



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXIX SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2025**

Cartas selecionadas do padre jesuíta português, Manuel da Nóbrega, escritas no Brasil entre 1549 e 1560: aplicação de teste de inteligência artificial (IA) no processo de edição

Gabriela Fortalesa de Carvalho¹; Zenaide de Oliveira Novais Carneiro²

1. Gabriela Fortalesa de Carvalho, PIBIC/CNPq, LETRAS-LINGUA PORTUGUESA, e-mail: gabifortalesa@gmail.com

2. Zenaide de Oliveira Novais Carneiro, DEPARTAMENTO DE LETRAS E ARTES, e-mail: zoncarneiro@uefs.br

PALAVRAS-CHAVE: filologia, inteligência artificial, edição semidiplomática

INTRODUÇÃO

No âmbito do projeto “Gramáticas do português colonial (1500-1822): contato entre línguas e novos acervos do CE-DOHS”, vinculado ao “*Corpus* Eletrônico de Documentos Históricos do Sertão” (CE-DOHS) e ao “Núcleo de Estudos de Língua Portuguesa” (NELP/UEFS), desenvolveu-se um estudo filológico e sócio-histórico de duas cartas autógrafas e parcialmente autógrafas do padre jesuíta português Manuel da Nóbrega (1517-1570). Além da análise linguística e histórica, a pesquisa propôs a avaliação do uso de ferramentas de Inteligência Artificial (IA) aplicadas ao processo de edição, com foco no *software Transkribus*, já amplamente explorado em estudos de filologia e paleografia digital (Lose, 2020).

A constituição da coleção documental de escritos de missionários jesuítas e de outras ordens religiosas no Brasil Colonial insere-se nas agendas do Projeto para a História do Português Brasileiro (PHPB), sistematizadas por Castilho (1998) e consolidadas em campos interligados (Lobo; Oliveira, 2012). Este trabalho contribui, sobretudo, para os campos histórico-filológico e sócio-histórico, ao reunir documentos inéditos e avaliar as condições de produção e circulação desses textos.

Nesse contexto, destaca-se a relevância da documentação jesuítica para a Linguística Histórica, uma vez que possibilita estudos que buscam compreender a transplantação do português para o Brasil no século XVI (Castro, 1996; Mattos e Silva, 2008). Além disso, a aplicação de IA no processo de edição aproxima a pesquisa das Humanidades Digitais, ampliando o diálogo entre a tradição filológica e os recursos tecnológicos contemporâneos.

MATERIAL E MÉTODOS

A pesquisa fundamentou-se em uma abordagem interdisciplinar que articula Linguística Histórica, Filologia, Paleografia e Humanidades Digitais. O *corpus* é composto por duas cartas autógrafas ou parcialmente autógrafas de Manuel da Nóbrega (1517-1570), escritas entre 1549 e 1560 e localizadas no Arquivo Nacional da Torre do Tombo e no *Archivum Romanum Societatis Iesu* (Arquivo Romano da Companhia de

Jesus). A busca documental foi realizada a partir da prospecção em catálogos especializados, como a coleção *Monumenta Brasiliae* (Leite, 1956), e de edições já consolidadas (Leite, 1931; 1938-1950), já as cartas selecionadas, de 12 inicialmente previstas para o desenvolvimento do plano, reduziram-se a 2, uma vez que muitos originais estão perdidos, muitos são em outras línguas que não a portuguesa e outros são apógrafos.

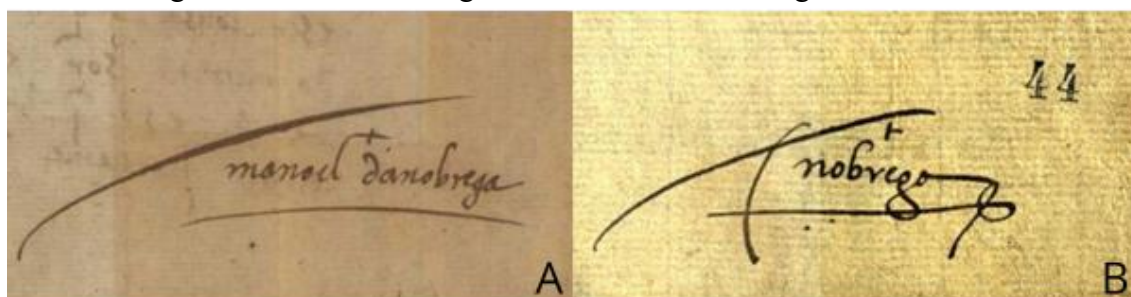
Os procedimentos metodológicos envolveram: (i) levantamento bibliográfico e arquivístico para identificação das fontes; (ii) construção do perfil histórico-social dos remetentes com base em critérios propostos por Carneiro (2005); (iii) aplicação de critérios editoriais semidiplomáticos, adaptados por Lemos (2024) a manuscritos do século XVI, visando à preservação das características gráficas originais; e (iv) experimentação com a plataforma *Transkribus*, utilizada para o reconhecimento e transcrição automática dos manuscritos (Lose, 2020).

O uso da ferramenta de IA buscou avaliar a eficácia do reconhecimento automático de texto manuscrito, comparando a transcrição bruta gerada pelo sistema com a edição semidiplomática final. Essa etapa possibilitou mensurar o desempenho da plataforma em termos de acurácia, bem como discutir suas potencialidades e limitações no contexto da paleografia digital.

RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)

O processo de prospecção documental possibilitou identificar duas cartas originais de Manuel da Nóbrega adequadas à edição filológica, originais e autógrafos (ou parcialmente): a carta ao Rei Dom João III, escrita em 1551, e a carta para Padre Miguel de Torres, de 1557. A análise paleográfica confirmou a autenticidade dos manuscritos, com destaque para elementos como caligrafia regular, assinatura autógrafa (Figura 1) e características gráficas recorrentes em documentos jesuítas do período.

Figura 1: assinatura autógrafa de Manoel da Nóbrega nos dois documentos



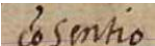
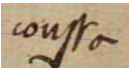
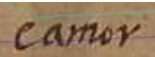
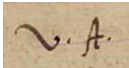
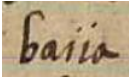
Legenda 1: Assinatura do padre Manoel da Nóbrega presente nas cartas. A: Assinatura da carta a Dom João III, B: Assinatura da carta ao Padre Miguel de Torres.

A edição semidiplomática, realizada conforme critérios definidos por Lose (2000) e adaptados por Lemos (2024), preservou aspectos formais da escrita original, respeitando abreviaturas, disposição gráfica e convenções do *scriptor*. Essa metodologia assegurou a validade dos dados para estudos em Linguística Histórica e Filologia. As edições estão em processo de publicação no site do CE-DOHS.

Na aplicação da plataforma *Transkribus* para a transcrição e edição, o teste foi feito com a primeira carta editada. Foram processados cerca de 1.300 vocábulos, com

taxa de erro aproximada de 5,9%. Entre os erros mais frequentes, destacaram-se confusões gráficas como <e>/<c>, <, e ss>/<st> e <i>/<r>, influenciadas por particularidades paleográficas. Apesar dessas limitações, o sistema forneceu uma transcrição inicial consistente, reduzindo o tempo de trabalho editorial e permitindo ao pesquisador concentrar-se na revisão crítica. É importante ressaltar que os documentos estavam em bom estado de conservação, bem como eram boas digitalizações, em alta qualidade, o que pode ter influenciado no sucesso do programa.

Tabela 1: Erros mais frequentes no reconhecimento de texto pelo modelo escolhido no

Confusão gráfica	Número de ocorrências	Exemplo de excerto da carta	Interpretação do modelo	Interpretação correta
Troca de <c> por <l>	9		lo sentio	cõsentio
Troca de <ss> por <st>	8		cousta	coussa
Troca de <c> por <e>	6		camor	eamor
Troca de <A> por <R>	4		V. R.	V.A.
Troca de <i> por <r>	4		baria	baia

Transkribus.

Legenda 1: 1. Troca de <c> por <l>; 2. Troca de <s> por <st>; 3. Troca de <c> por <e>; 4. Troca de <A> por <R>; 5. Troca de <i> por <r>. São indicados o número de ocorrências, o excerto da carta, a interpretação do modelo e a interpretação correta.

Os resultados demonstram que a integração de ferramentas de Inteligência Artificial ao trabalho filológico pode otimizar a etapa de transcrição, sem substituir a análise especializada. Dessa forma, reafirma-se a pertinência da combinação entre Humanidades Digitais e métodos tradicionais da Filologia, contribuindo para a constituição de *corpora* diacrônicos e para a investigação da história social da língua portuguesa (Castro, 1996; Mattos e Silva, 2008).

CONSIDERAÇÕES FINAIS

A pesquisa evidenciou a relevância das cartas de Manuel da Nóbrega para os estudos em Linguística Histórica e Filologia, ao mesmo tempo em que demonstrou o

potencial das ferramentas de Inteligência Artificial no apoio à edição de manuscritos coloniais. A experiência com o *Transkribus* revelou que, embora apresente limitações relacionadas à acurácia do reconhecimento de caracteres, a plataforma oferece um ponto de partida eficiente para a transcrição e análise de textos antigos, contribuindo para otimizar o trabalho filológico.

Ao integrar métodos tradicionais de edição com recursos digitais, este estudo reforça a importância das Humanidades Digitais como campo de diálogo entre tecnologia e pesquisa linguística. Além de fornecer subsídios para a constituição de *corpora* diacrônicos, os resultados abrem caminho para novas investigações na área dos estudos linguísticos que ampliem a compreensão de quais variedades da língua portuguesa chegaram ao Brasil com as caravelas, bem como da história social da língua portuguesa no Brasil colonial.

REFERÊNCIAS

- CASTILHO, A. T. de. 1998. Gramática do português falado: introdução. Campinas: Editora da UNICAMP.
- CARNEIRO, Z. O. N. 2005. Cartas brasileiras (1809-1904): um estudo linguístico-filológico. Campinas: UNICAMP, Tese (Doutorado).
- CASTRO, I. 1996. Para uma história do português clássico. In: Actas do Congresso Internacional sobre o Português, v. II. Lisboa: Associação Portuguesa de Linguística, p. 135-150.
- LEITE, S. 1931. Cartas do Brasil e mais escritos do P. Manuel da Nóbrega (Opera Omnia). Rio de Janeiro: Officina Industrial Graphica.
- LEITE, S. 1938-1950. História da Companhia de Jesus no Brasil. Lisboa: Portugália; Rio de Janeiro: Instituto Nacional do Livro, 10 v.
- LEITE, S. (Ed.). 1956. Monumenta Brasiliae I (1538-1553). Roma: Monumenta Historica Societatis Iesu.
- LEMOS, R. C. 2024. Critérios de edição de manuscritos quinhentistas. Salvador: EDUFBA.
- LOSE, A. D. 2020. Paleografia digital e ferramentas de inteligência artificial. Revista de Filologia e Linguística Portuguesa, 22(1), p. 45-60.
- LOBO, T.; OLIVEIRA, K. 2012. Linguística histórica do português brasileiro: avanços e perspectivas. São Paulo: Contexto.
- MATTOS E SILVA, R. V. 2004. Ensaio para uma sócio-história do português brasileiro. São Paulo: Parábola.
- MATTOS E SILVA, R. V. 2008. Caminhos da linguística histórica: ouvir o inaudível. São Paulo: Parábola.