Tabellião abaixo assignado	1841	010	012

Fonte: (Elaborado pelas autoras)

A partir da análise destes dados foram topografados 134 documentos e transcritas 5387 palavras de cabeçalhos e fechos sendo 30 cartas de liberdade, 77 escrituras e 27 procurações (cf. figura 2).

FIGURA 2: TIPOS DE DMis TOPOGRAFADOS NO LIVRO DE NOTAS E1



Fonte: Gráfico elaborado pelas autoras

As folhas imagens do Livro de Notas E1 foram essenciais no processo de topografia. Sobre isso, Santos; Brito (2021) revelam que a fotografia cientificamente controlada do método Lapelinc, além de preservar o documento físico e facilitar seu acesso, podem também viabilizar a conservação dos materiais, graças à edição fotográfica.

CONCLUSÕES/CONSIDERAÇÕES

Ao longo do período de Iniciação Científica foi realizada a topografia dos documentos do Livro de Notas E1 (1841-1848) do corpus DOVic, do qual foram analisadas 131 folhas imagens, dentre as quais havia 30 cartas de liberdade, 77 escrituras e 27 procurações, resultando em 134 documentos, conforme demonstrou a figura 2 gerada a partir do controle de dados.

Conclui-se que o método LAPELINC traz inovações essenciais para o desenvolvimento do trabalho do formador de *corpora* digitalmente anotados, tendo em vista o aumento da praticidade e da qualidade do trabalho produzido. Trata-se, portanto, de um conjunto de etapas controladas sistemática e cientificamente, nas quais destaca-se a fotografia científica, proporcionando maior praticidade para a análise dos documentos.

Através do método LAPELINC, colaboramos com a construção do corpus DOVIC (Documentos Oitocentistas de Vitória da Conquista e Região), garantindo fidedignidade aos documentos originais, particularidade essencial para a realização de pesquisas científicas baseadas neste corpus.

REFERÊNCIAS BIBLIOGRÁFICAS

COSTA, Aline; NAMIUTI, Cristiane. WebSinC: buscas on-line em corpora sintaticamente anotados. In: Humanidades digitais e o mundo lusófono. Organização Ricardo M. Pimenta, Daniel Alves. Rio de Janeiro: Editora FGV, 2021, págs. 37-47.
Faria, Pablo e Charlotte Galves. 2016. «Criando «Bancos de Árvores»: o Sistema de Anotação e o Processo Automático». Cadernos de Estudos Linguísticos 58(2): 299–315.

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.

O futuro das Humanidades Digitais é o passado. In: Carrilho, E.; Martins, A.M.; Pereira, S.; Silvestre J.P., Organizadores. Estudos Linguísticos e Filológicos Oferecidos a Ivo Castro. Lisboa: CLUL-Centro de Linguística da Universidade de Lisboa; 2019. p. 1381-1403.

PAIXÃO DE SOUSA, Maria Clara. O Corpus Tycho Brahe: contribuições para as humanidades digitais no Brasil. Filologia e Linguística Portuguesa, [S. l.], v. 16, n. esp., p. 53–93, 2014.

PAIXÃO DE SOUSA, Maria Clara. Memórias do Texto. Revista Texto Digital, v. 2, n. 1, 2006

XXIX Seminário de Iniciação Científica e Tecnológica 2025

SARDINHA, Tony. Linguística de corpus: histórico e problemática. In RETROSPECTIVA/RETROSPECTIVE D.E.L.T.A., Vol. 16, N.º 2, 2000(323-367)

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.

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