



Sociolinguística Histórica, Filologia e Humanidades Digitais: um olhar interdisciplinar e transdisciplinar sobre manuscritos oitocentistas do Rio Grande de São Pedro

Historical Sociolinguistics, Philology and Digital Humanities: an interdisciplinary and transdisciplinary approach to nineteenth-century manuscripts from Rio Grande de São Pedro

*Evellyne Patrícia Figueiredo de Sousa COSTA**

*Leici Landherr MOREIRA***

RESUMO: Este artigo tem como objetivo construir o perfil sociocultural do primeiro secretário da capitania do Rio Grande de São Pedro analisando a documentação indireta e as variações gráficas presentes nos manuscritos oitocentistas sob o arcabouço teórico da Sociolinguística Histórica e sua relação com a Filologia e Humanidades Digitais. Os resultados obtidos por meio da utilização das ferramentas do software AntConc possibilitaram a identificação das taxas de acerto/erro de etimologizações no *corpus*, identificando o conhecimento de língua latina do escrevente.

PALAVRAS-CHAVE: Sociolinguística Histórica. Filologia. Humanidades Digitais. Interdisciplinaridade. Transdisciplinaridade.

ABSTRACT: This article aims to build the sociocultural profile of the first secretary of the captaincy of Rio Grande de São Pedro by analyzing the indirect documentation and graphic variations present in nineteenth-century manuscripts. The theoretical framework of this study is Historical Sociolinguistics and its relationship with Philology and Digital Humanities. The results obtained through the use of the AntConc software tools made it possible to identify the hit/miss rates of etymology in the corpus. The results identify the writer's knowledge of the Latin language.

KEYWORDS: Historical Sociolinguistics. Philology. Digital Humanities. Interdisciplinarity. Transdisciplinarity.

Artigo recebido em: 31.07.2023

Artigo aprovado em: 21.12.2023

* Doutorado em Teoria e Análise Linguística pela PUC/RS. Professora Associada da UFSM. evellyne.costa@ufsm.br

** Doutoranda em Linguística pela UFSM. Professora na Antonio Meneghetti Faculdade (AMF). leicimoreira@hotmail.com

1 Introdução

Este artigo tem como objetivo discutir as contribuições teórico-metodológicas das Humanidades Digitais para a Filologia e a Sociolinguística Histórica. As Humanidades Digitais podem ser compreendidas, desde o surgimento nos anos quarenta do século XX, como uma área de investigação que relaciona humanidades e computação, do ponto de vista do uso de ferramentas e dados digitais, metodologias ou métodos digitais ou a investigação vinculada aos fenômenos culturais e artefatos digitais (Piotrowski, 2020, p. 3 *apud* Costa; Ramos; Almeida; Campos, 2021, p. 1). Nesse sentido, aspectos como liberação dos manuscritos; acesso, partilha e preservação de patrimônio; organização do conhecimento no domínio dos *Big Data* e investigação científica são questões em voga nas discussões sobre a aproximação das Humanidades Digitais e a Filologia. Nosso trabalho adiciona as imbricadas relações entre o linguístico e o social em sincronias passadas a essas discussões na medida em que propõe a reconstrução do perfil sociocultural do primeiro secretário da Capitania do Rio Grande de São Pedro, atual estado do Rio Grande do Sul, por meio da documentação indireta em *corpora* diacrônicos eletrônicos e a aplicabilidade da ferramenta computacional AntConc na análise de variações gráficas desse escrevente em um *corpus* de manuscritos oitocentistas, sob a ótica da Sociolinguística Histórica. Esse pressuposto teórico-metodológico concebe a língua como “inerentemente variável” tanto do ponto de vista sincrônico quanto diacrônico e condiciona a variação a fatores estruturais e sociais. Desse modo, pretendemos demonstrar como a ocorrência de determinadas variações gráficas nos manuscritos que compõem o *corpus* de pesquisa podem ser motivadas pelo emprego de formas linguísticas escritas no padrão da língua latina (etimologizações¹), em virtude da exigência do conhecimento

¹ Apesar de o termo “etimologização” compreender a origem em diferentes línguas, neste trabalho, é utilizado como equivalente à expressão “latinismo”, isto é, diz respeito somente às marcas relativas à grafia de origem latina.

de latim para a função de secretário (Freire, 1746, 1756) e de o escrevente já ter exercido o ofício de professor de gramática latina.

Partindo da perspectiva filológica, nossa investigação entende que a Filologia desempenha um importante papel na “libertação dos textos” em vários aspectos: (i) os documentos do nosso *corpus* são coletados em museus, arquivos públicos e, muitas vezes, não há uma política de acesso ou publicização do material; (ii) a leitura de manuscritos não é uma tarefa fácil para o público distante da prática filológica, portanto entendemos que as transcrições possibilitam a leitura dos textos; (iii) o labor filológico possibilita, também, a partilha ou publicização de documentos via publicação de edições.

Com o advento de ferramentas digitais, a partilha dos documentos em mananciais, hipertextos, dicionários digitais, vocabulários, edições digitais ou plataformas potencializa as contribuições da Filologia no sentido de ampliar o acesso aos textos, de contribuir para a preservação do patrimônio cultural e de possibilitar construção de *corpora* para estudos linguísticos. Nesse âmbito, encontram-se diversas pesquisas em Lexicologia, História da Língua, Linguística de *Corpora*, dentre outras. Os trabalhos de Quaresma (2013) e Lazzari e Finatto (2023) são exemplos de pesquisas que aproximam os estudos linguísticos do uso de ferramentas computacionais na seara dos *corpora* diacrônicos.

No contexto da construção do nosso *corpus* de estudo de uma sincronia passada, nossa pesquisa se aproxima da Filologia por conta do tratamento dos manuscritos, das decisões no tocante às normas de transcrição, da análise paleográfica, dentre outros aspectos. Ao mesmo tempo, neste trabalho, nos aproximamos das Humanidades Digitais ao utilizarmos acervos documentais históricos em formato digital para a reconstrução do perfil sociocultural² do escrevente e o aparato metodológico da

² Na perspectiva da Sociolinguística Histórica, compreendemos perfil sociocultural conforme a conceituação proposta por Lopes *et alii* (2010) em relação ao cruzamento de perspectivas que relacionam a trajetória de vida dos sujeitos, o contexto de produção dos textos e o mapeamento e descrição das redes de escrita.

ferramenta digital AntConc para extrair automaticamente dos manuscritos informações importantes para a análise linguística.

Nesse sentido, inicialmente apresentamos a Sociolinguística Histórica como pressuposto teórico-metodológico à luz das Humanidades Digitais para explicar o fenômeno de variação pela correlação entre fatores linguísticos e sociais analisados em registros escritos antigos e como as Humanidades Digitais contribuem para a reconstrução de variáveis sociais e análise de *corpora*; em seguida, discorremos sobre o contexto sócio-histórico da produção e circulação dos manuscritos, sobretudo, ao que se refere ao perfil sociocultural do escrevente e descrevemos o *corpus* de pesquisa. A seguir, explicamos a metodologia de trabalho para constituição do *corpus* na perspectiva filológica da edição dos manuscritos e o método de seleção e análise dos dados para verificar a taxa de uso e acerto dos latinismos nas variantes gráficas dos manuscritos. Por fim, apresentamos os resultados. Com este trabalho, será possível perceber, portanto, que as relações entre Humanidades Digitais, Filologia e Sociolinguística Histórica são marcadas pela interdisciplinaridade e transdisciplinaridade.

2 Pressupostos teóricos da Sociolinguística Histórica à luz das Humanidades Digitais

A Sociolinguística Histórica é uma área de estudo transdisciplinar e interdisciplinar, em que vários paradigmas e orientações de pesquisa são reunidos, pois são necessárias várias perspectivas para compreender a língua no passado: sociolinguística, dialetologia, sociopragmática, estudos discursivos, Linguística Histórica, História Social, Linguística de *Corpus* e Filologia (Nevalainen; Raumolin; Brunberg, 2012; Nevalainen, 2015; Auer; Peersan; Pickl; Rutten; Vosters, 2015). Em seu escopo, concebe a língua como uma heterogeneidade ordenada (Labov, 1972), englobando a variação e mudança, aspectos da Linguística Histórica, e questões da Sociolinguística Variacionista, como a articulação entre fatores linguísticos e sociais.

Desse modo, rompe-se a fronteira entre sincronia e diacronia ao se usar o presente para explicar o passado (Princípio do Uniformitarismo).

De fato, é a concepção de língua que embasa a Teoria da Variação Linguística, a aplicação de seus métodos aos estudos de estágios de língua do passado e a construção de *corpora* que esboçam a Sociolinguística Histórica. No entanto, devido à especificidade do *corpus* de registros antigos, pesquisadores propõem um refinamento teórico-metodológico para efetivar o objetivo de descrever e explicar os fenômenos de variação e de mudança pela correlação entre fatores linguísticos e sociais analisados em registros escritos antigos; assim, desde o trabalho pioneiro de Suzane Romaine (1982), linguistas tais como Menéndez (1995), com *Sociolinguística Histórica*, Roger Lass (2000), Fernández (2005), com a obra *Historia Social de las Lenguas de España*, Conde-Silvestre (2007) com *Sociolinguística Histórica*, dentre outros, vêm contribuindo para desenvolver os estudos na área.

Rutkowska e Rössler (2012, p. 213), por seu turno, defendem que a variabilidade inerente à língua se aplica também ao nível da ortografia, além dos níveis geralmente da fonologia, morfologia e sintaxe. Segundo os autores, o sistema ortográfico de uma língua está associado a essas áreas, logo, também é influenciado por uma variedade de fatores como aspectos diacrônicos, diatópicos, diafásicos e diastráticos (Rutkowska; Rössler, 2012, p. 217). Nesse sentido, para os linguistas, a definição mais adequada para variável ortográfica³ consistiria em:

uma variável ortográfica é uma característica de um sistema ortográfico de uma dada língua, relacionado aos níveis fonológico, morfológico ou lexical desse sistema linguístico e realizados por diferentes variantes sob circunstâncias extralinguísticas específicas (Rutkowska, Rössler, 2012, p. 219, tradução nossa).

³ Como não há uma norma ortográfica unificada no século XIX, período em que são produzidos os manuscritos que compõem o *corpus* de pesquisa, em nosso trabalho, empregamos o termo “variação gráfica” para designar a definição proposta por Rutkowska e Rössler (2012).

Na perspectiva da Sociolinguística Histórica, os fatores concernem ao tempo de produção do texto, localização geográfica, sexo, idade, aproximação com a escrita, redes sociais, tipo, estilo, registro do texto etc. Essa pesquisa, como a dos demais autores supracitados, vale-se de *corpora* histórico-diacrônicos disponíveis em acervos documentais, arquivos e museus. Mais recentemente, os pesquisadores assinalam as contribuições das Humanidades Digitais.

Nevalainen (2015), por exemplo, no artigo *What are historical sociolinguistics?* do primeiro volume do *Journal of Historical Sociolinguistics*, já enfatiza a importância das Humanidades Digitais na reconstrução de contextos macro e micro pela disponibilidade de *corpora* digitais de edições de textos muito maiores e de ferramentas para analisá-los. Realmente, segundo Conde-Silvestre (2007), a reconstrução das variáveis sociais do falante em seu contexto sócio-histórico consiste em uma das problemáticas enfrentadas pelo pesquisador da Sociolinguística Histórica, uma vez que a maioria dos registros antigos não fornece informações sobre o seu escrevente; nesse caso, intenta-se identificar as informações extralinguísticas sobre sua origem, escolaridade, faixa etária, profissão etc. em documentação indireta lotada em arquivos e acervos históricos. Nesse sentido, o acesso a *corpora* digitais de textos com mecanismos de busca, a exemplo do acervo de documentos datados do período colonial, lotados no Arquivo Histórico Ultramarino, mas disponibilizados pelo Projeto Resgate de Documentação Histórica Barão do Rio Branco, na Biblioteca Nacional Digital do Brasil, viabiliza a identificação de aspectos do perfil sociocultural do escrevente essenciais para a reconstrução biográfica e sua correlação com a análise linguística, como será destacado na próxima seção.

Sayli, Nurmer, Palander-Collin e Auer (2017), na mesma perspectiva de Nevalainen (2015), para responder à pergunta que intitula seu artigo *The future of historical sociolinguistics?*, propõem uma série de novos caminhos para a Sociolinguística Histórica como parte do campo mais amplo das Humanidades Digitais, referindo-se, principalmente, à utilização de ferramentas computacionais

para a seleção e tratamento dos dados. Essa abordagem, por sua vez, dialoga diretamente com outra problemática imposta ao pesquisador da língua em sincronias passadas: a construção do *corpus* com material linguístico histórico. Segundo Conde-Silvestre (2007, p. 48, tradução nossa), a Linguística de *Corpus* oferece a possibilidade de “lidar simultaneamente com a maioria dos textos preservados, de um mesmo período histórico, ou os analisa longitudinalmente em diferentes períodos contribuindo para resolver em parte o problema da natureza fragmentária dos materiais históricos”. Além disso, o autor (Conde-Silvestre, 2007, p. 47) enfatiza o uso das ferramentas computacionais, pois permite o trabalho com uma quantidade grande de dados, superando as limitações de um trabalho baseado somente na acuidade visual para detectar todos os processos e fenômenos linguísticos no estudo da mudança e variação da língua.

Nas próximas seções, nosso intuito é demonstrar essas implicações no estudo de determinadas variantes gráficas em manuscritos do primeiro secretário da Capitania do Rio Grande de São Pedro pela reconstrução do perfil sociocultural a partir da documentação indireta em *corpora* diacrônicos digitais e seleção e análise dos dados de variação por meio do software gratuito AntConc que permite obter de modo instantâneo listas de formas, frequência de uso e contextos de um grande número de textos ao mesmo tempo.

3 O ofício de secretário na Capitania de São Pedro: dos preceitos de Cândido Lusitano à prática de escrita da correspondência do governador Paulo Gama

No período colonial, compete à escrita a materialização da decisão administrativa, permitindo, assim, a guarda da memória político-administrativa da colônia (Hespanha, 2001). Nesse cenário, está situado o secretário, considerado pelo Conselho Ultramarino desde 1702 como um ofício de suma importância em todos os governos do ultramar. Como agentes oficiais da escrita junto aos governadores das capitanias, os secretários desempenhavam atividades de produção, circulação e

organização da informação administrativa, ocupando-se da redação dos atos comunicativos, provisões, cartas patentes e datas de sesmaria, além da organização de todos os documentos expedidos e recebidos a partir do processo de transcrição. Por essa razão, para desempenhar a função, circulavam manuais acerca do que se esperava do ofício de secretário e sua prática de escrita tal como “O Secretario Portuguez Compendiosamente Instruido no modo de Escrever Cartas” (1746) de Francisco José Freire, também conhecido como Cândido Lusitano.

Segundo Conceição (2011), o compêndio de Francisco José Freire, reeditado várias vezes ao longo dos séculos XVIII e XIX, é considerado a mais importante obra portuguesa na época moderna sobre a prática de escrita de cartas com grande permanência e circulação no Brasil. Araújo (1999) apontou a obra como o segundo livro mais presente em inventários e testamentos do Brasil Colonial, e Barbosa atesta sua presença em anúncios de livros dos periódicos das províncias brasileiras no século XIX (Barbosa, 2011). Neste trabalho, consideramos a edição de 1756, publicada na “Officina de Domingos Gonsalves” e composta das seguintes partes: capa; instrução preliminar; das perfeições do secretário (capítulo 1); das imperfeições do secretário (capítulo 2); das regras que o secretário deve praticar nas cartas de negócio (capítulo 3); índice de modelos de cartas composto de duas partes; formulário de tratamentos, muy necessario ao secretario portuguez; formulario de Sobrescritos.

Na “Instrução Preliminar”, Freire define a arte de escrever cartas como um exercício comum para a comunicação com os ausentes. Todavia, a composição dos documentos exige um método e boa forma; por isso, propõe diretrizes e exemplificações ao secretário principiante a fim de “facilitar o caminho aos que se applicarem a taõ nobre, como preciso emprego”. (Freire, 1756, §ii) para que “com respeito, e louvor sustentar o character da sua nobre occupação, e igualmente o da pessoa, a quem servir” (Freire, 1756, §ii). Apesar das instruções, é notório que o secretário apresenta vivo engenho, conhecimento das línguas latinas e materna,

conhecimentos dos autores que escreveram sobre o modo de compor as cartas. Desse modo, enfatiza:

Naõ temos que encommendar a perfeição da orthografia, e pontuação; porque naõ só supponho o Secretario perfeito nesta parte, mas ainda no interior conhecimento de lingua Materna, e Latina, como cousa taõ essencial. Se tiver a mesma noticia dos Idiomas mais polidos da Europa, mayor lustre dará ao seu nobre emprego (Freire, 1756, fol. 6r).

No primeiro e segundo capítulos, Freire discorre a respeito de dez regras sobre perfeições que o secretário deve ter e vícios que deve evitar. Estes seriam demora, prolixidade, aspereza, ignorância, escuridade; aquelas, segredo, erudição, generalidade, reflexão e eloquência. E encerra com os modelos de cartas e formulário de tratamento e sobrescrito.

Não há como afirmarmos a utilização desta obra pelos secretários do Rio Grande de São Pedro em sincronias pretéritas, no entanto, uma vez que o compêndio de Cândido Lusitano possui grande circulação no Brasil e pressupõe que o Secretário disponha de perfeito conhecimento da língua latina, propomos verificar tal conhecimento nas variantes gráficas da prática de escrita de correspondências do primeiro secretário nomeado por D. João, em 1806, para a Capitania do Rio Grande de São Pedro: Francisco José de Freitas.

Reconstruímos a trajetória de vida de Francisco de Freitas com o auxílio da pesquisa histórica bibliográfica e da consulta à documentação indireta (ESTEBAN-SEGURA, 2012). Apesar de ser uma pessoa pública, há uma escassez de informações nos estudos sobre a história do Rio Grande do Sul; em razão disso, no acervo digital disponibilizado pelo Projeto Resgate na Biblioteca Nacional Digital do Brasil, examinamos documentos redigidos ou que apenas fizessem menção a Francisco de Freitas. Dessa forma, reconstruímos sua trajetória, especialmente no Brasil, observando os ofícios ocupados.

O capitão de Milícias de Mato Grosso é português, mas vem ao Brasil por Provisão da Rainha Dona Maria de 1789, encaminhado para Villa Bela do Mato Grosso no intuito de preencher o cargo de professor substituto de gramática latina, pelo período de seis anos, assistido financeiramente pelo rendimento do subsídio literário.⁴ Em 17 de maio de 1806, Francisco Freitas é nomeado secretário do governo do Rio Grande de São Pedro para exercer o ofício durante três anos ou até que outro secretário fosse nomeado.⁵ Entre as suas atribuições, é responsável pela cópia das cartas dirigidas ao Ministério desde 1802 até 1808 dispostas no Códice A1.01 do fundo documental “Documentação dos Governantes”, lotado no Arquivo Histórico do Rio Grande do Sul. Assim, considerando o exercício da profissão de substituto de gramática latina e secretário, nas próximas seções apresentamos dez documentos do códice A1.01 a fim de explicar as variações gráficas dos textos pela taxa de uso e acerto de etimologizações, além de observar o quão próximo Francisco Freitas estaria das características previstas para um secretário, conforme postulado por Francisco José Freire.

4 *Corpus de pesquisa*

O *corpus* deste trabalho é composto de dez documentos manuscritos do códice A1.01 do fundo documental “Documentação dos Governantes”, localizado no acervo do Arquivo Histórico do Rio Grande do Sul (AHRS). Esse códice apresenta a cópia da correspondência expedida pelos governantes da Capitania de São Pedro ao Ministério de 1802 a 1808, copiadas do livro de registros sob a inspeção do secretário do governo Francisco José de Freitas em 1807.

Todos os documentos pertencem ao âmbito da administração pública e consistem em cópias de ofícios expedidos por Paulo José da Silva Gama, governador

⁴ AHU, Conselho Ultramarino, Brasil – Mato Grosso, 1803, caixa 41, documento nº 2051.

⁵ AHU, Conselho Ultramarino, Brasil – Rio Grande do Sul, 1806, caixa 10, documento nº 636.

da Capitania do Rio Grande de São Pedro no período de 1803 a 1807, ao secretário de estado da Marinha e Ultramar, João Rodrigues de Sá e Melo, nomeado Visconde de Anadia, a respeito dos seguintes assuntos:

Quadro 1 – Descrição do *corpus*.

Número do ofício	Data	Destinatário	Assunto
1	21 de março de 1803	Visconde de Anadia	Dúvidas da câmara da capitania sobre a nomeação de cirurgião-mor.
2	25 de julho de 1803	Visconde de Anadia	O estado da cultura do linho-cânhamo.
3	25 de julho de 1803	Visconde de Anadia	Lucros a obter com o arrendamento da fazenda do Bojuru e dos Povos.
4	4 de dezembro de 1803	Visconde de Anadia	Mapa demonstrativo do aumento das rendas da Fazenda Real da capitania.
5	4 de dezembro de 1803	Visconde de Anadia	Impossibilidade de enviar tábuas estatísticas da capitania.
6	5 de maio de 1804	Visconde de Anadia	Ordem da Secretaria de Estado da Fazenda para cobrar direitos sobre os artigos de luxo.
7	25 de agosto de 1804	Visconde de Anadia	Informação sobre a solicitação de licença do tenente de dragões Antonio Pinto da Fontoura.
8	25 de agosto de 1804	Visconde de Anadia	Relação da entrada de embarcações do comércio, estabelecimentos comerciais e casas construídas na vila de Porto Alegre desde 1803.
9	25 de setembro de 1806	Visconde de Anadia	Rendimento anual da extinta Provedoria e da Junta da Fazenda Real e das despesas anuais da Junta e da tropa da capitania.
10	4 de março de 1807	Visconde de Anadia	Suspeitas de falsificação do decreto que promove o alferes Francisco Osório do Amaral Sarmiento.

Fonte: elaborado pelas autoras.

Como cópias, os documentos não são de autoria intelectual de Francisco José de Freitas; ainda assim, ao que se refere ao processo de cópia, o secretário realiza alterações, dentre as quais destacamos a perspectiva gráfica tal como podemos verificar no cotejo entre a edição semidiplomática da segunda via do documento, lotada no Arquivo Histórico Ultramarino, disponibilizada pelo Projeto Resgate de Documentação Histórica Barão do Rio Branco na Biblioteca Nacional Digital do Brasil e a edição semidiplomática da cópia do manuscrito no códice A1.01 do AHRS:

Quadro 2 – Comparação entre as variações gráficas da segunda via e cópia do documento 2.

Fragmento da segunda Via do documento 2	Fragmento da cópia do documento 2
Naó remeto o Mappa da Popolação por que os Vigarios apezar de lhe ter escripto logo que tomei posse deste Governo, pedindo-lhe huma rella- ção circunstanciada dos individuos das Suas Freguezias, ainda alguns o naó tem cumprido, e logo que elles a tenhaó feito o porei presença de <i>Vossa Excelencia</i> .	Naó remetto o Mappa da População porque os Vigarios apezar de lhe ter es cripto logo que tomei posse deste Governo, pedindo-lhe huma rellação circunstanciada dos individuos das- Suas Freguezias, ainda alguns a naó deraõ, e logo <i>que</i> elles a tenhaõ feito o porei presença de <i>Vossa Excelencia</i> .

Fonte: elaborado pelas autoras.

Pelo contraste entre os fragmentos da segunda via e cópia do documento 2, salientamos que, mesmo consistindo em cópias, é a ocorrência de modificações de ordem gráfica, sobretudo, no âmbito dos latinismos pela mão do secretário que torna o *corpus* representativo para proposta de análise deste trabalho.

Além disso, essa variação pauta-se na ausência de uma normatização unificada para a escrita no século XIX, uma vez que somente em 1904 surge a primeira tentativa de unificação da ortografia, proposta por Gonçalves Viana. Assim, a instabilidade gráfica reflete os principais aspectos de períodos anteriores na história da grafia: fonético e pseudoetimológico (Coutinho, 1976). Embora o período fonético compreenda o século XII até o século XVI, em documentos do século XIX, ainda se

veem formas linguísticas que buscam representar na escrita os sons da fala, ao mesmo tempo que os latinismos também orientam a escrita das palavras. Nesse caso, é o período pseudoetimológico, iniciado na metade do século XVI até o século XX, em que se intensifica a grafia de formas latinas, ainda que as palavras nem sempre coincidam com a origem real na língua clássica. Por isso, uma vez que os documentos do *corpus* datam do período pseudoetimológico da língua, é pertinente analisar os usos e acertos das etimologizações pela mão do escrevente associado ao seu perfil sociocultural tal como será delineado na próxima seção.

5 Metodologia

Nesta seção, em primeiro lugar, apresentamos a metodologia de trabalho para a constituição do *corpus* de pesquisa desde a coleta dos manuscritos até a transcrição filológica e escolha da edição dos textos. Em seguida, discorreremos acerca da metodologia empregada para o tratamento dos dados, fundamentada nas Humanidades Digitais, pelo uso da ferramenta computacional AntConc para seleção e organização das formas linguísticas a fim de aplicar na perspectiva da Sociolinguística Histórica a proposta de Barbosa (2005, 2006), Barbosa e Lima (2019) e Monaretto (2021) para verificar as taxas de uso/acerto de etimologizações do secretário Francisco José de Freitas e aferir o quão próximo estaria de uma característica prevista para o ofício de secretário como o perfeito conhecimento da língua materna (língua portuguesa) e da língua latina, considerando o exercício da profissão de substituto de gramática latina. Dessa maneira, será possível delinear o perfil sociocultural desse escrevente.

No Arquivo Histórico do Rio Grande do Sul, realizamos a coleta dos manuscritos do códice A1.01 do fundo documental “Documentação dos Governantes” através da fotografia digital com uma câmera semiprofissional Fujifilm Finepix S3300, 14mp, Zoom 26x, sem o uso de flash e com a utilização de luvas descartáveis para não danificar o suporte e selecionamos 10 documentos.

A partir dos fac-símiles, classificamos os manuscritos conforme a espécie documental proposta por Bellotto (2002) para o período colonial com base na distinção entre carta para documentos endereçados ao Rei e ofício para documentos destinados à autoridade inferior ao soberano. Os documentos foram transcritos segundo as normas para edição semidiplomática, tendo em vista que tem se sobressaído na edição de documentos para o estudo da história da língua portuguesa (Cabraia, 2005, p. 96). Além disso, sua transcrição fiel viabiliza a identificação de variantes gráficas, especialmente no cotejo entre originais e cópias, tal como apontado no quadro 2 da seção anterior.

De acordo com Cabraia (2005), a edição semidiplomática diz respeito à edição com um grau médio de intervenção, uma vez que o editor efetua pequenas modificações tais como o desdobramento de abreviaturas, inserção ou supressão de elementos por conjectura, mas transcreve fielmente os caracteres romanos, diacríticos e sinais de pontuação. Seguimos, portanto, as normas do autor para a edição semidiplomática fazendo apenas duas adaptações: (i) a divisão das linhas do manuscrito será preservada; no entanto, optamos por assinalá-la pela marca de uma barra vertical | entre as linhas, indicando a mudança de fólio entre duas barras verticais || 1v. || 2r. || 2v. || 3r.; (ii) uniformizamos a separação vocabular segundo o sistema atual, baseado no vocábulo morfológico, assim, separamos hipossegmentações e juntamos hipersegmentações, pois entendemos que, nesses documentos, tais processos decorrem dos aspectos materiais da escrita (pena e tinta), sem motivações de ordem linguística.

Posteriormente, os arquivos com as edições semidiplomáticas foram convertidos individualmente para o formato .txt a fim de serem utilizados no software AntConc. O AntConc é um software desenvolvido pelo professor Laurence Anthony, Professor na Faculdade de Ciência e Engenharia da Waseda University, no Japão, para análise textual e Linguística de *Corpus*. Uma vez que a ferramenta computacional gera listas e frequências de palavras a partir de grandes quantidades de documentos,

possibilita a captação precisa e análise dos dados apurada para aferir a taxa de uso/acerto de etimologizações conforme proposto por Barbosa (2005, 2006), Barbosa e Lima (2019) e Monaretto (2021).

Barbosa (2005, 2006) apresenta uma proposta metodológica de construção de um parâmetro objetivo para averiguar a norma culta oitocentista fundamentada na etimologização gráfica. Trata-se da medição quantitativa do emprego de etimologias greco-latinas para verificar o grau de cultura dos redatores, haja vista que o século XIX é conhecido como período pseudoetimológico em virtude da aplicação errônea da relatinização de várias palavras portuguesas. Nessa perspectiva, a proposta de Barbosa (2005, 2006) pretende aferir o grau de cultura erudita dos redatores das cartas pessoais da família Ottoni por meio de vocábulos que evidenciam uma grafia original latina. De acordo com o linguista, o maior número de grafias latinas em um autógrafo oitocentista (taxa de uso) já indicaria o contato do redator com modelos de erudição; além disso, sendo a grafia consoante à forma original latina (taxa de acerto), maior seria o grau de conhecimento da norma culta da época. Para Barbosa (2005, 2006), portanto, o controle dessas taxas é o que permitiria avaliar a proximidade do redator com valores da norma culta de sua sincronia. Desse modo, o linguista estabelece os jornais como modelo textual da norma objetiva oitocentista, anotando, inicialmente as taxas de uso e de acerto nesses textos-modelo para, finalmente, averiguar se as taxas correspondem àquelas aferidas nas cartas dos avós da família Ottoni, de 1875, para seus dois netos, Misael e Christiano. A partir do referencial definido para o controle do grau de cultura por meio de jornais — taxa de uso de etimologizações (3,69%) e taxa de acerto de forma motivadas no significante latino (90,35%) —, classificou o grau de cultura de proximidade à cultura oitocentista do redator em: erudito, informante culto, informante de cultura mediana e semiculto.

Em 2019, Barbosa e Lima confirmam a aplicabilidade da metodologia a um conjunto de cartas pessoais entre Rui Barbosa e seu amigo Carlos de Aguiar, obtendo taxas muito próximas à do avô da família Ottoni e aos jornais. Contudo, os autores

propõem avanços no método para tornar os parâmetros como valores referência à gradiência de redator erudito, redator letrado, redator mediano e redator elementar para cartas pessoais de desconhecidos (Barbosa; Lima, 2019, p 180). Com o objetivo de controlar os graus de erudição escrita no século XIX, somado à aferição qualitativa da taxa de acertos, propõe o caminho qualitativo do contraste da taxa de uso por *tokens* e por *types*, medindo, assim, a penetração do uso de latinismos na variedade/riqueza vocabular do redator. A oposição diz respeito ao número total de ocorrências de vocábulos latinizados do *corpus* analisado (*tokens*) e o total de lexemas latinizados de mesma base de sentido e classe morfológica (*types*). Nas cartas pessoais entre Rui Barbosa e Carlos Aguiar, por exemplo, há nas grafias latinizadas de Rui Barbosa uma proporção de 249 *tokens*/115 *types*, enquanto nos latinismos de Carlos Aguiar há 162 *tokens*/61 *types*, revelando, portanto, a gradiência entre erudito e letrado (Barbosa; Lima, 2019, p. 184).

Monaretto (2021), por sua vez, ratifica o método de Barbosa (2005), aplicando o teste objetivo a um *corpus* de 78 correspondências manuscritas trocadas nos anos 1800, entre 22 remetentes gaúchos com grau de parentesco com o casal Júlio de Castilhos e Honorina Martins da Costa Castilhos. A fim de propor a estratificação desses redatores em termos de graus de cultura escrita, a autora fundamenta-se no levantamento de características gráficas apontadas por Marquilhas (2000) e espelha-se na aferição objetiva das habilidades de escrita nas cartas pessoais do casal de avós Ottoni por meio do uso de etimologizações gráficas (Barbosa, 2005). A partir da análise do *corpus*, Monaretto (2021, p. 83) define o grau de cultura por um teste quantitativo de uso e acerto de etimologizações motivadas pelo significante greco-latino e a associação com riqueza lexical do texto (Berber Sardinha, 2004) pela análise de frequência de ocorrências (*tokens*) e o tipo de certa estrutura (*types*). Os resultados demonstraram que Júlio de Castilhos apresenta o maior uso de formas latinizadas (4,49%), alto percentual de acertos (92,53%) e a maior riqueza vocabular (84,44%, 144 *types*/161 *tokens*), condizente com seu grau de instrução de nível superior e sua prática de escrita

enquanto jornalista do jornal porto-alegrense *A Federação* (Monaretto, 2021, p. 94). Em contrapartida, Ana Martins, sua sogra, revela a taxa de 22,22% de riqueza vocabular, possivelmente, em virtude do acesso restrito à educação escolar na época. Portanto, com esses resultados, Monaretto (2021) valida a proposta de aferição do grau de cultura pela relação uso/acerto de formas etimologizadas, articulada com a riqueza lexical tal como proposto por Barbosa (2005, 2006) e Barbosa e Lima (2019).

No presente trabalho, empregamos as propostas de Barbosa (2005, 2006), Barbosa e Lima (2019) e Monaretto (2021) no que tange à avaliação da taxa de uso/acerto de grafias latinas. No entanto, diferentemente dos autores supracitados, não pretendemos delimitar o grau de cultura e erudição do secretário, mas explicar as variações gráficas oriundas dos latinismos e aferir o quão próximo Francisco José de Freitas estaria de uma característica prevista para o secretário tal como o perfeito conhecimento da língua materna (língua portuguesa) e da língua latina, considerando o exercício da profissão de substituto de gramática latina.

6 Resultados

A partir do *corpus* de pesquisa, com a ferramenta computacional AntConc, geramos uma lista precisa de 2.801 palavras, considerando o número de frequência, dentre as quais selecionamos os seguintes latinismos: consoantes duplas (*ll, tt, mm, ff, pp, nn*) como em *bella, literatto, commercio, effcás, mappa, anno*, e encontros consonantais impróprios (*pt, ct, gm*) como em *escripto, actividade, augmento*. Não selecionamos palavras com <h> e <y> tais como *hum, huma, ley, rey* porque já constam em vocabulário da época (Bluteau, 1728). Essa primeira seleção resultou em 136 etimologizações, indicando uma taxa de uso de 4,85% de latinismos.

Para verificar a taxa de acerto e erro dos latinismos, examinamos a etimologia de cada uma das 136 palavras conforme “Dicionário Etimológico da Língua Portuguesa” (Cunha, 2010). A título de exemplificação, segue a amostra com algumas palavras investigadas:

Quadro 3 – Identificação de acertos e erros das etimologizações.

PALAVRA	FREQUÊNCIA	VERDADEIRO	FALSO	ETIMOLOGIA
actividade	1	X		<i>activitas</i>
actuaes	1	X		<i>actualis</i>
actualmente	2	X		<i>actualis</i>
affectadas	1	X		<i>affecto</i>
allega	1	X		<i>allego</i>
allegando	2	X		<i>allego</i>
allucionou	1		X	<i>alucinor</i>
anno	11	X		<i>annus</i>

Fonte: elaborado pelas autoras.

Os resultados revelaram que o escrevente efetuou 127 acertos e 9 erros para 136 formas latinizadas, correspondendo a uma taxa de acerto de 93,38% e a uma taxa de erro de 6,61%. Com efeito, o alto índice da taxa de acertos confirma que Francisco José de Freitas possui conhecimento da língua latina conforme previsto para a função de secretário na obra “O Secretario Portuguez Compendiosamente Instruido no modo de Escrever Cartas” (1745) de Francisco José Freire. Do mesmo modo, podemos dizer que o perfil sociocultural de Francisco Freitas no tocante ao exercício anterior do ofício de professor substituto de gramática latina também explica a variação gráfica concernente aos latinismos.

Embora a variação gráfica possa ser superficialmente vinculada ao período de produção dos textos, período pseudoetimológico, o arcabouço teórico da Sociolinguística Histórica permite reconstruir o perfil sociocultural do escrevente por caminhos metodológicos específicos para correlacioná-lo ao fenômeno de variação.

No contexto do nosso *corpus*, a variável profissão foi concebida pelo acesso aos documentos escritos pelo secretário disponibilizados digitalmente pelo projeto Resgate. Mesmo de posse dessas informações, para proceder a uma análise cuidadosa e precisa dos dados, foi de suma importância o tratamento dos documentos por meio das ferramentas computacionais do software AntConc, que supera as limitações de um trabalho baseado na acuidade visual para detectar os fenômenos linguísticos e

variações gráficas, principalmente na circunstância de um *corpus* com uma grande quantidade de textos.

Nosso trabalho reafirmou a estreita relação entre Sociolinguística Histórica e Filologia e sedimentou, também, a relação entre essas áreas e as Humanidades Digitais no que tange ao acesso aos manuscritos, à construção de *corpora*, ao tratamento e à análise dos dados. Essas aproximações ratificam o caráter interdisciplinar e transdisciplinar dessas perspectivas.

7 Considerações finais

Este trabalho empreendeu um estudo da variação gráfica em documentos oitocentistas do primeiro secretário da Capitania do Rio Grande de São Pedro com o objetivo de identificar as taxas de erros e de acertos dos latinismos e, com isso, preencher as lacunas do perfil sociocultural desse escrevente. Para tanto, dois importantes aspectos foram arrolados nessa pesquisa: (i) a análise da documentação indireta disponível na plataforma do projeto Resgate como percurso metodológico para a reconstrução do perfil sociocultural do escrevente; (ii) a utilização das ferramentas computacionais do software AntConc, que permitiu a obtenção automática de listas de formas e taxas de acerto/erro das etimologizações, demonstrando o conhecimento da língua latina do escrevente.

Pudemos demonstrar que as Humanidades Digitais desempenham um importante papel no sentido de dar acesso aos documentos através de plataformas digitais, possibilitar a construção de *corpora* e mineração dos textos nas análises linguísticas, contribuindo sobremaneira para ampliar a natureza interdisciplinar e transdisciplinar da Filologia, da Sociolinguística Histórica e da própria área das Humanidades Digitais.

Referências

ARAÚJO, J. de S. **Perfil do leitor colonial**. Salvador: UFBA; Ilhéus: UESC, 1999.

AUER, A.; PEERSMAN, C.; PICKL, S.; RUTTEN, G.; VOSTERS, R. Historical sociolinguistics: the field and its future. **Journal of Historical Sociolinguistics**, v.1, p. 1-12, 2015. Disponível em: <https://www.degruyter.com/document/doi/10.1515/jhsl-2015-0001/html>. Acesso em: 30 jul. 2023. DOI <https://doi.org/10.1515/jhsl-2015-0001>

BARBOSA, S. de F. P. A escrita epistolar, a literatura e os jornais do século XIX: uma história. **Revista da Anpoll**, v. 1, n. 30, 2011. Disponível em: <https://revistadaanpoll.emnuvens.com.br/revista/article/view/196>. Acesso em: 30 jul. 2023. DOI <https://doi.org/10.18309/anp.v1i30.196>

BARBOSA, A. G. Tratamento dos Corpora de sincronias passadas da língua portuguesa no Brasil: recortes grafológicos e lingüísticos. In: LOPES, C. R. dos S. (org.). **A Norma brasileira em construção: Fatos lingüísticos em cartas pessoais do século 19**. Rio de Janeiro: UFRJ, Pós-Graduação em Letras Vernáculas: FAPERJ, 2005. p. 45-66.

BARBOSA, A. G.; LIMA, A. X. O controle indireto de perfis sócio-históricos em *corpora* histórico-diacrônicos: a identificação de graus de letramento pela grafia etimológica do século XIX. In: CASTILHO, A. T. de. (coord.). **História do português brasileiro**. São Paulo: Contexto, 2019. p. 168-205.

BARBOSA, A. G. Tratamento dos Corpora de sincronias passadas da língua portuguesa no Brasil: recortes grafológicos e lingüísticos. In: LOBO, T.; RIBEIRO, I.; CARNEIRO, Z.; ALMEIDA, N. (org.). **Para a história do português brasileiro**. Volume VI: Novos dados, novas análises. Tomo II. Salvador: EDUFBA, 2006, v. 2. p. 761-780.

BELLOTTO, H. L. **Como fazer análise diplomática e análise tipológica de documento de arquivo**. São Paulo: Arquivo do Estado, Imprensa Oficial, 2002.

BLUTEAU, R. **Vocabulario portuguez, e latino, aulico, anatomico, architectonico, bellico, botanico: autorizado com exemplos dos melhores escritores portuguezes, e latinos; e offerecido a El Rey de Portugal D. João V**. Coimbra: Collegio das Artes da Companhia de Jesus, Lisboa: Officina de Pascoal da Sylva, 1712-1728.

CAMBRAIA, C. N. **Introdução à crítica textual**. São Paulo: Martins Fontes, 2005.

CONCEIÇÃO, A. A. **Sentir, escrever e governar: a prática epistolar e as cartas de D. Luís de Almeida, 2º marquês do Lavradio (1768-1779)**. 2011. Tese (Doutorado em História) – Programa de Pós-graduação em História Social, Universidade de São Paulo, São Paulo, 2011.

CONDE SILVESTRE, J. C. **Sociolinguística histórica**. Madrid: Gredos, 2007.

COSTA, R.; ALMEIDA, B.; RAMOS, M.; CAMPOS, M. I. B. O papel da linguística na era das humanidades digitais. **Linha D'Água**, São Paulo, v. 34, n. 02, p. 1-8. 2021. DOI <https://doi.org/10.11606/issn.2236-4242.v34i2p1-8>

COUTINHO, I. de L. **Pontos de gramática histórica**. Rio de Janeiro: Ao Livro Técnico, 1976.

CUNHA, A. G. da. **Dicionário etimológico da língua portuguesa**. Rio de Janeiro: Lexicon, 2010.

FERNÁNDEZ, F. M. **Historia social de las lenguas de España**. Barcelona: Ariel, 2005.

FREIRE, F. J. **O secretario portuguez compendiosamente instruido no modo de escrever cartas. Por meyo de huma instrucçam. Preliminar, regras de Secretaria, Formulario de tratamentos, e hum grande numero de Cartas em todas as especies, que tem mais uso**. Lisboa: Oficina de Miguel Rodrigues, 1746.

FREIRE, F. J. **O secretario portuguez compendiosamente instruido no modo de escrever cartas. Por meyo de huma instrucçam. Preliminar, regras de Secretaria, Formulario de tratamentos, e hum grande numero de Cartas em todas as especies, que tem mais uso**. Lisboa: Officina de Domingos Gonsalves, 1786.

ESTEBAN-SEGURA, L. Medical, Official, and Monastic Documents in Sociolinguistic Research. *In*: HERNÁNDEZ-CAMPOY, J. M.; CONDE SILVESTRE, J. C. (org.). **The handbook of historical sociolinguistics**. Oxford: Blackwell, 2012. p. 140-155. DOI <https://doi.org/10.1002/9781118257227.ch8>

HESPANHA, A. M. A constituição do Império português. Revisão de alguns enviesamentos correntes. *In*: FRAGOSO, J.; BICALHO, M. F.; GOUVÊA, M. de F. (org.). **O Antigo Regime nos trópicos: a dinâmica imperial portuguesa (séculos XVI-XVIII)**. Rio de Janeiro: Civilização Brasileira, 2001. p. 161-188.

LABOV, W. On the mechanism of linguistic change. *In*: GUMPERZ, J.; HYMES, D. **Directions in sociolinguistics: the ethnography of communication**. New York: Hold, Rinehart and Winstion, 1972. p. 512-538.

LASS, R. **Historical linguistics and language change**. Cambridge: Cambridge University Press, 2000.

LAZZARI, R. R.; FINATTO, M. J. B. Exame do vocabulário médico no Português no século XVIII: contribuições da lexicometria para o desenho de um dicionário histórico.

Mandinga, v. 7, p. 102-123, 2023. Disponível em: <https://revistas.unilab.edu.br/index.php/mandinga/article/view/1259>. Acesso em: 30 jul. 2023.

MENÉNDEZ, F. G. **Sociolinguística histórica**. Madrid: Visor, 1995.

MONARETTO, V. N. de O. Para uma estratificação de redatores oitocentistas por graus de cultura escrita. *In*: VANDRESEN, P.; MARTINS, M. A. R.; MONGILHOT, I. de O. e S. (org.). **Variação e mudança linguística no português falado e escrito na região Sul e outros temas**. São Paulo: Blucher, 2021, v. 1. p. 81-104.

NEVALAINEN, T.; RAUMOLIN-BRUNBERG, H. Historical Sociolinguistics: Origins, Motivations, and Paradigms. *In*: HERNÁNDEZ-CAMPOY, J. M.; CONDE SILVESTRE, J. C. **The handbook of historical sociolinguistics**. Oxford: Blackwell, 2012. p. 22-40. DOI <https://doi.org/10.1002/9781118257227>

NEVALAINEN, T. What are historical sociolinguistics? **Journal of Historical Sociolinguistics**. v. 1, p. 243-269, 2015. Disponível em: <https://www.degruyter.com/document/doi/10.1515/jhsl-2015-0014/html>. Acesso em: 30 jul. 2023. DOI <https://doi.org/10.1515/jhsl-2015-0014>

QUARESMA, P. Análise linguística de documentos da BPE: uma abordagem informática. *In*: GONÇALVES, F., M.; BANZA, A. P. (coord.). **Património textual e humanidades digitais da antiga à nova filologia**. Évora: CIDEHUS, 2013. p. 139-156. DOI <https://doi.org/10.4000/books.cidehus.1091>

ROMAINE, S. **Socio-historical linguistics**. Cambridge: Cambridge Press, 1982. DOI <https://doi.org/10.1017/CBO9780511720130>

RUTKOWSKA, H; RÖSSLER, P. Orthographic Variables. *In*: HERNÁNDEZ-CAMPOY, J. M.; CONDE SILVESTRE, J. C. **The handbook of historical sociolinguistics**. Oxford: Blackwell, 2012. p. 213-236. DOI <https://doi.org/10.1002/9781118257227.ch12>

RUTKOWSKA, H; RÖSSLER, P. Orthographic Variables. *In*: HERNÁNDEZ-CAMPOY, J. M.; CONDE SILVESTRE, J. C. **The handbook of historical sociolinguistics**. Oxford: Blackwell, 2012. p. 213-236. DOI <https://doi.org/10.1002/9781118257227.ch12>

SÄILY, T.; NURMI, A.; PALANDER-COLLIN, M.; AUER, A. The future of historical sociolinguistics? *In*: SÄILY, T.; NURMI, A.; PALANDER-COLLIN, M.; AUER, A. (ed.). **Exploring future paths for historical sociolinguistics**. Amsterdam and New York: Benjamins, 2017. p. 1-19. DOI <https://doi.org/10.1075/ahs.7.01sai>