A influência de fatores sociais sobre a variação entre expressão e não expressão de pronome sujeito no português brasileiro oitocentista

The influence of social factors on the variation between expression and non-expression of pronominal subject in Brazilian Portuguese in the 19th century

Alessandra Regina Guerra¹

Resumo: A língua portuguesa caracteriza-se por uma opcionalidade entre construir orações que têm como sujeito um pronome pessoal (*eu comprei um livro*) e construir orações sem expressão de pronome sujeito (*comprei um livro*). Na história do português brasileiro (PB), uma mudança linguística típica é o progressivo aumento na frequência de uso da primeira dessas opções, em detrimento da segunda. Este artigo filia-se a um projeto de pesquisa mais amplo que tem o propósito de investigar se tal mudança teria sido influenciada por fatores sociais. Nesse contexto, o artigo tem o objetivo de analisar se, no PB do século XIX, a variação entre expressão e não expressão de pronome sujeito é condicionada por variáveis sociais, considerando as seguintes variáveis: instrução, fase da vida, classe social, gênero/sexo. O quadro teórico é uma articulação entre a Sociolinguística Variacionista, a Sociolinguística Histórica e a Gramática Discursivo-Funcional. O trabalho emprega metodologia variacionista e utiliza, como *corpus*, um conjunto de peças teatrais brasileiras do século XIX. Os resultados sugerem que as variáveis instrução e fase da vida exercem influência sobre a mudança em questão: graus mais altos de instrução e a segunda fase da vida (16 a 50 anos) conservam índices mais altos de não expressão de sujeito.

Palavras-chave: Variação linguística. Mudança linguística. Pronome sujeito. Português brasileiro oitocentista.

Abstract: Portuguese language is characterized by an optionality between building a clause with a personal pronoun as subject (*eu comprei um livro – I bought a book*) and building it without pronominal subject (*comprei um livro – I bought a book*). In the history of Brazilian Portuguese (BP), a typical linguistic change is the progressive increase in the use of the first option, in opposition to the second one. This paper is part of a broader research project whose purpose is to investigate whether such change has been influenced by social factors. In this context, the article aims to analyze if, in BP, in the 19th century, the variation between expression and non-expression of pronominal subjects is influenced by social variables. The analysis addresses the following variables: education, stage of life, social class, gender/sex. The theoretical framework is an articulation among Variationist Sociolinguistics, Historical Sociolinguistics and Functional Discourse Grammar. The paper is based on variationist methodology, and it uses a corpus that is composed of Brazilian theatre plays produced in the 19th century. The results suggest that the variables education and stage of life are relevant: higher levels of education and the second stage of life maintain higher rates of non-expression of subject.

Keywords: Linguistic variation. Linguistic change. Pronominal subject. 19th century Brazilian Portuguese.

¹ Universidade Estadual Paulista (UNESP), *campus* de Araraquara. Núcleo de Pesquisas em Sociolinguística de Araraquara.

Considerações iniciais

Este artigo apresenta e discute resultados de investigação conduzida no interior do Núcleo de Pesquisas em Sociolinguística de Araraquara (SoLAR), sediado na Universidade Estadual Paulista (UNESP), *campus* de Araraquara, e inserido no campo dos estudos sociolinguísticos. Nos últimos anos, entre outras frentes de atuação, o grupo vem analisando o impacto de fatores sociais sobre diferentes processos de mudança diacrônica ocorridos no português brasileiro (PB). Nesse contexto, temos desenvolvido um projeto de pesquisa dedicado a uma mudança que, de acordo com a perspectiva sócio-funcionalista assumida no projeto (e também aqui), consiste no aumento da incidência de expressão pronominal de argumento-sujeito (como em *eu comprei um livro*), em oposição à opção por não expressão pronominal (*comprei um livro*).

Trata-se de mudança correspondente, no geral, ao que, nos termos da Gramática Gerativa, é tratado como aumento nos índices de uso de sujeito pronominal pleno (ou diminuição do emprego de sujeito nulo). O processo é amplamente tomado como fundamental na história do PB, sendo visto por Tarallo (1993), no âmbito da Sociolinguística Paramétrica, como parte de um conjunto de mudanças que teriam alterado drasticamente o PB, transformando-o num novo sistema gramatical e o diferenciando do português europeu.

Em nosso projeto de pesquisa, examinamos se a mudança em questão é impactada por fatores sociais no decorrer do século XIX ao XXI. No presente artigo, expomos parte dos resultados desse projeto, no caso resultados relativos ao século XIX. Assim, o objetivo do artigo é analisar se, no PB do século XIX, a variação entre expressão e não expressão pronominal de argumento-sujeito é condicionada por fatores sociais, considerando as seguintes variáveis sociais: grau de instrução do falante, fase da vida, classe social, gênero/sexo.

Dessa forma, o artigo procura não só contribuir para o propósito mais amplo do grupo SoLAR de investigar a correlação entre fatores sociais e processos de mudança linguística, mas também colaborar, particularmente, para a elaboração de um largo conjunto de estudos sobre a mudança relativa à (não) expressão de pronomes sujeito, tendo em vista sua relevância para a história do PB.

Nesse sentido, em trabalhos recentes temos analisado essa mudança em relação a diferentes aspectos, procurando contribuir para explicá-la de forma cada vez mais completa e específica. Em Guerra (2022), apresentamos dados quantitativos diacrônicos sobre a ocorrência da mudança, levantados a partir de amostras do PB do século XIX ao XXI; em Guerra (2021), descrevemos a influência que a mudança acarreta sobre a alteração diacrônica do grau de transparência linguística do PB; em Guerra (2025), tratamos também da diacronia da transparência do PB, focando, dessa vez, as alterações decorrentes da interação entre o aumento diacrônico no uso de expressão pronominal de argumento-sujeito e a passagem do uso do pronome *tu* para o predomínio do pronome *você*, processos

diacrônicos diretamente interligados. No presente artigo, damos continuidade à meta de propiciar uma ampla descrição do fenômeno diacrônico relativo à (não) expressão de pronome sujeito, explorando-o agora no que tange a sua relação com fatores sociais.

O artigo encontra-se organizado da seguinte forma: na próxima seção, sintetizamos os fundamentos teóricos e empíricos do artigo; em seguida, explicamos os procedimentos metodológicos; na sequência, procedemos à análise de dados; a última seção é dedicada às considerações finais.

Fundamentos teóricos e empíricos

O quadro teórico de nossa pesquisa é composto por uma articulação entre a Sociolinguística Variacionista, a Sociolinguística Histórica, que fundamentam o que diz respeito ao estudo de variação e mudança linguística em correlação com aspectos sociais, e a Gramática Discursivo-Funcional, empregada principalmente para embasar o tratamento da categoria gramatical de argumento-sujeito. O trabalho firma-se ainda em estudos sobre (não) expressão pronominal de argumento-sujeito no PB, como fundamento empírico da mudança linguística a que remete.

A Sociolinguística Variacionista, ou Teoria da Variação e Mudança Linguística (TVM), é uma abordagem social para estudo da estrutura e evolução das línguas. Historicamente, a abordagem desponta como reação ao estruturalismo e ao gerativismo, os dois paradigmas linguísticos mais bem estabelecidos na época do surgimento da teoria. Em termos gerais, a TVM opõe-se a dois pressupostos fundamentais manifestos nessas duas perspectivas: o objeto de estudo linguístico é um construto homogêneo (não apresenta variação); descrever adequadamente esse objeto implica a consideração das regras e relações internas dos componentes do sistema e a exclusão de possíveis fatores externos sobre o funcionamento desse sistema.

O programa de estudos da TVM encontra-se reunido principalmente em duas obras centrais: *Fundamentos empíricos para uma teoria da mudança linguística* (Weinreich; Labov; Herzog, 2006 [1968]) e *Padrões sociolinguísticos* (Labov, 1972). Essas obras propõem postulados teóricos e metodológicos para trabalho linguístico empírico. A nova abordagem assume como fundamental o reconhecimento da existência de estreita inter-relação entre língua e sociedade e, por isso, postula que “não se pode entender o desenvolvimento de uma mudança linguística sem levar em conta a vida social da comunidade em que ela ocorre” (Labov, 1972, p. 3). Assume ainda que o sistema linguístico é heterogêneo e que a variação linguística é propriedade regular desse sistema: “numa comunidade de fala, não há dois falantes que se expressem da mesma forma ou um mesmo falante que se expresse do mesmo modo nas diversas situações comunicativas” (Coelho, 2012, p. 22). Em outras palavras, a noção de variação

inerente às línguas implica que as línguas são sistemas heterogêneos (e não homogêneos) e que essa heterogeneidade ou variação pode ser ordenada ou organizada, já que há também aqui a ideia de “sistema”. Na TVM, o fato de existirem, em uma mesma comunidade linguística ou mesmo na fala de um único indivíduo, formas linguísticas alternativas não é considerado como marginal ou irrelevante.

A língua entendida enquanto sistema heterogêneo contém não apenas regras categóricas (como ocorre na visão de língua como sistema homogêneo), mas também regras variáveis. A noção de regra variável se contrapõe à de variação livre (ou de não condicionamento das variantes), comum em abordagens estruturalistas. O estudo das regras variáveis é central para uma abordagem sociolinguística: trata-se de apreender quais fatores linguísticos ou sociais estariam condicionando a aplicação ou não de tais regras. O instrumental metodológico da TVM permite, então, analisar e sistematizar os diversos tipos de variação linguística. Para a TVM, as construções em variação ocorrem em todos os níveis de análise de uma língua e não comprometem o funcionamento do sistema linguístico, mas atendem a demandas comunicativas e veiculam significados sociais específicos dos falantes, como local de origem ou de vivência, faixa etária aproximada, grupos (político, social etc.) com os quais se identificam etc. A observação de regularidades envolvendo a coexistência de variantes linguísticas e suas probabilidades de uso são captadas nessa abordagem pelo uso de técnicas quantitativas, o que confere ao sistema pressuposto pela TVM uma natureza probabilística.

Segundo Romaine (2005), a obra de Weinreich, Labov e Herzog (1968), pela sua ênfase na necessidade de incorporar fatores extralinguísticos em uma teoria da mudança linguística e por transcender a velha dicotomia entre sincronia e diacronia, além de marcar o surgimento da Sociolinguística enquanto um novo campo de investigação, também pode ser citada como obra que lança as bases para a Linguística Sócio-Histórica, também conhecida como Sociolinguística Histórica. A Sociolinguística Histórica é comumente entendida como estudo da relação entre linguagem e sociedade em sua dimensão histórica e pode ser definida como “a reconstrução da história de uma dada língua em seu contexto sociocultural” (Hernández-Campoy; Conde-Silvestre, 2012, p. 1).

O estabelecimento da Sociolinguística Histórica enquanto um campo linguístico independente e sua expansão ocorrem a partir do estudo de Romaine (1982). Em seu livro *Linguística sócio-histórica: seu status e metodologia*, a autora defende o uso do aparato teórico-metodológico da Sociolinguística em análises linguísticas históricas e faz uma das primeiras menções ao uso do termo *sócio-histórico* (Nevalainen; Raumolin-Brunberg, 2012).

Segundo Auer et al. (2015), na Sociolinguística Histórica, uma tarefa fundamental reside em encontrar respostas para as questões teóricas levantadas em Weinreich, Labov e Herzog (1968), a saber: (i) o problema dos fatores condicionantes (identificar os fatores motivadores da mudança linguística); (ii) o problema da *transição* (descrever como a mudança linguística se propaga e se difunde na sociedade); (iii) o problema do *encaixamento* (analisar o ajuste da mudança nas matrizes linguística e social); (iv) o problema da *avaliação* (identificar os correlatos subjetivos/a apreciação dos falantes acerca das mudanças linguísticas em estudo); (v) o problema da *implementação* (investigar como a mudança se estabelece e é incorporada pela comunidade). Conforme esclarecem Auer et al. (2015), na Sociolinguística Histórica os problemas de transição, encaixamento e avaliação alcançaram maior atenção. Assim, questões relacionadas a “como e quando as mudanças são transmitidas de um falante para outro”, “como novas formas se estabelecem nas comunidades de fala, em todas as faixas etárias, profissões ou estratos sociais”, e “como o prestígio, as normas de correção e as atitudes dos falantes em relação a formas específicas podem afetar as mudanças” são centrais para a disciplina (Auer et al., 2015, p. 4).

Para responder a perguntas como essas, na Sociolinguística Histórica equacionam-se o uso do princípio uniformitarista e a atenção às armadilhas do “anacronismo ideacional” (Bergs, 2012): ou seja, considera-se o princípio de que “as forças que operam para produzir mudanças linguísticas hoje são do mesmo tipo e ordem de magnitude que aquelas que operavam há cinco ou dez mil anos” (Labov, 1972, p. 275), com atenção ao risco de transposição de conceitos cuja aplicabilidade e validade são restritas às sociedades ocidentais modernas (classe social, gênero, prestígio) para cenários históricos.

Como estabelecido anteriormente, este artigo tem o objetivo de identificar possíveis influências de fatores sociais (instrução, fase da vida, classe social e gênero/sexo) sobre a variação entre expressão e não expressão pronominal de argumento-sujeito no PB do século XIX. Essa análise fundamenta-se nos preceitos fundamentais da TVM e da Sociolinguística Histórica discutidos acima. A partir de uma visão de língua como heterogeneidade ordenada, investigamos se fatores extralinguísticos estariam condicionando a aplicação de regras variáveis e se significados sociais (como prestígio social) podem estar sendo codificados por meio do uso das diferentes construções.

Além de firmar-se no quadro teórico exposto, o artigo recorre ainda ao modelo da Gramática Discursivo-Funcional (Hengeveld; Mackenzie, 2008) para caracterizar o conceito gramatical específico de argumento-sujeito e a opcionalidade entre sua expressão e não

expressão pronominal. A Gramática Discursivo-Funcional (GDF) é uma teoria funcionalista da estrutura gramatical das línguas. O modelo abrange quatro níveis de análise linguística: o nível interpessoal (NI), o nível representacional (NR), o nível morfossintático (NM) e o nível fonológico (NF). O NI descreve aspectos pragmáticos dos enunciados, incluindo unidades como *atos discursivos*, *ilocuções*, *subatos de atribuição* (que evocam propriedades) e *subatos de referência* (que evocam referentes). O NR descreve aspectos semânticos dos enunciados, mediante unidades como *conteúdos proposicionais* e *estados de coisas*; ainda o NR descreve os enunciados em termos de esquemas de predicação, com predicados e argumentos, sendo também responsável pela marcação de funções semânticas, como agente e paciente. O NM trata da estrutura morfossintática dos enunciados, com unidades como *orações*, *sintagmas*, *palavras*, *radicais* e *afixos*. O NF analisa a estrutura fonológica dos enunciados, mediante unidades como *sintagmas entoacionais*, *palavras fonológicas*, *pés* e *sílabas*.

Como exemplificação, considere-se a sentença em (1), observando o NI, o NR e o NM, mais relevantes para apreensão da noção de argumento-sujeito:

(1) Eu comprei um livro.

Na GDF, a análise desse exemplo passa por descrever, no NI, que a sentença, entre outros aspectos, é formada por um ato discursivo declarativo, composto por um subato atributivo (para indicar que a sentença evoca uma propriedade, a ação de comprar), dois subatos de referência que evocam o próprio falante (um a ser realizado pelo pronome *eu* e outro, pela desinência *i*), e um novo subato de referência (a ser preenchido, nos níveis seguintes, pelo sintagma *um livro*) que evoca um referente não identificável pelo ouvinte. O NR registra, por exemplo, o esquema de predicação, que inclui um predicado (preenchido pelo item lexical *comprar*, que descreve semanticamente a evocação de propriedade feita no subato atributivo do NI), seu primeiro argumento (que funciona como agente e corresponde ao falante) e seu segundo argumento (que funciona como paciente e é, então, preenchido pelo lexema *livro*, que propicia a descrição semântica da evocação de referente não identificável feita no NI). O NM registra, por exemplo, que a sentença é composta por uma oração, constituída por um primeiro sintagma nominal (com a função sintática de sujeito), um sintagma verbal e um segundo sintagma nominal (que atuará como objeto), nessa ordem; aí é também inserido o artigo *um* (do sintagma *um livro*), codificando a significação interpessoal de referente não identificável.

Essa rápida ilustração do modelo permite delimitar a categoria do *argumento-sujeito*, que se refere ao argumento (do esquema de predicação do NR) que, na implementação de uma

oração, é selecionado para a função de sujeito (marcada no NM). No exemplo em (1), que está na voz ativa, o argumento-sujeito é o argumento 1.

No PB, em oração finita, o argumento-sujeito é expresso por desinência verbal (no que diz respeito a informação de pessoa e número).² Quando envolve informação lexical, é também expresso por sintagma nominal na posição sintática de sujeito da oração. Já quando envolve apenas informação de pessoa e número (ou de pessoa, número e gênero), o argumento-sujeito pode ser expresso por pronome pessoal na posição de sujeito, como em (1), ou pode não ser expresso por pronome (sendo sua expressão realizada apenas por desinência verbal), como em *comprei um livro*. Essa é a opcionalidade entre *expressão e não expressão pronominal de argumento-sujeito* abordada no artigo, a que também nos referimos, de modo simplificado, como *(não) expressão de pronome sujeito* ou mesmo *(não) expressão de sujeito* (a qual, como citado, é tratada em termos de *sujeito pronominal pleno e nulo*, no quadro gerativista).³

A fundamentação do trabalho é complementada ainda por um conjunto de estudos segundo os quais se verifica no PB, diacronicamente, um aumento no índice de expressão pronominal de argumento-sujeito, passando a língua do predomínio da não expressão para a prevalência da expressão pronominal. Por exemplo, Barbosa, Duarte e Kato (2001) fornecem dados do PB que indicam a progressiva ocorrência da mudança. As autoras apuram as seguintes frequências de sujeito pleno: 20% (em 1845), 23% (1882), 25% (1918), 46% (1937), 50% (1955), 67% (1975), 74% (1992). Tarallo (1993), igualmente atestando a mudança, identifica os seguintes índices de sujeito pleno no PB: 23% (1725), 27% (1775), 16% (1825), 33% (1880), 79% (1981). No mesmo sentido, em trabalho anterior (Guerra, 2022), levantamos os seguintes dados percentuais de expressão pronominal de argumento-sujeito no PB, os quais também sugerem a mudança em questão: 21% (primeira metade do século XIX), 23,5% (segunda metade do século XIX), 43% (primeira metade do século XX), 58% (segunda metade do século XX), 67,5% (início do século XXI).⁴

Como explicado, o presente artigo assume a ocorrência do aumento diacrônico da expressão, *versus* a não expressão, pronominal de argumento-sujeito no PB e procura contribuir para que se possa analisar se fatores sociais teriam influenciado essa mudança, disponibilizando dados empíricos relativos ao século XIX, conforme tratado nas seções seguintes.

² O fenômeno aqui referido, com base em Leufkens (2015), como expressão de argumento-sujeito por desinência verbal número-pessoal (que, em termos mais específicos, constitui um subato de referência à entidade correspondente ao argumento-sujeito) corresponde ao que, tradicionalmente, é conhecido como concordância verbal.

³ Para um tratamento funcionalista da expressão pronominal do argumento-sujeito no PB, ver Oliveira (2018).

⁴ Em Guerra (2022), relatamos e discutimos dados de várias pesquisas que apontam para a referida mudança.

Procedimentos metodológicos

Para averiguar se fatores sociais exercem algum papel condicionador sobre o aumento diacrônico da frequência de expressão pronominal de argumento-sujeito (em prejuízo da não expressão pronominal), nossa pesquisa levantou dados em *corpus* de investigação. O *corpus* utilizado consistiu em uma subamostra do banco de dados elaborado no âmbito do projeto coletivo de pesquisa *Língua EmCena* (financiado pelo CNPq, processo 405181/2021-3), atualmente desenvolvido pelo grupo SoLAR. Esse banco de dados é composto por peças teatrais brasileiras dos séculos XIX a XXI e pela caracterização do perfil social dos personagens, no que diz respeito a fatores sociais relevantes para pesquisas sociolinguísticas, como gênero/sexo, idade/fase da vida, escolaridade/instrução, classe social e etnia. No banco de dados, as peças são distribuídas entre quatro recortes sincrônicos: século XIX (considerado, especificamente, de 1831 a 1889) e três fases do século XX (XX.1, 1890 a 1930; XX.2, 1931 a 1964; XX.3, 1965 a 2010), períodos que recobrem fatos históricos e sociais importantes para o país, segundo estudos sociológicos e históricos (Schwarcz, 2011).

O material do século XIX (fase aqui em análise) é composto por peças de quatro autores (Artur de Azevedo, França Junior, Machado de Assis e Martins Penna). A subamostra adotada em nossa pesquisa para levantamento de ocorrências de (não) expressão de pronome sujeito foi constituída por duas ou três peças de cada autor (o banco de dados conta com várias peças de cada um deles), mais especificamente pelas três primeiras cenas de cada peça. Tal delimitação foi feita para obtenção de um quantitativo de ocorrências viável para análise, já que o levantamento de todas as ocorrências de todas as peças inteiras resultaria num volume excessivamente grande de casos para análise. A escolha das peças foi orientada de modo a se obter personagens representativos dos diferentes valores previstos pelas variáveis sociais controladas na pesquisa, por exemplo personagens das diferentes faixas de idade previstas na variável *fases da vida*, como personagens infantis.⁵ Assim, foram utilizadas as seguintes peças:

Quadro 1: Peças teatrais integrantes do *corpus* da pesquisa.

Ano	Autor	Peça
1860	Artur de Azevedo	<i>Abel e Helena</i>
1888	Artur de Azevedo	<i>A Almanjarra</i>
1897	Artur de Azevedo	<i>A capital Federal</i>
1861	França Junior	<i>Meia hora de cinismo</i>
1882	França Junior	<i>Caiu o Ministério</i>
1861	Machado de Assis	<i>Desencantos</i>

⁵ Mesmo com esse critério, não foi possível encontrar personagens de algumas faixas de determinadas variáveis, como personagens com 50 anos ou mais, conforme informado adiante.

1863	Machado de Assis	<i>O caminho da porta</i>
1840	Martins Penna	<i>Judas em Sábado de Aleluia</i>
1845	Martins Penna	<i>O Noviço</i>
1845	Martins Penna	<i>Os três médicos</i>

Fonte: elaboração própria.

O estudo também delimitou-se à 1ª e à 2ª pessoa do discurso (singular e plural), não incluindo a 3ª pessoa (singular e plural). Como é geralmente concebido nos estudos gramaticais, 1ª e 2ª pessoas constituem as chamadas pessoas do discurso, em oposição à 3ª, normalmente tomada como não pessoa. Assim, essa delimitação circunscreve nossa pesquisa a um subsistema particular dentro da gramática do PB, o domínio da referência dêitica.

O levantamento de ocorrências foi ainda conduzido conforme os seguintes critérios:

- (i) foram coletadas ocorrências de expressão e não expressão de argumento-sujeito apenas em orações principais (orações absolutas, orações matrizes de subordinadas, orações núcleos de adverbiais e primeira oração de estruturas coordenadas);
- (ii) no contexto da 2ª pessoa do singular, o levantamento considerou ocorrências relativas tanto ao pronome *tu* quanto ao pronome *você*; igualmente, no âmbito da 1ª pessoa do plural, o levantamento incorporou casos de *nós* e de *a gente*; da mesma forma, no domínio da 2ª pessoa do plural, o levantamento considerou casos referentes às formas *vós* e *vocês*;
- (iii) foram excluídas ocorrências compostas pela coordenação entre formas pronominais e outros sintagmas, como em *eu e meu irmão fomos ao cinema*, *você e Maria fazem a mesma faculdade* etc.;
- (iv) foram descartadas ocorrências de orações com verbos no imperativo.

A partir das delimitações e dos critérios estabelecidos, a coleta de dados forneceu 449 ocorrências entre expressão e não expressão pronominal de argumento-sujeito, que foram, então, cruzadas com as variáveis sociais consideradas na pesquisa. Foram selecionadas quatro variáveis tradicionalmente reconhecidas como relevantes na análise sociolinguística, conforme especificado a seguir:

(4) Variáveis sociais controladas na pesquisa:⁶

(i) instrução:⁷

⁶ Em nossa pesquisa, essas variáveis sociais foram tomadas como variáveis independentes e a opcionalidade entre expressão e não expressão de pronome sujeito, como variável dependente.

⁷ No banco de dados do qual foi extraído o *corpus* desta pesquisa, a identificação do perfil social dos personagens das peças (por exemplo, a identificação de que um dado personagem tem instrução básica ou média) é realizada

- a) sem instrução (personagens analfabetos);
 - b) básica (personagem sabe ler (e escrever));
 - c) média (personagem exerce algum ofício e/ou tem algum tipo de conhecimento especializado);
 - d) avançada (personagem está se preparando para ingressar no ensino superior ou já exerce uma profissão que demanda formação superior).
- (ii) fase da vida:
- a) fase 1 (criança e jovem; 0-15 anos);
 - b) fase 2 (adulto; 16-50 anos);
 - c) fase 3 (idoso/velho; 51 anos ou mais).
- (iii) classe social:
- a) classe 1 (alta burguesia, alta sociedade, dominantes);
 - b) classe 2 (classes médias, classe emergente, intermediária);
 - c) classe 3 (proletariado, camponeses, classe subalterna);
 - d) classe 4 (oprimidos, marginais).
- (iv) gênero/sexo:
- a) masculino;
 - b) feminino.

O cruzamento de dados foi realizado eletronicamente, mediante emprego do programa R, uma linguagem de programação voltada à análise de dados, que pode ser utilizada para realizar computações estatísticas e gráficas, compilar e anotar *corpora*, produzir listas de frequências, entre diversas outras tarefas (Oushiro, 2014). Para cada cruzamento de variáveis, o programa gera um valor, chamado *valor p*, que indica a significância estatística dos dados em análise. Quanto menor for o valor, maior é a probabilidade estatística de que as variáveis analisadas tenham correlação. É comum utilizar o valor 0.05 como valor limítrofe para determinar a relevância estatística. Se o valor *p* é inferior a 0.05, há boa chance de que haja correlação entre as variáveis.

A seção seguinte apresenta, então, os resultados do cruzamento das variáveis sociais selecionadas com a opcionalidade entre expressão e não expressão pronominal de argumento-sujeito.

Análise de dados

A Tabela 1 exhibe resultados do cruzamento entre (não) expressão pronominal de argumento-sujeito e grau de instrução:

com base em informações contidas, nas próprias peças, acerca de seus personagens e mediante diálogo interdisciplinar com fontes e pesquisadores das áreas de História, Sociologia, Antropologia e Teatro.

Tabela 1: Cruzamento entre (não) expressão de sujeito e grau de instrução.

(Não) expressão de sujeito \ Instrução	Não expressão % (N)	Expressão % (N)	Total ⁸ % (N)
Básica	69,7% (147)	30,3% (64)	100% (211)
Média	76,2% (77)	23,8% (24)	100% (101)
Avançada	77,3% (75)	22,7% (22)	100% (97)
Valor p: 0.08 (> 0.05)			

Fonte: Elaboração própria.

Como mostra a tabela, o valor p, embora se encontre bastante próximo do valor limítrofe de 0.05, ainda ultrapassa esse limite, o que, a princípio, sugere que as variáveis em análise não mantenham entre si uma correlação mais sistemática. De todo modo, entendemos que os dados percentuais da tabela permitem formular algumas hipóteses (a serem confirmadas na continuação da pesquisa, mediante ampliação da amostra e comparação com dados de outras sincronias).⁹

A esse respeito, pode-se notar que, em todos os graus de instrução, predomina a variante não expressão pronominal – resultado, aliás, compatível com o que outras pesquisas (que fornecem dados gerais, sem controle de variáveis sociais) apontam para o período (século XIX), como é o caso dos dados relatados em seção anterior, segundo os quais, ao longo de todo o século XIX, ainda prevalece a opção por não expressão de pronome sujeito.

No entanto, verifica-se, entre os graus de instrução, uma diferença que pode estar apontando na direção de um papel condicionador da variável instrução sobre a (não) expressão pronominal. É possível notar, por um lado, as instruções avançada e média, com taxas de não expressão pronominal muito próximas entre si (77,3% e 76,2%), e, por outro lado, a instrução básica, com taxa relativamente menor (69,7%) (o que equivale a dizer que as instruções

⁸ Como explicado na seção de metodologia, o total de ocorrências levantadas no *corpus*, entre expressão e não expressão pronominal, foi de 449 casos. Porém, na análise da variável instrução, o total de ocorrências analisadas é um pouco menor (409 casos), pois estão excluídas as ocorrências produzidas por personagens cujo grau de instrução não pode ser apreendido na peça teatral. Mais adiante no artigo, pela mesma razão, os totais de ocorrências, na análise das variáveis fase da vida e classe social, também são menores que 449 casos. Já na análise da variável gênero/sexo, estão incluídas todas as ocorrências, pela viabilidade de se apreender o gênero/sexo de todos os personagens. Cabe também esclarecer que a Tabela 1 não contém dados relativos à faixa “sem instrução” porque o *corpus* analisado não contemplou esse tipo de personagem, distribuindo-se todos os dados pelas três outras faixas.

⁹ Como se verá adiante, nos demais cruzamentos de variáveis, expostos nas tabelas seguintes, apura-se o mesmo tipo de resultado acerca do valor p (que fica acima de 0.05). Também nesses casos permanece válido nosso entendimento de que os dados percentuais permitem a formulação de hipóteses a serem testadas posteriormente.

avançada e média têm índices menores de expressão pronominal, enquanto a instrução básica tem taxa maior de expressão de sujeito). Esses dados sugerem o seguinte caráter condicionador da instrução sobre a (não) expressão de pronome sujeito, no século XIX: graus mais altos de instrução estariam correlacionados a índices mais altos de não expressão pronominal de argumento-sujeito, enquanto graus mais baixos de instrução estariam correlacionados a índices relativamente menores de não expressão pronominal. Em outros termos, quanto maior o grau de instrução, maior a tendência para emprego de não expressão pronominal (o que fica, de fato, indicado pelos percentuais de não expressão exibidos pelas instruções básica, média e alta, que aumentam progressivamente, sendo 69,7%, 76,2% e 77,3% respectivamente).¹⁰

Com efeito, a correlação entre instrução avançada e maior preferência por não expressão pronominal pode ser vista como sendo condizente com uma tendência geral apurada em estudos sociolinguísticos. Tavares (2016) elenca um conjunto de tendências, constatadas na área, acerca da influência de fatores sociais sobre fenômenos de variação e mudança, entre as quais a autora lista a seguinte tendência relativa à variável escolaridade (correspondente ao que aqui chamamos de *instrução*):

níveis mais altos de escolaridade se correlacionam com traços linguísticos avaliados positivamente pela comunidade de fala – os traços linguísticos ditos “prestigiados”; em contraparte, níveis mais baixos de escolaridade se correlacionam com traços linguísticos avaliados negativamente pela comunidade de fala – os traços linguísticos tidos como informais ou mesmo estigmatizados (Tavares, 2016, p. 118).

Essa generalização permite esboçar uma possível explicação para os dados expostos na Tabela 1: a instrução avançada exibe índice maior de não expressão pronominal e a instrução básica, índice menor porque, no século XIX, a não expressão seria a variante de (maior) prestígio social. Em princípio, a validação dessa explicação é dificultada, pela natural impossibilidade de se examinar a avaliação (positiva ou negativa) dos falantes acerca das variantes em questão, já que se trata aqui do PB do século XIX. Porém, parece plausível cogitar que, no período, poderia ter havido certa correlação entre não expressão pronominal e (maior) prestígio social (i) assumindo que a não expressão (em relação à expressão pronominal) teria sido uma variante (mais) associada à norma-padrão da época (a variante prototípica da norma-

¹⁰ Conforme sugerem nossos dados, a oposição entre instrução mais alta e mais baixa não chega a correlacionar-se a uma oposição entre preferência por não expressão *versus* preferência por expressão pronominal, o que certamente significaria um condicionamento mais forte e evidente da instrução sobre a (não) expressão pronominal, mas se correlaciona a uma oposição entre maior e menor preferência por não expressão, o que, a nosso ver, não deixa de representar uma influência da instrução sobre a variação em questão.

padrão) e (ii) assumindo que formas e construções linguísticas típicas da norma-padrão tendem a desfrutar de prestígio social.

A possibilidade de que a não expressão pronominal, no período, tenha sido uma variante (mais) associada à norma-padrão pode ser ancorada, primeiramente, no entendimento de que a norma-padrão, diante de fenômenos de variação linguística, tende a comportar a variante conservadora, não a inovadora.¹¹ No caso da opcionalidade entre não expressão e expressão pronominal de argumento-sujeito, nossa compreensão é a de que não seja pertinente considerar, simplesmente, que uma variante (a não expressão) seja conservadora e a outra (a expressão pronominal) seja inovadora, uma vez que ambas já coexistiriam desde a formação da língua portuguesa (os próprios dados de Tarallo (1993) acima apontados atestam a ocorrência da variação um século antes, em 1725; ademais, como lembra Olbertz (2020), a opcionalidade de especificação do sujeito, no português, é herdada já do latim). Não obstante, é legítimo admitir que não expressão e expressão pronominal estejam ligadas aos traços de conservação e inovação no que diz respeito à *frequência de uso* de uma e outra dessas variantes. No século XIX, o uso mais frequente da não expressão e o uso pouco frequente da expressão (isto é, a preferência pela não expressão) constituem o que se pode tomar como uso linguístico conservador; o fato inovador não seria a expressão pronominal em si, mas o uso cada vez mais frequente de expressão pronominal *versus* o uso cada vez menos frequente de não expressão. Nesses termos é possível associar a não expressão pronominal à noção de variante conservadora e, por extensão, à norma-padrão.

A conexão entre não expressão pronominal e norma-padrão pode também ser aventada do fato de que, como lembra Faraco (2008, p. 80), no Brasil, no século XIX, a elite letrada conservadora empenhou-se em fixar, como norma-padrão, “certo modelo lusitano de escrita, praticado por alguns escritores portugueses do romantismo”. Considerando que, no português europeu, no século XIX (e mesmo posteriormente – cf. Duarte (1995)) predomina a não expressão pronominal de argumento-sujeito, é plausível inferir que essa é a variante (preferencialmente) incorporada pela norma-padrão brasileira do período.

Ademais, no Brasil, em gramáticas e manuais de redação do decorrer do século XX (e mesmo mais atuais), é comum figurar a recomendação (ou a prescrição) de emprego de sujeito oculto (elíptico, desinencial), em vez do uso de pronome sujeito, sobretudo na modalidade escrita formal – preferência que, de fato, se vê até hoje, por exemplo, em gêneros acadêmicos

¹¹ No que tange particularmente ao século XIX, ver, por exemplo, Faraco (2008, p. 81), acerca da “atitude excessivamente conservadora” da elite letrada brasileira responsável pela constituição da norma-padrão do país.

escritos, como o artigo científico. Em grande medida, gramáticas e manuais com esse tipo de orientação firmam-se na norma-padrão oitocentista; como reporta Pagotto (2013, p. 31), o Brasil do século XIX “é reconhecido como o período em que se fixa a [...] norma padrão recomendada pelos manuais de gramática atuais”.¹² Ou seja, a instrução, por parte de tais materiais, de emprego de sujeito oculto e a ancoragem dessas obras na norma oitocentista são, enfim, indicativos de conexão entre não expressão pronominal e norma-padrão no século XIX.

Por sua vez, a atribuição, pelos usuários da língua, de prestígio social à norma-padrão pode ser presumida a partir da relação que geralmente existe entre a norma-padrão e a norma culta.¹³ Em uma sociedade, esta última costuma ser caracterizada por prestígio social, sendo denominada, por vezes, justamente de *norma de prestígio*, como faz e recomenda Bagno (2003). Também Faraco (2008, p. 73) destaca o caráter de prestígio da norma culta, afirmando que seu “vínculo com os usos monitorados e com as práticas da cultura escrita leva os falantes a lhe atribuir um valor social positivo, a recobri-la com uma capa de prestígio social”. Ao mesmo tempo, ainda conforme explica Faraco (2008), para se estabelecer a norma-padrão em um país, normalmente (embora não sempre) toma-se como base a(s) norma(s) culta(s).¹⁴ Isso, de fato, ocorre no Brasil do século XIX, onde a norma-padrão é baseada em variedade culta, no caso em variedade culta do português europeu (conforme mencionado), como parte, segundo o autor, de uma tentativa da elite letrada brasileira de europeização do país. Portanto, estando a norma-padrão, no século XIX, baseada em variedade culta, é razoável supor que esta tenha concorrido para conferir prestígio àquela.¹⁵

No mesmo sentido, o caráter de prestígio da norma-padrão oitocentista pode ser cogitado considerando-se o fato de tratar-se de forma cultivada no ensino, tomada como uso correto da língua, a ser ensinado e aprendido. A esse respeito, cabe retomar as palavras de Votre (2012), para quem:

¹² Na mesma direção, Coelho et al. (2015, p. 145) afirmam que “a norma padrão do século XIX ainda está presente em boa parte de nossas gramáticas escolares atuais”.

¹³ A título de explicitação da distinção entre *norma-padrão* e *norma culta*, cabe lembrar o que explica Faraco (2008, p. 75). Para ele, “se a norma culta/comum/*standard* é a variedade que os letrados usam correntemente em suas práticas mais monitoradas de fala e escrita, a norma-padrão não é propriamente uma variedade da língua, mas [...] um construto sócio-histórico que serve de referência para estimular um processo de uniformização”. Ainda segundo o autor, “enquanto a norma culta/comum/*standard* é a expressão viva de certos segmentos sociais em determinadas situações, a norma-padrão é uma codificação relativamente abstrata, uma baliza extraída do uso real para servir de referência [...] a projetos políticos de uniformização linguística”.

¹⁴ Como exemplo de caso em que a norma-padrão não é baseada na norma culta, Faraco (2008) cita o que ocorreu com o tok pisin na Papua Nova Guiné, onde foram preferidas, como referência padronizadora, variedades rurais, o que, segundo o autor, dá uma direção inusitada ao processo de constituição da norma-padrão, se o consideramos pelo viés da experiência dos continentes europeu e americano.

¹⁵ Faraco (2008, p. 79) refere-se às formas da norma-padrão como “as formas cultuadas na [...] norma-padrão”, o que indica justamente o prestígio social a ela conferido.

as formas socialmente prestigiadas são semente e fruto da literatura oficial, que *as transforma em língua padrão*. Estão reguladas e codificadas nas gramática normativas, em que adquirem o estatuto de formas corretas, a serem ensinadas, aprendidas e internalizadas através de longo processo escolar (Votre, 2012, p. 51-52, grifos nossos).

Assim, por esse caminho (digamos *indireto*) de reconhecimento de relação entre não expressão pronominal de argumento-sujeito e (maior) prestígio (que passa pela relação entre não expressão e norma-padrão e entre norma-padrão e prestígio), e considerando a tendência de correlação entre maior escolaridade e variante de prestígio, parece-nos plausível, de fato, a explicação esboçada acima para os dados expostos na Tabela 1: no século XIX, a instrução avançada exibiria índice maior de não expressão pronominal e a instrução básica, índice menor porque, no período, a não expressão seria a variante de (maior) prestígio social.¹⁶

Trata-se de explicação que, enfim, aponta para a pertinência de se reconhecer possível correlação entre instrução mais alta e índices maiores de não expressão pronominal (e instrução mais baixa e índices menores de não expressão), o que significa, em outros termos, admitir a existência, no século XIX, de possível condicionamento social (no caso, por parte da variável instrução), sobre a (não) expressão pronominal de argumento-sujeito.

Também quanto à variável fase da vida, nossa avaliação é a de que, a partir dos dados, é pertinente reconhecer certo condicionamento social, no século XIX, sobre a (não) expressão de pronome sujeito. A Tabela 2 reúne os dados do cruzamento entre (não) expressão de sujeito e a referida variável:

Tabela 2: Cruzamento entre (não) expressão de sujeito e fase da vida.

(Não) expressão de sujeito	Não expressão % (N)	Expressão % (N)	Total % (N)
Fase da vida			
Fase 1 (criança e jovem; 0-15 anos)	61,5% (8)	38,5% (5)	100% (13)
Fase 2 (adulto; 16-50 anos)	78,2% (179)	21,8% (50)	100% (229)
Fase 3	68,9% (91)	31,1% (41)	100% (132)

¹⁶ Cabe ressaltar que a admissão de que possa haver associação entre não expressão pronominal e (maior) prestígio não significa reconhecer que a expressão pronominal seja uma variante estigmatizada; ela pode apresentar, na verdade, um caráter “neutro”, em termos de prestígio social, ou um caráter de menor prestígio que a não expressão pronominal. De fato, Duarte (1995) (embora se referindo a dados do século XX) avalia que, entre sujeito nulo e pleno, não haveria variante estigmatizada. Nesse sentido, nossa visão sobre uma possível associação, no século XIX, entre não expressão de pronome sujeito e prestígio pressupõe que não expressão e expressão de sujeito possam se opor em termos de maior *versus* menor (ou nenhum) prestígio, não em termos de prestígio *versus* estigma social.

(idoso/velho; 51 anos ou mais)			
			Valor p: 0.17 (> 0.05)

Fonte: Elaboração própria.

Nos dados da tabela, destaca-se a fase 2, em relação às outras duas, por preservar, de forma mais acentuada, a opção pela não expressão de sujeito. Aliás, a fase 2, com 78,2% de não expressão, supera os índices de não expressão das demais fases da vida de forma mais expressiva do que a diferença entre as faixas de instrução da variável anteriormente discutida, superando a fase 1 (61,5%) em 16,7 pontos percentuais. Já a fase 1 (principalmente) e a fase 3 mostram-se mais avançadas no emprego da expressão de sujeito. Inclusive, a fase 1, ao exibir índice próximo a 40% de expressão, parece despontar na direção de um percurso de aumento diacrônico da frequência de expressão de pronome sujeito, enquanto a fase 2 exibe pouco mais de 20%, taxa situada no patamar dos valores gerais de não expressão que ainda caracterizam o século XIX – valores de 20% (em 1845) e 23% (em 1882) nos dados de Barbosa, Duarte e Kato (2001) por exemplo.¹⁷

Desse modo, pode-se admitir a existência, no século XIX, de condicionamento da variável fase da vida sobre a (não) expressão pronominal de argumento-sujeito, o que pode ser colocado nos seguintes termos: a fase 2 está correlacionada a índices mais altos de não expressão pronominal de argumento-sujeito, enquanto as fases 1 e 3 correlacionam-se a índices menores de não expressão pronominal; ou seja, a fase 2 mantém, mais que as outras duas fases, a não expressão de pronome sujeito.

A apuração de que a fase 2, no século XIX, conserva mais a não expressão de sujeito vai ao encontro de resultados aferidos em trabalhos como o de Freitag (2005), o que se pode concluir assumindo-se novamente a não expressão como variante prototípica da norma-padrão no período. A autora argumenta haver uma tendência de correlação entre falantes de idade adulta (em oposição a crianças, adolescentes e idosos) e preferência por variantes linguísticas mais típicas da norma-padrão, porque indivíduos dessa faixa etária sofrem forte pressão do mercado de trabalho, a qual já não seria exercida sobre indivíduos das demais idades. Para ela:

passada a fase de rebeldia [adolescência], na idade adulta o estilo de fala do indivíduo tende a se acomodar [...]. Mas não sem antes sofrer a forte pressão exercida pelo mercado de trabalho no seu uso linguístico. É quando a norma linguística se impõe (ou melhor, é imposta, ainda que escamoteadamente) ao uso linguístico do indivíduo (Freitag, 2005, p. 114).

¹⁷ Convém, é verdade, relativizar os dados da fase 1, pelo baixo quantitativo de ocorrências (13 casos no total), decorrente da carência de personagens representativos dessa fase da vida nas peças componentes do *corpus*.

Freitag (2005) sustenta empiricamente a tendência, discutindo resultados de pesquisa sociolinguística (Coan, 1997) que analisa o emprego de pretérito perfeito e pretérito mais-que-perfeito composto na expressão de anterioridade temporal a um ponto de referência passado e segundo a qual a faixa etária de 25 a 49 anos (adultos) tem comportamento diferenciado, ao apresentar concentração mais elevada da segunda opção. Tomando tal opção como mais típica da norma, Freitag (2005) explica o resultado da seguinte forma:

Pode-se associar esse comportamento de uso ao fato de que os informantes dessa faixa etária são aqueles que estão atuando no mercado de trabalho e, devido às pressões sociais, privilegiam a norma, ainda que o uso de pretérito perfeito na expressão da anterioridade a um ponto de referência passado não seja estigmatizado. Assim que saem do mercado de trabalho, a situação parece voltar à linha estável traçada [menor emprego de pretérito mais-que-perfeito composto]. É somente na faixa etária de 25 a 49 anos que a concentração de ocorrências aumenta (Freitag, 2005, p. 119).

Essa generalização destacada por Freitag (2005) de correlação entre idade adulta e norma-padrão é uma tendência à qual se alinha nossa constatação de que, no século XIX, a segunda fase da vida (que envolve a idade adulta) conserva (mais que as outras fases) a não expressão de sujeito (a qual pode ser entendida justamente como variante prototípica da norma-padrão), o que, enfim, reforça a pertinência da constatação e de seu entendimento como uma forma de condicionamento da variável fase da vida sobre a (não) expressão de argumento-sujeito no período.¹⁸

Já no que concerne às outras duas variáveis sociais, os dados apurados em nossa pesquisa não possibilitam identificar algum tipo de influência dessas variáveis sobre a (não) expressão de sujeito. A Tabela 3 apresenta resultados relativos à variável classe social:

Tabela 3: Cruzamento entre (não) expressão de sujeito e classe social.¹⁹

(Não) expressão de sujeito \ Classe social	Não expressão % (N)	Expressão % (N)	Total % (N)
Classe 1 (alta burguesia, alta sociedade, dominantes)	74,0% (125)	26,0% (44)	100% (169)
Classe 2	75,4% (135)	24,6% (44)	100% (179)

¹⁸ É necessário ressaltar que, ao trazer comentários sobre o comportamento de adultos (Freitag, 2005) em uma sincronia do século XX, não pressupomos que o século XIX tenha, necessariamente, a mesma organização social, educacional e cultural. Sabe-se que não é assim em termos históricos e sociais. A comparação é aqui esboçada para levantar reflexões e hipóteses a serem averiguadas com mais detalhes na continuação do trabalho.

¹⁹ A Tabela 3 não contém dados relativos à classe 4 porque o *corpus* não apresentou esse tipo de personagem.

(classes médias, classe emergente, intermediária)			
Classe 3 (proletariado, camponeses, classe subalterna)	80,0% (16)	20,0% (4)	100% (20)
Valor p: 0.85 (> 0.05)			

Fonte: Elaboração própria.

Como se vê, as classes 1 (classe privilegiada) e 2 (classe intermediária) apresentam resultados muito próximos entre si, enquanto a classe 3 (classe não privilegiada) mostra certa diferença, preservando, um pouco mais que as outras duas, a opção pela não expressão pronominal de argumento-sujeito; em outras palavras, as classes 1 e 2 exibem taxas ligeiramente maiores de expressão de sujeito.²⁰

A princípio, poder-se-ia visualizar uma possível influência da variável classe social sobre a (não) expressão pronominal: a classe 3 conservaria, mais que as outras, a opção pela não expressão de sujeito. No entanto, alguns fatores (inclusive de cunho metodológico) inibem, a nosso ver, o reconhecimento dessa correlação. Em primeiro lugar, é relativamente baixo o quantitativo de dados disponíveis acerca da classe 3 (20 ocorrências no total). É também pequena a diferença percentual entre os resultados das três classes (menor, inclusive, que a diferença entre as faixas de cada uma das duas variáveis sociais anteriores); e a classe 3, que se destaca com maior diferença percentual das demais, é justamente a classe sobre a qual o quantitativo de dados é baixo. Ainda, comparando entre si apenas as classes 1 e 2, sobre as quais há quantitativo mais robusto de dados, os percentuais de não expressão (bem como os de expressão) são muito próximos (74% e 75,4%), sugerindo ausência de condicionamento da variável classe social sobre a (não) expressão de sujeito.

Ademais, é curioso notar que, diferentemente do que ocorre com as duas variáveis sociais anteriores, o resultado sobre a variável classe social não se alinha a tendência sociolinguística geral. Conforme lembram Coelho et al (2015), os trabalhos de Labov e seu grupo de pesquisa sobre o inglês nova-iorquino apontam que o grupo social mais privilegiado tende a optar pelo uso de variante padrão, enquanto o grupo menos privilegiado favorece o uso de variantes não padrão da língua. Levando em conta essa tendência, e considerando novamente a não expressão como variante típica da norma-padrão no século XIX, possivelmente variante de maior prestígio social, seria esperado que as classes mais

²⁰ Também aqui, cabe relativizar os resultados referentes à classe 3, pelo quantitativo relativamente baixo de casos (20 ocorrências).

privilegiadas preservassem mais essa variante do que as classes menos privilegiadas, o que não se verifica em nossos dados.

Diante disso, entendemos que, por ora, nossos dados não permitem uma avaliação mais acurada relativamente à questão do condicionamento da variável classe social sobre a (não) expressão de pronome sujeito, questão que, na verdade, requer mais exploração, especialmente com ampliação do conjunto de dados acerca da classe 3.

Finalmente, a Tabela 4 contém resultados atinentes à variável gênero/sexo:

Tabela 4: Cruzamento entre (não) expressão de sujeito e gênero/sexo.

(Não) expressão de sujeito \ Gênero/sexo	Não expressão % (N)	Expressão % (N)	Total % (N)
Feminino	74,2% (98)	25,8% (34)	100% (132)
Masculino	74,4% (236)	25,6% (81)	100% (317)
Valor p: 1.0 (> 0.05)			

Fonte: Elaboração própria.

A tabela mostra resultados praticamente iguais para falantes femininos e masculinos, o que sugere não haver condicionamento da variável gênero/sexo sobre a (não) expressão pronominal de argumento-sujeito no século XIX.

Paiva (2012) relata que diversos estudos de orientação sociovariacionista puderam corroborar a existência de um padrão bastante regular no que diz respeito à variável gênero/sexo: falantes do gênero/sexo feminino demonstram maior preferência por variantes linguísticas mais prestigiadas socialmente. Considerando essa tendência, e admitindo, uma vez mais, a possibilidade de que, no século XIX, a não expressão pronominal (comparativamente à expressão pronominal) seja uma variante com (maior) prestígio, seria de esperar que o gênero/sexo feminino exibisse índice de não expressão maior que o apurado para o gênero/sexo masculino, refletindo a referida tendência, o que, todavia, não ocorre em nossos dados.

Paiva (2012) esclarece que a análise da correlação entre gênero/sexo e variação linguística deve levar em conta não só o prestígio atribuído pela comunidade às variantes linguísticas, mas também a forma de organização social da comunidade. Segundo a autora, a consistência do padrão que correlaciona gênero/sexo feminino e variantes mais prestigiadas advém de análises de variações em comunidades de fala ocidentais. Ela explica que esse padrão pode ser revertido em comunidades marcadas por outros valores culturais e cita, como exemplo, a variação entre oclusiva uvular, oclusiva glotal e oclusiva velar em árabe: em

diferentes comunidades muçulmanas, a variante uvular, forma de prestígio baseada no árabe literário, predomina entre os homens, enquanto as mulheres estão associadas a maior uso das formas não prestigiadas. De forma similar, diferenças na forma de organização da sociedade podem ser um fator interveniente na relação entre gênero/sexo e emprego de variante mais prestigiada, nos dados do século XIX. De fato, a correlação apontada por Paiva (2012) entre gênero/sexo feminino e uso de formas prestigiadas é apurada em trabalhos realizados no avançar da segunda metade do século XX, época em que, no Brasil, a emancipação social da mulher adquire avanços em relação ao século XIX. É possível que a condição social da mulher, no século XIX, possa comprometer (ou afetar de alguma forma) a correlação entre gênero/sexo feminino e variantes de prestígio, concorrendo para explicar por que, em nossos dados oitocentistas, essa correlação não se verifica.²¹

Conclusão

O objetivo deste artigo foi apresentar e discutir dados do PB do século XIX, para averiguar se, nesse período, haveria influência de fatores sociais sobre a variação entre não expressão e expressão pronominal de argumento-sujeito. Foram analisados dados do cruzamento entre (não) expressão de pronome sujeito e quatro variáveis sociais: instrução, fase da vida, classe social e gênero/sexo. Os resultados sugerem que as duas primeiras variáveis exercem influência sobre a (não) expressão de sujeito: graus mais altos de instrução conservam índices mais altos de não expressão de sujeito, enquanto graus mais baixos de instrução correlacionam-se a índices relativamente menores de não expressão; a segunda fase da vida preserva índices mais altos de não expressão de sujeito, enquanto a primeira e a terceira fase correlacionam-se a índices menores de não expressão de sujeito. Por outro lado, os dados não permitem uma apuração mais assertiva sobre um possível papel condicionador da variável classe social, indicando ainda que a variável gênero/sexo não exerce influência sobre a (não) expressão de pronome sujeito.

O propósito mais amplo do trabalho foi oferecer dados que, futuramente reunidos a resultados de outros estudos, possam contribuir para verificar se, e de que modo, o processo diacrônico de aumento na frequência de uso de pronome sujeito é influenciado por fatores sociais. A continuidade de nossa pesquisa deverá ser orientada justamente nessa direção,

²¹ Cabe reconhecer que é difícil falar de diferenças linguísticas de sexo/gênero no século XIX, já que há poucos estudos ainda sobre a questão. A esse respeito, ver Lopes (2005).

envolvendo, entre outras questões, aplicar o tipo de análise praticada aqui a materiais de sincronias posteriores ao século XIX.

No que concerne, por exemplo, à variável instrução, tal exame poderá revelar que, com o decorrer do tempo, todas as faixas de instrução vão passando a empregar, com mais frequência, a expressão de pronome sujeito (verificando-se, em todas elas, a referida mudança diacrônica), porém, com a instrução avançada (a exemplo do que apuramos para o século XIX) mantendo sempre um índice de não expressão de sujeito relativamente maior que as outras faixas de instrução. Seria um cenário em que o aumento diacrônico na frequência de expressão de sujeito seria influenciado pela variável instrução, de um modo continuamente similar ao que apuramos para o século XIX. Ou, com o passar do tempo, a instrução avançada pode manter o índice alto de não expressão de sujeito do século XIX (concentrando-se a mudança diacrônica nas faixas de menor instrução), o que significaria um impacto ainda mais intenso e evidente da instrução sobre a mudança em questão. Ou ainda, em sincronias posteriores ao século XIX, as diferentes faixas de instrução, comparadas entre si, podem exibir percentuais (quase) iguais de não expressão de sujeito (todas as faixas similarmente diminuindo a taxa de não expressão e aumentando o índice de expressão de sujeito), o que revelaria que, diacronicamente, a variável instrução não exerceria influência sobre a (não) expressão de sujeito.

A análise dessas possibilidades constitui, enfim, a meta da continuidade deste artigo e justamente o propósito mais amplo para o qual ele procura contribuir.

Referências

AUER, A; PEERSMAN, C; PICKL, S; RUTTEN, G.; VOSTERS, R. Historical Sociolinguistics: the field and its future. *Journal of Historical Sociolinguistics*, v. 1, n. 1, p. 1-12, 2015.

BAGNO, M. Norma linguística e preconceito social: questões de terminologia. *Veredas: Revista de Estudos Linguísticos*, v. 5, n. 2, p. 71-83, 2003.

BARBOSA, P.; DUARTE, M. E. L.; KATO, M. A. A distribuição do sujeito nulo no português europeu e no português brasileiro. *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, 2001. p. 539-550.

BERGS, A. The uniformitarian principle and the risk of anachronisms in language and social history. In: HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C (eds.). *The Handbook of Historical Sociolinguistics*. London: Balckwell, 2012. p. 80-98.

COAN, M. *Anterioridade a um ponto de referência passado: pretérito (mais-que) perfeito*. 1997. 183f. Dissertação (Mestrado em Linguística). Universidade Federal de Santa Catarina, Florianópolis, 1997.

COELHO, I., L. et al. *Para conhecer Sociolinguística*. São Paulo: Contexto, 2015.

COELHO, I. L. et al. *Sociolinguística*. Florianópolis: LLV/CCE/UFSC, 2012.

DUARTE, M. E. L. *A perda do princípio “evite pronome” no português brasileiro*. 1995. 151f. Tese (Doutorado em Linguística), Universidade Estadual de Campinas, Campinas, 1995.

FARACO, C. A. *Norma culta brasileira: desatando alguns nós*. São Paulo: Parábola, 2008.

FREITAG, R. M. K. Idade: uma variável sociolinguística complexa. *Língua e Letras*, v. 6, n. 2, p. 105-121, 2005.

GUERRA, A. R. Variação diacrônica do grau de transparência linguística do português brasileiro: o papel do aumento percentual da expressão pronominal do argumento-sujeito. *Cadernos de Estudos Linguísticos*, v. 63, p. 1-16, e021011, 2021. DOI: <https://doi.org/10.20396/cel.v63i00.8657215>.

GUERRA, A. R. Descrição diacrônica da expressão pronominal do argumento-sujeito no português brasileiro. *Revista do GEL*, v. 19, n. 3, p. 203-230, 2022a. DOI: <http://dx.doi.org/10.21165/gel.v19i3.3449>.

GUERRA, A. R. Variação diacrônica do grau de transparência linguística do português brasileiro: o impacto do aumento percentual do uso de “a gente” em lugar de “nós”, 2025 (submetido).

HENGEVELD, K.; MACKENZIE, J. L. *Functional Discourse Grammar*. New York: Oxford University Press, 2008.

HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C. Introduction. In: HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C. (eds.). *The Handbook of Historical Sociolinguistics*. London: Balckwell, 2012, p. 1-8.

LABOV, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LEUFKENS, S. *Transparency in language: a typological study*. Utrecht: LOT, 2015.

LOPES, C. R. S. (org.). *A norma brasileira em construção: fatos linguísticos em cartas pessoais do século XIX*. Rio de Janeiro: UFRJ, FAPERJ, 2005.

NEVALAINEN, T.; RAUMOLIN-BRUNBERG, H. Historical Sociolinguistics: origins, motivations, and paradigms. In: HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C. (eds.). *The Handbook of Historical Sociolinguistics*. London: Balckwell, 2012. p. 22-40.

- OLBERTZ, H. A perda do sujeito nulo no português do Brasil: uma abordagem discursivo-funcional. In: CASSEB-GALVÃO, V. C.; ROSA, G. L.; SILVA, C. A.; BERTOQUE, L. A. P. *O sujeito gramatical no português brasileiro: expressão, concordância, ergatividade e afetamento*. São Paulo: Parábola, 2020. p. 15-45.
- OLIVEIRA, T. P. Subject expression in Brazilian Portuguese. In: KEIZER, E.; OLBERTZ, H. (org.). *Recent Developments in Functional Discourse Grammar*. Amsterdam: John Benjamins Publishing, 2018. p. 208-232.
- OUSHIRO, L. Tratamento de dados com o R para análises sociolinguísticas. In: FREITAG, R. M. K. (org.). *Metodologia de coleta e manipulação de dados em Sociolinguística*. São Paulo: Editora Edgard Blücher, 2014. p. 134-177.
- PAGOTTO, E. G. A norma das constituições e a constituição da norma no século XIX. *Revista Letras. Linguagem & Preconceito*, 2013. p. 31-50.
- PAIVA, M. C. A variável gênero/sexo. In: MOLLICA, M. C., BRAGA, M. L. *Introdução à Sociolinguística: o tratamento da variação*. 4ª ed. São Paulo: Contexto, 2012. p. 33-42.
- ROMAINE, S. 1982. *Socio-historical linguistics*. Its status and methodology. Cambridge: Cambridge University Press.
- ROMAINE, S. Historical sociolinguistics/historische soziolinguistik. In: AMMON, U; DITTMAR, N.; MATTHEIER, K. J.; TRUDGILL, P. (eds.). *Sociolinguistics: an international handbook of the science of language and society*. Berlin; New York: De Gruyter, 2005. p. 1696-1703.
- TARALLO, F. Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX. In: ROBERTS, I.; KATO, M. A. (org.). *Português brasileiro: uma viagem diacrônica*. Campinas: Ed. da Unicamp, 1993. p. 69-105.
- TAVARES, M. A. Uma análise sociolinguística comparativa de conectores sequenciadores: foco na escolaridade. *Linguística*, v. 32, n. 2, p. 113-128, 2016. DOI: 10.5935/2079-312X.20160020.
- SCHWARCZ, L. M (org.). *História do Brasil Nação: 1808-2010*. Volumes 1 a 5. Rio de Janeiro: Objetiva, 2011.
- VOTRE, S. J. Relevância da variável escolaridade. In: MOLLICA, M. C.; BRAGA, M. L. (org.). *Introdução à Sociolinguística: o tratamento da variação*. 4ª ed. São Paulo: Contexto, 2012, p. 51-57.
- WEINREICH, U.; LABOV, W.; HERZOG, M. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola, 2006 [1968].

ISSN: 1984-4921

DOI: <http://dx.doi.org/10.18226/19844921.v17.n38.05>

Submetido em: 15/05/2025

Aprovado em: 11/08/2025