

TOPOGRAFIA DE DOCUMENTOS DOS LIVROS DE NOTAS E15 (1881-1882)  
DO CORPUS DOVIC (DOCUMENTOS OITOCENTISTAS DE VITÓRIA DA  
CONQUISTA E REGIÃO)

Rebeca Mendes da Silva Gomes<sup>1</sup>, Cristiane Namiuti<sup>2</sup>

RESUMO

Esta pesquisa faz parte do subprojeto “Do português pré-Clássico às variantes modernas: contribuições para o estudo da sintaxe e interfaces.”, o qual vem sendo desenvolvida no Laboratório de Pesquisa em Linguística de Corpus (Lapellic). Dada a importância da fidedignidade entre o documento físico (DF) e a sua versão digital (DDI) para os estudos diacrônicos (Namiuti; Santos, 2019), utilizamos o Método LAPELINC para construção de *corpora* digitais anotados cientificamente controlados (Santos e Namiuti 2016). Este trabalho objetiva dar continuidade à vertente linguística do projeto temático, contribuindo para a construção do Corpus de Documentos Oitocentistas de Vitória da Conquista e Região (DOVIC), por meio da transcrição e topografia dos cabeçalhos de todos os 62 documentos que compõem o Livro de Notas E15 (1881-1882). A análise descritiva apontou para uma predominância de procurações (46) e escrituras (12), com ocorrências menos frequentes de cartas de liberdade (2), título (1) e ata (1). Além disso, os resultados demonstraram a eficácia do processo, que alcançou uma produtividade média de 186 palavras transcritas por hora, comprovando a rastreabilidade e o controle científico para a construção de *corpora* digitais anotados.

PALAVRAS-CHAVE: Linguística de Corpus; Topografia; LAPELINC; DOVIC; Diacronia.

TOPOGRAPHY OF DOCUMENTS FROM NOTEBOOKS E15 (1881-1882) FROM THE  
DOVIC CORPUS (19TH CENTURY DOCUMENTS FROM VITÓRIA DA CONQUISTA  
AND REGION)

ABSTRACT

This research is part of the subproject "From pre-Classical Portuguese to modern variants: contributions to the study of syntax and interfaces," developed at the Corpus Linguistics Research Laboratory (Lapellic). Given the importance of fidelity between the physical document (PD) and its digital version (DDI) for diachronic studies (Namiuti; Santos, 2019), we employed the LAPELINC Method for the construction of scientifically controlled annotated digital corpora (Santos and Namiuti 2016). This work aims to contribute to the thematic project's linguistic focus by constructing the Corpus of Nineteenth-Century Documents from Vitória da Conquista and Region (DOVIC). This was achieved through the transcription and topography of the headers of all 62 documents comprising the Notary Book E15 (1881-1882). The descriptive analysis pointed to a predominance of power of attorney documents (46) and deeds (12), with

---

<sup>1</sup> Graduada em Letras Vernáculas pela Universidade Estadual do Sudoeste da Bahia (UESB), *campus* Vitória da Conquista. Aluna voluntária de Iniciação Científica. E-mail: rebecamsdgonmes5@hotmail.com

<sup>2</sup> Professora do Departamento de Estudos Linguísticos e Literários (DELL) e do Programa de Pós-Graduação em Linguística (PPGLin) da Universidade Estadual do Sudoeste da Bahia (UESB). E-mail: cristianenamiuti@uesb.edu.br.

less frequent occurrences of letters of freedom (2), titles (1), and minutes (1). Furthermore, the results demonstrated the efficacy of the process, which achieved an average productivity of 186 transcribed words per hour, proving the traceability and scientific control required for the construction of annotated digital corpora.

KEYWORDS: Corpus Linguistics; Topography; LAPELINC Method; DOVIC; Diachrony.

## INTRODUÇÃO

A materialidade das fontes documentais, segundo Namiuti e Santos (2021), impõe um desafio significativo para a Linguística de *Corpus*. O acesso a esses documentos é limitado, exigindo que o pesquisador, para consultá-los, esteja no mesmo ambiente físico e atente para as rigorosas condições de manuseio e conversação. Desse modo, o suporte material do objeto, ao mesmo tempo em que limita o acesso, “caracteriza um modo de fazer humanidades” (Namiuti; Santos, 2021).

Nesse íterim, os avanços tecnológicos ocorridos no fim do século XX proporcionaram o surgimento de suportes digitais para essas fontes documentais, ampliando as possibilidades e alcance de consulta e manuseio. Essa migração requer um cuidado metodológico, pois, como afirma Paixão de Sousa (2006), é de suma importância integrar distintos planos de análise para garantir uma abordagem global do texto.

Os estudos diacrônicos realizados com base em textos antigos no meio eletrônico prezam pela fidedignidade entre o documento físico (DF) e a sua versão digital (DDI) (Namiuti; Santos, 2019). Em razão disso, os pesquisadores Cristiane Namiuti e Jorge Santos desenvolveram um método que viabiliza o resgate de informações do DF no ambiente digital. Para construir *corpora* eletrônicos anotados, cientificamente controlados, o Método LAPELINC desenvolve-se em três grandes etapas: Transposição, Transcrição e Compilação. Nessas etapas, são desenvolvidas atividades de “coleta de material, levantamento, organização, armazenamento e divulgação de dados” (Namiuti; Santos, 2021), além de promover a divulgação desse material para pesquisadores interessados.

O presente trabalho, desenvolvido no Lapelinc<sup>3</sup>, insere-se em um projeto maior denominado “Do português pré-clássico às variantes modernas: contribuições para o estudo da sintaxe e interfaces”, e tem como objetivo descrever as atividades de transcrição e topografia realizadas para desenvolvimento do Corpus de Documentos Oitocentistas de Vitória da Conquista (DOVIC). Para tal, focalizamos na análise do Livro de Notas E15 (1881-1882).

---

<sup>3</sup> Sigla para Laboratório de Pesquisa em Linguística de Corpus (Lapelinc/UESB).

## MATERIAIS E MÉTODOS

A pesquisa de Iniciação Científica seguiu uma metodologia que integrou a fundamentação teórica e o procedimento prático, a fim de que conhecêssemos melhor as atividades desenvolvidas no laboratório e o referencial teórico correspondente. Na primeira fase, o trabalho pautou-se na realização de encontros destinados à leitura e discussão de um referencial teórico fundamental, a saber, textos de Costa e Namiuti (2021), Namiuti e Santos (2019; 2021), Santos e Brito (2021), Sardinha (2021) e Paixão e Souza (2006). Na fase prática, por sua vez, realizou-se a transcrição e topografia dos cabeçalhos dos documentos do Livro de Notas E15 (1881-1882) e, ainda, a tabulação dos dados levantados a partir dessas atividades.

Como dito anteriormente, o fluxo de trabalho do Método LAPELINC estrutura-se em três etapas: (i) Transposição - passagem do DF para um DDI, por meio de fotografia cientificamente controlada (FCC); (ii) Transcrição - materialização das técnicas paleográficas para a digitalização fidedigna do texto; (iii) Compilação - anotação digital do documento em linguagem XML (Extensible Markup Language).

Com os documentos do Livro de Notas E15 (1881-1882) já transpostos para DDI, para o presente trabalho, nosso foco metodológico recaiu sobre a etapa de transcrição, por meio da aplicação de técnicas de topografia no referido livro, a partir das folhas imagens, compilando as seguintes informações: (i) Número sequencial do documento localizado no livro de notas - Documento Micro (DMi); (ii) Transcrição do cabeçalho; (iii) Identificação da tipologia do DMi; (iv) Ano; (v) Folha imagem inicial; e, por fim, (vi) Folha imagem final.

A tabulação dos dados, etapa realizada em paralelo aos avanços da topografia, visou uma análise quantitativa e descritiva. Nessa fase, foram registrados não apenas aspectos do próprio Livro de Notas E15 (a exemplo do número total e da frequência tipológica dos documentos), mas também aspectos do processo metodológico (como o tempo de topografia de cada documento).

## RESULTADOS E DISCUSSÃO

Com base nas atividades executadas no projeto de Iniciação Científica, elaboramos uma planilha de controle de dados destinada a pontuar e registrar todas as ações relativas ao processo de topografia, conforme detalhado na Tabela 1.

**TABELA 1:** Recorte da planilha de topografia do Livro de Notas E15

## XXIX Seminário de Iniciação Científica e Tecnológica 2025

**Topografia de documentos micro (DMi) no Livro de Notas – Documento Macro (DMa) E15-C11**

DMi	Cabeçalho	Tipo	Fecho	Ano	Folha- imagem inicial	Folha- imagem final
0	Este livro servirá para <u>nottas</u> do primeiro <u>Tabellião</u> desta <u>Villa</u> , e vai por mim <u>numera</u> do e rubricado com minha rubrica que fiz <u>JGFseilas</u> ; essa fim leva a termo de en[?][?]amento -	Termo de abertura		8 de novembro de 1881	0004	0004
1	Procuração bastante que <u>fasem</u> <u>Alfrãdo</u> Xavier Soares e João Fely de Araujo Costa, á seus procuradores como abaixo se declara.	Procuração		08 de novembro 1881.	0005	0006
2	Procuração bastante que faz Antonio [?]ues[?]hi[?] Lima Lopes, promotor publico desta Comarca da Victoria, como abaixo se declara.	Procuração		12 de novembro de 1881	0006	0007
3	Procuração bastante que faz Ismael José <u>Liol</u> [?][?]sis Mares á seus procuradores como abaixo se declara.	Procuração		26 de novembro de 1881	0007	0008
4	Procuração bastante que faz Dona Rachel Maria de São José, a seus procuradores como abaixo declara.	Procuração		Data do cabeçalho: 26 de novembro de 1881	0009	0010

Fonte: Elaborado pelas autoras (2025).

Os resultados práticos do projeto demonstram a realização da topografia dos cabeçalhos dos 62 documentos que compõem o Livro de Notas E15 (1881-1882), totalizando 1159 palavras transcritas. Quanto à distribuição tipológica, predominam as procurações (46 ocorrências), seguidas pelas escrituras (12 ocorrências), e por outras ocorrências menos frequentes (2 cartas de liberdade, 1 título e 1 ata).

Além disso, a análise do controle de dados revelou uma média de aproximadamente 186 palavras transcritas por hora. A tabulação dos dados permitiu registrar um indicador de produtividade: a média de palavras transcritas por hora avançou significativamente ao longo dos dias (cf. tabela 2). Este resultado, quantificado pela planilha de controle, é essencial para garantir o controle científico na construção de *corpora* eletrônicos fidedignos.

**TABELA 2:** Média de palavras transcritas por hora

	Dia1	Dia2	Dia3	Dia4
Palavras	68	228	533	330
Horas	1,07	2,15	1,51	1,49
Palavras/Hora	63,55140187	106,0465116	352,9801325	221,4765101

Fonte: Elaborado pelas autoras (2025).

### CONCLUSÕES/CONSIDERAÇÕES

O trabalho de Iniciação Científica alcançou seu objetivo ao realizar a transcrição e a topografia dos cabeçalhos dos 62 documentos que compõem o Livro de Notas E15 (1881-1882), um volume documental do *corpus* DOVIC. A análise descritiva desses

resultados evidenciou uma predominância de procurações e escrituras, indicando a riqueza da fonte para estudos diacrônicos da língua no contexto sociopolítico oitocentista.

Por meio dessa pesquisa, colaboramos com a construção do corpus DOVIC, assegurando a fidedignidade do material digitalizado em relação aos originais. Além disso, vale salientar que a utilização da planilha de controle e a mensuração da produtividade (aproximadamente 186 palavras/hora) corroboraram o Método LAPELINC como um conjunto de etapas controladas de modo sistemático e científico. Desse modo, o Método se consolida como um grande avanço nas Humanidades Digitais, uma vez que viabiliza a realização de coletas, levantamentos, organização, armazenamento e divulgação de dados para a construção de *corpora* eletrônicos anotados.

#### REFERÊNCIAS BIBLIOGRÁFICAS

COSTA, Aline Silva; NAMIUTI, Cristiane. WebSinC: buscas on-line em corpora sintaticamente anotados. *In*: PIMENTA, Ricardo M.; ALVES, Daniel. (org.). **Humanidade digitais e o mundo lusófono**. Rio de Janeiro: Editora FGV, 2021, p. 37-47.

NAMIUTI, Cristiane; SANTOS, Jorge Viana. Novos desafios para antigas fontes: a experiência DOVIC na nova linguística histórica. *In*: PIMENTA, Ricardo M.; ALVES, Daniel. (org.). **Humanidade digitais e o mundo lusófono**. Rio de Janeiro: Editora FGV, 2021, p. 69-89.

\_\_\_\_\_. O futuro das humanidades digitais é o passado. *In*: CARRILHO, E.; MARTINS, A. M.; PEREIRA, S.; SILVESTRE, J.P. (org.) **Estudos Linguísticos e Filológicos Oferecidos a Ivo Castro**. Lisboa: 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**, v. 16, n. esp., p. 53-93, 2014.

\_\_\_\_\_. **Memórias do Texto**. Revista Texto Digital, n. 2, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/textodigital/article/view/1420>. Acessado em: 29 ago. de 2025.

SARDINHA, Tony. **Linguística de corpus**: histórico e problemático. *In*: Retrospectiva/Retrospective D.E.L.T.A., vol. 16, n. 2, p. 323-367, 2000.

SANTOS, Jorge; BRITO, Giovani. A fotografia digital na transposição de documentos manuscritos históricos. *In*: PIMENTA, Ricardo M.; ALVES, Daniel. (org.). **Humanidade digitais e o mundo lusófono**. Rio de Janeiro: Editora FGV, 2021, p. 103-115.

\_\_\_\_\_. Fotografia técnica de documentos para formatação de corpora digitais eletrônicos: o método desenvolvido no Lapelinc. **Letras & Letras**, v. 30, n. 2, p. 421-430, 2014.