laborhistórico

Revista LaborHistórico
ISSN 2359-6910
v.12, n.1, e70066, 2026
DOI: 10.24206/lh.v12i1.70066

Artigo Dossiê

Do arquivo ao dado: considerações sobre a viabilidade dos jornais na pesquisa sociolinguística histórica

From Archive to Data: Considerations about the Viability of Newspapers in Historical Sociolinguistic Research

Arthur Ronald Brasil Terto[1]

arthur.terto@fale.ufal.br

Josefa Arruda Silva Neta[2]

josefaarruda35@gmail.com

Elyne Giselle de Santana Lima Aguiar Vitório[3]

elyne.vitorio@gmail.com

[1, 2 e 3] Universidade Federal de Alagoas

Editores-chefes
Marcus Dores
Célia Lopes

Editoras convidadas
Zenaide Carneiro
Mariana Lacerda
Alícia Lose

Dossier:
Metodologia de Organização de Corpora Diacrônicos para Pesquisas em Sociolinguística Histórica

Recebido: 26/10/2025
Aceito: 20/03/2026

Como citar:
TERTO, A. R. B.; SILVA NETA, J. A.; VITÓRIO, E. G. de S. L. Do arquivo ao dado: considerações sobre a viabilidade dos jornais na pesquisa sociolinguística histórica. *Revista LaborHistórico*, v.12, n.1, e70066, 2026. doi: https://doi.org/10.24206/lh.v12i1.70066

Resumo

Neste artigo, discutimos a viabilidade dos jornais históricos como fonte de dados primários para a pesquisa em Sociolinguística Histórica, campo interdisciplinar que investiga a relação entre fatores sociais e históricos e processos de variação e de mudança linguística em sincronias passadas. Partindo do pressuposto teórico de que a língua é um sistema heterogêneo e inerentemente variável, o estudo tem como objetivo central analisar as potencialidades e os desafios inerentes à utilização desse tipo de documento, que, embora mediado por instâncias editoriais, registra usos linguísticos menos monitorados e reveladores de práticas sociais. A fundamentação

This content is licensed under a Creative Commons Attribution 4.0 International License.

teórica ancora-se nos contributos da Sociolinguística Variacionista e da Linguística Histórica, articulando-os à reflexão sobre os gêneros textuais. Metodologicamente, o estudo baseia-se em uma revisão bibliográfica de pesquisas anteriores que utilizaram jornais como *corpus*, destacando a necessidade de critérios rigorosos para classificar grafias significativas (Lass, 2000) e para lidar com o *bad data problem*. Resultados de pesquisas feitas nesse campo do conhecimento demonstram que os jornais permitem observar a circulação de ideologias linguísticas, processos de normatização, emergência de variedades urbanas e fenômenos de mudança em curso, funcionando como um espaço híbrido entre a norma e o uso. Conclui-se que, apesar dos desafios, os jornais são uma fonte significativa para a reconstrução de ecologias linguísticas passadas. Para pesquisas futuras, propõe-se a constituição de *corpora* mais amplos e diversificados, a integração de ferramentas de Processamento de Linguagem Natural (PLN) para análise em larga escala e a adoção de abordagens interdisciplinares que articulem linguística, história cultural e estudos sociais.

Palavras-chave

Sociolinguística histórica. Jornais. Variação linguística. Mudança linguística. *Corpus*.

Abstract

In this article, we discuss the feasibility of historical newspapers as primary data sources for research in Historical Sociolinguistics, an interdisciplinary field that investigates the relationship between social and historical factors and processes of linguistic variation and change in past synchronic stages. Based on the theoretical assumption that language is a heterogeneous and inherently variable system, the central objective of this study is to analyze the potentialities and challenges inherent in the use of this type of document, which, although mediated by editorial instances, records less monitored linguistic uses and reveals social practices. The theoretical framework is grounded in the contributions of Variationist Sociolinguistics and Historical Linguistics, articulated with reflections on textual genres. Methodologically, the study relies on a literature review of previous research that used newspapers as corpus data, highlighting the need for rigorous criteria to classify significant spellings (Lass, 2000) and to address the bad data problem. Results from studies in this field demonstrate that newspapers make it possible to observe the circulation of linguistic ideologies, processes of standardization, the emergence of urban varieties, and ongoing processes of linguistic change, functioning as a hybrid space between norm and use. It is concluded that, despite the challenges, newspapers are a significant source for the reconstruction of past linguistic ecologies. For future research, we propose the construction of a broader and more diversified corpora, the integration of Natural Language Processing (NLP) tools for large-scale analysis, and the adoption

of interdisciplinary approaches that bring together linguistics, cultural history, and social studies.

Keywords

Historical sociolinguistics. Newspapers. Linguistic variation. Linguistic change. *Corpus*.

Introdução

A Sociolinguística Histórica consolida-se como campo interdisciplinar ao articular construtos teórico-metodológicos de diferentes ciências sociais - como a Sociologia, a História Social, a Psicologia Social, a Sociolinguística, a Filologia, entre outras - para compreender processos de variação e mudança linguística em sincronias passadas. Herdando sua identidade epistemológica, principalmente, de contribuições da Sociolinguística Variacionista e da Linguística Histórica, essa disciplina investiga de que maneira fatores sociais e históricos influenciaram fenômenos variáveis na língua ao longo do tempo. Um desafio central, no entanto, se coloca para esse campo: como investigar a influência de fatores sociais em períodos históricos da língua quando seus falantes não estão mais à disposição do pesquisador?

Com o objetivo de superar esse impasse, a Sociolinguística Histórica tem recorrido a registros escritos, de modo a inferir, através de pistas textuais, históricas e sociais, a realidade linguística de uma estrutura social passada. Essa alternativa, no entanto, implica outro desafio: como garantir que os dados coletados a partir dos registros escritos sejam suficientemente representativos da linguagem em uso por todos os estratos sociais em um dado momento histórico? Grande parte da documentação histórica disponível reflete, em geral, práticas de escrita monitoradas, distantes da oralidade e caracterizadas pela padronização linguística própria da modalidade escrita da língua.

Os jornais históricos se configuram como uma alternativa promissora de fonte de dados linguísticos históricos para a pesquisa em Sociolinguística Histórica, pois, ao mesmo tempo em que fazem uso de formas linguísticas menos monitoradas (sobretudo em seções como cartas de leitores, colunas de opinião, anúncios e notas policiais), também impõem camadas de mediação institucional, editorial e ideológica que precisam ser consideradas na análise linguístico-sócio-histórica. Além disso, a partir dos jornais históricos, também é possível identificar práticas linguísticas direcionadas a grupos sociais específicos (a depender de qual era o público leitor), permitindo análises que entrelaçam linguagem, ideologia e estratificação social. Embora o desafio de medir graus de formalidade e de representatividade do vernáculo se imponha não apenas a textos da esfera jornalística, mas a qualquer fonte escrita na pesquisa em Sociolinguística Histórica, o uso de jornais exige mormente

critérios rigorosos de seleção, recorte e interpretação. Isso porque, diferentemente de suportes textuais mais homogêneos, estes documentos reúnem, em um mesmo espaço, uma pluralidade de gêneros textuais com diferenças significativas entre si quanto ao distanciamento da fala espontânea.

Diante dessas questões, o objetivo deste artigo é discutir as potencialidades e os desafios do uso de jornais históricos como fonte de dados empíricos para as pesquisas em Sociolinguística Histórica. Partindo dos pressupostos teórico-metodológicos gerais da Sociolinguística Histórica, apontamos vantagens e desvantagens vinculadas ao uso desses materiais e demonstramos, através de experiências de pesquisas anteriores e de lacunas ainda existentes na literatura, como podem revelar padrões de variação linguística em diferentes períodos históricos.

1. Sociolinguística Histórica: Definição e Objeto de Estudo

A Sociolinguística moderna surgiu na década de 1960 como um contraponto às duas abordagens teórico-metodológicas predominantes na Linguística da primeira metade do século XX: o estruturalismo e o gerativismo. Tanto o estruturalismo quanto o gerativismo compreendem a língua como um sistema homogêneo, fechado e, portanto, desassociado de seus falantes reais. De encontro a essa concepção, a Sociolinguística, naquele momento, propunha um reposicionamento dos estudos linguísticos: propunha-se agora uma concepção de língua enquanto sistema inerentemente complexo, dinâmico e heterogêneo, não desvinculado dos falantes e dos usos reais que estes fazem dele sincrônica e diacronicamente. Em outras palavras, objetivava-se uma investigação empírica dos fatos linguísticos e uma análise linguística feita à luz de fatores sociais cuja atuação tinha implicação direta nas formas da língua. Portanto, a disciplina rompeu com a ideia de língua e falante ideais e reconheceu a heterogeneidade como parte constitutiva dos sistemas linguísticos.

Ao longo da segunda metade do século XX, diversas vertentes teórico-metodológicas constituíram-se dentro da Sociolinguística, cada uma adotando enfoques epistemológicos e objetivos de investigação específicos. Uma vertente clássica é aquela inaugurada por William Labov (1966), a Sociolinguística Variacionista, ou Teoria da Variação e Mudança. Em seus estudos na década de 1960, considerados seminais dentro da Sociolinguística, Labov demonstrou, em pesquisas realizadas em comunidades urbanas, que o processo de variação e mudança linguística é associado a variações sociais sistemáticas relacionadas à estratificação da sociedade a partir de diversos fatores, como classe social, idade, sexo e estilo de fala (LABOV, 1966, 1972a, 1972b, entre outros). Cabe destacar, contudo, que, para além da vertente variacionista, há outras igualmente relevantes na consolidação da Sociolinguística como ciência autônoma, sendo a Sociolinguística Histórica uma delas.

Partindo do pressuposto comum a todas as vertentes da Sociolinguística, (isto é, o pressuposto de que as línguas são sistemas inerentemente complexos, dinâmicos e heterogêneos), a Sociolinguística Histórica objetiva descrever e explicar processos de variação e de mudança linguística que ocorreram no passado, ou seja, em sincronias remotas (Romaine, 2009[1982]). Diferentemente da Sociolinguística Variacionista, porém, que utiliza amostras de língua falada para investigar fenômenos de variação na língua, a Sociolinguística Histórica ancora suas análises em amostras de língua escrita, visto que, por analisar estágios remotos da língua, a única fonte de dados disponível ao pesquisador são os textos escritos (Nevalainen, 2015). Diz-se que essa vertente teórico-metodológica surgiu, em um primeiro momento, a partir da proposta de Suzanne Romaine (2009[1982]) de unir procedimentos metodológicos da Sociolinguística Variacionista com procedimentos da Linguística Histórica, de modo a investigar fenômenos sociolinguísticos do passado. No entanto, ao longo do tempo, a Sociolinguística Histórica, conforme Conde Silvestre (2007) e Lopes (2020), delineou seus próprios pressupostos, objeto de estudo e métodos, consolidando-se como disciplina científica autônoma.

Assim como a Sociolinguística Variacionista, a Sociolinguística Histórica (i) concebe a variação linguística como elemento constitutivo e não periférico do sistema linguístico; (ii) reconhece o papel dos fatores sociais nos diversos usos que os falantes fazem da língua; (iii) valoriza dados empíricos como material legítimo para a análise linguística; e (iv) critica as abordagens linguísticas tradicionais que desconsideram a heterogeneidade inerente às línguas humanas. Podemos dizer, por conseguinte, que esses são os pontos convergentes entre essas duas abordagens sociolinguísticas. No entanto, embora essas abordagens compartilhem o princípio de que a variação linguística não é aleatória, mas influenciada por fatores sociais, elas também têm características peculiares que as distanciam (especialmente quanto aos seus objetos de estudo, aos seus recortes temporais, às suas metodologias e aos tipos de dados que analisam).

No que respeita ao objeto de estudo, por exemplo, enquanto a Sociolinguística Variacionista busca compreender de que maneira fatores sociais e linguísticos condicionam a variação linguística contemporânea para prever potenciais mudanças, a Sociolinguística Histórica busca reconstruir trajetórias de mudança linguística a partir da análise de variação linguística no passado para explicar como certos traços se desenvolveram ou desapareceram ao longo do tempo na língua. No que se refere aos seus procedimentos metodológicos, enquanto a Sociolinguística Variacionista utiliza, majoritariamente, métodos estatísticos e quantitativos para correlacionar variáveis sociais e linguísticas, a Sociolinguística Histórica pode utilizar métodos qualitativos, filológicos, paleográficos e também quantitativos, mas com ênfase na interpretação contextual (social e histórica) dos dados coletados. Por fim, no que diz respeito ao tipo de dados que analisam, a Sociolinguística Variacionista tem

suas pesquisas realizadas em comunidades de fala contemporâneas ao pesquisador, o que garante uma coleta direta dos dados de fala; já a Sociolinguística Histórica tem suas pesquisas empreendidas em comunidades do passado, acessíveis somente, portanto, indiretamente por meio de registros escritos escassos. Vê-se, assim, que a Sociolinguística Histórica tem uma identidade epistemológica própria.

Na seção seguinte, apresentamos brevemente os principais procedimentos e desafios metodológicos enfrentados por aqueles que decidem investigar fenômenos de variação e mudança linguística a partir do enfoque social e histórico da língua.

2. Métodos e Desafios na Investigação da Sociolinguística Histórica

A Sociolinguística Histórica tem se consolidado como uma disciplina científica de orientação empírica voltada para o estudo da variação e da mudança linguística em sincronias passadas, sempre em articulação com fatores sociais e históricos que influenciaram os usos linguísticos. Assim como na Sociolinguística Variacionista, os sociolinguistas históricos combinam abordagens quantitativas e qualitativas em suas análises, valendo-se de diferentes tipos de fontes escritas a fim de compreender como os usos linguísticos eram moldados pela estrutura social em períodos anteriores. Embora os métodos quantitativos sejam predominantes - com uso crescente de modelos estatísticos nas análises sócio-históricas - , a análise qualitativa é igualmente (talvez até mais) relevante na Sociolinguística Histórica. A abordagem qualitativa, na Sociolinguística Histórica, assume um caráter filológico, exigindo do sociolinguista histórico uma leitura atenta dos textos e a reconstrução detalhada de contextos culturais e históricos (Nevalainen, 2015).

Um desafio que se impõe aos sociolinguistas históricos é a questão da relação entre língua escrita e língua falada. Como eles não dispõem de registros diretos da oralidade em tempos passados, a reconstrução dos usos orais da língua a partir de textos escritos torna-se uma tarefa consideravelmente complexa, exigindo inferências fundamentadas nos traços linguísticos de diferentes níveis gramaticais presentes em textos que preservam características do discurso falado. Romaine (2009[1982]), entretanto, assegura que é possível investigar e estudar a variação através da língua escrita. A fundadora da Sociolinguística Histórica ressalta, porém, ser necessário que o sociolinguista histórico saiba diferenciar o que pode ser considerado, por exemplo, uma representação de variação puramente gráfica daquilo que pode ser uma representação de variação fonético-fonológica dos segmentos da língua. Esse mesmo princípio hermenêutico aplica-se a outros níveis de análise: no plano morfossintático, é preciso distinguir entre variação genuinamente gramatical e meras convenções de escrita ou arcaísmos gráficos; no âmbito lexical, importa discernir entre escolhas estilísticas do redator e indícios de mudança no vernáculo da época.

Nesse sentido, Lass (2000), no plano fonético-fonológico, por exemplo, propõe uma classificação para definir se o dado coletado do registro escrito é significativo ou não para o estudo da variação e da mudança linguísticas:

- **Lixo** (Garbage): grafias que provavelmente resultam de deslizes na escrita ou de erros cometidos pelo escrevente. Um exemplo seria a escrita *xapéu* em vez de *chapéu*.
- **Variação puramente gráfica**: grafias que refletem certas convenções ortográficas da época ou tentativas de recuperar aspectos etimológicos da palavra. Um exemplo seria a escrita de consoantes geminadas, como em *lettra* em vez de *letra*.
- **Grafias fonologicamente significativas**: grafias que refletem traços de pronúncia, por exemplo, isto é, que representam variação fonético-fonológica. Um exemplo seria a escrita de *buneca* em vez de *boneca*.

Desde o surgimento da Linguística de Corpus na década de 1980, diversos *corpora* de textos escritos têm sido compilados com o intuito de servir como fonte de dados para os estudos histórico-sociolinguísticos. Um marco importante nesse cenário foi a construção do *Corpus of Early English Correspondence* (Nevalainen; Raumolin-Brunberg, 2003), considerado o primeiro *corpus* histórico elaborado com informações sociais e históricas detalhadas sobre os autores dos textos escritos. Contudo, a despeito dos avanços em direção à construção e à disponibilização gratuita de *corpora* de textos escritos, um outro desafio que se impõe aos sociolinguistas históricos é a disponibilidade limitada de dados que representem adequadamente a estrutura social e, portanto, a ecologia sociolinguística do passado.

Esse desafio, conhecido *bad data problem*, refere-se ao fato de que grande parte dos registros escritos disponíveis nos *corpora* de textos históricos provém de setores letrados da sociedade, o que resulta em um viés educacional, de classe e de gênero. Quer dizer, como a alfabetização historicamente foi privilégio das elites, mulheres e classes populares permanecem sub-representadas nos arquivos textuais convencionais. Essa limitação dificulta a compreensão da distribuição social dos traços linguísticos.

Nesse contexto, uma alternativa proposta para garantir a validade ecológica dos dados linguísticos representativos de sincronias passadas é a adoção de textos pessoais escritos por figuras não ilustres, visto que permitem aos sociolinguistas históricos explicar o papel dos setores populares na mudança e na padronização linguística. Cartas pessoais, diários, testemunhos de tribunais são exemplos de materiais que possibilitam reconstruir a língua dos setores subalternos da sociedade. Por refletirem uma perspectiva de primeira pessoa e por serem gêneros textual-discursivos menos monitorados (no sentido de serem registros de atos de fala reais, capturados em

contextos de interação dialógica), esses documentos são considerados registros escritos particularmente próximos à oralidade (Sevic, 1999; Van Der Wal; Rutten, 2013). Além desses, gêneros como peças teatrais e sermões também têm sido explorados como meios de acesso às formas vernaculares em contextos históricos (Culpeper; Kitö, 2010).

Dentre os textos compilados nos *corpora* de textos históricos, os textos jornalísticos têm se destacado como fontes relevantes para a investigação histórico-sociolinguística, sobretudo a partir do século XIX, com a popularização da imprensa no Brasil. Por refletirem o uso linguístico em contextos públicos e semiformais, esses textos permitem ao pesquisador observar processos de normatização, emergência de variedades urbanas, circulação de ideologias linguísticas e transformações nos registros escritos. Além disso, jornais regionais, de cunho popular ou político, podem revelar marcas do contato linguístico, da variação regional e das atitudes sociais em relação às formas linguísticas, funcionando como "testemunhos" indiretos de usos linguísticos que não eram registrados *a priori* em textos pertencentes a outras esferas textual-discursivas (Percy, 2012).

Na seção seguinte, apresentamos algumas pesquisas que utilizaram os jornais como fonte de dados primários em investigações da Sociolinguística Histórica e contribuições para a realização de estudos utilizando esse dado.

3. Jornais como Fonte de Dados em Sociolinguística Histórica

Embora seja possível iniciar investigações sobre variação e mudança linguística de longo prazo a partir de textos oriundos da língua falada, esse tipo de abordagem inevitavelmente requer a ampliação do recorte temporal para abranger períodos anteriores da história da língua – momentos nos quais o acesso a registros orais é inexistente. Diante dessa limitação, supracitada nas seções anteriores, os estudos de natureza diacrônica passam a se concentrar, necessariamente, na análise de fenômenos linguísticos tal como registrados em textos escritos.

Apesar de dialogarmos com a tradição que reconhece o *corpus* como um recurso fundamental para descrições linguísticas de base empírica e para o desenvolvimento de reflexões teóricas ancoradas em usos reais da língua (Barbosa; Marine, 2007), ampliamos essa perspectiva ao defender que, tanto em estudos sincrônicos quanto em investigações diacrônicas, a centralidade não está, necessariamente, na modalidade (oral ou escrita) adotada. O aspecto determinante reside no tipo e no gênero textual selecionados para a análise linguística em questão. É preciso compreender as especificidades do gênero e do tipo textual em foco, uma vez que tais elementos trazem consigo marcas estruturais próprias, temas mais frequentes, níveis distintos de formalidade, além de intenções comunicativas variadas. O domínio dessas

características por parte do pesquisador conduz a uma abordagem metodológica mais adequada no trato com o *corpus*, seja ele de natureza oral ou escrita, permitindo, assim, uma análise mais precisa e coerente com os objetivos da pesquisa linguística.

Por isso, conforme Barbosa e Marine (2007), é recomendável estimular a aplicação do modelo teórico-metodológico laboviano na análise de textos escritos. Isso porque, ao se identificarem fenômenos linguísticos em variação nesse tipo de material, há um forte indicativo de que tais processos encontram-se em um estágio avançado, possivelmente já configurando uma mudança efetiva na língua. Como destaca Faraco, as transformações linguísticas

> costumam se desencadear na fala informal de grupos socioeconômicos intermediários; avançam pela fala informal de grupos mais altos na estrutura socioeconômica; chegam a situações formais de fala e só então começam a ocorrer na escrita (Faraco, 2005, p. 26).

Tendo em vista a busca por compreender aspectos linguísticos de sincronias passadas do português brasileiro, especialmente em contextos em que não há registros orais disponíveis, a linguagem da imprensa - mais especificamente os textos jornalísticos - se apresenta como uma fonte para estudos sociolinguístico-históricos. Embora os jornais sigam convenções editoriais e apresentem certo grau de padronização, eles também registram usos reais da língua, com traços da variação linguística que circula social e regionalmente em determinados períodos históricos (Percy, 2012).

Além de refletirem tendências normativas, os textos jornalísticos carregam indícios de usos linguísticos em trânsito, muitas vezes antecipando ou consolidando processos de mudança. Por isso, sua análise permite observar fenômenos emergentes, que podem representar fases intermediárias de reestruturação gramatical, por exemplo (Barbosa; Marine, 2012). Os jornais, nesse sentido, oferecem um espaço híbrido entre a norma e o uso, funcionando como uma janela para as práticas linguísticas de seu tempo e, consequentemente, como um instrumento metodológico para investigações linguísticas.

Diante da relevância histórica e sociolinguística dos jornais como registros textuais que refletem práticas linguísticas e contextos sociais específicos, diversas pesquisas têm recorrido a esses materiais para investigar fenômenos de variação e mudança na língua. Ao explorar jornais antigos, pesquisadores têm conseguido acessar informações sobre usos concretos da língua escrita, frequentemente relacionados a dinâmicas sociais, políticas e culturais, que contribuem para ampliar o entendimento sobre como as variações linguísticas se manifestam e evoluem ao longo do tempo.

O estudo realizado por Barbosa e Balsalobre (2008), por exemplo, reflete sobre a utilização de jornais do início do século XX como uma fonte valiosa para pesquisas linguísticas. O foco principal deste estudo é investigar as características do gênero jornalístico, estabelecendo conexões entre essas características e o contexto linguístico

da época. As autoras realizam uma análise comparativa das formas pronominais de tratamento *tu* e *você*. Essa investigação é conduzida a partir de periódicos da imprensa negra[1] e da imprensa oficial do município de Araraquara.

De modo semelhante, Jorge (2018) investigou a variação na sintaxe da concordância verbal em manchetes de jornais eletrônicos sob a perspectiva da Teoria da Variação. Seu estudo revelou que, de modo geral, esses meios de comunicação têm procurado, conscientemente ou não, influenciar o estilo de escrita de seus leitores. Considerando que cada veículo se destina a um público distinto - composto por leitores de diferentes classes sociais e níveis de escolaridade -, os autores observaram que tanto a estrutura quanto o estilo linguístico adotado por cada jornal são moldados em função de seu público-alvo. A pesquisa, portanto, investigou de que modo essas diferentes formas de uso da linguagem influenciam a percepção e a recepção das mensagens pelos leitores. Observa-se, assim, que a variação linguística presente na concordância sintática das manchetes é uma estratégia dos redatores para encontrar a forma mais eficaz de comunicar a notícia ao público.

Passos (2023) investigou o uso dos verbos *ter* e *haver* na cidade de Goiás sob a ótica da Sociolinguística Histórica e utilizou jornais disponíveis no sistema da Hemeroteca Digital Nacional como *corpus*, adotando uma abordagem qualitativa para identificar padrões de uso e compreender o contexto sociocultural que permeavam esses usos verbais. Os periódicos constituíram, nessa pesquisa, fonte primária para viabilizar a análise diacrônica, uma vez que permitiram acessar diferentes sincronias da comunidade de fala local por meio de um suporte textual com datação precisa.

Outra contribuição foi o estudo de Barros (2021) que discute o uso dos jornais como fontes históricas, propondo um conjunto de procedimentos metodológicos para a análise historiográfica desse tipo de documento. O estudo focou no contexto do Brasil do início do século XIX, período em que os jornais tinham um caráter predominantemente político e circulavam em tiragens reduzidas, frequentemente controladas por pequenos grupos. Além disso, o artigo também aborda períodos posteriores da história da imprensa, nos quais os jornais passaram a atuar em mercados mais amplos, com tiragens significativamente maiores.

No capítulo "Early Advertising and Newspapers as Sources of Sociolinguistic Investigation" do livro *The Handbook of Historical Sociolinguistics*, Percy (2012) investigou o potencial dos jornais - tanto online quanto impressos - e dos anúncios publicitários antigos como fontes valiosas para pesquisas em sociolinguística histórica, especialmente no contexto da língua inglesa. A autora defende que esses materiais, muitas vezes negligenciados como evidência linguística, podem fornecer dados

[1] Imprensa negra é o nome dado ao conjunto de jornais, revistas e outros meios de comunicação produzidos por pessoas negras e voltados para pessoas negras com o objetivo de responder à exclusão da população negra da grande imprensa e de ser instrumento de luta política, afirmação identitária e denúncia do racismo.

relevantes sobre a variação linguística, a mudança linguística ao longo do tempo e as atitudes sociais em relação à linguagem. Ela argumenta que os anúncios, por sua natureza persuasiva e orientada para o público, revelam não apenas aspectos gramaticais e lexicais, mas também normas culturais, ideologias linguísticas e a construção de identidades sociais (como gênero, classe e nacionalidade).

Percy (2012) propõe uma abordagem metodológica que combina análise linguística rigorosa com atenção ao gênero textual, ao contexto histórico de circulação e aos aspectos gráficos e tipográficos do material impresso e online. O estudo também inclui exemplos de *corpora* construídos a partir de jornais do século XVIII e XIX, discutindo os desafios e possibilidades de se trabalhar com esse tipo de documento. Além disso, Percy (2012) enfatiza a importância da interdisciplinaridade e da integração entre a sociolinguística, a história da linguagem e os estudos culturais para a análise desses textos.

Além disso, há também um projeto desenvolvido pelo grupo GEVAR, cujo objetivo é a descrição de fenômenos linguísticos que foram ou estão em processo de variação no português falado em Minas Gerais, mas que ainda carecem de análise detalhada sob perspectivas sincrônicas e/ou diacrônicas. O projeto organiza um banco de dados composto por jornais locais – entre eles o *Lavoura e Comércio*, *Correio Católico* e *A Gazeta* de Uberaba, e o *Correio de Uberlândia*. A digitalização parcial do *Lavoura e Comércio* (fundado em 1889) já permite análises do português escrito da época, revelando características linguísticas peculiares do período. É o caso, por exemplo, da variação no uso dos verbos ter e haver com sentido existencial: ao lado de construções como "havia muitas festas na cidade", encontram-se usos como "tinha muitos problemas na lavoura", flagrando a competição entre as duas formas já em sincronias passadas e evidenciando um processo de mudança linguística em curso no português brasileiro. De forma similar, o *Correio de Uberlândia*, com exemplares digitalizados desde sua fundação em 1938, está sendo preparado para utilização linguística, com a transcrição para arquivos editáveis e a organização por seções e gêneros textuais. A constituição desse *corpus* possibilita estudos detalhados da variação linguística regional, contribuindo para a compreensão da identidade do português brasileiro a partir de sua diversidade interna.

O resgate de exemplares antigos se justifica por diversas razões:

- muitos desses documentos são peças únicas ou de circulação extremamente limitada, em virtude da perda ou deterioração ao longo do tempo;
- seus conteúdos tratam de aspectos fundamentais da experiência humana – como nascimentos, mortes, ordens, notícias, avisos e acontecimentos cotidianos – que, independentemente do tema, produzem efeitos concretos sobre os leitores e registram práticas sociais significativas; e

- são textos atravessados por uma marca enunciativa própria do discurso jornalístico, cuja forma de escrita carrega traços estilísticos e ideológicos que os distinguem de outros gêneros textuais.

Nesse contexto, acredita-se que a valorização e o estudo sistemático de acervos jornalísticos antigos contribuem não apenas para a preservação da memória cultural e histórica de localidades específicas, mas também para a constituição de um patrimônio discursivo mais amplo, de interesse nacional.

4. Perspectivas futuras

Diante dos argumentos apresentados ao longo deste artigo, a relevância do uso dos jornais como prática investigativa é inesgotável, na medida em que os impressos, bem como os online, oferecem múltiplas possibilidades de análise a cada nova abordagem, seja histórica, linguística ou discursiva. Assim, dado o acesso crescente a *corpora* históricos, espera-se que os jornais continuem sendo mais explorados como fonte de dados

Com o crescimento contínuo da digitalização de acervos em bibliotecas e hemerotecas, supõe-se uma ampliação significativa da acessibilidade não só aos jornais online, mas também aos jornais antigos, o que permitirá a constituição de *corpora* mais extensos, representativos e diversificados em termos cronológicos, geográficos e sociais. Essa ampliação também favorecerá a realização de estudos comparativos, possibilitando a observação de dinâmicas linguísticas em diferentes regiões e momentos históricos.

Paralelamente, a incorporação de ferramentas de análise linguística automatizada, como aquelas oriundas do campo do Processamento de Linguagem Natural (PLN)[2] (Ladeira, 2010), tende a desempenhar um papel central nas investigações futuras. Tais recursos permitirão a detecção de padrões de variação e mudança com maior precisão, facilitando o rastreamento de formas emergentes ou em declínio, além de possibilitar o cruzamento de dados linguísticos com metadados editoriais, sociais e históricos. A automação também pode colaborar na identificação de recorrências lexicais, sintáticas e ortográficas em larga escala, sem, no entanto, dispensar o olhar interpretativo e contextual do pesquisador.

No campo teórico, o uso continuado e aprofundado de textos jornalísticos pode contribuir para reavaliar noções consolidadas, como a distinção entre mudanças acima e abaixo do nível da consciência (Weinreich; Labov; Herzog, 2006 [1968]),

[2] O Processamento de Linguagem Natural (PLN) é uma área da inteligência artificial e da linguística computacional que se dedica a desenvolver métodos e sistemas capazes de compreender, interpretar, gerar e analisar a linguagem humana de forma automática por computadores.

uma vez que o discurso jornalístico, embora normatizado, frequentemente revela marcas de uso popular e local.

Além disso, as abordagens interdisciplinares despontam como outro eixo relevante para os estudos futuros. A articulação com campos como a história cultural, a sociologia da linguagem, os estudos de gênero e os estudos pós-coloniais podem aprofundar a compreensão sobre as ideologias linguísticas veiculadas pela imprensa, os processos de apagamento ou visibilização de vozes sociais, e a relação entre linguagem, poder e identidade.

Não obstante tais avanços, permanecem desafios metodológicos importantes, como a dificuldade de interpretar registros escritos como reflexo direto da oralidade, a escassez de exemplares em determinadas regiões ou períodos, e as limitações impostas pela normatização editorial, como os problemas de layout nas páginas em jornais online. Tais questões demandam atenção crítica e constante revisão das práticas analíticas, garantindo que a interpretação dos dados respeite suas condições de produção e circulação e que a análise não seja tendenciosa ou sem validação científica.

Considerações finais

Ao considerar os jornais como espaço de circulação de práticas discursivas marcadas por diferentes níveis de formalidade, intenções comunicativas e construções identitárias, foi possível destacar sua utilidade como fonte empírica para investigações que se debruçam sobre a variação e a mudança linguística. Mais do que veículos de opinião ou de notícias, os jornais carregam marcas do tempo, refletindo não apenas os acontecimentos que narram, mas também os modos de dizer, as escolhas linguísticas e os padrões discursivos de grupos sociais inseridos em contextos históricos específicos e comunidades específicas.

A diversidade de estudos aqui revisitados demonstra que o trabalho com jornais pode lançar luz sobre fenômenos linguísticos ainda pouco explorados ou compreendidos, principalmente quando se adota uma perspectiva metodológica atenta à materialidade do texto, à heterogeneidade dos gêneros jornalísticos e aos efeitos de sentido produzidos. Além disso, reforça-se a necessidade de se aliar o rigor analítico à sensibilidade histórica, reconhecendo que a linguagem registrada nesses impressos dialoga constantemente com processos sociais mais amplos – sejam eles de natureza política, ideológica ou cultural.

Entretanto, apesar de já existirem pesquisas que utilizam jornais como fonte de dados para investigações linguísticas, observa-se que esse tipo de abordagem ainda é relativamente restrita, especialmente no âmbito da Sociolinguística Histórica. Embora os periódicos ofereçam um vasto potencial para o estudo da variação e da mudança, muitas vezes são subutilizados, seja por desafios metodológicos (Percy, 2012), seja pela predominância de outras fontes mais consolidadas, como cartas pessoais ou documentos oficiais (Barbosa; Marine, 2007). A escassez de trabalhos

que exploram sistematicamente os jornais sob a perspectiva variacionista e histórica revela a necessidade de um investimento maior na construção de *corpora* a partir desse tipo de material, bem como no desenvolvimento de metodologias adequadas à análise de seus gêneros e condições de circulação.

Por fim, espera-se que este trabalho possa contribuir para a ampliação das possibilidades de investigação dentro da Sociolinguística Histórica, incentivando não apenas o uso de fontes como os jornais, mas também a reflexão crítica sobre seus alcances e limitações. Em um momento em que os estudos de língua e sociedade demandam abordagens cada vez mais interdisciplinares e ancoradas em dados reais, retornar aos arquivos da imprensa não significa apenas olhar para o passado, mas também compreender, com maior profundidade, os caminhos da língua em sua trajetória histórica e social.

Referências

BARBOSA, J. B.; BALSALOBRE, S. R. G. A imprensa como fonte para pesquisas linguísticas. Revista da ANPOLL, Brasília, n. 25, p. 62-89, 2008. DOI:https://doi.org/10.18309/anp.v1i25.55. Acesso em: 29 out 2025.

BARBOSA, J. B.; MARINE, T. de C. Organização de um Banco de Dados a partir de Jornais do Triângulo Mineiro: contribuições para o resgate da sócio história do português brasileiro. In: V Jornadas de Filologia y Lingüística - Identidades Dinámicas: Variación y Cambio en el Español de América. La Plata: FaHCE-UNLP, 2012. Disponível em: http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.3743/ev.3743.pdf. Acesso em: 29 out 2025.

BARBOSA, J. B.i; MARINE, T. de C.. O uso das Cartas como Corpus: uma nova possibilidade às Ciências Humanas. In: Anais do V Encontro de Linguística de Corpus. São Paulo: 2007, v. 5. p 1-18.

BARROS, J. d'A. Sobre o uso dos jornais como fontes históricas – uma síntese metodológica. Revista Portuguesa de História, v. 52, p. 421-443, 2021. DOI: https://doi.org/10.14195/0870-4147_52_17. Acesso em: 29 out 2025.

CONDE SILVESTRE, J. C. Sociolingüística Histórica. Madri: Gredos, 2007.

CULPEPER, J.; KYTÖ, M. 2010. Early Modern English Dialogues: spoken interaction as writing (studies in English Language). Cambridge: Cambridge University Press, 2010.

FARACO, C. A. Linguística histórica: uma introdução ao estudo da história das línguas. São Paulo: Parábola Editorial, 2005.

JORGE, A. A variação na sintaxe de concordância verbal em manchetes de jornais eletrônicos: uma mudança em progresso. Revista de Educação do Vale do Arinos - RELVA, v. 6, n. 1, p. 115–126, 2019. DOI: https://doi.org/10.30681/relva.v6i1.3770. Acesso em: 29 out 2025.

LABOV, W. The Social Stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics, 1966.

LABOV, W. Language in the Inner City: studies in the black English vernacular. Philadelphia: University of Pennsylvania Press, 1972.

LABOV, W. Padrões sociolinguísticos. Tradução de Marcos Bagno et al. São Paulo: Parábola Editorial, 2008 [1972].

LADEIRA, A. P. Processamento de Linguagem Natural: caracterização da produção científica dos pesquisadores brasileiros. 259f. Tese (Doutorado em Ciência da Informação). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

LASS, R. Historical Linguistics and Language Change. Cambridge: Cambridge University Press, 2000.

LOPES, Ca. Sociolinguística Histórica no Brasil: caminhos e desafios. Mesa redonda debatida por Rosane de Andrade Berlinck, Célia Regina dos Santos Lopes e Huda da Silva Santiago, sob mediação de Rosane de Andrade Berlinck, 2020. 1 vídeo (1h 49min 20s). Publicado pelo Canal da Associação Brasileira de Linguística. Disponível em: https://www.youtube.com/watch?v=nE4i53QDacE&t=4054s. Acesso em: 26 jun. 2024.

NEVALAINEN, T.; RAUMOLIN-BRUNBERG, H. Historical Sociolinguistics: language change in Tudor and Stuart England. London: Longman, 2003.

NEVALAINEN, T. What are historical sociolinguistics? Journal of Historical Sociolinguistics, v. 1, n. 2. p. 243-269. 2015. DOI: https://doi.org/10.1515/jhsl-2015-0014. Acesso em: 29 out 2025.

PASSOS, A. G. A. Ter e Haver em Jornais da Cidade de Goiás-GO: uma abordagem sociolinguística. 20f. Trabalho de Conclusão de Curso (Graduação em Letras). Faculdade de Letras, Universidade Estadual de Goiás, Cora Coralina, 2023.

PERCY, C. Early Advertising and Newspapers as Sources of Sociolinguistic Investigation. In: HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C. (Orgs.). The Handbook of Historical Sociolinguistics. Nova Jersey: Wiley-Blackwell, 2012, p. 191-210.

ROMAINE, S. Socio-historical Linguistics: its status and methodology. Cambridge: Cambridge University Press, 2009 [1982].

ROSA, E. Sociolinguística Histórica. Revista de Letras, Curitiba, v. 17, n. 21, p. 1-17, 2015.

SEVIC, R. Early Collections of Private Documents: the missing link in the diachronic corpora. In: BEEDHAM, Christopher (Org.). Langue and Parole in Synchronic and Diachronic Perspective. Selected Proceedings of the XXXIst Annual Meeting of the Societas Linguistica Europaea. St. Andrews, 1999, p. 337–347.

VAN DER WAL, M.; RUTTEN, G. Ego-documents in a Historical-sociolinguistic Perspective. In: VAN DER WAL, M.; RUTTEN, G. (Orgs.). Touching the Past: studies in the historical sociolinguistics of ego-documents. Amsterdam & Philadelphia: Benjamins, 2013, p. 1-17.

WEINREICH, U.; LABOV, W.; HERZOG, M. Fundamentos Empíricos para uma Teoria da Mudança Linguística. São Paulo: Parábola Editorial, 2006 [1968].