

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS**

ERENILDO QUEIROZ DE SOUZA

**AS FORMAS DE REFERÊNCIA AO SUJEITO DE 2ª PESSOA DO SINGULAR
EM MISSIVAS MINEIRAS DOS SÉCULOS XIX E XX:
UMA ANÁLISE LINGÜÍSTICO-SOCIAL.**

**BELO HORIZONTE
FACULDADE DE LETRAS DA UFMG
2021**

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS**

ERENILDO QUEIROZ DE SOUZA

**AS FORMAS DE REFERÊNCIA AO SUJEITO DE 2ª PESSOA DO SINGULAR
EM MISSIVAS MINEIRAS DOS SÉCULOS XIX E XX:
UMA ANÁLISE LINGUÍSTICO-SOCIAL.**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Estudos Linguísticos. Área de Concentração: Linguística Teórica e Descritiva. Linha de Pesquisa: 1A - Estudo da Variação e Mudança Linguística.

Orientadora: Prof^a. Dr^a. Márcia Cristina de Brito Rumeu.

**BELO HORIZONTE
FACULDADE DE LETRAS DA UFMG
2021**

Ficha catalográfica elaborada pelo Bibliotecário Israel Jose da Silva – CRB/6-2128

S729f

Souza, Erenildo Queiroz de.

As formas de referência ao sujeito de 2ª pessoa do singular em missivas mineiras dos séculos XIX e XX [manuscrito] : uma análise linguístico-social / Erenildo Queiroz de Souza. – 2021.

198 f., enc. : il., grafs., tabs., p&b., color.

Orientadora: Márcia Cristina de Brito Rumeu.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudo da Variação e Mudança Linguística.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras

Bibliografia: p. 121-127.

Apêndices: p. 128-198.

1. Língua portuguesa – Variação – Teses. 2. Mudanças linguísticas – Teses. 3. Sociolinguística – Teses. 4. Língua portuguesa – Pronomes – Teses. 5. Língua portuguesa – Sintaxe – Teses. I. Rumeu, Márcia Cristina de Brito. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD : 469.798



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS

FOLHA DE APROVAÇÃO

AS FORMAS DE REFERÊNCIA AO SUJEITO DE 2ª PESSOA DO SINGULAR EM MISSIVAS
MINEIRAS DOS SÉCULOS XIX E XX: UMA ANÁLISE LINGÜÍSTICO-SOCIAL

ERENILDO QUEIROZ DE SOUZA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudo da Variação e Mudança Linguística.

Aprovada em 22 de fevereiro de 2021, pela banca constituída pelos membros:

Prof(a). Márcia Cristina de Brito Rumeu - Orientadora
UFMG

Prof(a). Thiago Laurentino de Oliveira
UFRJ

Prof(a). Heliana Ribeiro de Mello
UFMG

Belo Horizonte, 22 de fevereiro de 2021.



Documento assinado eletronicamente por **Thiago Laurentino de Oliveira, Usuário Externo**, em 22/02/2021, às 17:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Heliana Ribeiro de Mello, Professora do Magistério Superior**, em 22/02/2021, às 18:34, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcia Cristina de Brito Rumeu, Professora do Magistério Superior**, em 22/02/2021, às 22:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0521263** e o código CRC **1030F765**.

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS

Dissertação de Mestrado intitulada **As formas de referência ao sujeito de 2ª pessoa do singular em missivas mineiras dos séculos XIX e XX: uma análise linguístico-social** apresentada por Erenildo Queiroz de Souza, em 26 de fevereiro de 2021, à Banca Examinadora constituída pelos seguintes Membros:

Prof.^a Dr.^a Márcia Cristina de Brito Rumeu – Orientadora
Universidade Federal de Minas Gerais – UFMG

Prof. Dr. Thiago Laurentino de Oliveira – Examinador
Universidade Federal do Rio de Janeiro – UFRJ

Prof.^a Dr.^a Heliana Ribeiro de Mello – Examinadora
Universidade Federal de Minas Gerais – UFMG

Prof.^a Dr.^a Maria Cândida Trindade Costa de Seabra – Examinadora Suplente
Universidade Federal de Minas Gerais – UFMG

AGRADECIMENTOS

Às professoras Maria José e Almira Lins, por marcarem, respectivamente, minha infância nas séries iniciais na Escola Parque e minha juventude na Faculdade de Ciências Humanas da UEPB.

Aos professores da Faculdade de Letras e do Programa de Pós-Graduação em Estudos Linguísticos da UFMG, pela qualidade do ensino.

À minha orientadora Prof.^a Dr.^a Márcia Rumeu, pelo cuidado e dedicação incansáveis. Sempre me lembrarei com carinho das videoconferências em razão da pandemia do Covid-19. Com as novas tecnologias da informação e da comunicação, o tempo não parou para os nossos estudos.

Aos professores da banca, Prof. Dr. Thiago Laurentino de Oliveira – Examinador (UFRJ), Prof.^a Dr.^a Heliana Ribeiro de Mello – Examinadora (UFMG), Prof.^a Dr.^a Maria Cândida Trindade Costa de Seabra – Examinadora Suplente (UFMG) e Prof.^a Dr.^a Márcia Cristina de Brito Rumeu – Orientadora (UFMG), por terem aceitado o convite para a composição desta Banca em uma época tão complexa e pelas excelentes contribuições para o aprimoramento desta Dissertação.

Não poderia deixar de fazer um agradecimento especial à família Guimarães e à família Vieira da Silva por terem cedido o acesso às cartas pessoais de seus inestimáveis queridos entes mineiros cujas missivas foram analisadas nesta pesquisa. Agradeço também especialmente aos Arquivos Públicos (AEM, AN, APCBH, APM, IHGMG, MAB) não só por guardarem cuidadosamente os manuscritos históricos, mas também por me terem permitido acessá-los.

Ao Exército Brasileiro, pela fantástica experiência profissional que me possibilitou e, em especial, aos meus comandantes e aos meus chefes, no Colégio Militar de Belo Horizonte, pela confiança no meu trabalho e por sempre me incentivarem ao aprimoramento profissional.

À minha amada Família Queiroz, especialmente aos meus pais, Erenaldo e Maria Lúcia, pela educação para o bem, e aos meus irmãos Ezequias, Edison, Edmilson, Edvaldo, Érika e Elayne, pela união que sempre permeou nossas vidas.

À minha filha Rebeca, meu melhor presente, e ao meu genro, Arthur, pela parceria.

À minha mulher Karla, pelo amor, compreensão e cumplicidade nesses anos.

Em especial, ao meu Deus, pelo Filho, e pelas muitas misericórdias.

Muito obrigado.

SOUZA, Erenildo Queiroz de. **As formas de referência ao sujeito de 2ª pessoa do singular em missivas mineiras dos séculos XIX e XX: uma análise linguístico-social**. 2021. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

RESUMO

O objetivo desta investigação é descrever as formas de referência ao sujeito de 2SG (*vossa mercê, você, tu*) correlacionadas às relações sociais que as embasam na produção escrita mineira dos séculos XIX e XX, à luz da Teoria do Poder e da Solidariedade (BROWN & GILMAN, 1960), dos pressupostos teórico-metodológicos da Sociolinguística Histórica (HERNÁNDEZ-CAMPOY & CONDE SILVESTRE, 2012; HERNÁNDEZ-CAMPOY & SHILLING, 2012) e da Sociolinguística Laboviana (WEINREICH, LABOV & HERZOG, 2008 [1968]; LABOV, 1994). Parte-se da hipótese de que o *você* seria mais produtivo do que o *tu*, cf. Lopes & Rumeu (2015), Rumeu, Cruz & Cardoso (2018). Busca-se saber em que nível de variação o *tu* e o *você* estariam na escrita de sincronias passadas e como seriam expressas as relações de Poder e de Solidariedade. O ponto de partida desta pesquisa é o entendimento de que a função sintática de *sujeito* representa a principal porta de acesso assumida por pronomes gramaticalizados como o *você* e o *a gente*, ao se inserirem no sistema pronominal do Português Brasileiro, conforme já observado em estudos diacrônicos (RUMEU, 2013; SOUZA, 2012; LOPES & CAVALCANTE, 2011; LOPES & VIANNA, 2012). Esta pesquisa toma o presente como base para a reconstrução do caminho percorrido pela mudança linguística no passado (LABOV, 1994) embasada em 332 cartas pessoais (amorosa, familiar e de amizade) e em 26 cartas de comércio, disponíveis em acervos públicos e pessoais. Os resultados das formas *vossa mercê, você* e *tu* estão distribuídos em relação à *expressão da forma de referência ao sujeito de 2SG (nulo/pleno)*, à *pessoa verbal (concordância)*, ao *paralelismo formal e semântico*, ao *gênero da carta*, ao *sexo do informante*, à *faixa etária*, à *relação interpessoal*, às *relações sociais* e ao *período das cartas*. As formas *tu* e *você* foram submetidas ao RBrul para a geração dos índices percentuais e probabilísticos da regra variável em discussão: *tu versus você*. Os resultados surpreendem em relação à hipótese principal, tendo em vista que o *tu* se mostrou mais produtivo do que o *você*, ainda que em uma disputa acirrada. As formas verbais de 3SG (concordância), o *você-sujeito*, as formas clíticas de 3SG, as possessivas de 3SG e as formas de *você não-sujeito* mostram-se propulsoras do *você* nas missivas mineiras. Considerando a interação entre os interlocutores, as formas *tu* e *você* predominam nas relações simétricas, o que sugere o encaminhamento da sociedade mineira pelos domínios da Solidariedade (BROWN & GILMAN, 1960), ao passo que o *vossa mercê* mostra-se restrito às relações assimétricas ascendentes (de inferior para superior), cf. Lopes & Rumeu (2015).

PALAVRAS-CHAVE: *variação tu/você, relações sociais, Sociolinguística Histórica.*

SOUZA, Erenildo Queiroz de. **As formas de referência ao sujeito de 2ª pessoa do singular em missivas mineiras dos séculos XIX e XX: uma análise linguístico-social**. 2021. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

ABSTRACT

The objective of this research is to describe the forms of reference to the 2SG subject (*vossa mercê, você, tu*) correlated to the social relations that underlie them in the handwritten letters of Minas Gerais in the 19th and 20th centuries, based on the Theory of Power and Solidarity (BROWN & GILMAN, 1960), the theoretical-methodological assumptions of historical Sociolinguistics (HERNÁNDEZ-CAMPOY & CONDE SILVESTRE, 2012; HERNÁNDEZ-CAMPOY & SHILLING, 2012) and Labovian Sociolinguistics (WEINREICH, LABOV & HERZOG, 2008 [1968]; LABOV, 1994). It is based on the hypothesis that *você* would be more productive than *tu*, according to Lopes & Rumeu (2015), Rumeu, Cruz & Cardoso (2018). One seeks to know at what level of variation *você* and *tu* would be in the writing of past synchronies and how the relationships of Power and Solidarity would be expressed. The starting point of this research is the understanding that the syntactic function of the subject represents the main gateway assumed by grammatical pronouns such as *você* and *a gente*, when inserted in the pronominal system of Brazilian Portuguese as already observed in diachronic studies (RUMEU, 2013; SOUZA, 2012; LOPES & CAVALCANTE, 2011; LOPES & VIANNA, 2012). This research takes the present as the basis for the reconstruction of the path traveled by linguistic change in the past (LABOV, 1994) based on 332 personal letters (loving, familiar and friendship letters) and 26 trade letters available in public and personal collections. The results of “vossa mercê” (“You”), “você” and “tu” (“You) are distributed in relation to the *expression of the form of reference to the subject of 2SG (null/full)*, the *verbal person (concordance)*, the *formal and semantic parallelism*, the *gender of the letter*, the *gender of the informant*, the *age*, the *interpersonal relationship*, the *social relations* and the *period* of the letters. The data were submitted to RBrul for the generation of the percentage and probabilistic indexes of the variable rule under discussion: *tu versus você*. The results achieved were surprising in relation to the main hypothesis, considering that *tu* as a subject proved to be more productive than *você* in the analyzed letters from Minas Gerais, even in a fierce dispute. The verbal forms of 3SG (concordance), the *você-subject* (“you-subject”), the clitic forms of 3SG, the possessive of 3SG and the forms of the *você* non-subject (“you non-subject”) show themselves to be propellers of the “você” in the letters of Minas Gerais. Considering the interaction between interlocutors, *tu* and *você* were more productive in symmetrical social relations. This fact suggests the forwarding of Mineira society through the domains of Solidarity, whereas “vossa mercê” is shown to be restricted to ascending asymmetrical relations (from bottom to top) according to Lopes & Rumeu (2015).

KEYWORDS: *variation tu/você, social relations, historical sociolinguistics.*

LISTA DE ABREVIATURA E SIGLAS

2PL -	Segunda pessoa do plural
2SG -	Segunda pessoa do singular
3SG -	Terceira pessoa do singular
AEM -	Acervo dos Escritores Mineiros
ALFAL -	Asociación de Lingüística y Filología de América Latina
AN -	Arquivo Nacional
AP -	Acervo pessoal
APCBH -	Arquivo Público da Cidade de Belo Horizonte
APM -	Arquivo Público Mineiro
CF(cf)	Conforme
IHGGMG -	Instituto Histórico e Geográfico de Minas Gerais
MAB -	Museu Abílio Barreto
NA -	Não aplicável
OCO -	Ocorrência(s)
PB -	Português brasileiro
PE -	Português europeu
PR -	Peso relativo

LISTAS DE GRÁFICOS, IMAGENS, QUADROS E TABELAS

Lista de Gráficos

Gráfico 1 - Distribuição das formas <i>você</i> e <i>tu</i> ao longo do tempo, cf. Souza (2012)	26
Gráfico 2 - A utilização do <i>você</i> nas relações sociais em três períodos diferentes	28
Gráfico 3 - As formas de referência ao sujeito 2SG em função das relações sociais no decorrer do tempo	99
Gráfico 4 - As formas <i>vossa mercê</i> , <i>você</i> e <i>tu</i> distribuídas no eixo do tempo (principalmente os séculos XIX e XX)	107

Lista de Imagens

Imagem 1 - Seis subsistemas dos pronomes de segunda pessoa <i>você</i> e <i>tu</i> no PB	23
Imagem 2 - Pesquisas realizadas sobre as variantes <i>Tu</i> e <i>Você</i> em Minas Gerais	36
Imagem 3 - Dado de <i>você</i> -sujeito em carta pessoal (amizade)	42
Imagem 4 - Dado de <i>vosmecê</i> em carta de amizade	43
Imagem 5 - Uso de <i>você</i> em carta pessoal	43
Imagem 6 - Evidência de dúvida do autor entre as formas <i>tu</i> e <i>você</i>	45
Imagem 7 - Ficha de identificação de carta pessoal	45
Imagem 8 - Assinatura do autor em carta de amizade (documento autógrafo)	46
Imagem 9 - Uso de abreviaturas como indício da habilidade do autor em sua expressão escrita	48
Imagem 10 - Bilhete de ACNP endereçado ao pai dela, o Barão de Cocais, escrito logo após o fechamento da carta do marido, BNP	51
Imagem 11 - Transcrição de bilhete de ACNP endereçado ao pai dela, o Barão de Cocais, escrito logo após o fechamento da carta	52
Imagem 12 - Carta de BNP ao sogro, o Barão de Cocais, fazendo referência à esposa e às filhas	53
Imagem 13 - Planilha Excel/Cartas pessoais mineiras	68

Imagem 14 - Dado de <i>tu nulo</i> em carta amorosa	71
Imagem 15 - Dado de <i>você</i> -sujeito em carta de amizade	72
Imagem 16 - Dado de <i>você</i> em carta familiar	73
Imagem 17 - Trecho de carta de comércio	73
Imagem 18 - Gênero carta pessoal e seus subgêneros	74
Imagem 19 - As cartas de comércio ao Barão de Cocais	75
Imagem 20 - Interface do RStudio – R versão 3.6.1	79
Imagem 21 - Carta de ST, com exemplo de uso de <i>você</i> com forma de 2SG em uma carta do século XX	87
Imagem 22 - Carta de PP, exemplo de uso de <i>você</i> com forma de 2SG	87
Imagem 23 - Carta de MIDV, exemplo de uso de <i>tu</i> com forma de 3SG	88
Imagem 24 - Carta de RT com exemplo de uso de <i>tu</i> com forma de 3SG	88

Lista de Quadros

Quadro 1 - Diferenças entre comunicação oral e escrita, cf. Conde Silvestre (2007, p. 43) ...	41
Quadro 2 - Amostras históricas selecionadas	50
Quadro 3 - Relação de missivas de comércio do Arquivo Nacional (AN)	56
Quadro 4 - Relação de missivas pessoais do Acervo dos Escritores Mineiros	57
Quadro 5 - Relação de missivas pessoais do Arquivo Público Mineiro (APM)	59
Quadro 6 - Relação de missivas pessoais do IHGMG	60
Quadro 7 - Relação de missivas pessoais do Museu Abílio Barreto	62
Quadro 8 - Relação de missivas pessoais do APCBH	62
Quadro 9 - Relação de missivas pessoais do Acervo Pessoal.....	64

Lista de Tabelas

Tabela 1 - Uso das formas de tratamento ao rei	16
Tabela 2 - A utilização da forma pronominal <i>você</i> em cada tipo de relação social em cartas cariocas (SOUZA, 2012)	27
Tabela 3 - As formas de referência ao sujeito de 2SG nas cartas mineiras analisadas	82
Tabela 4 - A expressão dos pronomes-sujeito de 2SG: nulo ou pleno	84
Tabela 5 - As formas de referência ao sujeito de 2SG e a pessoa verbal (<i>concordância</i>)	85
Tabela 6 - As formas de referência ao sujeito de 2SG e o paralelismo formal e semântico .	90
Tabela 7 - As formas de referência ao sujeito de 2SG correlacionadas ao sexo e à faixa etária dos escreventes	93
Tabela 8 - As formas de referência ao sujeito de 2SG correlacionadas às relações sociais simétricas e assimétricas (ascendentes e descendentes)	96
Tabela 9 - As formas <i>Vossa Mercê</i> , <i>Tu</i> e <i>Você</i> distribuídas pelas relações sociais ao longo do tempo	99
Tabela 10 - Distribuição geral das formas de referência ao sujeito de 2SG por carta no eixo do tempo (principalmente entre os séculos XIX e XX)	106
Tabela 11 - O gênero textual “carta”: os pronomes-sujeito de 2SG em relação às cartas pessoais e de comércio em cena	108
Tabela 12 - O efeito da pessoa verbal (<i>concordância</i>) sobre o <i>você-sujeito</i>	114
Tabela 13 - O efeito do paralelismo formal e sintático sobre o <i>você-sujeito</i>	114
Tabela 14 - O efeito da faixa etária sobre o <i>você-sujeito</i>	115
Tabela 15 - Síntese das variáveis independentes estatisticamente relevantes para a aplicação do <i>você-sujeito</i>	117

SUMÁRIO

CONSIDERAÇÕES INICIAIS	12
1 REVISÃO DO TEMA	15
1.1 A história da inserção do <i>Você</i> no quadro de pronomes pessoais do português brasileiro ..	15
1.2 A variação <i>tu/você</i> na fala mineira: mapeamento linguístico e social	29
SÍNTESE DO CAPÍTULO	37
2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA	39
2.1 O método de trabalho no âmbito da Sociolinguística Histórica: questões e soluções	39
2.1.1 Questões enfrentadas no levantamento de <i>corpora</i> históricos confiáveis	44
2.2 O trabalho com cartas históricas: as cartas mineiras dos séculos XIX e XX	48
2.3 Descrição das amostras de cartas históricas oitocentistas e novecentistas	50
2.3.1 O Arquivo Nacional (RJ)	51
2.3.2 O Acervo dos Escritores Mineiros	56
2.3.3 O Arquivo Público Mineiro	59
2.3.4 O Instituto Histórico e Geográfico de Minas Gerais	60
2.3.5 O Museu Abílio Barreto	61
2.3.6 O Arquivo Público da Cidade de Belo Horizonte	62
2.3.7 Acervos Pessoais	63
2.4 A Teoria do Poder e Solidariedade à luz de Brown & Gilman (1960)	64
2.5 As formas nominais e pronominais de referência à 2SG: os parâmetros de uma análise linguístico-social	66
2.5.1 Os grupos de fatores linguísticos e sociais	67
2.5.1.1 As formas de referência ao sujeito de 2SG: <i>vossa mercê/tu/você</i>	69
2.5.1.2 Expressão da forma de referência ao sujeito de 2SG: nulo/pleno	69
2.5.1.3 Pessoa verbal (<i>concordância</i>)	70
2.5.1.4 Paralelismo formal e semântico	70
2.5.1.5 Os gêneros “carta pessoal” e “carta de comércio”	71
2.5.1.6 O sexo do redator	75
2.5.1.7 A faixa etária do redator	75
2.5.1.8 As relações interpessoais e sociais	76

2.5.1.9 Período	77
2.6 O método de levantamento, submissão e análise quantitativa dos dados históricos: o RBrul em cena	77
SÍNTESE DO CAPÍTULO	81
3 RESULTADOS GERAIS: DESCRIÇÃO E ANÁLISE DOS DADOS EM TERMOS DE PERCENTUAIS	82
3.1 Os fatores linguísticos: a expressão do sujeito, a pessoa verbal e o paralelismo formal e semântico	83
3.2 Os fatores extralinguísticos: o sexo, a faixa etária, as relações sociais, o tempo e o gênero textual “carta”	92
SÍNTESE DO CAPÍTULO	110
4 A VARIAÇÃO TU/VOCÊ: DESCRIÇÃO E ANÁLISE DOS RESULTADOS ESTATISTICAMENTE RELEVANTES	112
SÍNTESE DO CAPÍTULO	116
CONSIDERAÇÕES FINAIS	118
REFERÊNCIAS BIBLIOGRÁFICAS	121
APÊNDICE (Rodada nos níveis <i>stepping up</i> e <i>stepping down</i>)	128

CONSIDERAÇÕES INICIAIS

Esta dissertação tem como foco as formas de referência ao sujeito de 2ª pessoa do singular (doravante 2SG) em missivas mineiras oitocentistas e novecentistas do português brasileiro (doravante PB). Considerando o fato de o contexto de *sujeito* ser interpretado como a principal *porta de acesso* assumida pelos pronomes gramaticalizados *você* e *a gente*, ao se inserirem no sistema pronominal do PB (RUMEU, 2013; SOUZA, 2012; LOPES & CAVALCANTE, 2011; LOPES & VIANNA, 2012; RUMEU, 2004; LOPES, 1999), pretende-se descrever as estratégias de referência ao interlocutor, tendo em vista os âmbitos da simetria e da assimetria das relações sociais (BROWN & GILMAN, 1960) que as embasam em sincronias passadas (séculos XIX e XX). Nesse sentido, esta investigação parte do princípio de que as línguas estão em constante mudança e de que as estratégias de referência ao sujeito de 2SG também podem sofrer a influência dos fatores linguísticos e sociais que permeiam a expressão linguística (falada ou escrita) dos informantes, cf. as orientações da Sociolinguística Laboviana (LABOV, 1994; WEINREICH, LABOV & HERZOG, 2008 [1968]).

É importante levar em conta que, em Minas Gerais, prevalece o *você* (COELHO, 1999; PERES, 2006; HERÊNIO, 2006; GONÇALVES, 2008; MOTA, 2008; SCHERRE *ET AL.*, 2015; SILVA, 2017; REIS, 2019), ainda que se tenha evidências do *tu*, em localidades específicas do espaço geográfico mineiro (São João da Ponte, Ressaquinha, Lontra), cf. discutido, respectivamente, por Mota (2008), Silva (2017) e Reis (2019). Nesse sentido, à luz de um panorama de variação em relação à referência ao sujeito de 2SG na atual realidade de Minas Gerais, volta-se o foco desta dissertação para a discussão da alternância *tu/você* à luz dos pressupostos teórico-metodológicos da Sociolinguística Histórica (CONDE SILVESTRE, 2007; HERNÁNDEZ-CAMPOY & CONDE SILVESTRE, 2012) e da *Teoria do Poder e Solidariedade* (BROWN & GILMAN, 1960) na escrita culta de redatores mineiros nos séculos XIX e XX.

Este trabalho também é fomentado pelas seguintes questões:

- I. Em que nível de produtividade estariam as formas *tu* e *você* na escrita mineira de sincronias passadas (séculos XIX e XX)?
- II. Quais os tipos de relações sociais que subsidiariam o *tu* e o *você* nas cartas mineiras dos séculos XIX e XX?

III. Quais os contextos linguísticos e sociais que suscitariam o *você* na produção escrita mineira oitocentista e novecentista?

Uma vez justificado o tema deste trabalho e expostas as questões motivadoras desta investigação, passa-se, de (a) a (c), à exemplificação das estratégias de referência ao sujeito de 2SG, monitoradas neste estudo (*vossa mercê, você e tu*)¹.

(a) [...] Oxalá Ellas Lhe encontrem restabelecido de seus encommodos, e que *Vossa mercê* já mais abuze do Vigor que felizmente em sua idade ainda goza! Ouvi o João de Aguiar que *Vossa mercê* espera o avizo da Tia Senhora D. Maria para hir fazer o Baptizado [...]. (BNP. Paciencia, 14.09.1856.)

(b) [...] Presentí, nas minhas conversas com Lucinha, o quanto *Você* sofreria com a perda de Maria Luiza. Ainda na hora em que esperávamos o ônibus em que viria a Lúcia, - ignorando ainda o desenlace, - estivemos falando longamente na amizade de Maninha a *Você* [...]. (JCL. RJ, 26.10.1948)

(c) [...] Muito te agradeço a parte que *estás* tomando em meus soffrimentos [...] Tenho fé em Deus que *você* há de ser muito feliz em tua melindrosa carreira sacerdotal [...]. (FAP. Caeté, 03.07.1917)

Esta dissertação está orientada pelos princípios teóricos da Sociolinguística Histórica (CONDE SILVESTRE, 2007; HERNÁNDEZ-CAMPOY & CONDE SILVESTRE, 2012) principalmente no que se refere à *autenticidade*, à *autoria* e à *validade social e histórica* das fontes (HERNÁNDEZ-CAMPOY & SHILLING, 2012). As amostras de missivas históricas foram levantadas não só em acervos públicos (em sua maior parte) tais como o *Arquivo Nacional* (RJ), o *Acervo dos Escritores Mineiros*, o *Arquivo Público Mineiro*, *Instituto Histórico e Geográfico de Minas Gerais*, o *Museu Abílio Barreto*, o *Arquivo Público da Cidade de Belo Horizonte*, mas também em Acervos Privados (não públicos). Trabalha-se com amostras de cartas (pessoais e comerciais) produzidas preponderantemente por redatores que nasceram (a maior parte) e, principalmente, viveram no espaço geográfico de Minas Gerais, entre os séculos XIX e XX principalmente (1840 a 2000). Ainda em termos de orientação teórica, traz-se à cena a descrição-analítica das formas de referência ao sujeito de 2SG em função das relações sociais que as embasam na produção escrita mineira, tendo em vista a *Teoria do Poder e Solidariedade* (BROWN & GILMAN, 1960). O controle quantitativo das formas *vossa mercê, você e tu* está amparado nas potencialidades do RStudio (interface gráfica do RBrul), tendo em vista especificamente a geração dos pesos relativos para a regra variável *tu versus você*. Em termos de

¹ Nesta investigação, optou-se por expor em itálico os exemplos das formas de referência ao sujeito de 2SG (*vossa mercê, você e tu*) nos contextos das conservadoras transcrições das cartas oitocentistas e novecentistas analisadas.

análise da regra variável, os resultados, de um modo geral, surpreendem, considerando que a hipótese principal é a de que as cartas mineiras seriam um contexto de prevalência do *você*, mas é o *tu* a forma que predomina em termos percentuais.

Esta dissertação está estruturada em quatro capítulos. Nas considerações iniciais, apresentam-se o tema, a sua justificativa, as questões motivadoras, os pressupostos teóricos, os acervos e o método de tratamento dos dados. O capítulo 1 é reservado ao resgate de trabalhos que também tiveram o foco voltado para as estratégias de referência ao sujeito de 2SG em sincronias passadas e em sincronias recentes. No capítulo 2, não somente são apresentados alguns dos princípios da Sociolinguística Histórica, mas também se expõe a *Teoria do Poder e Solidariedade*. Além disso, descrevem-se as amostras históricas, os parâmetros linguísticos e extralinguísticos da análise dos dados (grupos de fatores) e apresentam-se algumas das potencialidades e os encaminhamentos para o trabalho com o RBrul, especificamente na interface do RStudio. Os capítulos 3 e 4 são reservados à descrição-analítica dos resultados. Enquanto expõe-se, no capítulo 3, a distribuição dos dados em termos percentuais (*vossa mercê*, *você* e *tu*) em função das variáveis independentes (linguísticas e extralinguísticas), descrevem-se, no capítulo 4, os resultados estatisticamente relevantes em termos de pesos relativos para a aplicação da regra variável *tu versus você* nas cartas mineiras analisadas (séculos XIX e XX). Por fim, chega-se às considerações finais com a proposta de algumas sistematizações dos principais resultados acerca da dinâmica *tu/você* em função das questões propulsoras deste trabalho.

CAPÍTULO 1. REVISÃO DO TEMA

A proposta deste capítulo é a de uma breve retomada da história de formação do *você* na língua portuguesa (subseção 1.1) até alcançar a alternância *tu/você* na fala mineira (subseção 1.2).

1.1. A história da inserção do *Você* no quadro de pronomes pessoais do português brasileiro

O percurso histórico do pronome *você* em uso no PB passa por um recuo a sincronias passadas do português europeu (doravante PE). O presente capítulo revisita a “odisseia” pela qual a língua portuguesa passa no decorrer do tempo, tendo em vista especificamente as formas pronominais *tu* e *você*, cf. Lopes *et al.* (2018), Faraco (2017 [1996]), Scherre *et al.* (2015), Souza (2012), Rumeu (2011) e Scherre *et al.* (2009).

Ao voltar o foco para a origem do *você*, Faraco (2017 [1996], p. 51-82), em *O tratamento Você em português: uma abordagem histórica*, discute a respeito da formação do *você* português, considerando aspectos formais, históricos e pragmáticos. Na perspectiva de Faraco (2017 [1996], p. 115) à luz de Weinreich *et al.* (1968), Meillet (1906, 1965), o contexto social conduziu as mudanças linguísticas no sistema pronominal da língua portuguesa. O autor, ao fazer um levantamento histórico, desde a origem latina das formas tratamentais, explica que o sistema latino tardio era composto pelo paradigma do pronome *tu*, usado para a referência singular menos formal, pelo paradigma do pronome *vós*, usado para a referência formal a apenas um interlocutor e para a referência plural através da qual se pode evocar mais de uma pessoa, independentemente do grau de formalidade (FARACO, 2017 [1996], p. 116).

No século XIII, Portugal passou por uma fase de reorganização na estrutura social, que era dominada pelo Clero e pela Nobreza. A partir de 1254, com a ascensão econômica e social da burguesia, esta passou a ter representantes na Corte Portuguesa. Dessa forma, a burguesia começou a disputar espaço com a Nobreza, inclusive com a aquisição de terras e a conquista de cargos públicos. Nesse contexto, o Poder que estava mais concentrado nas mãos do

Rei e da Aristocracia rural sofreu um enfraquecimento, ao mesmo tempo em que surgiu uma nova Aristocracia urbana formada pela burguesia (FARACO, 2017 [1996], p. 116). Iniciava-se uma nova fase das relações tratamentais na sociedade portuguesa. As mudanças na estrutura social não tardariam a gerar mudanças também na língua.

Segundo Faraco (2017 [1996], p. 118), o primeiro registro da forma tratamental *vossa mercê* data de 1331, no texto das Cortes. A forma *vós*, forma que já não era mais de uso exclusivo para tratamento ao rei, foi substituída por diversas formas tratamentais criadas em virtude das *leis de cortesia*, no reinado de Afonso V, entre 1438 e 1481. Essas *leis de cortesia* determinavam as formas tratamentais exclusivas para se dirigir ao rei, determinando o *status* de membros da Nobreza, Clero e burguesia. Dentre as formas para o tratamento exclusivo ao Rei estão *vossa senhoria* (1434), *vossa majestade* (1442), *vossa alteza* (1450) e *vossa excelência* (1455) ao lado do *vossa mercê* (1331), cf. Faraco (2017 [1996], p. 118). A TABELA 1 evidencia a rápida alteração do valor semântico dado às formas tratamentais *vossa alteza*, *vossa senhoria* e *vossa mercê*, entre 1455 e 1490 (SANTOS LUZ, 1956, p. 362). Ao final do século XV, a forma *vossa mercê* já não mais possuía relevância para referência exclusiva à realeza.

TABELA 1
Uso das formas de tratamento ao rei (em %)

	1455	1472-3	1477	1481-2	1490
Vossa Alteza	44%	50%	54%	69%	99%
Vossa Senhoria	37%	13%	28%	24%	1%
Vossa Mercê	19%	37%	18%	7%	-

Fonte: SANTOS LUZ, 1956, p. 362.

Conforme os textos das Cortes, em 1434, há o primeiro registro da forma *vossa senhoria*. Em 1442, ocorre a criação do tratamento *vossa majestade*, seguido por *vossa alteza*, em 1450. No ano de 1455, Afonso V decreta o emprego da forma tratamental *vossa excelência* para o tratamento ao rei. No entanto, já não se tratava apenas da deferência exclusiva ao monarca português, pois a estrutura social havia mudado e, nessa nova sociedade, as pessoas que ocupavam posições hierárquicas de prestígio passaram a exigir, especialmente dos inferiores, a forma tratamental equivalente à sua posição social. Nesse sentido, o decreto de Felipe II, em

1597, e o de João V, em 1739, tiveram o objetivo de listar as formas tratamentais e os respectivos destinatários. Dessa forma, à medida que determinada forma tratamental tinha o uso banalizado, novas formas tratamentais foram criadas para assegurar o *status* social das classes sociais que faziam jus a essa deferência.

Com a expansão da burguesia, Portugal sofreu o que Brown e Gilman chamaram de “crise de tratamento” (BROWN & GILMAN, 1960, p. 270). O mundo havia mudado, enquanto a sociedade portuguesa ainda estava apegada ao antigo sistema tratamental medieval. No século XVI, já desgastada, a forma *vossa mercê* já não possuía o valor honorífico, passando a ser autorizado o seu uso, cf. decreto real de 1597, para o fechamento de cartas, cf. Faraco (2017 [1996], p. 120). Essa “crise” tornou-se alvo das comédias portuguesas dos séculos XVII e XVIII, cf. Lindley Cintra (1972, p. 30).

O sistema tratamental continuou em evolução e a forma *vossa mercê* assumiu, nos séculos XVII e XVIII, um irreversível processo de arcaicização. Com perdas semântica e formal, o *vossa mercê* evoluiu, gradual e paulatinamente, até a forma *você* (FARACO, 2017 [1996], p. 121). Isso quer dizer que o *vossa mercê* não se extinguiu abruptamente, especialmente no PE. Nessa direção, em algumas regiões de Portugal, sobretudo em zonas rurais, a forma *você* carrega uma marca negativa, cf. Paiva Boléo (1946 *apud* FARACO, 2017 [1996], p. 121). No PB, a forma *vossa mercê*, além de dar origem à forma plena do pronome *você*, continuou em processo de mudança morfológica até dar origem à forma *cê*, utilizada em quase todas as regiões do Brasil, e à forma *ocê*, característica também do estado de Minas Gerais.

Considerando a oposição entre os pronomes *tu* e *você*, no PE e no PB, Faraco (2017 [1996], p. 121) destaca que, em Portugal, o pronome *tu* é usado correntemente nas relações sociais entre pessoas íntimas, ao passo que a forma *você* tanto é empregada para o tratamento entre iguais não solidários, quanto nas relações sociais assimétricas descendentes (de superior para inferior). Por outro lado, no PB atual, a realidade linguística é diferente, uma vez que o *tu* se mostra mais produtivo em algumas regiões do país como é o caso da região Sul e de algumas localidades das regiões Norte e Nordeste do Brasil, ao passo que o pronome *você* chegou a um alto nível de produtividade em todas as regiões brasileiras (SCHERRE *et al.*, 2015).

Como explicar o fato de a forma *você* fazer referência semântica à segunda pessoa do discurso e concordar formalmente com o verbo na terceira pessoa do singular (doravante 3SG)? A explicação poderia ser dada pela origem histórica da forma *vossa mercê*, como sintagma

nominal criado para reverenciar o rei, cuja concordância verbal se dava na terceira pessoa do singular. Ao empregar a forma *vossa mercê* para se dirigir ao monarca, o súdito não estava fazendo uma referência direta ao rei, ou seja, à segunda pessoa, e sim, de forma metonímica, à *mercê* (graça) do Rei, ou seja, à terceira pessoa, à benesse que poderia ser concedida ou não pelo monarca português (FARACO, 2017 [1996], p. 122). Essa argumentação a respeito da concordância do pronome *você* com a terceira pessoa do singular localiza historicamente a “revolução da terceira pessoa” nos termos de Faraco (2017 [1996]). Esse fenômeno gerou repercussões² gramaticais no sistema linguístico, como a conhecida “mistura de tratamento” entre pronomes nominativos, acusativos, dativos e possessivos, “mistura” extremamente condenada na perspectiva da tradição gramatical. Em síntese, Faraco (2017 [1996]) discute a evolução das formas tratamentais no português desde a origem latina, passando pelas variedades brasileira e europeia do português, à luz também das dinâmicas sociais nos contextos históricos específicos.

Rumeu (2011) estuda as relações sociais, à luz da Teoria do Poder e Solidariedade (BROWN & GILMAN, 1960), interpretando-as através dos usos das formas pronominais e nominais de tratamento, em cartas cariocas setecentistas e oitocentistas. Para isso, a pesquisadora selecionou amostras de cartas manuscritas e editadas de modo fac-similar e diplomático-interpretativamente em sua dissertação de Mestrado (RUMEU, 2004). O estudo reuniu cartas escritas no Rio de Janeiro da segunda metade do século XVIII e do século XIX, compondo amostras com quinze cartas oficiais e quinze cartas não-oficiais de cada um dos dois séculos, compondo um total de sessenta cartas cariocas (RUMEU 2011, p. 116). Em sintonia com Brown & Gilman (1960) e também com Faraco (2017 [1996]), a respeito da origem das diferenças entre as formas *tu* e *vos*, Rumeu (2011) registra que a reforma decretada por Constantino³, no século IV, determinou a semântica implícita de pluralidade à forma latina *vos*, empregada para deferência ao rei. Ao se dirigir ao seu imperador, o emprego da forma *vos* englobava metaforicamente o monarca, todo o povo e todo o Poder do império concentrado nas mãos do seu líder maior (RUMEU, 2011, p. 117). Outra possível interpretação para essa semântica de pluralidade da forma *vos* envolve a divisão administrativa do império romano em império do

² Como exemplos de pesquisas voltadas às repercussões gramaticais em função dos arranjos do sistema pronominal do PB, têm-se as seguintes: Oliveira (2014), Souza (2014), Rumeu (2015), Rumeu & Oliveira (2016), Cardoso (2017), Cruz (2017), Lopes *et al.* (2018), Rumeu (2020), que também passaram pela relação entre a referência ao sujeito de 2SG (através das formas *tu* e *você*) e os pronomes-complemento nos contextos de acusativo (*te*, *lhe*, *você*), dativo (*te*, *lhe*, *a você/para você*) e oblíquo (*contigo*, *com você*).

³ Constantino, imperador romano entre 272 d.C e 337 d.C.

norte e império do sul, com a existência de dois imperadores, um ao norte e outro ao sul. Ao envolver metafórica e semanticamente os dois imperadores, buscava-se fortalecer a unidade do império romano. Nesse sentido, quando alguém se dirigia, por exemplo, ao imperador do norte, ao empregar a forma tratamental *vos*, tal reverência se estendia também ao imperador do sul e vice-versa.

No tocante à oposição entre a *Semântica do Poder* e a *Semântica da Solidariedade*, Wardhaugh (2006 [1986]) defende que houve uma evolução conceitual em direção à ampliação da Solidariedade nas relações sociais. Segundo Wardhaugh (2006 [1986]), essa evolução se deu “do assimétrico *tu/vous* para o polido e simétrico *vous/vous* e para o mútuo e simétrico *tu/tu*”. Na interpretação de Wardhaugh (2006 [1986]), as manifestações democráticas (como a Revolução Francesa, por exemplo) modificaram as relações sociais das sociedades modernas em direção à *Semântica da Solidariedade*. Por outro lado, ainda que se tenha observado o crescimento das relações simétricas fortalecidas pela dinâmica da Solidariedade, nas sociedades contemporâneas, as relações sociais ainda continuam sendo movidas pela *Semântica do Poder* (WARDHAUGH, 2006 [1986]).

Para a análise dos dados, Rumeu (2011) selecionou, nos grupos de fatores extralinguísticos, a distribuição temporal entre os séculos XVIII e XIX, o tipo de texto (entre cartas oficiais e cartas não oficiais) e as relações sociais (simétricas e assimétricas). Os dados foram submetidos ao controle estatístico oferecido pelo Pacote de Programas do Goldvarb (Pacote de Programas VARBRUL – VARiABLe RULE – adaptado ao ambiente Windows), a fim de se obter as frequências brutas das formas tratamentais e pronominais de referência ao interlocutor que se mostraram variáveis nas amostras de cartas dos séculos XVIII e XIX. Rumeu (2011) e Gomes & Lopes (2016, p. 141-142) analisam as relações sociais assimétricas ascendentes (de inferior para superior), assimétricas descendentes (de superior para inferior) e as relações simétricas (entre iguais da alta classe social).

O estudo realizado por Rumeu (2011) registrou uma mudança quanto ao emprego da forma tratamental *vossa excelência* do século XVIII para o século XIX. No século XVIII, as relações sociais ascendentes apresentaram maior produtividade, em 71% das ocorrências de uso, com a forma *vossa excelência*, ao passo que, no século XIX, a maior produtividade de dados se deu nas simétricas. Quanto à forma *vossa mercê*, foi nas relações sociais simétricas que tal forma apresentou maior produtividade, com frequência de 63%, no contexto do século XVIII. Quanto à

forma *você*, a autora constatou maior produtividade, em 94% das ocorrências, nas relações assimétricas descendentes (relações de superior para inferior). O *tu* evidenciou, por outro lado, maior produtividade, em 92% das ocorrências, em cartas particulares trocadas entre missivistas com maior proximidade entre si. Segundo Rumeu (2011), o *vossa mercê* passou por um acelerado processo de dessemantização, aumentando o percentual de 63%, no século XVIII, para 100%, no século XIX. Entre o período do século XVIII para o século XIX, as formas *você* e *tu* aumentam a produtividade nas relações solidárias de um mesmo grupo social (classe social alta): de 6% para 100%, para a forma *você*; e de 92% para 100%, para o pronome *tu*. Em síntese, Rumeu (2011) considera o estudo como preliminar, em razão do reduzido número de dados, mas consegue tecer algumas considerações com respeito à correlação entre as frequências de uso das formas tratamentais e o tipo de relação social estabelecida entre autor e interlocutor. No que tange ao século XVIII, a investigação aponta que, na amostra analisada, as relações assimétricas ascendentes e descendentes, perfazem 69% das ocorrências, indicando uma organização social movida pela *Semântica do Poder*. Por outro lado, a alta produtividade das relações simétricas, em 68% do total de ocorrências das cartas do século XIX, permite inferir que a *Semântica da Solidariedade* parece encaminhar a organização da sociedade brasileira.

No capítulo *A reorganização do sistema pronominal de 2ª pessoa na história do Português Brasileiro: a posição de sujeito*, Lopes *et al.* (2018) assumem o objetivo de discutir os sistemas dos pronomes de referência ao sujeito de 2SG produtivos no PB de sincronias passadas (séculos XIX e XX). Os autores registram que o quadro pronominal do PB “reflete um sincretismo entre a segunda e a terceira pessoas gramaticais” (LOPES *et al.* 2018, p. 26) que pode ser uma continuidade da “revolução da terceira pessoa” ocorrida no século XIV, cf. Faraco (2017 [1996], p. 116). Entre os exemplos desse sincretismo no PB, destaca-se o uso das formas pronominais *te*, *lhe* e *você*, na função de complemento, em combinação, ora com o pronome-sujeito *você*, ora com o pronome-sujeito *tu*.

Da forma nominal *vossa mercê*, que concorda formalmente com o verbo na terceira pessoa do singular, a forma gramaticalizada *você* herdou características morfológicas. É o que argumenta Hopper (1991) no sentido de ter havido “a persistência da especificação original de 3ª pessoa, embora a interpretação semântico-discursiva seja a de 2ª pessoa” (HOPPER 1991 *apud* LOPES *et al.* 2018, p. 27). Segundo os autores, quanto ao emprego dos pronomes *você* e *tu*, a partir do século XX, o PB e o PE seguem caminhos distintos. Partindo dessa constatação, Lopes

et al. (2018) discutem a referida reorganização do sistema pronominal numa perspectiva diatópico-diacrônica. O trabalho de Lopes *et al.* (2018) se mostrou motivador para esta investigação no nível de Mestrado também pelo fato de se tratar de uma análise sociolinguística embasada em “cartas pessoais” para a análise dos pronomes-sujeito de 2SG, enfocando prioritariamente na alternância *tu/você*. Trata-se de um amplo mapeamento das formas de referência ao sujeito de 2SG no PB escrito em sincronias passadas, o que tende a permitir uma perspectiva de análise comparativa entre a realidade mineira e a de outros estados brasileiros. Lopes *et al.* (2018) abordam, em uma primeira parte, os aspectos históricos, teóricos e metodológicos quanto ao processo de mudanças das formas tratamentais de 2SG na língua portuguesa e discutem, na segunda parte, o emprego das formas tratamentais de 2SG.

Quanto aos aspectos históricos, pode-se destacar que o pronome *vós* herdou o caráter de respeito e cortesia da forma latina *vos*, empregado em relações sociais ascendentes, de inferior para superior. Conforme constatado por Faraco (2017 [1996]) e Rumeu (2011), o *vos*, que inicialmente foi usado para a deferência exclusiva ao rei, com o tempo perdeu esse *status* e foi substituído pelo *vossa mercê* no português medieval. Essa preocupação com a manifestação (marcação) linguística do prestígio real (extensivo também à nobreza) não parou por aí. Em consequência, criaram-se as formas tratamentais *vossa senhoria*, *vossa alteza*, *vossa excelência* e *vossa majestade*, conforme já abordado neste capítulo. À medida que determinada forma tratamental deixava de ser exclusiva para tratamento ao rei e passava a ser usada para a deferência a outras posições de prestígio, essa forma entrava em processo de desgaste quanto à semântica de cortesia, cf. Santos Luz (1956), Lindley Cintra (1972) e Faraco (2017 [1996]). Segundo Lopes *et al.* (2018, p. 31), em relação ao período quinhentista do português, “sabe-se que no período inicial de ocupação do Brasil, no século XVI, o desgaste semântico sofrido por *vós*, a redução fonética do *vossa mercê* e o seu uso generalizado como *você* já estavam em etapa bastante avançada”.

Segundo Lopes *et al.* (2018), no século XVIII, nas cartas escritas no Brasil, o emprego do pronome *você* era marcado em relações sociais assimétricas descendentes (relações de superior para inferior). No mesmo período, a forma nominal de tratamento *vossa mercê*, ainda em uso, era empregada em relações simétricas entre pessoas da classe alta e também em relações sociais assimétricas ascendentes (relações de inferior para superior). Por outro lado, as cartas brasileiras do século XX evidenciam que a forma pronominal *você* começa a competir com o

pronome *tu*, nas relações sociais simétricas mais íntimas, seja entre os menos escolarizados, seja entre os mais escolarizados, assumindo inicialmente maiores frequências de uso na produção escrita das mulheres. Nesse sentido, “a era oitocentista do PB tem sido assumida como ponto de partida para a anuência alcançada pelo *você*, como pronome de referência ao sujeito de (segunda pessoa) do singular” (SOTO, 2001 *apud* LOPES *et al.*, 2018, p. 31). Conforme Lopes *et al.* (2018, p. 31), as cartas do século XIX já apresentavam registros da coexistência do pronome-sujeito *você* com pronomes-complemento e possessivos de 2SG e 3SG, tais como *te/o, a, lhe; tua/sua*, produtividade essa que avança pelo século XX. Essa coexistência pode ser observada também nas cartas que embasam esta Dissertação de Mestrado, cf. a sentença (1), transcrita de uma carta pessoal familiar, em uma relação social assimétrica descendente (de superior para inferior), através da qual a mãe tem o filho como interlocutor.

(1) “[...] fiquei muito satisfeita Com a tua resolução de apresentar-se candidato, estou certa de que *você* sahirá victorioso [...]” (MBCS. S.I., 21.03.1903).

Conforme Lopes *et al.* (2018), até o século XVIII, as relações sociais eram movidas pela *Semântica do Poder* também através do uso do *você* nas relações sociais assimétricas descendentes. A partir dos séculos XIX e XX, o *você*, além de já ser usado nas relações simétricas, passou a ser usado também nas relações assimétricas ascendentes. Através do seguinte trecho de uma carta mineira do início do século XX, transcrito em (2), exemplifica-se uma relação assimétrica ascendente (de inferior para superior) por meio das palavras dirigidas pela afilhada CLB à sua madrinha HL.

(2) “Querida madrinha, Fazendo-lhe uma visitinha muito afetuosa venho convida-la para assistir à festinha da entrega dos diplomas no próximo dia 30. Terei muito gosto que *você* venha. [...]” (CLB. Lambari, 25.11.1941).

Segundo Lopes *et al.* (2018, p. 34), no século XIX, o *você* era uma estratégia híbrida e ambígua, pois, ao mesmo tempo em que conservava traços de distanciamento polido, disputava espaço com o pronome *tu*, particularmente em relações solidárias entre pessoas íntimas. Nessa direção, os autores investigaram as relações sociais, tendo como referência as relações assimétricas descendentes (de superior para inferior, tais como pai-filho, mãe-filho etc.), as relações assimétricas ascendentes (de inferior para superior tais como filho-pai, filho-mãe etc.) e as relações simétricas (tais como as relações entre amigos etc.), cf. Lopes *et al.* (2018, p. 41).

Quanto aos dados relativos ao *corpus* de cartas de Minas Gerais usado por Rumeu, Cruz & Cardoso (cf. LOPES *et al.*, 2018), estiveram em foco 89 cartas mineiras, circunscritas no período de 1850 até 1989. Considerando as ocorrências de *vossa mercê*, *você* e *tu*, como estratégias de referência ao sujeito de 2SG, o pronome-sujeito *você* obteve maior produtividade, com a frequência de uso de 84% (189/226), seguido pelo *tu*, com 12% (28/226) e pelo *vossa mercê*, com 4% (9/226). Entre 1850 e 1989, essas três formas de referência ao sujeito de 2SG sofreram mudanças. No período de 1850 a 1879, as formas em competição são o *vossa mercê* (60%) e o *você* (40%). De 1890 até 1930, o pronome-sujeito *você*, com mais de 80%, mostra-se bem produtivo. Por fim, entre 1940 e 1989, observa-se a alternância *tu/você* em favor do inovador *você* que chega, no final do período, a 100% dos dados. De um modo geral, os dados em Minas Gerais indicaram que a forma *vossa mercê* foi uma estratégia usada majoritariamente no século XIX. Na primeira metade do século XX, o uso do pronome-sujeito *tu* foi baixo (5%) e, na segunda metade do século XX, o emprego do pronome-sujeito *você*, por outro lado, sobressaiu-se (91% (1940-49), 98% (1950-59), 100% (1960-69), 100% (1970-79), 100% (1980-89)). Segundo Rumeu, Cruz & Cardoso (2018 *apud* Lopes *et al.* 2018, p. 77), esses resultados que mostram a produtividade do *você* em sincronias passadas (séculos XIX e XX) parecem confirmar o fato de o subsistema de *você* prevalecer na fala mineira atual.

Para o estudo dos pronomes de 2SG *tu* e *você* no PB de sincronias passadas, é importante mencionar o fato de Rumeu, Cruz & Cardoso (cf. LOPES *et al.*, 2018) terem relacionado os seus resultados também aos seis subsistemas, cf. Imagem 1, propostos por Scherre *et al.* (2009), atualizados por Scherre *et al.* (2015) e reduzidos a três subsistemas por Lopes & Cavalcante (2011).

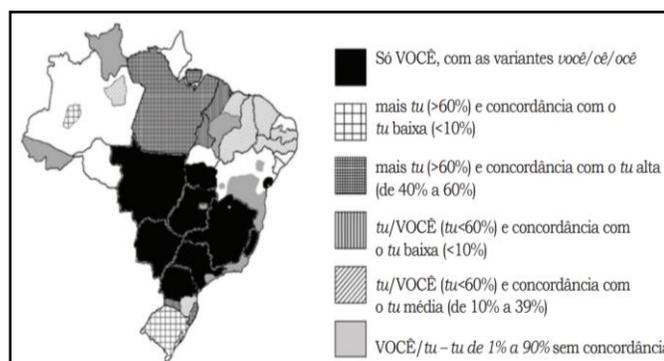


Imagem 1 – Seis subsistemas dos pronomes de segunda pessoa *você* e *tu* no PB

Fonte: SCHERRE *et al.* 2015, p. 142.

O subsistema *Só você*⁴, com as variantes *você/cê/ocê* engloba os estados de Goiás, Mato Grosso e Mato Grosso do Sul (Centro-Oeste), o estado do Tocantins (Norte), a cidade de Salvador (Nordeste), o estado do Paraná (Sul) e grande parte dos estados de Minas Gerais e de São Paulo (Sudeste), à exceção da cidade de São João da Ponte⁵, em Minas Gerais. A partir do mapeamento elaborado por Scherre *et al.* (2009), atualizado por Scherre *et al.* (2015), em que vigoram seis subsistemas tratamentais, Lopes & Cavalcante (2011) reduzem-nos a três, cf. também discutido por Lopes *et al.* (2018): (I) o subsistema de *você* (regiões Centro-Oeste, Sudeste); (II) o subsistema de *tu* (Nordeste, Norte, Sul); e (III) o subsistema da alternância *você/tu* (regiões Centro-Oeste, Sudeste, Sul, Nordeste, Norte).

Ainda numa perspectiva histórica, têm-se os resultados de Souza (2012) que também volta o seu foco de análise para os pronomes-sujeito de 2SG *tu* e *você* com o objetivo de observar o comportamento do *você*-sujeito em missivas de cariocas e de pessoas que moravam no Rio de Janeiro, nos séculos XIX e XX. À luz dos princípios da Sociolinguística Histórica que, por sua vez, foram inspirados na Sociolinguística Variacionista (LABOV, 1994; WEINREICH, LABOV & HERZOG, 2008 [1968]), o estudo de Souza (2012) partiu da hipótese de que o *tu*-sujeito se manteve como estratégia de 2SG durante o século XIX e início do século XX, com predominância de ocorrências de *tu*-sujeito nulo nas relações simétricas de maior intimidade, entre os missivistas do sexo masculino. Além da Teoria da Variação, a autora orientou o seu trabalho de análise da dinâmica variável das formas pronominais *tu/você* à luz também da Teoria de Poder e Solidariedade (BROWN & GILMAN, 1960). O *corpus* analisado por Souza (2012) é composto por 354 cartas pessoais, partindo do entendimento de que o referido gênero “estabelece uma comunicação marcada pela espontaneidade” (SOUZA, 2012, p. 59). Para a análise da influência e da significância de cada grupo de fatores (variáveis independentes) em relação à variável dependente, Souza (2012) utilizou o Pacote de Programas Goldvarb (GUY & ZILLES, 2007).

A investigação de Souza (2012) está voltada para o pronome-sujeito de 2SG, levando em conta também a expressão do sujeito e partindo da hipótese de que, no século XIX, o *você* seria preferencialmente preenchido na posição de sujeito, ao contrário do *tu*-sujeito, que ocorreria

⁴ Neste trabalho, ao invés de empregar a nomenclatura subsistema *Só você* proposta por Scherre *et al.* (2015), propõe-se o emprego do termo subsistema de *você*-sujeito *exclusivo*.

⁵ Além da cidade de São João da Ponte (MOTA, 2008), essa revisão do tema registra que investigações realizadas nas cidades de Ressaquinha (SILVA, 2017) e Lontra (REIS, 2019) também constataram a presença do uso da forma *tu* sem concordância.

prioritariamente nulo. No século XX, a partir da década de 1930, seria possível observar que o *você*-sujeito suplantaria o pronome *tu* e o PB passaria a ser uma língua [- sujeito nulo]. Uma segunda hipótese seria a de que as ocorrências verbais de 3SG favoreceriam a forma *você* e as de 2SG favoreceriam o *tu*. Os casos de concordância não canônica ocorreriam com o pronome *tu*, ainda na primeira metade do século XX, quando a alternância teria se tornado maior. Uma terceira hipótese seria a de que o emprego da forma inovadora *você* seria mais produtivo nas cartas femininas, ao passo que o *tu*-sujeito seria mais recorrente nas missivas masculinas. Uma quarta hipótese considerada por Souza (2012, p. 86) seria a de que o *você*-sujeito ocorreria de forma mais produtiva nas relações assimétricas ascendentes (de inferior para superior), marcando o distanciamento e respeito herdados do arcaico *vossa mercê*. Por outro lado, nas relações descendentes (de superior para inferior), o *tu* seria o esperado. Quando as relações sociais fossem marcadas pela simetria entre os interlocutores, a expectativa seria a disputa entre as variantes *você* e *tu*. A quinta hipótese apontada pela autora seria a de que quanto maior o grau de intimidade e de vínculo familiar, maior a tendência ao emprego do pronome *tu*, ao passo que o emprego do respeitoso e distante *você* se daria quando fosse menor o grau de familiaridade e intimidade. Por fim, a sexta e última hipótese apontada por Souza (2012) diz respeito à variável período. Nesse sentido, o emprego do *tu* seria mais produtivo no século XIX, porém, no início do século XX, a forma inovadora passaria a ser mais frequente, marcando, ainda, o distanciamento e atenuação característicos do século XIX. Com o aumento das ocorrências de *você*, este passaria a disputar o mesmo espaço com o pronome *tu*. Essa disputa entre *você* e *tu* estaria relacionada às mudanças vividas pela sociedade brasileira na primeira metade do século XX (SOUZA, 2012, p. 86).

Os resultados alcançados por Souza (2012, p. 6) evidenciaram que a partir da década de 1930, o emprego de *tu* foi superado pelo *você* (DUARTE, 1993, 1995; SOUZA, 2012). A análise foi iniciada com a distribuição geral dos dados de *tu* e *você*, por década, de 1870 até 1970, controlando o número de cartas e o número de dados, num total de 354 e 1381, respectivamente. Os resultados registraram o predomínio do *tu*, no século XIX, ao passo que, no século XX, os dados de *tu* entraram em declínio e os de *você* aumentaram em relação ao século XIX (SOUZA, 2012, p. 90), cf. o seu GRÁFICO 1.

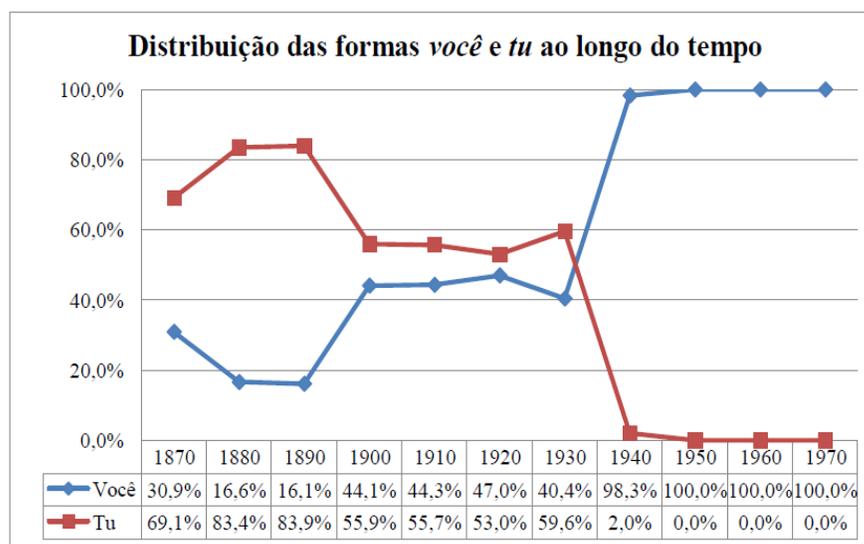


GRÁFICO 1 – Distribuição de frequência de *tu* e *você* em 100 anos de produção epistolar: de 1870 a 1970.

Fonte: SOUZA, 2012, p. 90.

De forma geral, Souza (2012, p. 90) sistematiza os resultados da alternância *tu/você* em três momentos distintos. São eles:

- Período I (1870-1890): o *tu* é mais frequente que o *você*.
- Período II (1900-1930): o *tu* e o *você* apresentam frequências próximas.
- Período III (1930-1970): o *você* torna-se a estratégia preferida.

Considerando as frequências de uso e os pesos relativos dos pronomes-sujeito de 2SG, tendo a forma *você* como valor de aplicação, Souza (2012 p. 93) observou que o *você* pleno apresentou frequência de 77,4% e 0.744 de peso relativo (PR), para a realização plena da forma inovadora, ao passo que a expressão do *você*-sujeito nulo apresentou frequência de 25,1% e 0.344 de peso relativo. Esses resultados corroboram os resultados de Rumeu (2008) e Machado (2006, 2011). Quanto à pessoa verbal, os dados observados por Souza (2012) evidenciaram que as formas de 3SG favoreceram o *você*, com PR 0.99, ao contrário das formas de 2SG, com PR 0.018. Esses resultados mostram que a marca de 3SG, herdada do *vossa mercê* na concordância do *você* com a 3SG, permaneceu nos dados históricos. Os resultados apontaram, também, que a concordância não canônica apresentou poucas ocorrências no *corpus* de escrita analisado, com apenas cinco ocorrências em um universo de 1381 dados (SOUZA, 2012, p. 99). Quanto às

relações sociais, as formas *tu* e *você* distribuíram-se da seguinte forma (cf. TABELA 2), descrita e analisada por Souza (2012, p. 103).

TABELA 2
A utilização da forma pronominal *você*
em cada tipo de relação social em cartas cariocas

RELAÇÃO SOCIAL			
	<i>Você</i>	%	P.R.
Superior-Inferior	96/324	29,6%	0.342
Simétrica	515/1018	73,7%	0.571
Inferior-Superior	8/39	2,8%	0.117

Valor de aplicação: *você*.

Fonte: SOUZA, 2012, p. 103.

O peso relativo 0.571 para o *você* indica que as relações simétricas favorecem levemente o emprego da forma *você*, ao passo que as relações assimétricas de superior para inferior e de inferior para superior mostraram-se como contextos propulsores do *tu*. As relações sociais ascendentes (de inferior para superior) contrariam a hipótese de Souza (2012), tendo em vista o fato de que o *tu* apresentou maior produtividade nesse contexto social. Na amostra analisada por Souza (2012, p. 103), as relações assimétricas ascendentes (de inferior para superior) mostraram-se regidas pela Solidariedade, visto que o *tu* foi a forma pronominal que prevaleceu nesse tipo de relação social das cartas cariocas.

Souza (2012) analisou ainda o emprego do *você* nos três momentos históricos (1870-90 (I), 1900-30 (II), 1930-70 (III)) a partir dos quais é possível perceber que o pronome *você* apresentou menor produtividade no período I (1870-1890), aumentou suas frequências de uso ao longo dos períodos II (1900-30) e III (1930-70) até o *você* mostrar-se, no decorrer do século XX, como a estratégia pronominal preferida pelo redator carioca.

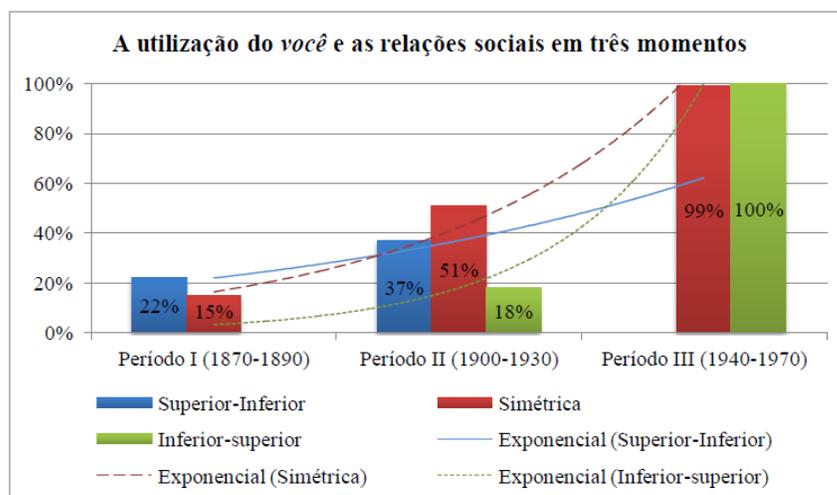


GRÁFICO 2 - A utilização do *você* nas relações sociais em três períodos diferentes
 Fonte: SOUZA, 2012, p. 111.

Ao analisar o resultado das relações interpessoais, Souza (2012, p. 111) interpretou que a forma inovadora *você* foi mais produtiva nas missivas trocadas no âmbito da família ampliada (simetria social marcada pela relação entre parentes e amigos), com PR 0.785, ao passo que, entre os casais cariocas, o *você* teve a sua força retraída (PR 0.186), enquanto o contexto das cartas familiares trocadas entre integrantes da família nuclear, com 0.522 de peso relativo, evidenciou produtividade acima das cartas entre casais e abaixo da família ampliada. A autora concluiu que o emprego da forma *você* é menor nas relações sociais de maior intimidade.

Por fim, ao analisar o grupo de fatores gênero *carta pessoal*, os dados da investigação de Souza (2012), entre outros resultados, apontaram que o *você*-sujeito apresentou maior frequência (64%) nas cartas pessoais, o que a levou a interpretar que quanto maior o distanciamento entre os interlocutores, maior a produtividade do *você*. Em sentido oposto, parece seguir, nas cartas amorosas, uma vez que a forma *você* apresentou menor produtividade, com apenas 16,8%, em oposição a 83,2% de ocorrências de *tu*. Cruzando o gênero *carta* com o *sexo*, observa-se que as mulheres apresentaram maior frequência de *você* (61%) nas cartas familiares, contra 31% de *você* nas ocorrências dos homens. Mesmo nas cartas amorosas, que apresentaram baixa produtividade do pronome *você* (16,8%), as mulheres usaram a forma inovadora com maior frequência, em 34% dos dados, em oposição aos 5% de produtividade do *você* nas cartas amorosas escritas por homens. Para a melhor compreensão dessa variação das formas *tu* e *você*, passa-se ao mapeamento linguístico e social na fala mineira.

1.2. A variação *tu/você* na fala mineira: mapeamento linguístico e social

A discussão acerca da variação *tu/você* em sua expressão falada está fundamentada no objetivo principal deste trabalho, que é a descrição-analítica das evidências históricas da alternância *tu/você* à luz das relações sociais que as embasaram nas cartas mineiras (séculos XIX e XX). A análise de resultados de investigações baseadas em dados de fala pode oferecer um interessante parâmetro de análise acerca da dinâmica *tu/você* na fala mineira, cf. Reis (2019), Silva (2017), Scherre *et al.* (2015), Gonçalves (2008), Mota (2008), Peres (2006), Herênio (2006) e Coelho (1999), servindo como ponto de partida para a discussão dos resultados desta investigação voltados para sincronias passadas.

Considerando o mapeamento linguístico do PB, Scherre *et al.* (2015, p. 171) asseguram que “o *tu* brasileiro, assim como o *você* brasileiro, é plural”. Para Minas Gerais, os resultados apontam para a prevalência do pronome *você* (e variações *ocê* e *cê*). Nesse sentido, serão expostas algumas peculiaridades do uso do *você*-sujeito especificamente no espaço geográfico de Minas Gerais que surpreendentemente apresenta algumas evidências do *tu*. Nesse sentido, a investigação de Peres (2006) analisou duas amostras de fala, sendo uma de 1982 e outra de 2002. Submetendo os dados ao Pacote de Programas Goldvarb, revelou produtividade variável para as formas *você*, *ocê* e *cê*. Ao *cê* coube apenas a sua produtividade na função de sujeito pré-verbal (VITRAL, 1996 *apud* PERES, 2006, p. 22). A forma *ocê* mostrou-se mais produtiva como complemento verbal. Na amostra de fala de 1982, o *cê* evidenciou maior produtividade em 77,5% (455/587) dos dados, seguido pelo *você*, com 15,9% (93/587) e *ocê*, com 6,6% (39/587). Mostraram-se significativas a *função sintática*, a *idade* e a *classe social* dos informantes para o *cê*. A forma *ocê* mostrou-se mais produtiva como complemento verbal. Na amostra de fala de 1982, o *cê* evidenciou maior produtividade em 77,5% (455/587) dos dados, seguido pelo *você*, com 15,9% (93/587) e *ocê*, com 6,6% (39/587). A amostra de fala do ano de 2002 reuniu 1453 dados distribuídos entre as formas *cê*, que obteve com 72,6% (1055/1453), seguido pela forma *você*, com 23,5% (342/1453), e *ocê*, com 3,9% (56/1453). Os resultados evidenciaram como significativas a *função sintática* (favoreceu o *cê* em função de sujeito), a *contiguidade em relação ao verbo* (favoreceu o *cê* pré-verbal), a *referência* (favoreceu o *cê* como referência definida) e a *focalização contrastiva* (favoreceu o *você*).

Segundo Peres (2006), os dados comparativos entre as amostras de fala de 1982 e 2002 indicaram a complexidade dos usos das formas *você*, *ocê* e *cê* em Belo Horizonte. Os resultados indicaram a implementação da forma *cê* (sujeito), em Belo Horizonte, na fala das mulheres, o que se aplica também ao *você*, cf. Peres. Quanto ao *ocê*, o estudo concluiu que o seu uso se mostrou restrito à função de sujeito da oração principal. Além disso, a autora, à luz dos resultados de Ramos (2000 *apud* PERES, 2006) e de Coelho (1999), ressaltou o fato de a produtividade da forma *ocê* ter se deixado evidenciar nas áreas rurais de Minas Gerais.

Partindo da capital mineira, passa-se à análise de Herênio (2006) que, à luz da perspectiva da Sociolinguística paramétrica e de 86 entrevistas, sendo 43 delas em Imperatriz (MA) e mais 43 aplicadas aos falantes de Uberlândia (MG), enfocou a dinâmica *tu/você*. Os resultados indicaram que a forma *você* apresentou frequência de uso superior ao *tu*, nas duas cidades, evidenciando 73%, em Imperatriz, e 100%, em Uberlândia, o que corrobora o mapeamento linguístico de Scherre *et al.* (2015) a partir do qual Minas Gerais está inserido no subsistema linguístico *você*-sujeito exclusivo.

Para o centro-oeste mineiro, tem-se a pesquisa de Gonçalves (2008) voltada à produtividade das formas *você*, *ocê* e *cê* na cidade de Arcos. Quanto aos fatores controlados, o estudo analisou o *tipo de contexto*, o *ambiente fonológico* precedente, a *função sintática* e o *tipo de frase*, a *procedência geográfica*, a *classe social*, a *faixa etária* e o *gênero* dos informantes. Os dados de fala contam com aproximadamente 26 horas de gravação, constituídos por 100 narrativas orais espontâneas de falantes da cidade de Arcos (MG). Foram coletadas 50 narrativas da zona urbana e 50, da zona rural. Do total, 82 narrativas apresentaram ocorrências das formas *você*, *ocê* e *cê* (GONÇALVES, 2008, p. 159). A partir das narrativas individuais orais espontâneas, foram levantados 510 dados da fala de 40 informantes (20 da zona urbana e 20 da zona rural), distribuídos entre as faixas etárias de 15 a 30 anos, de 31 a 59 anos e de 60 anos acima. A investigação fundamentou-se nos pressupostos teóricos da Sociolinguística Interacional (GOFFMAN, 1976 *apud* GONÇALVES, 2008, p. 47) e da Teoria do Poder e Solidariedade (BROWN & GILMAN, 1960, p. 178-179). O objetivo principal foi identificar indícios que comprovem se a forma *você* é um pronome movido pela *Semântica do Poder* e se as formas variantes *ocê* e *cê* seriam movidas, por outro lado, pela dinâmica da Solidariedade (BROWN & GILMAN, 1960, 1978). Os resultados demonstraram que, na variação dos pronomes *você* versus *ocê/cê*, as formas *você*, *ocê* e *cê* estão em variação estável, identificando a forma *você* em

relações de Poder e as formas *ocê* e *cê* em dinâmicas de Solidariedade. Com base na Sociolinguística Interacional, a investigação de Gonçalves (2008, p. 231-234) concluiu que a situação interacional entre o falante e seu interlocutor define as relações estabelecidas de Poder ou de Solidariedade que envolvem traços de polidez ou de intimidade, respectivamente, que tendem a interferir nas escolhas pronominais.

Localizada em Minas Gerais, a 563 km ao norte de Belo Horizonte, a cidade de São João da Ponte, com 25.358 habitantes⁶, foi ambiente de pesquisa em estudos linguísticos realizada por Mota (2008). A pesquisadora investigou se há tendência de mudança em relação à variação pronominal *tu* e *você*, no português oral do norte de Minas. O estudo está fundamentado nos pressupostos teóricos da Teoria de Variação de Labov (1972) para a análise de 24 entrevistas com informantes dos sexos feminino e masculino, distribuídos em quatro faixas etárias (7 a 14 anos, 15 a 25 anos, 26 a 49 anos e acima de cinquenta anos). O estudo concluiu que o *tu* é mais produtivo nas relações marcadas por intimidade entre os interlocutores. Trata-se de um fenômeno de variação estável em que o *tu* foi observado em 10% dos dados (49/509). A maior produtividade do *tu* mostrou-se entre os informantes mais jovens (entre 15 a 25 anos), ainda que tal forma variante (“*tu*”) tenha se mostrado funcional também entre falantes de meia idade e idosos, levando a autora a interpretar o *tu* na fala dos jovens como uma “estratégia linguística utilizada pelos jovens para a construção de uma identidade de grupo” (MOTA, 2008, p. 84).

A investigação de Silva (2017) para a cidade de Ressaquinha (MG), localizada na mesorregião do Campo das Vertentes, a 129 km da capital mineira evidenciou o *tu*, em 38% dos dados, ao lado de *você* (10%), *ocê* (17,1%) e *cê* (34,4%). Considerando os fatores externos, Silva observou que a forma *você* apresentou maior produtividade tanto na zona rural (63,5%), quanto na zona urbana (59,7%). Quanto aos fatores internos, o *tu* foi favorecido pelo contexto das frases interrogativas (53,3%), ao passo que a forma *você* mostrou-se mais produtiva nas frases negativas (69,7%). Quando comparamos os pronomes *você* versus *tu*, tendo em vista os pesos relativos, o pronome *tu* é favorecido pelo contexto de referência específica (0.602, 45,4%, 199/438), ao passo que o pronome *você* é favorecido em uso de referência genérica. O *tu* também foi favorecido pelo contexto sintático de sujeito (0.544, 40,8%, 221/542) inclusive no discurso direto (0.548, 40,1%, 172/429). A soma da forma *você* e suas variantes *ocê* e *cê*, com frequência de 61,5%, em

⁶ IBGE/2010.

oposição à forma *tu*, com frequência de 38,5%, deixa claro que a produtividade da forma inovadora (*você*) é predominante, confirmando a hipótese inicial da autora.

Reis (2019) dá uma interessante contribuição ao mapeamento acerca da variação *tu* e *você* em Minas Gerais, mais especificamente em parte do norte de Minas Gerais. A pesquisa teve o objetivo de verificar a alternância de uso das formas *tu* e *você* na escrita e na oralidade dos alunos de duas turmas do 8º ano do ensino fundamental de uma escola pública estadual, em Lontra⁷, localizada no norte de Minas Gerais. A investigação teve como principal hipótese a alternância *tu* e *você*, na fala e na escrita em Lontra (MG), como um caso de variação linguística cuja alternância é influenciada por fatores estruturais e não estruturais (REIS, 2019, p. 28). Reis (2019) argumenta que a forma tratamental *tu*, embora tenha uma origem historicamente anterior ao pronome *você*, é objeto, em Lontra, de preconceito. Como peculiaridade do uso das formas *você* e *tu* em Lontra, a pesquisadora destaca que o *você* é usado nas relações sociais travadas entre desconhecidos e o *tu* é empregado para o tratamento a pessoas íntimas, familiares ou conhecidas.

Reis (2019) analisou um *corpus* de oralidade composto por dados coletados de 48 alunos de duas turmas (4 e 5) do 8º ano do ensino fundamental. A idade dos alunos está na faixa etária entre 13 e 14 anos. A turma 4 é composta por 27 alunos (19 meninas e 8 meninos), dos quais 22 são da zona urbana e 5, da rural. A turma 5 é composta por 21 alunos (8 meninas e 13 meninos), dos quais 18 são da zona urbana e 3, da área rural. Os dados foram coletados a partir da gravação de duas mensagens de áudio produzidas por cada aluno/informante. Na fase diagnóstica da pesquisa, foi elaborado um *corpus* de escrita com os dados coletados a partir da atividade “Fato conhecido”, na qual cada informante elaborou uma narrativa sobre uma tragédia ocorrida em determinado local e também a partir da elaboração de um bilhete endereçado a um amigo do aluno/informante. (REIS, 2019, p. 93). Na primeira atividade escrita, considerando as duas turmas, os resultados mostraram 77,5% de *você* (e variantes *ocê* e *cê*) em oposição ao *tu* (22,5%). Na segunda atividade do *corpus* de escrita, “Um bilhete para um amigo”, o pronome *você* (e variantes *ocê* e *cê*) obteve um maior percentual em 64,7% das ocorrências, superior ao *tu* (35,3%). Os dados de língua escrita expuseram ocorrências das formas variantes *tu* e *você* e também das formas *ucê*, *ocê* e *cê*, formas variáveis do “você”, porém, com menores frequências.

⁷ Cf. o Censo IBGE 2010, Lontra possui uma população de 8.397 pessoas, que tem como um de seus vizinhos a cidade de São João da Ponte. É uma cidade jovem, cujo povoamento foi iniciado em 1920, aproximadamente, tornando-se oficialmente uma cidade no ano de 1992.

Constatou-se também que a produtividade da variável dependente foi influenciada pelos fatores externos *sexo* e *grau de intimidade*. Foram detectadas ainda marcas de oralidade na escrita, tais como o uso do *tu* em concordância com o verbo na 3SG e o emprego das variantes do *você* através das formas *ocê*, *ucê* e *cê* nos textos escritos pelos informantes.

Os resultados estatísticos relativos aos *corpora* de escrita e de oralidade revelaram que os informantes usaram o pronome *tu* para se referir ao amigo com uma frequência bem menor do que a forma *você*, cuja produtividade foi de 74,5%. Mesmo assim, o pronome *tu* apresentou alguma produtividade (25,5%). Considerando apenas a variação *tu* e *você* na oralidade, Reis (2019) constatou que o *você* ocorreu em 79% dos dados em oposição ao *tu*, que foi usado em 21% dos dados. Assim como observado para a amostra de língua escrita, também para a amostra de língua falada foram observadas evidências das variações do *você* (*ocê* e *ucê*) motivadas pelas variáveis sociais *gênero* (*sexo*), *grau de intimidade* e *classe social*. Embora a investigação de Reis (2019) tenha contado tão somente com menos de 50 informantes, além de ter se limitado à faixa etária entre 13 e 14 anos, quando somada a outras pesquisas realizadas com base em dados do interior de Minas Gerais, a pesquisa se mostra relevante, tendo em vista os resultados observados para Ressaquinha (SILVA, 2017) e para São João da Ponte (MOTA, 2008).

Encerrando esse capítulo de revisão do tema, discute-se os resultados de Coelho (1999), cuja análise se voltou para o emprego do *você* no português falado no norte de Minas, sob a perspectiva da abordagem da Teoria da Variação. Além da forma plena do pronome *você*, a pesquisa sociolinguística também controlou o uso das variantes *ocê* e *cê*, na fala dos moradores da cidade de São Francisco⁸. A pesquisa de Coelho (1999) teve o objetivo de identificar os fatores linguísticos e os fatores extralinguísticos que condicionam a alternância entre as variantes *você*, *ancê*, *ocê* e *cê* em São Francisco (MG). Nessa direção, entende-se que a variação sincrônica, em determinada comunidade de fala, dá continuidade a um fato diacrônico a partir do qual a mudança é necessariamente precedida por evidências de variação, cf. Wenreich, Labov e Herzog (1968 *apud* COELHO, 1999, p. 34). Na composição da amostra, foram coletados 629 dados, a partir de entrevistas orais com 24 pessoas residentes em São Francisco, sendo 12

⁸ Em conformidade com o Censo do IBGE 2010, São Francisco tem uma população de 53.828 pessoas cuja maioria reside na zona rural. A cidade de São Francisco é banhada pelo rio de mesmo nome, está localizada a 589 km de Belo Horizonte e está nas proximidades das cidades de Montes Claros e de São João da Ponte.

moradores da zona rural e 12 da zona urbana. Aspecto interessante destacado por Coelho (1999, p. 2) foi a percepção de que, na zona rural, os sanfranciscanos idosos usam a variante *ancê/ancêis*. Essa observação levou a autora à conjectura da hipótese de que a forma *você* e suas variantes não estejam sendo empregadas da mesma forma na cidade e no campo, conforme dados expostos de (3) a (6).

- (3) ...se'eu te pagá três pra *você* batê embaixo... (E-22)
- (4) ...eu achei que *cê* ia volta logo pá Belorzonte... (E-617)
- (5) ...o que qui *ocê* acha... meu menino tomou bomba... (E-90)
- (6) ... e o oto já ta... inflorado... *ancê* chega lá... (E-520)

Para justificar sua investigação, Coelho (1999) observa que as formas estigmatizadas como “erradas” pelas classes prestigiadas coincidem com as formas que têm herança do português arcaico. Os moradores das áreas mais afastadas da área urbana, especialmente os idosos, tendem a manter em uso essas formas desvalorizadas, em razão das diferenças naturais das dinâmicas entre a zona rural e a cidade (CRUZ, 1999 *apud* COELHO, 1999, p. 3). Essa variação parece não ocorrer apenas na variação morfológica. Por exemplo, o pronome *você* é classificado como definido na gramática tradicional, embora seja produtivo o seu emprego de forma indefinida, como no exemplo abaixo (ILARI *et al.*, 1996 *apud* COELHO, 1999, p. 3).

- (7) “Antigamente *você* ia ao Cine Ipiranga, eram umas poltronas ótimas tinha lá em cima *você* ficava bem acomodado.” (DID-SP-234:578-579)

Nessa direção, a forma *você* tem cedido lugar à forma *cê* em interpretação definida:

- (8) “[...] De que lado *cê* tá? (E-3)” (COELHO, 1999, p. 4).

A investigação realizada por Coelho (1999) tem como pressuposto o modelo sociolinguístico teórico-metodológico de Labov (1972) para a análise de 24 entrevistas orais, gravadas a partir da sua condução à luz de um roteiro com temas variados. A transcrição das entrevistas foi realizada com base nas normas do Projeto NURC/SP. Trata-se de 629 dados submetidos ao Pacote de Programas Varbrul para os cálculos das frequências de usos e dos pesos relativos. Os resultados evidenciaram a forma *cê*, em 56% das ocorrências, seguida das formas *ocê* (23%), *você* (20%) e *ancê* (1%). As frequências indicaram que a forma *cê* já está implementada no PB. Os poucos dados (sete ocorrências) de *ancê* foram coletados da fala de

idosos não escolarizados. Devido aos reduzidos dados da variante *ancê*, esta variante foi excluída da análise. Tendo em vista que havia três variantes a serem calculadas (*você* ~ *ocê* ~ *cê*), as limitações do Varbrul não possibilitaram a rodada com três variáveis, para a identificação dos índices probabilísticos da regra variável em questão. Quanto aos pesos relativos, os resultados dos fatores internos indicaram que a variante *cê* é a mais relevante e que tende a se manter produtiva na *função de sujeito*. Apenas o fator *função de objeto* favoreceu a variante *você* e nenhum fator favoreceu a variante *ocê* (COELHO, 1999, p. 58). Em relação aos grupos de fatores sociais, a variante *você* foi favorecida pela *faixa etária*, com PR 0.45, em oposição ao PR 0.32 para a variante *cê*, e ao PR 0.23, para *ocê*. Os jovens mostraram-se propícios ao uso das formas *ocê* (PR = 0.40) e *cê* (PR = 0.39), e desfavoreceram o *você* (PR = 0.21). Dentre os idosos, as formas *você* (PR = 0.36) e *ocê* (PR = 0.37) mostraram-se profícuas, enquanto a forma *cê* (PR = 0.27) mostrou-se desfavorecida. Ao colocar-se em discussão o fator *classe social*, foi possível observar que a forma *você* (22% e PR = 0.29) tem frequência menor do que a variante *ocê* (26% e PR = 0.38), mesmo na classe privilegiada. Quanto à frequência, o fator *procedência geográfica* indicou maior percentual para a variante *cê* (57%, na zona urbana; e 58%, na zona rural). No entanto, a forma *cê* (PR = 0.26, na zona urbana e PR = 0.34, na zona rural) apresentou menor PR do que a forma *ocê* (PR = 0.49), na zona rural, e menor PR do que o pronome *você* (PR = 0.55), na zona urbana. Os resultados da análise proposta por Coelho (1999) apontaram a variante *cê* como mais produtiva do que o *você* e do que *ocê* e *ancê*. Além disso, o *cê* demonstrou alta produtividade nas zonas urbana e rural.

Os trabalhos revisados neste capítulo dão uma visão geral a título de amostragem do que já tem sido feito com vistas, principalmente, ao mapeamento do PB em sua expressão vernacular, especialmente em terras mineiras. Nesse sentido, a Imagem 2 indica as localidades mineiras de Belo Horizonte, Ressaquinha, Arcos, Uberlândia, São Francisco e Lontra, cujos dados linguísticos serviram como campo de pesquisa das formas pronominais de 2SG em Minas Gerais.

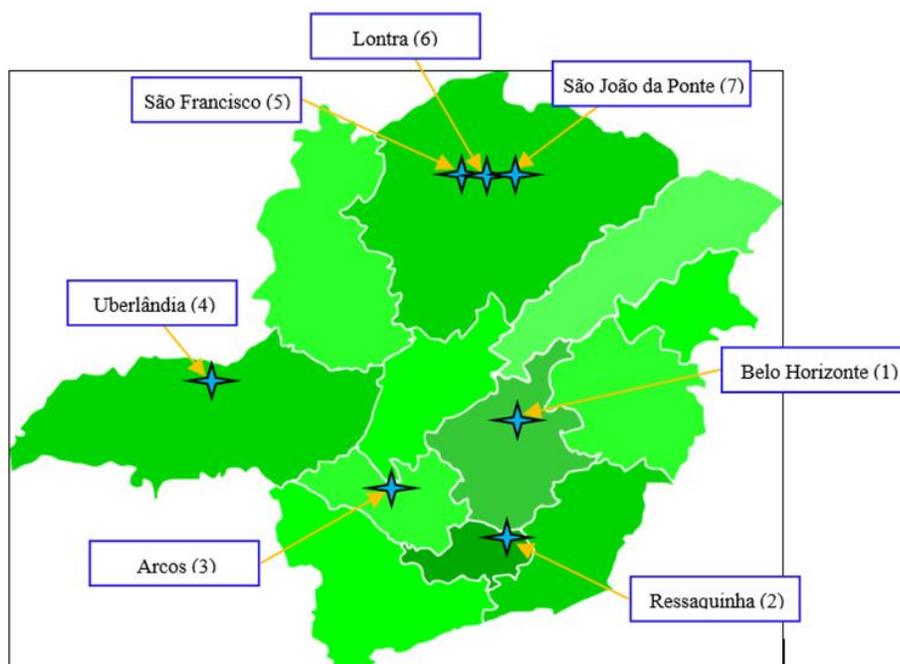


Imagem 2: Pesquisas realizadas sobre as variantes *Tu* e *Você* em Minas Gerais⁹
 Fonte: Mapa das mesorregiões de Minas Gerais (adaptado).

Por fim, neste capítulo, tem-se a preocupação com o delineamento de um panorama das formas de referência ao sujeito de 2SG. Nesse percurso, como em uma “odisseia”, o pronome *você* sofreu transformações (formais, semânticas e pragmáticas) no processo histórico pelo qual passou o PE e chegou ao Brasil com as primeiras naus que aqui aportaram. Em terras brasileiras, as formas *tu* e *você* tiveram os seus usos expandidos em função também das configurações histórico-sociais e culturais que as embasavam dentro e fora do espaço geográfico mineiro. Descrever esse processo é um grande desafio para os linguistas, embora muitos estudos já tenham sido feitos, como é o caso do *mapeamento linguístico* do PB proposto por Scherre *et al.* (2015).

⁹ O mapa das áreas sobre as quais já foram realizadas pesquisas a respeito da 2SG em Minas Gerais foi elaborado a partir de imagem da *internet* adaptada aos dados da pesquisa. Imagem do mapa disponível em <https://pt.wikipedia.org/wiki/Ficheiro:Mesorregi%C3%B5es_de_Minhas_Gerais.svg>. Acesso em 28 jun 2020. Pesquisas revistas no capítulo 1 deste trabalho: 1 (PERES, 2006), 2 (SILVA, 2017), 3 (GONÇALVES, 2008), 4 (HERÊNIO, 2006), 5 (COELHO, 1999), 6 (REIS, 2019) e 7 (MOTA, 2008).

SÍNTESE DO CAPÍTULO:

Neste capítulo 1, percorre-se, à luz de Faraco (2017 [1996]), a história de formação do *você*, cujas evidências morfossintáticas, semânticas e pragmáticas revelam tratar-se de um fenômeno linguístico encaixado nas matrizes linguística e social. Assume-se como ponto de partida o fato de figurarem, no sistema latino tardio, o *tu* e o *vós*, sendo o *tu* empregado em contextos menos formais e o *vós* para a referência a apenas um interlocutor ou a mais de uma pessoa, independentemente do grau de formalidade (FARACO, 2017 [1996], p. 116).

A estrutura social portuguesa passou por um processo de reorganização, com a ascensão econômica e social da burguesia, que começou a disputar espaço de poder com o Clero e a Nobreza. Nesse contexto, o poder do Rei e da Aristocracia rural começou a enfraquecer (FARACO, 2017 [1996], p. 116), dando origem à necessidade de se criar uma forma tratamental que fosse exclusiva para deferência ao Rei. Nesse sentido, surgem, através das *leis de cortesia*, o *vossa mercê* e outras formas treatmentais específicas para a Nobreza. Na perspectiva de Santos Luz (1956 *apud* FARACO, 2017 [1996], p. 118), o *vossa mercê* teria sido a forma tratamental mais antiga de Portugal, cf. registrado no texto das Cortes de 1331.

A forma *vossa mercê* perdeu prestígio e deixou de ser empregada para deferência exclusiva ao Rei. Com o passar do tempo, passou a ser usada para referência aos integrantes das classes superiores que estavam abaixo do monarca. E assim o *vossa mercê* foi se popularizando até alcançar a Plebe, o que aponta para o seu desgaste fonético e semântico, entre os séculos XV e XVI, e a conseqüente origem do *você* através de uma dinâmica social de cima para baixo (*change from above*), cf. Labov (1994, *apud* RUMEU, 2013, p. 22-23). Resiste no *você* a referência semântica à 2SG em concordância formal com a 3SG, cf. Lopes *et al.* (2018), Souza (2012), Rumeu (2015), Lopes & Rumeu (2007).

Ainda neste capítulo de revisão do tema, fez-se um percurso exploratório da capital mineira para o interior, abordando sete pesquisas realizadas em Minas Gerais com foco nas formas treatmentais de 2SG *tu* e *você* (e suas variantes). Peres (2006), ao investigar a produtividade das formas variantes *você*, *ocê* e *cê* em Belo Horizonte, embasada em análises nos âmbitos do tempo aparente e do tempo real, constatou a *função* sujeito, a *contigüidade ao verbo*

(o contexto pré-verbal), a *referência definida* como contextos favorecedores do *cê*, ao passo que a *focalização contrastiva* impulsionou o *você*.

O estudo de Herênio (2006) voltou-se também para as variantes *tu* e *você*, o que lhe permitiu evidenciar o uso categórico do *você* em Uberlândia (MG). Gonçalves, em 2008, também se volta à referência ao sujeito de 2SG através das formas *você*, *ocê* e *cê* na fala mineira de Arcos (MG), o que o levou a constatar que *você*, *ocê* e *cê* estão em variação estável e que a forma *você* se deixa evidenciar no âmbito da *Semântica do Poder*, ao passo que as formas *ocê* e *cê* funcionam na dinâmica social da Solidariedade (GONÇALVES, 2008, p. 228).

Mota (2008) investigou a variação dos pronomes *tu* e *você* no português oral de São João da Ponte (MG). O resultado da pesquisa revelou que o *tu* coexiste com o *você* no norte de Minas. As pesquisas de Silva (2017), voltada para a referência à 2SG, em Ressaquinha, de Reis (2019), voltada para Lontra, e de Mota (2008), voltada para a fala mineira de São João da Ponte, revelam dados do *tu* no estado de Minas Gerais, onde prevalece o *você*, cf. discutido por Scherre *et al.* (2015).

A investigação de Coelho (1999) cujos dados são os da fala de São Francisco teve a sua análise também voltada para o *você* e suas formas variantes *ocê* e *cê*. Os dados evidenciaram a alta produtividade do *cê-sujeito* em relação às formas *você* e *ocê*.

Uma vez revisto o caráter variável das formas de referência ao sujeito de 2SG que está expresso não somente na história de formação do *você*, mas também na fala mineira de sincronias mais recentes, constata-se a prevalência do *você* em meio às raras ocorrências do *tu* detectadas nos trabalhos analisados.

CAPÍTULO 2. FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA

Neste capítulo 2, volta-se o foco não só para a apresentação dos encaminhamentos teórico-metodológicos direcionados às especificidades da Sociolinguística Histórica (HERNÁNDEZ-CAMPOY & SCHILLING, 2012) e da *Teoria do Poder e da Solidariedade* (BROWN & GILMAN, 1960), mas também para as descrições das amostras (missivas históricas mineiras), dos parâmetros da análise linguística e do trabalho com o RStudio e com o RBrul.

2.1 O método de trabalho no âmbito da Sociolinguística Histórica: questões e soluções.

No âmbito da Sociolinguística cabe ao pesquisador controlar também os fatores sociais, partindo do entendimento de que fenômenos de variação e mudança são influenciados por fatores linguísticos e extralinguísticos. Ao considerar a relevância dos fatores sociais, a investigação sociolinguística busca o diálogo com outros ramos da Linguística e de outras áreas de estudo, tais como a Antropologia linguística, a Psicologia social e a Etnografia que contribuem, por exemplo, com o emprego de técnicas de análise de fontes documentais históricas. Nesse sentido, destacam-se os estudos de investigadores britânicos e americanos com foco na língua inglesa, tais como William Labov, Peter Trudgill, Jack Chambers, Walt Wolfram e Ralph Fasold, cf. Conde Silvestre (2007, p. 11), importantes referências para análise de fenômenos de variação e mudança.

Esta investigação parte do pressuposto de que os estudos sociolinguísticos de sincronias recentes podem fomentar os estudos de sincronias passadas, cf. Labov (1994), e orientar as investigações baseadas em fontes históricas. Assim, a Sociolinguística Histórica promove o diálogo entre o presente e o passado, de modo que a investigação do contexto em que se desenvolve o fenômeno de mudança linguística em andamento na atualidade pode ajudar a explicar o passado, cf. Labov (1975 *apud* CONDE SILVESTRE, 2007, p. 10). A ideia é a de que os impulsos de variação e mudança que influenciam um fenômeno linguístico do presente podem também ter atuado em sincronias passadas como evidência do Princípio do Uniformitarismo

(LABOV, 1994). Com origem na Geologia, a partir da noção do uniformitarismo da mudança é possível inferir determinado conhecimento a respeito de algum fenômeno geológico do passado, à luz da observação do mesmo fenômeno no momento presente (CHRISTY, 1833; 1983 *apud* CONDE SILVESTRE, 2007). Ao aplicar tal princípio aos estudos no âmbito da Sociolinguística, é possível inferir que determinado fenômeno de variação linguística observado na atualidade pode refletir sua ocorrência no passado, conforme Milroy (1992) e Labov (1994).

Conde Silvestre (2007, p. 35) discute os desafios enfrentados pelo linguista-pesquisador que se orienta por análises realizadas à luz da Sociolinguística Histórica. Um deles é o de fazer o melhor uso de “dados ruins” (LABOV 1994). Isso quer dizer que cabe ao linguista aproveitar, da melhor forma possível, da produção escrita dos informantes de sincronias passadas, os registros que sobreviveram à ação do tempo. Além das limitações impostas pela “sobrevivência” das fontes históricas que resistiram à ação do tempo, ainda há questões relacionadas, por exemplo, à quantidade dos manuscritos históricos e aos níveis sociais e intelectuais dos escreventes. Os Estudos Labovianos das décadas de 1950 e 1960 para o inglês¹⁰ discutem a respeito da dificuldade de se reconstruir variáveis independentes históricas plenamente confiáveis, que poderiam ter influenciado processos de variação ou mudança da língua em períodos históricos, tais como a estrutura socioeconômica e as relações interpessoais (LABOV, 1994 *apud* CONDE SILVESTRE, 2007, p. 13). Nesse sentido, Conde Silvestre (2007, p. 39) argumenta que, na pesquisa sincrônica, o linguista pode intuir que, em determinada comunidade, há uma mudança em andamento e os resultados dos seus estudos podem apontar ao encaminhamento da mudança, embora não possa concluir, com exatidão, a respeito da direção e do resultado final da mudança. Numa investigação histórica, entretanto, o pesquisador já sabe o resultado final do processo de mudança, bem como a reconstrução do processo de variação e mudança em si, o que passa pela reconstrução não só dos perfis sociais dos informantes, mas também das estruturas sociais de sincronias passadas (CONDE SILVESTRE, 2007, p. 39).

No âmbito da Sociolinguística Histórica, cabe ao linguista-pesquisador fazer o melhor uso dos registros históricos que restaram de sociedades passadas, mesmo que tenham restado apenas fontes escritas deterioradas e incompletas, mas que, ainda assim, podem ser muito úteis

¹⁰ Segundo Veloso (2014, p. 1742), no XVII *Congreso Internacional Asociación de Lingüística y Filología de América Latina* (ALFAL 2014), realizado em João Pessoa (PB), Labov (1966) investigou o inglês da cidade de Nova York e atestou que o uso das variáveis linguísticas estava associado às categorias sociais primárias, tais como classe socioeconômica, sexo, idade, entre outras, variáveis sociais cuja descrição possibilitou a observação de um padrão regular e replicável.

para a explicação de fenômenos linguísticos contemporâneos (LABOV, 1994). Nas palavras de Conde Silvestre (2007), “*el material de que dispone el investigador en sociolingüística histórica está compuesto fundamentalmente de textos escritos en el pasado que se han conservado por azar*” (CONDE SILVESTRE, 2007, p. 42-44), o que vai ao encontro do objetivo principal deste trabalho dissertativo: a análise das estratégias de referência ao sujeito de 2SG (*vossa mercê, você, tu*) a partir das relações sociais que as embasaram nas amostras da produção escrita mineira de sincronias passadas (séculos XIX e XX, principalmente).

A valorização das fontes orais em oposição às fontes escritas desconsidera que as diferenças entre oralidade e escrita se devem, primeiramente, à substância de cada manifestação gráfica ou fônica e, em segundo lugar, à necessidade comunicativa de cada sistema linguístico, seja oral, seja escrito. A respeito das manifestações falada e escrita, Conde Silvestre (2007, p. 43) destaca as principais diferenças estruturais, gramaticais e léxicas:

QUADRO 1
Diferenças entre comunicação oral e escrita

COMUNICAÇÃO ORAL ¹¹	COMUNICAÇÃO ESCRITA
<ul style="list-style-type: none"> - desenvolvida em uma dimensão temporal, de forma dinâmica e efêmera; - requer a presença dos participantes na interação comunicativa; - caráter mais espontâneo e flexível do discurso oral (repetição de expressões, frases sem terminar ou sem reiniciar sem afetar a comunicação); - as ambiguidades podem ser superadas considerando o contexto comunicativo e a relação entre os interlocutores; e - uso de expressões dêiticas que exigem um contexto imediato para serem compreendidas. 	<ul style="list-style-type: none"> - manifestada no espaço, de forma estática e permanente; - geralmente é produzida quando o emissor não está em contato direto com o receptor; - construção mais cuidada e estável da mensagem emitida; - não há um contexto situacional que especifique as circunstâncias em que o texto foi elaborado, caracterizando a diferença entre o contexto de produção e o contexto de recepção da mensagem; e - caráter estilístico mais formal da linguagem, inspirada na norma padrão e em expressões típicas de um sistema ou outro.

Fonte: Conde Silvestre, 2007, p. 43.

Após apontar as diferenças acima relacionadas, Conde Silvestre (2007, p. 44) argumenta que a comunicação escrita independe da oral e não registra a oralidade. A respeito

¹¹ Esta pesquisa busca a caracterização da produção escrita informal de indivíduos letrados no Brasil dos séculos XIX e XX, à luz de Rumeu (2013, p. 58), considerando as restrições impostas aos trabalhos baseados em registros escritos de sincronias passadas. No entanto, é válido destacar que, além das diferenças entre comunicação oral e escrita (quadro 1) propostas por Conde Silvestre (2007, p. 43), um outro aspecto deve ser acrescentado na descrição da comunicação oral. Trata-se da importância da prosódia, que “se constitui como o veículo necessário das ilocuções”, cf. Raso, Mello e Mittmann (2014, p. 47). Sobre esse assunto, Mello (2014, p. 30) indica alguns parâmetros que caracterizam a fala espontânea, entre os quais estão o canal comunicativo, o contexto sociológico e as características sociolinguísticas dos interlocutores (sexo, idade, origem geográfica etc.).

dessa ideia, o autor defende que a língua escrita pode ser analisada independentemente da língua oral e que, embora haja grandes diferenças entre escrita e oralidade, a língua escrita não é isenta de variabilidade, pois a variação linguística se manifesta tanto no meio oral, quanto no meio escrito (BIBER, 1988 *apud* SILVESTRE, 2007, p. 44). Adverte-se, por outro lado, quanto ao fato de que nem todos os textos escritos no passado histórico servem à pesquisa linguística, pois nem todos evidenciam variação, uma vez que podem sugerir influência da norma-padrão, como ocorre, por exemplo, nos textos oficiais e literários. A respeito da seleção de textos escritos propícios às análises no âmbito da Sociolinguística Histórica, destaca-se a *carta pessoal*, gênero textual prevalente nesta investigação.

Registros escritos cuya relación con las variedades vernáculas orales del pasado se puede intuir, como determinadas expresiones que aparecen en cartas privadas o en los diarios personales, especialmente si sus autores son hablantes semialfabetizados en cuyo caso la influencia del estándar es menor.

(SCHNEIDER, 2002 *apud* CONDE SILVESTRE, 2007, p. 46)

Alinhada à proposta de Schneider (2002) e de Conde Silvestre (2007), tem-se esta pesquisa, que embasada em cartas mineiras está munida de alguns protocolos metodológicos, tais como as buscas não somente por registros de nascimento dos escreventes, mas também por dados genealógicos levantados nas próprias cartas pessoais do *corpus* em análise, para a constatação de que se trata de missivas produzidas por punhos brasileiros, mais especificamente, mineiros. Considerando que o foco desta dissertação são as formas de referência ao sujeito de 2SG e as relações sociais que as impulsionam, exemplifica-se, em (9), cf. Imagem 3, trecho da carta pessoal através da qual o autor opta por se dirigir ao seu amigo (CA) através do *você*.

(9) “[...] É da maior necessidade que *você* o conclúa e publique, contribuindo para que se rectifique o conceito actual do romance entre nós. [...]”. (CDA. RJ, 04.08.1936.)

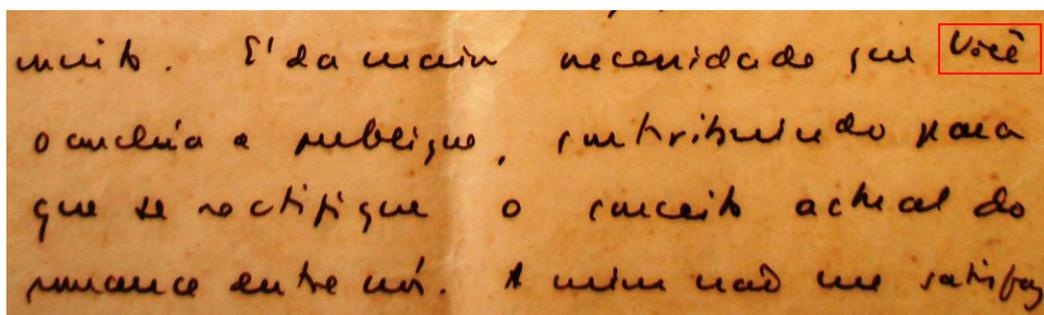


Imagem 3: Dado de *você*-sujeito em carta pessoal (amizade).
(CDA. RJ, 04.08.1936.)

Por outro lado, nessa mesma carta, que é de *você-sujeito* (imagens 3 e 5), o autor opta por tratar o seu interlocutor, em pleno século XX, por *vosmecê* (imagem 4). Trata-se de uma forma nominal cujo uso não é produtivo no PB do século XX (1936), o que justificou que tal dado não tenha sido levado em consideração nesta análise.

(10) “[...] Sem saber, velho Cyro, *vosmecê* me realizou um milagre homérico. [...]”. (CDA. RJ, 04.08.1936.)

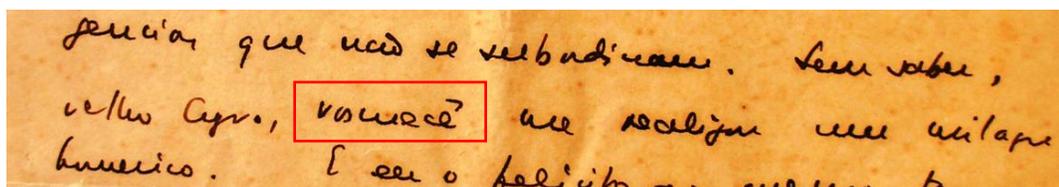


Imagem 4: Dado de *vosmecê* em carta de amizade. (CDA. RJ, 04.08.1936.)

(11) “[...] *Você* sorrirá, quem sabe, dessa ternura alarmante, que não se fartou ainda depois de tantos objectos propostos á sua fruição [...]”. (CDA. RJ, 04.08.1936.)

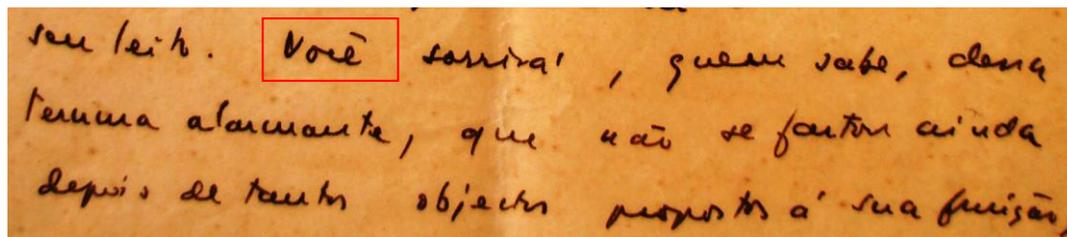


Imagem 5: Uso de *você* em carta pessoal. (CDA. RJ, 04.08.1936.)

Alguns motivos levaram à análise qualitativa da ocorrência da forma “*vosmecê*” (10) e à opção por não computar a sua ocorrência na etapa quantitativa/estatística. Ainda que tenha sido escrita em papel timbrado do gabinete do ministro do Ministério da Educação e Saúde Pública, o teor da carta evidencia que se trata de uma missiva pessoal de amizade, escrita em 04.08.1936, momento em que o *vossa mercê* (*vosmecê*) já estava em desuso no PB. O autor da missiva (CDA) é um ícone da literatura brasileira que parece ter usado a forma *vosmecê* na sentença (10) mais em um tom jocoso do que como expressão da norma de uso da época. Observe-se que no contexto da ocorrência de *vosmecê*, o missivista emprega os termos “velho Cyro”, para se dirigir ao amigo interlocutor, e “milagre homérico”, para se referir a algum feito do interlocutor em episódio anterior, que não está explícito na carta, mas que é do conhecimento de ambos, o que parece indicar uma relação de intimidade entre eles. Em termos quantitativos,

esclarece-se também que dentre o total de 36 cartas de CDA incluídas na amostra, todas as ocorrências de referência ao sujeito de 2SG são da forma *você*, exceto a única ocorrência de *vosmecê* já referida. Para evitar o enviesamento dos resultados, optou-se, pois, por não computar, nesta análise dos dados, a sentença (10), tendo em vista o uso intencional do *vosmecê* em missiva da década de 30 do século XX (1936).

O trabalho de levantamento, codificação e análise de dados históricos é um desafio ao linguista-pesquisador justamente porque não pode ser uma coleta acrítica de dados. Nessa direção, ao ler a carta, o linguista deve levar em conta também, dentre outros aspectos, o período em que se deu a produção da carta, a linguagem empregada pelo autor, a recorrência do dado em análise nas produções do mesmo autor etc. Esse cuidado demanda a atenção do investigador, especialmente quando se pretende utilizar uma grande quantidade de dados, de modo a assegurar a coerência dos resultados voltados para sincronias passadas. Assim sendo, a análise qualitativa dos dados é fundamental para identificar as possíveis inconsistências. Tendo em vista as questões teóricas e metodológicas que se impõem a uma análise linguística embasada em dados históricos, há de se atentar, à luz de Hernández-Campoy & Schilling (2012), principalmente à *autenticidade*, à *autoria* e à *validade social e histórica* das amostras de missivas históricas em análise.

2.1.1 Questões enfrentadas no levantamento de *corpora* históricos confiáveis

A *autenticidade* das fontes está relacionada ao nível de pureza dos textos (HERNÁNDEZ-CAMPOY & SCHILLING, 2012, p. 68) em relação às possíveis evidências de hipercorreções, mistura dialetal e “erros” do escriba, cf. discutido também por Labov (1994). Há de se ter clareza de que as fontes (históricas) escritas não são transcrições da fala em uma sincronia passada. Acrescente-se a isso também a questão de o texto escrito produzido por escreventes cultos (como é o caso das cartas mineiras) ser necessariamente conduzido pelo filtro da norma-padrão. Assim sendo, cabe ao linguista-pesquisador depreender uma possível influência da norma-padrão em textos marcados pela intimidade, como é o caso das cartas pessoais, separando-as, pois, do uso que mais se aproxime do vernáculo do escritor. A Imagem 6, transcrita em (12), parece exemplificar o titubeio do autor entre as formas *tu* e *você*, o que parece

evidenciar o *você* rasurado como a manifestação do vernáculo do autor para a referência ao interlocutor (sua noiva).

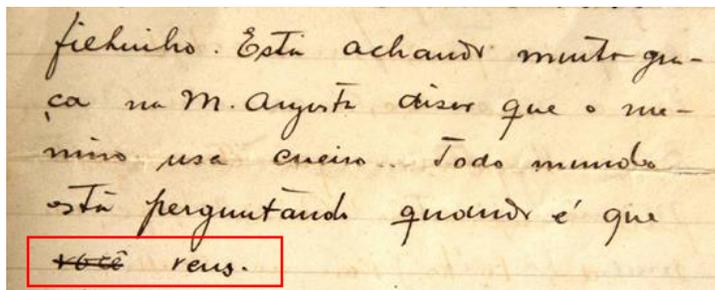


Imagem 6: Evidência de dúvida do autor entre as formas *tu* e *você*. (JAVS. Machado, MG, 16.03.1926.)

(12) “[...] Está achando muita graça na M. Augusta dizer que o menino usa cueiro... Todo mundo está perguntando quando é que ~~você~~ vem. [...]” (JAVS. Machado, MG, 16.03.1926.)

Na tentativa de entrever-se o vernáculo à luz de dados escritos, são usadas, neste trabalho, fontes manuscritas originais e autógrafas que permitam reconhecer ou reconstruir a autenticidade dos registros históricos, imprimindo confiabilidade não somente aos dados históricos, mas também às conclusões respaldadas na análise de tais dados. Ainda em relação ao cuidado do linguista-pesquisador quanto à *autenticidade* das fontes, as cartas utilizadas na análise foram transcritas à luz das conservadoras normas de transcrição propostas por Rumeu (2013). A transcrição conservadora da carta manuscrita é precedida por uma ficha de identificação do documento em que estão registrados o acervo, o local e a data da carta, o nome do autor, a data e o local de nascimento, cf. a Imagem 7.

PHPB – Século XX – Não-oficiais / Carta. Edição: RUMEU, Márcia.; REZENDE, Martha.; FUSCO, Beatriz.
Local e data: Belo Horizonte, 13 de novembro de 1961.
Fonte: Acervo dos Escritores Mineiro.
Autor: Henriqueta Lisboa.
Local e Data de Nascimento: Lambari, MG, 15.07.1901, cf. Fls <u>114V^o</u> , sob o n.342 do Livro A-1 do Registro Civil de Pessoas Naturais da Comarca de Lambari.
Nacionalidade: Brasileira.
Naturalidade: Minas Gerais.
Referência: Coleção Henriqueta Lisboa
Idade (na ocasião da escritura desta carta): 60 anos.
Conteúdo: Henriqueta Lisboa agradece favores feitos por seu irmão José Carlos.

Imagem 7: Ficha de identificação de carta pessoal (HL. BH, 13.11.1961.)

A questão da *autoria* dos documentos que compõem o objeto de estudo da investigação também é um desafio a ser enfrentado pela pesquisa em Sociolinguística Histórica (HERNÁNDEZ-CAMPOY & SCHILLING, 2012, p. 68). Se o propósito principal é resgatar traços do vernáculo dos escritores de sincronias passadas, é importante ter-se certeza acerca do punho que redigiu o documento, o que justifica a busca por cartas autógrafas. No processo de coleta dos dados, foram incluídas cartas autógrafas de missivistas brasileiros nascidos em Minas Gerais, à exceção de dois missivistas. Um deles é o italiano Ernesto Tassini¹², que chegou da Itália ao Brasil, mais especificamente a Minas Gerais, ainda no início da juventude, aos 18 anos, ao final do século XIX. Outro missivista Bernardo Nascentes Pinto, que nasceu no Rio de Janeiro (SILVA LEME, 1904, p. 341), casou-se com a mineira Anna Casimira, filha do Barão de Cocais, e morou em Minas Gerais (Ouro Preto, Paciência e São Vicente), como gestor dos negócios da família, como se pode observar nas cartas analisadas nesta investigação. Nas amostras de missivistas analisadas neste trabalho, todas as cartas são autógrafas (cf. Imagem 8), o que pode ser constatado pela comparação entre a letra da redação da carta e a letra de sua assinatura cujas evidências também, por vezes, podem ser resgatadas em outras missivas do mesmo autor.

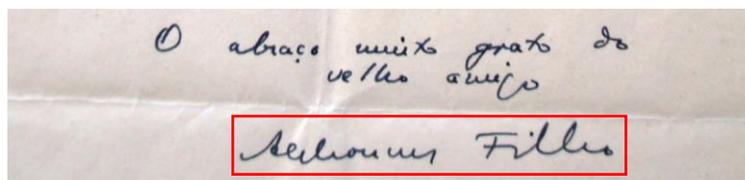


Imagem 8: Assinatura do autor em carta de amizade (documento autógrafo). (AGF. BH, 31.12.1947.)

A *validade social e histórica* (HERNÁNDEZ-CAMPOY & SCHILLING, 2012, p. 70) remete à relevância de se saber não só acerca da posição social dos informantes de sincronias pretéritas (à exceção dos redatores cultos com alguma progressão social), mas também acerca da estrutura social da qual faziam parte. Isso quer dizer que se faz necessária a reconstrução não somente da história de vida dos escreventes, mas também do contexto social dos missivistas de sincronias passadas. Assim sendo, a reconstituição do perfil biográfico dos informantes pressupõe

¹² Segundo consta no livro de registro da Hospedaria Horta Barbosa, a família Tassini chegou à cidade de Juiz de Fora (MG), no dia 9 de novembro de 1894, cf. o Livro de Registro de Entrada e Saída dos Imigrantes e na Lista de Bordo do Vapor. O então jovem Ernesto Tassini, após morar alguns meses em Juiz de Fora, seguiu para Belo Horizonte, onde trabalhou na área de marcenaria e constituiu família, conforme Alves (2008, p. 141).

o levantamento de informações acerca do local de nascimento, da filiação, da idade no momento de escritura da carta, da classe social e da escolarização.

Para ilustrar uma fonte de coleta de dados biográficos, a imagem (9) demonstra que a própria carta pessoal pode fornecer pistas a respeito das características do autor. No caso do missivista BNP, genro do Barão de Cocais, poucos dados biográficos sobre ele sobreviveram ao tempo ou estão disponíveis para a reconstrução do perfil social do informante. Nesse sentido, a constatação do emprego produtivo de abreviaturas, por exemplo, em todas as cartas do BNP permite inferir que se trata de um informante habilidoso em relação à expressão escrita. Nesse sentido, a própria carta pessoal apresenta-se como uma importante fonte histórica acerca do grau de intimidade do redator com a escrita. Com base nas datas das missivas (2ª metade do século XIX) e também no fato de o interlocutor ser o seu sogro, o Barão de Cocais, é possível fazer algumas inferências. A primeira é que se pode marcar o nascimento do BNP na primeira metade do século XIX, uma vez que a maioria das suas cartas é da 2ª metade do século XIX. A segunda diz respeito ao fato de que o conteúdo das cartas, que majoritariamente trata do comércio e da lavra de minério, permite a dedução de que se trata de um autor em sua fase adulta. Mesmo sem as devidas comprovações documentais principalmente acerca do grau de letramento do autor, é possível observar, com base em parâmetros da época da escritura (século XIX), nos termos de Lopes & Rumeu (2018), que se trata de um ¹³redator que parece ter mantido um maior nível de contato com modelos históricos de escrita. Corroboram essa inferência o uso abundante de índices abreviativos e o conhecimento da estrutura composicional do gênero carta pessoal, principalmente em relação aos registros do local, da data e da captação de benevolência, cf. Paredes Silva (1988), Bakhtin (1997), Lopes & Rumeu (2018). No corpo da carta, podem ser observados o vocabulário variado e as narrativas coesas em relação aos contextos tratados nas cartas de BNP. Acrescente-se ainda que a pontuação e a caligrafia de BNP parecem evidenciá-lo como um escrevente habilidoso em suas produções escritas. Ainda no que se refere à sua produção escrita, o emprego recorrente de abreviaturas parece ser mais um indício de que o missivista possuía um bom domínio da expressão escrita. Observe-se, à título de exemplificação, na imagem 9, o uso da fórmula “*Meu Prezadissimo Pai, Compadre, e Amigo*” como expressão

¹³ Cf. Lopes e Rumeu (2018), o grau de escolaridade de períodos históricos pode também ser observado com base nos parâmetros de escrita em uso na época em que o texto foi escrito. A análise pode também se basear no grau de letramento do redator, que pode ser erudito, pouco hábil ou inábil, ou se basear no grau de contato com modelos de escrita históricos (MARQUILHAS, 2000 *apud* LOPES & RUMEU, 2018, p. 151).

cerimoniosa de referência ao sogro, o Barão de Cocais, opção devidamente prevista pelos Manuais de abreviaturas, tais como Borges Nunes (1981) e Flexor (2008).

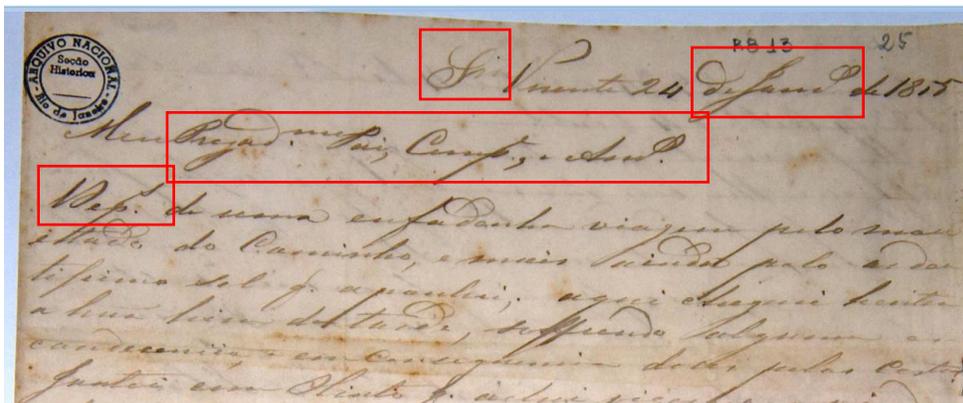


Imagem 9: Uso de abreviaturas como indício da habilidade do autor em sua expressão escrita. (BNP. Sam Vicente, 24.01.1855.)

Consideradas essas questões (*autenticidade, autoria, validade social e histórica*), assume-se a preferência pelo trabalho com cartas pessoais (familiares, amorosas, de amizade) e cartas de comércio produzidas por escreventes mineiros que conduziram suas vidas (ainda que por um período) em MG. Nesse sentido, acredita-se na possibilidade de se entrever, através do acentuado nível de intimidade das missivas em análise, o vernáculo dos mineiros cultos.

2.2 O trabalho com cartas históricas: as cartas mineiras dos séculos XIX e XX

Rumeu (2013), embasada em cartas familiares dos séculos XIX e XX, analisa a dinâmica *tu/você* nas cartas cariocas oitocentistas e novecentistas. Ao argumentar acerca do desafio do linguista-pesquisador por confiáveis fontes históricas do PB, a autora conduz a sua discussão à luz de Lucchesi (1988 *apud* RUMEU, 2013, p. 53) acerca da polarização de normas, tendo em vista a heterogenidade social e a pluralidade de normas. Em um extremo da realidade linguística estão as normas populares que caracterizam o português não-padrão, utilizado, de forma geral, pelos falantes não-escolarizados. No outro extremo, estão as normas cultas que representam o português padrão, acessível em sua maioria à elite social. Assim sendo, Rumeu

(2013) volta-se ao português *standard* através dos registros da escrita culta do PB dos séculos XIX e XX, embasada na análise de cartas pessoais cariocas da família Pedreira Ferraz-Magalhães (RUMEU, 2013, p. 53-55). Inspirado em Rumeu (2013), que investigou a história do pronome *você* no PB, em cartas pessoais cariocas, está também este estudo fundamentado em missivas pessoais (principalmente) de mineiros cultos, orientado pelos princípios da Sociolinguística Histórica para a análise das dinâmicas sociais de Poder e de Solidariedade (BROWN & GILMAN, 1960) que conduziram os usos das formas de referência ao sujeito de 2SG (*vossa mercê, você e tu*).

Em termos estruturais, a *carta pessoal* é constituída na tríade *autor – leitor – tema íntimo* (RUMEU, 2013, p. 60) como expressão padronizada, há muito tempo no mundo ocidental, de seções específicas para o local, a data, a saudação inicial, o corpo da carta, a despedida e a assinatura, cf. discutido também por Lopes *et. al.* (2010, p. 242). Em termos da semântica da carta pessoal, trata-se de uma expressão escrita mais “descuidada” em relação às prescrições normativas, possibilitando, mesmo que difusamente, a detecção de traços da oralidade. No entanto, o linguista-pesquisador precisa ter em mente que, da mesma forma que as cartas pessoais aproximam-se, em alguma medida, da fala espontânea, a oralidade também pode ser influenciada por traços da expressão escrita. Embora se caracterizem pela informalidade, as cartas pessoais podem não evidenciar tão claramente os traços da oralidade do PB dos séculos XIX e XX, em razão da grande influência da norma-padrão em função de serem concretizadas em meio escrito. Devido a essa restrição, ao invés de conduzir a pesquisa na percepção da oralidade através da escrita, Rumeu (2013) direciona a sua investigação para a “caracterização da produção escrita informal de indivíduos letrados no Brasil dos séculos XIX e XX” (RUMEU, 2013, p. 58).

Partindo dessa concepção de *carta pessoal* caracterizada pela produção escrita informal dos missivistas, as fontes históricas resgatadas representam um desafio à investigação, pois o desgaste provocado pelo tempo, as condições de armazenamento, o extravio de folhas, o vocabulário utilizado pelos informantes são importantes questões que devem ser enfrentadas e superadas. Assim sendo, acredita-se ser importante ancorar as análises linguísticas em cuidadosas e conservadoras edições (edições fac-similares e diplomáticas de manuscritos históricos) preparadas para embasar estudos linguísticos como são as cartas mineiras que fundamentam esta dissertação e também conduziram, por exemplo, as análises de Lopes & Rumeu (2015), Cruz (2017), Cardoso (2017) e Carvalho (2020).

Uma vez justificada, ainda que brevemente, a opção pelo trabalho com cartas históricas, passa-se à descrição das amostras deste trabalho.

2.3 Descrição das amostras de cartas históricas oitocentistas e novecentistas

Esta pesquisa reúne amostras de cartas pessoais mineiras, formadas por correspondências trocadas nos séculos XIX e XX, ainda que haja tão somente uma carta pessoal do início do século XXI, chegando a um total de 358¹⁴, das quais 324 cartas foram selecionadas com dados de *vossa mercê*, *você* e *tu* em função de sujeito de 2SG. As amostras de cartas reúnem a produção escrita de 57 missivistas que, em sua maior parte, são informantes cultos, ainda que alguns deles não tenham tido os seus perfis biográficos completamente reconstituídos.

Na sequência, estão minuciosamente descritas as amostras dos acervos históricos, cf. quadro 2. Em cada acervo, as cartas estão distribuídas pela autoria (através da sigla ou do código do autor), pelos locais de nascimento dos autores, pela indicação do(s) ano(s) das cartas, acompanhadas, entre parênteses, pela quantidade de cartas por ano.

QUADRO 2
Amostras históricas selecionadas

ACERVOS ¹⁵	SEXO		PERÍODOS	GÊNERO DA MISSIVA				TOTAL	
	H	M		CARTA		CARTA PESSOAL		AUTORES	CARTAS
				COMERCIAL	AMOR	FAMILIAR	AMIZADE		
AEM	16	13	1911 a 2007	-	06	67	77	29	150
AN	1*	0	1848 a 1858	26	-	-	-	01	26
AP	01	02	1924 a 1969	-	59	01	-	03	60
APCBH	06	01	1890 a 1915	-	-	21	05	07	26
APM	01	0	1868 a 1908	-	05	07	13	01	25
IHGGMG	06	03	1907 a 1986	-	-	24	25	09	49
MAB	03	04*	1930 a 1956	-	01	20	01	07	22
TOTAL	34	23	-	26	71	140	121	57	358

¹⁴ Das 358 cartas, 34 missivas foram excluídas da amostra durante a fase de análise estatística por terem apresentado somente ocorrências imperativas, que foram controladas, mas não incluídas na análise em termos de índices percentuais e probabilísticos, tendo em vista não serem foco desta investigação.

¹⁵ LEGENDA: AEM: Acervo dos Escritores Mineiros; AN: Arquivo Nacional; AP: Acervo Pessoal; APCBH: Arquivo Público da Cidade de Belo Horizonte; APM: Arquivo Público Mineiro; IHGGMG: Instituto Histórico e Geográfico de Minas Gerais; MAB: Museu Abílio Barreto.

2.3.1 O Arquivo Nacional (RJ)

O Arquivo Nacional¹⁶ (AN) faz parte do Ministério da Justiça e Segurança Pública e está localizado na cidade do Rio de Janeiro-RJ. A amostra relativa ao acervo do AN é composta por cerca de 30 cartas de comércio enviadas, no período de 1848 a 1856, por Bernardo Nascentes Pinto (BNP) ao sogro, o Coronel José Feliciano Pinto Coelho da Cunha (JFPCC), mais conhecido pelo título de Barão de Cocais, cf. SOUZA (2013). Do total de cartas disponíveis no acervo, 26 missivas (escritas entre 1848 a 1858) apresentaram ocorrências do *vossa mercê* que são produtivas, nas cartas de Bernardo ao sogro. As cartas de 21.07.1848 e de 07.09.1854¹⁷ são finalizadas com bilhetes de Anna Cassimira Nascentes Pinto (ACNP), esposa do BNP, endereçados ao Barão de Cocais, nos quais há dados de *tu* e de *vós*, cf. imagens 10 e 11.

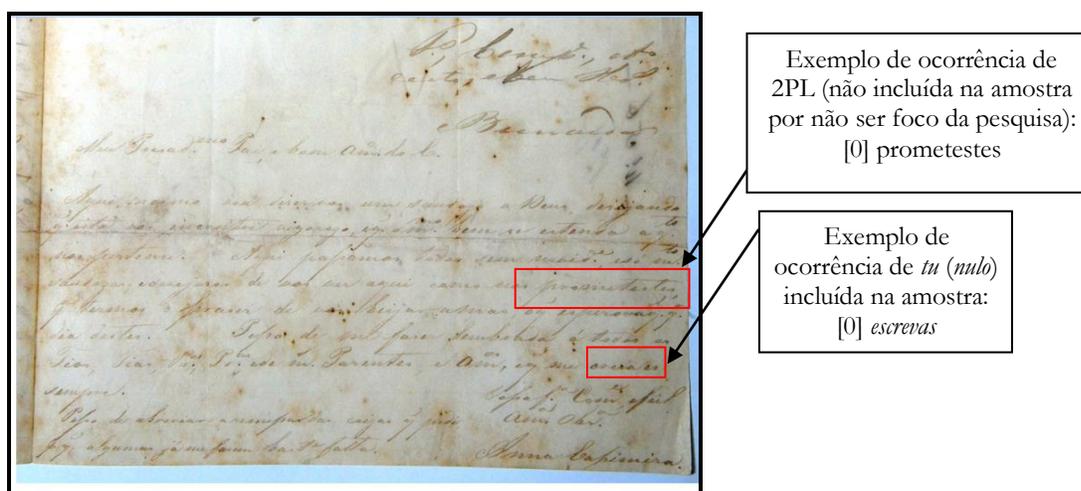


Imagem 10: Bilhete de ACNP endereçado ao pai dela, o Barão de Cocais, escrito logo após o fechamento da carta do marido. (BNP. São Vicente, 07.09.1854.)

¹⁶ O Arquivo Nacional pode ser acessado através da *internet*, por meio do endereço eletrônico <<http://www.arquivonacional.gov.br/br/>>, que permite consultas, tais como a verificação de entrada de estrangeiros no Brasil, registro civil, dentre outras opções.

¹⁷ Arquivo Nacional [R8 1]. BNP [ACNP]. Cocais, 21.07.1848; e AN [R8 7]. BNP [ACNP]. Sam Vicente, 07.09.1854.

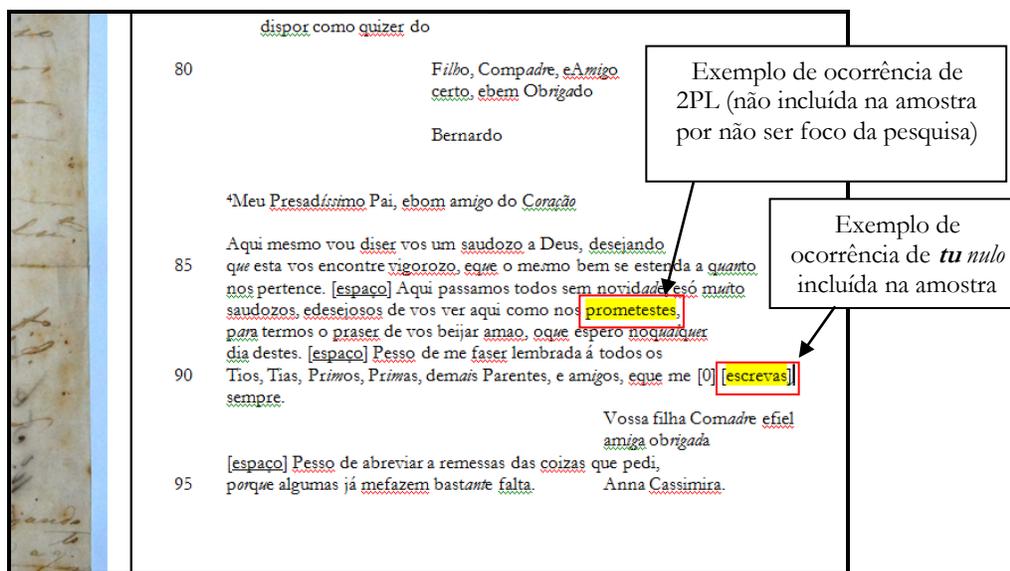


Imagem 11: Transcrição de bilhete de Anna Cassimira endereçado ao pai dela, o Barão de Cocais, escrito logo após o fechamento da carta do marido, BNP. (BNP [ACNP]. São Vicente, 07.09.1854.)

Para que as considerações finais de uma análise linguística sejam historicamente validadas (*validade social e histórica*), é mais do que necessária a reconstrução do perfil biográfico do missivista à luz de ¹⁸Bueno & Barata (2000). As cartas de Bernardo ao Barão de Cocais são cartas de comércio que têm como contexto as relações comerciais com os ingleses, as negociações voltadas para a mineração do ouro em terras brasileiras, os prejuízos financeiros, tendo em vista os revezes da vida econômica e social de Minas Gerais no século XIX, cf. contextualização histórica proposta por Silva (2012).

Nas missivas enviadas por BNP ao sogro, há o registro do local e da data onde as cartas foram redigidas. Na amostra em análise, estão em cena as cartas produzidas em locais como Cocais, São Vicente, Paciência e Machado, todas localizadas em Minas Gerais. Por exemplo, em 21 de julho de 1848, de Cocais, Bernardo escreve ao sogro sobre os negócios relacionados à lavra de minérios na mina. No vocativo da carta, Bernardo, ao mesmo tempo em que demonstra respeito ao sogro, pelo emprego da forma tratamental *Meu Prezadíssimo*, tratando-o como *Pai, Compadre e Amigo*. Na instância de captação de benevolência, o missivista

¹⁸ Segundo Barata & Bueno (2000), os Nascentes Pinto (a família Nascentes Pinto) foram uma antiga e importante família de origem portuguesa que se estabeleceu no Rio de Janeiro no início do século XVIII. A origem dos dois sobrenomes vem do casamento entre Manuel Nascentes Porto e Maria Pinto de Paiva, de cuja união nasceu Manuel Nascentes Pinto. Este último se casou com Antônia Maria, filha de Francisco Lopes Braga, e faleceu em 1697, deixando numerosa descendência do seu casamento (cf. BARATA & BUENO, 2000, p. 1600-1601).

busca atrair a boa vontade do interlocutor e revela uma relação movida pela *Semântica da Solidariedade* com o sogro conforme exemplificado em (13):

(13) [...] *Meu Prezadíssimo Pai, Compadre, e Amigo*¹⁹
Tenho presente a sua de 28 do passado, e com ella a
certeza deque com quanto ha nosso ahi passa sem maior novi
dade, oque Deos queira continue. [...]
 (BNP. Cocais, 21.07.1848.)

O grau de aproximação entre BNP e o Barão de Cocais também pode ser constatado na sentença (14), na qual o autor usa a expressão “Minha mãe” para se referir à sogra. Nessa mesma carta, faz referência à própria esposa pelo diminutivo carinhoso “Anninha”, sentença (15), (Imagem 12) e também através da menção às “Pequenas” para se referir às filhas “Júlia” e “Anna”, o que parece indicar certa intimidade entre BNP e o sogro nas cartas em análise.

(14) [...] Minha Mãe, e sua Familia paixão como sempre, e a mesma Nunca (o que fiz sua recommendação quanto a continuar com os remedios) vai [...] agradecendo suas vizitas [...] (BNP. Cocais, 21.07.1848.)

(15) [...] Anninha tambem escreve, e dirá de si, as Pequenas pasão bem, e falao ambas no vôvô. [...]” (BNP. Cocais, 21.07.1848.)



Imagem 12: Carta de BNP ao sogro, o Barão de Cocais, fazendo referência à esposa e às filhas. (BNP. Cocais, 21.07.1848.)

Em 30 de agosto de 1853, BNP tomou posse como procurador do sogro em relação ao Córrego do Mirguita, desde a nascente, na Serra do Cacao, em todas as suas vertentes e lagoinhas, até a bacia com o Rio das Velhas (BNP. Cocais, 21.07.1848). Esse registro feito pelo próprio BNP em sua carta enviada ao sogro é uma indicação da relação de confiança entre os interlocutores. As mais de 30 cartas de comércio enviadas por BNP ao sogro são evidências de que, em razão da vida política do Barão de Cocais, tanto na gestão, em Cocais, da mineração como em outras questões financeiras, BNP estava à frente e diretamente envolvido nos negócios da família. Nesse sentido, segundo Silva (2012), parece que, motivado pela eclosão da Revolução Liberal em Minas, em 1842, José Feliciano Pinto Coelho da Cunha foi “aclamado pela Guarda

¹⁹ No texto original, o vocativo inicial vem marcado pela presença de abreviaturas, que caracterizam as cartas de Bernardo em todo o texto das cartas: *Meu Prezad^{mo} Pai, Comp^e, e A^o.*

Nacional, em Barbacena, chefe do movimento rebelde e presidente da província” de Minas. Quando o movimento foi derrotado, em agosto de 1842, José Feliciano fugiu para não ser preso pelas tropas de Caxias, “através de uma passagem subterrânea secreta que saía da cozinha de sua residência”, para se esconder nas grutas da serra da Conceição (SILVA, 2012, p. 117-118). Em novembro de 1842, BNP representou o sogro na cobrança de providências, junto ao inglês Jorge Morgan, para a resolução de questões que envolviam as ações da Companhia de Cocais. BNP não obteve sucesso em sua cobrança, tendo em vista que Morgan alegou não ter conseguido vender as ações por preço satisfatório, as quais permaneceram em nome do inglês Edward Oxenford.

Em carta de 1854, pode-se conhecer um pouco mais do lado humano de BNP, pela demonstração de preocupação com a situação de saúde da esposa que parecia ter, naquele contexto da carta, uma saúde debilitada, cf. (16).

(16) [...] Tenho aqui de rogar lhe *que* se Anninha continuar a soffrer, me não ocultar, pois que eu já mais deixaria de o sentir, *athe por que* tarda a noticia boa, ou má eu desejo *que* me deem com exactidão, *que* vivendo na desconfiança de *que* occultão de mim alguã cousa soffrerei *muíto* mais. Como Amigo lhe faço este pedido. [...] (BNP. Sam Vicente, 07.07.1854.)

No comando dos negócios do sogro no Engenho de Cocais, BNP cuidava das questões dos trabalhadores contratados, controle dos escravos, disciplina no serviço, saúde de familiares, empregados e escravos, compra de materiais e de animais, condições de trabalho na mina, entre muitas outras atividades. Mensalmente, por meio de carta, BNP fazia a prestação de contas da produção de minérios (ouro) ao sogro. Por vezes, a produção favorecia os lucros, como o próprio BNP conta em uma longa carta, em fevereiro de 1855, na qual o missivista relata a frustração quanto aos resultados da mina em um mês e a grata satisfação com os lucros no mês seguinte, conforme (17):

(17) Sam Vicente 15 de 7br.º de 1855
 MeuPrezadissimo Pai, Compadre, eAmigo
 [...]

 He pois agora *que* lhe direi
que em Julho apenas colhemos 66¾
 oitavas de ouro e mais 7 vintens!!
 O mez de Agosto porem obtivemos
 o producto de 335 oitavas e 2 vintens *que*
 veio reanimar-nos.
 [...]

 Filho Compadre
 Amigo Obrigadissimo
 Bernardo
 (BNP. Sam Vicente, 10.02.1855.)

A prosperidade de Cocais chamou a atenção de investidores ingleses e talvez essa tenha sido a triste sorte de tão bela e rica região mineira. A respeito da paisagem local, o inglês Gardner não poupou elogios ao dizer que em Cocais

as casas tinham todas um ar de asseio e elegância, caiadas e cercadas de pequenos pomares com laranjeiras, cafeeiros, bananeiras etc. A igreja salientava-se fortemente entre as demais construções, rodeada de altas palmeiras que dão àquele sítio um aspecto verdadeiramente tropical (SILVA, 2012, p. 111).

Segundo Silva (2012, p. 171), o Barão de Cocais era um dos maiores acionistas da *National Brazilian Mining Association* e havia nomeado como seu representante, e legítimo procurador em Londres, o inglês Edward Oxenford. Assim, José Feliciano entregou nas mãos dos ingleses o controle de suas ações e o destino de um patrimônio que jamais voltaria às mãos da família *Pinto Coelho*. Sobre esse assunto, José Feliciano escreveu ao genro Bernardo, em 1861, expressando seu cansaço e insatisfação por estar sendo enganado pelos seus representantes em Londres. Desse modo, o “Barão de Cocais tentou, exaustivamente, resolver a questão referente a seus direitos na companhia” (SILVA, 2012, p. 172).

Em 1861, ele [o Barão de Cocais] recebeu uma carta do genro, que, entre outras coisas, dizia o seguinte: “Paciência, 21 de junho de 1861 [...] principiaremos em dizer que bem estimaremos que se realize o negócio das ações da Companhia de Cocais, e que não sejamos ainda mais desta vez burlados. Vimos a carta do visconde de Barbacena, e temos refletido sobre sua exigência”. Em 9 de julho de 1869, José Feliciano Pinto Coelho da Cunha, abatido pela tuberculose, faleceria em Cocais, desprovido de recursos materiais, sem ter conseguido fazer prevalecer seus direitos perante a companhia inglesa (SILVA, 2012, p. 172).

Anos após a morte do Barão de Cocais, seus descendentes constituíram um advogado inglês para representar seus interesses quanto à herança que teriam direito, tendo em vista as riquezas geradas pelo ouro extraído da Companhia de Cocais. Sem sucesso. José Feliciano Pinto Coelho da Cunha, o Barão de Cocais, faleceu sem jamais ter reavido sua riqueza cuja administração fora confiada aos seus representantes ingleses. Bernardo Nascentes Pinto, por meio de suas cartas endereçadas ao sogro, eternizadas nas linhas traçadas em tantas missivas recheadas de abreviaturas, deixou ao PB a herança linguística marcada pelo arcaico *vossa mercê*. Com o tempo, a herança do Barão de Cocais “adquiriu contornos de lenda e realidade”²⁰. As missivas

²⁰ Cf. Silva, 2012, p. 198.

enviadas por Bernardo ao sogro está sintetizada no quadro 3 como forma de registro das cartas de comércio do acervo do Arquivo Nacional efetivamente utilizadas nesta dissertação.

QUADRO 3
Relação de missivas de comércio do Arquivo Nacional (AN-RJ)

MISSIVISTA	LOCAL E DATA DE NASCIMENTO	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS DE COMÉRCIO
		1848 ²¹			2
		1853			4
		1854	Ascendente		3
BNP	-	1855	(de inferior para superior)	genro-sogro	4
		1856			6
		1857			2
		1858			5
		TOTAL			26 ²²

2.3.2 O Acervo dos Escritores Mineiros

A coleta de documentos históricos passou também pelo Acervo dos Escritores Mineiros (AEM) no qual foram levantadas 150 cartas pessoais, abrangendo 29 ilustres escreventes, tais como Abgar Renault, Cyro dos Anjos, Henriqueta Lisboa, Murilo Mendes, Otto Lara Resende, dentre outras efemérides mineiras. Na sequência, estão descritas as informações relativas à quantidade de cartas por autor, local de nascimento, indicação do seu local e ano de nascimento, ano de redação da carta, relação social, relação interpessoal e a distribuição de cartas em relação ao subgênero da carta, conforme exposto no quadro 4.

²¹ As cartas de 21.07.1848 e de 07.09.1854 contêm bilhetes de ACNP, esposa de BNP, enviados ao pai dela, o Barão de Cocais (relação social ascendente filha-pai).

²² Embora todas as cartas da amostra do Arquivo Nacional tenham Bernardo Nascentes Pinto como remetente, entre as 26 cartas da amostra, há as 2 cartas acima citadas em que se pode acessar bilhetes escritos por Anna Cassimira com dados de *tu*.

QUADRO 4
Acervo dos Escritores Mineiros (AEM)

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	AMOROSA	FAMILIAR	AMIZADE
AGF	Mariana 03.06.1918	1947(2)	Simétrica	amigo-amigo	-	-	-	2
		1955(1)	Simétrica	amigo-amigo	-	-	-	1
		1956(2)	Simétrica	amigo-amigo	-	-	-	2
		1959(1)	Simétrica	amigo-amigo	-	-	-	1
		1961(1)	Simétrica	amigo-amigo	-	-	-	1
		1963 (1)	Simétrica	amigo-amigo	-	-	-	1
		1965 (1)	Simétrica	amigo-amigo	-	-	-	1
		1968 (2)	Simétrica	amigo-amigo	-	-	-	2
		1969(1)	Simétrica	amigo-amigo	-	-	-	1
AL	Lambari 16.08.1906	1931(1)	Simétrica	irmã-irmã	-	-	1	-
ALO	Lambari 22.04.1904	1962 (1)	Simétrica	irmã-irmã	-	-	1	-
AM	Sabará 09.12.1894	1941 (1)	Simétrica	irmão-irmão	-	-	1	-
		1944(2)	Simétrica	irmão-irmão	-	-	2	-
		1945(1)	Simétrica	irmão-irmão	-	-	1	-
		1946(1)	Simétrica	irmão-irmão	-	-	1	-
		NA (1)	Simétrica	irmão-irmão	-	-	1	-
AR	Barbacena 15.04.1901	1926(1)	Simétrica	amigo-amigo	-	-	-	1
		1930(1)	Simétrica	irmão-irmão	-	-	1	-
		1931(1)	Simétrica	irmão-irmão	-	-	1	-
		1935(2)	Simétrica	irmão-irmão	-	-	2	-
		1936(2)	Simétrica	irmão-irmão	-	-	2	-
		1937(1)	Simétrica	irmão-irmão	-	-	1	-
		1940(1)	Simétrica	irmão-irmão	-	-	1	-
		1989(1)	Simétrica	irmão-irmão	-	-	1	-
NA (1)	Simétrica	irmão-irmão	-	-	1	-		
AVP	Campanha 13.09.1957	1962(2)	Ascendente	sobrinha-tia	-	-	2	-
CA	Montes Claros 05.10.1906	1979(1)	Simétrica	amigo-amigo	-	-	-	1
		1981(1)	Simétrica	amigo-amigo	-	-	-	1
		1931(1)	Simétrica	amigo-amigo	-	-	-	1
		1932(2)	Simétrica	amigo-amigo	-	-	-	2
		1944(1)	Simétrica	amigo-amigo	-	-	-	1
		1948(1)	Simétrica	amigo-amigo	-	-	-	1
CDA	Itabira 31.10.1902	1951(1)	Simétrica	amigo-amigo	-	-	-	1
		1954(1)	Simétrica	amigo-amigo	-	-	-	1
		1970(1)	Simétrica	amigo-amigo	-	-	-	1
		1972(1)	Simétrica	amigo-amigo	-	-	-	1
		1977(1)	Simétrica	amigo-amigo	-	-	-	1
		1981(1)	Simétrica	amigo-amigo	-	-	-	1
		1983(1)	Simétrica	amigo-amigo	-	-	-	1
		1986(1)	Simétrica	amigo-amigo	-	-	-	1
CLB	Lambari 04.04.1931	1941(1)	Ascendente	afilhada-madrinha	-	-	1	-
EAP	Caeté 17.07.1892	1911(1)	Simétrica	primo-primo	-	-	1	-
FRA	NA	1968(1)	Ascendente	sobrinho-tia	-	-	1	-
		1933(2)	Simétrica	irmã-irmã	-	-	2	-
		1961(1)	Simétrica	irmã-irmão	-	-	1	-
HL	Lambari 15.07.1901	1973(1)	Simétrica	amiga-amigo	-	-	-	1
		1975(1)	Simétrica	amiga-amigo	-	-	-	1
		1975(1)	Descendente	amiga-amigo	-	-	-	1
		1977(1)	Descendente	amiga-amigo	-	-	-	1
		1979(3)	Descendente	amiga-amigo	-	-	-	3

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	PESSOAL		
						AMOROSA	FAMILIAR	AMIZADE
HL	Lambari 15.07.1901	1982(1)	Simétrica	amiga-amigo	-	-	-	1
		NA (1)	Descendente	amiga-amigo	-	-	-	1
		NA (1)	Simétrica	irmã-irmão	-	-	1	-
JAG	Mariana/ Ouro Preto 1921	1943(1)	Simétrica	amigo-amigo	-	-	-	1
		NA (1)	Simétrica	amigo-amigo	-	-	-	1
		1947 (1)	Simétrica	marido- esposa	-	1	-	-
		1962(2)	Descendente	pai-filho	-	-	2	-
		1969(3)	Simétrica	marido- esposa	-	3	-	-
JCL	Lambari 04.11.1902	1924(1)	Simétrica	irmão-irmã	-	-	1	-
		1927(1)	Simétrica	irmão-irmã	-	-	1	-
		1941(1)	Simétrica	irmão-irmã	-	-	1	-
		1945(1)	Simétrica	irmão-irmã	-	-	1	-
		1947(1)	Simétrica	irmão-irmã	-	-	1	-
		1948(1)	Simétrica	irmão-irmã	-	-	1	-
		1960(1)	Simétrica	irmão-irmã	-	-	1	-
JLJ	Lambari/NA Belo	1971(1)	Simétrica	irmão-irmã	-	-	1	-
		1924(1)	Simétrica	irmão-irmã	-	-	1	-
LBVM	Horizonte 26.09.1934	2007(1)	Descendente	prima/avó- prima/neta	-	-	1	-
LMA	Lapa 04.05.1910	1945(1)	Descendente	mãe-filho	-	-	1	-
LR	NA	1947(1)	Descendente	tio-sobrinho	-	-	1	-
LUC	NA	1925(1)	Simétrica	prima-prima	-	-	1	-
		1953(1)	Ascendente	sobrinha-tia	-	-	1	-
MAVP	Campanha 15.04.1932	1957(1)	Ascendente	sobrinha-tia	-	-	1	-
		1959(1)	Ascendente	sobrinha-tia	-	-	1	-
		1961(1)	Ascendente	sobrinha-tia	-	-	1	-
		1983(1)	Ascendente	sobrinha-tia	-	-	1	-
		1941(1)	Simétrica	irmã-irmã	-	-	1	-
MJLB	Lambari 18.09.1898	1944(2)	Simétrica	irmã-irmã	-	-	2	-
		1945(1)	Simétrica	irmã-irmã	-	-	1	-
		1948(1)	Simétrica	irmã-irmã	-	-	1	-
MJM	NA	1951(1)	Simétrica	irmã-irmã	-	-	1	-
		1938(1)	Simétrica	prima-prima	-	-	1	-
MM	Juiz de fora 13.05.1901	1939(1)	Simétrica	amigo-amigo	-	-	-	1
		1956(1)	Simétrica	amigo-amigo	-	-	-	1
		1959(1)	Simétrica	amigo-amigo	-	-	-	1
		1960(1)	Simétrica	amigo-amigo	-	-	-	1
		1961(2)	Simétrica	amigo-amigo	-	-	-	2
		1964(1)	Simétrica	amigo-amigo	-	-	-	1
		1937(2)	Descendente	mãe-filha	-	-	2	-
MRVL	Campanha 1879	1939(1)	Descendente	mãe-filha	-	-	1	-
		1940(1)	Descendente	mãe-filha	-	-	1	-
		1941(1)	Descendente	mãe-filha	-	-	1	-
		1944(2)	Descendente	mãe-filha	-	-	2	-
		1946(1)	Descendente	mãe-filha	-	-	1	-
		1951(1)	Descendente	mãe-filha	-	-	1	-
		NA (2)	Descendente	mãe-filha	-	-	2	-
		1947(1)	Simétrica	amigo-amigo	-	-	-	1
		1948(2)	Simétrica	amigo-amigo	-	-	-	2
		1949 (5)	Simétrica	amigo-amigo	-	-	-	5
OLR	São J. Del- Rei 01.05.1922	1950(3)	Simétrica	amigo-amigo	-	-	-	3
		1951(5)	Simétrica	amigo-amigo	-	-	-	5
		1957(1)	Simétrica	amigo-amigo	-	-	-	1
		1958(5)	Simétrica	amigo-amigo	-	-	-	5
		1959(2)	Simétrica	amigo-amigo	-	-	-	2
PP	NA	1955(1)	Simétrica	amigo-amigo	-	-	-	1
		NA (1)	Simétrica	amigo-amigo	-	-	-	1
RCAM	BH 01.03.1957	1976(1)	Simétrica	amiga-amigo	-	-	-	1

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	PESSOAL		
						AMOROSA	FAMILIAR	AMIZADE
RCAM	BH 01.03.1957	1977(1)	Simétrica	amiga-amigo	-	-	-	1
		1978(1)	Simétrica	amiga-amigo	-	-	-	1
		NA (2)	Simétrica	amiga-amigo	-	-	-	2
WAL	NA	NA (1)	Simétrica	irmão-irmão	-	-	1	-
WL	Lambari 1912	1960(1)	Simétrica	irmão-irmão	-	-	1	-
QUANTIDADE		SUBTOTAL POR SUBGÊNERO			-	6	67	77
29		TOTAL DE CARTAS DO ACERVO					150	

2.3.3 O Arquivo Público Mineiro

Do Arquivo Público Mineiro (APM), há cartas pessoais de João Pinheiro da Silva (JPS)²³, missivas já editadas e publicadas por Luz (2015). Trata-se de 7 cartas familiares, 13 cartas de amizade e 5 cartas amorosas, cf. a síntese exposta no quadro 5.

QUADRO 5
Relação de missivas pessoais do Arquivo Público Mineiro (APM)

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	PESSOAL		
						AMOROSA	FAMILIAR	AMIZADE
JPS	Serro 16.12.1860	1868(1)	Ascendente	sobrinho-tio	-	-	1	-
		1869(1)	Ascendente	sobrinho-tio	-	-	1	-
		1882(1)	Ascendente	sobrinho-tio	-	-	1	-
		1884(1)	Ascendente	sobrinho-tio	-	-	1	-
		1890(1)	Ascendente	sobrinho-tio	-	-	1	-
		1890 (1)	Simétrica	marido-mulher	-	1	-	-
		1891(3)	Simétrica	marido-mulher	-	3	-	-
		1891(3)	Descendente	amigo-amigo	-	-	-	3
		1892(1)	Descendente	amigo-amigo	-	-	-	1
		1892(1)	Simétrica	marido-mulher	-	1	-	-
		1896(1)	Simétrica	amigo-amigo	-	-	-	1
		1899(1)	Descendente	amigo-amigo	-	-	-	1
		1900(1)	Ascendente	sobrinho-tio	-	-	1	-
		1901(1)	Descendente	amigo-amigo	-	-	-	1
		1903(2)	Descendente	amigo-amigo	-	-	-	2
		1904(1)	Simétrica	amigo-amigo	-	-	-	1
		1905(2)	Simétrica	amigo-amigo	-	-	-	2

²³ “João Pinheiro (JP), cf. Luz (2015, p. 44-53), nasceu na cidade do Serro (MG), em 16 de dezembro de 1860, e faleceu em Belo Horizonte (MG), em 25 de outubro de 1908. Foi professor, advogado, político e industrial. Fez o ginásio no Seminário de Mariana e estudou engenharia na Escola de Minas de Ouro Preto (MG). Coursou direito na Faculdade de Direito de São Paulo e trabalhou como professor da Escola Normal Superior da mesma cidade. Anos mais tarde abriu um escritório de advocacia em Ouro Preto. Entrou para a política envolvido com a causa republicana, sendo nomeado Secretário e, posteriormente, Presidente do Estado de Minas Gerais. Foi eleito deputado para a Assembleia Constituinte da República, mas optou por retirar-se da política e dedicar-se à atividade industrial com a qual trabalhou até o fim da vida.” (CARVALHO, 2020, p. 811-82).

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	PESSOAL		
					AMOROSA	FAMILIAR	AMIZADE	
JPS	Serro 16.12.1860	1908(1) NA(1)	Simétrica Ascendente	amigo-amigo sobrinho-tio	- -	- -	- 1	1 -
QUANTIDADE	SUBTOTAL POR SUBGÊNERO				-	5	7	13
1	TOTAL DE CARTAS DO ACERVO				25			

2.3.4 O Instituto Histórico e Geográfico de Minas Gerais

As 49 cartas do Instituto Histórico e Geográfico de Minas Gerais (IHGMG) estão distribuídas entre cartas familiares (24) e cartas de amizade (25), cf. o quadro 6. Trata-se de missivas de amizade produzidas por um importante representante mineiro da literatura brasileira, Carlos Drummond de Andrade (CDA), e por familiares e amigos do Pe. Agenor (PA)²⁴.

QUADRO 6
Relação de missivas pessoais do
Instituto Histórico e Geográfico de Minas Gerais (IHGMG)

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO					
					COMERCIAL	PESSOAL				
					AMOROSA	FAMILIAR	AMIZADE			
AFF	NA	1917(1)	Simétrica	primo-primo	-	-	1	-		
		1932(1)	Simétrica	amigo-amigo	-	-	-	1		
		1936(3)	Simétrica	amigo-amigo	-	-	-	3		
		1938(1)	Simétrica	amigo-amigo	-	-	-	1		
		1939(1)	Simétrica	amigo-amigo	-	-	-	1		
		1946(1)	Simétrica	amigo-amigo	-	-	-	1		
		1949(1)	Simétrica	amigo-amigo	-	-	-	1		
		1951(2)	Simétrica	amigo-amigo	-	-	-	2		
		CDA	Itabira 31.10.1902	1952(1)	Simétrica	amigo-amiga	-	-	-	1
				1953(2)	Simétrica	amigo-amigo	-	-	-	2
				1954(4)	Simétrica	amigo-amigo	-	-	-	4
				1959(2)	Simétrica	amigo-amigo	-	-	-	2
				1960(1)	Simétrica	amigo-amigo	-	-	-	1
				1963(1)	Simétrica	amigo-amiga	-	-	-	1
Desconhecido	NA	1986(1)	Simétrica	amigo-amigo	-	-	-	1		
		NA(1)	Simétrica	amigo-amigo	-	-	-	1		
		1912(1)	Simétrica	amigo-amigo	-	-	-	1		

²⁴ “Pe. Agenor (PA), cf. Corrêa (2010, p. 16-18), nasceu em Caeté (MG), em 30 de novembro de 1890 e morreu em Belo Horizonte (MG), em 12 de fevereiro de 1968. Foi um sacerdote que se destacou na Igreja por sua inteligência e probidade (...). Filho de professores, iniciou seus estudos em casa em Ibertioga (MG), depois matriculou-se no Asilo Arquiepiscopal Nossa Senhora do Patrocínio, mudando-se, anos mais tarde para Mariana, onde foi ordenado. O padre desde cedo revelou-se um aluno aplicado, merecendo o apoio e a admiração de seus pares. No sacerdócio, desempenhou diversas funções eclesiais nas Arquidioceses de Mariana, Belo Horizonte e Sete Lagoas, dedicando sua vida aos fiéis e à Igreja.” (CARVALHO 2020, p. 83.).

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	AMOROSA	FAMILIAR	AMIZADE
FAPJ	Senhora do Porto	1913(1)	Descendente	tio-sobrinho	-	-	1	-
		1916(1)	Descendente	tio-sobrinho	-	-	1	-
	23.02.1875	1917(3)	Descendente	tio-sobrinho	-	-	3	-
HMV	NA	1913(1)	Simétrica	amigo-amigo	-	-	-	1
JP	NA	1917(1)	Descendente	tio-sobrinho	-	-	1	-
MAP	NA	1918(1)	Descendente	tio-sobrinho	-	-	1	-
OAAP	Ibertioga	1917(1)	Simétrica	prima-primo	-	-	1	-
	12.09.1892	1918(1)	Simétrica	irmã-irmão	-	-	1	-
		1907(1)	Descendente	pai-filho	-	-	1	-
		1912(1)	Descendente	pai-filho	-	-	1	-
		1913(4)	Descendente	pai-filho	-	-	4	-
RAAP	Senhora do Porto	1914(1)	Descendente	pai-filho	-	-	1	-
		1915(1)	Descendente	pai-filho	-	-	1	-
	11.05.1836	1917(2)	Descendente	pai-filho	-	-	2	-
		1923(1)	Descendente	pai-filho	-	-	1	-
		1924(1)	Descendente	pai-filho	-	-	1	-
		NA(1)	Descendente	pai-filho	-	-	1	-
RAP	NA	1915(1)	Descendente	tia-sobrinho	-	-	1	-
QUANTIDADE	SUBTOTAL POR SUBGÊNERO				-	-	24	25
10	TOTAL DE CARTAS DO ACERVO						49	

2.3.5 O Museu Abílio Barreto

As missivas do Museu Abílio Barreto (MAB) apresentam cartas de amizade e familiares produzidas por integrantes das famílias Tassini²⁵ e Barreto²⁶, ambos representantes mineiros como falantes cultos da língua portuguesa, tendo em vista as suas respectivas funções sociais, cf. quadro 7.

²⁵ “Raul Tassini (RT), cf. Andrade (2008, p. 215), nasceu em Belo Horizonte (MG), em 1909, onde também faleceu em 1992. Foi desenhista, ilustrador poeta e museólogo. Estudou na Escola de Belas Artes de Belo Horizonte e na Academia de Belas Artes, em Roma. Trabalhou como técnico em museologia do Museu Histórico Abílio Barreto (MG) e no Museu Nacional de Belas Artes (RJ).” (ANDRADE, 2008 *apud* CARVALHO, 2020, p. 82-83).

²⁶ “Abílio Velho Barreto (AVB), cf. Duarte (2010, p. 18-19), nasceu em Diamantina (MG), em 22 de outubro de 1883, e morreu em Belo Horizonte (MG), em 17 de julho de 1959. Foi escritor, poeta, jornalista e homem público. Trabalhou na imprensa recém-fundada de Belo Horizonte como tipógrafo, revisor e, posteriormente, como redator do jornal Minas Gerais. Ocupou o cargo de Primeiro Oficial do Arquivo Público Mineiro e foi convidado, anos mais tarde, a organizar o Museu Histórico de Belo Horizonte que futuramente receberia seu nome. Além de poesias e romances, também publicou obras sobre a história belo-horizontina.” (DUARTE, 2010 *apud* CARVALHO, 2020, p. 82).

QUADRO 7
Relação de missivas pessoais do Museu Abílio Barreto (MAB)

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	AMOROSA	FAMILIAR	AMIZADE
AB	Diamantina 22.10.1883	1956(1)	Descendente	pai-filho	-	-	1	-
CT	NA	1930(1)	Descendente	mãe-filho	-	-	1	-
		1932(1)	Descendente	mãe-filho	-	-	1	-
		1931(3)	Descendente	mãe-filho	-	-	3	-
CT/ET	NA	1932(1)	Descendente	mãe-filho	-	-	1	-
ET ²⁷	Veneto, Itália 17.02.1876	1932(1)	Descendente	pai-filho	-	-	1	-
MOC	NA	NA (1)	Simétrica	amiga-amiga	-	-	-	1
		1931(4)	Ascendente	filho-pais	-	-	4	-
		1931(1)	Simétrica	amigo-amiga	-	1 ²⁸	-	-
RT	Belo Horizonte 03.07.1909	1932(3)	Ascendente	filho-pais	-	-	3	-
		1932(1)	Simétrica	irmão-irmão	-	-	1	-
		1954(1)	Simétrica	irmão-irmão	-	-	1	-
ST	NA	NA(1)	Ascendente	filho-pais	-	-	1	-
		1932(1)	Simétrica	irmã-irmão	-	-	1	-
VT	NA	1931(1)	Simétrica	irmã-irmão	-	-	1	-
QUANTIDADE		SUBTOTAL POR SUBGÊNERO			-	1	20	1
7		TOTAL DE CARTAS DO ACERVO					22	

2.3.6 O Arquivo Público da Cidade de Belo Horizonte

A amostra do Arquivo Público da Cidade de Belo Horizonte (APCBH) engloba um total de 26 cartas pessoais familiares, cf. descrito no QUADRO 8.

QUADRO 8
Relação de missivas pessoais do
Arquivo Público da Cidade de Belo Horizonte (APCBH)

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	REL. SOCIAL	INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	AMOROSA	FAMILIAR	AMIZADE
AS	São João Evangelista 27.03.1886	1904(1)	Ascendente	sobrinho-tio	-	-	1	-
CCR	Guanhães NA	1890(1)	Descendente	tio-sobrinho	-	-	1	-
		1910(1)	Descendente	tio-sobrinho	-	-	1	-
GF	Mariana 03.01.1865	1906(1)	Simétrica	amigo-amigo	-	-	-	1

²⁷ O redator ET é o único informante não nascido no Brasil. Embora ET tenha nascido em Veneto, Itália, as cartas de sua autoria foram incluídas na amostra em análise, considerando que o redator chegou, ainda muito jovem, à cidade de BH onde constituiu família e se desenvolveu profissionalmente.

²⁸ Embora seja uma carta amorosa, a relação interpessoal está codificada como amigo-amiga por não estar claro que o autor seja namorado ou noivo, mas um admirador da interlocutora.

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	REL. SOCIAL	INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	AMOROSA	FAMILIAR	AMIZADE
MBCS	São Miguel e Alma de Guanhões 25.01.1847	1891(1)	Descendente	mãe-filho	-	-	1	-
		1901(1)	Descendente	mãe-filho	-	-	1	-
		1903(7)	Descendente	mãe-filho(6) sogra-nora(1)	-	-	7	-
		1904(5)	Descendente	mãe-filho	-	-	5	-
		1905(1)	Descendente	mãe-filho	-	-	1	-
		1895(1)	Ascendente	filho-mãe	-	-	1	-
NCS	Serro 11.01.1876	1920(1)	Simétrica	amigo-amigo	-	-	-	1
		1938(1)	Simétrica	amigo-amigo	-	-	-	1
		NA(1)	Simétrica	amigo-amigo	-	-	-	1
PB	Serro 17.01.1861	1908(1)	Simétrica	amigo-amigo	-	-	-	1
PSN	Januária 25.11.1863	1903 (1)	Simétrica	irmão-irmão	-	-	1	-
		1915(1)	Simétrica	irmão-irmão	-	-	1	-
QUANTIDADE		SUBTOTAL POR SUBGÊNERO			-	-	21	5
7		TOTAL DE CARTAS DO ACERVO					26	

2.3.7 Acervos Pessoais

Além dos dados coletados de cartas acessadas nos arquivos públicos acima citados, a amostra do Acervo Pessoal (AP) contém 60 cartas pessoais arquivadas em acervos privados²⁹, cujos informantes terão as suas identidades resguardadas neste trabalho. Para isso, foram mantidas as letras iniciais do nome completo dos missivistas ou as três primeiras letras do nome, nos casos em que o sobrenome do autor da carta não foi identificado. No total, são 3 informantes, entre eles, o informante JAVS e sua noiva MIDV com 44 cartas amorosas, trocadas no período de 1924 a 1927, quando os redatores ainda eram noivos. Trata-se de 12 cartas de autoria de MIDV, nascida em Fazenda Cachoeirinha, em Alfenas e 32 cartas de JAVS, nascido em Machado, MG. Além do casal acima referido, a amostra contém cartas pessoais da família Guimarães. Trata-se de 15 cartas de autoria de CPG, nascida em Jequitinhonha. Em uma das cartas de CPG, há um bilhete de RAU, filho de CPG, cf. síntese descrita no QUADRO 9.

²⁹ O acesso às cartas de acervos pessoais foi possível mediante o contato junto aos responsáveis pelos Acervos dos documentos históricos, tendo sido assegurado às famílias dos missivistas que seria resguardada a identidade dos informantes.

QUADRO 9
Relação de missivas dos Acervos Pessoais

MISSIVISTA	LOCAL E DATA DE NASC.	ANO DA CARTA	RELAÇÃO SOCIAL	RELAÇÃO INTERPESSOAL	QUANTIDADE DE CARTAS POR GÊNERO			
					COMERCIAL	PESSOAL AMOROSA	FAMILIAR	AMIZADE
CPG	Jequitinhonha 20.03.1910	1960 (2)	Simétrica	esposa-marido	-	2	-	-
		1961 (6)	Simétrica	esposa-marido	-	6	-	-
		1962 (3)	Simétrica	esposa-marido	-	3	-	-
		1963 (2)	Simétrica	esposa-marido	-	2	-	-
		1969 (2)	Simétrica	esposa-marido	-	2	-	-
JAVS	Machado 01.04.1893	1924(9)	Simétrica	noivo-noiva	-	9	-	-
		1925(18)	Simétrica	noivo-noiva	-	18	-	-
		1926(2)	Simétrica	noivo-noiva	-	2	-	-
		1927(1)	Simétrica	noivo-noiva	-	1	-	-
		NA(2)	Simétrica	noivo-noiva	-	2	-	-
MIDV	Alfenas 17.08.1901	1924(1)	Simétrica	noiva-noivo	-	1	-	-
		1925(11)	Simétrica	noiva-noivo	-	11	-	-
		NA (1)	Descendente	avó-neta	-	-	1	-
Quantidade		Subtotal por subgênero		-	59	1		
3		Total de cartas do acervo				60		

2.4 A Teoria do Poder e Solidariedade à luz de Brown & Gilman (1960)

A *Teoria do Poder e Solidariedade* de Brown e Gilman (1960) é discutida no artigo intitulado *The Pronouns of Power and Solidarity* em que os autores explicam a origem de dois pronomes singulares de tratamento de referência à 2ª pessoa, T (*tu* latino) e V (*vos* latino), passando não somente em relação à sua origem latina (*tu* e *vos*), mas também aos seus usos no italiano (*tu* e *voi*), espanhol (*tu* e *vos*, depois *usted*), francês (*tu* e *vous*), alemão (*du* e *Ihr*, depois *er* e, mais tarde, *Sie*) e inglês (*thou* e *ye*, depois *you*).

A discussão de Brown & Gilman (1960, p. 254) trata das formas pronominais latinas *tu* e *vos* pela contextualização histórica dos seus usos. Segundo os referidos autores, com as reformas do imperador romano Diocleciano, a partir do século IV, o *vós* (2PL) em oposição ao *tu* (2SG) passou a assumir um uso de deferência exclusiva ao imperador como expressão de polidez e distanciamento entre uma classe dominante e outra, dominada. Entre os séculos XI e XIV, cristalizou-se a *Semântica do Poder* descrita por Brown e Gilman (1960, p. 255), dimensão social

que estabelece as relações assimétricas e envolve, no mínimo, duas pessoas, com a máxima³⁰ “*Se A é mais antigo que B, B não é mais antigo que A*” (BROWN & GILMAN, 1960, p. 257). A título de exemplificação, observe-se, em (18), o uso do *tu* numa dinâmica de assimetria descendente (de superior para inferior) travada numa relação entre pai e filho. Por outro lado, em (19), tem-se uma relação de assimetria ascendente (de inferior para superior) travada, com o uso do *vossa mercê*, entre sobrinho e tio.

(18) “Meu presado filho Agenor [...], acho que *tens* feito muito progresso, embora lutando sempre com dificuldades. [...]” (RAAP. Lagoa Santa, 30.10.1912.)

(19) “[...] Vai pois a Carta para o Senhor Luiz Jose, aberta para *Vossa mercê* a ler, e depois deixa la e fazer entregar. [...]” (BNP. Paciência, 10.05.1856.)

Em algumas sociedades modernas, a oposição *tu/vós* permite que a *Semântica do Poder* sofra a influência da *Semântica da Solidariedade*, dando origem a sistemas simétricos como o *vós/vós* e o *tu/tu*, cf. Wardhaugh (2006 [1986]). Nesse sentido, a dimensão da *Solidariedade* pode ser compreendida com base na relação familiar³¹: “*Se A tem os mesmos pais que B, B tem os mesmos pais que A*” (BROWN & GILMAN, 1960, p. 258), cf. está ilustrado em (20) com base na interação travada entre irmãos (relação de igualdade) a partir da convivência entre as formas pronominais *tu* e *você* numa dinâmica da *Solidariedade*.

(20) “[...] Se poudesse *voce* Tinho fazer uma excursão tão grande como a que eu fiz, *haverias* de extasiar-te ante as bellezas historicas desta Italia maravilhosa [...]” (RT. Roma, 25.02.1932).

A Teoria do Poder e Solidariedade nasceu de um estudo de Brown & Gilman (1960) embasado em entrevistas a estudantes de diversas nacionalidades. Foram selecionados dois informantes de cada país (Espanha, Argentina, Chile, Dinamarca, Noruega, Suécia, Israel, África do Sul, Índia, Suíça, Holanda, Áustria e Iugoslávia), tendo em vista a apreensão das formas pronominais de 2SG produtivas nos contextos das relações simétricas e assimétricas (ascendentes e descendentes). Os informantes responderam sobre a possibilidade de uso das formas pronominais de 2SG (T (*tu*) ou V (*vous*)) nas relações sociais com sua mãe, seu pai, seu avô, sua

³⁰ “*If A is older than B, B is not older than A*” (BROWN & GILMAN, 1960, p. 257).

³¹ “*If A has the same parents as B, B has the same parents as A.*” (BROWN & GILMAN, 1960, p. 258).

esposa, um irmão mais velho ou mais jovem, um primo, a babá que conheceu desde a infância, colegas da faculdade, entre garçom e cliente, colegas de escritório, militares, chefe e empregado etc. Brown & Gilman (1960, p. 280) sugerem que a dimensão da Solidarietà tende a ser ampliada na atualidade, o que está em conformidade com Wardhaugh (2006 [1986]) em relação ao fato de os movimentos sociais e as sociedades modernas se encaminharem em direção à Solidarietà ainda que também evidenciem a *Semântica do Poder* (WARDHAUGH, 2006 [1986]). Nesse sentido, Rumeu (2011, p. 124) constata que as cartas cariocas setecentistas mostram sinais da *Semântica do Poder*, ao passo que as missivas oitocentistas parecem refletir a *Semântica da Solidarietà*.

Nesta dissertação, parte-se do pressuposto de que as línguas humanas expressam-se numa dinâmica de heterogeneidade ordenada cuja expressão variável pode ser descrita e analisada nos eixos estrutural e social (WEINREICH *et al.*, 1968). Assim sendo, volta-se o foco para os contextos das relações sociais (assimétricas (ascendentes e descendentes) e simétricas) que sedimentam os usos de *tu* e *você* nas cartas oitocentistas e novecentistas.

2.5 As formas nominais e pronominais de referência à 2SG: os parâmetros de uma análise linguístico-social

A proposta principal desta dissertação é conduzida pelos princípios teóricos e metodológicos da Sociolinguística Histórica (ROMAINE, 1982; CONDE SILVESTRE, 2007; HERNÁNDEZ-CAMPOY & CONDE SILVESTRE, 2012) orientada também pelos pressupostos teórico-metodológicos da Sociolinguística Variacionista (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1994). A regra variável *você/tu* é o foco deste trabalho, considerando o contexto das relações sociais nos âmbitos da simetria e da assimetria social (ascendente e descendente) que fundamentam as escolhas pronominais dos missivistas mineiros nos séculos XIX e XX.

Uma primeira etapa desta investigação foi a organização das amostras³² de cartas de missivistas mineiros. Uma segunda etapa foi o levantamento das formas de referência ao sujeito de 2SG com o foco voltado especificamente para as formas *vossa mercê*, *você* e *tu* em 324 cartas, o que necessariamente passa pela codificação de 1236 dados em relação aos parâmetros linguísticos e extralinguísticos (sociais), pensados a partir dos resultados de pesquisas linguísticas acerca da dinâmica *tu/você* em sincronias recentes (em Minas Gerais) e em sincronias passadas (cf. capítulo 1). Uma terceira etapa foi a codificação dos dados tendo em vista as formas de pronomes-sujeito de 2SG como variável dependente (*vossa mercê*, *você* e *tu*) analisada em função das seguintes variáveis independentes.

- a. Expressão da forma de referência ao sujeito de 2SG: nulo ou pleno.
- b. Pessoa verbal (concordância);
- c. Paralelismo formal e semântico;
- d. O gênero da carta pessoal (familiar, amorosa, de amizade) e da carta de comércio;
- e. Sexo do informante;
- f. Faixa etária;
- g. Relação interpessoal;
- h. Relações sociais;
- i. Períodos das cartas.

Passa-se à descrição dos critérios linguísticos e extralinguísticos, considerando as hipóteses e/ou objetivos que os fomentaram nesta análise, tendo sempre em vista os resultados de análises exploradas no capítulo 1 (revisão do tema) desta dissertação.

2.5.1 Os grupos de fatores linguísticos e sociais

O fenômeno em estudo envolve as estratégias pronominais de referência ao sujeito de 2SG na escrita mineira dos séculos XIX e XX. Para o controle dos dados foram definidos, à luz de hipóteses e/ou objetivos embasados nos resultados de pesquisas linguísticas, nove grupos de fatores (variáveis independentes). Os dados de referência ao sujeito de 2SG foram levantados e organizados em uma planilha do *Excel* através da qual todas as ocorrências foram categorizadas

³² Os dados de formas de referência ao sujeito de 2SG foram levantados em 358 cartas, incluindo, a princípio, as ocorrências de imperativo de 2SG. Ao retirá-los da análise, passa-se à análise de uma variável enéria formada pelas formas *vossa mercê*, *você* e *tu*, em uma amostra de 324 cartas, com 1236 ocorrências,

em função das variáveis independentes. A organização dos dados na planilha do *Excel* permite a sua codificação, correção e análise desde o início do processo, passando pela composição das tabelas dinâmicas, responsáveis pela geração dos índices percentuais até a verificação dos pesos relativos, nas rodadas nos níveis *step up & step-down* do RBrul. O trabalho do linguista conta com os recursos estatísticos que tendem a facilitar o tratamento desses dados, por meio do controle de frequências de uso (em percentuais), pesos relativos, *p-value*, *input* e *logodds*³³. Em suma, os dados foram submetidos ao Programa RBrul na interface do RStudio (OUSHIRO, 2014, p. 136) e também ao editor de planilhas *Microsoft Excel* (pacote *Office*), cf. Imagem 13, cuja tabela dinâmica também se mostrou essencial ao controle, à correção e à análise dos dados levantados e categorizados em relação às variáveis independentes.

B	C	D	E	F	G	H	I	J	K	L	M
1	expressao	paralelismo	carta	genero	fretaria	interpessoal	social	periodo	ooo		
2	p	3	1 f	h	v	p	d	c	(...), sua mãe e eu, que apesar da viagem fustigante [você] [chegou] bem (...). MAB. AB. BH. 24.01.1956		
3	p	3	4 f	h	v	p	d	c	(...) da < sua > carta, que vou responder. Vou escrever a ele hoje, dizendo que [você] [está] em Vitória (...).		
4	p	3	4 f	h	v	p	d	c	(...) < você > está em Vitória o < teu > endereço. (...) Eu continuo como [você] [sabe] (...). MAB. AB. BH. 24.		
5	p	3	4 f	h	v	p	d	c	(...) Por estes dias < sua > Mãe escreverá em resposta. Ela e eu fazemos votos para que [você] [seja] bem		
6	n	2	1 f	h	x	n	s	7	(...) Pegou-te [0] não [levares] a mal esse meu modo de procede (...). IHGMG. Afonso. BH. 22.03.1917.		
7	p	3	1 z	h	j	b	s	b	(...) Suas palavras me deram (...) alegria (...) mas [você], grande artista que é, [sabe] da solidão em que		
8	n	3	2 z	h	j	b	s	b	(...) Suas palavras me deram (...) alegria (...) mas < você >, grande artista que [0] [é], sabe da solidão em		
9	n	3	2 z	h	j	b	s	b	(...) mas < você >, (...), sabe da solidão em que andamos (...). Jamais me esquecerei do estímulo que [0] m		
10	p	3	1 z	h	j	b	s	b	(...) coração sensível e sincero como o < seu >. Uma vez ou duas, [você] me [falou] que este quadro um di		
11	p	3	1 z	h	k	b	s	d	(...) Logo que surja, [você] [será], como sempre, das primeiras a receber um exemplar. (...). AEM. AGF. Br		
12	p	3	2 z	h	k	b	s	d	(...) < você > será, como sempre, (...) "Vigília Política" está magnífica, tal como tudo que [você] [faz], em		
13	p	3	1 z	h	k	b	s	d	(...) [Você] já entregou o meu voto? (...). AEM. AGF. Brasília, 19.06.1963.		
14	n	3	2 z	h	k	b	s	d	(...) < Você > já entregou (...) Se [0] não o [fiz], consulte o Oscar (...). AEM. AGF. Brasília, 19.06.1963.		
15	p	3	4 z	h	k	b	s	d	(...) Estou daqui torcendo pela < sua > eleição, mas (...) [você] não [deve] < se > impressionar (...). AEM. AG		
16	p	3	4 z	h	k	b	s	d	(...) se houver, derrota < sua >. (...) Se [você] [tiver] necessidade de fazer me (...). AEM. AGF. Brasília, 19.0		
17	n	3	4 z	h	k	b	s	d	(...) Se < você > tiver necessidade de [0] [fazer] me (...). AEM. AGF. Brasília, 19.06.1963.		
18	p	3	1 z	h	k	b	s	d	(...) seu caso foi resolvido (...). Meus parabéns, pois, e nossos votos, meus e de Hygiene, para que [você]		
19	n	3	2 z	h	k	b	s	d	(...) que você seja (...) na condição, em que [0] < se > [encontra], de (...) AEM. AGF. Brasília, 22.05.1968.		
20	n	3	2 z	h	k	b	s	d	(...) que você seja (...) na condição, em que < se > encontra, de [0] [poder] gozar o ócio com dignidade. (...)		
21	p	3	1 z	h	k	b	s	d	(...) de que já lhe falei - no portão do Cemitério do Bonfim. (...) [Você] [poderia] consultar o Lourenço (...)		

Imagem 13: Planilha Excel/Cartas Pessoais Mineiras. Arquivo Pessoal.
Fonte: Microsoft Office Excel, Windows 2010.

Na sequência, passa-se à descrição dos contextos linguísticos e extralinguísticos a que foram submetidos os dados de pronomes-sujeito das cartas mineiras (séculos XIX e XX).

³³ A linguagem R permite a realização de “computações estatísticas e gráficas, compilar e anotar *corpora*, produzir listas de frequências”, além de poder ser gratuitamente instalada em distintas plataformas (UNIX, Windows e MacOS), cf. Oushiro (2014, p. 134).

2.5.1.1 As formas de referência ao sujeito de 2SG: *vossa mercê, tu e você*

Em um primeiro momento desta análise, o grupo “pronomes-sujeito de 2SG” funciona como a variável (dependente) eneária (cf. GUY & ZILLES, 2007), tendo em vista a hipótese de que o *você* já se mostraria, em sincronias passadas (século XIX), mais produtivo, em conformidade ao que foi observado para a fala mineira de sincronias recentes (COELHO, 1999; PERES, 2006; HERÊNIO, 2006; GONÇALVES, 2008; MOTA, 2008; SILVA, 2017 e REIS, 2019).

No tocante às relações sociais, buscou-se verificar se o *você* se mostraria produtivo no contexto de assimetria descendente (de superior para inferior), considerando ainda que, no Brasil do século XIX, o *tu* e o *você* passaram a concorrer principalmente em relações solidárias mais íntimas (RUMEU, 2013, p. 51). Em relação ao *vossa mercê*, a hipótese inicial é a de que a sua produtividade ficaria restrita aos contextos das relações sociais menos solidárias e transacionais em conformidade com Barcia (2006 *apud* RUMEU, 2013, p. 28).

2.5.1.2 Expressão da forma de referência ao sujeito de 2SG: nulo ou pleno

No grupo de fatores *Expressão da forma de referência ao sujeito de 2SG*, a análise da expressão (nula ou plena) das formas *tu* e *você* como sujeitos pronominais de 2SG é motivada pela ideia de que, à medida que o *você* se pronominaliza, tenda a assumir os mesmos campos funcionais que ocupa um legítimo pronome do PB, cf. discutido por Rumeu (2013) com base nas cartas cariocas da família Pedreira Ferraz-Magalhães. Parte-se da hipótese de que o *tu* ocorreria preferencialmente nulo (21), ao passo que o *você* ocorreria como sujeito pleno (22), cf. Souza (2012, p. 86). Em relação às missivas cariocas, Rumeu (2013) observou que a expressão nula do sujeito é categoricamente favorecida pelo *tu*, ao passo que o *você* se mostrou preferencialmente pleno, ainda que em acirrada concorrência com a sua expressão nula, cf. Rumeu (2013, p. 169). Para o *vossa mercê*, a hipótese é a de sua expressão plena, nas cartas mineiras analisadas (23), acompanhando os resultados de Rumeu (2004) não só para o século XVIII, mas também para as missivas cariocas do século XIX (RUMEU, 2013).

(21) [...] Desejo que esta vá te encontrar gozando perfeita saude ao par de grandes felicidades. Eu como *sabes* gozo saude mas sempre pelejando com os filhos na maior dificuldade da vida [...]. (OAAP. Lagoa Santa, 28.02.1918.)

(22) [...] Estou espantado em ver que este ano está chegando ao termo. Que velocidade a do tempo ultimamente! *Você* não o sente também? [...]. (AM. RJ, 01.10.1945.)

(23) [...] Diceme elle que *Vossa Mercê* o havia feito voltar por elle estar com a roupa muito ordinária e não levasse outra para mudar; contando que só hia a Ouro Preto. [...]. (BNP. Paciencia, 22.12.1853.)

2.5.1.3 Pessoa verbal (concordância)

A análise da pessoa verbal (concordância) parte da hipótese de que o *você* seria favorecido por formas verbais de 3SG (24), ao passo que o *tu* (25) teria, em contextos verbais de 2SG, maior produtividade, cf. Souza (2012, p. 86). Nesse sentido, ao investigar cartas cariocas, Souza (2012) constatou “a persistência formal (herdada do *vossa mercê*) presente na concordância *você* + 3ª pessoa, mesmo o *você* fazendo referência à segunda pessoa do discurso”, cf. Souza (2012, p. 99).

(24) [...] Ao chegar soube que seo Pae vendera a casa ao Belisario, como *Você* já sabe, e baracta, foi precipitado demais, elle me disse que escreveo-lhes sobre tudo menunciosamente. [...]. (PB. S.I., 19.08.1908.)

(25) [...] Para nós, em qualquer hipótese, é um prazer ter-te aqui; *tu* resolverás. Pretendemos seguir para Minas na primeira quinzena de fevereiro. [...]. (AM. RJ, jan. 1944.)

2.5.1.4 Paralelismo formal e semântico

O controle dos dados em relação ao *Paralelismo formal e semântico* parte da hipótese de que a sequência discursiva iniciada pela forma pronominal *você* seria um contexto promissor para a sua produtividade, cf. Omena (2003) e Lopes & Vianna (2012, p. 148). Deste modo, pretende-se descrever até que ponto o emprego de formas de 2SG ou de 3SG influenciaria ou não os pronomes-sujeito da produção escrita mineira. O monitoramento do paralelismo formal e semântico está legitimado no intuito de controle da uniformidade no tratamento, o que significa

entender que o foco está voltado para detectar se há o uso exclusivo de formas do paradigma de *tu* no contexto de *tu-sujeito*, o uso exclusivo de formas do paradigma de *você* no contexto de *você-sujeito* ou ainda se há formas do paradigma de *tu* no contexto de *você-sujeito* ou formas de *você* no contexto de *tu-sujeito*.

2.5.1.5 Os gêneros *carta pessoal* e *carta de comércio*

A análise dos gêneros *carta pessoal* e *carta de comércio* assume o controle dos dados de referência ao sujeito de 2SG não só em relação aos três subgêneros das cartas pessoais (amorosa, amizade e familiar), mas também em função das cartas de comércio. A opção pelo trabalho de análise dos dados de formas de referência ao sujeito de 2SG embasado em cartas (principalmente as cartas pessoais) está justificada no fato de se tratar de fontes mais transparentes e livres da pressão da norma-padrão, possibilitando, então, a observação de traços linguísticos da oralidade (BARBOSA, 1999 *apud* RUMEU, 2013, p. 56). Nessa direção, Rumeu (2013, p. 57-58) propõe que o foco da análise seja menos audacioso do que captar a oralidade através da língua escrita, mas, por outro lado, prevê a possível descrição de traços linguísticos em variação com base na escrita culta de sincronias passadas.

As missivas classificadas como amorosas são as que têm como interlocutores os casais de noivos, namorados e as missivas trocadas entre marido e mulher, cf. exemplo da sentença (26) e da Imagem 14. Na investigação de Souza (2012, p. 116), o *tu* foi a forma mais recorrente nas cartas amorosas, as quais pressupõem um alto grau de intimidade entre os interlocutores.

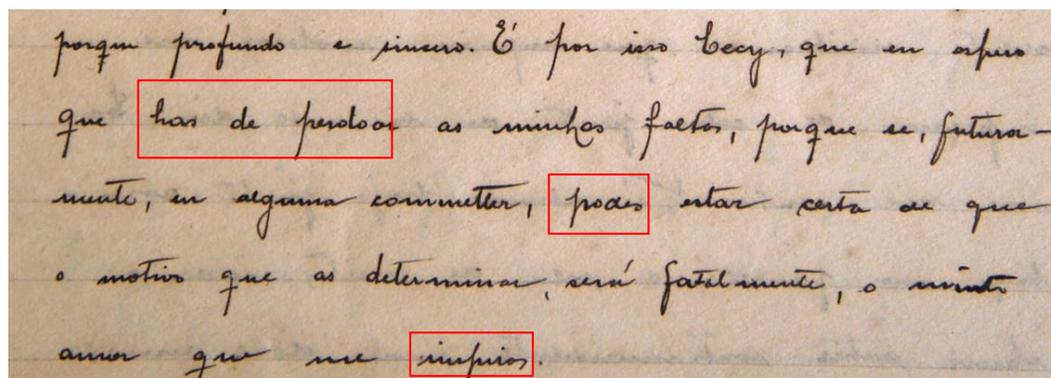


Imagem 14: Dado de *tu* nulo em carta amorosa. (JAVS. Machado, 01.01.1925.)

(26) [...] É por isso Cecy, que eu espero que *has* de perdoar as minhas faltas, porque se, futuramente, em alguma *commetter*, *podes* estar certa de que o motivo que as determinar, será fatalmente, o muito amor que me *inspiras*. [...] (JAVS. Machado, 01.01.1925.)

As cartas de amizade refletem a relação afetiva entre amigos que se conhecem com maior ou menor profundidade, e que às vezes compartilham sonhos, convicções e momentos importantes da vida, como no seguinte trecho de missiva enviada por CDA ao amigo C., cf. a Imagem 15 e a sentença (27).

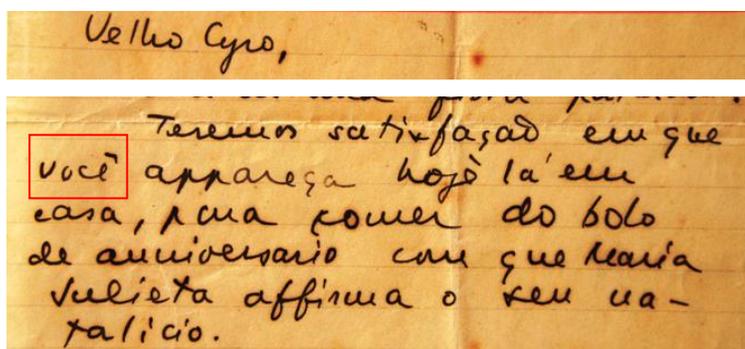


Imagem 15: Dado de *você-sujeito* em carta de amizade. (CDA. BH, 04.03.1931.)

(27) [...] Velho Cyro, [...] Teremos satisfação em que *você* apareça hoje lá em casa, para comer do bolo de aniversário com que Maria Julieta afirma o seu natalício. [...] (CDA. BH, 04.03.1931.)

As missivas familiares são as enviadas aos pais, mães, irmãos, primos, sobrinhos, tios e familiares em geral. Na perspectiva de Souza (2012, p. 114), as “cartas familiares seriam consideradas medianas no grau de intimidade entre os indivíduos, diferentemente das cartas amorosas, nas quais os interlocutores apresentam alto grau de intimidade”. As missivas familiares revelam momentos do dia a dia da família, como na relação entre mãe e filho, ilustrada em (28), através da qual a mineira LMA, ao empregar o *você*, em carta familiar (1945), produzida no final da Segunda Guerra Mundial, traz ensinamento ao filho F., cf. Imagem 16 e sentença (28).

Rio, 5 Setembro 1945
 Querido Fernando
 Recebemos sua cartinha e gostamos
 muito do desenho do fundo do mar.
 Ficou lindo. **Você** sabe que hoje à
 noite a cidade toda vai ficar escura,
 sem luz nenhuma? Sabe porque?
 É para que os aviões inimigos
 não saibam onde está a cidade...

Imagem 16: Dado de *você* em carta familiar (LMA. RJ, 05.09.1945.)

(28) [...] Querido Fernando [...] Recebemos sua cartinha e gostamos muito do desenho do fundo do mar. Ficou lindo. *Você* sabe que hoje à noite a cidade toda vai ficar escura, sem luz nenhuma? Sabe porque? É para que os aviões inimigos não saibam onde está a cidade...[...]. (LMA. RJ, 05.09.1945.)

Em síntese, as cartas de comércio voltadas para a descrição das atividades da lavra de minério (mais especificamente “ouro”) estão restritas às cartas enviadas por BNP ao sogro em meados do século XIX, cf. Imagem 17 e sentença (29).

...o serviço em bom estado dando o mez passado o producto de 286 ¼ Oitavas, pela maneira seguinte: De 4 (incluzive) athe 22 – 128 ¾ oitavas, e os 8 ultimos dias socando também as noites 157 ½ [espaço] A pedra continua boa, e se não houver alguma mudança será esta apuração ainda melhor. [...].

Imagem 17: Trecho de carta de comércio. (BNP. Sam Vicente, 11.12.1854.)

(29) [...] Achei nosso serviço em bom estado dando o mez passado o producto de 286 ¼ Oitavas, pela maneira seguinte: De 4 (incluzive) athe 22 – 128 ¾ oitavas, e os 8 ultimos dias socando também as noites 157 ½ [espaço] A pedra continua boa, e se não houver alguma mudança será esta apuração ainda melhor. [...]. (BNP. Sam Vicente, 11.12.1854.)

Segundo Souza (2012), o termo carta pessoal, em oposição à carta de comércio, seria uma generalização, um hiperônimo, que engloba diferentes tipos de carta tais como a carta

peçoal (ou menos familiar), a carta familiar e a carta amorosa (SOUZA, 2012, p. 114), cf. a Imagem 18:

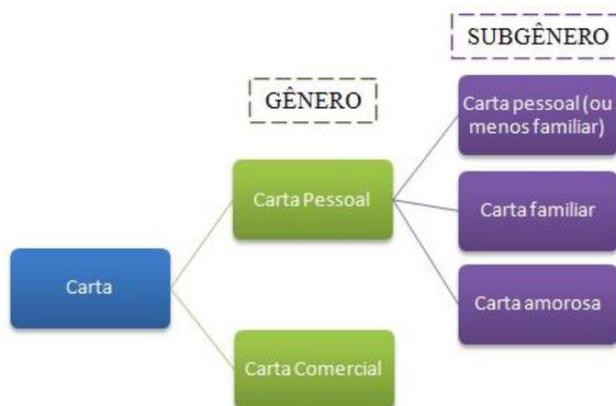


Imagem 18: Gênero carta pessoal e seus subgêneros
Fonte: SOUZA, 2012, p. 114.

Souza (2012) faz referência à carta comercial, colocando-a em oposição à carta pessoal, à luz da argumentação de Marcuschi (2008, p. 155), distinguindo o gênero *carta comercial* de *carta pessoal*, considerando que a *carta pessoal* “corresponde a uma combinação paramétrica relativamente próxima ao polo do imediato comunicativo” (OESTERREICHER, 2006, p. 256 *apud* SOUZA, 2012, p. 113). Dessa maneira, a *carta pessoal* se aproximaria mais de uma escrita informal do que a *carta comercial*. No entanto, ainda que se caracterizem pelo uso de expressões formulaicas e mais formais e também marcadas pela polido *vossa mercê*, as 26 cartas de comércio escritas por BNP apresentam trechos nos quais é possível perceber indícios de informalidade. Nessa direção, observe-se, em (30) e Imagem 19, além da apresentação de relatos regulares quantos aos negócios na mina, o missivista também faz comentários negativos acerca do cunhado (“[...] servirá ao menos para o Cunhado esbanjar!! [...]”), além de fazer menção explícita ao processo de escrita como se fosse expressão da sua fala (“Falando nisto, ocorre-me *que talvez* [...]”), o que parece evidenciar traços de informalidade também nas cartas comerciais destinadas ao Barão de Cocais analisadas.

(30) Deos dê juízo *aquem* o não tem!... Quem sabe se as cobranças da Ponte Nova não servindo para pagamentos de dividas, servirá ao menos para o Cunhado esbanjar!! Falando nisto, ocorre-me *que talvez* amigavelmente e por sua via pudesse eu ao menos obter alguma segurança por meio de hipoteca de Escravos, pois constame *que* dois ainda estão fora de *qual quer* comprometimento, mas *que* já ha quem tenha visto n’ellas. [...] (BNP. Paciencia, 01.04.1857.)

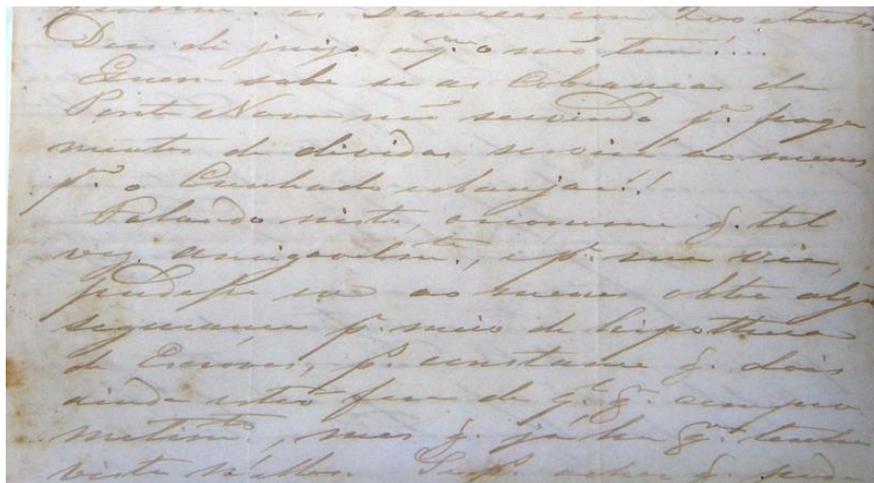


Imagem 19: As cartas de comércio ao Barão de Cocais.
(BNP. Paciência, 01.04.1857.)

2.5.1.6 O sexo do redator

Pretende-se observar se o sexo (homem/mulher) do escrevente se mostraria ou não relevante no processo de variação entre as formas *tu/você* nas missivas mineiras dos séculos XIX e XX. Parte-se da hipótese de que as mulheres mineiras (assim como as mulheres cariocas) empregariam mais o *você*, que é a forma inovadora não legitimada pela tradição gramatical, mas que apesar disso não é socialmente rechaçada (RUMEU 2013, p. 93-94). Teríamos, nas cartas mineiras, esse mesmo comportamento observado para as escreventes cariocas que pareceram se mostrar propulsoras do inovador *você*?

2.5.1.7 A faixa etária do redator

O controle da *faixa etária* do escrevente é motivado pelo intuito de verificar se os mais jovens já empregariam mais o inovador *você*, apontando assim para a realidade mineira atual que é a de prevalência do *você* no espaço mineiro, segundo Scherre *et al.* (2015). Parte-se

da hipótese confirmada por Rumeu (2013) de que os escreventes jovens tendem a assumir o *você*, o que parece oferecer indícios de um gradual e paulatino processo de mudança linguística, principiado em fins do século XIX e instaurado no período que vai dos anos 20 a 30 do século XX (LOPES & RUMEU, 2007 *apud* RUMEU, 2013, p. 150).

2.5.1.8 As relações interpessoais e sociais

O controle da *relação interpessoal* parte da hipótese de que nas relações em que se tem maior grau de intimidade o missivista recorreria preferencialmente ao *tu*, ao passo que nas relações menos íntimas tenderia a empregar o *você*, cf. Souza (2012, p. 87), no tocante às cartas cariocas oitocentistas e novecentistas. Nesta análise, as *relações interpessoais* controladas são as seguintes: afilhado(a)-madrinha, amigo(a)-amigo(a), avô(ó)-neto(a), esposa-marido, filho(a)-mãe, filho(a)-pai, genro-sogro(a), irmão(ã)-irmão(ã), mãe-filho(a), marido-esposa, neto(a)-avô, noivo(a)-noivo(a), pai-filho(a), primo(a)-primo(a), sobrinho(a)-tio(a), sogro(a)-nora e tio(a)-sobrinho(a).

Ainda que se tenha optado por descrever os *tipos de relações sociais* profícuas em relação às formas *tu* e *você*, é importante esclarecer que tais relações interpessoais travadas entre remetente e destinatário das cartas podem ser sintetizadas, à luz de Brown & Gilman (1960), nos âmbitos da assimetria social (*ascendente* e *descendente*) e da simetria social (entre *iguais*). Neste trabalho, parte-se da hipótese de que, nas relações simétricas e nas assimétricas descendentes (de superior para inferior), o *você* prevaleceria, cf. Lopes & Rumeu (2015, p. 23). Ainda no que se refere às relações de simetria social, a hipótese é a de que as formas *tu* e *você* estariam em variação, cf. Souza (2012, p. 86). No contexto das assimétricas ascendentes (de inferior para superior), a hipótese é a de que o *tu* seria a forma preferida, considerando as constatações de Souza (2012, p.139) embasadas nas cartas cariocas (XIX e XX), ainda que, nas relações de superior para inferior, o *tu* seja a forma esperada na dinâmica do Poder, cf. Brown e Gilman (1960).

2.5.1.9 Período

A partir do controle da variável tempo em períodos (1840-1849; 1850-1859; 1860-1869; 1870-1879; 1880-1889; 1890-1899; 1900-1909; 1910-1919; 1920-1929; 1930-1939; 1940-1949; 1950-1959; 1960-1969; 1970-1979; 1980-1989; 1990-1999; e 2000-2009), pretende-se acompanhar a produtividade das formas de referência ao sujeito de 2SG entre os anos de 1840 e 2009, tendo em vista as hipóteses de inserção do *você* em fins do século XIX e de seu avanço gradual e paulatino no decorrer do século XX, cf. as constatações de Souza (2012, p.137). Quanto ao *vossa mercê*, a hipótese seria a de que as cartas do século XIX possivelmente ainda seriam marcadas pela sua presença.

Uma vez descritos os parâmetros linguísticos e extralinguísticos de análise das formas de referência ao sujeito de 2SG, passa-se à apresentação do método para a análise quantitativa dos dados através da sua submissão ao RBrul.

2.6 O método de levantamento, submissão e análise quantitativa dos dados históricos: o RBrul em cena

Com o intuito de descrever o método de levantamento e submissão dos dados ao RBrul, passa-se à discussão de algumas de suas potencialidades à luz de Gomes (2012), Freitag (2014) e Oushiro (2014). Gomes (2012), no artigo “*Para além dos pacotes estatísticos Varbrul/Goldvarb e RBrul: qual a concepção de gramática?*” compara modelos teóricos relevantes à análise de regras através da comparação entre os programas Varbrul/Goldvarb e RBrul. A autora parte do entendimento de que o estudo sociolinguístico busca descrever o caráter sistemático de uma dada regra variável a partir de seus condicionamentos internos e externos. Desse modo, uma questão enfrentada na investigação é a identificação do efeito dos diversos fatores que concorrem entre si e influenciam no uso de determinada variável e a percepção de como cada fator contribui nesse processo. Para isso, Gomes (2012) faz uma revisão do tema da evolução dos modelos matemáticos empregados na investigação de dados variáveis a partir da

comparação entre os Programas Varbrul/Goldvarb e RBrul, o que a leva a concluir que o RBrul oferece mais recursos ao processo de análise dos dados.

O RBrul é uma ferramenta criada por Daniel Ezra Johnson cujo *software* pode ser adquirido diretamente na internet (<http://www.danielezrajohnson.com/rbrul.html>) para o acesso e o *download* gratuitos dos dados relativos à instalação e à atualização. Para instalar o RBrul, o computador já deve possuir o programa R instalado, plataforma através da qual o RBrul realiza suas rodadas. Entre os objetivos do RBrul, além das funções já realizadas pelo Goldvarb e Varbrul, destaca-se a possibilidade de dar conta de modelos mistos, bem como de estimar efeitos *between-groups* – gênero/sexo e *within groups* – indivíduos. O RBrul tem, ainda, o objetivo de empregar os recursos gráficos do R na análise dos dados variáveis. Nesse sentido, dentre os recursos do RBrul estão, por exemplo, a geração dos pesos relativos e a inserção de um número ilimitado de fatores por grupo. O RBrul reconhece arquivos de dados em formato *.tkn*, usados pelo Goldvarb, no formato *.txt*, usados pelo Varbrul e *.csv*, usados pelo Programa *Excel*, cf. orientações específicas para cada programa (GOMES, 2012, p. 262). A submissão dos dados às rodadas de geração dos pesos relativos pressupõe a conversão do arquivo de dados do *Excel* no formato *.csv* (*comma separate values*). Esta última destaca o ganho metodológico e teórico com o RBrul, ao possibilitar a diferenciação entre variáveis fixas e aleatórias em determinada rodada:

Se dois efeitos são aninhados, isto é, as características de um são mapeadas em outra variável, então o efeito aninhado será o efeito aleatório e o outro será o efeito fixo. Se, por exemplo, as seguintes variáveis estiverem sob consideração, falante e sexo/gênero, falante está aninhado em sexo/gênero e será o efeito aleatório (GOMES, 2012, p. 264).

Como importante referência metodológica nesta pesquisa, Freitag (2014) organizou uma coletânea de textos baseados na metodologia de coleta e manipulação de dados em Sociolinguística. Dentre os textos organizados está o de Oushiro (2014) que trata do emprego de ferramentas computacionais para a manipulação de dados linguísticos. Sob o título *Tratamento de dados com o R para análises linguísticas*, no capítulo 10, Oushiro (2014) recomenda o R como uma estratégia para tratamento estatístico dos dados em análises sociolinguísticas. A autora destaca que o sociolinguista trabalha com amostras com acentuado número de dados cujo tratamento pode se dar, por exemplo, em atividades como análise estatística, verificação de frequência e elaboração de gráficos. Com o RBrul o usuário pode ajustar os seus recursos com os códigos ou *scripts* registrados nas linhas de comando, podendo ser reaproveitados em atividades

posteriores. Para iniciar o emprego da ferramenta, o usuário deve carregar o sistema com alguns pacotes que tornam possível a realização de diversas tarefas no sistema. Tais pacotes são indicados nos manuais disponíveis e podem ser baixados de *sites* que trabalham com o R. Para isso, o linguista deve acessar o Projeto R, no site <<http://cran.r-project.org/>>, instalar o programa, conforme o sistema operacional no qual será realizado o *download* (Linux, MacOS ou Windows).

Considerando as potencialidades da linguagem R, tem-se a sua interface no RStudio que amplia as ferramentas disponíveis através da lista de pacotes instalados e o histórico das linhas de comando (FREITAG, 2014, p. 136). A instalação do RStudio pode ser realizada no seguinte endereço <<http://www.rstudio.com/ide/download/>>. Após a instalação, quando estiver aberto, o RStudio apresentará as seguintes janelas:

- i. *Source*, para visualização e edição de *scripts*;
- ii. *Environment* e *History*, com os objetos carregados na memória do R para a presente sessão e com o histórico de linhas de comando executadas;
- iii. *Console*, no qual as funções e os *scripts* são executados;
- iv. *Files*, *Plots*, *Packages*, *Help* e *Viewer*, respectivamente, para arquivos, gráficos, pacotes, ajuda e visualizador (FREITAG, 2014, P. 136).

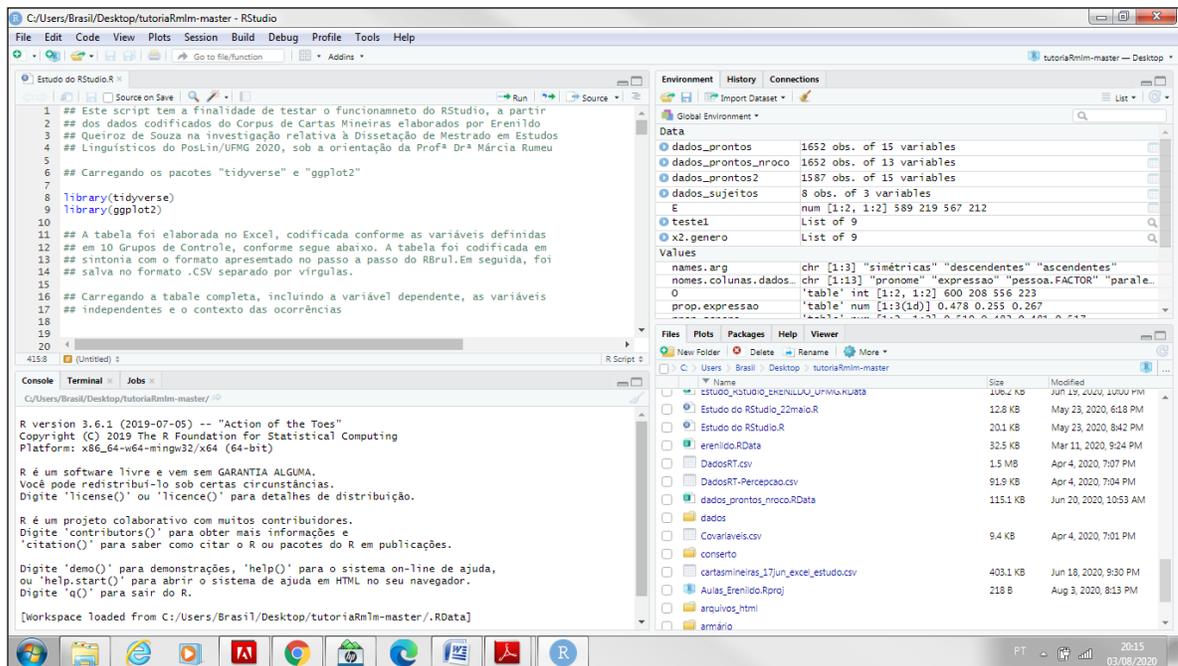


Imagem 20: Interface do RStudio – R versão 3.6.1 (2019-07-05)

Em sintonia com Freitag (2014, p. 137), a imagem (20) ilustra a interface do RStudio, instalado na plataforma *Windows* que serviu para esta análise conduzida também pelos direcionamentos propostos por Oushiro (2017).

Oushiro (2017) propõe uma introdução à estatística voltada para linguistas, tratando de conceitos básicos tais como *tipos de variáveis, tabelas, gráficos, testes de proporção, qui-quadrado, teste-t, correlação, regressão, regressão linear e regressão logística*. A proposta da autora vai no sentido de permitir a identificação de códigos que podem ser utilizados em outros estudos sociolinguísticos. Além disso, a autora descreve comandos que podem facilitar a condução do estudo sociolinguístico tais como o *teste de proporção* e o de *qui-quadrado* cujos códigos foram utilizados com o arquivo de dados em formato *.csv*. No caso específico da proporção, o *script* do RStudio gera uma tabela que realiza a normalização dos dados, principalmente no caso de formas variantes enéarias.

Por fim, levando em conta a força dos pesos relativos para a compreensão da influência das variáveis independentes em relação às escolhas pronominais dos missivistas, vale destacar a contribuição do *Passo a passo para uso do RBrul* de autoria de Avelheda Bandeira & Souza Guerreiro (2019). Trata-se de uma ferramenta também voltada para a operacionalização dos dados a partir do RStudio. Nesse sentido, a primeira parte do *Passo a passo do RBrul* contém a explicação da codificação dos dados baseada no *Microsoft Excel* com orientações quanto à elaboração de uma planilha *Excel* com os grupos de fatores, contemplando a variável dependente, as variáveis independentes e a organização dos dados através da tabela dinâmica, a partir da qual, por sua vez, conduziu-se a elaboração de tabelas e gráficos, organização e correção dos possíveis problemas de codificação dos dados. Após o ajuste dos dados da primeira parte, a planilha está em condições de ser aplicada ao RBrul. De forma geral, a ferramenta possibilita o ajuste dos dados, transformando variáveis fatoriais em numéricas (e vice-versa) bem como a exclusão de dados, entre outros recursos. No menu principal, pode-se comandar também o cruzamento de dados e a função “modelar”. Nesta última, pode-se indicar a rodada multivariacionista a partir dos níveis *step-up & step-down*³⁴ nos quais o investigador tem acesso aos índices probabilísticos (pesos relativos).

³⁴ Cf. Guy e Zilles (2007, p. 62), a rodada multivariacionista *step-up & step-down* corresponde a um método automático de seleção de modelos, que também recebe outras denominações, como *stepwise, forward e backward*. Quanto ao funcionamento, descrito no *Passo a passo* do RBrul, a rodada *step-up & step-down* “inicia o *step-up* apenas com uma dentre as variáveis independentes e, a cada ‘etapa’, acrescenta uma nova variável independente dentre as selecionadas, até chegar à melhor rodada, que é aquela que apresenta a melhor relação entre as variáveis

SÍNTESE DO CAPÍTULO:

O capítulo da fundamentação teórico-metodológica é iniciado com a apresentação do método de trabalho no que se refere à Sociolinguística Histórica (subseção 2.1), enfocando especificamente as noções de *autenticidade*, *autoria* e *validade social e histórica* (HERNÁNDEZ-CAMPOY & SCHILLING, 2012). Na sequência, passa-se, em 2.2 e 2.3, à discussão acerca do gênero “carta” histórica, chegando à descrição das amostras utilizadas nesta dissertação advindas de distintos acervos.

Na seção 2.4, apresenta-se a *Teoria do Poder e Solidariedade*, tendo em vista a abordagem de Brown & Gilman (1960) a partir da dinâmica variável, por exemplo, das formas *tu* e *vous* (francês) dinamizadas pelas relações sociais que as sedimentam em relações sociais assimétricas e simétricas, respectivamente. Ainda no âmbito deste segundo capítulo, passa-se, na seção 2.5, à descrição dos parâmetros linguísticos (expressão do sujeito de 2SG, pessoa verbal, paralelismo formal e semântico) e extralinguísticos (os gêneros “carta pessoal” e “carta de comércio”, sexo, faixa etária, relações interpessoais, sociais e período) que conduziram a análise das formas *vossa mercê*, *você* e *tu* na referência ao sujeito de 2SG.

Por fim, apresenta-se, na seção 2.6, o método de trabalho com a linguagem R, principalmente em relação à sua interface no RStudio através do RBrul capaz de gerar, em rodada multivariacionista, os índices probabilísticos de uma dada regra variável nos níveis *step-up* & *step-down*.

envolvidas no que diz respeito ao condicionamento da variante que se está analisando; em seguida, inicia o *step-down* com todas as variáveis independentes e, a cada ‘etapa’, exclui uma variável que não se mostrou significativa, até chegar idealmente às mesmas variáveis significativas a que se chegou durante o *step-up*.” (AVELHEDA BANDEIRA & SOUZA GUERREIRO, 2019, p. 29).

CAPÍTULO 3. RESULTADOS GERAIS: DESCRIÇÃO E ANÁLISE DOS DADOS EM TERMOS PERCENTUAIS

A proposta deste capítulo é a de apresentação da distribuição dos pronomes-sujeito de 2SG a partir de uma rodada geral feita através do Excel, responsável pela geração dos índices percentuais para a variável dependente eneária em análise: *vossa mercê*, *você* e *tu*. Assim sendo, este capítulo subdivide-se em duas seções principais voltadas para as análises dos fatores linguísticos (a *expressão do sujeito*, a *pessoa verbal* e o *paralelismo formal e semântico*) e extralinguísticos (o *sexo*, a *faixa etária*, as *relações sociais*, o *tempo* e a *carta*).

Considerando que, neste capítulo, estão em discussão os resultados, em termos percentuais, das formas de referência ao sujeito de 2SG, a hipótese é a de que o *você* seria a forma preferida pelos mineiros também em sincronias passadas, considerando principalmente a atual prevalência do *você* em Minas Gerais (SCHERRE *et al.*, 2015; GONÇALVES, 2008; PERES, 2006; HERÊNIO, 2006; COELHO, 1999), em detrimento dos usos mais esporádicos do *tu* na fala mineira como, por exemplo, em São João da Ponte (MOTA, 2008), em Ressaquinha (SILVA, 2017) e em Lontra (REIS, 2019). Para dar conta de tal hipótese, partiu-se para a análise acerca da distribuição das formas *vossa mercê*, *você* e *tu* como uma regra variável eneária (GUY & ZILLES, 2007).

TABELA 3
As formas de referência ao sujeito de 2SG nas cartas mineiras analisadas
(sécs. XIX e XX)

FORMAS DE REFERÊNCIA AO SUJEITO DE 2SG		
VOSSA MERCÊ	VOCÊ	TU
79/1236	569/1236	588/1236
(6,39%)	(46,04%)	(47,57%)

Em termos de resultados gerais, observa-se uma acirrada concorrência entre as formas *tu* (588/1236, 47,57%) e *você* (569/1236, 46,04%) com a surpreendente preferência pelo *tu* (47,57%) nas cartas mineiras, cabendo, em 6,39% dos dados (79/1236), algumas evidências do

vossa mercê que se mostraram restritas ao século XIX. De (31) a (33), ilustram-se os dados das formas de referência ao sujeito de 2SG produtivas nas cartas mineiras analisadas.

(31) [...] Bem sensível me he a noticia do estado do *Padre Americo*, tanto mais *que* avista do *que Vossa Mercê* me diz não me animo a esperar *que* melhore. [...]. (BNP. Paciencia, 07.09.1853.)

(32) [...] Estou daqui torcendo pela sua eleição, mas sempre pensando que *você* não deve se impressionar com um desfecho que lhe seja acaso desfavorável [...]. (AGF. Brasília, 19.06.1963.)

(33) [...] Que maravilhosa és *tu*, minha boa Cecy, que fizeste do mais desiludido dos homens, o mais esperançoso, o mais crente, o mais vibrante? [...]. (JAVS. Machado, 31.12.1924.)

Esses resultados passam pela exclusão das construções imperativas de 2SG. A motivação para a exclusão das construções imperativas desta análise está justificada no fato de que as construções imperativas inflacionariam as evidências de expressão nula do sujeito principalmente no século XX, momento em que o PB passou por uma mudança de parâmetro do sujeito nulo no PB a partir dos anos 30 (DUARTE, 1993, 1995). Outro motivo para a exclusão dos dados de imperativo de 2SG é o fato de se tratar de um fenômeno movido pela alternância entre as formas do indicativo (“vem (*tu*)”), do subjuntivo (“venha (*você*)”) e do imperativo abasileirado (“vem *você*”), como já discutido por Carvalho (2020) inclusive para as cartas mineiras que embasam esta dissertação.

3.1 Os fatores linguísticos: a expressão do sujeito, a pessoa verbal e o paralelismo formal e semântico

Passa-se à exposição dos resultados percentuais para o controle da variável dependente eneária (*vossa mercê*, *você* e *tu*) em relação às variáveis independentes linguísticas (a expressão do sujeito de 2SG, a pessoa (concordância verbal) e o paralelismo formal e semântico), considerando as hipóteses para cada variável conjecturadas no capítulo 2.

TABELA 4
A expressão dos pronomes-sujeito de 2SG: nulo ou pleno.

REFERÊNCIA AO SUJEITO DE 2SG	EXPRESSÃO	
	NULO	PLENO
VOSSA MERCÊ	23/79 (29,11%)	56/79 (70,89%)
VOCÊ	145/569 (25,48%)	424/569 (74,52%)
TU	570/588 (96,94%)	18/588 (3,06%)
TOTAL	738/1236 (60%)	498/1236 (40%)

A análise da expressão das formas *vossa mercê*, *ocê* e *tu* é motivada pela ideia de que apesar de o inovador *ocê* sincronicamente atuar como um legítimo pronome, o seu percurso histórico resguarda evidências da sua origem de base nominal (*vossa mercê*). Uma delas está na expressão (nula ou plena) da forma nominal ou pronominal de referência ao sujeito de 2SG. Assim sendo, faz todo o sentido constatar, com base na TABELA 4, a produtividade do *ocê*, nas cartas mineiras em análise, como sujeito pleno, em **74,52%** dos dados (424 oco), o que também se repete com a forma nominal de tratamento *vossa mercê* que também preferencialmente se mostra plena, em 70,89% dos dados (56 oco), como é típico dos nomes. Em caminho oposto, segue o *tu*, potencialmente nulo, em 96,94% dos dados (570 oco), de modo a conservar a sua origem etimológica que é a de pronome-sujeito latino. De (34) a (36), ilustram-se evidências do comportamento prevalente em relação à expressão do sujeito de 2SG: *vossa mercê* pleno, *ocê* pleno e *tu* nulo.

(34) [...] Bem sensível me he a noticia do estado do Padre Americo, tanto mais que avista do que *Vossa Mercê* me diz não me animo a esperar que melhore. [...]. (BNP. Paciencia, 07.09.1853.)

(35) [...] *Você* pediu a nomeação de dous officiaes [...]. (JPS. Ouro Preto, 07.01.1892.)

(36) [...] para te descançar um pouco. Como *vaes* de arrumação? [...]. (JAG. BH, 27.12.1947.)

Os resultados desta dissertação acompanham os já encontrados em análises sociolinguísticas sobre a configuração pronominal do PB em sincronias passadas. Em relação à expressão do *vossa mercê*, há evidências históricas da sua produtividade como sujeito pleno, em 74% e 65%, respectivamente, nas cartas cariocas dos séculos XVIII e XIX, cf. Rumeu (2004, p. 112). Em relação às missivas cariocas dos séculos XIX e XX, Rumeu (2013) observou que o contexto de sujeito *nulo* mostra-se como um contexto favorecido pelo conservador *tu*, ao passo que o contexto de sujeito *pleno* é acionado pelo *você*, cf. Rumeu (2013, p. 169). Também os resultados de Souza (2012, p. 93) voltados para uma ampliada amostra de cartas cariocas que percorrem um século (1870 a 1970) também apontam para a produtividade do *você* pleno, em 77,4% dos dados, evidenciando tal contexto como um contexto propulsor do *você*, o que está respaldado no PR de 0.744. Em suma, observa-se que os resultados sobre a análise da expressão do sujeito de 2SG (*vossa mercê*, *você* e *tu*) mostram não só os dados de *vossa mercê* pleno, mas também os expõem restritos ao século XIX. Os dados de *você* que se mostram preferencialmente *plenos* repercutem um traço que era peculiar à forma nominal de tratamento originária (*vossa mercê*) como uma evidência de um traço formal que se mantém na forma pronominal *você*, cf. discutido também por Rumeu (2004), Rumeu (2006), Lopes & Rumeu (2007), Lopes & Cavalcante (2011, p. 46). Por outro lado, os dados de *tu* mostram-se preferencialmente nulos em conformidade com o sistema flexional da língua portuguesa que pode contar com uma desinência número-pessoal (DNP) específica de 2SG na forma verbal.

TABELA 5
As formas de referência ao sujeito de 2SG e a pessoa verbal (concordância)

FORMAS DE REFERÊNCIA AO SUJEITO DE 2SG	PESSOA VERBAL (CONCORDÂNCIA)	
	2SG	3SG
VOSSA MERCÊ	-	79/79 (100%)
VOCÊ	2/569 (0,35%)	567/569 (99,65%)
TU	586/588 (99,66%)	2/588 (0,34%)
TOTAL	588/1236 (60%)	648/1236 (40%)

O intuito principal para o controle da *pessoa verbal* é determinar as evidências do *tu* com formas verbais de 3SG (“tu fez”) já na produção escrita mineira dos séculos XIX e XX, ainda que se saiba da preferência dos mineiros pelo *você* (e variações “ocê” e “cê” principalmente) tanto em sua expressão falada (COELHO, 1999; HERÊNIO, 2006; PERES, 2006; MOTA, 2008; GONÇALVES, 2008; SILVA, 2017; REIS, 2019), quanto em sua expressão escrita de sincronias passadas (LOPES & RUMEU, 2015). No que se refere à pessoa verbal, conjectura-se a prevalência do *tu* com formas verbais de 2SG e das formas *vossa mercê* e *você* com formas verbais de 3SG, considerando também o fato de se tratar de cartas produzidas por hábeis redatores. Em termos de repercussões históricas do processo de gramaticalização do *vossa mercê* está o fato de o produtivo e atual *você* harmonizar-se morfossintaticamente às formas verbais de 3SG como uma evidência da sua origem histórica na forma nominal de tratamento *vossa mercê* cujo núcleo é o substantivo *mercê*, responsável por acionar tal concordância verbal, cf. discutido por Faraco (2017 [1996], p. 122), Rumeu (2004), Lopes & Machado (2005), Rumeu (2006).

Tendo em vista o panorama de acirrada alternância entre as formas *tu* e *você* com frequências de 47,57% e 46,04%, respectivamente, em favor das evidências de *tu* na escrita mineira (TABELA 3), observou-se em relação à pessoa (concordância) um quadro de simetria sintática, uma vez que o *você* mostrou-se, em 99,65% dos dados (569oco), harmonizado com formas verbais de 3SG, o que também se aplica categoricamente ao *vossa mercê*, cf. ilustrado em (37) e (38). O *tu* mostrou-se, quase categoricamente, em concordância com formas verbais de 2SG (99,66%, 586oco), cf. (39), ainda que, na maioria das vezes, em sua forma nula. Esses resultados vão ao encontro das constatações de Souza (2012, p. 99) em relação às formas verbais de 3SG como um contexto propulsor do *você* (0.993) nas cartas cariocas.

Dado de *vossa mercê* com verbo na 3SG:

(37) [...] Estimaria bem que *Vossa Mercê trouxesse* (ou se ainda tiver demora que mandasse) <↑~~mandasse~~> a gente que diz ter, pois que virião adiantar tanta couza que vejo ainda bem atrasadas [...]. (BNP. Paciencia, 07.09.1853.)

Dado de *você* com verbo na 3SG:

(38) [...] Cheguei a trez dias, tendo encontrado muitos trabalhos principalmente de correspondencias e mais agora com o nosso projeto que tem excitado discussões e interesses contrarios, tenho mi visto atormentado. Ao chegar soube que seo Pae vendera a casa ao Belisario, como *Você* já sabe, e baracta, foi precipitado demais, elle me disse que escreveo-lhes sobre tudo menunciosamente. [...]. (PB. S.I, 19 de agosto de 1908.)

Dado de *tu* com verbo na 2SG:

(39) [...] Nenhuma outra ambição vi|ve dentro em mim senão a de | merecer o teu amor [...] Digam tudo quanto quiserem, nada me faria [...] se *tu achares* que eu sou bom, honesto, digno de tua estima [...]. (JAVS. Machado, 19.01.1925.)

Por outro lado, chamam a atenção as evidências históricas não só de *tu* com formas verbais de 3SG, mas também as ocorrências de *você* com formas verbais de 2SG. De (40) a (43), ilustram-se tais dados de ausência de concordância verbal na produção escrita dos mineiros: duas ocorrências de *você* com formas de 2SG, em (40) e (41), ilustrados nas imagens 21 e 22, e duas ocorrências de *tu* com formas verbais de 3SG, em (42) e (43). Os dados de *tu* harmonizado às formas de 3SG foram inexpressivos, nesta investigação, como estão ilustrados em (23 e (24), o que se coaduna às baixas frequências de uso do *tu* nas cidades mineiras de São João da Ponte (MOTA, 2008), Lontra (REIS, 2019) e Ressaquinha (SILVA, 2017).

(40) [...] mas se é rico não faz mal, *você* não achas? [...]. (ST. BH, 18.01.1932.)

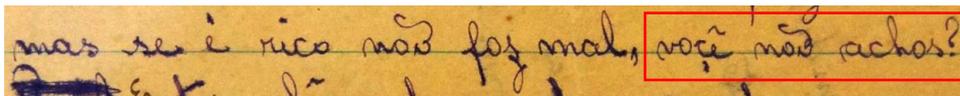


Imagem 21: Carta de ST, com exemplo de uso de *você* com forma de 2SG em uma carta do século XX. (ST. BH, 18.01.1932.)

(41) [...] Terei tôda a benevolência ao tratar do caso no C.T.A da Faculdade e disso *você* estás certa. Mas a lei é muito explícita [...]. (PP. s.l, s.d.)

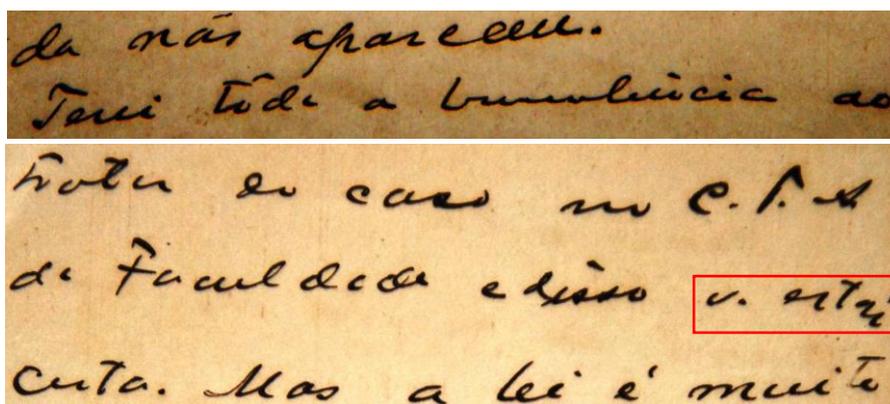


Imagem 22: Carta de PP com exemplo de uso de *você* com forma de 2SG. (PP. s.l, s.d.)

A sentença (40) foi escrita e enviada pela missivista mineira ST, de Belo Horizonte para o irmão, também mineiro, RT, quando este estava morando na Itália, no ano de 1932. Pelo

período da carta, que coincide com o período de reorganização do sistema pronominal do PB (década de 30 do século XX), interpreta-se a convivência das formas *você* e *tu* (“*voçê não achas*”) como um claro indício da instabilidade do sistema em relação às formas de referência ao sujeito de 2SG, na década de 30 do século XX (DUARTE, 1993, 1995). Evidência que também parece se refletir na carta de amizade trocada entre PP e HL (“*voçê estás certa*”), ainda que, nesta carta, não se tenha menção à data da sua produção, cf. (41) ilustrado também na Imagem 22.

(42) [...] porque não te escrevo mais [...] se me deixasse guiar pela penna, não haveria papel que me dêsse, mas é, meu querido José, unicamente defeito de educação que só *tu* poderá corrigil-o [...]. (MIDV. Fazenda, 25.02.1925.)

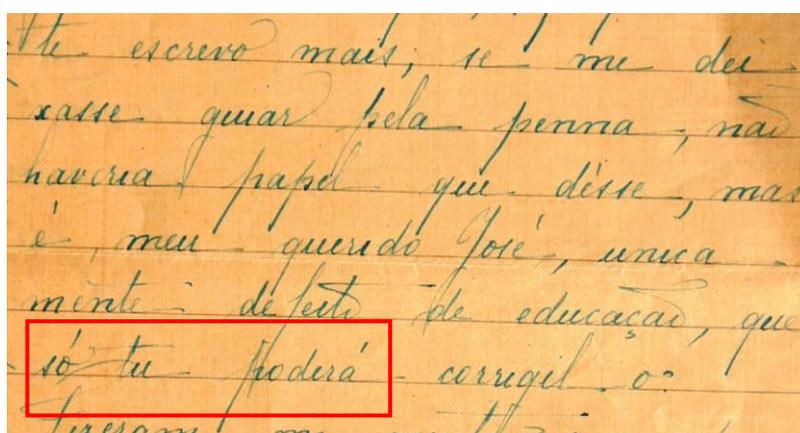


Imagem 23: Carta de MIDV com exemplo de uso de *tu* com forma de 3SG. (MIDV. Fazenda, 25.02.1925.)

(43) [...] E Estella está namorando muito? | Diz a ella que aproveite mesmo a olhos abertos, porque este mundo é horroroso, e a mais agente está é passando por elle, por isso gosa mesmo bicha senão *tu* arrepende. | Arranja um pequeno cheio do “burro” que te possa confortar ao menos como voce está ahi em casa, não achas? [...]. (RT. Roma, 08.12.1931.)

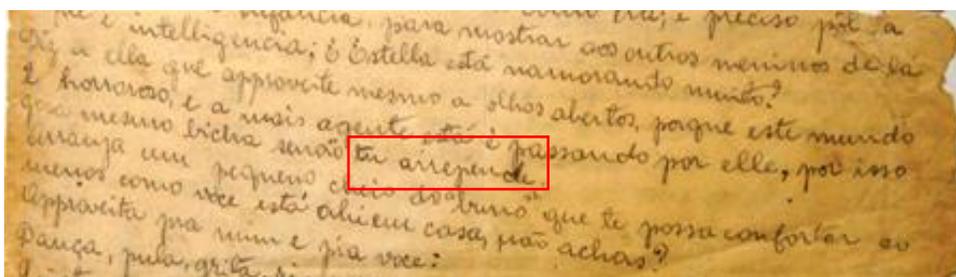


Imagem 24: Carta de RT com exemplo de uso de *tu* com forma de 3SG. (RT. Roma, 08.12.1931.)

As sentenças (42) e (43) apresentam o curioso emprego do *tu* em concordância não-canônica com formas verbais de 3SG, ainda que a literatura registre a preferência dos mineiros pelo *você*, tanto em sua expressão falada, cf. Peres (2006), quanto em sua expressão escrita de períodos históricos, cf. Lopes & Rumeu (2015). São ocorrências quantitativamente inexpressivas

e fortuitas, mas que merecem uma apreciação qualitativa. Em (42), às vésperas do casamento, a mineira MIDV escreve ao noivo JAVS, tratando de assuntos típicos das cartas amorosas trocadas entre namorados, tais como a demora nas respostas e a baixa frequência das missivas enviadas ao noivo. Vale destacar que nessa carta datada de 25.02.1925, MIDV se dirige ao interlocutor em 15 ocorrências, das quais 14 são ocorrências de *tu*-sujeito nulo, ou seja, sujeito marcado pela desinência verbal. Considerando ainda o total de ocorrências relativas às cartas de MIDV na amostra, de um total de 65 dados, 64 são ocorrências de *tu*-sujeito nulo, com apenas uma ocorrência plena, na qual a missivista emprega o *tu*-sujeito pleno combinando com a 3SG. Na sentença (43), tem-se um exemplo de uma carta pessoal familiar enviada da Itália, pelo mineiro RT aos pais dele, que moravam em Belo Horizonte. Na missiva, RT escreve sobre as belezas da Itália e expressa a expectativa quanto ao retorno ao Brasil e a tristeza por estar tão distante da família. É interessante observar que o autor não emprega formas tratamentais para se dirigir aos pais. No entanto, ao pedir notícias a respeito dos irmãos, em (43) RT faz uma referência específica à irmã Estella e pergunta: “E Estella está namorando muito?”. Em seguida, o autor parece dirigir-se diretamente à irmã, fazendo combinações do imperativo afirmativo (que não é foco deste trabalho) com uma sequência de orações nas quais combina o *tu* com a flexão verbal em 3SG ([...] senão *tu arrepende* [...]), emprega o *você*-sujeito concordando com a 3SG ([...] ao menos como *você está* aí em casa, [...]) e usa também o *tu*-sujeito nulo ([...] não *achas?* [...]). Enfim, as sentenças de (40) a (43) foram produzidas entre as décadas de 1920 e 1930, uma fase decisiva para a emergência da forma inovadora *você* no PB, um período da história do Brasil marcado por eventos como a Semana de Arte Moderna em 1922, as revoltas do Movimento Tenentista, a Revolução de 1930, da qual Minas Gerais foi um dos articuladores, e a aprovação do decreto que instituiu o voto feminino no Brasil, em 1932.

TABELA 6
As formas de referência ao sujeito de 2SG e o paralelismo formal e semântico

PARALELISMO FORMAL E SEMÂNTICO	FORMAS DE REFERÊNCIA AO SUJ. DE 2SG		
	VOSSA MERCÊ	VOCÊ	TU
1ª Ocorrência	28/79 (35,44%)	174/569 (30,58%)	114/588 (19,39%)
Formas precedidas por <i>Você</i> -sujeito pleno/nulo	-	226/569 (39,72%)	15/588 (2,55%)
Formas precedidas por <i>Você</i> não-sujeito	-	31/569 (5,45%)	5/588 (0,85%)
Formas precedidas por formas de 3SG (clíticos e possessivos)	12/79 (15,19%)	69/569 (12,13%)	5/588 (0,85%)
Formas precedidas por <i>Tu</i> -sujeito pleno/nulo	-	14/569 (2,46%)	234/588 (39,80%)
Formas precedidas por <i>Tu</i> não-sujeito	-	-	17/588 (2,89%)
Formas precedidas por formas de 2SG (clíticos e possessivos)	-	7/569 (1,23%)	142/588 (24,15%)
Formas precedidas por formas de 2PL (clíticos e possessivos)	3/79 (3,80%)	2/569 (0,35%)	5/588 (0,85%)
Formas precedidas por <i>Vossa Mercê</i> (suj. e não-suj.)	35/79 (44,30%)	-	-
Formas precedidas por Vocativo	-	3/569 (0,53%)	14/588 (2,38%)
Formas precedidas pelo imperativo de 2SG (Subj.)	1/79 (1,27%)	36/569 (6,33%)	8/588 (1,36%)
Formas precedidas pelo imperativo de 2SG (Ind.)	-	7/569 (1,23%)	29/588 (4,93%)
TOTAL	79/1236 (6,39%)	569/1236 (46,04%)	588/1236 (47,57%)

Para o controle do paralelismo formal e semântico, a hipótese é a de que a sequência discursiva iniciada por *você* seja um contexto promissor para a sua produtividade, cf. pensado, inicialmente, por Omena (2003) em relação ao *a gente* e constatado também, posteriormente, por Lopes & Vianna (2012, p. 148) e também por Rumeu (2013) em relação à implementação do *você* no sistema pronominal do PB. Resta saber se as opções de referência ao sujeito de 2SG dos escreventes mineiros combinarão ou não com os pronomes-sujeito em análise (*vossa mercê*, *você* e *tu*) com formas de 2ª e 3ª pessoas gramaticais. Além disso, controlou-se a primeira ocorrência do item pronominal em si. Em síntese, o controle do paralelismo está fundamentado também no intuito de averiguar a questão da presença ou ausência da uniformidade no tratamento através do emprego exclusivo de formas do paradigma de *tu* (forma de 2ª pessoa do singular) ou do paradigma de *vossa mercê* (formas de 3ª pessoa do singular), cf. Brito (2001).

Em relação ao *paralelismo formal e semântico*, constata-se que o *vossa mercê* mostra-se mais produtivo, respectivamente, quando precedido por *vossa mercê* nas funções de sujeito e não-sujeito (44,30%), como a 1ª ocorrência (35,44%) e precedido por formas de 3SG (15,19%), cf. ilustram os dados de (44) a (46). O *você* também se mostra mais produtivo nos contextos de *você-sujeito* (39,72%), de 1ª ocorrência (30,58%) e precedido por formas pronominais de 3SG (12,13%) como se pode verificar de (47) a (49). O *tu* mostra-se mais produtivo precedido por *tu-sujeito* (39,80%), por formas de 2SG (24,15%) e em contexto de 1ª ocorrência (19,39%) como se verifica de (50) a (52). Em síntese, esses resultados indicam que as evidências de *vossa mercê* e *você* estão preferencialmente combinadas com formas do paradigma de 3SG, ao passo que o *tu* harmoniza-se com formas vinculadas ao paradigma de 2SG, confirmando as discussões de Omena (2003), Lopes & Vianna (2012, p. 148), Rumeu (2013) e também sugerindo a escrita mineira mais próxima dos parâmetros da norma-padrão.

Dado de *vossa mercê* precedido de formas de *vossa mercê* (não-sujeito):

(44) [...] Qual quer quantia que lhe adiante, queira comunicarmos por sêr aqui levada em carta, e inderezarse a Vossa mercê. [...] e aceitando³⁵ *Vossa mercê* recomendações de suas Netas, aceite³⁶ também os protestos de estima dos Filhos, Compadres, e Amigos bem Obedientes [...] (BNP. Paciencia, 01.04.1857.)

Dado de *vossa mercê* em contexto de 1ª ocorrência:

(45) [...] Amanhã seguem para ahi minha Mulher, e Filhas que bem saudozo me deixão, mas que fogem se Deos assim quer! Sugeitome pois irremediavelmente á sua Vontade. Oxalá Ellas Lhe encontrem restabelecido de seus encommodos, e que *Vossa mercê* já mais abuze do Vigor que felismente em sua idade ainda goza! [...] (BNP. Paciencia, 14.09.1856.)

Dado de *vossa mercê* precedido por formas de 3SG:

(46) [...] Olinto tem Lhe esperado para fazerLhe sua proposta sobre o Paciencia, e como ancio que *Vossamercê* ainda se demora, disse-me hontem *que* breve ahi aparecerá. [...] (BNP. Sam Vicente, 15.09.1855.)

³⁵ Na sentença (44), a ocorrência de *vossa mercê*-sujeito pleno da oração “[...] aceitando *Vossa mercê* [...]” foi incluída na amostra analisada nesta investigação, tendo como verbo da oração a forma gerundiva “aceitando”.

³⁶ A oração de *vossa mercê*-sujeito nulo “[...] aceite também os protestos [...]” consta como contexto linguístico da sentença (44), porém, não foi selecionada para a amostra, por se tratar de uma ocorrência imperativa.

Dado de *você* no contexto de *você-sujeito*:

(47) [...] Você poderia consultar o Lourenço não só sôbre a corre-|ção da frase como <↑sôbre> sua [...] autoria? Estamos mais esperançosos agora de que Brasília ganhe nôvo impulso com o Prefeito Sette Câmara, meu melhor amigo da infância. De qualquer forma, esperamos que *você* se anime a vir até aqui, para conhecer uma cidade que, mesmo imatura, tem o que mostrar ao visitante. [...]. (AGF. Brasília, 22.11.1961.)

Dado de *você* no contexto de 1ª ocorrência:

(48) [...] Recebi a conta do telefone. *Você* não me deve nada. E eu não lhe deveria a caso? [...]. (OLR. RJ, 28.04.1950.)

Dado de *você* precedido por formas pronominais de 3SG:

(49) [...] Por estes dias sua Mãe escreverá em resposta. Ela e eu fazemos votos para que *você* seja bem sucedido [...]. (AB. BH, 24.01.1956.)

Dado de *tu* precedido por *tu-sujeito*:

(50) [...] como sabes somos em tantos irmãos e o feliz precisa ajudar o infeliz, não *achas*? [...]. (RT. Roma, 25.02.1932.)

Dado de *tu* precedido por formas de 2SG (clíticos e possessivos):

(51) [...] Esperei-te com um apetitoso jantar e fui mais uma vez lograda!... Depois acertaremos as contas!. Tenho recebido as tuas cartas. Quando *vens*? [...]. (MIDV. Fazenda, 02.01.1925.)

Dado de *tu* em contexto de 1ª ocorrência:

(52) [...] Hoje em dia, eu bem podia diser: a unica verdade é amar. Não fôras *tu*, minha terna companheira e a vida para mim seria detestavel! [...]. (JPS. RJ, 14.02.1891)

Passa-se na próxima subseção deste capítulo à descrição-analítica da distribuição dos dados em relação às variáveis extralinguísticas.

3.2 Os fatores extralinguísticos: o sexo, a faixa etária, as relações sociais, o tempo e o gênero textual *carta*

Pretende-se observar se o sexo do escrevente se mostrou ou não relevante à alternância *tu/você* nas cartas mineiras (séculos XIX e XX). Em relação às cartas cariocas, Rumeu (2013) observou as mulheres como as propulsoras do inovador *você*. Seria evidente, nas cartas mineiras, esse mesmo comportamento por parte das escreventes mineiras? Em relação à

faixa etária, a ideia é observar se os redatores mais jovens já empregariam ou não o *você*, o que funcionaria como uma evidência de uma mudança concretizada na 2ª metade do século XIX (cf. TARALLO, 1993, p.99) a partir da inserção gradual e paulatina do *você* ao sistema pronominal do PB.

TABELA 7
As formas de referência ao sujeito de 2SG correlacionadas
ao sexo e à faixa etária dos escreventes

FAIXA ETÁRIA SEXO	FORMAS DE REFERÊNCIA AO SUJEITO DE 2SG					
	VOSSA MERCÊ		VOCÊ		TU	
	HOMEM	MULHER	HOMEM	MULHER	HOMEM	MULHER
CRIANÇA (menor de 14 anos)	5/5 (100%)	-	-	1/1 (100%)	-	-
JOVEM (de 14 a 30 anos)	5/5 (100%)	-	91/116 (78%)	25/116 (22%)	42/107 (39%)	65/107 (61%)
ADULTO (de 31 a 50 anos)	69/69 (100%)	-	187/215 (87%)	28/215 (13%)	297/317 (94%)	20/317 (6%)
IDOSO (mais de 50 anos)	-	-	79/180 (44%)	101/180 (56%)	55/93 (59%)	38/93 (41%)
TOTAL	79/1108 (7%)	-	357/1108 (32%)	155/1108 (14%)	394/1108 (36%)	123/1108 (11%)

Considerando que este trabalho está embasado em *corpora* históricos, há de se trazer à cena o fato de não ser possível uma amostra equânime em relação à distribuição dos tipos de cartas pelas variáveis sociais *sexo* e *faixa etária* dos informantes. Resguardadas as especificidades das amostras em questão, passa-se à descrição dos pronomes-sujeito correlacionados ao *sexo* e à *faixa etária* dos missivistas.

As evidências históricas de *vossa mercê* mostram-se restritas ao século XIX nas amostras de cartas mineiras analisadas. Basicamente, os dados de *vossa mercê* prevalecem entre os adultos, visto que se teve acesso a um número maior de cartas produzidas por adultos na realidade oitocentista. Para o *vossa mercê*, observa-se a sua prevalência nas cartas dos homens adultos (69oco), ao passo que os seus usos esporádicos como consequência também do escasso número de missivas mostram-se entre as crianças (5oco) e os jovens (5oco) do sexo masculino. De (52) a (54), têm-se dados do *vossa mercê*, produzidos por escreventes crianças, jovens e adultos, respectivamente, com destaque para a sentença (53), única ocorrência de *vossa mercê* no século XX.

(52) [...] Meu tio. Depois que para este lugar vim jamais tive a satisfação de vos ver, sem duvida é por que *assim Vossa merce* tem querido, pois não vos era difficil chegar *Vossa merce* té aqui a fim de dar a mim e a todos os nossos muito prazer. Deos permitta que *Vossa mercê* mappareça por cá brevemente. [...]. (JPS. Nossa Senhora do Porto, 10.01.1868.)

(53) [...] Estou muito satisfeito, porque acho que é a materia³⁷ mais difficil que tenho estudado. Mas, a quem devo isto, senão a *Vossa Mercê* que com tanta bondade e desinteressadamente vae me educando! [...]. (AS. BH, 30.08.1904.)

(54) [...] As 6 horas da tarde de hoje aqui chegou Romualdo *que* nos trouxe a sua estimada de hontem; aqual nos causou não pequena surpresa! Na *que Vossa mercê* nos dirigio em data de 18 do passado, dava como decidida a hida de Ovidio para Santa Quitheria, e por isso ficamos certos de *que* elle reconhecendo vantagem desse passo, de bôa mente o tinha aceitado, e persuadimo nos *que* nisso não hia mal. [...]. (BNP. Paciencia, 06.09.1858.)

Para os dados de *você* e *tu*, observa-se um panorama mais diversificado em relação ao *sexo* e à *faixa etária* dos escreventes. Os homens adultos (87%), os homens jovens (78%) e também as mulheres idosas (56%) preferem o *você* para a referência ao interlocutor. Dentre os escreventes mineiros, os homens jovens também se sobressaem em relação ao *você* (em 78% dos dados), o que parece ser sintomático, considerando a integração desse pronome de 2SG ao sistema do PB em legítima variação com o *tu* que, por sua vez, resiste na produção escrita das mulheres jovens (61%), dos homens adultos (94%) e dos homens idosos (59%). Diferentemente dos resultados alcançados por Rumeu (2013) em relação às mulheres cariocas que preferiram o *você*, as mulheres jovens mineiras preferiram o *tu*, o que pode ser interpretado principalmente não só em função do conservadorismo da língua escrita, mas também em virtude do acentuado nível de experiência dessas redatoras com a norma-padrão. De (55) a (59), ilustram-se evidências do *você* na produção escrita de jovens e adultos, bem como do *tu* na produção escrita das mulheres jovens, dos homens adultos e idosos, respectivamente.

Dado de *você* produzido por mulher jovem:

(55) [...] Não espero o fim do mês para mandar-lhe minha carta. *Você* sabe: gosto demais de escrever-lhe. E minha alegria é dobrada, quando recebo sua amável resposta. [...]. (RCAM. BH, 20.10.1977.)

Dado de *você* produzido por mulher adulta:

(56) [...] Querida Henriqueta *Você* tem razão: preciso continuar a viver, embora sem a grande alegria que me dava o amor conjugal, verdadeiro, profundo, inabalável, construído sobre o alicerce da fé em Deus e da confiança naquele que eu, em hora feliz e abençoada, escolhi por companheiro. [...]. (MJLB. Lambari, 25.06.1951.)

³⁷ O missivista estava se referindo a um exame escolar de geometria.

Dado de *tu* produzido por mulher jovem:

(57) [...] Desejo que esta vá te encontrar gozando perfeita saúde ao par de grandes felicidades. Eu como *sabes* gozo saúde mas sempre pelejando com os filhos na maior dificuldade da vida [...]. (OAAP. Lagoa Santa, 28.02.1918.)

Dado de *tu* produzido por homem adulto:

(58) [...] Lucia, se *estás* disposta a não ter conforto, vem desde já; se não, espera que a nova [*tribu*] Delio – Celina volta, o que se dará dentro de poucos dias. Para nós, em qualquer hipótese, é um prazer ter-te aqui; *tu resolverás*. Pretendemos seguir para Minas na primeira quinzena de fevereiro. [...]. (AM. RJ, jan. 1944.)

Dado de *tu* produzido por homem idoso:

(59) [...] *Tu*, nunca te *podias* ter feito uma édeia nem apossimativa do que é realmente na minha terra... Não é verdade? Mas agora *tu podés* certamente apreciar os raros incantos Italicos [...]. (ET. BH, 20.01.1932.)

Neste trabalho, para a discussão acerca da interferência das relações sociais na referência ao interlocutor, parte-se da hipótese de que o *você* seria produtivo não somente nas assimétricas descendentes (de superior para inferior), mas também nas simétricas, cf. discutido por Lopes & Rumeu (2015, p. 23), nas cartas mineiras dos séculos XIX e XX. Por outro lado, o *tu* teria a sua produtividade resguardada principalmente pelas relações simétricas (LOPES & RUMEU, 2015, p. 23). No que se refere às simétricas, a hipótese é a de que o *tu* e o *você* estariam, em algum nível, em variação, tendo também em vista os resultados de Souza (2012, p. 86) para as cartas cariocas oitocentistas e novecentistas. Considerando a origem tratamental do *vossa mercê*, prevê-se que também o seu uso esteja marcado, nas cartas mineiras oitocentistas, pelas relações assimétricas (de inferior para superior). Passa-se, na TABELA 8, à análise da distribuição dos dados de *vossa mercê*, *você* e *tu* nas relações sociais simétricas e assimétricas ascendentes e descendentes nas cartas mineiras em análise.

TABELA 8
As formas de referência ao sujeito de 2SG correlacionadas
às relações sociais simétricas e assimétricas (ascendentes e descendentes)

RELAÇÕES SOCIAIS ³⁸	FORMAS DE REFERÊNCIA AO SUJEITO DE 2SG		
	VOSSA MERCÊ	VOCÊ	TU
ASCENDENTE (de inferior para superior)	79/79 (100%)	17/569 (2,99%)	6/588 (1,02%)
DESCENDENTE (de superior para inferior)	-	107/569 (18,8%)	183/588 (31,12%)
SIMÉTRICA (entre iguais)	-	445/569 (78,21%)	399/588 (67,86%)
TOTAL	79/1236 (6,39%)	569/1236 (46,04%)	588/1236 (47,57%)

De modo geral, considerando a oposição entre os contextos das relações sociais simétricas e assimétricas (ascendentes e descendentes), em relação a cada variante (*vossa mercê*, *você* e *tu*), verifica-se, com base na TABELA 8, que, nas relações assimétricas ascendentes (60), observa-se a prevalência do *vossa mercê* (100%). Por outro lado, tanto o *você* (61) quanto o *tu* (62) mostraram-se produtivos nas relações simétricas com índices percentuais de 78,21% e 67,86%, respectivamente.

Dado de *vossa mercê* em relação assimétrica ascendente (de inferior para superior):

(60) [...] Reconheço que não devo amargurala, contando-lhe cousas que vão incidir dolorosamente no seu espirito; [...] porque estando em longe e ausente do lar, tenho a certeza de que *Vossa Merce* fica inquieta e acabrunhada [...]. (NCS. Ouro Preto, 07.02.1895.)

Dado de *você* em relação simétrica:

(61) [...] Logo que surja, *você* será, como sempre, das primeiras a receber um exemplar. [...]. (AGF. Brasília, 06.11.1968.)

Dado de *tu* em relação simétrica:

(62) [...] Porque não me *escreveste* mais? A tua letra é muito grande. Por mais que me *escreves*, sempre me *escreves* menos do que eu a ti. [...]. (AR. 10.09.1925)

³⁸ Reflete o universo de dados, que não é balanceado, levando-se em conta o fato de, em sociolinguística histórica, o pesquisador ficar restrito aos dados que sobreviveram à ação do tempo.

O fato de o *vossa mercê* mostrar-se categoricamente produtivo nas relações marcadas pela assimetria ascendente (de inferior para superior), em 100% dos dados, vai ao encontro de resultados de pesquisas embasadas também em missivas pessoais mineiras (séculos XIX e XX) cujos dados de *vossa mercê* manifestaram-se também categoricamente nas relações de inferior para superior, cf. discutido por Lopes & Rumeu (2015, p. 19). Por outro lado, as formas *você* e *tu* se distribuem, ainda que com distintos níveis de produtividade, pelas relações simétricas e assimétricas (ascendentes e descendentes). Nesse sentido, observa-se que tanto o *você*, quanto o *tu* prevalecem, em primeiro lugar, nas relações simétricas, alcançando, respectivamente, os índices de 78,21% e 67,86%, e, em segundo lugar, nas assimétricas descendentes (de superior para inferior), com os percentuais de 18,8% e 31,12%, respectivamente. Esses resultados mostram-se interessantes ao foco principal desta pesquisa, uma vez que evidenciam não só o fato de o *você* ter se espreado por relações simétricas e assimétricas, mas também o fato de ter prevalecido nas relações simétricas (78,21%), assim como o *tu* (67,86%). Essas distribuições das formas *tu* e *você* parecem já apontar não só para um *você* que assume os mesmos campos funcionais do *tu* na escrita mineira, mas também para a prevalência do *você* nas relações simétricas, corroborando os resultados alcançados por Souza (2012, p. 139) em relação às cartas cariocas. Em (63) e (64), observam-se evidências de *tu* e *você* em contexto de assimetria social descendente, contexto de alta produtividade de tais formas pronominais de 2SG.

Dado de *tu* em relação assimétrica descendente (de superior para inferior):

(63) [...] Aqui em casa não será como nos annos anteriores pois falta-nos uma pessoa, e está *es tu* para completa felicidade nossa [...]. (CT. BH, 13.12.1931.)

Dado de *você* em relação assimétrica descendente (de superior para inferior):

(64) [...] faço votos a Deus para que *voce* tenha muitas felicidades, paz e saude no novo ano. [...]. (AEM. MRVL. S.l., 29.12.1944.)

No âmbito das relações assimétricas ascendentes, contexto historicamente fomentador das formas nominais de tratamento (como é o caso do *vossa mercê*), também estão as seis ocorrências de *tu*, marcadas pelas relações entre filho e pai e sobrinho e tio, cf. está exemplificado de (65) a (70). Como interpretar esses dados do tão propalado íntimo *tu* em contexto de assimetria social ascendente (de inferior para superior)? Ao correlacionar tais

resultados voltados às cartas mineiras aos observados por Lopes & Rumeu (2015, p. 19) para as cartas cariocas, constata-se que também para os escreventes cariocas foram levantadas, na 1ª metade do século XX (1900-39), evidências de *tu*, em 63% dos dados, no contexto de assimetria ascendente. Na escrita mineira, também foram levantados, já na 1ª metade do século XIX (1848 e 1854), dados de *tu* nas relações de inferior para superior (6oco), o que parece evidenciar a propagação de uma dinâmica social de maior proximidade mesmo nas dinâmicas de assimetria social ascendente manifestas na relação entre filho e pai e entre sobrinho(a) e tio. Ainda no contexto de assimetria social ascendente, estão as dezessete evidências de *você* calcadas na dinâmica sobrinha e tia, expostas em (71) e (72), o que parece expor o *você* em franca disseminação também pelas relações sociais assimétricas ascendentes na expressão escrita mineira do século XIX.

Dado de *tu* na relação entre filha e pai:

(65) [...] Aninha esta as voltas [...] e pergunta como já *vaes*? Papai [...]. (ACNP. Cocais, 21.07.1848.)

Dado de *tu* na relação entre filha e pai:

(66) [...] como nos prometestes, eque me *escrevas* sempre. [...]. (ACNP. Sam Vicente, 07.09.1854.)

Dados de *tu* na relação entre sobrinho e tia(o):

(67) [...] Como *vais* passando de saúde. Espero que tudo bem. Eu por aqui estou ótimo e com saudades da senhora [...]. (FRA. RJ, 11.12.1968.)

(68) [...] *Podes* reformar o Dirictorio consulte o Dr Jose Pedro Araujo [...]. (JPS. s.l., s.d.)

(69) [...] Faz paralisar todas as medidas [...] Fico a espera do que *decidires*. [...]. (JPS. s.l., s.d.)

(70) [...] Sendo isto possível, *farás* um acto de justiça. [...]. (JPS. SP, 08 e 09.06.1884.)

Dado de *você* na relação entre sobrinha e tia:

(71) [...] *Você* já deve saber que estamos em condições de hospedá-la. [...]. (AVP. Santiago, 29.03.1962.)

Dado de *você* na relação entre sobrinha e tia:

(72) [...] Queria dizer-lhe tia, que o Neruda voltou [...] *Você* deve ter recebido meu recado quanto ao endereço do Tiago [...]. (AVP. Santiago, 06.11.1962)

Passa-se, no GRÁFICO 3 e na TABELA 9, à distribuição das formas de referência ao sujeito de 2SG pelas relações sociais no eixo do tempo.

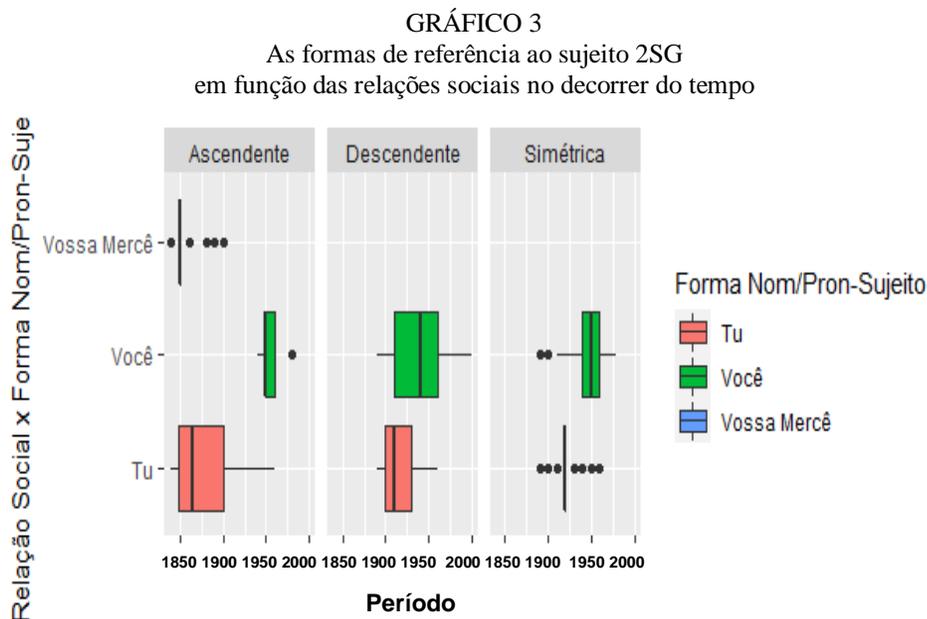


TABELA 9
As formas *Vossa Mercê*, *Tu* e *Você* distribuídas pelas relações sociais ao longo do tempo

REL. SOCIAL	VD	PERÍODO														TOTAL GERAL		
		1840-1849	1850-1859	1860-1869	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	1960-1969	1970-1979	1980-1989			2000-2009
ASSIM. ASCENDENTES	VM ^{CE}	2 2%	67 67%	5 5%	2 2%	2 2%	1 1%	-	-	-	-	-	-	-	-	-	79 79%	100/ 1194 (8%)
	TU	1 1%	1 1%	-	1 1%	-	-	-	-	-	-	-	1 1%	-	-	-	4 4%	
	VOCÊ	-	-	-	-	-	-	-	-	-	1 1%	8 8%	7 7%	-	1 1%	-	17 17%	
ASSIM. DESCENDENTES	TU	-	-	-	-	15 5,5%	49 18%	52 19%	8 2,9%	49 18%	5 1,84%	-	2 0,74%	-	-	-	180 66,18%	272/ 1194 (23%)
	VOCÊ	-	-	-	-	10 3,68%	10 3,68%	9 3,31%	-	9 3,31%	12 4,41%	5 1,84%	21 7,72%	9 3,31%	1 0,37%	6 2,21%	92 35,82%	
SIMÉTRICAS	TU	-	-	-	-	23 2,8%	11 1,34%	5 0,61%	284 34,6%	39 4,74%	4 0,49%	16 1,95%	7 0,85%	-	-	-	389 47,32%	822/ 1194 (69%)
	VOCÊ	-	-	-	-	5 0,61%	9 1,1%	5 0,61%	5 0,61%	84 10,2%	75 9,12%	121 14,7%	93 11,3%	28 3,41%	8 0,97%	-	433 52,68%	
TOTAL GERAL		3	68	5	3	55	80	71	297	181	97	150	131	37	10	6	1194	

Em análise estão 1194 ocorrências (1194/1236oco), uma vez que foram excluídas as 42 produzidas em cartas sem datação. O GRÁFICO 3, gerado a partir de uma planilha no formato

.csv do RStudio (RBrul), pode ter a sua interpretação clarificada também à luz dos dados quantitativos (dados brutos e percentuais) das formas *vossa mercê*, *você* e *tu* distribuídos pelas relações sociais no eixo do tempo (cf. TABELA 9).

Considerando que a maior distribuição dos dados está nas relações simétricas, em 69% dos dados, inicia-se a análise das formas de referência ao sujeito de 2SG conforme o GRÁFICO 3 em correlação à TABELA 9. Nas relações sociais simétricas, as ocorrências de *tu* estão concentradas entre os anos de 1920-29, o que está indicado pela posição da mediana do *boxplot* e também na TABELA 9 através do índice de 34,6% de *tu* (284/822oco) em relação à superior incidência percentual de *você* com 0,61% (5/822oco). Embora em menor quantidade, a linha horizontal no 1º quartil indica que há dados de *tu*, nas relações simétricas, distribuídos a partir da década de 1890 (23/822oco). Para os dados de *você*, a posição da mediana evidencia um maior nível de simetria, o que indica que as ocorrências estão distribuídas de forma mais equilibrada ao longo do tempo (principalmente entre as décadas de 1930 e 1960), com uma maior concentração de dados na década de 50 (121/822oco = 14,7%). Os dois pontos isolados à esquerda da linha horizontal representam dados de *você* discrepantes em relação aos demais dados na virada do século XIX para o século XX, cf. (73), (73a), (73b), (73c), (73d).

Dados de *você* na década de 1890 (de um total de 5oco de *você*):

(73) [...] Precisas viver; acho conveniente que a Dona Nicota e o Neusinho venhão ficar uns tempos em tua casa cuidando dos meninos e que *Você* venha passar uns tempos commigo. [...]. (JPS. Caeté, 29.12.1896.)

(a) [...] comprei um corte de seda para você; [...] por que *você* precisa de um bom vestido [...]. (JPS. RJ, 15.02.1891.)

(b) [...] você manda fazer a mudança [...] *Você* não quis mandar medida para um vestido [...]. (JPS. RJ, 15.02.1891.)

(c) [...] Vou passar-te telegrammas [...] Recebeu o berço? [...] Como é que *você* diz que eu não lembro [...]. (JPS. RJ, 09.11.1890.)

(d) [...] da veneração por tua esposa, [...] e que *Você* não deixaria tambem soffrer nunca um filho meo... [...]. (JPS. Caeté, 29.12.1896.)

Dados de *você* na década de 1900 (de um total de 9oco de *você*), início do século XX, cf. (74 a-g):

(74) [...] como *Você* já sabe [...] representações que *Você* prometteo mandar [...]. (PB. S.I., 19.08.1908.)

(a) [...] Vejo que não podes demorar longe dos seus [...] tudo isto eu desejaria que *você* examinasse bem [...]. (JPS. BH, 23.05.1908.)

- (b) [...] eu desejaria que *você* examinasse bem, não para *fazer* um relatorio [...]. (JPS. BH, 23.05.1908.)
- (c) [...] eu desejaria que *você* examinasse bem, não para fazer um relatorio, mas para *ficar* sabendo bem nitidamente [...]. (JPS. BH, 23.05.1908.)
- (d) [...] que *você* examinasse bem, não para fazer um relatorio, mas para ficar sabendo bem nitidamente e *poder* dar uma opiniao [...]. (JPS. BH, 23.05.1908.)
- (e) [...] que *você* examinasse bem [...] que quantidade de dinheiro *precisa* [...]. (JPS. BH, 23.05.1908.)
- (f) [...] Olha Calogeras, isso de codigos de autonomias [...] *Voce* tem-no largo e forte [...]. (JPS. Caeté, 25.02.1905.)
- (g) [...] sabes aquellas nomeações de supplentes a juizes [...] respondeu-me *Você* [...]. (GF. Mariana, 27.12.1906.)

Em relação especificamente aos dados de *tu* das relações simétricas, os dados mostram-se discrepantes nas amostras de carta em análise. Observe-se a ausência do retângulo no painel e a presença marcante da mediana (linha vertical espessa), evidenciando que os dados de *tu* estão concentrados na década de 1920, cf. ilustrado em (75). Trata-se de 284 ocorrências concentradas na década de 1920-1929 (284/389). Por outro lado, observa-se que *tu* e *você* parecem disputar o espaço funcional desde o final do século XIX até a década de 1969, sendo que a partir da década de 1930, o *você* apresenta maior produtividade do que o *tu*. Confirma-se assim a hipótese de que o *você* (433/822, 52,68%) se mostraria mais produtivo do que o *tu* (389, 47,32%) nas relações simétricas, cf. também observado por Lopes & Rumeu (2015, p. 23) para as cartas mineiras e por Souza (2012, p. 86) para as cartas cariocas. O fato de as evidências de *tu* nas relações simétricas restringirem-se aos *outliers* do *boxplot* revela que há discrepância entre os dados de *tu*, cf. já comentado em relação à concentração dos dados de *tu* na década de 1920.

Dado de *tu* da década de 1920 (de um total de 284oco de *tu*):

(75) [...] *podes* ficar certa de que eu sou digno da confiança que em mim *depositas* [...]. (JAVS. Machado, 28.01.1925.)

A década de 1920-29 destaca-se, no âmbito desta amostra, por apresentar um maior número de ocorrências de formas pronominais. Trata-se de 297 ocorrências (297/1194) distribuídas pelas relações simétricas (289 oco) e assimétricas descendentes (8 oco). As 289 oco de *tu* foram produzidas por 7 autores, sendo que 216 das ocorrências foram produzidas por tão somente um único autor, o mineiro JAVS, em suas cartas amorosas à noiva MIDV. Embora os

dados quantitativos da década de 1920 destoem numericamente dos demais dados, parecem estar em sintonia com o período histórico do PB (2ª década do século XX), marcado linguisticamente pela produtividade do conservador *tu*. Nesse sentido, os dados respaldam o fato de se lidar com fontes históricas que sobreviveram ao tempo, configurando, pois, o desafio do sociolinguista trabalhar da melhor forma possível com os dados que sobreviveram à ação do tempo (LABOV, 1994).

Nas relações assimétricas descendentes (de superior para inferior), a referência ao sujeito de 2SG se dá através das formas *tu* e *você*. No *boxplot* do *você*, a posição da mediana indica que há certa disparidade na distribuição dos dados, concentrados, por sua vez, entre os anos de 1940-49 (1940=12/272oco) e de 1960-69 (1960=21/272oco). Embora em menor quantidade, o 1º quartil (cujo limite máximo é representado pela linha vertical esquerda do *boxplot* de *você* descendente) indica que há dados de *você*, nas relações assimétricas descendentes, distribuídos a partir da década de 1890 (10/272oco) até a de 1910 (9/272oco); o 2º quartil está representado pela mediana, entre 1930 e 1940. O GRÁFICO 3 não aponta dados isolados de *você* nem de *tu* nas relações assimétricas descendentes. No entanto, há 6 décadas com dados abaixo de 10 oco de *você* (1910, 1920, 1950, 1970, 1980 e 2000) dentre as quais se destaca a década de 1980 com apenas 1 oco, cf. (76).

Dado de *você* na década de 1980 (única ocorrência de *você*):

(76) [...] *Você* poderia colaborar conosco? [...]. (HL. BH, 19.05.1982.)

Os dados de *tu* nas relações assimétricas descendentes também evidenciam algum nível de disparidade na distribuição dos dados (GRÁFICO 3). A posição da mediana revela que o maior percentual de dados está concentrado na década de 1910, em 19% dos dados (52/272), cf. TABELA 9. De forma geral, os dados de *tu* estão mais equanimemente distribuídos nos anos de 1900-09 (49/272, 18%) e 1930-09 (49/272, 18%), circunscritas pelas linhas verticais (esquerda e direita, respectivamente) da caixa do GRÁFICO 3. A partir da década de 1940 (TABELA 9), os percentuais de *tu*, exemplificados em (77) e (78), mostram-se em declínio, passando de 18% (49/272), na década de 30, para 1,84% (5/272), na década de 40 até alcançar, na década de 60, 0,74% (2/272).

Dados de *tu* na década de 1960 (2/23 oco de *tu* para 21/23 de *você*):

(77) [...] Bernardino de *teres* chegado ahi com bôa viagem graças a Deus, e das boas noticias que me *dás* dos nossos negócios [...]. (JAG. BH, 18.12.1962.)

(78) [...] Paulo Voce já está entrando no negocio [...]; venda mais [...] Fiquei satisfeito Bernardino de *teres* chegado ahi com bôa viagem [...]. (JAG. BH, 18.12.1962.)

Nas relações assimétricas descendentes (de superior para inferior), as formas *tu* e *você* disputam o espaço de atuação funcional com maior concentração de dados entre as décadas de 1910 (*tu*) e de 1940 (*você*), levando-se em conta a posição da mediana. Os dados de *você* (desde 1890 até 2000) distribuem-se mais equilibradamente ao longo do tempo do que os dados de *tu* (1890 até 1960), ainda que, no cômputo geral, o *tu* (180/272) prevaleça em relação ao *você* (92/272) no âmbito das relações assimétricas descendentes. No entanto, da década de 1970 em diante, a forma *você* é categórica nas assimétricas descendentes, confirmando a hipótese de que, nas relações sociais assimétricas descendentes, a partir da década de 1960, o *você* seria a forma mais produtiva, cf. discutido por Lopes & Rumeu (2015, p. 23), ainda que, até a década de 1930, o *tu* tenha prevalecido nas assimétricas descendentes das cartas mineiras em análise.

Nas relações assimétricas ascendentes (de inferior para superior), a quantidade de ocorrências é baixa (100/1194, 8,38%), especialmente quando o enfoque são apenas as formas *tu* e *você*. De um total de 100 ocorrências, apenas 21% dados (21/100) são da alternância *tu/você*. As demais 79 ocorrências são de *vossa mercê* cuja distribuição dos dados também discrepa em relação aos dados de *tu* e de *você*. Isso quer dizer que os dados de *vossa mercê* não estão distribuídos equanimemente ao longo do tempo, mas estão concentrados (cf. mediana) na década de 50 do século XIX (1850), o que corresponde à ausência de caixa no *boxplot* de *vossa mercê* ascendente. Tais dados referem-se às ocorrências das cartas de BNP ao Barão de Cocais. Além de concentrarem-se em uma única década (1850) remetem a um único missivista. O ponto isolado à esquerda da mediana possivelmente representa os 2 dados das sentenças (79) e (80), ambos da mesma carta. Os demais dados discrepantes de *vossa mercê* estão representados pelos 4 pontos à direita da mediana, que representam dados isolados de *vossa mercê*, cf. exemplificado de (81) a (84). Confirma-se, na amostra analisada, a hipótese de que o *vossa mercê* apresentaria maior produtividade no século XIX do que no século XX como também observado, entre os anos de 1850-79, por Lopes & Rumeu (2015, p. 19) também para as cartas mineiras, tendo em vista se

tratar de uma forma nominal que já teria se tornado arcaica desde os séculos XVII e XVIII, cf. Faraco (1996 [2017]).

Dados de *vossa mercê* na década de 1840 (2/3 oco de *vossa mercê* para 1/3 de *tu*):

(79) [...] que *Vossa Mercê* si satisfizesse com o preço [...] foram vendidas na Itabira a 3760 réis como verá da Conta junta [...]. (BNP. Cocais, 21.07.1848.)

(80) [...] Estimei muito que *Vossa Mercê* satisfizesse com o preço que [...] tirado em Maio, e agora [...]. (BNP. Cocais, 21.07.1848.)

Dado de *vossa mercê* na década de 1860:

(81) [...] que estas tortas linha vá encontrar a *Vossa merce*; gozando perfeita saude [...]. (JPS. Ouro Preto, 21.12.1869.)

Dado de *vossa mercê* na década de 1880:

(82) [...] *Vossa merce* naturalmente deve [...]; e a certidão *Vossa merce* há de tirar em duplicata. [...]. (JPS. Ouro Preto, 10.02.1882.)

Dado de *vossa mercê* na década de 1890:

(83) [...] certeza de que *Vossa Merce* fica [...] entregou a *Vossa Merce* todos os livros que mandei? Peço que entregue ao José [...]. (NCS. Ouro Preto, 07.02.1895.)

Dado de *vossa mercê* na década de 1900 (única ocorrência no século XX, na amostra):

(84) [...] Mas, a quem devo isto, senão a *Vossa Mercê* que com tanta bondade e desinteressadamente vae me educando! [...]. (AS. BH, 30.08.1904.)

Nas relações ascendentes, ainda que se tenha em cena poucas ocorrências do *tu* (4/100, 4%), a mediana do GRÁFICO 3 indica a concentração dos dados na década de 1850-1859, cf. ilustrado em (85). Observem-se também as evidências do *tu* expostas de (86) até (88).

Dados de *tu* nas décadas de 1840, 1850, 1860 e 1880:

(85) [...] como nos prometestes, eque me *escrevas* sempre. [...]. (BNP [ACNP]. Sam Vicente, 07.09.1854.)

(86) [...] Aqui mesmo vos direi um [...] Aninha esta as voltas [...] e pergunta como já *vaes*? Papai [...]. (BNP [ACNP]. Cocais, 21.07.1848.)

(87) [...] pedir-lhe [...] de emprestar-me 200:000 reis [...] Sendo isto possível, *farás* um acto de justiça [...]. (JPS. SP, 08 e 09.06.1884.)

(88) [...] Como *vais* passando de saúde. [...]. (FRA. RJ, 11.12.1968.)

(89) [...] Como *você* tem passado? [...]. (MAVP. Campanha, 13.12.1983.)

(90) [...] Fazendo-lhe uma visitinha [...]. Terei muito gôsto que *você* venha. [...]. (CLB, Lambari, 25.11.1941.)

O *boxplot* da forma *você* nas relações assimétricas ascendentes (de inferior para superior) apresenta uma configuração um pouco diferente das demais. A posição da mediana coincidindo com a linha vertical que marca o limite esquerdo do *boxplot* indica que as ocorrências não estão distribuídas de forma equilibrada ao longo do tempo. O ponto isolado à direita do retângulo indica dados discrepantes na década de 1980, cf. sentença (89). Além desse dado isolado na década de 1980, há também a década de 1940 com apenas uma ocorrência de *você* também na dinâmica de assimetria ascendente (TABELA 9), cf. está ilustrado em (90). Ainda nas assimétricas ascendentes, confirma-se também a hipótese de que o *você* apresentaria maior produtividade do que o *tu* nas cartas mineiras, cf. conjecturado por Souza (2012, p. 86) em relação às cartas cariocas, ainda que os resultados de Souza (2012, p. 139) tenham evidenciado as relações assimétricas ascendentes como um contexto desfavorecedor do *você*.

Nesta análise, considerando a hipótese de que, no século XIX, a variante *tu* seria a mais profícua e que, ao longo do século XX, o *você* se tornaria mais produtivo, cf. observado Lopes & Rumeu (2015, p. 19) para as cartas mineiras, a partir do ano 1900, e por Souza (2012, p. 87) a partir da década de 30 do século XX, para as cartas cariocas (sécs. XIX e XX), passa-se especificamente aos dados correlacionados ao tempo e ao número de cartas mineiras (cf. TABELA 10). Pretende-se expor, através da TABELA 10, a distribuição dos dados, principalmente em relação ao número de cartas, o que é consequência deste trabalho lidar com os documentos históricos que chegaram (até o momento) às mãos do linguista-pesquisador, cabendo a este fazer o melhor uso possível dos escassos dados de sincronias passadas.

Conforme a TABELA 10, dentre as 1194 ocorrências de formas de referência ao sujeito de 2SG, tendo em vista as cartas datadas das amostras de cartas em análise, observam-se as evidências de *vossa mercê*, restritas ao século XIX (79 oco), de *você* (542 oco) e *tu* (573 oco) distribuídas principalmente pelo decorrer do século XX. Em termos gerais, constata-se que há uma acirrada concorrência entre as formas *tu* (573 oco) e *você* (542 oco) nas cartas mineiras analisadas. Considerando também o número de cartas levantadas, observa-se que o *tu* prevalece, principalmente, nas 1^{as} décadas do século XX (entre 1900 e 1930 com 497 oco), ao passo que o *você* predomina entre 1930 e 1960 (436 oco).

TABELA 10
Distribuição geral das formas de referência ao sujeito de 2SG por carta no eixo do tempo
(principalmente entre os séculos XIX e XX)

SÉCULOS	XIX						XX										XXI	TOTAL	
ANOS	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000		
CARTAS	1	22	2	-	2	13	22	21	44	36	38	43	38	16	8	-	1	307 ³⁹	
DADOS (OCO)	3	68	5	-	3	55	80	71	297	181	97	150	131	37	10	-	6	1194 ⁴⁰	
TOTAL DE CARTAS	40						266										1	307	
DADOS	VM ^{CE}	2	67	5	-	2	2	1	-	-	-	-	-	-	-	-	-	-	79/1194
	TU	1	1	-	-	1	38	60	57	292	88	9	16	10	-	-	-	-	573/1194
	VOCÊ	-	-	-	-	-	15	19	14	5	93	88	134	121	37	10	-	6	542/1194

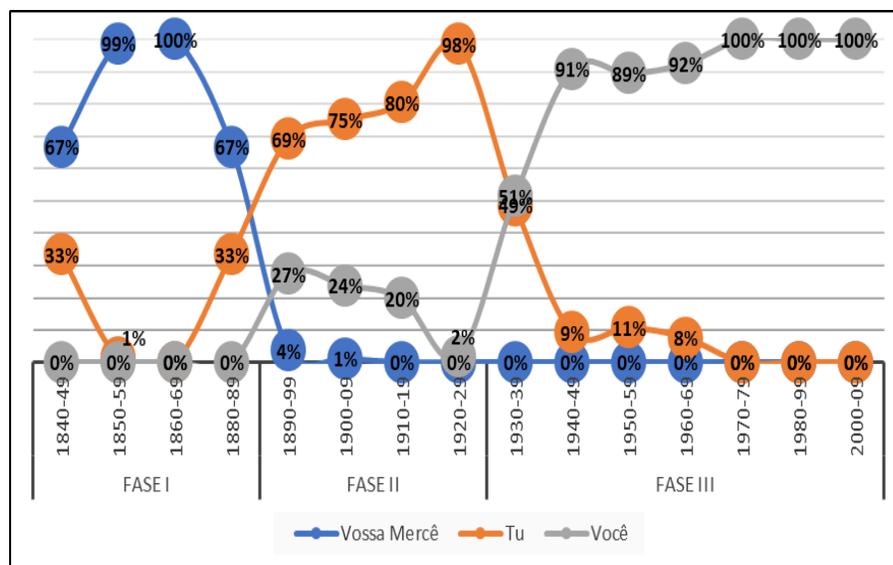
Apesar do esforço do linguista-pesquisador para equilibrar a amostra em relação ao tempo (séculos), ao sexo e à faixa etária dos informantes, trata-se, pois, de uma tarefa que é naturalmente obstaculizada pelo fato de os trabalhos em Sociolinguística Histórica restringirem-se sempre às missivas que se deixam encontrar no interior dos acervos públicos e privados. Além disso, acrescenta-se o fato de as cartas em análise estarem sob o crivo da reconstituição dos perfis sociais dos seus redatores, o que necessariamente dialoga com a questão de se tratar de autores nascidos e residentes (ainda que por algum período de suas vidas) no espaço mineiro. O GRÁFICO 4 evidencia a distribuição dos dados (em termos de índices percentuais) de *vossa mercê*, *você* e *tu* no decorrer do tempo.

De um modo geral, a análise do GRÁFICO 4 expõe um panorama de alternância entre as formas *vossa mercê*, *você* e *tu*, o que permitiu a segmentação das décadas dos séculos XIX (2ª metade principalmente) e XX em análise em três fases. É importante esclarecer que as evidências das formas *vossa mercê*, *você* e *tu* com percentuais de avanço e de recuo de tais formas estão necessariamente vinculadas ao número de cartas históricas em análise que necessariamente resplandecem os textos a que se teve acesso no interior dos acervos.

³⁹ Além das 307 cartas acima computadas, há mais 17 missivas que fazem parte da amostra (de um total de 324 cartas) e que não foram incluídas na tabela em questão por estarem sem data completa.

⁴⁰ Além dos 1194 dados acima computados, há mais 42 dados que fazem parte da amostra (de um total de 1236 oco) e que não foram incluídas na tabela em foco por estarem sem referência exata ao ano em que foram redigidas.

GRÁFICO 4
As formas *vossa mercê*, *you* e *tu* distribuídas no eixo do tempo
(principalmente os séculos XIX e XX)



Na fase I (1840-1889), estão, na cena da variação, as formas *vossa mercê* e *tu*. O *vossa mercê* manteve-se com as frequências de uso de 67%, 99%, 100% e 67% entre os anos de 1840-49, 1850-59, 1860-69 e 1880-89, respectivamente. A partir da década de 90 do século XIX (1890), o *vossa mercê* decresce para 4%, entre 1890-99, até alcançar, na 1ª década do século XX (1900-09), 1% de produtividade. Nessa fase I, observa-se a prevalência do *vossa mercê*, ao passo que o *tu* que se mostrou com índices percentuais de 33%, 1%, 0% e 33% entre os anos de 1840-49, 1850-59, 1860-69 e 1880-89, respectivamente.

Na fase II (1890-1929), tem-se em evidência dados do inovador *you*, entre os anos de 1890-99, 1900-09, 1910 e 1920-29, ainda que com baixas frequências de uso de 27%, 24%, 20%, 2%, respectivamente. À medida que o *you*, entre 1890 e 1929, parece tomar fôlego de uso nas missivas mineiras analisadas, o *vossa mercê*, por outro lado, começa a perder o seu campo de atuação (4%, 1% 0%, 0%) e o *tu*, em sentido oposto, parece funcionar como uma produtiva forma de referência ao sujeito de 2SG, que com os percentuais 69%, 75%, 80% e 98%, respectivamente, mostra-se em variação com o *you*. A partir dessa fase II, as formas em variação são o *tu* e o *you* em função também do desuso do *vossa mercê*.

Na cena da variação estão, entre os anos de 1930-2009 (fase III), as formas *you* e *tu*. À proporção que o *tu* decresce com os índices de 49% (1930-39), 9% (1940-49), 11% (1950-59), 8% (1960-69) até a sua ausência nos anos 2000, o *you* passa a assumir, gradual e

paulatinamente, os altos índices de produtividade de 51% (1930-39), 91% (1940-49), 89% (1950-59) e 92% (1960-69) até manter-se em uso categórico entre as décadas de 70 do século XX e os anos 2000. Muito sintomático é o fato de os anos 30 do século XX terem se mostrado, nas cartas mineiras analisadas, como o marco para a fixação do *você*, já que também os anos 30 são interpretados como marco temporal para a mudança de parâmetro do sujeito nulo no PB (DUARTE, 1993). Trata-se de uma mudança na marcação do parâmetro do sujeito motivada justamente pela reorganização do sistema pronominal com a inserção das formas *você* e *a gente* formas de P2 e P4 em relação de concordância com a 3SG.

Passa-se, na TABELA 11, à análise da distribuição dos dados de *vossa mercê*, *você* e *tu* pelas cartas amorosas, de amizade, familiares e de comércio.

TABELA 11
O gênero textual “carta”: os pronomes-sujeito de 2SG
em relação às cartas pessoais e de comércio em cena

AS CARTAS PESSOAIS (SUBTIPOS) E DE COMÉRCIO	FORMAS DE REFERÊNCIA AO SUJEITO DE 2SG			TOTAL
	VOSSA MERCÊ	VOCÊ	TU	
Cartas amorosas	-	71/569 (12,47%)	302/588 (51,36%)	373/1236 (30%)
Cartas familiares	10/79 (12,65%)	208/569 (36,56%)	229/588 (38,94%)	447/1236 (36%)
Cartas de amizade	-	290/569 (50,97%)	55/588 (9,35%)	345/1236 (28%)
Cartas de comércio	69/79 (87,35%)	-	2/588 (0,35%)	71/1236 (6%)
TOTAL	79/1236 (6,39%)	569/1236 (46,04%)	588/1236 (47,57%)	1236

Os dados de *vossa mercê* mostram-se profícuos, em 87,35% dos dados (69/79 oco), nas cartas de comércio ao Barão de Cocais (91), mas há dez evidências de *vossa mercê* também em missivas familiares, cf. ilustrado de (92) a (101).

Dado de *vossa mercê* em carta de comércio:

(91) [...] pois que *Vossa mercê* foi aceitar (...). Terá *Vossa mercê* tomado outro a cordo sobre avinda do Camillo? [...]. (BNP. Paciencia, 17.08.1858.)

Dados de *vossa mercê* em cartas familiares:

(92) [...] jamais tive a satisfação de vos ver, sem duvida é por que assim *Vossa merce* tem querido [...]. (APM. JPS. Nossa Senhora do Porto, 10.01.1868.)

(93) [...] assim *Vossa merce* tem querido, pois não vos era difficil chegar *Vossa merce* té aqui [...]. (JPS. Nossa Senhora do Porto, 10.01.1868.)

(94) [...] assim *Vossa merce* tem querido, pois não vos era difficil chegar *Vossa merce* té aqui a fim de dar a mim e a todos [...]. (JPS. Nossa Senhora do Porto, 10.01.1868.)

(95) [...] Deos permitta que *Vossa merce* appareça por cá brevemente [...]. (JPS. Nossa Senhora do Porto, 10.01.1868.)

(96) [...] *Vossa merce* naturalmente deve admirar-se de tal resolução [...]. (JPS. Ouro Preto, 10.02.1882.)

(97) [...] e a certidão *Vossa merce* há de tirar em duplicata. [...]. (JPS. Ouro Preto, 10.02.1882.)

(98) [...] tenho a certeza de que *Vossa Mercê* fica inquieta [...]. (NCS. Ouro Preto, 07.02.1895.)

(99) [...] A *Dona* Anna, quando passou, entregou a *Vossa Mercê* todos os livros que mandei? Peço que entregue ao José Pedro [...]. (NCS. Ouro Preto, 07.02.1895.)

(100) [...] que estas tortas linha vá encontrar a *Vossa merce*; gozando perfeita saude [...]. (JPS. Ouro Preto, 21.12.1869.)

(101) [...] Mas, a quem devo isto, senão a *Vossa Mercê* que com tanta bondade e desinteressadamente vae me educando! [...]. (AS. BH, 30.08.1904.)

O *tu* prevalece nas cartas amorosas, cf. sentença (102), em 51,36% dos dados (302/588), ao passo que o *você* predomina nas cartas de amizade (103), em 50,97% (290/569), confirmando o já sabido caráter íntimo do *tu* e o de menor proximidade entre remetente e destinatário do *você* (103), cf. já atestado por Souza (2012) para as cartas cariocas. O fato de o *você* ter se mostrado produtivo nas missivas de amizade parece reverberar um traço da sua origem nominal que é o seu caráter respeitoso, ainda que tenha passado a assumir o comportamento de um legítimo pronome em alternância com o *tu*.

Dado de *tu* em missiva amorosa:

(102) [...] Como *tens* passado meu querido José? Não *tens* esquecido da tua Iracy? [...]. (MIDV. RJ, 21.03.1925.)

Dado de *você* em missiva de amizade:

(103) [...] De qualquer forma, esperamos que *você* se anime a vir até aqui [...]. (AGF. Brasília, 22.11.1961.)

SÍNTESE DO CAPÍTULO

Em termos percentuais, os resultados apontam para uma variável dependente eneária composta pelas formas *vossa mercê*, *você* e *tu* para a referência ao sujeito de 2SG. Considerando as variáveis linguísticas (a *expressão do sujeito*, a *pessoa verbal* e o *paralelismo formal e semântico*) e extralinguísticas (o *sexo*, a *faixa etária*, as *relações sociais*, o *tempo* e a *carta*), resumam-se alguns dos principais resultados, tendo em vista as hipóteses conjecturadas.

A análise da expressão do sujeito de 2SG em relação ao *vossa mercê* o evidencia preferencialmente pleno. O fato de o *você* também prevalecer *pleno* expõe a fixação de um traço peculiar à forma nominal de tratamento originária (*vossa mercê*) como um traço que ainda resiste no *você* (RUMEU, 2004; RUMEU, 2006; LOPES & RUMEU, 2007; LOPES & CAVALCANTE, 2011). Por outro lado, a prevalência do *tu*-nulo está de acordo com o esperado, ou seja, com a sua origem etimológica, uma vez que assume uma desinência número-pessoal (DNP) específica de 2SG acionada na forma verbal.

Em relação à pessoa verbal (concordância), o *vossa mercê* mostra-se categoricamente harmonizado com formas verbais de 3SG. Nesse sentido, é possível interpretar tal fato, em harmonia com Lopes & Rumeu (2007), Lopes & Cavalcante (2011), como mais uma evidência de um traço formal do *vossa mercê* que se fixou no pronome *você*, forma de referência semântica à 2SG (traço de pessoa), ainda que em harmonia formal com a 3SG (LOPES & RUMEU, 2007). Por outro lado, o *tu* mostra-se, nas cartas mineiras analisadas, preferencialmente harmonizado com formas verbais morfológicamente estruturadas a partir de desinências específicas de 2SG.

Em suma, os traços de *expressão do sujeito* e *pessoa* (concordância) evidenciam o *vossa mercê* e o *você* preferencialmente plenos e harmonizados com formas verbais de 3SG, ao passo que o *tu* apresenta-se preferencialmente *nulo* e em concordância com formas verbais de 2SG. A partir dos contextos de *expressão do sujeito* e de *pessoa verbal* (concordância) é possível interpretar que as formas *você* e *tu* acompanham os comportamentos etimológicos da forma nominal de tratamento *vossa mercê* e do pronome latino *tu*, o que parece evidenciar uma produção escrita que, apesar de íntima (pessoal), mantém-se em harmonia com a norma-padrão.

No que se refere ao *paralelismo formal e semântico*, observa-se que, de um modo geral, o *vossa mercê* e o *você* estão preferencialmente combinados com formas do paradigma de

3SG, ao passo que o *tu* se ajusta às formas do paradigma de 2SG (*tu*), o que parece indicar que, também neste contexto, a escrita mineira mostra-se sintonizada com os princípios da norma-padrão.

Ao cruzar *sexo* e *faixa etária*, é possível chegar a algumas generalizações. Os dados de *você* distribuem-se por todas as faixas etárias (homens jovens, adultos e idosos), ainda que os homens jovens tenham sido os que pareceram impulsionar o *você* nas cartas mineiras. O *tu*, por outro lado, foi a preferência das mulheres jovens, o que parece apontá-las mais conservadoras. O *vossa mercê* prevalece na produção escrita dos homens adultos, uma vez que tais dados se limitam à produção escrita oitocentista, momento para o qual não foram levantadas missivas produzidas por mulheres.

No que diz respeito às relações sociais, os resultados apontam para o fato de o *você* ter-se sobressaído nas relações simétricas (entre iguais), assim como o *tu*, aproximando o *você* dos mesmos campos funcionais do *tu*, conforme também observado por Souza (2012) para as missivas cariocas. O *vossa mercê* mostrou-se restrito ao âmbito das relações assimétricas ascendentes (de inferior para superior) em conformidade com os resultados de Lopes & Rumeu (2015) também para as cartas mineiras.

A análise da distribuição dos dados no eixo do tempo evidencia um cenário de variação principalmente entre as formas *tu* e *você*, já que as evidências de *vossa mercê* ficaram restritas ao século XIX. Em síntese, é possível estabelecer três fases. São elas: na fase I (1840-1889), o *vossa mercê* prevalece *pari passu* o *tu* decresce; na fase II (1890-1929), observa-se alternância *tu/você* a partir do avanço do *você* e do *tu* e decréscimo do *vossa mercê*; na fase III (1930-2009), as formas *tu* e *você* estão no cenário da variação, cabendo ao *você* a perspectiva de ascensão gradual e paulatina à medida que o *tu* decresce.

Em relação às cartas pessoais (amorosas, amizade e familiares) e às cartas de comércio, o *tu* prevalece nas cartas amorosas, enquanto o *você* predomina nas cartas de amizade, constatações que se alinham às de Souza (2012) para as cartas cariocas oitocentistas e novecentistas. O *vossa mercê* prepondera nas cartas de comércio, ainda que tenha mostrado algum uso nas cartas familiares.

CAPÍTULO 4. A VARIAÇÃO TU/VOCÊ: DESCRIÇÃO E ANÁLISE DOS RESULTADOS ESTATISTICAMENTE RELEVANTES

Neste capítulo, passa-se à descrição-analítica dos resultados da alternância *tu/você*, assumindo o *você* como valor de aplicação nos níveis de *step-up* & *step-down* do RBrul na interface do RStudio (OUSHIRO, 2014, p. 136). É importante esclarecer que os resultados probabilísticos estão expostos, considerando o *p-value*, o *input* e os *logodds*.

O *p-value* diz respeito à hipótese de não haver correlação entre a variável dependente e as variáveis independentes. Caso o *p-value* seja menor do que 0,05 (5%), recusa-se a ideia da não-correlação entre a variável preditora (dependente) e as variáveis independentes (modelo nulo), o que significa que uma dada variável condiciona significativamente a aplicação da regra variável em análise. Em sendo o *p-valor* igual ou maior do que 0,05, assume-se a conjectura da não-correlação entre a variável preditora e as independentes em análise.

O *input* equivale ao nível geral de produtividade do valor de aplicação para uma dada regra variável. Um valor alto (>0,40) para o *input* significa que a regra variável é produtiva, considerando sempre o valor de aplicação em questão, ao passo que um valor baixo (<0,40) para o *input* evidencia um menor índice de produtividade para a variável dependente. Nesta análise, o valor de aplicação assumido para a regra variável *tu/você* é o *você*.

Os *logodds* expõem a possibilidade de aplicação de uma dada regra variável. A interpretação dos *logodds* deve também encontrar respaldo na interpretação dos índices dos pesos relativos. Em termos gerais, *logodds* especificados a partir de valores positivos apontam para a aplicação da regra variável em questão, ao passo que valores negativos sugerem o seu desfavorecimento.

Para que a rodada multivariada no RBrul fosse implementada com sucesso, algumas recodificações, tendo em vista o *você* como valor de aplicação, tiveram de ser feitas. São elas: 1º) os dados de *vossa mercê* foram excluídos neste momento da análise; 2º) os dados das formas pronominais das cartas sem referência ao seu ano de produção foram também excluídos (42 oco), considerando que a variável “tempo” é uma das variáveis independentes (sociais) em cena nesta rodada; 3º) os redatores não inseridos em faixa etária alguma, uma vez que não se conseguiu resgatar exatamente a data específica de nascimento, foram mantidos na rodada. Uma possível

retirada de tais dados de pronomes cujos autores são mineiros que não tiveram a sua data de nascimento resgatada impossibilitava, por sua vez, que se alcançasse os níveis de *step up and step-down* com o devido *match*. Feitos esses ajustes que conduziram ao total de 1115 ocorrências⁴¹ à luz de um *input* de 0,094 de aplicação geral da regra variável em análise (*tu versus você* com o *você* como valor de aplicação), chega-se à seleção dos seguintes grupos de fatores na própria ordem de seleção do RBrul⁴²: 1º) *pessoa* (~0); 2º) *paralelismo formal e semântico* (0.0175); e 3º) *faixa etária* (0.0376). Temos em discussão uma regra variável com o *você* como valor de aplicação para a qual o *input* é baixo (0,094), o que está em sintonia com o índice geral de uso do *você* (46,04%) que é menor ao alcançado pelo *tu* (47,57%). Nessa mesma direção, segue o *log-likelihood* (logaritmo de verossimilhança). Trata-se de um índice revelador do nível de aproximação entre o modelo e os dados. Na rodada em análise, o *log-likelihood* assumiu o valor de -4.682, caracterizando a regra variável em análise com inexpressividade para o *você*, o que também está de acordo com o baixo *input* (0.094) para a aplicação do *você* nas cartas mineiras analisadas. Uma vez apresentados os índices gerais para a aplicabilidade do *você*, passa-se à discussão dos resultados específicos das variáveis selecionadas (*pessoa*, *paralelismo*, *faixa etária*).

A notação científica atribuída ao parâmetro *pessoa verbal* (concordância) é de ~0 como *p-value*. Isso quer dizer que a *pessoa verbal* se mostra como um contexto condicionante à aplicação da regra variável em análise. Em outros termos, esse *p-value* evidencia que há praticamente 100% de chance de o contexto da *pessoa verbal* impulsionar o valor de aplicação (*você*) da regra variável em questão (*tu versus você*), o que justifica a sua seleção como o 1º fator condicionante para o *você-sujeito*.

⁴¹ As recodificações feitas levam em conta o número total de **1236** ocorrências com a exclusão das 79 ocorrências de *vossa mercê* e das 42 ocorrências de cartas sem referência ao ano da sua produção, o que reduz os dados ao número total de **1115** ocorrências para a análise probabilística sobre a variação *tu/você*.

⁴² Reprodução das informações dadas pelo RBrul (RStudio): “BEST STEP-UP MODEL OF RESPONSE pronomes IS WITH PREDICTOR(S): *pessoa* (~0) + *paralelismo* (0.0175) + *fxetaria* (0.0376) [p-values building from null model]”.

TABELA 12
O efeito da pessoa verbal (concordância) sobre o *você-sujeito*

PESSOA (CONCORDÂNCIA)	VOCÊ (oco)	%	PR	logodds
3SG	541/543	99,63%	0.999	44.477
2SG	1/572	0,002%	0.001	- 44.477

Valor de aplicação: *você*.
p-value: ~ 0
input: 0.094

O contexto de *pessoa* (concordância) assume o *p-value* de ~ 0, o que permite interpretar esse contexto como altamente significativo para a aplicação do *você*. A conjugação dos valores em termos percentuais, probabilísticos (PR) e de *logodds* evidencia a 3ª pessoa verbal como um contexto propulsor do *você* nas cartas mineiras analisadas (séculos XIX e XX), cf. TABELA 12. Essa constatação está fundamentada principalmente no alto PR de 0.999 (= 99,63% de 541oco) conjugado ao alto índice de *logodds* de 44.477 para a construção do modelo nulo que, por sua vez, se mostra em caminho diametralmente oposto ao da 2ª pessoa verbal (- 44.477), permitindo interpretar o verbo na 2SG como um contexto inibidor do *você-sujeito*. Confirmou-se a previsão de que também a produção escrita mineira estaria mais propensa à harmonização morfossintática entre o *você-sujeito* e o verbo na 3ª pessoa verbal, como também constatado por Souza (2012, p. 99) em relação à produção escrita dos cariocas (séculos XIX e XX).

TABELA 13
O efeito do paralelismo formal e sintático sobre o *você-sujeito*

PARALELISMO FORMAL E SINTÁTICO	VOCÊ (oco)	%	PR	logodds
Precedido por <i>Você-Sujeito Pleno/Nulo</i>	216/231	93,5%	0.999	21.042
Precedido por formas de 2PL (<i>vós-suj.</i> , clíticos (<i>vos</i>) e possessivos de 2PL (<i>vosso</i>))	2/7	28,6%	0.999	16.162
Precedido por formas de 3SG (clíticos e possessivos)	66/71	93%	0.51	0.039
Precedido por formas de <i>Você não-Sujeito</i>	30/34	88,2%	0.505	0.021
Precedido por formas de <i>Tu não-Sujeito</i>	0/17	0%	0.999	16.664

Valor de aplicação: *você*.
p-value: 0.0175
input: 0.094

O contexto do paralelismo formal e semântico mostra-se com *p-value* de 0.0175, valor muito próximo do zero, o que ratifica a seleção desse grupo de fatores como o segundo contexto condicionante à aplicação do *você-sujeito*. Em termos probabilísticos (PR), estatísticos e de *logodds*, evidencia-se que o *você* é condicionado pelo contexto de *você-sujeito* (pleno ou nulo). Isso quer dizer que o *você-sujeito* tende a ser produtivo, se precedido pelo *você-sujeito*, tendo em vista os altos índices de probabilidade (0.999), valor muito maior que 0.5 e muito próximo de 1.0), percentual (93,5% = 216/231) e de *logodds* (21.042) que acentuam a força do *você-sujeito* como seu próprio contexto propulsor. O *você-sujeito*, se precedido por formas de 3SG (clíticos e possessivos) e por formas de *você não-sujeito*, mostra-se com índices probabilísticos (0.51 e 0.505), *logodds* (0.039 e 0.021) e percentuais (93% e 88,2%), respectivamente, o que evidencia que o *você* é fomentado também pelas formas de 3SG (clíticos e possessivos) e pelas formas de *você não-sujeito*. Por outro lado, ainda que os índices probabilísticos e de *logodds* sejam altos para as formas pronominais de 2PL (0.999 e 16.162) e de formas de *tu não-sujeito* (0.999 e 16.664) como contextos precedentes ao *você-sujeito*, há de se atentar ao fato de que se trata de valores brutos extremamente ínfimos (2oco, 0oco, respectivamente), o que evidencia tais contextos do paradigma da 2ª pessoa (formas de 2PL e formas de *tu não-sujeito*) como contextos inibidores do *você* e, conseqüentemente, propulsores do *tu* nas cartas analisadas (séculos XIX e XX).

TABELA 14
O efeito da faixa etária sobre o *você-sujeito*

FAIXA ETÁRIA	Você (oco)	%	PR	<i>logodds</i>
Adultos (31 a 50 anos)	215/532	40,4%	0.98	3.872
Idosos (acima de 50 anos)	180/273	65,9%	0.977	3.767
Jovens (14 a 30 anos)	116/221	52,5%	0.001	-17.817
Crianças (menos de 14 anos)	1/1	100%	0.001	-14.836

Valor de aplicação: *você*.
p-value: 0.0376
input: 0.094

A *faixa etária* mostra-se como o terceiro contexto motivador do *você-sujeito* nas cartas mineiras analisadas (séculos XIX e XX), amparado no *p-value* de 0.0376 como expressão do modelo nulo. Apesar de os índices probabilísticos, expressos através de pesos relativos (0.98 e 0.977) e de *logodds* (3.872 e 3.767), parecerem apontar adultos e os idosos como os redatores que, na amostra analisada, preferem o *você*, há de se ter cautela para a análise dos dados especificamente no que se refere à faixa etária dos escreventes. Faz-se necessário ponderar principalmente sobre o fato de que este trabalho está embasado em amostras de língua escrita de sincronias passadas que não estão equanimemente distribuídas em relação à produção escrita dos jovens, adultos e idosos também no decorrer dos séculos XIX e XX, uma vez que dificilmente se tem acesso, no interior dos acervos públicos e privados, a um número equilibrado de cartas pessoais brasileiras (manuscritas), produzidas entre os anos de 1840 e 2009 (cf. TABELA 10).

SÍNTESE DO CAPÍTULO:

Na TABELA 15, estão sumarizados os resultados probabilísticos para as variáveis independentes *pessoa*, *paralelismo formal e semântico* e *faixa etária* que se mostram estatisticamente relevantes à aplicação do *você*.

TABELA 15

Síntese das variáveis independentes estatisticamente relevantes para a aplicação do *você-sujeito*

VARIÁVEIS LINGUÍSTICAS	VARIÁVEL EXTRALINGUÍSTICA
1º) PESSOA (CONCORDÂNCIA) <ul style="list-style-type: none"> • 3ª pessoa do singular (3SG): 0.999 (99,63%, 541/543oco) <p><i>p-value</i>: ~ 0</p>	3º) FAIXA ETÁRIA <p>Adultos (31 a 50 anos):</p> <ul style="list-style-type: none"> • 0.98 (40,4%, 215/532oco) <p>Idosos (acima de 50 anos):</p> <ul style="list-style-type: none"> • 0.977 (65,9%, 180/273oco) <p><i>p-value</i>: 0.0376</p>
2º) PARALELISMO FORMAL E SEMÂNTICO <ul style="list-style-type: none"> • Precedido por <i>Você-Sujeito Pleno/Nulo</i>: 0.999 (93,5%, 216/231oco) • Precedido por formas de 3SG (clíticos e possessivos): 0.51 (93%, 66/71oco) • Precedido por formas de <i>Você não-Sujeito</i>: 0.505 (88,2%, 30/34oco) <p><i>p-value</i>: 0.0175</p>	
Valor de aplicação: <i>você</i> . <i>input</i> : 0.094	

A *concordância* com forma verbal de 3ª pessoa mostrou-se como um contexto propulsor do *você*, confirmando a hipótese inicial respaldada nos resultados de Souza (2012, p. 99) para as cartas oitocentistas e novecentistas. Em relação ao *paralelismo formal e semântico*, o *p-value* é de 0.0175, o que evidencia tal contexto também significativo para a aplicação do *você*.

Em relação às especificidades do *paralelismo formal e semântico*, constata-se o *você-sujeito* probabilisticamente favorecido pelos contextos de *você-sujeito* (0.999), de formas clíticas e possessivas de 3SG (0.51) e de formas de *você não-sujeito* (0.505), corroborando a hipótese de que estruturas formal e semanticamente relacionadas ao paradigma de *você* tendam a promover o *você-sujeito* como conjecturado por Rumeu (2013) para as cartas cariocas à luz de Omena (2003), Lopes & Vianna (2012).

O único contexto extralinguístico selecionado foi a *faixa etária* com um *p-value* de 0.0376 (valor também bem próximo de zero) para a expressão do modelo nulo, o que o evidencia como um contexto relevante à produtividade do *você* nas cartas analisadas.

CONSIDERAÇÕES FINAIS

Este trabalho investiga o uso das formas *vossa mercê*, *você* e *tu* em referência ao sujeito de 2SG em missivas mineiras dos séculos XIX e XX. Inicialmente, os dados foram analisados em termos percentuais (capítulo 3), a fim de verificar a influência das variáveis linguísticas e extralinguísticas em relação à variável dependente enéaria (*vossa mercê*, *você* e *tu*). Na sequência, o foco está na regra variável *tu versus você*, assumindo o *você* como variável dependente para a análise probabilística (capítulo 4).

O fato de se ter em cena um trabalho embasado em *corpora* históricos conduz o linguista-pesquisador a amostras pouco equânimes no que se refere à distribuição dos dados pelas categorias extralinguísticas *sexo*, *faixa etária do escrevente*, *tempo* (período) e ao *gênero textual* “carta” quer pessoal (amorosa, amizade e familiar), quer comercial. Uma vez expostas essas considerações acerca dos limites impostos por análises históricas, que estão em relação direta com a busca por fazer o melhor uso dos dados históricos (LABOV, 1994), recuperam-se as questões norteadoras desta dissertação, a fim de respondê-las.

I. Em que nível de produtividade estariam as formas *tu* e *você* na escrita mineira de sincronias passadas (séculos XIX e XX)?

A baixíssima produtividade do *vossa mercê*, em tão somente **6,39%** dos dados (79/1236) restritos ao século XIX, é compreensível, tendo em vista a celeridade do processo de gramaticalização em questão no PB nos períodos analisados (séculos XIX e XX), cf. Rumeu (2013), Lopes & Rumeu (2007), Faraco (2017 [1996], p. 121), Lopes *et al.* (2018), dentre outros. De um modo geral, em cena estão o *tu* e o *você* que concorrem acirradamente, assumindo índices percentuais de **47,57%** (588/1236) e de **46,04%** (569/1236), respectivamente. Na produção escrita mineira, a proeminência do *tu* mostra-se dissonante em relação à atual realidade sincrônica da fala mineira cuja realidade é a de prevalência do *você* (COELHO, 1999; PERES, 2006; HERÊNIO, 2006; GONÇALVES, 2008; MOTA, 2008; SCHERRE *ET AL.*, 2015; SILVA, 2017; REIS, 2019). O conservadorismo da língua escrita parece ter motivado a dinamização do *tu*,

infirmar a hipótese inicial de prevalência do *você* nas cartas mineiras oitocentistas e novecentistas analisadas.

II. Quais os tipos de relações sociais que subsidiariam o *tu* e o *você* nas cartas mineiras dos séculos XIX e XX?

As formas pronominais *você* e *tu* predominam nas relações simétricas. Acrescente-se o fato de que a prevalência das formas *você* (**78,21%**) e *tu* (**67,86%**) nas relações simétricas permite interpretar o encaminhamento da sociedade mineira pelos domínios da Solidariedade (BROWN & GILMAN, 1960). Por outro lado, o *vossa mercê*, que se manteve restrito aos dados do século XIX, conserva a *Semântica do Poder*, uma vez que se deixou evidenciar tão somente nas relações de assimetria ascendente (de inferior para superior), cf. discutido por Lopes & Rumeu (2015) também em relação às missivas mineiras.

III. Quais os contextos linguísticos e sociais que suscitariam o *você* na produção escrita mineira oitocentista e novecentista?

Nas cartas mineiras, o *você* está impulsionado pela *pessoa* (concordância), pelo *paralelismo formal e semântico* e pela *faixa etária* do informante, o que sugere algumas sistematizações. São elas:

- (a) O contexto de 3ª pessoa verbal é extremamente significativo à aplicação do *você* (**0.999**), conformando-se aos resultados de Souza (2012, p. 99) para as cartas oitocentistas e novecentistas;
- (b) Os contextos de *você-sujeito* (**0.999**), de formas clíticas e possessivas de 3SG (**0.51**) e de formas de *você não-sujeito* (**0.505**) apresentam-se como propulsores do *você*. Isso quer dizer que estruturas formalmente idênticas devem ser levadas em consideração na apreciação de uma dada regra variável, como também observado por Rumeu (2013) em relação às formas do paradigma de *tu* como propulsores do *tu-sujeito* nas cartas cariocas;

(c) Ainda que os índices probabilísticos apontem para a produtividade do *você* na produção escrita de adultos e idosos (faixa etária), é importante ter-se em mente o fato de que tais resultados não encontram respaldo em uma amostra totalmente equilibrada em relação aos informantes jovens, adultos e idosos, distribuídos por todas décadas/fases da 2ª metade do século XIX e do século XIX (cf. TABELA 10).

REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, C. R. A. **Preciosas memórias, belos fragmentos: Abílio Barreto e Raul Tassini – a ordenação do passado na formação do acervo do Museu Histórico de Belo Horizonte (1935-1956)**. Dissertação (Mestrado em História) - Universidade Federal de Minas Gerais, Belo Horizonte, 2008.
- AVELHEDA BANDEIRA, A. C. C.; SOUZA, S. C. G. **Passo a passo para uso do R-Brul**. 2017 (Curso de curta duração ministrado/Extensão).
- BAKHTIN, M. Os gêneros do discurso. In: **Estética da Criação Verbal**. São Paulo: Martins Fontes, 1997, p. 261-306.
- BARATA, C. E. A. BUENO, A. H. C. **Dicionário das Famílias Brasileiras**. Volumes I e II. São Paulo: Ibero-América, 2000.
- BARBOSA, A. Tratamento dos corpora de sincronias passadas da língua portuguesa no Brasil: recortes grafológicos e linguísticos. In: LOPES, C. R. dos S. **A Norma Brasileira em construção: fatos linguísticos do século 19**. Rio de Janeiro: Faculdade de Letras, UFRJ, FAPERJ, 2005, p. 25-43.
- BERGS, A. The Uniformitarian Principle and the Risk of Anachronisms in Language and Social History. In: HERNÁNDEZ-CAMPOY; CONDE SILVESTRE. **The Handbook of Historical Sociolinguistics**. Oxford, Wiley-Blackwell, 2012, p. 80-98.
- BIBER, D. **Variation across speech and writing**. Cambridge University Press, New York, 1988.
- BORGES NUNES, E. **Abreviaturas paleográficas portuguesas**. Lisboa, 1981.
- BROWN, R.; GILMAN, A. The pronouns of power and solidarity. In: Sebeok, T.A. (ed.) 1960. **Style in language**. Cambridge- Mass: MIT Press, 2003, p.253-276.
- CARDOSO, N. D. **As estratégias de dativo de 2ª pessoa em cartas pessoais (séculos XIX e XX)**. Dissertação (Mestrado em Letras) – Universidade Federal de Minas Gerais, 2017.
- CARVALHO, L. F. de. **O estatuto variável do imperativo de 2ª pessoa do singular em missivas mineiras: um estudo sociolinguístico de cunho histórico (séculos XIX e XX)**. Dissertação (Mestrado em Letras) – Universidade Federal de Minas Gerais, Belo Horizonte, 2020.
- COELHO, M. S. V. **Uma abordagem variacionista do uso da forma você no Norte de Minas**. 1999. 95 f. Dissertação (Mestrado em Letras: Linguística) - Universidade Federal de Minas Gerais, Belo Horizonte, 1999.
- CONDE SILVESTRE, J. C. **Sociolinguística Histórica**. Madrid: Gredos, 2007.

CRUZ, I. A. **A alternância Tu/Você em contextos sintáticos de complementação e de adjunção: estudo de cartas pessoais dos séculos XIX e XX.** Dissertação (Mestrado em Letras) – Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

CUNHA, C.; CINTRA, L. F. L. **Nova Gramática do Português Contemporâneo.** Rio de Janeiro: Nova Fronteira, [1985] 2007.

DANIEL, E. J. **R-BRUL** version 3.1.4, 2020. Disponível em <<http://www.danielezrajohnson.com/Rbrul.R>>. Acesso em 7 jan. 2021.

DUARTE, M. E. L. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In: I. Roberts; M. Kato. (orgs.) **Português brasileiro: uma viagem diacrônica.** São Paulo, Campinas, Editora da Unicamp: 1993, p. 107-128.

DUARTE, M. E. L. **A perda do princípio ‘Evite pronome’ no português brasileiro.** Tese de Doutorado em Linguística. Universidade Estadual de Campinas, Campinas. Inédita, 1995.

FARACO, C. A. “O tratamento *você* em português: uma abordagem histórica.” In: **Fragmenta 13**, Publicação do Curso de Pós-Graduação em Letras da UFPR. Curitiba, Editora da UFPR, 2017 [1996].

FLEXOR, M. H. O. **Abreviaturas: manuscritos dos séculos XVI ao XIX. 2ª ed. aum.** São Paulo: Arquivo do Estado de São Paulo, Editora UNESP, Secretaria de Estado da Cultura, 1991.

GOMES, V. S. **Para além dos pacotes estatísticos Varbrul, Goldvarb e Rbrul: qual a concepção de gramática.** *Revista Gelne*, v. 14 Especial, 2012, p. 259-272.

GOMES, V. S., LOPES, C. R. S. **Formas tratamentais em cartas escritas em Pernambuco (1869-1969): tradição discursiva e sociopragmática.** *Revista de Estudos da Linguagem*, v. 24, 2016, p. 137-165.

GONÇALVES, C. R. **Uma abordagem sociolinguística do uso das formas você, ocê e cê no português.** Tese (Doutorado em Letras) – Universidade de São Paulo, São Paulo, 2008.

GUY, G; ZILLES, A. **Sociolinguística Quantitativa – instrumental de análise.** São Paulo: Parábola Editorial, 2007.

HERÊNIO, K. K. P. **Tu e você em uma perspectiva intra-linguística.** Dissertação (Mestrado em Letras) – Universidade Federal de Uberlândia, Uberlândia, 2006.

HERNÁNDEZ-CAMPOY, J. M.; SCHILLING, N. The Application of the Quantitative Paradigm to Historical Sociolinguistics: Problems with the Generalizability Principle. In: HERNÁNDEZ-CAMPOY, J. M.; CONDE-SILVESTRE, J. C. **The Handbook of Historical Sociolinguistics.** Wiley-Blackwell, 2012, p. 63-79.

IBGE – **INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA**, 2010. Resultado dos Dados Preliminares do Censo – 2010. www.ibge.gov.br/cidade@

LABOV, W. **Padrões sociolinguísticos**. Tradução de Marcos Bagno, Maria Marta Pereira Scherre, Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008 (1972).

LABOV, W. **Principles of Linguistic Change: Internal Factors**. Cambridge: Blackwell Publishers, 1994, v. I.

LINDLEY CINTRA, L. F. **Sobre formas de tratamento na língua portuguesa**. Lisboa: Horizonte, 1972.

LOPES, C. R. S.; RUMEU, M. C. B. O quadro de pronomes pessoais do português: as mudanças na especificação dos traços intrínsecos. In: CASTILHO, A. T.; TORRES MORAIS, M. A.; LOPES, R. E. V. & CYRINO, S. M. L. (Org.). **Descrição, história e aquisição do português brasileiro**. Campinas: Pontes/FAPESP, 2007, p. 419-436.

LOPES, C. R. S.; CAVALCANTE, S. R. **A cronologia do voçamento no português brasileiro**: expansão de você-sujeito e retenção do clítico-te. In: *Linguística*, 2011, v.25, p. 30-65.

LOPES, C. R. S.; MARCOTULIO, L. L.; RUMEU, M. C. B.; LIMA, A. **Reflexões metodológicas para a análise sociocultural de redatores em corpora históricos**. Gragoatá (UFF), 2010, v. 29, p. 239-253.

LOPES, C. R. S. Pronomes pessoais. In: BRANDÃO, S F & VIEIRA, S. R. (Orgs). **Ensino de gramática: descrição e uso**. São Paulo: Contexto, 2007, p. 103-114.

LOPES, C. R. S.; MARCOTULIO, L. L.; RUMEU, M. C. B.; COELHO, I. L.; ANDRADE, A.; MARTINS, M. A.; LACERDA, M. O.; GOMES, V. S.; MONTE, V. M.; CARNEIRO, Z. O. N.; SOUZA, C. M. N.; BALSALOBRE, S.; SOUZA, J. P.; OLIVEIRA, T. L.; MOURA, K. K.; CRUZ, I. A.; CARDOSO, N. D. A reorganização do sistema pronominal de 2ª pessoa na história do português brasileiro: a posição de sujeito. In: CASTILHO, Ataliba T. de. (Org.). **História do português brasileiro**: mudança sintática das classes de palavra – perspectiva funcionalista. São Paulo, Contexto, 2018, v. 4, p. 24-140.

LOPES, C. R. S.; MACHADO, A. C. M. Tradição e inovação: indícios do sincretismo entre a segunda e a terceira pessoas nas cartas dos avós. In: LOPES, C. R. S. (Org.). **A norma brasileira em construção**: fatos lingüísticos do século 19. Rio de Janeiro: UFRJ; FAPERJ, 2005. p. 45-66.

LOPES, C. R. S.; RUMEU, M. C. B. **A difusão do você pelas estruturas sociais carioca e mineira dos séculos XIX e XX**. Labor Histórico, v. 1, 2015, p. 12-25.

LOPES, C. R. S.; RUMEU, M. C. B. **A identificação dos perfis socioculturais dos redatores de corpora históricos**: encaminhamentos metodológicos. Diadorim, Rio de Janeiro, v. 20, 2018, p. 147-168.

LOPES, C. R. S.; VIANNA, J. B. S. A competição entre *nós* e *a gente* nas funções de complemento e adjunto: desvendando outras portas de entrada para o pronome inovador. **Caligrama: Revista de Estudos Românicos**, v. 17, n. 2, 2012, p. 137-161.

LUZ, R. D. **O tratamento na produção epistolar de João Pinheiro da Silva: análise sociopragmática de tu x você e respectivas formas gramaticais.** Dissertação (Mestrado em Letras: Estudos da Linguagem) - Universidade Federal de Ouro Preto, Mariana, 2015.

MACHADO, A. C. M. **A implementação de “Você” no quadro pronominal: as estratégias de referência ao interlocutor em peças teatrais no século XX.** (Dissertação de Mestrado em Língua Portuguesa) – Curso de Pós-Graduação em Letras Vernáculas. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2006.

MACHADO, A. C. M. **As formas de tratamento nos teatros brasileiro e português dos séculos XIX e XX.** (Tese de Doutorado em Língua Portuguesa) – Curso de Pós-Graduação em Letras Vernáculas. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

Mapa de Minas Gerais elaborado a partir de imagem da *internet* adaptada aos dados da pesquisa. Imagem do mapa disponível em <[https://pt.wikipedia.org/wiki/Ficheiro:Mesorregi%C3%B5es de Minas Gerais.svg](https://pt.wikipedia.org/wiki/Ficheiro:Mesorregi%C3%B5es_de_Minhas_Gerais.svg)>. Acesso em 28 jun. 2020.

MELO, H. Methodological issues for spontaneous speech corpora compilation: The case of C-ORAL-BRASIL. In: **Spoken and Linguistic Studies.** John Benjamins Publishing Company, 2014, p. 27-68.

MENON, O. **O Sistema Pronominal do Português do Brasil.** Revista Letras, Curitiba: Editora da UFPR, n. 44, 1995, p. 91-106.

MOTA, M. A. **A variação dos pronomes ‘tu’ e ‘você’ no português oral de São João da Ponte (MG).** Dissertação (Mestrado em Linguística) – Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

OLIVEIRA, T. L. **Entre o linguístico e o social: complementos dativos de 2ª pessoa em cartas cariocas (1880-1980).** Dissertação (Mestrado em Letras) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

OMENA, N. P. A referência à primeira pessoa do plural: variação ou mudança? In: PAIVA, M. da C.; DUARTE, M. E. L. (org.) **Mudança linguística em tempo real.** Rio de Janeiro: Faperj/Contracapa, 2003.

OUSHIRO, L. Tratamento de dados com o R para análises sociolinguísticas, p.133-176. In Raquel Meister Ko. Freitag (Org.). **Metodologia de Coleta e Manipulação de Dados em Sociolinguística,** São Paulo: Editora Edgard Blücher, 2014. Disponível em: <<http://dx.doi.org/10.5151/BlucherOA-MCMDS-10cap>>. Acesso em 8 jan. 2021.

OUSHIRO, L. **Introdução à estatística para linguistas.** v1.01, dez. UNICAMP, 2017.

PERES, E. P. **O uso do você, ocê, cê em Belo Horizonte: um estudo em tempo aparente e em tempo real.** 2006. 247 f. Tese (Doutorado em Letras: Linguística) Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

RASO, T.; MELLO, H.; MITTMANN, M. **O projeto C-ORAL-BRASIL**. CHIMERA. *Romance Corpora and Linguistic Studies* 1, 2014, p. 31-67.

R CRAN. **R: The Comprehensive R Archive Network**. Disponível em <<http://cran.r-project.org/>>. Acesso em 7 jan. 2021.

REIS, Z. M. **A variação ‘Tu’ e ‘Você’ no português falado e escrito em Lontra (MG)**. Dissertação (Mestrado Profissional em Letras) – Universidade Estadual de Montes Claros, Montes Claros, 2019.

ROMAINE, S. **Socio-historical linguistics: its status and methodology**. Cambridge University Press. New York, 2010 [1982].

RStudio. Take control of your R code. Disponível em: <<http://www.rstudio.com/ide/download/>>. Acesso em 7 jan. 2021.

RUMEU, M. C. B. Variation in the paradigms of *tu* and *você* subject and complements in letters from Minas Gerais, Brazil, 1860-1989. In: HUMMEL, Martin; LOPES, Célia Regina dos Santos. (Org.). **Address in Portuguese and Spanish**. 1ed. GmbH, Berlin/Munich/Boston: De Gruyter, 2020, v. 1, p. 227-249. (<https://www.degruyter.com/view/title/573880>)

RUMEU, M. C. B.; CRUZ, I. A.; CARDOSO, N. D. Formas de tratamento em cartas de Minas Gerais. In: LOPES, C. R. S. *et al.* A reorganização do sistema pronominal de 2ª pessoa na história do português brasileiro: a posição de sujeito. In: CASTILHO, Ataliba T. (Org.). **História do português brasileiro: mudança sintática das classes de palavra – perspectiva funcionalista**. São Paulo, Contexto, 2018, v. 4, p. 42-53.

RUMEU, M. C. B. **Formas variantes do imperativo de segunda pessoa nos séculos XIX e XX**: a expressão do social. *Signum: Estudos da Linguagem*, 2016. v. 19, p. 310-341.

RUMEU, M. C. B. OLIVEIRA, T. L. **A expressão da 2ª pessoa do singular em contextos de complementação e de adjunção**: retratos do encaixamento estrutural e social. *Revista Linguística (Online)*, 2016. v. 32, p. 25-46.

RUMEU, M. C. B.; LOPES, C. R. S. **A difusão do Você pelas estruturas sociais carioca e mineira dos séculos XIX e XX**. *Revista Labor Histórico*, Rio de Janeiro, 2015; 1 (1): 12-25, jan. / jun.

RUMEU, M. C. B. **Língua e sociedade**: a história do pronome 'Você' no português brasileiro. Rio de Janeiro: Ítaca (FAPERJ), 2013.

RUMEU, M. C. B. **As relações de poder e de solidariedade na sociedade carioca dos séculos XVIII e XIX**. São Paulo: Todas as Letras, 2011, v. 13, n. 2, p. 115-126. Disponível em: <<http://editorarevistas.mackenzie.br/index.php/tl/article/viewFile/4013/3204>>. Acesso em 12 jul. 2018.

RUMEU, M. C. B. **Vestígios da pronominalização de ‘Vossa Mercê’ > ‘Você’ em missivas cariocas e mineiras: uma incursão pelo português brasileiro escrito nos séculos XIX e XX.** Veredas - Revista de Estudos Linguísticos, v.16, 2012, p. 36-55.

RUMEU, M. C. B. **A implementação do *Você* no português brasileiro oitocentista e novecentista: um estudo de painel.** Tese (Doutorado em Letras Vernáculas – Língua Portuguesa). Faculdade de Letras/UFRJ, Rio de Janeiro, 2008.

RUMEU, M. C. B. **Traços formais e semântico-discursivos no processo de gramaticalização de ‘Vossa Mercê’ > ‘Você’.** Revista do GEL (Araraquara), v.3, 2006, p. 67-82.

RUMEU, M. C. B. **Para uma história do português no Brasil: formas pronominais e nominais de tratamento em cartas setecentistas e oitocentistas.** Dissertação (Mestrado em Letras), UFRJ, Rio de Janeiro, 2004.

SANTOS LUZ, M. dos. **Fórmulas de tratamento no português arcaico.** Revista portuguesa de filologia. v. 7, 1956, p. 251-363.

SCHERRE, M. M. P.; LUCCA, N. N. G.; DIAS, E. P.; ANDRADE, C. Q.; MARTINS, G. F. Usos dos pronomes *você* e *tu* no português brasileiro. In: **SIMELP: SIMPOSIO MUNDIAL DE ESTUDOS DE LINGUA PORTUGUESA**, 2009, Évora. Apresentação de Trabalho. Évora: Universidade de Évora, 2009.

SCHERRE, M. M. P.; DIAS, E. P.; ANDRADE, C. Q.; MARTINS, G. F. Variação dos pronomes “tu” e “você”. In: MARTINS, M. A.; ABRAÇADO, J (Org). **Mapeamento sociolinguístico do português brasileiro.** São Paulo, Contexto, 2015, p.133-172.

SILVA, F. C. **Barões do ouro e aventureiros britânicos no Brasil.** São Paulo, Editora da Universidade de São Paulo; Belém, Editora da Universidade do Pará, 2012.

SILVA, S. C. **A variação dos pronomes *tu* e *você* na fala mineira de Ressaquinha (MG).** Dissertação (Mestrado em Letras) – Universidade Federal de Ouro Preto, *Ouro Preto*, 2017.

SILVA LEME, L. G da. Genealogia paulistana. São Paulo: DUPRAT & COMP, v. 4, 1904.

SOUZA, C. D. **Eu te amo, eu lhe adoro, eu quero você: a variação das formas de acusativo de 2ª pessoa em cartas pessoais (1880-1980).** Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

SOUZA, J. P. F. **Mapeando a entrada do *você* no quadro pronominal: análise de cartas familiares dos séculos XIX-XX.** Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.

SOUZA, R. C. **Espada e ouro no Império Brasileiro: uma revisão biográfica do Barão de Cocais, 1792-1869.** Monografia (Especialização em Gestão Pública) – Universidade Federal de Ouro Preto, Barão de Cocais, 2013.

VELOSO, R. As três ondas da Sociolinguística e um estudo em comunidades de práticas. In: **XVII CONGRESO INTERNACIONAL ASOCIACIÓN DE LINGÜÍSTICA Y FILOLOGÍA DE AMÉRICA LATINA (ALFAL)**, 2014, João Pessoa, Brasil. p. 1740-1749.

WARDHAUGH, R. **An introduction to sociolinguistics**. Oxford: Blackwell Publishers Ltd, 2006.

WEINREICH, U.; LABOV, W.; HERZOG, M. I. Empirical foundations for a theory of language change. In: LEHMANN, W.; MALKIEL, Y. **Directions for historical linguistics**. University of Texas Press. 1968.

APÊNDICE (Rodada nos níveis *stepping up* e *stepping down*)

STEPPING UP...

STEP 0 - Run 0 - model with no fixed predictors

```
$misc.1
  n df intercept overall proportion
1115 1   -0.056           0.486
```

```
$misc.2
log.likelihood      AIC      AICc Dxy R2
-772.428 1546.856 1546.86  0  0
```

STEP 1 - adding to Run 0 model with no fixed predictors

Trying with *expressao*...

```
$expressao43
factor logodds tokens v/v+t centered factor weight
  p    2.276    427 0.958           0.907
  n   -2.276    688 0.193           0.093
```

```
$misc.1
  n df intercept overall proportion centered input prob
1115 2    0.847           0.486           0.7
```

```
$misc.2
log.likelihood      AIC      AICc  Dxy  R2
-412.414 828.827 828.838 0.723 0.598
```

Run 1 (above) with *expressao* is better than Run 0 without *expressao*, $p = 1.32e-158$

Trying with *pessoa*...

```
$pessoa44
factor logodds tokens v/v+t centered factor weight
 3sg    5.974    543 0.996           0.997
 2sg   -5.974    572 0.002           0.003
```

```
$misc.1
  n df intercept overall proportion centered input prob
1115 2   -0.374           0.486           0.408
```

```
$misc.2
log.likelihood      AIC      AICc  Dxy  R2
```

⁴³ Expressão da forma de referência ao sujeito de 2SG (nulo ou pleno) – códigos utilizados no grupo de fatores: p: pleno; n: nulo.

⁴⁴ Pessoa verbal (concordância) – códigos utilizados no grupo de fatores: 3sg: 3ª pessoa do singular; 2sg: 2ª pessoa do singular.

-20.553 45.105 45.116 0.995 0.916

Run 2 (above) with *peessoa* is better than Run 0 without *peessoa*, $p = 0$

Trying with *paralelismo*...

\$paralelismo⁴⁵

factor	logodds	tokens	v/v+t	centered	factor	weight
2	4.210	231	0.935			0.985
4	4.123	71	0.930			0.984
3	3.558	34	0.882			0.972
y	3.094	40	0.825			0.957
1	1.951	273	0.601			0.876
9	0.627	7	0.286			0.652
z	0.157	35	0.200			0.539
t	0.002	17	0.176			0.501
5	-1.260	245	0.057			0.221
7	-1.438	145	0.048			0.192
6	-15.023	17	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	11	-1.543			0.486		0.176

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-399.449	820.899	821.138	0.82	0.736

Run 3 (above) with *paralelismo* is better than Run 0 without *paralelismo*, $p = 8.48e-154$

Trying with *carta*...

\$carta⁴⁶

factor	logodds	tokens	v/v+t	centered	factor	weight
z	5.012	326	0.837			0.993
f	3.259	424	0.472			0.963
a	1.923	363	0.190			0.872
c	-10.194	2	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	4	-3.372			0.486		0.033

⁴⁵ Paralelismo formal e semântico - códigos utilizados no grupo de fatores: 2 - forma antecedida por *você-suj pleno* ou *nulo*; 4 - forma precedida por formas de 3SG [clíticos e possessivos]: *se, o/a, lhe* (clíticos), *seu/sua* (possessivos); 3 - forma antecedida por *você* não-sujeito; y - forma antecedida pelo imperativo de 2SG no subjuntivo; 1 - 1^a ocorrência; 9 - forma precedida por formas de 2PL: [pronomes sujeito *vós*, clíticos e possessivos]: *vós, vos, vossos, vossas, convosco*; z - forma antecedida pelo imperativo de 2SG no indicativo; t - forma precedida por vocativo; 5 - forma antecedida por *tu-suj* (nulo ou pleno); 7 - forma precedida por formas de 2SG [clíticos e possessivos]: *te* (clíticos), *teu(s)/tua(s)*, 'teus' (possessivos); 6 - forma antecedida por *tu* não-sujeito (*em ti, de ti, para ti, contigo*).

⁴⁶ Carta - códigos utilizados no grupo de fatores: z - amizade; f - familiar; a - amorosa; c - comércio.

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-614.475	1236.949	1236.985	0.552	0.354

Run 4 (above) with carta is better than Run 0 without carta, $p = 3.59e-68$

Trying with genero...

\$genero⁴⁷

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.091	348	0.517			0.523
h	-0.091	767	0.472			0.477

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	2	-0.022			0.486		0.495

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-771.446	1546.892	1546.903	0.039	0.002

Run 5 (above) with genero is better than Run 0 without genero, $p = 0.161$

Trying with fxetaria...

\$fxetaria⁴⁸

factor	logodds	tokens	v/v+t	centered	factor	weight
c	10.110	1	1.000			> 0.999
v	-1.795	273	0.659			0.142
j	-2.356	221	0.525			0.087
k	-2.844	532	0.404			0.055
x	-3.115	88	0.341			0.042

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	5	2.456			0.486		0.921

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-743.413	1496.826	1496.88	0.239	0.097

Run 6 (above) with fxetaria is better than Run 0 without fxetaria, $p = 7.52e-12$

Trying with interpessoal...

⁴⁷ Sexo do autor da carta – códigos utilizados no grupo de fatores: m – mulher; h – homem.

⁴⁸ Faixa etária do autor da carta – códigos utilizados no grupo de fatores: c – criança (< de 14 anos); v – idoso (mais de 50 anos); j – jovem (de 14 a 30 anos); k – adulto (de 31 a 50 anos); x – dá conta dos casos nos quais não é possível resgatar a ficha biográfica do informante, mas se sabe que é brasileiro.

`$interpessoal49`

factor	logodds	tokens	v/v+t	centered	factor	weight
j	6.375	70	0.900			0.998
l	6.257	18	0.889			0.998
n	6.257	18	0.889			0.998
b	5.836	325	0.840			0.997
i	5.176	130	0.731			0.994
h	4.024	13	0.462			0.982
m	3.631	120	0.367			0.974
s	3.113	39	0.256			0.957
p	2.771	61	0.197			0.941
k	2.723	37	0.189			0.938
a	-15.388	281	0.000			< 0.001
e	-15.388	2	0.000			< 0.001
u	-15.388	1	0.000			< 0.001

`$misc.1`

n	df	intercept	overall	proportion	centered	input	prob
1115	13	-4.178			0.486		0.015

`$misc.2`

log.likelihood	AIC	AICc	Dxy	R2
-412.155	850.31	850.641	0.795	0.96

Run 7 (above) with interpessoal is better than Run 0 without interpessoal, $p = 1.76e-146$

Trying with social...

`$social50`

factor	logodds	tokens	v/v+t	centered	factor	weight
a	1.153	21	0.810			0.76
s	-0.187	822	0.527			0.453
d	-0.965	272	0.338			0.276

`$misc.1`

n	df	intercept	overall	proportion	centered	input	prob
1115	3	0.294			0.486		0.573

`$misc.2`

log.likelihood	AIC	AICc	Dxy	R2
-752.855	1511.711	1511.732	0.16	0.045

Run 8 (above) with social is better than Run 0 without social, $p = 3.16e-09$

Trying with periodo...

⁴⁹ Relação interpessoal - códigos utilizados no grupo de fatores: j – esposa-marido; l – sobrinho(a)-tio(a); n – primo (a)-primo(a); b – amigo (a)-amigo (a); i – irmão(ã)-irmão(ã); h – marido-esposa; m – mãe-filho(a); s – afilhado (a)-madrinha; p – pai-filho(a); k – tio(a)-sobrinho(a); a – noivo(a)-noivo(a); e – filho(a)-pai; u – sogro(a)-nora.

⁵⁰ Relação social - códigos utilizados no grupo de fatores: a – relação social assimétrica ascendente; s - relação social simétrica; d - relação social assimétrica descendente.

\$periodo⁵¹

factor	logodds	tokens	v/v+t	centered	factor	weight
f	17.609	37	1.000			> 0.999
g	17.609	10	1.000			> 0.999
j	17.609	6	1.000			> 0.999
d	2.536	131	0.924			0.927
b	2.323	97	0.907			0.911
c	2.168	150	0.893			0.897
9	0.098	181	0.514			0.524
5	-0.887	53	0.283			0.292
6	-1.107	79	0.241			0.248
7	-1.361	71	0.197			0.204
8	-4.025	297	0.017			0.018
0	-17.523	1	0.000			< 0.001
1	-17.523	1	0.000			< 0.001
4	-17.523	1	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	14	-0.043			0.486		0.489

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-377.401	782.802	783.184	0.844	0.869

Run 9 (above) with periodo is better than Run 0 without periodo, $p = 1.86e-160$

Adding pessoa...

STEP 2 - adding to Run 2 model with pessoa

Trying with expressao...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	0.49	427	0.958			0.62
n	-0.49	688	0.193			0.38

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	5.679	543	0.996			0.997
2sg	-5.679	572	0.002			0.003

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	3	-0.227			0.486		0.443

\$misc.2

⁵¹ Período (tempo) - códigos utilizados no grupo de fatores: f – 1970-1979; g – 1980-1989; d – 1960-1969; b – 1940-1949; c – 1950-1959; 9 – 1930-1939; 5 – 1890-1899; 6 – 1900-1909; 7 – 1910-1919; 8 – 1920-1929; 0 – 1840-1849; 1 – 1850-1859; 4 – 1880-1889.

```

log.likelihood    AIC    AICc    Dxy    R2
      -20.304 46.607 46.629 0.996 0.917

```

Run 10 (above) with pessoa + expressao is better than Run 2 without expressao, $p = 0.481$

Trying with paralelismo...

\$pessoa

```

factor logodds tokens v/v+t centered factor weight vif
  3sg   23.249   543 0.996                > 0.999 >20
  2sg  -23.249   572 0.002                < 0.001 >20

```

\$paralelismo

```

factor logodds tokens v/v+t centered factor weight vif
  2   22.622   231 0.935                > 0.999 >20
  4    2.979    71 0.930                0.952 >20
  3    2.817    34 0.882                0.944 >20
  y    2.665    40 0.825                0.935 >20
  1    2.241   273 0.601                0.904 >20
  9    1.727     7 0.286                0.849 >20
  t    1.501    17 0.176                0.818 >20
  5    1.128   245 0.057                0.755 >20
  6    0.764    17 0.000                0.682 >20
  7  -19.222   145 0.048                < 0.001 >20
  z  -19.222    35 0.200                < 0.001 >20

```

\$misc.1

```

n df intercept overall proportion centered input prob
1115 12      -2.081                0.486                0.111

```

\$misc.2

```

log.likelihood    AIC    AICc    Dxy    R2
      -9.769 43.538 43.821    1 0.997

```

Run 11 (above) with pessoa + paralelismo is better than Run 2 without paralelismo, $p = 0.0175$

Trying with carta...

\$pessoa

```

factor logodds tokens v/v+t centered factor weight
  3sg   5.936   543 0.996                0.997
  2sg  -5.936   572 0.002                0.003

```

\$carta

```

factor logodds tokens v/v+t centered factor weight
  z   3.994   326 0.837                0.982
  f   3.115   424 0.472                0.958
  a   1.350   363 0.190                0.794
  c  -8.459     2 0.000                < 0.001

```

\$misc.1

```

n df intercept overall proportion centered input prob

```

1115 5 -3.172 0.486 0.04

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-18.998	47.995	48.049	0.998	0.929

Run 12 (above) with pessoa + carta is better than Run 2 without carta, p = 0.375

Trying with genero...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	6	543	0.996			0.998
2sg	-6	572	0.002			0.002

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.298	348	0.517			0.574
h	-0.298	767	0.472			0.426

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	3	-0.263			0.486		0.435

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-20.444	46.889	46.911	0.995	0.917

Run 13 (above) with pessoa + genero is better than Run 2 without genero, p = 0.642

Trying with fxetaria...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.174	543	0.996			> 0.999	>20
2sg	-23.174	572	0.002			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	19.027	88	0.341			> 0.999	>20
c	0.305	1	1.000			0.576	>20
v	0.021	273	0.659			0.505	>20
k	-0.152	532	0.404			0.462	>20
j	-19.201	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	6	0.087			0.486		0.522

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-15.207	42.414	42.49	0.999	0.995

Run 14 (above) with pessoa + fxetaria is better than Run 2 without fxetaria, $p = 0.0303$

Trying with interpessoal...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.088	543	0.996			> 0.999	>20
2sg	-23.088	572	0.002			< 0.001	>20

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	23.113	39	0.256			> 0.999	>20
j	4.256	70	0.900			0.986	>20
l	4.216	18	0.889			0.985	>20
n	4.216	18	0.889			0.985	>20
b	4.066	325	0.840			0.983	>20
h	3.327	13	0.462			0.965	>20
m	3.162	120	0.367			0.959	>20
p	2.813	61	0.197			0.943	>20
k	2.795	37	0.189			0.942	>20
e	1.914	2	0.000			0.871	>20
u	1.914	1	0.000			0.871	>20
i	-15.142	130	0.731			< 0.001	>20
a	-40.649	281	0.000			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	14	-3.392			0.486		0.033

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.943	47.887	48.269	1	0.998

Run 15 (above) with pessoa + interpessoal is better than Run 2 without interpessoal, $p = 0.0473$

Trying with social...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	5.976	543	0.996			0.997
2sg	-5.976	572	0.002			0.003

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	0.732	21	0.810			0.675
d	-0.330	272	0.338			0.418
s	-0.403	822	0.527			0.401

\$misc.1

	n	df	intercept	overall	proportion	centered	input	prob
	1115	4	-0.007			0.486		0.498

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-20.523	49.046	49.082	0.995	0.916

Run 16 (above) with pessoa + social is better than Run 2 without social, p = 0.971

Trying with periodo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	5.667	543	0.996			0.997
2sg	-5.667	572	0.002			0.003

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
f	15.066	37	1.000			> 0.999
g	15.066	10	1.000			> 0.999
j	15.066	6	1.000			> 0.999
d	1.419	131	0.924			0.805
b	1.312	97	0.907			0.788
c	1.234	150	0.893			0.774
9	0.194	181	0.514			0.548
5	-0.300	53	0.283			0.426
6	-0.410	79	0.241			0.399
7	-0.538	71	0.197			0.369
8	-3.911	297	0.017			0.02
0	-14.732	1	0.000			< 0.001
1	-14.732	1	0.000			< 0.001
4	-14.732	1	0.000			< 0.001

\$misc.1

	n	df	intercept	overall	proportion	centered	input	prob
	1115	15	-0.167			0.486		0.458

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-16.025	62.051	62.487	0.999	0.957

Run 17 (above) with pessoa + periodo is better than Run 2 without periodo, p = 0.7

Adding paralelismo...

STEP 3 - adding to Run 11 model with pessoa + paralelismo

Trying with expressao...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	0.683	427	0.958			0.664

```

      n  -0.683    688 0.193
                                0.336

$peessoa
  factor logodds tokens v/v+t centered factor weight vif
    3sg  22.721    543 0.996                > 0.999 >20
    2sg -22.721    572 0.002                < 0.001 >20

$paralelismo
  factor logodds tokens v/v+t centered factor weight vif
    2  22.437    231 0.935                > 0.999 >20
    4   3.070     71 0.930                0.956 >20
    3   2.961     34 0.882                0.951 >20
    y   2.829     40 0.825                0.944 >20
    1   2.011    273 0.601                0.882 >20
    9   1.939      7 0.286                0.874 >20
    t   1.495     17 0.176                0.817 >20
    5   1.067    245 0.057                0.744 >20
    6   0.591     17 0.000                0.644 >20
    z -19.086     35 0.200                < 0.001 >20
    7 -19.313    145 0.048                < 0.001 >20

$misc.1
  n df intercept overall proportion centered input prob
 1115 13    -1.889                0.486                0.131

$misc.2
  log.likelihood    AIC    AICc Dxy    R2
    -9.357 44.714 45.045    1 0.997

Run 18 (above) with pessoa + paralelismo + expressao is better than Run 11
without expressao, p = 0.364

Trying with carta...

$peessoa
  factor logodds tokens v/v+t centered factor weight vif
    3sg  23.196    543 0.996                > 0.999 >20
    2sg -23.196    572 0.002                < 0.001 >20

$paralelismo
  factor logodds tokens v/v+t centered factor weight vif
    2  22.038    231 0.935                > 0.999 >20
    3   2.845     34 0.882                0.945 >20
    4   2.825     71 0.930                0.944 >20
    y   2.553     40 0.825                0.928 >20
    1   2.070    273 0.601                0.888 >20
    t   1.930     17 0.176                0.873 >20
    6   1.785     17 0.000                0.856 >20
    5   1.180    245 0.057                0.765 >20
    9   1.141      7 0.286                0.758 >20
    z -18.997     35 0.200                < 0.001 >20
    7 -19.370    145 0.048                < 0.001 >20

$carta

```

factor	logodds	tokens	v/v+t	centered	factor	weight
z	1.061	326	0.837			0.743
f	0.898	424	0.472			0.711
a	-0.956	363	0.190			0.278
c	-1.003	2	0.000			0.268

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	15	-2.436			0.486		0.08

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.093	48.186	48.623	1	0.997

Run 19 (above) with pessoa + paralelismo + carta is better than Run 11 without carta, $p = 0.717$

Trying with genero...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.322	543	0.996			> 0.999	>20
2sg	-23.322	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	22.695	231	0.935			> 0.999	>20
4	2.992	71	0.930			0.952	>20
3	2.872	34	0.882			0.946	>20
y	2.515	40	0.825			0.925	>20
1	2.232	273	0.601			0.903	>20
9	1.899	7	0.286			0.87	>20
t	1.381	17	0.176			0.799	>20
5	1.158	245	0.057			0.761	>20
6	0.796	17	0.000			0.689	>20
7	-19.167	145	0.048			< 0.001	>20
z	-19.372	35	0.200			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.456	348	0.517			0.612
h	-0.456	767	0.472			0.388

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	13	-1.927			0.486		0.127

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.544	45.087	45.418	1	0.997

Run 20 (above) with pessoa + paralelismo + genero is better than Run 11 without genero, $p = 0.502$

Trying with fxetaria...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.477	543	0.996			> 0.999	>20
2sg	-44.477	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.042	231	0.935			> 0.999	>20
6	16.664	17	0.000			> 0.999	>20
9	16.162	7	0.286			> 0.999	>20
4	0.039	71	0.930			0.51	>20
3	0.021	34	0.882			0.505	>20
y	-0.748	40	0.825			0.321	>20
1	-0.804	273	0.601			0.309	>20
t	-1.514	17	0.176			0.18	>20
5	-2.083	245	0.057			0.111	>20
7	-24.388	145	0.048			< 0.001	>20
z	-24.388	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.014	88	0.341			> 0.999	>20
k	3.872	532	0.404			0.98	>20
v	3.767	273	0.659			0.977	>20
c	-14.836	1	1.000			< 0.001	>20
j	-17.817	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	16	-2.271			0.486		0.094

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.682	41.364	41.86	1	0.999

Run 21 (above) with pessoa + paralelismo + fxetaria is better than Run 11 without fxetaria, $p = 0.0376$

Trying with interpessoal...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	43.03	543	0.996			> 0.999	>20
2sg	-43.03	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.763	231	0.935			> 0.999	>20
6	21.416	17	0.000			> 0.999	>20
3	1.959	34	0.882			0.876	>20
4	0.917	71	0.930			0.714	>20
y	0.101	40	0.825			0.525	>20

1	-0.015	273	0.601		0.496	>20
5	-0.717	245	0.057		0.328	>20
9	-1.148	7	0.286		0.241	>20
t	-1.337	17	0.176		0.208	>20
7	-20.256	145	0.048		< 0.001	>20
z	-22.682	35	0.200		< 0.001	>20

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	20.915	39	0.256			> 0.999	>20
e	16.128	2	0.000			> 0.999	>20
u	16.128	1	0.000			> 0.999	>20
p	16.012	61	0.197			> 0.999	>20
h	13.620	13	0.462			> 0.999	>20
j	0.392	70	0.900			0.597	>20
n	0.207	18	0.889			0.552	>20
k	-0.521	37	0.189			0.373	>20
b	-0.736	325	0.840			0.324	>20
m	-1.557	120	0.367			0.174	>20
l	-16.847	18	0.889			< 0.001	>20
i	-20.005	130	0.731			< 0.001	>20
a	-43.735	281	0.000			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	24		0.351		0.486		0.587

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-3.296	54.592	55.693	1	0.999

Run 22 (above) with pessoa + paralelismo + interpessoal is better than Run 11 without interpessoal, p = 0.373

Trying with social...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.269	543	0.996			> 0.999	>20
2sg	-23.269	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	22.603	231	0.935			> 0.999	>20
4	2.979	71	0.930			0.952	>20
3	2.823	34	0.882			0.944	>20
y	2.647	40	0.825			0.934	>20
1	2.229	273	0.601			0.903	>20
9	1.758	7	0.286			0.853	>20
t	1.494	17	0.176			0.817	>20
5	1.118	245	0.057			0.754	>20
6	0.827	17	0.000			0.696	>20

z	-19.211	35	0.200		< 0.001	>20
7	-19.269	145	0.048		< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
d	0.116	272	0.338			0.529
a	0.012	21	0.810			0.503
s	-0.127	822	0.527			0.468

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	14	-2.013			0.486		0.118

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.754	47.508	47.89	1	0.997

Run 23 (above) with pessoa + paralelismo + social is better than Run 11 without social, $p = 0.985$

Trying with periodo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	49.822	543	0.996			> 0.999	>20
2sg	-49.822	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	37.038	231	0.935			> 0.999	>20
6	16.939	17	0.000			> 0.999	>20
1	15.751	273	0.601			> 0.999	>20
5	14.084	245	0.057			> 0.999	>20
t	13.676	17	0.176			> 0.999	>20
9	11.863	7	0.286			> 0.999	>20
3	-8.624	34	0.882			< 0.001	>20
4	-8.680	71	0.930			< 0.001	>20
y	-9.638	40	0.825			< 0.001	>20
7	-23.845	145	0.048			< 0.001	>20
z	-58.566	35	0.200			< 0.001	>20

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
d	16.653	131	0.924			> 0.999	>20
7	16.259	71	0.197			> 0.999	>20
9	16.259	181	0.514			> 0.999	>20
6	15.336	79	0.241			> 0.999	>20
0	13.001	1	0.000			> 0.999	>20
1	13.001	1	0.000			> 0.999	>20
4	13.001	1	0.000			> 0.999	>20
b	0.968	97	0.907			0.725	>20
5	-0.716	53	0.283			0.328	>20
c	-9.242	150	0.893			< 0.001	>20
f	-10.511	37	1.000			< 0.001	>20

j	-10.532	6	1.000	< 0.001	>20
g	-33.799	10	1.000	< 0.001	>20
8	-39.677	297	0.017	< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	25	-5.495		0.486			0.004

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-6.181	62.363	63.557	1	0.999

Run 24 (above) with pessoa + paralelismo + periodo is better than Run 11 without periodo, $p = 0.893$

Adding fxetaria...

STEP 4 - adding to Run 21 model with pessoa + paralelismo + fxetaria

Trying with expressao...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	0.302	427	0.958			0.575
n	-0.302	688	0.193			0.425

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.094	543	0.996			> 0.999	>20
2sg	-44.094	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	20.868	231	0.935			> 0.999	>20
6	16.529	17	0.000			> 0.999	>20
9	15.874	7	0.286			> 0.999	>20
4	0.090	71	0.930			0.523	>20
3	-0.087	34	0.882			0.478	>20
y	-0.615	40	0.825			0.351	>20
1	-0.806	273	0.601			0.309	>20
t	-1.220	17	0.176			0.228	>20
5	-2.084	245	0.057			0.111	>20
7	-24.124	145	0.048			< 0.001	>20
z	-24.426	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	24.830	88	0.341			> 0.999	>20
v	3.939	273	0.659			0.981	>20
k	3.808	532	0.404			0.978	>20
c	-14.816	1	1.000			< 0.001	>20
j	-17.762	221	0.525			< 0.001	>20

\$misc.1

```

      n df intercept overall proportion centered input prob
1115 17      -2.208                0.486                0.099

```

```
$misc.2
```

```

log.likelihood      AIC      AICc Dxy      R2
      -4.632 43.263 43.821   1 0.999

```

Run 25 (above) with pessoa + paralelismo + fxetaria + expressao is better than Run 21 without expressao, p = 0.751

Trying with carta...

```
$pessoa
```

```

factor logodds tokens v/v+t centered factor weight vif
3sg  76.985     543 0.996                > 0.999 >20
2sg -76.985     572 0.002                < 0.001 >20

```

```
$paralelismo
```

```

factor logodds tokens v/v+t centered factor weight vif
  2  25.054     231 0.935                > 0.999 >20
  6  22.268      17 0.000                > 0.999 >20
  9  19.355       7 0.286                > 0.999 >20
  3   7.441      34 0.882                0.999 >20
  1   4.043     273 0.601                0.983 >20
  4   2.606      71 0.930                0.931 >20
  y   1.959      40 0.825                0.876 >20
  5   1.694     245 0.057                0.845 >20
  t   1.188      17 0.176                0.766 >20
  7 -33.596     145 0.048                < 0.001 >20
  z -52.013      35 0.200                < 0.001 >20

```

```
$carta
```

```

factor logodds tokens v/v+t centered factor weight vif
  c  29.963       2 0.000                > 0.999 >10
  f  12.257     424 0.472                > 0.999 >10
  z -17.538     326 0.837                < 0.001 >10
  a -24.682     363 0.190                < 0.001 >10

```

```
$fxetaria
```

```

factor logodds tokens v/v+t centered factor weight vif
  x  44.594      88 0.341                > 0.999 >20
  v  25.229     273 0.659                > 0.999 >20
  k  22.897     532 0.404                > 0.999 >20
  j -31.615     221 0.525                < 0.001 >20
  c -61.105       1 1.000                < 0.001 >20

```

```
$misc.1
```

```

      n df intercept overall proportion centered input prob
1115 19      -5.613                0.486                0.004

```

```
$misc.2
```

```

log.likelihood      AIC      AICc Dxy R2
      -3.296 44.592 45.286   1 1

```

Run 26 (above) with pessoa + paralelismo + fxetaria + carta is better than Run 21 without carta, $p = 0.428$

Trying with genero...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.543	543	0.996			> 0.999	>20
2sg	-44.543	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.076	231	0.935			> 0.999	>20
6	16.752	17	0.000			> 0.999	>20
9	16.251	7	0.286			> 0.999	>20
4	-0.099	71	0.930			0.475	>20
3	-0.175	34	0.882			0.456	>20
y	-0.766	40	0.825			0.317	>20
1	-0.890	273	0.601			0.291	>20
t	-1.218	17	0.176			0.228	>20
5	-2.006	245	0.057			0.119	>20
7	-24.311	145	0.048			< 0.001	>20
z	-24.613	35	0.200			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.302	767	0.472			0.575
m	-0.302	348	0.517			0.425

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.095	88	0.341			> 0.999	>20
v	3.934	273	0.659			0.981	>20
k	3.521	532	0.404			0.971	>20
c	-14.551	1	1.000			< 0.001	>20
j	-17.999	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	17	-2.233			0.486		0.097

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.632	43.263	43.821	1	0.999

Run 27 (above) with pessoa + paralelismo + fxetaria + genero is better than Run 21 without genero, $p = 0.751$

Trying with interpersoal...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	55.398	543	0.996			> 0.999	>20
2sg	-55.398	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
6	52.591	17	0.000			> 0.999	>20
2	30.962	231	0.935			> 0.999	>20
9	24.295	7	0.286			> 0.999	>20
4	-6.640	71	0.930			0.001	>20
3	-7.793	34	0.882			< 0.001	>20
1	-8.532	273	0.601			< 0.001	>20
y	-8.734	40	0.825			< 0.001	>20
5	-9.391	245	0.057			< 0.001	>20
t	-10.626	17	0.176			< 0.001	>20
7	-26.118	145	0.048			< 0.001	>20
z	-30.013	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	27.985	88	0.341			> 0.999	>20
k	15.396	532	0.404			> 0.999	>20
v	3.476	273	0.659			0.97	>20
j	-11.107	221	0.525			< 0.001	>20
c	-35.750	1	1.000			< 0.001	>20

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
p	33.846	61	0.197			> 0.999	>20
u	33.483	1	0.000			> 0.999	>20
e	22.419	2	0.000			> 0.999	>20
h	21.690	13	0.462			> 0.999	>20
s	15.046	39	0.256			> 0.999	>20
n	4.957	18	0.889			0.993	>20
j	-6.314	70	0.900			0.002	>20
l	-9.012	18	0.889			< 0.001	>20
b	-12.594	325	0.840			< 0.001	>20
i	-14.682	130	0.731			< 0.001	>20
k	-24.410	37	0.189			< 0.001	>20
m	-25.207	120	0.367			< 0.001	>20
a	-39.223	281	0.000			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	28		0.404		0.486		0.6

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	58.773	60.268	1	0.999

Run 28 (above) with pessoa + paralelismo + fxetaria + interpessoal is better than Run 21 without interpessoal, $p = 0.883$

Trying with social...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
--------	---------	--------	-------	----------	--------	--------	-----

3sg	86.721	543	0.996		> 0.999	>20
2sg	-86.721	572	0.002		< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	45.646	231	0.935			> 0.999	>20
6	41.006	17	0.000			> 0.999	>20
9	40.660	7	0.286			> 0.999	>20
y	7.905	40	0.825			> 0.999	>20
5	7.120	245	0.057			0.999	>20
t	3.611	17	0.176			0.974	>20
1	2.861	273	0.601			0.946	>20
4	-23.943	71	0.930			< 0.001	>20
3	-25.269	34	0.882			< 0.001	>20
7	-49.799	145	0.048			< 0.001	>20
z	-49.799	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	59.196	88	0.341			> 0.999	>20
v	20.442	273	0.659			> 0.999	>20
k	19.634	532	0.404			> 0.999	>20
j	-36.898	221	0.525			< 0.001	>20
c	-62.374	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	12.548	822	0.527			> 0.999	>2.5
a	11.930	21	0.810			> 0.999	>2.5
d	-24.478	272	0.338			< 0.001	>2.5

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	18	-12.572			0.486		0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-2.773	41.545	42.169	1	1

Run 29 (above) with pessoa + paralelismo + fxetaria + social is better than Run 21 without social, $p = 0.148$

Trying with periodo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	69.439	543	0.996			> 0.999	>20
2sg	-69.439	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	27.181	231	0.935			> 0.999	>20
6	26.732	17	0.000			> 0.999	>20
9	23.922	7	0.286			> 0.999	>20

4	6.123	71	0.930		0.998	>20
1	5.999	273	0.601		0.998	>20
y	3.505	40	0.825		0.971	>20
3	3.222	34	0.882		0.962	>20
5	3.106	245	0.057		0.957	>20
t	2.546	17	0.176		0.927	>20
7	-32.278	145	0.048		< 0.001	>20
z	-70.059	35	0.200		< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	16.926	88	0.341			> 0.999	>20
k	16.499	532	0.404			> 0.999	>20
v	15.253	273	0.659			> 0.999	>20
c	-23.964	1	1.000			< 0.001	>20
j	-24.713	221	0.525			< 0.001	>20

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
0	63.712	1	0.000			> 0.999	>20
4	63.712	1	0.000			> 0.999	>20
9	27.458	181	0.514			> 0.999	>20
1	22.500	1	0.000			> 0.999	>20
5	22.234	53	0.283			> 0.999	>20
7	8.655	71	0.197			> 0.999	>20
d	8.388	131	0.924			> 0.999	>20
6	6.471	79	0.241			0.998	>20
b	-22.783	97	0.907			< 0.001	>20
c	-23.191	150	0.893			< 0.001	>20
f	-24.229	37	1.000			< 0.001	>20
8	-29.409	297	0.017			< 0.001	>20
j	-61.234	6	1.000			< 0.001	>20
g	-62.286	10	1.000			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	29		-2.126		0.486		0.107

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-2.773	63.545	65.149	1	1

Run 30 (above) with pessoa + paralelismo + fxetaria + periodo is better than Run 21 without periodo, $p = 0.993$

No significant improvement this step, best model from last step is Run 21

BEST STEP-UP MODEL OF RESPONSE pronome IS WITH PREDICTOR(S): pessoa (~0) + paralelismo (0.0175) + fxetaria (0.0376)
[p-values building from null model]

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.477	543	0.996			> 0.999	>20

2sg -44.477 572 0.002 < 0.001 >20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.042	231	0.935			> 0.999	>20
6	16.664	17	0.000			> 0.999	>20
9	16.162	7	0.286			> 0.999	>20
4	0.039	71	0.930			0.51	>20
3	0.021	34	0.882			0.505	>20
y	-0.748	40	0.825			0.321	>20
1	-0.804	273	0.601			0.309	>20
t	-1.514	17	0.176			0.18	>20
5	-2.083	245	0.057			0.111	>20
7	-24.388	145	0.048			< 0.001	>20
z	-24.388	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.014	88	0.341			> 0.999	>20
k	3.872	532	0.404			0.98	>20
v	3.767	273	0.659			0.977	>20
c	-14.836	1	1.000			< 0.001	>20
j	-17.817	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	16	-2.271			0.486		0.094

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.682	41.364	41.86	1	0.999

STEPPING DOWN...

STEP 0 - Run 0 - full model with expressao + pessoa + paralelismo + carta + genero + fxetaria + interpessoal + social + periodo

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	6.755	427	0.958			0.999
n	-6.755	688	0.193			0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	38.792	543	0.996			> 0.999
2sg	-38.792	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	33.925	17	0.000			> 0.999
2	22.894	231	0.935			> 0.999
9	18.335	7	0.286			> 0.999

4	1.408	71	0.930		0.803
y	-3.056	40	0.825		0.045
5	-5.122	245	0.057		0.006
t	-5.956	17	0.176		0.003
3	-6.896	34	0.882		0.001
1	-10.928	273	0.601	<	0.001
7	-20.962	145	0.048	<	0.001
z	-23.642	35	0.200	<	0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
c	59957878	2	0.000		> 0.999
z	-19985952	326	0.837		< 0.001
a	-19985959	363	0.190		< 0.001
f	-19985967	424	0.472		< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight
m	4.128	348	0.517		0.984
h	-4.128	767	0.472		0.016

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight
x	11.272	88	0.341		> 0.999
v	10.735	273	0.659		> 0.999
k	7.420	532	0.404		0.999
j	-14.224	221	0.525		< 0.001
c	-15.203	1	1.000		< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor weight
u	6149549	1	0.000		> 0.999
l	6149547	18	0.889		> 0.999
n	6149530	18	0.889		> 0.999
h	6149528	13	0.462		> 0.999
s	6149528	39	0.256		> 0.999
p	6149526	61	0.197		> 0.999
i	6149512	130	0.731		> 0.999
b	6149509	325	0.840		> 0.999
m	6149507	120	0.367		> 0.999
j	6149506	70	0.900		> 0.999
k	6149502	37	0.189		> 0.999
a	6149460	281	0.000		> 0.999
e	-73794204	2	0.000		< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor weight
s	6.301	822	0.527		0.998
a	-6.301	21	0.810		0.002
d	NA	272	0.338		<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor weight
4	41.945	1	0.000		> 0.999

7	26.527	71	0.197	>	0.999
5	25.462	53	0.283	>	0.999
8	25.214	297	0.017	>	0.999
b	15.137	97	0.907	>	0.999
c	13.149	150	0.893	>	0.999
9	12.908	181	0.514	>	0.999
f	2.486	37	1.000		0.923
d	1.041	131	0.924		0.739
6	0.793	79	0.241		0.688
j	-20.373	6	1.000	<	0.001
0	-61.323	1	0.000	<	0.001
1	-82.967	1	0.000	<	0.001
g	NA	10	1.000		<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	46	13836433		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	94.773	98.821	1	0.999

STEP 1 - dropping from Run 0 model with expressao + pessoa + paralelismo + carta + genero + fxetaria + interpessoal + social + periodo

Trying without expressao...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	51.93	543	0.996			> 0.999
2sg	-51.93	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	36.684	17	0.000			> 0.999
2	32.869	231	0.935			> 0.999
9	26.212	7	0.286			> 0.999
4	-5.635	71	0.930			0.004
y	-6.527	40	0.825			0.001
3	-6.758	34	0.882			0.001
1	-7.353	273	0.601			0.001
5	-8.033	245	0.057			< 0.001
t	-8.603	17	0.176			< 0.001
7	-26.147	145	0.048			< 0.001
z	-26.708	35	0.200			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
c	312658645	2	0.000			> 0.999
z	-104219538	326	0.837			< 0.001
f	-104219552	424	0.472			< 0.001
a	-104219555	363	0.190			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.177	348	0.517			0.544
h	-0.177	767	0.472			0.456

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
v	25.892	273	0.659			> 0.999
x	12.642	88	0.341			> 0.999
k	9.540	532	0.404			> 0.999
c	-21.566	1	1.000			< 0.001
j	-26.509	221	0.525			< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight
e	171.776	2	0.000			> 0.999
l	22.624	18	0.889			> 0.999
u	17.377	1	0.000			> 0.999
n	-4.840	18	0.889			0.008
h	-8.038	13	0.462			< 0.001
p	-16.600	61	0.197			< 0.001
b	-19.156	325	0.840			< 0.001
j	-19.176	70	0.900			< 0.001
i	-23.940	130	0.731			< 0.001
m	-35.567	120	0.367			< 0.001
k	-38.029	37	0.189			< 0.001
a	-46.429	281	0.000			< 0.001
s	NA	39	0.256			<NA>

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
s	4.595	822	0.527			0.99
a	-4.595	21	0.810			0.01
d	NA	272	0.338			<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	59554083	1	0.000			> 0.999
5	59554068	53	0.283			> 0.999
8	59554066	297	0.017			> 0.999
7	59554065	71	0.197			> 0.999
9	59554062	181	0.514			> 0.999
b	59554049	97	0.907			> 0.999
f	59554048	37	1.000			> 0.999
c	59554048	150	0.893			> 0.999
6	59554030	79	0.241			> 0.999
d	59554030	131	0.924			> 0.999
g	59554014	10	1.000			> 0.999
j	59554000	6	1.000			> 0.999
0	-357324264	1	0.000			< 0.001
1	-357324300	1	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	45	44665510		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	92.773	96.645	1	0.999

Run 1 (above) with pessoa + paralelismo + carta + genero + fxetaria + interpessoal + social + periodo is worse than Run 0 also including expressao, p = 1

Trying without pessoa...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	5.002	427	0.958			0.993
n	-5.002	688	0.193			0.007

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	15946.562	17	0.000			> 0.999
t	-1587.483	17	0.176			< 0.001
9	-1591.275	7	0.286			< 0.001
4	-1593.261	71	0.930			< 0.001
y	-1593.341	40	0.825			< 0.001
2	-1593.654	231	0.935			< 0.001
5	-1594.868	245	0.057			< 0.001
z	-1597.503	35	0.200			< 0.001
7	-1597.742	145	0.048			< 0.001
3	-1598.146	34	0.882			< 0.001
1	-1599.288	273	0.601			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
a	5.62950e+14	363	0.190			> 0.999
f	5.62950e+14	424	0.472			> 0.999
z	5.62950e+14	326	0.837			> 0.999
c	-1.68885e+15	2	0.000			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	3.291	348	0.517			0.964
h	-3.291	767	0.472			0.036

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
c	14.509	1	1.000			> 0.999
v	0.621	273	0.659			0.65
k	-4.385	532	0.404			0.012
x	-5.031	88	0.341			0.006
j	-5.714	221	0.525			0.003

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight
--------	---------	--------	-------	----------	--------	--------

b	24.413	325	0.840		> 0.999
e	22.197	2	0.000		> 0.999
n	12.391	18	0.889		> 0.999
h	8.365	13	0.462		> 0.999
i	5.260	130	0.731		0.995
k	0.748	37	0.189		0.679
l	-1.502	18	0.889		0.182
m	-6.102	120	0.367		0.002
j	-8.614	70	0.900		< 0.001
p	-9.715	61	0.197		< 0.001
a	-22.813	281	0.000		< 0.001
u	-24.626	1	0.000		< 0.001
s	NA	39	0.256		<NA>

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	1.064	21	0.810			0.743
s	-1.064	822	0.527			0.257
d	NA	272	0.338			<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
1	2.2518e+15	1	0.000			> 0.999
f	1.6311e+01	37	1.000			> 0.999
d	-6.5280e+00	131	0.924			0.002
7	-1.1178e+01	71	0.197			< 0.001
b	-1.3517e+01	97	0.907			< 0.001
8	-1.3753e+01	297	0.017			< 0.001
9	-1.4272e+01	181	0.514			< 0.001
c	-1.5436e+01	150	0.893			< 0.001
5	-1.6820e+01	53	0.283			< 0.001
6	-1.7253e+01	79	0.241			< 0.001
4	-3.0049e+01	1	0.000			< 0.001
j	-3.7004e+01	6	1.000			< 0.001
0	-2.2518e+15	1	0.000			< 0.001
g	NA	10	1.000			<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	44	-5.6295e+14		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1122.838	2333.677	2337.377	0.887	1

Run 2 (above) with expressao + paralelismo + carta + genero + fxetaria + interpessoal + social + periodo is worse than Run 0 also including pessoa, p = 0

Trying without paralelismo...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	19.433	427	0.958			> 0.999

n	-19.433	688	0.193		< 0.001
---	---------	-----	-------	--	---------

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor weight
3sg	53.521	543	0.996		> 0.999
2sg	-53.521	572	0.002		< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
a	1588134	363	0.190		> 0.999
f	1588116	424	0.472		> 0.999
z	1588102	326	0.837		> 0.999
c	-4764352	2	0.000		< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight
h	3.127	767	0.472		0.958
m	-3.127	348	0.517		0.042

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight
v	43.283	273	0.659		> 0.999
x	33.823	88	0.341		> 0.999
k	30.015	532	0.404		> 0.999
j	-30.745	221	0.525		< 0.001
c	-76.376	1	1.000		< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor weight
e	5864180.0	2	0.000		> 0.999
u	-488628.7	1	0.000		< 0.001
b	-488638.2	325	0.840		< 0.001
s	-488644.8	39	0.256		< 0.001
l	-488647.8	18	0.889		< 0.001
n	-488651.2	18	0.889		< 0.001
h	-488652.3	13	0.462		< 0.001
k	-488664.6	37	0.189		< 0.001
m	-488687.6	120	0.367		< 0.001
i	-488711.5	130	0.731		< 0.001
j	-488722.6	70	0.900		< 0.001
p	-488726.0	61	0.197		< 0.001
a	-488804.7	281	0.000		< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor weight
a	2.838	21	0.810		0.945
s	-2.838	822	0.527		0.055
d	NA	272	0.338		<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor weight
8	109.689	297	0.017		> 0.999

4	99.567	1	0.000	> 0.999
f	57.889	37	1.000	> 0.999
b	55.719	97	0.907	> 0.999
c	51.949	150	0.893	> 0.999
9	50.271	181	0.514	> 0.999
7	49.353	71	0.197	> 0.999
5	48.703	53	0.283	> 0.999
d	48.445	131	0.924	> 0.999
6	15.554	79	0.241	> 0.999
j	-17.293	6	1.000	< 0.001
0	-254.544	1	0.000	< 0.001
1	-315.304	1	0.000	< 0.001
g	NA	10	1.000	<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	36	-1099496		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.91	75.819	78.29	1	1

Run 3 (above) with expressao + pessoa + carta + genero + fxetaria + interpessoal + social + periodo is worse than Run 0 also including paralelismo, p = 1

Trying without carta...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	6.558	427	0.958			0.999
n	-6.558	688	0.193			0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	39.083	543	0.996			> 0.999
2sg	-39.083	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	34.039	17	0.000			> 0.999
2	22.931	231	0.935			> 0.999
9	18.267	7	0.286			> 0.999
4	1.041	71	0.930			0.739
y	-3.381	40	0.825			0.033
5	-5.437	245	0.057			0.004
3	-5.928	34	0.882			0.003
t	-6.369	17	0.176			0.002
1	-11.042	273	0.601			< 0.001
7	-20.182	145	0.048			< 0.001
z	-23.939	35	0.200			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	3.156	348	0.517			0.959
h	-3.156	767	0.472			0.041

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
v	12.309	273	0.659			> 0.999
x	11.620	88	0.341			> 0.999
k	6.987	532	0.404			0.999
j	-15.422	221	0.525			< 0.001
c	-15.494	1	1.000			< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight
e	128.868	2	0.000			> 0.999
u	18.043	1	0.000			> 0.999
l	18.040	18	0.889			> 0.999
h	-0.685	13	0.462			0.335
n	-1.122	18	0.889			0.246
s	-1.388	39	0.256			0.2
b	-5.250	325	0.840			0.005
p	-6.196	61	0.197			0.002
j	-16.693	70	0.900			< 0.001
i	-18.973	130	0.731			< 0.001
m	-23.438	120	0.367			< 0.001
k	-29.861	37	0.189			< 0.001
a	-61.344	281	0.000			< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
s	5.767	822	0.527			0.997
a	-5.767	21	0.810			0.003
d	NA	272	0.338			<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	42.586	1	0.000			> 0.999
7	28.731	71	0.197			> 0.999
8	27.953	297	0.017			> 0.999
5	27.754	53	0.283			> 0.999
b	17.623	97	0.907			> 0.999
9	16.245	181	0.514			> 0.999
c	15.616	150	0.893			> 0.999
f	6.964	37	1.000			0.999
d	3.500	131	0.924			0.971
6	2.772	79	0.241			0.941
j	-18.227	6	1.000			< 0.001
0	-74.554	1	0.000			< 0.001
1	-96.963	1	0.000			< 0.001
g	NA	10	1.000			<NA>

\$misc.1

n df intercept overall proportion centered input prob

u	6975961	1	0.000	>	0.999
s	6975941	39	0.256	>	0.999
n	6975937	18	0.889	>	0.999
h	6975937	13	0.462	>	0.999
p	6975931	61	0.197	>	0.999
b	6975929	325	0.840	>	0.999
i	6975920	130	0.731	>	0.999
m	6975917	120	0.367	>	0.999
j	6975912	70	0.900	>	0.999
k	6975906	37	0.189	>	0.999
a	6975870	281	0.000	>	0.999
e	-83711123	2	0.000	<	0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
s	4.891	822	0.527			0.993
a	-4.891	21	0.810			0.007
d	NA	272	0.338			<NA>

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	43.174	1	0.000			> 0.999
5	33.650	53	0.283			> 0.999
8	32.792	297	0.017			> 0.999
7	32.776	71	0.197			> 0.999
b	22.282	97	0.907			> 0.999
9	22.216	181	0.514			> 0.999
c	20.320	150	0.893			> 0.999
f	18.645	37	1.000			> 0.999
6	7.216	79	0.241			0.999
d	6.805	131	0.924			0.999
j	-11.603	6	1.000			< 0.001
0	-101.696	1	0.000			< 0.001
1	-126.577	1	0.000			< 0.001
g	NA	10	1.000			<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	45	15695861			0.486		NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	92.773	96.645	1	0.999

Run 5 (above) with expressao + pessoa + paralelismo + carta + fxetaria + interpessoal + social + periodo is worse than Run 0 also including genero, p = 1

Trying without fxetaria...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	6.7	427	0.958			0.999
n	-6.7	688	0.193			0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	38.872	543	0.996			> 0.999
2sg	-38.872	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	24.511	231	0.935			> 0.999
6	23.215	17	0.000			> 0.999
9	18.775	7	0.286			> 0.999
3	-1.248	34	0.882			0.223
4	-1.970	71	0.930			0.122
y	-2.791	40	0.825			0.058
1	-3.668	273	0.601			0.025
5	-3.827	245	0.057			0.021
t	-4.135	17	0.176			0.016
7	-23.281	145	0.048			< 0.001
z	-25.580	35	0.200			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
c	186196032	2	0.000			> 0.999
z	-62065314	326	0.837			< 0.001
a	-62065344	363	0.190			< 0.001
f	-62065374	424	0.472			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	16.512	348	0.517			> 0.999
h	-16.512	767	0.472			< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight
p	19097075	61	0.197			> 0.999
k	19097060	37	0.189			> 0.999
l	19097060	18	0.889			> 0.999
n	19097057	18	0.889			> 0.999
u	19097049	1	0.000			> 0.999
s	19097037	39	0.256			> 0.999
m	19097026	120	0.367			> 0.999
i	19097023	130	0.731			> 0.999
h	19097004	13	0.462			> 0.999
j	19096993	70	0.900			> 0.999
b	19096981	325	0.840			> 0.999
a	19096930	281	0.000			> 0.999
e	-229164295	2	0.000			< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
s	24.507	822	0.527			> 0.999
d	6.658	272	0.338			0.999
a	-31.166	21	0.810			< 0.001

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	59.495	1	0.000			> 0.999
5	15.589	53	0.283			> 0.999
b	11.168	97	0.907			> 0.999
c	10.808	150	0.893			> 0.999
8	5.246	297	0.017			0.995
6	-0.149	79	0.241			0.463
d	-0.898	131	0.924			0.289
7	-1.126	71	0.197			0.245
f	-1.621	37	1.000			0.165
9	-1.972	181	0.514			0.122
0	-24.276	1	0.000			< 0.001
1	-24.276	1	0.000			< 0.001
j	-47.988	6	1.000			< 0.001
g	NA	10	1.000			<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	43	42968325			0.486		NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	88.773	92.306	1	0.999

Run 6 (above) with expressao + pessoa + paralelismo + carta + genero + interpessoal + social + periodo is worse than Run 0 also including fxetaria, p = 1

Trying without interpessoal...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	9.684	427	0.958			> 0.999
n	-9.684	688	0.193			< 0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	46.256	543	0.996			> 0.999
2sg	-46.256	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	24.041	231	0.935			> 0.999	>5
6	23.263	17	0.000			> 0.999	>5
9	19.057	7	0.286			> 0.999	>5
3	8.368	34	0.882			> 0.999	>5
4	5.934	71	0.930			0.997	>5
y	3.504	40	0.825			0.971	>5
5	0.918	245	0.057			0.715	>5
t	0.499	17	0.176			0.622	>5
1	-9.947	273	0.601			< 0.001	>5
7	-30.017	145	0.048			< 0.001	>5
z	-45.620	35	0.200			< 0.001	>5

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
c	84992924	2	0.000			> 0.999
f	-28330964	424	0.472			< 0.001
z	-28330965	326	0.837			< 0.001
a	-28330995	363	0.190			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
m	4.962	348	0.517			0.993	>20
h	-4.962	767	0.472			0.007	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
v	30.674	273	0.659			> 0.999	>20
x	11.964	88	0.341			> 0.999	>20
k	7.467	532	0.404			0.999	>20
j	-20.449	221	0.525			< 0.001	>20
c	-29.656	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	12.643	822	0.527			> 0.999	>5
a	9.145	21	0.810			> 0.999	>5
d	-21.788	272	0.338			< 0.001	>5

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	16189179	1	0.000			> 0.999
7	16189150	71	0.197			> 0.999
5	16189142	53	0.283			> 0.999
9	16189126	181	0.514			> 0.999
6	16189120	79	0.241			> 0.999
b	16189120	97	0.907			> 0.999
c	16189118	150	0.893			> 0.999
d	16189117	131	0.924			> 0.999
8	16189111	297	0.017			> 0.999
f	16189107	37	1.000			> 0.999
g	16189089	10	1.000			> 0.999
j	16189088	6	1.000			> 0.999
0	-97134719	1	0.000			< 0.001
1	-97134747	1	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	36	12141841		0.486			1

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	74.773	77.244	1	0.999

Run 7 (above) with expressao + pessoa + paralelismo + carta + genero + fxetaria + social + periodo is worse than Run 0 also including interpersoal, p = 1

Trying without social...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	6.828	427	0.958			0.999
n	-6.828	688	0.193			0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	38.896	543	0.996			> 0.999
2sg	-38.896	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	34.135	17	0.000			> 0.999
2	22.835	231	0.935			> 0.999
9	18.294	7	0.286			> 0.999
4	1.436	71	0.930			0.808
y	-2.969	40	0.825			0.049
5	-4.974	245	0.057			0.007
t	-5.857	17	0.176			0.003
3	-6.886	34	0.882			0.001
1	-11.168	273	0.601			< 0.001
7	-21.279	145	0.048			< 0.001
z	-23.567	35	0.200			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
c	54669576	2	0.000			> 0.999
z	-18223184	326	0.837			< 0.001
a	-18223192	363	0.190			< 0.001
f	-18223200	424	0.472			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	4.083	348	0.517			0.983
h	-4.083	767	0.472			0.017

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
x	13.815	88	0.341			> 0.999
v	13.366	273	0.659			> 0.999
k	10.109	532	0.404			> 0.999
j	-11.922	221	0.525			< 0.001
c	-25.368	1	1.000			< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor	weight
u	5607157	1	0.000			> 0.999
l	5607148	18	0.889			> 0.999
h	5607147	13	0.462			> 0.999
n	5607144	18	0.889			> 0.999
s	5607142	39	0.256			> 0.999
p	5607134	61	0.197			> 0.999

i	5607125	130	0.731		> 0.999
b	5607123	325	0.840		> 0.999
j	5607119	70	0.900		> 0.999
m	5607115	120	0.367		> 0.999
k	5607109	37	0.189		> 0.999
a	5607073	281	0.000		> 0.999
e	-67285537	2	0.000		< 0.001

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	42.281	1	0.000			> 0.999
7	26.679	71	0.197			> 0.999
8	25.428	297	0.017			> 0.999
5	21.366	53	0.283			> 0.999
b	15.187	97	0.907			> 0.999
c	13.208	150	0.893			> 0.999
9	12.872	181	0.514			> 0.999
f	2.748	37	1.000			0.94
d	0.917	131	0.924			0.714
6	0.375	79	0.241			0.593
j	-26.934	6	1.000			< 0.001
0	-56.048	1	0.000			< 0.001
1	-78.079	1	0.000			< 0.001
g	NA	10	1.000			<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	45	12616056			0.486		NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	92.773	96.645	1	0.999

Run 8 (above) with expressao + pessoa + paralelismo + carta + genero + fxetaria + interpessoal + periodo is worse than Run 0 also including social, p = 1

Trying without periodo...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	5.683	427	0.958			0.997
n	-5.683	688	0.193			0.003

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	44.203	543	0.996			> 0.999
2sg	-44.203	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	42.366	17	0.000			> 0.999
2	22.638	231	0.935			> 0.999
9	18.798	7	0.286			> 0.999

4	-0.815	71	0.930		0.307
y	-5.279	40	0.825		0.005
5	-7.102	245	0.057		0.001
1	-7.708	273	0.601	<	0.001
t	-8.193	17	0.176	<	0.001
3	-8.622	34	0.882	<	0.001
7	-17.449	145	0.048	<	0.001
z	-28.633	35	0.200	<	0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
c	142129134	2	0.000		> 0.999
a	-47376372	363	0.190		< 0.001
f	-47376380	424	0.472		< 0.001
z	-47376382	326	0.837		< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight
h	2.121	767	0.472		0.893
m	-2.121	348	0.517		0.107

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight
x	23.639	88	0.341		> 0.999
k	6.319	532	0.404		0.998
v	2.390	273	0.659		0.916
j	-12.263	221	0.525		< 0.001
c	-20.084	1	1.000		< 0.001

\$interpessoal

factor	logodds	tokens	v/v+t	centered	factor weight
u	14577377	1	0.000		> 0.999
p	14577365	61	0.197		> 0.999
h	14577364	13	0.462		> 0.999
l	14577358	18	0.889		> 0.999
s	14577357	39	0.256		> 0.999
n	14577353	18	0.889		> 0.999
b	14577343	325	0.840		> 0.999
j	14577335	70	0.900		> 0.999
i	14577332	130	0.731		> 0.999
m	14577331	120	0.367		> 0.999
k	14577327	37	0.189		> 0.999
a	14577296	281	0.000		> 0.999
e	-174928136	2	0.000		< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor weight
s	3.618	822	0.527		0.974
a	-3.618	21	0.810		0.026
d	NA	272	0.338		<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	34	32799033		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	70.773	72.976	1	0.999

Run 9 (above) with expressao + pessoa + paralelismo + carta + genero + fxetaria + interpessoal + social is worse than Run 0 also including periodo, p = 1

Dropping interpessoal...

STEP 2 - dropping from Run 7 model with expressao + pessoa + paralelismo + carta + genero + fxetaria + social + periodo

Trying without expressao...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor weight	vif
3sg	59.962	543	0.996		> 0.999	>20
2sg	-59.962	572	0.002		< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor weight
6	35.470	17	0.000		> 0.999
2	34.624	231	0.935		> 0.999
9	29.840	7	0.286		> 0.999
4	-3.480	71	0.930		0.03
3	-4.780	34	0.882		0.008
1	-5.238	273	0.601		0.005
y	-5.357	40	0.825		0.005
5	-5.875	245	0.057		0.003
t	-7.238	17	0.176		0.001
7	-24.533	145	0.048		< 0.001
z	-43.433	35	0.200		< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
c	192116606	2	0.000		> 0.999
z	-64038858	326	0.837		< 0.001
f	-64038860	424	0.472		< 0.001
a	-64038887	363	0.190		< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight	vif
m	2.254	348	0.517		0.905	>20
h	-2.254	767	0.472		0.095	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight	vif
v	38.826	273	0.659		> 0.999	>20
x	25.610	88	0.341		> 0.999	>20
k	6.237	532	0.404		0.998	>20

j	-30.637	221	0.525		< 0.001	>20
c	-40.036	1	1.000		< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
a	18.487	21	0.810			> 0.999	>5
s	9.965	822	0.527			> 0.999	>5
d	-28.453	272	0.338			< 0.001	>5

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
4	36593691	1	0.000			> 0.999
5	36593666	53	0.283			> 0.999
7	36593648	71	0.197			> 0.999
9	36593645	181	0.514			> 0.999
b	36593629	97	0.907			> 0.999
8	36593628	297	0.017			> 0.999
c	36593627	150	0.893			> 0.999
6	36593627	79	0.241			> 0.999
f	36593624	37	1.000			> 0.999
d	36593620	131	0.924			> 0.999
j	36593596	6	1.000			> 0.999
g	36593593	10	1.000			> 0.999
0	-219561779	1	0.000			< 0.001
1	-219561816	1	0.000			< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	35	27445222			0.486		1

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	72.773	75.108	1	1

Run 10 (above) with pessoa + paralelismo + carta + genero + fxetaria + social + periodo is worse than Run 7 also including expressao, p = 1

Trying without pessoa...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	1.095899e+15	427	0.958			> 0.999
n	-1.095899e+15	688	0.193			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	1.742809e+15	231	0.935			> 0.999
4	1.694915e+15	71	0.930			> 0.999
3	9.960930e+14	34	0.882			> 0.999
y	9.572709e+14	40	0.825			> 0.999
5	8.138453e+14	245	0.057			> 0.999
1	2.863594e+14	273	0.601			> 0.999
z	9.761715e+13	35	0.200			> 0.999
9	-6.656083e+14	7	0.286			< 0.001

t	-8.443120e+14	17	0.176		< 0.001
7	-2.096106e+15	145	0.048		< 0.001
6	-2.982884e+15	17	0.000		< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
z	5.720321e+15	326	0.837		> 0.999
f	4.839156e+15	424	0.472		> 0.999
a	4.664316e+15	363	0.190		> 0.999
c	-1.522379e+16	2	0.000		< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight
m	3.724288e+14	348	0.517		> 0.999
h	-3.724288e+14	767	0.472		< 0.001

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight
c	7.494574e+14	1	1.000		> 0.999
k	1.451416e+14	532	0.404		> 0.999
x	-1.831476e+14	88	0.341		< 0.001
j	-2.893836e+14	221	0.525		< 0.001
v	-4.220678e+14	273	0.659		< 0.001

\$social

factor	logodds	tokens	v/v+t	centered	factor weight
a	9.863636e+14	21	0.810		> 0.999
s	-2.679417e+14	822	0.527		< 0.001
d	-7.184219e+14	272	0.338		< 0.001

\$periodo

factor	logodds	tokens	v/v+t	centered	factor weight
0	1.293470e+16	1	0.000		> 0.999
1	1.250018e+16	1	0.000		> 0.999
j	1.173949e+15	6	1.000		> 0.999
b	-1.351690e+15	97	0.907		< 0.001
d	-1.824284e+15	131	0.924		< 0.001
c	-1.974006e+15	150	0.893		< 0.001
f	-2.150337e+15	37	1.000		< 0.001
8	-2.256431e+15	297	0.017		< 0.001
7	-2.285521e+15	71	0.197		< 0.001
9	-2.495281e+15	181	0.514		< 0.001
5	-2.623698e+15	53	0.283		< 0.001
6	-3.264188e+15	79	0.241		< 0.001
4	-6.383390e+15	1	0.000		< 0.001
g	NA	10	1.000		<NA>

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	34	-2.474377e+15		0.486			NA

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1838.226	3744.453	3746.656	0.908	1

Run 11 (above) with `expressao + paralelismo + carta + genero + fxetaria + social + periodo` is worse than Run 7 also including `pessoa`, $p = 0$

Trying without `paralelismo`...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	1.132	427	0.958			0.756
n	-1.132	688	0.193			0.244

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	112.424	543	0.996			> 0.999	>20
2sg	-112.424	572	0.002			< 0.001	>20

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
z	3.573750e+13	326	0.837			> 0.999
f	3.573750e+13	424	0.472			> 0.999
a	3.573750e+13	363	0.190			> 0.999
c	-1.072125e+14	2	0.000			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
m	27.219	348	0.517			> 0.999	>20
h	-27.219	767	0.472			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	106.707	88	0.341			> 0.999	>20
v	53.025	273	0.659			> 0.999	>20
k	10.946	532	0.404			> 0.999	>20
j	-61.443	221	0.525			< 0.001	>20
c	-109.236	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	33.725	822	0.527			> 0.999	>20
a	5.351	21	0.810			0.995	>20
d	-39.076	272	0.338			< 0.001	>20

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight
0	1.225286e+14	1	0.000			> 0.999
1	1.225286e+14	1	0.000			> 0.999
4	-2.042143e+13	1	0.000			< 0.001
5	-2.042143e+13	53	0.283			< 0.001
6	-2.042143e+13	79	0.241			< 0.001
b	-2.042143e+13	97	0.907			< 0.001
8	-2.042143e+13	297	0.017			< 0.001
7	-2.042143e+13	71	0.197			< 0.001
d	-2.042143e+13	131	0.924			< 0.001
9	-2.042143e+13	181	0.514			< 0.001

f	-2.042143e+13	37	1.000	< 0.001
c	-2.042143e+13	150	0.893	< 0.001
j	-2.042143e+13	6	1.000	< 0.001
g	-2.042143e+13	10	1.000	< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	26	-1.531607e+13		0.486			0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-5.888	63.775	65.066	1	1

Run 12 (above) with expressao + pessoa + carta + genero + fxetaria + social + periodo is worse than Run 7 also including paralelismo, p = 0.532

Trying without carta...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
p	9.308	427	0.958			> 0.999	>15
n	-9.308	688	0.193			< 0.001	>15

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	59.76	543	0.996			> 0.999	>20
2sg	-59.76	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
9	38.255	7	0.286			> 0.999	>20
6	32.149	17	0.000			> 0.999	>20
2	30.267	231	0.935			> 0.999	>20
4	9.071	71	0.930			> 0.999	>20
y	7.762	40	0.825			> 0.999	>20
5	6.782	245	0.057			0.999	>20
t	6.610	17	0.176			0.999	>20
1	-6.748	273	0.601			0.001	>20
3	-9.547	34	0.882			< 0.001	>20
7	-47.483	145	0.048			< 0.001	>20
z	-67.119	35	0.200			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
m	1.051	348	0.517			0.741	>20
h	-1.051	767	0.472			0.259	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
v	34.236	273	0.659			> 0.999	>20
x	13.934	88	0.341			> 0.999	>20
k	12.562	532	0.404			> 0.999	>20
j	-25.536	221	0.525			< 0.001	>20
c	-35.195	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
a	17.361	21	0.810			> 0.999	>15
s	10.171	822	0.527			> 0.999	>15
d	-27.532	272	0.338			< 0.001	>15

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
4	59.321	1	0.000			> 0.999	>20
0	57.218	1	0.000			> 0.999	>20
7	33.530	71	0.197			> 0.999	>20
1	19.120	1	0.000			> 0.999	>20
5	17.292	53	0.283			> 0.999	>20
9	15.311	181	0.514			> 0.999	>20
6	-2.677	79	0.241			0.064	>20
b	-18.127	97	0.907			< 0.001	>20
c	-19.022	150	0.893			< 0.001	>20
f	-23.608	37	1.000			< 0.001	>20
d	-25.178	131	0.924			< 0.001	>20
8	-25.696	297	0.017			< 0.001	>20
j	-41.979	6	1.000			< 0.001	>20
g	-45.505	10	1.000			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	33	-0.844			0.486		0.301

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	68.773	70.848	1	1

Run 13 (above) with expressao + pessoa + paralelismo + genero + fxetaria + social + periodo is worse than Run 7 also including carta, p = 1

Trying without genero...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	9.529	427	0.958			> 0.999
n	-9.529	688	0.193			< 0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	47.639	543	0.996			> 0.999	>7.5
2sg	-47.639	572	0.002			< 0.001	>7.5

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	26.136	17	0.000			> 0.999
2	24.515	231	0.935			> 0.999
9	19.966	7	0.286			> 0.999
3	5.991	34	0.882			0.998
4	5.904	71	0.930			0.997

y	3.538	40	0.825		0.972
5	-2.269	245	0.057		0.094
t	-3.507	17	0.176		0.029
1	-10.616	273	0.601		< 0.001
7	-22.713	145	0.048		< 0.001
z	-46.946	35	0.200		< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor weight
c	72736789	2	0.000		> 0.999
f	-24245587	424	0.472		< 0.001
z	-24245588	326	0.837		< 0.001
a	-24245614	363	0.190		< 0.001

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight	vif
v	37.696	273	0.659		> 0.999	>20
x	19.135	88	0.341		> 0.999	>20
k	6.042	532	0.404		0.998	>20
j	-24.225	221	0.525		< 0.001	>20
c	-38.648	1	1.000		< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor weight	vif
a	20.118	21	0.810		> 0.999	>20
s	7.682	822	0.527		> 0.999	>20
d	-27.800	272	0.338		< 0.001	>20

\$periodo

factor	logodds	tokens	v/v+t	centered	factor weight
4	13854666	1	0.000		> 0.999
5	13854643	53	0.283		> 0.999
7	13854641	71	0.197		> 0.999
9	13854627	181	0.514		> 0.999
b	13854619	97	0.907		> 0.999
6	13854619	79	0.241		> 0.999
c	13854617	150	0.893		> 0.999
f	13854615	37	1.000		> 0.999
d	13854613	131	0.924		> 0.999
8	13854611	297	0.017		> 0.999
g	13854589	10	1.000		> 0.999
j	13854589	6	1.000		> 0.999
0	-83127710	1	0.000		< 0.001
1	-83127740	1	0.000		< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	35	10390966		0.486			1

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	72.773	75.108	1	0.999

Run 14 (above) with `expressao + pessoa + paralelismo + carta + fxetaria + social + periodo` is worse than Run 7 also including `genero`, $p = 1$

Trying without `fxetaria`...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	10.492	427	0.958			> 0.999
n	-10.492	688	0.193			< 0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	78.709	543	0.996			> 0.999
2sg	-78.709	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	53.382	231	0.935			> 0.999
6	51.780	17	0.000			> 0.999
9	48.883	7	0.286			> 0.999
t	33.040	17	0.176			> 0.999
5	31.850	245	0.057			> 0.999
3	-15.272	34	0.882			< 0.001
y	-15.950	40	0.825			< 0.001
4	-17.620	71	0.930			< 0.001
1	-18.883	273	0.601			< 0.001
z	-67.773	35	0.200			< 0.001
7	-83.438	145	0.048			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
z	32682625	326	0.837			> 0.999
f	32682603	424	0.472			> 0.999
a	32682574	363	0.190			> 0.999
c	-98047803	2	0.000			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
m	27.196	348	0.517			> 0.999	>7.5
h	-27.196	767	0.472			< 0.001	>7.5

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	27.852	822	0.527			> 0.999	>2.5
d	10.613	272	0.338			> 0.999	>2.5
a	-38.465	21	0.810			< 0.001	>2.5

\$periodo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
0	112054723	1	0.000			> 0.999	>20
1	112054723	1	0.000			> 0.999	>20
4	-18675629	1	0.000			< 0.001	>20
7	-18675742	71	0.197			< 0.001	>20
5	-18675768	53	0.283			< 0.001	>20

b	-18675782	97	0.907	< 0.001	>20
d	-18675797	131	0.924	< 0.001	>20
9	-18675798	181	0.514	< 0.001	>20
6	-18675799	79	0.241	< 0.001	>20
c	-18675808	150	0.893	< 0.001	>20
g	-18675811	10	1.000	< 0.001	>20
f	-18675825	37	1.000	< 0.001	>20
8	-18675827	297	0.017	< 0.001	>20
j	-18675859	6	1.000	< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	32	-14006827		0.486			0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	66.773	68.725	1	1

Run 15 (above) with expressao + pessoa + paralelismo + carta + genero + social + periodo is worse than Run 7 also including fxetaria, p = 1

Trying without social...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	18.448	427	0.958			> 0.999
n	-18.448	688	0.193			< 0.001

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight
3sg	61.858	543	0.996			> 0.999
2sg	-61.858	572	0.002			< 0.001

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
6	31.883	17	0.000			> 0.999
9	26.737	7	0.286			> 0.999
2	18.520	231	0.935			> 0.999
y	13.357	40	0.825			> 0.999
4	12.772	71	0.930			> 0.999
t	8.734	17	0.176			> 0.999
3	4.988	34	0.882			0.993
1	-10.881	273	0.601			< 0.001
7	-18.835	145	0.048			< 0.001
5	-20.079	245	0.057			< 0.001
z	-67.195	35	0.200			< 0.001

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight
c	105056027	2	0.000			> 0.999
z	-35018658	326	0.837			< 0.001
f	-35018660	424	0.472			< 0.001
a	-35018708	363	0.190			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor weight
m	4.398	348	0.517		0.988
h	-4.398	767	0.472		0.012

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor weight	vif
k	33.321	532	0.404		> 0.999	>20
v	23.380	273	0.659		> 0.999	>20
x	17.911	88	0.341		> 0.999	>20
j	-29.728	221	0.525		< 0.001	>20
c	-44.885	1	1.000		< 0.001	>20

\$periodo

factor	logodds	tokens	v/v+t	centered	factor weight
4	20010754	1	0.000		> 0.999
5	20010688	53	0.283		> 0.999
7	20010684	71	0.197		> 0.999
d	20010684	131	0.924		> 0.999
9	20010676	181	0.514		> 0.999
6	20010653	79	0.241		> 0.999
b	20010653	97	0.907		> 0.999
c	20010653	150	0.893		> 0.999
8	20010647	297	0.017		> 0.999
f	20010640	37	1.000		> 0.999
g	20010625	10	1.000		> 0.999
j	20010591	6	1.000		> 0.999
0	-120063942	1	0.000		< 0.001
1	-120064005	1	0.000		< 0.001

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	34	15008005			0.486		1

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	70.773	72.976	1	1

Run 16 (above) with expressao + pessoa + paralelismo + carta + genero + fxetaria + periodo is worse than Run 7 also including social, p = 1

Trying without periodo...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor weight	vif
p	4.054	427	0.958		0.983	>10
n	-4.054	688	0.193		0.017	>10

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor weight	vif
3sg	69.738	543	0.996		> 0.999	>20
2sg	-69.738	572	0.002		< 0.001	>20

\$paralelismo

```

factor logodds tokens v/v+t centered factor weight vif
  6  40.242    17 0.000                > 0.999 >20
  2  35.989   231 0.935                > 0.999 >20
  9  30.764    7 0.286                > 0.999 >20
  3   3.078   34 0.882                0.956 >20
  y   3.051   40 0.825                0.955 >20
  5   1.381  245 0.057                0.799 >20
  1  -0.183  273 0.601                0.454 >20
  t  -1.348   17 0.176                0.206 >20
  4 -20.337   71 0.930                < 0.001 >20
  7 -32.862  145 0.048                < 0.001 >20
  z -59.774   35 0.200                < 0.001 >20

$carta
factor logodds tokens v/v+t centered factor weight vif
  c  46.871     2 0.000                > 0.999 >20
  f  -0.730   424 0.472                0.325 >20
  z  -5.158   326 0.837                0.006 >20
  a -40.983   363 0.190                < 0.001 >20

$genero
factor logodds tokens v/v+t centered factor weight vif
  h   2.715   767 0.472                0.938 >20
  m  -2.715   348 0.517                0.062 >20

$fxetaria
factor logodds tokens v/v+t centered factor weight vif
  x  38.963    88 0.341                > 0.999 >20
  v  36.629   273 0.659                > 0.999 >20
  k   0.753   532 0.404                0.68 >20
  j -29.378   221 0.525                < 0.001 >20
  c -46.967     1 1.000                < 0.001 >20

$social
factor logodds tokens v/v+t centered factor weight vif
  s  15.332   822 0.527                > 0.999 >15
  a   5.327    21 0.810                0.995 >15
  d -20.659   272 0.338                < 0.001 >15

$misc.1
  n df intercept overall proportion centered input prob
1115 23    -1.957                0.486                0.124

$misc.2
log.likelihood    AIC    AICc Dxy R2
    -1.386 48.773 49.785    1    1

Run 17 (above) with expressao + pessoa + paralelismo + carta + genero +
fxetaria + social is worse than Run 7 also including periodo, p = 1

Dropping periodo...

STEP 3 - dropping from Run 17 model with expressao + pessoa + paralelismo
+ carta + genero + fxetaria + social

```

Trying without expressao...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	78.265	543	0.996			> 0.999	>20
2sg	-78.265	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
6	50.676	17	0.000			> 0.999	>20
2	40.005	231	0.935			> 0.999	>20
9	34.086	7	0.286			> 0.999	>20
3	3.252	34	0.882			0.963	>20
y	1.292	40	0.825			0.784	>20
5	0.795	245	0.057			0.689	>20
1	0.276	273	0.601			0.569	>20
t	-2.220	17	0.176			0.098	>20
4	-30.756	71	0.930			< 0.001	>20
7	-38.321	145	0.048			< 0.001	>20
z	-59.083	35	0.200			< 0.001	>20

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
c	39.475	2	0.000			> 0.999	>20
f	0.465	424	0.472			0.614	>20
z	-0.645	326	0.837			0.344	>20
a	-39.296	363	0.190			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
h	0.192	767	0.472			0.548	>7.5
m	-0.192	348	0.517			0.452	>7.5

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	46.947	88	0.341			> 0.999	>20
v	32.808	273	0.659			> 0.999	>20
k	7.386	532	0.404			0.999	>20
j	-29.272	221	0.525			< 0.001	>20
c	-57.869	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	13.935	822	0.527			> 0.999	>10
a	10.122	21	0.810			> 0.999	>10
d	-24.057	272	0.338			< 0.001	>10

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	22	-4.501			0.486		0.011

\$misc.2

```
log.likelihood    AIC    AICc Dxy R2
      -1.386 46.773 47.699  1  1
```

Run 18 (above) with pessoa + paralelismo + carta + genero + fxetaria + social is worse than Run 17 also including expressao, p = 1

Trying without pessoa...

\$expressao

```
factor logodds tokens v/v+t centered factor weight
  p    3.201    427 0.958                0.961
  n   -3.201    688 0.193                0.039
```

\$paralelismo

```
factor logodds tokens v/v+t centered factor weight vif
  2    4.809    231 0.935                0.992 >2.5
  3    4.580     34 0.882                0.99 >2.5
  4    3.716     71 0.930                0.976 >2.5
  y    3.350     40 0.825                0.966 >2.5
  9    1.818      7 0.286                0.86 >2.5
  1   -0.042    273 0.601                0.489 >2.5
  t   -0.181     17 0.176                0.455 >2.5
  5   -0.833    245 0.057                0.303 >2.5
  z   -1.123     35 0.200                0.245 >2.5
  7   -1.533    145 0.048                0.178 >2.5
  6  -14.561     17 0.000                < 0.001 >2.5
```

\$carta

```
factor logodds tokens v/v+t centered factor weight
  z    6.477    326 0.837                0.998
  f    4.506    424 0.472                0.989
  a    1.766    363 0.190                0.854
  c  -12.749      2 0.000                < 0.001
```

\$genero

```
factor logodds tokens v/v+t centered factor weight
  m    0.739    348 0.517                0.677
  h   -0.739    767 0.472                0.323
```

\$fxetaria

```
factor logodds tokens v/v+t centered factor weight
  c    8.906      1 1.000                > 0.999
  v   -1.036    273 0.659                0.262
  k   -2.537    532 0.404                0.073
  j   -2.596    221 0.525                0.069
  x   -2.738     88 0.341                0.061
```

\$social

```
factor logodds tokens v/v+t centered factor weight
  a    2.003     21 0.810                0.881
  s    0.197    822 0.527                0.549
  d   -2.200    272 0.338                0.1
```

\$misc.1

	n	df	intercept	overall	proportion	centered	input	prob
	1115	22	-1.748			0.486		0.148

\$misc.2

	log.likelihood	AIC	AICc	Dxy	R2
	-125.462	294.924	295.851	0.984	0.927

Run 19 (above) with expressao + paralelismo + carta + genero + fxetaria + social is worse than Run 17 also including pessoa, $p = 6.57e-56$

Trying without paralelismo...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	1.123	427	0.958			0.755
n	-1.123	688	0.193			0.245

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.843	543	0.996			> 0.999	>20
2sg	-23.843	572	0.002			< 0.001	>20

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
z	2.681	326	0.837			0.936	>2.5
c	0.484	2	0.000			0.619	>2.5
f	0.173	424	0.472			0.543	>2.5
a	-3.337	363	0.190			0.034	>2.5

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.029	348	0.517			0.507
h	-0.029	767	0.472			0.493

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	20.482	88	0.341			> 0.999	>20
v	1.624	273	0.659			0.835	>20
k	0.141	532	0.404			0.535	>20
c	-1.268	1	1.000			0.22	>20
j	-20.979	221	0.525			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	0.756	21	0.810			0.68
s	0.553	822	0.527			0.635
d	-1.309	272	0.338			0.213

\$misc.1

	n	df	intercept	overall	proportion	centered	input	prob
	1115	13	-0.089			0.486		0.478

\$misc.2

```
log.likelihood    AIC    AICc Dxy    R2
      -10.56 47.121 47.452    1 0.996
```

Run 20 (above) with expressao + pessoa + carta + genero + fxetaria + social is worse than Run 17 also including paralelismo, p = 0.0494

Trying without carta...

\$expressao

```
factor logodds tokens v/v+t centered factor weight vif
  n    0.583     688 0.193                0.642 >15
  p   -0.583     427 0.958                0.358 >15
```

\$pessoa

```
factor  logodds tokens v/v+t centered factor weight vif
  3sg  121.685     543 0.996                > 0.999 >20
  2sg  -121.685     572 0.002                < 0.001 >20
```

\$paralelismo

```
factor logodds tokens v/v+t centered factor weight vif
  6   59.478      17 0.000                > 0.999 >20
  9   58.672       7 0.286                > 0.999 >20
  2   53.198     231 0.935                > 0.999 >20
  y   14.090      40 0.825                > 0.999 >20
  5   13.535     245 0.057                > 0.999 >20
  t   10.628      17 0.176                > 0.999 >20
  4  -20.117      71 0.930                < 0.001 >20
  1  -22.352     273 0.601                < 0.001 >20
  3  -25.361      34 0.882                < 0.001 >20
  7  -62.392     145 0.048                < 0.001 >20
  z  -79.379      35 0.200                < 0.001 >20
```

\$genero

```
factor logodds tokens v/v+t centered factor weight vif
  h   17.617      767 0.472                > 0.999 >10
  m  -17.617      348 0.517                < 0.001 >10
```

\$fxetaria

```
factor logodds tokens v/v+t centered factor weight vif
  x   96.778      88 0.341                > 0.999 >20
  v   23.269     273 0.659                > 0.999 >20
  k   10.123     532 0.404                > 0.999 >20
  c  -62.906       1 1.000                < 0.001 >20
  j  -67.264     221 0.525                < 0.001 >20
```

\$social

```
factor logodds tokens v/v+t centered factor weight vif
  a   24.477      21 0.810                > 0.999 >10
  s   24.061     822 0.527                > 0.999 >10
  d  -48.538     272 0.338                < 0.001 >10
```

\$misc.1

```
  n df intercept overall proportion centered input prob
1115 20   -16.138                0.486                0
```

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	42.773	43.54	1	1

Run 21 (above) with expressao + pessoa + paralelismo + genero + fxetaria + social is worse than Run 17 also including carta, p = 1

Trying without genero...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
p	3.106	427	0.958			0.957	>5
n	-3.106	688	0.193			0.043	>5

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	70.974	543	0.996			> 0.999	>20
2sg	-70.974	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	36.786	231	0.935			> 0.999	>20
6	35.316	17	0.000			> 0.999	>20
9	30.946	7	0.286			> 0.999	>20
3	6.149	34	0.882			0.998	>20
1	2.957	273	0.601			0.951	>20
y	2.671	40	0.825			0.935	>20
5	2.319	245	0.057			0.91	>20
t	-0.262	17	0.176			0.435	>20
4	-21.889	71	0.930			< 0.001	>20
7	-35.000	145	0.048			< 0.001	>20
z	-59.992	35	0.200			< 0.001	>20

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
c	45.104	2	0.000			> 0.999	>15
f	1.820	424	0.472			0.861	>15
z	-5.025	326	0.837			0.007	>15
a	-41.900	363	0.190			< 0.001	>15

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	37.287	88	0.341			> 0.999	>20
v	36.915	273	0.659			> 0.999	>20
k	2.761	532	0.404			0.941	>20
j	-27.108	221	0.525			< 0.001	>20
c	-49.855	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	15.860	822	0.527			> 0.999	>2.5
a	2.225	21	0.810			0.902	>2.5
d	-18.085	272	0.338			< 0.001	>2.5

```
$misc.1
  n df intercept overall proportion centered input prob
1115 22      -4.661                0.486                0.009
```

```
$misc.2
log.likelihood      AIC      AICc Dxy R2
      -1.386 46.773 47.699   1   1
```

Run 22 (above) with expressao + pessoa + paralelismo + carta + fxetaria + social is worse than Run 17 also including genero, p = 1

Trying without fxetaria...

```
$expressao
factor logodds tokens v/v+t centered factor weight
  p    0.554    427 0.958                0.635
  n   -0.554    688 0.193                0.365
```

```
$pessoa
factor logodds tokens v/v+t centered factor weight vif
 3sg  57.775    543 0.996                > 0.999 >20
 2sg -57.775    572 0.002                < 0.001 >20
```

```
$paralelismo
factor logodds tokens v/v+t centered factor weight vif
  2  32.858    231 0.935                > 0.999 >20
  3  12.316     34 0.882                > 0.999 >20
  6  11.417     17 0.000                > 0.999 >20
  y  11.286     40 0.825                > 0.999 >20
  1  10.977    273 0.601                > 0.999 >20
  5  10.699    245 0.057                > 0.999 >20
  t   9.998     17 0.176                > 0.999 >20
  9   9.227     7 0.286                > 0.999 >20
  4 -19.297     71 0.930                < 0.001 >20
  z -44.500     35 0.200                < 0.001 >20
  7 -44.980    145 0.048                < 0.001 >20
```

```
$carta
factor logodds tokens v/v+t centered factor weight vif
  c  31.518     2 0.000                > 0.999 >20
  z  13.476    326 0.837                > 0.999 >20
  f  -3.920    424 0.472                0.019 >20
  a -41.075    363 0.190                < 0.001 >20
```

```
$genero
factor logodds tokens v/v+t centered factor weight vif
  m  17.784    348 0.517                > 0.999 >20
  h -17.784    767 0.472                < 0.001 >20
```

```
$social
factor logodds tokens v/v+t centered factor weight vif
  s  13.021    822 0.527                > 0.999 >2.5
  d  12.797    272 0.338                > 0.999 >2.5
```

a -25.818 21 0.810 < 0.001 >2.5

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	19	-2.698		0.486			0.063

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-6.564	51.128	51.822	1	1

Run 23 (above) with expressao + pessoa + paralelismo + carta + genero + social is worse than Run 17 also including fxetaria, p = 0.0348

Trying without social...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
p	19.027	427	0.958			> 0.999	>5
n	-19.027	688	0.193			< 0.001	>5

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	86.329	543	0.996			> 0.999	>20
2sg	-86.329	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	18.244	231	0.935			> 0.999	>20
3	16.541	34	0.882			> 0.999	>20
6	16.483	17	0.000			> 0.999	>20
y	15.078	40	0.825			> 0.999	>20
4	14.806	71	0.930			> 0.999	>20
1	14.318	273	0.601			> 0.999	>20
9	12.925	7	0.286			> 0.999	>20
t	12.720	17	0.176			> 0.999	>20
5	-20.778	245	0.057			< 0.001	>20
7	-21.952	145	0.048			< 0.001	>20
z	-78.385	35	0.200			< 0.001	>20

\$carta

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
c	29.152	2	0.000			> 0.999	>20
f	25.817	424	0.472			> 0.999	>20
z	-7.174	326	0.837			0.001	>20
a	-47.795	363	0.190			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
h	0.633	767	0.472			0.653	>5
m	-0.633	348	0.517			0.347	>5

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	52.500	88	0.341			> 0.999	>20

v	51.433	273	0.659		> 0.999	>20
k	47.160	532	0.404		> 0.999	>20
j	-43.111	221	0.525		< 0.001	>20
c	-107.982	1	1.000		< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	21	-10.31			0.486		0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	44.773	45.618	1	1

Run 24 (above) with expressao + pessoa + paralelismo + carta + genero + fxetaria is worse than Run 17 also including social, p = 1

Dropping carta...

STEP 4 - dropping from Run 21 model with expressao + pessoa + paralelismo + genero + fxetaria + social

Trying without expressao...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	121.36	543	0.996			> 0.999	>20
2sg	-121.36	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
6	60.340	17	0.000			> 0.999	>20
9	59.921	7	0.286			> 0.999	>20
2	52.244	231	0.935			> 0.999	>20
y	13.644	40	0.825			> 0.999	>20
5	13.221	245	0.057			> 0.999	>20
t	10.114	17	0.176			> 0.999	>20
4	-20.033	71	0.930			< 0.001	>20
1	-22.766	273	0.601			< 0.001	>20
3	-24.474	34	0.882			< 0.001	>20
7	-62.169	145	0.048			< 0.001	>20
z	-80.042	35	0.200			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
h	17.951	767	0.472			> 0.999	>2.5
m	-17.951	348	0.517			< 0.001	>2.5

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	97.213	88	0.341			> 0.999	>20
v	24.002	273	0.659			> 0.999	>20
k	8.903	532	0.404			> 0.999	>20
c	-62.765	1	1.000			< 0.001	>20
j	-67.353	221	0.525			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
a	24.522	21	0.810			> 0.999	>10
s	23.917	822	0.527			> 0.999	>10
d	-48.439	272	0.338			< 0.001	>10

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	19	-15.834			0.486		0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	40.773	41.467	1	1

Run 25 (above) with pessoa + paralelismo + genero + fxetaria + social is worse than Run 21 also including expressao, p = 1

Trying without pessoa...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	3.061	427	0.958			0.955
n	-3.061	688	0.193			0.045

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	5.510	231	0.935			0.996	>2.5
4	4.620	71	0.930			0.99	>2.5
3	4.317	34	0.882			0.987	>2.5
y	4.132	40	0.825			0.984	>2.5
9	2.026	7	0.286			0.883	>2.5
1	0.480	273	0.601			0.618	>2.5
t	-0.537	17	0.176			0.369	>2.5
z	-0.860	35	0.200			0.297	>2.5
5	-1.021	245	0.057			0.265	>2.5
7	-1.810	145	0.048			0.141	>2.5
6	-16.858	17	0.000			< 0.001	>2.5

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.016	767	0.472			0.504
m	-0.016	348	0.517			0.496

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
c	9.885	1	1.000			> 0.999
v	-1.754	273	0.659			0.148
x	-2.574	88	0.341			0.071
j	-2.749	221	0.525			0.06
k	-2.808	532	0.404			0.057

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
--------	---------	--------	-------	----------	--------	--------

a	1.702	21	0.810	0.846
s	-0.203	822	0.527	0.449
d	-1.499	272	0.338	0.183

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	19	2.455		0.486		0.921	

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-157.292	352.585	353.279	0.969	0.905

Run 26 (above) with expressao + paralelismo + genero + fxetaria + social is worse than Run 21 also including pessoa, $p = 8.8e-70$

Trying without paralelismo...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	0.358	427	0.958			0.588
n	-0.358	688	0.193			0.412

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	22.992	543	0.996			> 0.999	>20
2sg	-22.992	572	0.002			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.493	767	0.472			0.621
m	-0.493	348	0.517			0.379

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	19.913	88	0.341			> 0.999	>20
v	0.735	273	0.659			0.676	>20
k	-0.523	532	0.404			0.372	>20
c	-0.599	1	1.000			0.354	>20
j	-19.526	221	0.525			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	1.108	21	0.810			0.752
s	0.122	822	0.527			0.53
d	-1.230	272	0.338			0.226

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	10	0.201		0.486		0.55	

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-14.24	48.479	48.678	0.999	0.995

Run 27 (above) with expressao + pessoa + genero + fxetaria + social is worse than Run 21 also including paralelismo, $p = 0.00416$

Trying without genero...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
n	17.574	688	0.193			> 0.999	>2.5
p	-17.574	427	0.958			< 0.001	>2.5

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	163.792	543	0.996			> 0.999	>20
2sg	-163.792	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	84.819	231	0.935			> 0.999	>20
6	79.681	17	0.000			> 0.999	>20
9	78.793	7	0.286			> 0.999	>20
y	12.669	40	0.825			> 0.999	>20
5	11.955	245	0.057			> 0.999	>20
t	7.811	17	0.176			> 0.999	>20
1	7.527	273	0.601			0.999	>20
4	-53.060	71	0.930			< 0.001	>20
3	-54.526	34	0.882			< 0.001	>20
z	-79.080	35	0.200			< 0.001	>20
7	-96.589	145	0.048			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	120.579	88	0.341			> 0.999	>20
k	46.919	532	0.404			> 0.999	>20
v	12.272	273	0.659			> 0.999	>20
j	-60.893	221	0.525			< 0.001	>20
c	-118.876	1	1.000			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
s	24.529	822	0.527			> 0.999	>15
a	22.471	21	0.810			> 0.999	>15
d	-46.999	272	0.338			< 0.001	>15

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	19	-30.772			0.486		0

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-1.386	40.773	41.467	1	1

Run 28 (above) with expressao + pessoa + paralelismo + fxetaria + social is worse than Run 21 also including genero, $p = 1$

Trying without fxetaria...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight
p	0.647	427	0.958			0.656
n	-0.647	688	0.193			0.344

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	22.839	543	0.996			> 0.999	>20
2sg	-22.839	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	22.422	231	0.935			> 0.999	>20
4	3.086	71	0.930			0.956	>20
3	2.983	34	0.882			0.952	>20
y	2.704	40	0.825			0.937	>20
9	2.080	7	0.286			0.889	>20
1	2.022	273	0.601			0.883	>20
t	1.449	17	0.176			0.81	>20
5	1.080	245	0.057			0.747	>20
6	0.764	17	0.000			0.682	>20
z	-19.270	35	0.200			< 0.001	>20
7	-19.320	145	0.048			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
m	0.333	348	0.517			0.582
h	-0.333	767	0.472			0.418

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
d	0.299	272	0.338			0.574
s	-0.119	822	0.527			0.47
a	-0.180	21	0.810			0.455

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	16	-1.763			0.486		0.146

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.192	50.384	50.879	1	0.997

Run 29 (above) with expressao + pessoa + paralelismo + genero + social is worse than Run 21 also including fxetaria, p = 0.00359

Trying without social...

\$expressao

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
p	9.256	427	0.958			> 0.999	>20

```

      n  -9.256    688 0.193                < 0.001 >20

$peessoa
  factor logodds tokens v/v+t centered factor weight vif
    3sg   51.439    543 0.996                > 0.999 >20
    2sg  -51.439    572 0.002                < 0.001 >20

$paralelismo
  factor logodds tokens v/v+t centered factor weight vif
     9   28.798     7 0.286                > 0.999 >20
     2   19.772   231 0.935                > 0.999 >20
     t   15.381    17 0.176                > 0.999 >20
     6   13.525    17 0.000                > 0.999 >20
     y    0.053    40 0.825                0.513 >20
     4   -1.155    71 0.930                0.24 >20
     1   -2.195   273 0.601                0.1 >20
     5   -2.387   245 0.057                0.084 >20
     3   -4.835    34 0.882                0.008 >20
     7  -24.222   145 0.048                < 0.001 >20
     z  -42.734    35 0.200                < 0.001 >20

$genero
  factor logodds tokens v/v+t centered factor weight vif
     h    9.256    767 0.472                > 0.999 >20
     m   -9.256    348 0.517                < 0.001 >20

$fxetaria
  factor logodds tokens v/v+t centered factor weight vif
     x   33.699     88 0.341                > 0.999 >20
     v   14.014   273 0.659                > 0.999 >20
     k   -1.881   532 0.404                0.132 >20
     c  -20.647     1 1.000                < 0.001 >20
     j  -25.186   221 0.525                < 0.001 >20

$misc.1
  n df intercept overall proportion centered input prob
 1115 18    -2.032                0.486                0.116

$misc.2
  log.likelihood    AIC    AICc Dxy    R2
    -4.159 44.318 44.942    1 0.999

Run 30 (above) with expressao + pessoa + paralelismo + genero + fxetaria
is worse than Run 21 also including social, p = 0.0625

Dropping expressao...

STEP 5 - dropping from Run 25 model with pessoa + paralelismo + genero +
fxetaria + social

Trying without pessoa...

```

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	4.303	231	0.935			0.987
4	3.885	71	0.930			0.98
3	3.650	34	0.882			0.975
y	3.272	40	0.825			0.963
1	1.820	273	0.601			0.861
9	0.646	7	0.286			0.656
z	0.093	35	0.200			0.523
t	-0.265	17	0.176			0.434
5	-1.138	245	0.057			0.243
7	-1.271	145	0.048			0.219
6	-14.993	17	0.000			< 0.001

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.022	767	0.472			0.505
m	-0.022	348	0.517			0.495

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
c	12.165	1	1.000			> 0.999
v	-2.012	273	0.659			0.118
x	-2.981	88	0.341			0.048
j	-3.514	221	0.525			0.029
k	-3.658	532	0.404			0.025

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	0.937	21	0.810			0.718
s	0.377	822	0.527			0.593
d	-1.314	272	0.338			0.212

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	18	1.667			0.486		0.841

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-371.401	778.801	779.425	0.853	0.764

Run 31 (above) with paralelismo + genero + fxetaria + social is worse than Run 25 also including pessoa, $p = 5.91e-163$

Trying without paralelismo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.214	543	0.996			> 0.999	>20
2sg	-23.214	572	0.002			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.458	767	0.472			0.613

```

      m  -0.458      348 0.517
                                0.387

$fxetaria
factor logodds tokens v/v+t centered factor weight vif
  x  19.886      88 0.341
                                > 0.999 >20
  v   0.753     273 0.659
                                0.68 >20
  c  -0.487       1 1.000
                                0.381 >20
  k  -0.559     532 0.404
                                0.364 >20
  j -19.594     221 0.525
                                < 0.001 >20

$social
factor logodds tokens v/v+t centered factor weight
  a   1.147       21 0.810
                                0.759
  s   0.141     822 0.527
                                0.535
  d  -1.288     272 0.338
                                0.216

$misc.1
  n df intercept overall proportion centered input prob
1115 9      0.149                0.486                0.537

$misc.2
log.likelihood      AIC      AICc      Dxy      R2
      -14.37 46.739 46.902 0.999 0.995

Run 32 (above) with pessoa + genero + fxetaria + social is worse than Run
25 also including paralelismo, p = 0.00378

Trying without genero...

$pessoa
factor logodds tokens v/v+t centered factor weight vif
  3sg  86.721     543 0.996
                                > 0.999 >20
  2sg -86.721     572 0.002
                                < 0.001 >20

$paralelismo
factor logodds tokens v/v+t centered factor weight vif
  2  45.646     231 0.935
                                > 0.999 >20
  6  41.006      17 0.000
                                > 0.999 >20
  9  40.660       7 0.286
                                > 0.999 >20
  y   7.905      40 0.825
                                > 0.999 >20
  5   7.120     245 0.057
                                0.999 >20
  t   3.611      17 0.176
                                0.974 >20
  1   2.861     273 0.601
                                0.946 >20
  4 -23.943      71 0.930
                                < 0.001 >20
  3 -25.269      34 0.882
                                < 0.001 >20
  7 -49.799     145 0.048
                                < 0.001 >20
  z -49.799      35 0.200
                                < 0.001 >20

$fxetaria
factor logodds tokens v/v+t centered factor weight vif
  x  59.196      88 0.341
                                > 0.999 >20
  v  20.442     273 0.659
                                > 0.999 >20
  k  19.634     532 0.404
                                > 0.999 >20
  j -36.898     221 0.525
                                < 0.001 >20

```

```

      c -62.374      1 1.000      < 0.001 >20

$social
factor logodds tokens v/v+t centered factor weight vif
  s  12.548     822 0.527      > 0.999 >2.5
  a  11.930      21 0.810      > 0.999 >2.5
  d -24.478     272 0.338      < 0.001 >2.5

$misc.1
  n df intercept overall proportion centered input prob
1115 18   -12.572           0.486           0

$misc.2
log.likelihood      AIC      AICc Dxy R2
      -2.773 41.545 42.169   1 1

Run 33 (above) with pessoa + paralelismo + fxetaria + social is worse than
Run 25 also including genero, p = 0.0959

Trying without fxetaria...

$pessoa
factor logodds tokens v/v+t centered factor weight vif
  3sg  23.334     543 0.996      > 0.999 >20
  2sg -23.334     572 0.002      < 0.001 >20

$paralelismo
factor logodds tokens v/v+t centered factor weight vif
  2  22.677     231 0.935      > 0.999 >20
  4   2.993      71 0.930      0.952 >20
  3   2.865      34 0.882      0.946 >20
  y   2.507      40 0.825      0.925 >20
  1   2.238     273 0.601      0.904 >20
  9   1.898       7 0.286      0.87 >20
  t   1.433      17 0.176      0.807 >20
  5   1.150     245 0.057      0.759 >20
  6   0.821      17 0.000      0.694 >20
  7 -19.200     145 0.048      < 0.001 >20
  z -19.380      35 0.200      < 0.001 >20

$genero
factor logodds tokens v/v+t centered factor weight
  m   0.448     348 0.517           0.61
  h  -0.448     767 0.472           0.39

$social
factor logodds tokens v/v+t centered factor weight
  d   0.208     272 0.338           0.552
  s   0.096     822 0.527           0.524
  a  -0.304      21 0.810           0.425

$misc.1
  n df intercept overall proportion centered input prob
1115 15   -2.048           0.486           0.114

```

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.54	49.081	49.517	1	0.997

Run 34 (above) with pessoa + paralelismo + genero + social is worse than Run 25 also including fxetaria, $p = 0.00263$

Trying without social...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.543	543	0.996			> 0.999	>20
2sg	-44.543	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.076	231	0.935			> 0.999	>20
6	16.752	17	0.000			> 0.999	>20
9	16.251	7	0.286			> 0.999	>20
4	-0.099	71	0.930			0.475	>20
3	-0.175	34	0.882			0.456	>20
y	-0.766	40	0.825			0.317	>20
1	-0.890	273	0.601			0.291	>20
t	-1.218	17	0.176			0.228	>20
5	-2.006	245	0.057			0.119	>20
7	-24.311	145	0.048			< 0.001	>20
z	-24.613	35	0.200			< 0.001	>20

\$genero

factor	logodds	tokens	v/v+t	centered	factor	weight
h	0.302	767	0.472			0.575
m	-0.302	348	0.517			0.425

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.095	88	0.341			> 0.999	>20
v	3.934	273	0.659			0.981	>20
k	3.521	532	0.404			0.971	>20
c	-14.551	1	1.000			< 0.001	>20
j	-17.999	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	17	-2.233			0.486		0.097

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.632	43.263	43.821	1	0.999

Run 35 (above) with pessoa + paralelismo + genero + fxetaria is worse than Run 25 also including social, $p = 0.039$

Dropping genero...

STEP 6 - dropping from Run 33 model with pessoa + paralelismo + fxetaria + social

Trying without pessoa...

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	4.303	231	0.935			0.987
4	3.886	71	0.930			0.98
3	3.644	34	0.882			0.975
y	3.272	40	0.825			0.963
1	1.821	273	0.601			0.861
9	0.650	7	0.286			0.657
z	0.094	35	0.200			0.523
t	-0.264	17	0.176			0.434
5	-1.138	245	0.057			0.243
7	-1.270	145	0.048			0.219
6	-14.998	17	0.000			< 0.001

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
c	12.162	1	1.000			> 0.999
v	-2.015	273	0.659			0.118
x	-2.992	88	0.341			0.048
j	-3.509	221	0.525			0.029
k	-3.646	532	0.404			0.025

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	0.922	21	0.810			0.715
s	0.387	822	0.527			0.596
d	-1.309	272	0.338			0.213

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	17	1.661			0.486		0.84

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-371.419	776.837	777.395	0.853	0.764

Run 36 (above) with paralelismo + fxetaria + social is worse than Run 33 also including pessoa, $p = 2.33e-162$

Trying without paralelismo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.104	543	0.996			> 0.999	>20
2sg	-23.104	572	0.002			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
--------	---------	--------	-------	----------	--------	--------	-----

x	19.610	88	0.341		> 0.999	>20
v	0.644	273	0.659		0.656	>20
k	-0.270	532	0.404		0.433	>20
c	-0.750	1	1.000		0.321	>20
j	-19.234	221	0.525		< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
a	1.050	21	0.810			0.741
s	0.159	822	0.527			0.54
d	-1.210	272	0.338			0.23

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	8	0.161			0.486		0.54

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-14.596	45.191	45.321	0.999	0.995

Run 37 (above) with pessoa + fxetaria + social is worse than Run 33 also including paralelismo, $p = 0.0086$

Trying without fxetaria...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.269	543	0.996			> 0.999	>20
2sg	-23.269	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	22.603	231	0.935			> 0.999	>20
4	2.979	71	0.930			0.952	>20
3	2.823	34	0.882			0.944	>20
y	2.647	40	0.825			0.934	>20
1	2.229	273	0.601			0.903	>20
9	1.758	7	0.286			0.853	>20
t	1.494	17	0.176			0.817	>20
5	1.118	245	0.057			0.754	>20
6	0.827	17	0.000			0.696	>20
z	-19.211	35	0.200			< 0.001	>20
7	-19.269	145	0.048			< 0.001	>20

\$social

factor	logodds	tokens	v/v+t	centered	factor	weight
d	0.116	272	0.338			0.529
a	0.012	21	0.810			0.503
s	-0.127	822	0.527			0.468

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	14	-2.013			0.486		0.118

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-9.754	47.508	47.89	1	0.997

Run 38 (above) with pessoa + paralelismo + social is worse than Run 33 also including fxetaria, $p = 0.00742$

Trying without social...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	44.477	543	0.996			> 0.999	>20
2sg	-44.477	572	0.002			< 0.001	>20

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
2	21.042	231	0.935			> 0.999	>20
6	16.664	17	0.000			> 0.999	>20
9	16.162	7	0.286			> 0.999	>20
4	0.039	71	0.930			0.51	>20
3	0.021	34	0.882			0.505	>20
y	-0.748	40	0.825			0.321	>20
1	-0.804	273	0.601			0.309	>20
t	-1.514	17	0.176			0.18	>20
5	-2.083	245	0.057			0.111	>20
7	-24.388	145	0.048			< 0.001	>20
z	-24.388	35	0.200			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.014	88	0.341			> 0.999	>20
k	3.872	532	0.404			0.98	>20
v	3.767	273	0.659			0.977	>20
c	-14.836	1	1.000			< 0.001	>20
j	-17.817	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	16	-2.271			0.486		0.094

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.682	41.364	41.86	1	0.999

Run 39 (above) with pessoa + paralelismo + fxetaria is worse than Run 33 also including social, $p = 0.148$

Dropping social...

STEP 7 - dropping from Run 39 model with pessoa + paralelismo + fxetaria

Trying without pessoa...

\$paralelismo

factor	logodds	tokens	v/v+t	centered	factor	weight
2	4.207	231	0.935			0.985
4	4.026	71	0.930			0.982
3	3.573	34	0.882			0.973
y	3.109	40	0.825			0.957
1	1.873	273	0.601			0.867
9	0.654	7	0.286			0.658
z	0.114	35	0.200			0.529
t	-0.002	17	0.176			0.499
5	-1.191	245	0.057			0.233
7	-1.441	145	0.048			0.191
6	-14.924	17	0.000			< 0.001

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight
c	12.977	1	1.000			> 0.999
v	-2.730	273	0.659			0.061
j	-3.253	221	0.525			0.037
k	-3.465	532	0.404			0.03
x	-3.529	88	0.341			0.029

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	15		1.716		0.486		0.848

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-393.062	816.124	816.561	0.833	0.746

Run 40 (above) with paralelismo + fxetaria is worse than Run 39 also including pessoa, $p = 6.1e-171$

Trying without paralelismo...

\$pessoa

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
3sg	23.174	543	0.996			> 0.999	>20
2sg	-23.174	572	0.002			< 0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	19.027	88	0.341			> 0.999	>20
c	0.305	1	1.000			0.576	>20
v	0.021	273	0.659			0.505	>20
k	-0.152	532	0.404			0.462	>20
j	-19.201	221	0.525			< 0.001	>20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	6		0.087		0.486		0.522

\$misc.2

```

log.likelihood    AIC  AICc  Dxy  R2
      -15.207 42.414 42.49 0.999 0.995

```

Run 41 (above) with pessoa + fxetaria is worse than Run 39 also including paralelismo, $p = 0.0207$

Trying without fxetaria...

\$pessoa

```

factor logodds tokens v/v+t centered factor weight vif
  3sg  23.249   543 0.996                > 0.999 >20
  2sg -23.249   572 0.002                < 0.001 >20

```

\$paralelismo

```

factor logodds tokens v/v+t centered factor weight vif
  2  22.622   231 0.935                > 0.999 >20
  4   2.979    71 0.930                0.952 >20
  3   2.817    34 0.882                0.944 >20
  y   2.665    40 0.825                0.935 >20
  1   2.241   273 0.601                0.904 >20
  9   1.727     7 0.286                0.849 >20
  t   1.501    17 0.176                0.818 >20
  5   1.128   245 0.057                0.755 >20
  6   0.764    17 0.000                0.682 >20
  7 -19.222   145 0.048                < 0.001 >20
  z -19.222    35 0.200                < 0.001 >20

```

\$misc.1

```

n df intercept overall proportion centered input prob
1115 12      -2.081                0.486                0.111

```

\$misc.2

```

log.likelihood    AIC  AICc  Dxy  R2
      -9.769 43.538 43.821  1 0.997

```

Run 42 (above) with pessoa + paralelismo is worse than Run 39 also including fxetaria, $p = 0.0376$

All remaining predictors are significant, best model from last step is Run 39

BEST STEP-DOWN MODEL OF RESPONSE pronome IS WITH PREDICTOR(S): pessoa (6.1e-171) + paralelismo (0.0207) + fxetaria (0.0376)
[p-values dropping from full model]

\$pessoa

```

factor logodds tokens v/v+t centered factor weight vif
  3sg  44.477   543 0.996                > 0.999 >20
  2sg -44.477   572 0.002                < 0.001 >20

```

\$paralelismo

```

factor logodds tokens v/v+t centered factor weight vif
  2  21.042   231 0.935                > 0.999 >20
  6  16.664    17 0.000                > 0.999 >20

```

9	16.162	7	0.286	>	0.999	>20
4	0.039	71	0.930		0.51	>20
3	0.021	34	0.882		0.505	>20
y	-0.748	40	0.825		0.321	>20
1	-0.804	273	0.601		0.309	>20
t	-1.514	17	0.176		0.18	>20
5	-2.083	245	0.057		0.111	>20
7	-24.388	145	0.048	<	0.001	>20
z	-24.388	35	0.200	<	0.001	>20

\$fxetaria

factor	logodds	tokens	v/v+t	centered	factor	weight	vif
x	25.014	88	0.341			>	0.999 >20
k	3.872	532	0.404				0.98 >20
v	3.767	273	0.659				0.977 >20
c	-14.836	1	1.000			<	0.001 >20
j	-17.817	221	0.525			<	0.001 >20

\$misc.1

n	df	intercept	overall	proportion	centered	input	prob
1115	16	-2.271			0.486		0.094

\$misc.2

log.likelihood	AIC	AICc	Dxy	R2
-4.682	41.364	41.86	1	0.999

BEST STEP-UP MODEL OF RESPONSE pronome IS WITH PREDICTOR(S): pessoa (~0) + paralelismo (0.0175) + fxetaria (0.0376)
[p-values building from null model]

BEST STEP-DOWN MODEL OF RESPONSE pronome IS WITH PREDICTOR(S): pessoa (6.1e-171) + paralelismo (0.0207) + fxetaria (0.0376)
 [p-values dropping from full model]

STEP-UP AND STEP-DOWN MATCH!