

Metodologia de constituição de corpora diacrônicos em Sociolinguística Histórica: Variação Pronominal em cartas pessoais Alagoanas

Methodology for the Constitution of Diachronic Corpora in Historical Sociolinguistics: Pronominal Variation in Personal Letters from Alagoas

Suziane de Oliveira Porto Silva*
Universidade Federal de Alagoas
Maceió, Alagoas, Brasil.

Resumo: Este artigo apresenta e discute os procedimentos metodológicos adotados na constituição de um *corpus* diacrônico de cartas pessoais alagoanas, desenvolvido no âmbito de uma pesquisa em Sociolinguística Histórica. A investigação teve como foco a análise da variação pronominal entre *tu* e *você* em posição de sujeito, com base em documentos produzidos entre as décadas de 1910 e 2000. O percurso metodológico contemplou a seleção criteriosa das fontes, a transcrição semidiplomática, a categorização das variáveis linguísticas e extralinguísticas e a aplicação de modelos estatísticos. Além disso, foram discutidos os desafios enfrentados durante a organização do *corpus* e as possibilidades analíticas que emergem desse tipo de material. Ao final, o estudo propõe uma sistematização metodológica voltada à constituição de corpora históricos que possa ser replicada em pesquisas de escopo semelhante, contribuindo para o fortalecimento empírico e interpretativo da Sociolinguística Histórica.

Palavras-chave: Sociolinguística Histórica. Corpus diacrônico. Variação pronominal. Cartas pessoais. Metodologia.

Abstract: This article presents and discusses the methodological procedures adopted in the construction of a diachronic corpus of personal letters from Alagoas, developed within the scope of a historical sociolinguistic study. The research focused on the analysis of pronominal variation between *tu* and *você* in subject position, based on documents produced between the 1910s and 2000s. The methodological path included the careful selection of sources, semi-diplomatic transcription, categorization of linguistic and extralinguistic variables, and the application of statistical models. Additionally, the article addresses the challenges encountered during corpus organization and highlights the analytical possibilities provided by such material. The study proposes a methodological framework for the constitution of historical corpora that can be replicated in similar investigations, contributing to the empirical and interpretative advancement of Historical Sociolinguistics.

Keywords: Historical Sociolinguistics. Diachronic corpus. Pronominal variation. Personal letters. Methodology.

1 INTRODUÇÃO

Nas últimas décadas, a constituição de corpora diacrônicos tem se consolidado como um dos principais alicerces das pesquisas em Sociolinguística Histórica, uma vez que permite a observação empírica de fenômenos linguísticos em perspectiva temporal e

* Doutora em Linguística pelo Programa de Pós-Graduação em Linguística e Literatura (PPGLL) da Universidade Federal de Alagoas (UFAL). E-mail: suziane.porto@hotmail.com

social. Com a ampliação dos estudos voltados à história das línguas naturais, especialmente no domínio do português brasileiro, tornou-se imprescindível o desenvolvimento de metodologias de organização documental que garantam tanto a autenticidade quanto a representatividade dos dados analisados. Neste contexto, a articulação entre a Linguística Histórica, a Sociolinguística Variacionista e disciplinas auxiliares como a Filologia e a Paleografia tem se mostrado fundamental para o fortalecimento epistemológico do campo (Romaine, 2009; Rumeu, 2012; Zilles; Faria, 2018).

A Sociolinguística Histórica, por sua própria natureza interdisciplinar, exige do pesquisador não apenas o domínio dos princípios teóricos da variação e mudança linguística (Weinreich; Labov; HErzog, 2006; Labov, 2008), mas também a familiaridade com os procedimentos técnicos e críticos de organização documental. Isso inclui desde a seleção criteriosa das fontes — considerando sua datação, autoria, circulação e legibilidade — até o tratamento paleográfico e a codificação linguística adequada às análises estatísticas. O corpus deixa, assim, de ser mero repositório e passa a funcionar como instrumento científico estruturado, cuja validade depende de decisões metodológicas transparentes e sistematizadas (Mattos e Silva, 2004; Martins et al., 2015).

Este artigo tem como objetivo apresentar e discutir os procedimentos metodológicos empregados na constituição de um corpus diacrônico de cartas pessoais alagoanas, elaborado no âmbito de uma pesquisa sociolinguística de cunho histórico. A investigação em questão analisou a variação pronominal entre *tu* e *você* em posição de sujeito em cartas manuscritas do século XX, com foco no estado de Alagoas, região ainda pouco contemplada nos estudos sobre o sistema de tratamento do português brasileiro. A escolha do gênero epistolar se justifica por sua natureza dialógica, pela recorrência de formas de tratamento e pelo potencial de revelar aspectos intersubjetivos como poder, solidariedade, afetividade e distanciamento social — fatores cruciais para a compreensão da escolha pronominal (Brown; Gilman, 1960; Marcuschi, 2001; Cardoso, 2010).

A proposta aqui delineada contempla desde os critérios de seleção dos documentos — considerando aspectos como autenticidade, condição física, proveniência e diversidade de remetentes — até os procedimentos de categorização das variáveis linguísticas e extralinguísticas, bem como o tratamento estatístico dos dados com o auxílio de software específico. Além disso, serão abordados os desafios enfrentados quanto à legibilidade dos manuscritos, à delimitação do tempo de produção das cartas, à atribuição de autoria e à classificação tipológica dos subgêneros epistolares.

Espera-se, com este relato metodológico, contribuir para o avanço das discussões sobre a organização de corpora históricos em Linguística, promovendo o intercâmbio de experiências e consolidando práticas que aliem rigor técnico, viabilidade empírica e sensibilidade documental. Ao evidenciar a complexidade envolvida na constituição de um corpus sociolinguístico de base diacrônica, o artigo busca não apenas expor um percurso de pesquisa, mas também suscitar reflexões sobre os modos de produção de conhecimento no campo das ciências da linguagem.

2. FUNDAMENTAÇÃO TEÓRICA

A construção de corpora diacrônicos voltados à análise sociolinguística histórica exige, antes de qualquer delineamento metodológico, a definição de um referencial teórico sólido, capaz de dar conta da complexidade do objeto investigado, ou seja, a variação e a mudança linguística ao longo do tempo e das especificidades dos dados mobilizados, como textos produzidos em sincronias pretéritas, atravessados por condicionantes gráficos, normativos, sociais e históricos. Partimos do princípio de que a análise linguística, especialmente em perspectiva histórica, apenas atinge interpretações consistentes quando sustentada por um arcabouço conceitual que articule língua, sociedade, tempo e uso.

Nesse sentido, adotamos como fundamento teórico central a Sociolinguística Variacionista, sistematizada por Weinreich, Labov e Herzog (2006), e a Sociolinguística Histórica, tal como proposta inicialmente por Romaine (1982) e ampliada, posteriormente, pelos trabalhos de Campoy e Conde Silvestre (2012), Rosa (2015) e outros autores que têm refletido sobre a aplicação de métodos sociolinguísticos à análise de textos diacrônicos. A articulação entre essas vertentes nos fornece os fundamentos necessários para a compreensão dos fenômenos linguísticos observados, ao mesmo tempo que orienta as etapas de seleção, organização e interpretação dos dados que compõem o corpus da pesquisa.

2.1 SOCIOLINGUÍSTICA VARIACIONISTA

A Sociolinguística Variacionista constitui uma ruptura paradigmática com os modelos linguísticos estruturalistas clássicos, particularmente com o modelo saussuriano que postulava a existência de uma língua abstrata, homogênea e invariante, centrada em um falante idealizado e desvinculada dos contextos reais de uso. A partir das formulações de Labov (2008) e da tríade fundacional composta por Weinreich, Labov e Herzog (2006), passa-se a considerar a língua como um sistema estruturado, mas fundamentalmente heterogêneo, cuja variação é regular, condicionada e observável.

Essa perspectiva teórica postula que as escolhas linguísticas dos falantes resultam da interação entre fatores internos ao sistema (como contexto fonológico, morfossintático e lexical) e externos a ele (como idade, gênero, escolaridade, grupo social, região de origem, estilo de fala, entre outros). Tais escolhas operam dentro de um conjunto de possibilidades regidas por regras variáveis nas quais os resultados de análise nos permitem identificar padrões de uso e tendências de mudança (COELHO et al., 2010).

Essa concepção de língua leva, inevitavelmente, à compreensão da mudança linguística como processo contínuo, progressivo e socialmente motivado. Embora nem toda variação implique mudança, toda mudança se dá a partir de um estado de variação (Paiva; Duarte, 2006). A coexistência entre formas concorrentes, em determinado estágio da história da língua, evidencia a natureza gradual do processo de substituição, como demonstra Faraco (2019), ao descrever a sucessão de fases em que variantes competem entre si até que uma se consolide como hegemônica. Nesse contexto, a observação da variação assume papel central na identificação dos vetores que impulsionam transformações linguísticas no tempo.

2.2 A EMERGÊNCIA DA SOCIOLINGÜÍSTICA HISTÓRICA

A ampliação do escopo metodológico da Sociolinguística Variacionista à análise de dados históricos deu origem à Sociolinguística Histórica, campo interdisciplinar que conjuga os pressupostos da Linguística Histórica com os métodos empíricos da Sociolinguística. Adotamos a proposta de Romaine (1982), que defende que os textos escritos antigos também carregam marcas de variação linguística e, portanto, podem ser analisados à luz dos mesmos princípios aplicáveis à língua falada.

Assentamo-nos, ainda, sobre o princípio do uniformitarismo, segundo o qual os mecanismos linguísticos e sociais observados no presente também atuaram no passado, com certo grau de estabilidade. Como observa Labov (2008), os processos que atuaram na mudança fonológica de séculos anteriores ainda podem ser observados nas mudanças em andamento. Tal premissa não implica homogeneidade absoluta entre contextos históricos, mas fornece uma base racional para a extrapolação de padrões e a interpretação de dados diacrônicos.

A Sociolinguística Histórica, ao lidar com registros escritos, enfrenta desafios metodológicos significativos: a escassez de dados, a falta de homogeneidade amostral, a ausência de metainformações sociais, as convenções gráficas da época e as ambiguidades próprias do gênero textual histórico. De acordo com Rosa (2015), é necessário que o pesquisador esteja atento à natureza do dado, distinguindo o que é variação linguística genuína do que é idiosincrasia do escriba ou imposição estilística. Para isso, recomenda-se, conforme as orientações de Lass (2000), que o pesquisador (i) seja nativo da língua-alvo, (ii) tenha conhecimento filológico das formas antigas da língua, e (iii) utilize, sempre que possível, pistas contextuais que permitam inferir os fatores sociais implicados na produção textual.

Montgomery (2007, apud Rosa, 2015) propõe uma abordagem multidimensional para a organização do corpus histórico, contemplando aspectos como: (a) a dimensão textual (proximidade da escrita com a oralidade), (b) a dimensão temporal (data de produção e recorte histórico), (c) a dimensão social (informações sobre o autor e sua posição na hierarquia social), (d) a dimensão espacial (local de produção e circulação do texto), e (e) a dimensão de representatividade (grau de abrangência dos dados em termos de grupo social e região). Esses critérios tornam-se imprescindíveis para assegurar a consistência e a validade analítica do corpus diacrônico.

2.3.1 Corpora diacrônicos e rigor metodológico

A constituição de corpora diacrônicos nos impõe o reconhecimento da materialidade dos dados e das limitações que os envolvem. Diferentemente dos corpora sincrônicos de fala, produzidos sob controle metodológico direto, os corpora históricos derivam de documentos preexistentes, cuja forma e conteúdo escapam ao controle do pesquisador. Como afirma Labov (1994, p. 11), “a pesquisa em Linguística Histórica exige fazer o melhor uso de dados ruins”. Assumimos essa máxima como um princípio orientador: mesmo que imperfeitos, os dados históricos oferecem testemunhos linguísticos valiosos, desde que tratados com rigor e responsabilidade metodológica.

Dessa forma, a constituição de um corpus histórico não pode prescindir de uma postura metodológica rigorosa, ancorada em critérios claros de seleção, categorização e interpretação. A organização dos dados deve contemplar, tanto quanto possível, a diversidade social da comunidade de fala representada, considerando fatores como classe social, gênero, faixa etária e grau de escolaridade. Do mesmo modo, a delimitação temporal do *corpus* deve estar alinhada aos objetivos analíticos da pesquisa, evitando anacronismos e sobreposições que possam comprometer a interpretação da variação.

Ao integrar os pressupostos da Sociolinguística Variacionista à análise de textos históricos, a Sociolinguística Histórica propõe um modelo metodológico que conjuga empiria e interpretação, observação estatística e reconstrução histórica. É nessa articulação que se insere a proposta deste trabalho: desenvolver uma metodologia de organização de corpora diacrônicos que não apenas atenda às exigências da pesquisa sociolinguística, mas também respeite as especificidades e contingências dos dados históricos, garantindo a validade e a solidez das inferências interpretativas produzidas.

3. METODOLOGIA DE ORGANIZAÇÃO DO CORPUS

A constituição de corpora diacrônicos no âmbito da Sociolinguística Histórica requer um percurso metodológico cuidadosamente delineado, que abarque desde os critérios de seleção documental até as estratégias de transcrição, categorização e codificação das ocorrências linguísticas. Diante da especificidade dos dados históricos, frequentemente condicionados por suportes escritos, acervos fragmentados e limitações contextuais, adotamos uma abordagem que alia rigor empírico e flexibilidade interpretativa, conforme recomendam Romaine (1982), Labov (1994, 2008), Rosa (2015) e Campoy e Conde Silvestre (2012).

3.1 SELEÇÃO E DELIMITAÇÃO DO CORPUS

A primeira etapa da organização de um corpus diacrônico consiste na seleção das fontes documentais. Dada a inexistência de registros orais sistemáticos anteriores ao século XX, a Sociolinguística Histórica se vale prioritariamente de textos escritos que, pela sua natureza discursiva, se aproximem da oralidade e revelem aspectos do uso efetivo da língua em contextos comunicativos reais. Entre os gêneros que melhor cumprem esse papel estão as cartas pessoais, os diários, as petições, os bilhetes e outros documentos manuscritos produzidos em situações não formais (Koch; Elias, 2014; Campoy; Conde Silvestre, 2012).

No presente modelo metodológico, optamos pelas cartas pessoais como objeto empírico preferencial. Tal escolha se justifica pelo caráter intersubjetivo desse gênero textual, que mobiliza estratégias linguísticas diretamente relacionadas às relações sociais entre os interlocutores. As cartas, especialmente aquelas trocadas entre sujeitos não institucionalizados, ou seja, não pertencentes ao aparato estatal ou eclesiástico, revelam traços de espontaneidade, estratégias de envolvimento, formas de tratamento e estruturas discursivas que se aproximam da oralidade, ainda que adaptadas à codificação escrita.

No delineamento do corpus, estabelecemos três eixos fundamentais para orientar os critérios de inclusão documental, os quais consideramos essenciais para assegurar a validade empírica e a consistência interpretativa da amostra. Esses eixos também podem servir como referência metodológica para outras investigações que se proponham a constituir corpora históricos a partir de documentos escritos. São eles:

- (i) a relevância linguística do material, considerando que os textos precisam conter ocorrências da variável em análise. Em nosso estudo, tratamos dos pronomes pessoais de segunda pessoa do singular em posição de sujeito, buscando garantir a pertinência dos dados em relação à hipótese investigativa;
- (ii) a delimitação geográfica e sociocultural da amostra, a partir da qual restringimos o corpus ao território do estado de Alagoas, buscando constituir uma comunidade de fala representativa do recorte regional definido;
- (iii) a marcação temporal confiável dos documentos, priorizando textos com datação explícita, mas também aceitando aqueles cuja cronologia pudesse ser inferida com base em elementos contextuais consistentes.

Documentos que não atendiam a um ou mais desses critérios, como os de autoria não identificada, sem marcação temporal verificável ou com forte grau de edição que comprometesse sua integridade linguística, foram excluídos da amostra. Essa seleção buscou garantir a representatividade, a autenticidade e a coerência metodológica necessárias à constituição de corpora diacrônicos em pesquisas de base empírica.

As fontes documentais foram diversas, buscando representar uma ampla gama de perfis sociais e situações comunicativas. Incluem-se no *corpus*:

- i) Cartas de Breno Accioly para Mário de Andrade, datadas entre 1942 e 1943, coletadas do acervo do IEB/USP. São cartas pessoais que apresentam relação de amizade e proximidade entre os missivistas, com um total de 12 cartas;
- ii) Cartas de missivistas de Coité do Nóia/AL, datadas entre 1970 e 2003. São cartas pessoais que apresentam proximidade entre os missivistas e, em sua maioria são cartas amorosas, coletadas através de doação dos proprietários, com um total de 10 cartas;
- iii) Cartas de missivistas não ilustres pertencentes ao acervo do Arquivo Público de Alagoas, datadas entre 1966 e 1997. São cartas que possuem maior e menor grau de proximidade entre os missivistas, com cartas que apresentam relação de amizade e trabalho, com um total de 67 cartas;
- iv) Cartas de Graciliano Ramos, datadas entre 1911 e 1930, pertencentes ao acervo do Arquivo Público de Alagoas. São cartas pessoais que apresentam relação de amizade com proximidade entre os missivistas, com um total de 14 cartas;
- v) Cartas de Graciliano Ramos, datadas entre 1910 e 1952, publicadas no livro “Cartas” da editora Record em 2013. São cartas pessoais que apresentam relações de amor e amizade com grau de proximidade entre os missivistas. Essas cartas foram disponibilizadas após 27 anos de seu falecimento, por sua esposa Heloísa Ramos, com um total de 101 cartas.

A representatividade do corpus foi construída a partir da combinação de registros manuscritos de diferentes décadas e regiões alagoanas, com variação de letramento, de subgêneros epistolares e de tipos de interlocução, assegurando a diversidade necessária à análise.

3.2 TRANSCRIÇÃO E PRESERVAÇÃO DA AUTENTICIDADE TEXTUAL

Constituído o conjunto documental, passamos à etapa de transcrição dos dados, buscando equilibrar dois objetivos: (i) garantir a fidelidade ao texto original, preservando traços relevantes para a análise linguística, como ortografia, pontuação, segmentação, erros de grafia e marcas de oralidade; e (ii) possibilitar a normalização mínima que permita a codificação estatística posterior. Para isso, utilizamos o uso de transcrição sem diplomática, conforme os parâmetros definidos pelo Projeto Para a História do Português Brasileiro (PHPB), os quais oferecem orientações consolidadas para a edição de documentos manuscritos e impressos em pesquisas diacrônicas.

Além disso, elaboramos um sistema de identificação padronizado dos textos, com numeração individual e registro sistemático das informações contextuais contendo: autor, data, local, relação entre interlocutores, armazenadas em planilhas digitais. Isso nos permitiu o cruzamento entre dados linguísticos e extralinguísticos, conforme os seguintes critérios:

- i) Leitura crítica do manuscrito, com identificação de elementos como data, local e interlocutores;
- ii) Transcrição literal com intervenções mínimas;
- iii) Registro paralelo de dados metalinguísticos como uso de abreviaturas, rasuras, apagamentos e interferências gráficas.

Esses procedimentos garantiram a preservação da autenticidade textual, sem prejuízo da análise linguística, respeitando a historicidade e a singularidade discursiva de cada carta.

3.3 CATEGORIZAÇÃO E ANOTAÇÃO LINGUÍSTICA

A partir do conjunto de cartas selecionado e transcrito, realizamos a identificação e categorização das ocorrências de pronomes pessoais de segunda pessoa do singular em posição de sujeito. Essa delimitação constituiu a variável dependente do estudo, cujo comportamento foi analisado em correlação com múltiplas variáveis independentes, de natureza linguística e social. A seguir, são descritas as variáveis controladas:

- (i) Período de escrita da carta: organizado em intervalos decenais (, com base na datação explícita do documento;

- (ii) Subgênero epistolar: categorização temática da carta como amorosa, familiar, de amizade ou de trabalho, com base na relação entre os interlocutores e no conteúdo predominante;
- (iii) Sexo do missivista e relação entre os sexos dos interlocutores: variáveis sociais associadas ao impacto da identidade de gênero nas formas de tratamento;
- (iv) Preenchimento do sujeito: distinção entre sujeitos pronominais explícitos e sujeitos elípticos, relevantes para o índice de visibilidade do pronome;
- (v) Concordância verbal: análise da concordância verbal, elemento fundamental para a investigação do uso das formas pronominais;
- (vi) Paralelismo pronominal: observação da alternância ou regularidade de uso dos pronomes dentro de uma mesma carta;
- (vii) Relação de poder e solidariedade: variável interpretativa baseada na Teoria de Brown e Gilman (1960), inferida a partir do conteúdo da mensagem, dos vocativos e da construção dos papéis sociais.

Cada ocorrência foi anotada em planilha com campos codificados para posterior análise estatística. A categorização foi realizada manualmente, por meio de leitura interpretativa de cada carta, garantindo a qualidade do mapeamento linguístico.

3.4 ANÁLISE ESTATÍSTICA

Para o tratamento estatístico dos dados, é fundamental que o pesquisador disponha as ocorrências linguísticas em um banco de dados estruturado, com codificação clara das variáveis dependente e independentes. Em nossa pesquisa, organizamos as 546 ocorrências em planilha digital, com todas as variáveis previamente categorizadas, o que viabilizou a aplicação de testes estatísticos de forma sistemática e interpretativamente controlada.

Adotamos o software R (R CORE TEAM, 2018) como ferramenta para a análise estatística, por meio da aplicação de modelos de regressão logística, adequados à investigação de variáveis categóricas. Essa modelagem permite estimar a influência relativa de fatores linguísticos, como a presença ou ausência do sujeito e a forma de concordância verbal, bem como de fatores extralinguísticos, como o sexo do missivista, o tipo de vínculo entre os interlocutores, o subgênero epistolar e o período de produção das cartas.

Ressaltamos que o tratamento estatístico não deve se restringir à descrição de frequências, mas deve ser concebido como etapa interpretativa do processo analítico. A análise quantitativa, quando aliada a critérios interpretativos bem definidos, contribui para a formulação de hipóteses sobre regularidades de uso e possíveis trajetórias de mudança linguística. No caso da variação pronominal entre *tu* e *voce*, os resultados nos permitiram compreender com maior precisão os condicionamentos sociais e estruturais envolvidos

em sua distribuição ao longo do século XX, no contexto específico das cartas pessoais alagoanas.

4. DESAFIOS E POSSIBILIDADES NA ORGANIZAÇÃO DE CORPORA DIACRÔNICOS

A constituição de corpora diacrônicos em pesquisas de base empírica, especialmente no âmbito da Sociolinguística Histórica, demanda não apenas decisões de ordem técnica, mas também um posicionamento epistemológico que fundamente as escolhas metodológicas adotadas. Ao longo da organização do corpus, enfrentamos desafios que nos parecem comuns a estudos dessa natureza, e cuja antecipação pode contribuir para orientar outras investigações semelhantes.

Um dos principais obstáculos refere-se à disponibilidade e ao estado de conservação das cartas pessoais. Observamos que muitos arquivos, públicos ou particulares, não oferecem condições adequadas para o armazenamento e a organização sistemática desses documentos, o que dificulta o acesso e compromete, em certos casos, a integridade dos dados. Como apontado por Rumeu (2012), a documentação epistolar ainda carece de políticas de preservação e valorização que reconheçam sua importância como fonte legítima de análise linguística e histórica.

No que se refere à materialidade dos documentos, a leitura revelou desafios decorrentes da caligrafia individual, da instabilidade gráfica e da ortografia não padronizada. A transcrição das cartas exigiu não apenas cuidado e atenção, mas também decisões metodológicas quanto ao grau de intervenção no texto original. Como sugere Mattos e Silva (2004), a transcrição em pesquisas linguísticas deve preservar o valor linguístico do dado, evitando normalizações que apaguem traços relevantes para a análise da variação. Por isso, optamos por uma transcrição semi-diplomática, preservando elementos como desvios ortográficos, segmentações não convencionais e marcas discursivas da oralidade escrita.

Outro ponto que exigiu atenção metodológica foi a atribuição das variáveis extralinguísticas. Informações como o sexo do missivista, o tipo de relação entre os interlocutores, o grau de proximidade e o contexto sociocomunicativo foram inferidas com base em pistas textuais e contextuais presentes nas cartas, como vocativos, pronomes de tratamento, estruturas de saudação e conteúdo temático. Esse tipo de inferência demanda uma leitura atenta às marcas sociais do discurso e ao funcionamento relacional da linguagem nas interações escritas. Essa etapa envolveu o exercício de uma interpretação histórica e discursiva, conforme propõem Zilles e Faraco (2008), para quem a análise linguística de documentos históricos requer a reconstrução do universo social em que se inserem os sujeitos epistolares.

A categorização da variável “relação de poder e solidariedade” foi orientada pela proposta de Brown e Gilman (1960), que relaciona o uso pronominal a eixos de autoridade e afetividade. No entanto, reconhecemos a necessidade de adaptar esse referencial à análise de dados históricos. Como defende Martins (2016), o pesquisador deve considerar as condições pragmáticas e sociais específicas de produção dos enunciados, evitando a imposição de modelos analíticos anacrônicos. Em nosso estudo, procuramos interpretar

o uso dos pronomes *tu* e *você* à luz das informações disponíveis em cada documento, buscando compreender como as escolhas linguísticas se articulam às dinâmicas sociais da época.

Outro desafio importante foi lidar com a heterogeneidade tipológica das cartas, tanto em relação ao subgênero epistolar (amoroso, familiar, profissional) quanto às estratégias discursivas empregadas. Em muitos casos, observamos a coexistência de registros com oscilação entre estilos formais e informais dentro de uma mesma carta, fenômeno que, conforme destaca Marcuschi (2001), evidencia a instabilidade genérica dos textos cotidianos e desafia classificações rígidas. Nesse sentido, a categorização das cartas foi feita de forma interpretativa, considerando o foco temático e o tipo de vínculo predominante, sem desconsiderar a fluidez inerente ao gênero.

Apesar das dificuldades enfrentadas, ressaltamos que a organização do corpus possibilitou o acesso a formas de uso da língua pouco visíveis em fontes normatizadas, revelando práticas linguísticas mais próximas do cotidiano. A espontaneidade relativa das cartas pessoais nos permitiu observar, com maior nitidez, as estratégias linguísticas envolvidas nas relações interpessoais e sua relação com variáveis sociais e discursivas.

Além disso, a estruturação desse *corpus* nos permitiu observar como a organização sistemática de dados históricos pode contribuir para o refinamento de hipóteses sobre mudança linguística, ao identificar padrões recorrentes e descontinuidades no uso dos pronomes de segunda pessoa ao longo do tempo. Como aponta Paiva (2013), o *corpus* histórico, quando criteriosamente construído, não apenas documenta o passado da língua, mas também se constitui como lugar de produção de conhecimento linguístico-teórico, pois desafia modelos explicativos ancorados em dados contemporâneos e permite testar hipóteses sobre a interação entre língua, tempo e sociedade.

CONSIDERAÇÕES FINAIS

A constituição de corpora diacrônicos representa um eixo metodológico central para os estudos empíricos em Sociolinguística Histórica, na medida em que possibilita a observação de fenômenos linguísticos em perspectiva temporal e social. Ao longo deste artigo, buscamos apresentar e discutir os procedimentos adotados na organização de um corpus epistolar alagoano, composto por cartas pessoais manuscritas produzidas entre as décadas de 1910 e 2000. Detalhamos os critérios estabelecidos para a seleção dos documentos, os procedimentos de transcrição e anotação, a categorização das variáveis e a análise estatística dos dados.

A estruturação desse corpus nos permitiu observar, com base empírica, o comportamento variável dos pronomes de segunda pessoa *tu* e *você* em contexto regional, lançando luz sobre as formas de tratamento em circulação no português alagoano ao longo do século XX. Para além dos resultados específicos, a experiência metodológica aqui relatada contribui para o debate mais amplo sobre os limites e possibilidades do uso de fontes manuscritas em pesquisas sociolinguísticas de base histórica.

Adotamos uma abordagem metodológica fundamentada nos princípios da Teoria da Variação e Mudança Linguística, da Sociolinguística Histórica e das práticas documentais voltadas à análise de textos antigos. Essa combinação revelou-se eficaz para

garantir a consistência interna do corpus, a coerência entre os dados e os objetivos analíticos e a possibilidade de interpretações fundamentadas nas relações entre linguagem, tempo e sociedade.

Entre os principais desafios enfrentados, destacamos a escassez de acervos organizados, as limitações na identificação de variáveis contextuais e a complexidade inerente à análise de documentos produzidos fora da esfera institucional. Essas dificuldades reforçam a necessidade de que o pesquisador adote uma metodologia sensível às características dos dados, capaz de equilibrar critérios técnicos com flexibilidade interpretativa. Ressaltamos, nesse ponto, a importância de decisões fundamentadas em princípios analíticos claros e em uma postura crítica frente às contingências que marcam a documentação histórica.

Como contribuição para o campo, consideramos que este estudo oferece um modelo possível de constituição e tratamento de corpus diacrônico, voltado à análise sociolinguística da variação pronominal em registros privados. Ao sistematizar uma amostra representativa de cartas pessoais de uma região ainda pouco explorada nos estudos de variação, buscamos colaborar com a ampliação dos referenciais metodológicos disponíveis para a área, incentivando a valorização da escrita cotidiana como fonte legítima de investigação linguística.

Esperamos que futuras pesquisas possam expandir esse modelo, incorporando materiais de outras regiões e períodos históricos, bem como aprofundando os cruzamentos entre análise linguística e interpretação sociocultural. A criação de bancos de dados compartilhados, com transcrições padronizadas e metadados acessíveis, configura-se como um caminho promissor para o fortalecimento da Sociolinguística Histórica como campo analítico atento à pluralidade dos usos da língua e à complexidade dos contextos de produção textual.

REFERÊNCIAS

BROWN, Roger; GILMAN, Albert. The pronouns of power and solidarity. In: SEBEOK, Thomas A. (Ed.). *Style in Language*. Cambridge, MA: MIT Press, 1960. p. 253–276.

CAMPOY, María Carmen; CONDE SILVESTRE, Juan José. *Historical sociolinguistics: Sociohistorical linguistics in Spanish context*. Bern: Peter Lang, 2012.

CARDOSO, Suzana Alice Marcelino. *Variação e mudança linguística: o papel das cartas pessoais*. Salvador: EDUFBA, 2010.

COELHO, Maria do Socorro Silva; SILVA, Roberta Pires de Oliveira e; FARACO, Carlos Alberto. *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2010.

FARACO, Carlos Alberto. *História sociopolítica da língua portuguesa*. São Paulo: Parábola Editorial, 2019.

KOCH, Ingedore Villaça; ELIAS, Vanda Maria. *Linguagem e discurso: modos de organização e gêneros textuais*. São Paulo: Cortez, 2014.

LABOV, William. *Principles of linguistic change: Cognitive and cultural factors*. Volume 3. Oxford: Wiley-Blackwell, 2008.

LABOV, William. *Principles of linguistic change: Internal factors*. Volume 1. Oxford: Blackwell, 1994.

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

MARCUSCHI, Luiz Antônio. *Gêneros textuais: definição e funcionalidade*. In: _____. *Gêneros textuais e ensino*. São Paulo: Parábola, 2001. p. 19–36.

MARTINS, Eliane Gouvêa et al. *Documentos e escrita: problemas de edição e análise*. São Paulo: Humanitas, 2015.

MARTINS, Eliane Gouvêa. *Análise do discurso e escrita: reflexões sobre documentos e práticas de leitura*. São Paulo: Humanitas, 2016.

MATTOS E SILVA, Rosa Virgínia. *Para a história do português brasileiro: metodologia e dados*. São Paulo: Humanitas, 2004.

MONTGOMERY, Michael. *Historical dialectology and linguistic diffusion*. In: HERNÁNDEZ-CAMPOY, Juan Manuel; CONDE-SILVESTRE, Juan José (Eds.). *The handbook of historical sociolinguistics*. Oxford: Wiley-Blackwell, 2012. p. 446–461. (cit. apud ROSA, 2015)

PAIVA, Maria da Conceição de. *História das ideias linguísticas: questões e perspectivas teóricas*. Recife: EDUPE, 2013.

PAIVA, Maria da Conceição de; DUARTE, Maria Eugênia Lammoglia. *Varição linguística: análise e ensino*. Rio de Janeiro: Lucerna, 2006.

R CORE TEAM. R: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2018. Disponível em: <https://www.R-project.org/>.

ROMAINE, Suzanne. *Historical sociolinguistics*. Cambridge: Cambridge University Press, 1982.

ROMAINE, Suzanne. *Variation in language: Variationist sociolinguistics*. Oxford: Oxford University Press, 2009.

ROSA, Maria Célia Lima. *Sociolinguística histórica: fundamentos, métodos e perspectivas*. São Paulo: Parábola, 2015.

RUMEU, Margarida. *As cartas privadas como documentos linguísticos: contribuições da filologia para a sociolinguística histórica*. In: FARACO, Carlos Alberto; ZILLES, Ana Maria; MENON, Elizeu (Orgs.). *Estudos de Sociolinguística Histórica: novas perspectivas sobre o português do Brasil*. Campinas: Mercado de Letras, 2012. p. 171–192.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. *Empirical foundations for a theory of language change*. In: WINFRED, Lehmann; MALONE, Joseph C. (Eds.). *Directions for historical linguistics: A symposium*. Austin: University of Texas Press, 2006 [1968]. p. 95–188.

ZILLES, Ana Maria; FARIA, Isis Chamma. *A escrita de não letrados no Brasil: considerações sobre edição de documentos históricos*. In: FARACO, Carlos Alberto; ZILLES, Ana Maria; MENON, Elizeu (Orgs.). *Estudos de Sociolinguística Histórica: novas perspectivas sobre o português do Brasil*. Campinas: Mercado de Letras, 2018. p. 69–94.

Submetido em 20 de maio 2025

Aceito para publicação em 28 de dezembro de 2025