



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE HUMANIDADES**  
**DEPARTAMENTO DE LETRAS VERNÁCULAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA**

**MARDEN ALYSON MATOS DE ARAUJO**

**PADRÕES DE ALTERNÂNCIA PRONOMINAL E DE CONCORDÂNCIA VERBAL  
DA PRIMEIRA PESSOA DO PLURAL NO FALAR CULTO DE FORTALEZA: UMA  
ANÁLISE EM TEMPO REAL DE CURTA DURAÇÃO**

**FORTALEZA**

**2024**

MARDEN ALYSON MATOS DE ARAUJO

**PADRÕES DE ALTERNÂNCIA PRONOMINAL E DE CONCORDÂNCIA VERBAL  
DA PRIMEIRA PESSOA DO PLURAL NO FALAR CULTO DE FORTALEZA: UMA  
ANÁLISE EM TEMPO REAL DE CURTA DURAÇÃO**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Linguística (PPGL) do Centro de Humanidades da Universidade Federal do Ceará como requisito parcial à obtenção do título de Doutor em Linguística.

Área de concentração: Descrição e Análise Linguística.

Orientadora: Profa. Dra. Hebe Macedo de Carvalho

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- A69p Araujo, Marden Alyson Matos de.  
Padrões de alternância pronominal e de concordância verbal da primeira pessoa do plural no falar culto de Fortaleza : uma análise em tempo real de curta duração / Marden Alyson Matos de Araujo. – 2024.  
406 f.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Humanidades, Programa de Pós-Graduação em Linguística, Fortaleza, 2024.  
Orientação: Profa. Dra. Hebe Macedo de Carvalho.
1. Variação linguística. 2. Pronomes nós/ a gente. 3. Tempo real. 4. Mudança linguística. I. Título.  
CDD 410
-

MARDEN ALYSON MATOS DE ARAUJO

**PADRÕES DE ALTERNÂNCIA PRONOMINAL E DE CONCORDÂNCIA VERBAL  
DA PRIMEIRA PESSOA DO PLURAL NO FALAR CULTO DE FORTALEZA: UMA  
ANÁLISE EM TEMPO REAL DE CURTA DURAÇÃO**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Linguística (PPGL) do Centro de Humanidades da Universidade Federal do Ceará como requisito parcial à obtenção do título de Doutor em Linguística.

Área de concentração: Descrição e Análise Linguística.

Aprovada em: 30/07/2024.

BANCA EXAMINADORA

---

Prof. Dra. Hebe Macedo de Carvalho (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dra. Aluiza Alves de Araujo  
Universidade Estadual do Ceará (UECE)

---

Prof. Dr. Izabel Larissa Lucena Silva  
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (Unilab)

---

Prof. Dr. Sávio André de Souza Cavalcante  
Universidade Estadual do Ceará (UECE)

---

Prof. Dra. Maria Elias Soares  
Universidade Federal do Ceará (UFC)

## AGRADECIMENTOS

À minha estimada orientadora, Profa. Dra. Hebe Macedo de Carvalho, por toda ajuda, encaminhamentos e todas as contribuições ao desenvolvimento desta tese e à minha formação como pesquisador.

Ao Reryson, meu companheiro de vida, que esteve ao meu lado durante todo o processo de desenvolvimento deste trabalho. Todo apoio, cuidado e atenção foram fundamentais para superar essa etapa da minha vida acadêmica com êxito.

À professora Dra. Aluiza Alves de Araujo, por quem eu tenho verdadeira estima, que desde a graduação foi um exemplo de professora e de pesquisadora para mim. Todas as contribuições durante o desenvolvimento desta tese foram valiosas e essenciais para o aperfeiçoamento da pesquisa.

À professora Dra. Maria Elias Soares, pelas discussões tão necessárias nos primeiros meses de realização deste doutorado. Todos os ensinamentos foram essenciais para que fosse possível materializar esta tese.

Ao professor Dr. Sávio André de Souza Cavalcante e à professora Dra. Izabel Larissa Lucena Silva, pela disposição em participar da banca de avaliação deste trabalho.

A admirável Profa. Dra. Socorro Aragão, que segurou minhas mãos nos momentos iniciais desta tese e me apontou uma direção a seguir. Muito obrigado.

À amável professora Dra. Rakel Beserra, por toda ajuda e suporte durante a pesquisa. Os inúmeros arquivos compartilhados e as discussões realizadas foram fundamentais para a conclusão do trabalho. Você foi fantástica.

À Maylle Freitas, estimada colega de pesquisa, pelas várias horas gastas discutindo os aspectos funcionais e estatísticos do *R Studio*. Todo o conhecimento compartilhado foi fundamental para o tratamento dos dados desta pesquisa.

Ao Renato Souza, querido colega de estudo, por ter trazido luz nos últimos meses de conclusão deste trabalho.

Aos queridos colegas do Grupo de Estudos de Sociolin-CE, pelos excelentes encontros e momentos de descontração.

À professora Dra. Lorena Rodrigues, pela competência e pelas excelentes contribuições dadas para o desenvolvimento desta pesquisa.

À Fundação Cearense De Apoio Ao Desenvolvimento Científico e Tecnológico – FUNCAP, pelo apoio financeiro fundamental para a realização desta pesquisa.

Aos informantes do Projeto PORCUFORT das décadas de 1990 e de 2020, que com suas doces histórias desenharam cada capítulo desta tese.

## RESUMO

Partindo do arcabouço teórico-metodológico utilizado pela Teoria da Variação e da Mudança Linguística, delineado por Weinreich, Labov e Herzog (1968) e por Labov (1997, 2001, 2003), esta pesquisa objetiva analisar e descrever, em tempo real, os padrões de alternância pronominal e os padrões de concordância verbal com os pronomes *nós* e *a gente* no falar culto de Fortaleza. Para tanto, utilizamos amostra composta por 104 informantes provenientes do Projeto Português Oral Culto de Fortaleza (PORCUFORT), sendo 50 informantes extraídos da Fase I do projeto, com gravações realizadas entre os anos de 1993 e 1995 (Década de 1990), e outros 54 informantes retirados da Fase II do projeto, com entrevistas realizadas entre 2018 e 2022 (Década de 2020). Nossa amostra foi socialmente estratificada de acordo com o tipo de inquérito (DID – diálogo entre informante e documentador; D2 – Diálogo entre dois informantes; e EF – elocuições formais), sexo (homem e mulher), faixa etária (22 a 35 anos; 36 a 55 anos; e a partir de 56 anos) e década de gravação (1990 e 2020). Para análise da alternância pronominal, controlamos os preditores *grau de referencialidade do pronome, tipo de verbo, paralelismo formal, tempo verbal e saliência fônica*. Em relação aos padrões de concordância verbal com os pronomes de primeira pessoa do plural, observamos a atuação das variáveis *tempo e paradigma verbal do modo indicativo, estrutura do verbo, conjugação verbal e tonicidade do verbo*. Através da linguagem de programação R, por meio do software *R Studio*, obtivemos um total de 4575 observações de *nós* e *a gente*. Na Amostra 1990, a análise apontou maior proporção de *a gente* (58,3%) em detrimento do *nós* (41,7%). Na década de 2020, a análise identificou uso ainda mais elevado de *a gente* (79,95%), e menor recorrência de *nós* (20,05%). Por outro lado, a análise da concordância verbal contabilizou 4574 ocorrências. Na década de 1990, foram flagrados 1957 dados distribuídos da seguinte forma: *a gente* sem *-mos* (58,2%), *nós* com *-mos* (41%), *nós* sem *-mos* (0,7%) e *a gente* com *-mos* (0,1%). Na década de 2020, se notou aumento significativo do padrão emergente, aumento discreto da concordância não padrão e diminuição expressiva no uso da estrutura padrão de concordância, distribuídos da seguinte forma: *a gente* sem *-mos* (79,5%), *nós* com *-mos* (19,1%), *nós* sem *-mos* (1%) e *a gente* com *-mos* (0,5%). De modo geral, houve diminuição da proporção de uso de *nós* com *-mos*, em todos os contextos observados. A análise em tempo real de curta duração indicou franca expansão do pronome inovador *a gente*, sendo beneficiado nos contextos analisados em maior intensidade na Amostra 2020, e mais discretamente nos dados da Amostra 1990. Esses resultados demonstram que a variante inovadora se torna mais evidente a cada década, sendo preferida em todas as faixas etárias da década atual. Apesar da forte expansão do pronome *a*

*gente* e sua estrutura emergente de concordância no falar culto de Fortaleza, os resultados apontam para uma certa estabilidade na variação.

**Palavras-chave:** Variação linguística; Pronomes *nós/ a gente*; Tempo real; Mudança linguística;

## ABSTRACT

Based on the theoretical-methodological framework used by the Theory of Linguistic Variation and Change, outlined by Weinreich, Labov and Herzog (1968) and by Labov (1997, 2001, 2003), this research aims to analyze and describe, in real time, the patterns of pronoun alternation and the patterns of verbal agreement with the pronouns *nós* and *a gente* in the cultured speech of Fortaleza. To this end, we used a sample composed of 104 informants from the Projeto Português Oral Culto de Fortaleza (PORCUFORT), with 50 informants taken from Phase I of the project, with recordings made between 1993 and 1995 (1990s), and another 54 informants taken from Phase II of the project, with interviews conducted between 2018 and 2022 (2020s). Our sample was socially stratified according to the type of survey (D1 – dialogue between informant and documenter; D2 – dialogue between two informants; and EF – formal utterances), sex (man and woman), age group (22 to 35 years old; 36 to 55 years old; and 56 years old and older), and decade of recording (1990 and 2020). To analyze pronoun alternation, we controlled the predictors degree of *pronoun referentiality*, *verb type*, *formal parallelism*, *verb tense*, and *phonic salience*. Regarding the patterns of verbal agreement with first-person plural pronouns, we observed the performance of the variables *tense and verbal paradigm of the indicative mood*, *verb structure*, *verb conjugation*, and *verb tonicity*. Through the *R* programming language, through the *R Studio* software, we obtained a total of 4,575 observations of *nós* and *a gente*. In the 1990 Sample, the analysis indicated a higher proportion of *a gente* (58.3%) to the detriment of *nós* (41.7%). In the 2020s, the analysis identified an even higher use of *a gente* (79.95%), and a lower recurrence of *nós* (20.05%). On the other hand, the analysis of verbal agreement counted 4,574 occurrences. In the 1990s, 1,957 data were recorded, distributed as follows: *a gente* without *-mos* (58.2%), *nós* with *-mos* (41%), *nós* without *-mos* (0.7%), and *a gente* with *-mos* (0.1%). In the 2020s, there was a significant increase in the emerging pattern, a slight increase in non-standard agreement, and a significant decrease in the use of the standard agreement structure, distributed as follows: *a gente* without *-mos* (79.5%), *nós* with *-mos* (19.1%), *nós* without *-mos* (1%), and *a gente* with *-mos* (0.5%). In general, there was a decrease in the proportion of use of *nós* with *-mos* in all contexts observed. The short-term real-time analysis indicated a clear expansion of the innovative pronoun *a gente*, benefiting in the contexts analyzed with greater intensity in the 2020 Sample, and more discreetly in the data from the 1990 Sample. These results demonstrate that the innovative variant becomes more evident with each decade, being preferred in all age groups in the current decade. Despite the

strong expansion of the pronoun *a gente* and its emerging agreement structure in the formal speech of Fortaleza, the results point to a certain stability in the variation.

**Keywords:** Linguistic variation; Pronouns *nós/ a gente*; Real time; Linguistic change.

## LISTA DE FIGURAS

Figura 1 – Mapa do Brasil com pesquisas variacionistas em cada região .....	57
Figura 2 – Imagem aérea do litoral de Fortaleza-CE .....	85
Figura 3 – Mapa da região metropolitana de Fortaleza .....	83
Figura 4 – Av. Domingos Olímpio antes do alargamento (08/01/1992) .....	85
Figura 5 – Coleta dos dados de nossa pesquisa .....	126
Figura 6 – Correspondência entre os valores em <i>odds</i> , <i>logodds</i> e probabilidade .....	139
Figura 7 – Mensagem de erro ao convergir modelo de regressão – Amostra 2020 .....	223
Figura 8 – Mensagem de erro ao convergir modelo de regressão – Amostra 1990 .....	317

## LISTA DE GRÁFICOS

Gráfico 1	– Queda do traço de plural do substantivo <i>gente</i> em tempo real .....	53
Gráfico 2	– Comparação da proporção de uso de <i>nós</i> e <i>a gente</i> nos estudos variacionistas do português brasileiro .....	74
Gráfico 3	– Crescimento da população de graduados nas décadas de 1990, 2000 e 2010 ..	89
Gráfico 4	– Evolução das matrículas realizadas entre 2010 e 2019 .....	90
Gráfico 5	– Proporção de uso das variantes <i>nós</i> e <i>a gente</i> na amostra da década de 1990 .....	149
Gráfico 6	– Comparação de nossos resultados com outras pesquisas com amostras de fala culta .....	150
Gráfico 7	– Comparação no uso de <i>nós</i> e <i>a gente</i> entre mulheres em estudos variacionista .....	154
Gráfico 8	– Proporção das <i>nós</i> e <i>a gente</i> por <i>tipo de inquérito</i> : Amostra de 1990 .....	157
Gráfico 9	– Proporção de uso de <i>nós</i> e <i>a gente</i> em função do grau de saliência fônica .....	166
Gráfico 10	– Proporção de uso das variantes <i>nós</i> e <i>a gente</i> após remoção de dados colineares .....	170
Gráfico 11	– Proporção de uso de <i>nós</i> e <i>a gente</i> por faixa etária (N = 1938) .....	178
Gráfico 12	– Proporção de uso dos pronomes <i>nós</i> e <i>a gente</i> : Amostra 2020 (N=2618) .....	200
Gráfico 13	– Comparação entre proporções de uso de <i>nós</i> e <i>a gente</i> em estudos variacionistas .....	201
Gráfico 14	– Comparação no uso de <i>nós</i> e <i>a gente</i> entre homens e mulheres em outros estudos .....	205
Gráfico 15	– Proporção das variantes por <i>faixa etária</i> na Amostra 2020 .....	208
Gráfico 16	– Proporção de uso de <i>nós</i> e <i>a gente</i> por <i>tipo de inquérito</i> na Amostra 2020 .....	210
Gráfico 17	– Proporção de <i>nós</i> e <i>a gente</i> em função do nível de <i>saliência fônica</i> .....	220
Gráfico 18	– Proporção de uso de <i>nós</i> e <i>a gente</i> após remoção de dados colineares .....	225

Gráfico 19	– Proporção de uso de <i>nós</i> e <i>a gente</i> em função do <i>grau de referencialidade</i> ....	236
Gráfico 20	– Proporção de uso de <i>a gente</i> e <i>nós</i> em tempo real: décadas de 1990 e 2020 ....	249
Gráfico 21	– Comparação do resultado de nosso estudo com outras pesquisas em tempo real .....	250
Gráfico 22	– Proporção uso de <i>nós</i> e <i>a gente</i> por <i>faixa etária</i> em tempo real (N = 4573) ....	257
Gráfico 23	– Tendência de uso de <i>a gente</i> por <i>faixa etária</i> em tempo real.....	259
Gráfico 24	– Tendência no uso de <i>nós</i> e <i>a gente</i> por <i>tipo de inquérito</i> em tempo real (N=4573) .....	263
Gráfico 25	– Tendência de uso de <i>nós</i> e <i>a gente</i> por <i>grau de referencialidade</i> em tempo real.....	267
Gráfico 26	– Uso de <i>nós</i> e <i>a gente</i> por verbos de <i>ação, dicendi e epistêmico</i> em tempo real .....	271
Gráfico 27	– Uso de <i>nós</i> e <i>a gente</i> por verbos de <i>estado e verbo ter</i> em tempo real.....	273
Gráfico 28	– Tendência de uso de <i>nós</i> e <i>a gente</i> por <i>saliência fônica</i> em tempo real.....	278
Gráfico 29	– Proporção de uso da concordância verbal na Amostra 1990 (N=1957) .....	288
Gráfico 30	– Proporção de uso da concordância verbal por <i>faixa etária</i> (Amostra 1990) ....	296
Gráfico 31	– Proporção de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> (N=1589) .....	316
Gráfico 32	– Proporção de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> por <i>faixa etária</i> (Amostra 1990) .....	325
Gráfico 33	– Proporções de uso de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> por <i>tipo de inquérito</i> .....	327
Gráfico 34	– Distribuição de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> por estrutura do verbo.....	336
Gráfico 35	– Distribuição dos padrões de concordância verbal: Amostra 2020 .....	339

Gráfico 36 – Proporções da concordância verbal em recorte binário – Amostra 2020 .....	359
Gráfico 37 – Proporção de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> por <i>tipo de inquérito</i> - Amostra 2020 .....	367
Gráfico 38 – Proporção das variantes sobre <i>o tempo e tipo de paradigma verbal</i> .....	368
Gráfico 39 – Proporção de uso das variantes em função da <i>estrutura verbal</i> .....	371
Gráfico 40 – Padrões de concordância verbal em tempo real .....	374
Gráfico 41 – Distribuição de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> em tempo real .....	375
Gráfico 42 – Distribuição das variantes de concordância por <i>faixa etária</i> em tempo real .....	378
Gráfico 43 – Proporção das variantes de concordância por <i>tipo de inquérito</i> em tempo real .....	381

## LISTA DE TABELAS

Tabela 1 – Atuação da função sintática sobre o pronome <i>a gente</i> em Araujo (2016) .....	48
Tabela 2 – Atuação da <i>faixa etária</i> sobre o pronome <i>a gente</i> em Araujo (2018) .....	134
Tabela 3 – Valores correspondentes entre probabilidades, <i>odds</i> e <i>logodds</i> .....	141
Tabela 4 – Resultados do modelo de regressão: Freitas, Rodrigues e Santos (2022) .....	144
Tabela 5 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>sexo/gênero</i> : Amostra 1990 (N=1956) ....	152
Tabela 6 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>faixa etária</i> : Amostra de 1990 (N=1956).	155
Tabela 7 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>tipo de inquérito</i> : Amostra 1990 (N=1956) .....	156
Tabela 8 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>grau de referencialidade</i> : Amostra 1990 (N=1956) .....	159
Tabela 9 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>tipo de verbo</i> : Amostra 1990 (N=1956) .....	160
Tabela 10 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>paralelismo</i> : amostras de 1990 (N=1956) .....	162
Tabela 11 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>Tempo verbal e formas nominais do verbo</i> : Amostra 1990 (N=1956) .....	164
Tabela 12 – Distribuição de <i>nós</i> e <i>a gente</i> segundo a <i>saliência fônica</i> : Amostra 1990 (N=1956) .....	165
Tabela 13 – Resultado do teste de colinearidade das variáveis preditoras através da função <i>vif</i> .....	171
Tabela 14 – Resultado do teste de seleção automática das variáveis por meio da função <i>step</i> .....	173
Tabela 15 – Resultado do teste multivariado de <i>nós</i> e <i>a gente</i> através da função <i>lrm</i> .....	174
Tabela 16 – Estimativas de uso de <i>a gente</i> modelo de regressão logística: Amostra 1990 (N = 1938) .....	175

Tabela 17 – Distribuição de <i>nós</i> e <i>a gente</i> por paradigma verbal do modo indicativo ....	196
Tabela 18 – Proporção de <i>nós</i> e <i>a gente</i> por <i>sexo/gênero</i> : Amostra 2020 (N= 2618) .....	204
Tabela 19 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>faixa etária</i> : Amostra 2020 (N=2618) ....	207
Tabela 20 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>tipo de inquérito</i> : Amostra 2020 (N=2618).	209
Tabela 21 – Distribuição de <i>nós</i> e <i>a gente</i> segundo a <i>referencialidade</i> : Amostra 2020 (N=2618) .....	212
Tabela 22 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>tipo de verbo</i> : Amostra 2020 (N=2618) ....	213
Tabela 23 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>paralelismo</i> : Amostra 2020 (N=2618) .....	215
Tabela 24 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o <i>Tempo verbal e formas nominais do verbo</i> : Amostra 2020 (N=2618) .....	217
Tabela 25 – Proporção de <i>nós</i> e <i>a gente</i> segundo a <i>saliência fônica</i> : Amostra 2020 (N=2618) .....	219
Tabela 26 – Resultado do teste de colinearidade das variáveis preditoras através da função <i>vif</i> .....	226
Tabela 27 – Resultado do teste de seleção automática das variáveis por meio da função <i>step</i> .....	226
Tabela 28 – Resultado do teste multivariado de <i>nós</i> e <i>a gente</i> através da função <i>lrm</i> .....	227
Tabela 29 – Estimativas de uso de <i>a gente</i> modelo de regressão logística: Amostra 2020 (N = 2596) .....	229
Tabela 30 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>faixa etária</i> em tempo real .....	256
Tabela 31 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>tipo de inquérito</i> em tempo real .....	261
Tabela 32 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>grau de referencialidade</i> em tempo real ..	266
Tabela 33 – Distribuição de <i>nós</i> e <i>a gente</i> por <i>tipo de verbo</i> em tempo real .....	271
Tabela 34 – Distribuição de <i>nós</i> e <i>a gente</i> segundo o grau de <i>saliência fônica</i> em tempo real .....	275

Tabela 35 – Distribuição das proporções da concordância uso em função do <i>sexo/gênero</i> .....	293
Tabela 36 – Distribuição das proporções de uso da concordância verbal por <i>faixa etária</i> .....	295
Tabela 37 – Distribuição das proporções da concordância verbal por <i>tipo de inquérito</i> .....	298
Tabela 38 – Distribuição das variantes de concordância verbal por tempo e paradigma verbal .....	302
Tabela 39 – Distribuição das proporções da concordância verbal por <i>estrutura do verbo</i> .....	309
Tabela 40 – Distribuição das proporções da concordância verbal por <i>conjugação verbal</i> .....	310
Tabela 41 – Distribuição das proporções da concordância verbal por <i>tonicidade do verbo</i> .....	312
Tabela 42 – Distribuição de concordância verbal com dados de modo indicativo .....	314
Tabela 43 – Resultado do teste de colinearidade <i>vif</i> da concordância verbal – Amostra 1990 .....	318
Tabela 44 – Resultado da seleção das variáveis de concordância por meio da função <i>step</i> .....	319
Tabela 45 – Resultado do teste multivariado de <i>nós</i> e <i>a gente</i> através da função <i>lrm</i> .....	320
Tabela 46 – Estimativas de <i>nós</i> com <i>-mos</i> no modelo de regressão logística: 1990 (N = 1586) .....	321
Tabela 47 – Efeito do <i>tempo e tipo de paradigma verbal</i> sobre concordância verbal em outras amostras do português brasileiro .....	331
Tabela 48 – Distribuição de <i>nós</i> com <i>-mos</i> e <i>a gente</i> sem <i>-mos</i> por tipo de verbo .....	334
Tabela 49 – Comparação do resultado de nosso estudo com pesquisas sociolinguísticas .....	340

Tabela 50 – Distribuição das proporções de uso da concordância por <i>sexo/gênero</i> .....	341
Tabela 51 . Distribuição das proporções de uso da concordância por <i>tipo de inquérito</i> ....	344
Tabela 52 – Proporções da concordância verbal por <i>tempo e tipo de paradigma verbal</i> .....	346
Tabela 53 – Distribuição das proporções da concordância verbal por <i>estrutura do verbo</i> .	351
Tabela 54 – Distribuição das proporções da concordância verbal por <i>conjugação verbal</i> ..	353
Tabela 55 – Distribuição das proporções da concordância verbal por <i>tonicidade do verbo</i> .....	355
Tabela 56 – Distribuição das proporções da concordância com dados no modo indicativo.....	358
Tabela 57 – Resultado do teste de colinearidade <i>vif</i> das variáveis preditoras - Amostra 2020 .....	360
Tabela 58 – Resultado do teste de seleção das variáveis através da função <i>step</i> - Amostra 2020 .....	361
Tabela 59 – Resultado do teste multivariado de <i>nós e a gente</i> através da função <i>lrm</i> .....	361
Tabela 60 – Estimativas de <i>nós</i> com <i>-mos</i> no modelo de regressão logística: 2020 (N = 2224) .....	362
Tabela 61 – Atuação da <i>faixa etária</i> sobre o uso de <i>nós</i> com <i>-mos</i> em tempo real .....	378
Tabela 62 – Atuação do <i>tipo de inquérito</i> sobre o uso de <i>nós</i> com <i>-mos</i> em tempo real .....	380
Tabela 63 – Atuação do <i>tempo e paradigma verbal</i> sobre <i>nós</i> com <i>-mos</i> em tempo real .....	382
Tabela 64 – Atuação da <i>estrutura verbal</i> sobre o uso de <i>nós</i> com <i>-mos</i> em tempo real .....	384

## LISTA DE QUADROS

Quadro 1	– Correlação entre indivíduo e comunidade em tempo real e aparente .....	40
Quadro 2	– Processo de gramaticalização de a gente .....	54
Quadro 3	– Evolução do Número de Instituições de Ensino Superior no Brasil - 1980-1998 .....	86
Quadro 4	– Comparação entre o número de vagas ofertadas e o número de inscrições no vestibular entre os anos de 1990 e 1998 .....	87
Quadro 5	– Comparativo entre vagas ofertadas, inscritos no vestibular e ingressos no ano de 2018 no estado do Ceará .....	91
Quadro 6	– Estratificação dos informantes do PORCUFORT Fase I .....	93
Quadro 7	– Estratificação dos informantes do PORCUFORT Fase II .....	95
Quadro 8	– Distribuição dos informantes nas amostras do PORCUFORT Fase I (década de 1990) e Fase II (década de 2020) .....	96
Quadro 9	– Detalhamento dos informantes do PORCUFORT Fase I selecionados para esta pesquisa .....	97
Quadro 10	– Detalhamento dos informantes do PORCUFORT Fase II selecionados para esta pesquisa .....	99
Quadro 11	– Comparação entre a tendência geral na amostra e nas análises de 1990 e de 2020 .....	254
Quadro 12	Comparação do resultado de nosso estudo com pesquisas sociolinguísticas .....	289
Quadro 13	Confronto entre os resultados das análises geral e específicas de concordância .....	376

## **LISTA DE ABREVIATURAS E SIGLAS**

ABNT	Associação Brasileira de Normas Técnicas
IBGE	Instituto Brasileiro de Geografia e Estatística
BNCC	Base Nacional Comum Curricular
DID	Diálogo Entre Informante e Documentador
D2	Diálogo Entre Dois Informantes
EF	Elocução Formal
PB	Português Brasileiro
1PP	Primeira Pessoa do Plural
3SL	Terceira Pessoa do Singular
DNP	Desinência Número Pessoal

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	24
2	<b>TEORIA DA VARIAÇÃO E MUDANÇA LINGUÍSTICA</b> .....	37
3	<b>O FENÔMENO DA PRIMEIRA PESSOA DO PLURAL</b> .....	45
3.1	<b>A inserção de <i>a gente</i> no quadro dos pronomes do português</b> .....	45
3.2	<b>Panorama dos estudos sobre <i>nós</i> e <i>a gente</i> no Brasil</b> .....	55
3.2.1	<b><i>Panorama dos estudos variacionistas sobre alternância pronominal</i></b> .....	56
3.2.1.1	<i>Estudos sobre a alternância pronominal em tempo aparente</i> .....	57
3.2.1.2	<i>Estudos sobre a alternância pronominal em tempo real</i> .....	66
3.2.2	<b><i>Panorama dos estudos sobre concordância verbal</i></b> .....	75
4	<b>METODOLOGIA</b> .....	79
4.1	<b>Método de pesquisa empregado</b> .....	79
4.2	<b>Descrição da comunidade de fala: os falantes cultos de Fortaleza</b> .....	81
4.2.1	<b><i>Um breve contexto sócio-histórico</i></b> .....	83
4.3	<b>Os corpora e as amostras</b> .....	91
4.4	<b>Variáveis</b> .....	101
4.4.1	<b><i>Variáveis previsoras: o fenômeno da alternância pronominal</i></b> .....	102
4.4.1.1	<i>Variáveis previsoras de ordem linguística</i> .....	103
4.4.1.2	<i>Variáveis previsoras de ordem extralinguística</i> .....	113
4.4.2	<b><i>Variáveis previsoras: o fenômeno da concordância verbal</i></b> .....	116
4.5	<b>Procedimentos de coleta e análise dos dados</b> .....	125
4.6	<b>Análise estatística: o modelo quantitativo</b> .....	126
4.7	<b>Estatística básica e inferencial para o tratamento de dados linguísticos</b> ....	132
4.7.1	<b><i>Alguns conceitos básicos da estatística inferencial</i></b> .....	133
4.7.2	<b><i>O modelo de regressão logística</i></b> .....	137
4.7.3	<b><i>A linguagem de programação R e a interface R Studio</i></b> .....	142

<b>5</b>	<b>DESCRIÇÃO DOS DADOS E ANÁLISE DOS RESULTADOS: O FENÔMENO DA ALTERNÂNCIA PRONOMINAL</b> .....	<b>147</b>
<b>5.1</b>	<b>A alternância pronominal de primeira pessoa do plural no falar culto de Fortaleza: análise descritiva dos dados de fala da década de 1990</b> .....	<b>148</b>
<b>5.1.1</b>	<i>Atuação das variáveis previsoras em termos de frequência e proporção de uso: a alternância pronominal de primeira pessoa do plural na Amostra 1990</i> .....	<b>152</b>
<b>5.1.2</b>	<i>Análise multivariada da atuação das variáveis linguísticas e sociais sobre a variável resposta: o modelo de regressão logística – Amostra 1990</i> .....	<b>168</b>
<b>5.1.3</b>	<i>Síntese dos resultados – A alternância pronominal na Amostra 1990</i> .....	<b>198</b>
<b>5.2</b>	<b>A alternância pronominal de primeira pessoa do plural no falar culto de Fortaleza: análise descritiva dos dados de fala da década 2020</b> .....	<b>200</b>
<b>5.2.1</b>	<i>Atuação das variáveis previsoras em termos de frequência e proporção de uso: a alternância pronominal na Amostra 2020</i> .....	<b>204</b>
<b>5.2.2</b>	<i>Análise multivariada da atuação das variáveis linguísticas e sociais sobre a variável resposta: o modelo de regressão logística – Amostra 2020</i> .....	<b>222</b>
<b>5.2.3</b>	<i>Síntese dos resultados – A alternância pronominal na Amostra 2020</i> .....	<b>245</b>
<b>5.3</b>	<b>Análise comparativa das décadas de 1990 e de 2020: um estudo em tempo real de curta duração de tendência</b> .....	<b>247</b>
<b>5.3.1</b>	<i>Síntese dos resultados – A alternância pronominal em tempo real de curta duração</i> .....	<b>282</b>
<b>6</b>	<b>DESCRIÇÃO DOS DADOS E ANÁLISE DOS RESULTADOS: O FENÔMENO DA CONCORDÂNCIA VERBAL</b> .....	<b>284</b>
<b>6.1</b>	<b>A concordância verbal de primeira pessoa do plural no falar culto de Fortaleza: análise descritiva dos dados de fala da década 1990</b> .....	<b>287</b>
<b>6.1.1</b>	<i>Atuação das variáveis previsoras em termos de frequência e proporção de uso: a concordância verbal na Amostra 1990</i> .....	<b>292</b>
<b>6.1.2</b>	<i>Análise multivariada em um modelo de regressão logística: a concordância verbal em um recorte binário – Amostra 1990</i> .....	<b>314</b>
<b>6.1.3</b>	<i>Síntese dos resultados – A concordância verbal na Amostra 1990</i> .....	<b>337</b>
<b>6.2</b>	<b>A concordância verbal de primeira pessoa do plural no falar culto de Fortaleza: análise descritiva dos dados de fala da década 2020</b> .....	<b>338</b>

6.2.1	<i>Atuação das variáveis previsoras em termos de frequência e proporção de uso: a concordância verbal na Amostra 2020</i> .....	340
6.2.2	<i>Análise multivariada em um modelo de regressão logística: a concordância verbal em um recorte binário – Amostra 2020</i> .....	357
6.2.3	<i>Síntese dos resultados – A concordância verbal na Amostra 2020</i> .....	372
6.3	<b>A concordância verbal nas décadas de 1990 e de 2020: um estudo em tempo real de curta duração de tendência</b> .....	373
6.3.1	<i>Síntese dos resultados – A concordância verbal em tempo real de curta duração</i> .....	386
7	<b>CONCLUSÃO</b> .....	387
	<b>REFERÊNCIAS</b> .....	397

## 1 INTRODUÇÃO

Um dos objetivos da Sociolinguística Variacionista é estudar, analisar, apreender e sistematizar os padrões de comportamento linguístico observáveis em uma comunidade de fala, visando à descrição de fenômenos variáveis empregados e compartilhados por falantes dentro desta mesma comunidade de fala (Lucchesi, 2004). Dentre estes fenômenos variáveis, estão inseridos os padrões de *alternância pronominal* e de *concordância verbal de primeira pessoa do plural*, processos que estão correlacionados, constituindo faces de um mesmo fenômeno variável, na medida em que a escolha entre os pronomes *nós* e *a gente* influencia a concordância verbal subsequente (Rubio, 2012; Scherre, Yacovenco e Naro, 2018; Carvalho, Freitas e Favacho, 2020).

A expressão *a gente* tem sua origem no substantivo *gente* (Menon, 1995; Lopes, 2003). Por extensão de uso, a forma passou a desempenhar função de pronome pessoal, por meio de um processo de pronominalização. De acordo com Menon (1996, p. 186), “a forma deixa de ser um substantivo e passa a integrar o sistema dos pronomes pessoais, conservando, porém, com o verbo a mesma relação sintática de terceira pessoa gramatical”. Nesse sentido, o pronome *a gente* manteve associação formal com o verbo na terceira pessoa do singular, mesmo resgatando a interpretação semântica de plural, e passou a disputar o domínio de primeira pessoa do plural com o pronome *nós*, criando uma estrutura emergente na língua (*a gente sem -mos*) que é favorecida em diversos contextos linguísticos e sociais.

As gramáticas, de modo geral, apresentam um quadro pronominal fixo e invariável que se limitam aos seguintes pronomes: *eu, tu, ele, nós, vós, eles* (Neves, 2000; Cegalla, 2005; Bechara, 2009; Perini, 2010). Estudos sociolinguísticos — Lopes (1993), Brustolin (2009), Silva (2010), Santos et al. (2011), Vítório (2015), Araújo (2016), Freitas et al. (2020) e Silva (2022), para citar alguns deles —, acerca dos pronomes pessoais de primeira pessoa do plural no português brasileiro, tanto na fala quanto na escrita, comprovam que em consequência de várias mudanças ocorridas na língua, esse quadro já não mais corresponde com a realidade, tendo em vista que o uso do pronome *a gente* para designar a primeira pessoa do plural é bastante frequente entre os falantes de Língua Portuguesa no Brasil.

Pesquisas sobre a variação no uso dos pronomes *nós* e *a gente* amparados pela Teoria da Variação e da Mudança Linguísticas (Weinreich; Labov; Herzog, 2006 [1968]) atestam significativa proporção de uso do pronome inovador *a gente* em detrimento do pronome conservador *nós* em diversas amostras do português brasileiro. Alguns desses trabalhos utilizam como amostra *corpora* de várias regiões do país, capturados em dado momento do tempo, com

dados sincrônicos, ou seja, em tempo aparente. Dentre outros, podemos localizar o trabalho de Omena (1996), sobre o falar carioca; Maia (2003), acerca do falar em Belo Horizonte; e Fernandes (2004), a respeito do falar de João Pessoa.

Além desses estudos, outras pesquisas consideram analisar o fenômeno variável em tempo real, isto é, com dados diacrônicos, considerando duas décadas distintas. Dentre esses trabalhos podemos destacar as pesquisas desenvolvidas por Lopes (2003) e por Nascimento (2013) quando estudaram a variação entre os pronomes *nós* e *a gente* em amostras do Projeto Norma Urbana Oral Culta — NURC, nas cidades do Rio de Janeiro e Salvador, com amostra das décadas de 1970 e 1990, respectivamente

Um ponto importante de salientar é que, no Brasil, *corpora* que possuem amostras em duas épocas distintas são escassos, fato que dificulta a realização de estudos com análises em tempo real sobre fenômenos variáveis. Dentre os bancos de dados de que temos conhecimento, destacamos o Programa de Estudos sobre o Uso da Língua - PEUL, o Projeto da Norma Urbana Oral Culta – NURC, o Projeto Censo da Variação Lingüística do Rio de Janeiro - CENSO, e o Projeto Descrição do Português Oral Culto de Fortaleza-CE, PORCUFORT em suas Fases I e II, pois possibilitam o desenvolvimento de análises comparativas com amostras de fala de duas décadas distintas, atendendo aos requisitos necessários para a realização de pesquisas em tempo real no Brasil.

Nosso objetivo, neste estudo, é analisar em tempo real de curta duração os padrões de *alternância pronominal* e de *concordância verbal de primeira pessoa do plural* na fala culta de Fortaleza – CE, a partir de dados provenientes do Projeto Português Oral Culto de Fortaleza - PORCUFORT Fase I (década de 1990) e Fase II (décadas de 2020). Pontuamos que as gravações da segunda fase do Projeto PORCUFORT foram iniciadas em 13 de abril de 2018, com a gravação do primeiro inquérito do tipo DID (diálogo entre informante e documentador), e foram finalizadas em 15 de fevereiro de 2022, com a coleta de um inquérito do tipo EF (elocução formal) (Araújo; Viana; Rodrigues, 2021). Por captar a fala culta fortalezense no período de transição da década de 2010 para 2020, optamos por adotar as nomenclaturas “Década 2020”, “Amostra 2020” ou ainda “Anos 2020” para fazer referência aos dados referentes à Fase II do projeto. Sob a óptica da Teoria da Variação e da Mudança Linguística (Weinreich; Labov; Herzog, 2006 [1968]), visamos verificar os efeitos de variáveis linguísticas e sociais sobre a realização das formas *nós* e *a gente* e as suas respectivas concordâncias padrão e não padrão. Para isso, observamos apenas pronomes na função de sujeito e expressamente

realizados. Ou seja, neste recorte, não consideramos os dados de pronomes implícitos<sup>1</sup> (ou apagados) para efeito de análise.

As variações na alternância e na concordância com os pronomes de primeira pessoa do plural compreendem realizações que demonstram faces distintas de um mesmo fenômeno linguístico (Scherre, Yacovenco e Naro, 2018; Carvalho, Freitas e Favacho, 2020). Sendo assim, analisamos neste estudo duas variáveis respostas distintas (ou variáveis dependentes): a realização dos pronomes *nós* e *a gente* e a *concordância verbal de primeira pessoa do plural*. Primeiramente, tratamos os dados de *alternância pronominal* (ver capítulo 5.1), sendo uma variável resposta constituída por duas variantes: *nós* e *a gente*. Os *supertokens*<sup>2</sup> seguintes, coletados da base de dados em estudo, ilustram as variantes adotadas para efeito de análise dos dados:

- (1) Inf.: isso é puro folclore tá? simplesmente ***a gente*** diz que ***nós*** trabalhamos um lado só do cérebro o lado direito (Inq. 21 – F1 – DID, homem, 25 anos)
- (2) Inf.: a diretora manda chamar a professora da sala vizinha porque ***a gente*** tava só... aí depois do outro dia ***nós*** ficamos TUDO com ela (Inq. 55 – F1 – D2, mulher, 45 anos)

Além da alternância entre os pronomes de primeira pessoa do plural, neste recorte analisamos os *padrões de concordância verbal* com os pronomes *nós* e *a gente*, distribuídos em função da presença ou ausência da desinência verbal de primeira pessoa do plural *-mos*, sendo esta uma variável que comporta quatro variantes: *nós* com *-mos* (*nós cantamos*), *nós* sem *-mos* (*nós canta*), *a gente* sem *-mos* (*a gente canta*) e *a gente* com *-mos* (*a gente cantamos*). As ocorrências dispostas de (3) a (6) ilustram a realização dos pronomes de primeira pessoa do plural com suas respectivas concordâncias padrão e não padrão:

#### ***Nós com -mos:***

- (3) Inf.: aí na hora que ***nós chegamos*** nesse local pra pegar o trem... aí eu digo ah não vou logo tomar meu chocolate quente ((risos)) o rapaz do trem chamando... (Inq. 24 – F2 – DID, mulher, 64 anos)

<sup>1</sup> Consideramos sujeito implícito aquele que não aparece na oração de forma explícita, expressamente marcada, como no seguinte exemplo extraído da base de dados analisada:

Inf.: professor de Educação... oh... técnico de E/ de basquetebol ***somos*** educador né? e mas ***somos*** também... professor de Educação Física (Inq. 34 - F1 - D2, homem, 41 anos).

<sup>2</sup> *Supertoken* é a alternância de duas ou mais variantes que ocorrem no mesmo trecho de sua fala/escrita (Tagliamonte, 2006).

***Nós sem -mos:***

- (4) Inf.: perto da minha vó era um terreno SÓ... né... ***nós morava*** num terreno grande... na frente ficava na casa da minha avó e atrás ficava a casa dos meus pais (Inq. 29 – F2 – DID, mulher, 62 anos)

***A gente sem -mos:***

- (5) Inf.: nossa maior barreira que ***a gente tem***... é a queimadura... principalmente a de terceiro grau destrói totalmente essa barreira ***a gente fica*** totalmente vulnerável... é em termo de comparação (Inq. 50 – F1 – DID, homem, 29 anos)

***A gente com -mos:***

- (6) Inf.: a nossa mente é uma folha em branca ***a gente somos*** um nem bom nem mau vai depender das nossas experiências e das nossas escoLHAS (Inq. 79 – F2 – EF, mulher, 36 anos)

Para a análise estatística, foi empregada a linguagem de programação *R*, por meio de ambiente *R Studio* (R Core Team, 2021), que fornece as frequências e proporções gerais das variantes em função de cada variável preditora, além de plotar gráficos representativos dos resultados em análise, e criar um modelo de regressão logística, por meio do qual analisamos os fenômenos linguísticos à luz da estatística inferencial, que nos forneceu os contextos sociais e linguísticos mais propensos, em termos de favorecimento, ao emprego das variantes.

Este estudo adota a metodologia de pesquisa em tempo real de curta duração em tendência (Labov, 2008; 1994) ao comparar amostras da fala fortalezense em duas sincronias, década de 1990 do século XX e década de 2020 do século XXI. Uma das formas de compreender os padrões de variação que caracterizam a comunidade de fala de Fortaleza é investigar os movimentos de mudança no seu curso de implementação, comparando duas épocas distintas, observando se o fenômeno apresenta indícios de mudança linguística. Nesse sentido, buscamos observar o estágio da variação e mudança em progresso já constatado em estudos sociolinguísticos (Lopes, 2003; Araujo, 2016; Souza, 2020; entre outros).

No Brasil, diversos trabalhos no ramo da sociolinguística são realizados com o objetivo de encontrar respostas sobre como determinada mudança linguística ocorre em uma comunidade de fala. Muitos desses estudos trazem contribuições para o conhecimento sobre fenômenos variáveis no nível morfosintático, que abrange os pronomes e as suas formas variantes. Nesses trabalhos, são explanadas questões pertinentes sobre o tratamento da variação,

como os fatores linguísticos ou sociais que podem influenciar na realização de dado fenômeno, na tentativa de construir um caminho mais consistente e fundamentado para a análise e descrição da língua.

Algumas pesquisas merecem destaque por analisar os pronomes de primeira pessoa do plural em tempo real, considerando duas décadas distintas. Como dito anteriormente, em nosso país, temos ciência de poucos *corpora* que permitem o estudo a partir de duas sincronias (o Programa de Estudos sobre o Uso da Língua - PEUL, o Projeto Norma Oral Culta – NURC, o Projeto Censo da Variação Linguística do Rio de Janeiro - CENSO, e o Projeto Descrição do Português Oral Culto de Fortaleza-CE, PORCUFORT em suas Fases I e II) e, por isso, estudos em tempo real sobre a realização dos pronomes *nós* e *a gente* com suas respectivas concordâncias verbais ainda são escassos.

Dentre esses trabalhos podemos destacar a pesquisa de Lopes (2003) e Nascimento (2013) quando analisaram as formas variáveis *nós* e *a gente* em amostras do Projeto NURC – Norma Urbana Oral Culta – nas cidades do Rio de Janeiro e Salvador, com amostra das décadas de 1970 e 1990. Ainda, mais recentemente, merece destaque estudo realizado por Freitas, Rodrigues e Santos (2021), publicado no livro intitulado *O Falar Culto de Fortaleza em Foco* (Araújo; Viana; Rodrigues, 2021), obra que traz uma rica coletânea de textos com estudos realizados com dados de fala provenientes do Projeto Descrição do Português Oral Culto de Fortaleza – PORCUFORT. No capítulo, os autores tratam de analisar a variação entre pronomes *nós* e *a gente* em duas sincronias, apresentando dados importantes sobre o comportamento do fenômeno na capital cearense nessas duas décadas. Vale destacar, no entanto, que o trabalho em questão analisa apenas a *alternância pronominal*, diferentemente desta tese, que além de analisar a *alternância* entre os pronomes, trata de analisar também os padrões de *concordância verbal de primeira pessoa do plural*.

A partir do que foi levantado, nos propomos analisar em tempo real a realização dos pronomes *nós* e *a gente*, bem como analisar os padrões de *concordância verbal de primeira pessoa do plural*, na fala culta de Fortaleza com dados provenientes do projeto PORCUFORT Fase I (década de 1990) e Fase II (década de 2020). Para tanto, adotamos o modelo teórico-metodológico da sociolinguística laboviana, que procura sistematizar os dados de fala, fazer a descrição da estrutura da língua dentro de determinada comunidade de fala, para estabelecer relações entre os fatores linguísticos e extralinguísticos que conduzem o processo de variação linguística. Para isso, nos apoiamos nos pressupostos teórico-metodológicos descritos pela Teoria da Variação e da Mudança Linguística (Weinreich; Labov; Herzog, [1968]2006).

Destacamos que as análises pautadas sob a luz da teoria variacionista podem ser

realizadas por meio de dois parâmetros metodológicos possíveis: análises com fenômenos variáveis a partir da observação de uma sincronia, denominados de estudo em tempo aparente; e as análises que comparam duas sincronias, chamado de estudo em tempo real.

Nas análises de tempo aparente, os dados obtidos na pesquisa não são comparados entre duas sincronias distintas. É feito um recorte no tempo em que os dados serão analisados dentro de um mesmo intervalo de tempo, buscando indícios de mudança linguística. Para isso, a variável *faixa etária* exerce uma função muito importante, pois, para Labov (2008), os informantes com maior faixa etária apresentam um comportamento linguístico que representa o falar da geração anterior e seu padrão de uso tende a desaparecer com eles, enquanto os mais jovens tendem a apresentar um comportamento linguístico que reflete o padrão de uso da geração atual e levarão esse padrão com eles no futuro. Portanto, analisar o falar de pessoas de idades diferentes em tempo aparente, ou seja, em uma determinada sincronia, pode revelar os diferentes estágios da língua, como uma possível mudança à vista. No entanto, o pesquisador necessita de atenção, pois nem toda característica linguística está relacionado à mudança, por isso é “preciso distinguir as diferenças etárias que indicam mudanças lingüísticas daquelas diferenças que simplesmente caracterizam a linguagem de jovens e velhos e se repetem em qualquer geração” (Paiva & Silva, 1998, P.353).

Já as análises em tempo real, por outro lado, envolvem o estudo da mesma comunidade de fala em dois ou mais momentos distintos no tempo. Isso permite observar diretamente as mudanças que ocorreram ao longo do tempo. Paiva e Duarte (2004) acreditam que um intervalo de uma geração já pode ser suficiente para observar o estágio de variação e mudança linguística de um fenômeno dentro de uma comunidade. Esse método pode revelar mudanças em progresso que não são evidentes na análise em tempo aparente, como aquelas que ocorrem gradualmente ou que afetam toda a comunidade de maneira uniforme.

A realização de uma análise em tempo real da variação na *alternância* e na *concordância verbal* entre os pronomes *nós* e *a gente* se justifica pela obtenção de dados importantes para a compreensão dos fenômenos, pois, além de outras questões, é possível averiguar se, na fala culta de Fortaleza, os fenômenos encontram-se em um processo de mudança em curso ou se configuram como casos de variação estável, nos termos de Labov (2008). Dados como esses são de extrema importância para a pesquisa sociolinguística, pois esta é uma das questões que dão norte a muitos trabalhos de cunho variacionista desenvolvidos no Brasil e no mundo.

Neste trabalho, como explicitado, tratamos da análise de processos distintos que são faces de um mesmo fenômeno, tendo em vista que a escolha entre os pronomes *nós* e *a gente*

influencia a concordância com o verbo. Por conta disso, trabalhamos com duas variáveis respondidas distintas. Inicialmente, apresentaremos as questões de pesquisa relacionadas à alternância entre as formas *nós* e *a gente*. Partimos da hipótese geral de que o pronome *a gente* é mais frequente que o pronome *nós* em nossa amostra de fala culta fortalezense tanto na década de 1990 quanto na década de 2020, tendo maior proporção de uso na década mais atual, como demonstram os trabalhos de Lopes (2003), Omena (2003) e Nascimento (2013). A partir dessas observações, percebemos que esta proposta de estudo sobre a alternância pronominal converge para os seguintes problemas:

- Qual a proporção de uso dos pronomes sujeitos *nós* e *a gente* na amostra do PORCUFORT Fase I (década de 1990)?
- Qual a proporção dos pronomes sujeitos (*nós* ou *a gente*) na amostra do PORCUFORT Fase II (década de 2020)?
- Em que medida os grupos de fatores linguísticos, a saber, *grau de referencialidade do pronome, tipo de verbo, paralelismo formal, tempo verbal e saliência fônica*, influenciam a *alternância pronominal de primeira pessoa do plural* nas décadas de 1990 e de 2020?
- Em que medida os fatores dentre as variáveis sociais *sexo/gênero, faixa-etária, tipo de inquérito e década de gravação* atuam sobre a realização dos pronomes de primeira pessoa do plural *nós* e *a gente* nas amostras de 1990 e de 2020?
- O fenômeno linguístico em questão apresenta-se como uma variação estável ou em um processo de uma mudança linguística em progresso, no sentido de *a gente* estar substituindo *nós* na comunidade analisada?
- Houve aumento ou diminuição na proporção de uso das formas pronominais *nós* e *a gente* de uma década para outra?

Para a variável resposta *padrões de concordância verbal de primeira pessoa do plural*, nossa hipótese é a de que o padrão emergente, *a gente* sem *-mos*, é mais recorrente em ambas as amostras, devido a sua relação direta com *os padrões de alternância pronominal*. Em seguida, o pronome padrão com a concordância padrão (*nós* com *-mos*) seria usado em grande proporção entre os falantes cultos fortalezenses, sendo as formas não padrão (*nós* sem *-mos* e *a gente* com *-mos*) pouco recorrentes, seguindo a tendência dos dados de fala de Fortaleza e de amostras do Português brasileiro (Scherre et al., 2018; Carvalho et al., 2021). Nesse sentido, a análise da *concordância verbal de primeira pessoa do plural* aponta para as seguintes questões:

- Qual a proporção da *concordância verbal* padrão e não padrão com sujeito explícito de primeira pessoa do plural em amostras das décadas de 1990 e de 2020?

- Quais e como as variáveis de ordem linguísticas (*tonicidade do verbo, conjugação, estrutura do verbo e tempo e tipo de paradigma verbal*) condicionam os padrões de concordância verbal com os pronomes *nós* e *a gente*?

- Quais e como as variáveis previsoras de ordem extralinguística entre *sexo/gênero, faixa-etária, tipo de inquirido e década de gravação* atuam sobre a *concordância verbal de primeira pessoa* com os pronomes *nós* e *a gente* nas amostras representativas das décadas de 1990 e de 2020?

- O fenômeno linguístico da concordância verbal apresenta-se como uma variação estável ou existem sinais de uma mudança em progresso?

Dadas as questões de pesquisa que conduziram este estudo e os objetivos geral e específicos sobre cada fenômeno variável, fomos guiados, durante os trabalhos desenvolvidos nesta tese, por hipóteses distintas para cada uma das variáveis resposta analisadas. Primeiro, em relação a *alternância* entre os pronomes *nós* e *a gente*, levantamos as seguintes hipóteses secundárias sobre a variação *nós* e *a gente* a partir de amostras do PORCUFORT da década de 1990 e de 2020:

- a) A forma pronominal *a gente* é mais produtiva que a forma *nós* entre os fortalezenses de nível superior da década de 1990 (PORCUFORT Fase I), seguindo a tendência dos estudos realizados em outras partes do país (Omena, 2003; Tamanine, 2010; Nascimento, 2013).
- b) Na década de 2020 (PORCUFORT Fase II) há um aumento significativo na proporção de uso da forma inovadora *a gente* comparada à década mais antiga, dada a forte expansão do pronome *a gente*, como observado em outras capitais (Lopes, 2003; Nascimento, 2013).
- c) No grupo de fatores *faixa-etária*, os informantes de faixa etária III (*com mais de 50 anos*) usam mais a variante padrão *nós*, enquanto os informantes mais jovens (*22 a 35 anos*) utilizam a variante inovadora *a gente* em maior proporção, conforme prevê a teoria laboviana (Labov, 2008 [1972]).
- d) No grupo de fatores *sexo/gênero*, as mulheres favorecem o uso de *a gente*, ao contrário dos homens, que beneficiam a forma *nós*, tendo em vista que a hipótese laboviana do Paradoxo do Gênero (Labov, 1990, 2001) prevê as mulheres sendo mais receptivas as

formas inovadoras.

- e) No que se refere à *década de registro*, em ambas as décadas há maior recorrência do emprego de *a gente*, sendo mais expressiva na década de 2020.
- f) Na variável *tipo de inquérito*, *a gente* é preferido em inquéritos menos formais (D2), enquanto o pronome *nós* é mais recorrente em registros mais formais (EF), já que ambientes menos formais são mais propícios à realização de variantes inovadoras (Coelho et al, 2015)
- g) Existe uma mudança linguística em curso, no sentido de *a gente* estar substituindo o pronome *nós* em todos os contextos.
- h) No grupo de fatores *tempo verbal*, os verbos expressos no pretérito imperfeito beneficiam a forma pronominal *a gente*, dada a natureza dos inquéritos, que recorre a narrativas de experiências pessoais (contexto fértil para esse tempo verbal).
- i) Na variável *paralelismo formal*, o uso do pronome *a gente* será beneficiado em contextos de manutenção do mesmo pronome, tendo em vista que marcas levam a marcas e zeros levam a zeros (Scherre e Naro, 1997).
- j) Na variável *Grau de referencialidade do pronome*, a forma pronominal *a gente* é mais produtiva em contextos de uso no sentido genérico, enquanto o pronome *nós* é mais usado com sentido específico, mantendo suas funções prototípicas.
- k) No grupo de fatores *Saliência fônica*, os verbos menos salientes condicionam a utilização de *a gente*, já os verbos com maior diferenciação fônica privilegiam o uso de *nós*.

Em relação a segunda variável resposta analisada neste estudo, os *padrões de concordância verbal de primeira pessoa do plural*, foram levantadas algumas hipóteses sobre a atuação de cada preditor de ordem linguística e social testada neste estudo sobre a realização de cada variante, são elas:

- a) O padrão emergente de concordância na língua *a gente* com *-mos* (*a gente canta*) é o mais recorrente entre as formas de concordância verbal em nossas amostras de 1990 e de 2020, seguindo a tendência atestada em outros estudos (Zilles, Maya e Silva, 2000; Vianna, 2006; Rubio, 2013).
- b) No grupo de fatores *sexo/gênero*, tanto homens quanto mulheres tendem a utilizar a o padrão emergente e a concordância padrão em maior proporção (*a gente* sem *-mos* e *nós* com *-mos*).

- c) Por se tratar de amostras de fala culta, há pouca recorrência de realizações de concordância não padrão, tendo em vista que as formas *a gente* com *-mos* (*a gente cantamos*) e *nós* sem *-mos* (*nós canta*) são consideradas informais e passíveis de estigma (Vitório, 2015; Freitas et al, 2020).
- d) Em relação a *tonicidade do verbo*, há favorecimento da forma *a gente* sem *-mos* com monossílabos e oxítonos, enquanto as formas paroxítonas privilegiam a concordância padrão *nós* com *-mos* (Borges, 2004; Tamanine, 2010; Franceschini, 2011).
- e) Na variável *tempo e tipo de paradigma verbal do modo indicativo*, há favorecimento do pronome *a gente* com *-mos* em verbos que possuem a mesma morfologia na relação presente/passado, uma vez que o uso da forma emergente (*a gente* com *-mos*) resolveria o problema da ambiguidade temporal entre esses verbos (Scherre et al, 2018)

Com o objetivo de buscar respostas para as questões que foram levantadas nesta seção, e de testar as hipóteses para cada variável, elegemos como objetivo geral desta tese analisar, em tempo real de curta duração, a variação do uso dos pronomes pessoais *nós* e *a gente* e suas respectivas concordâncias padrão e não padrão no falar culto de Fortaleza-CE à luz da teoria da variação e mudança linguística. A partir desse objetivo geral, formulamos os seguintes objetivos específicos:

- Verificar qual forma pronominal, *nós* ou *a gente*, é mais produtiva nas amostras dos *corpora* PORCUFORT em cada amostra analisada (1990 e 2020).
- Analisar qual *padrão de concordância verbal de primeira pessoa do plural* é mais fértil na fala culta de Fortaleza em duas sincronias distintas.
- Verificar quais e como condicionadores linguísticos (*grau de referencialidade do pronome, tipo de verbo, paralelismo formal, tempo verbal e saliência fônica*) atuam sobre a alternância pronominal entre *nós* e *a gente* em amostras dos *corpora* PORCUFORT em duas sincronias.
- Averiguar quais e como variáveis de ordem linguísticas dentre *tonicidade do verbo, conjugação verbal, estrutura do verbo e tempo e tipo de paradigma verbal* interferem significativamente na concordância verbal com os pronomes de primeira pessoa do plural.
- Investigar as variáveis extralinguísticas (*sexo/gênero, faixa etária, tipo de inquérito e década de gravação*) que influenciam os padrões de *alternância pronominal* entre os pronomes *nós* e *a gente* em amostras do PORCUFORT Fase I (Amostra 1990) e Fase II (Amostra 2020);
- Observar as variáveis previsoras de ordem extralinguísticas (*sexo/gênero, faixa etária,*

*tipo de inquérito e década de gravação*) que influenciam os padrões de *concordância verbal* com os pronomes *nós* e *a gente* em amostras da década 1990 e da década de 2020;

- Examinar, a partir de amostras de PORCUFORT das décadas de 1990 e 2020, se o fenômeno de *alternância* entre os pronomes *nós* e *a gente* se encontra em processo de mudança em curso ou se figuram como casos de variação estável.

- Analisar, com dados advindos das décadas de 1990 e 2020, se o fenômeno da *concordância verbal* com pronomes *nós* e *a gente* se apresenta como um fenômeno em variação estável ou se há indícios de uma mudança em curto.

Essas questões, hipóteses e objetivos foram investigadas com a análise dos dados de fala em situações concretas de uso da língua, a partir das conversações obtidas por meio de entrevistas promovidas. Neste sentido, este trabalho fornece uma caracterização do falar culto de Fortaleza em duas sincronias distintas, podendo contribuir para o ensino de língua materna, se levarmos em consideração o conceito de competência comunicativa, desenvolvido por Hymes (1971). Tal conceito, imprescindível para a construção de metodologias para o ensino de línguas, promove a ideia de que, na aquisição de um sistema gramatical, a criança incorpora também o sistema de uso relacionado a pessoas, lugares, objetivos e outras formas de comunicação juntamente com atitudes e crenças a eles relacionados. Assim, o ensino de língua materna, de acordo com a BNCC (Base Nacional Comum Curricular), deverá promover a discussão acerca do eixo USO => REFLEXÃO => USO, considerando a existência e a possibilidade da realização de uma forma pronominal de primeira pessoa do plural que não se encontra presente no quadro pronominal proposto pelas gramáticas: a forma *a gente*.

Se pensarmos no contexto pedagógico, o espaço da sala de aula consiste em um rico ambiente de ecologia linguística (Mufwene, 2004). São diversos sujeitos, cada qual com sua cultura, sua história de vida e suas formas de expressão linguística, que estão em contato e se reestruturam a todo o momento, através das interações. Todos esses aspectos sociais e culturais refletem na linguagem e nas formas variáveis compartilhadas entre membros de determinada comunidade. Essas formas nem sempre correspondem ao padrão linguístico estabelecido pela norma escolar e acabam, muitas vezes, não sendo exploradas em sala de aula.

O que se observa ao analisar vários estudos descritivos do comportamento da língua portuguesa no Brasil é que existe um espaço em branco bastante expressivo entre a prática pedagógica do ensino de língua materna e a Base Nacional Comum Curricular (BNCC), principalmente no que diz respeito ao estudo da variação linguística. Uma das habilidades e competências descritas pelo próprio BNCC aborda a necessidade de o aluno refletir e entender

as variações linguísticas como “inerentes a qualquer sistema linguístico” (Brasil, 2018, p. 81), devendo a questão da variação ser abordada em sala, conforme descreve o documento:

Analisar o fenômeno da variação linguística, em seus diferentes níveis (variações fonético-fonológica, lexical, sintática, semântica e estilístico-pragmática) e em suas diferentes dimensões (regional, histórica, social, situacional, ocupacional, etária etc.), de forma a ampliar a compreensão sobre a natureza viva e dinâmica da língua e sobre o fenômeno da constituição de variedades linguísticas de prestígio e estigmatizadas, e a fundamentar o respeito às variedades linguísticas e o combate a preconceitos linguísticos. (Brasil, 2018 p. 82).

Sendo a variação inerente às línguas humanas, abre-se uma questão que a sociolinguística pretende oferecer material para discussão e que, também, justifica esse trabalho: Se a própria BNCC reconhece a natureza viva e dinâmica das línguas, tratando da importância de se analisar as variações em seus diferentes níveis, por que as escolas e os gramáticos (pelo menos uma grande parte deles) insistem em categorizar, por exemplo, o quadro pronominal do português brasileiro de forma padrão e fixa, não passível de variação, tendo em vista que diversos trabalhos variacionistas atestam que a realidade linguística não é invariável?

No que diz respeito aos capítulos destinados aos pronomes, os livros didáticos, com raras exceções, não citam o pronome *a gente* como uma forma válida de expressão à primeira pessoa do plural, mesmo sendo empregado frequentemente com valor pronominal em exemplos de textos nos próprios livros. Por exemplo, em Cunha e Cintra (2001), a forma *a gente* é empregada como uma representação da primeira pessoa pronominal, com diferentes graus de referencialidade, em trechos de obras literárias:

- (I) Houve um momento entre *nós* / Em que *a gente* não falou. (Fernando Pessoa) (pág. 296, grifos nossos)
- (II) Você não calcula o que é *a gente* ser perseguida pelos homens. Todos me olham como se quisessem devorar-me. (Ciro dos Anjos) (pág. 296, grifo nosso)

Para além das reflexões pedagógicas, os resultados deste trabalho mostram-se relevantes para outras áreas do conhecimento científico, como a Linguística Forense, pois os estudos sociolinguísticos podem estar à disposição da justiça para auxiliar no que diz respeito a investigações criminais. Trabalhos como este desempenham papel importante na Linguística Forense, tendo em vista que fornecem uma descrição detalhada das características linguísticas de um indivíduo ou de um grupo, auxiliando na criação de um perfil sociolinguístico que pode ser usado para atribuir autoria a um discurso gravado ou a um texto e, dessa maneira, ter mais

um recurso para fundamentar uma investigação, se valendo da caracterização do falar dessa região, para auxiliar, por exemplo, na identificação de pessoas ou de suspeitos (Sousa-Silva e Coulthard, 2016).

De acordo com Gonçalves & Brescancini (2014, pág. 68), por exemplo, uma das técnicas realizadas pela Linguística Forense para a identificação de um interlocutor é proceder a comparação de uma “amostra de fala de um indivíduo desconhecido com um grupo de amostras de fala pertencentes a locutores de identidade sabida, reunidas para fins de confronto ou pertencentes a um determinado banco de dados de produções orais”. Para os autores, a Linguística Forense acredita que o uso da língua é moldado por uma série de variáveis, como região geográfica, gênero, etnia, idade, grau de instrução, formação acadêmica e profissão, e que é possível traçar um perfil sociolinguístico de um indivíduo ao analisar sua fala. Desta forma, Marengo, Souza e Fonseca (2019, p. 149) asseveram que “como a interação legal ocorre em ambientes sociais especializados, ricos em combinações dessas e de outras variáveis, ratificamos a grande importância da sociolinguística para a Linguística Forense”.

Apresentadas as hipóteses, para termos uma melhor organização retórica, esta tese está dividida em 7 capítulos. No capítulo 1 temos esta introdução, da qual foram tratadas questões sobre o nosso objeto de estudo, bem como as problemáticas, questões norteadoras e hipóteses levantadas no decorrer do trabalho. O capítulo 2 é destinada à fundamentação teórica que sustenta esta pesquisa. Nesta seção, refletimos sobre as principais questões presentes nos procedimentos teórico-metodológicos da Teoria da Variação e Mudança Linguística. No capítulo 3, apresentamos um breve histórico sobre a inserção do pronome *a gente* no quadro pronominal do português brasileiro ao longo do tempo, assim como uma revisão da literatura em que descrevemos os principais trabalhos relacionados a variação na *alternância* e na *concordância verbal de primeira pessoa do plural*. O capítulo 4 compreende os aspectos metodológicos empregados na realização deste trabalho, com questões pertinentes sobre a amostra e as variáveis preditoras controladas, o procedimento de coleta dos dados e a ferramenta estatística utilizada. No capítulo seguinte, temos os resultados referentes à análise estatística realizada por meio do ambiente *R Studio*, onde são descritos os principais fatores que atuam sobre a variação na *alternância pronominal* de primeira pessoa do plural. O capítulo 6 apresenta a análise e a descrição dos resultados referentes aos dados de *concordância verbal* com os pronomes *nós* e *a gente*. O capítulo 7 é dedicado à conclusão sobre o resultado da análise dos pronomes *nós* e *a gente* e suas respectivas concordâncias na fala culta de Fortaleza, em que destacamos os principais achados desse estudo.

## 2 TEORIA DA VARIAÇÃO E MUDANÇA LINGUÍSTICA

A Teoria da Variação e Mudança Linguística (Weinreich; Labov; Herzog, 2006 [1968], Labov, 2001, 2006 [1966], 2008 [1972], 2010), também conhecida como Sociolinguística Quantitativa, ou ainda Sociolinguística Variacionista e, ainda, Sociolinguística Laboviana, nasceu a partir da publicação, em 1968, de um conjunto de textos intitulados *Directons for Historical Linguistic: a symposium*. Entre esses textos, estava um ensaio realizado para uma conferência na Universidade do Texas (EUA) denominado de *Empirical foundations for a theory of language change*, escrito por Uriel Weinreinch, William Labov e Mavin Herzog, a partir do qual começou a se desenhar a nova teoria.

Contra-pondo-se a visão estruturalista difundida por Saussure (2002) de que a língua deveria ser concebida como um sistema homogêneo, ideal e abstrato, essa teoria objetivava analisar como os fatores linguísticos e, principalmente, os fatores sociais exercem influência sobre a realização de diversos fenômenos variáveis na língua. Dessa forma, a proposta da teoria da variação e da mudança linguística defendia que a língua era um sistema heterogêneo, colocando em destaque a comunidade de fala e a língua em situações reais de uso como objeto de estudo.

Com isso, uma enorme quantidade de pesquisas desenvolvidas sob o olhar da teoria sociolinguística variacionista atestaram que a variação linguística não acontece como fruto do acaso ou da escolha arbitrária do falante, nem é entendida como um caos na língua, mas é decorrente de um processo ordenado e sistematizado. Desta maneira, os princípios que norteiam a teoria pressupõem a existência de um processo de variação linguística que é inerente ao próprio sistema linguístico, além de ocorrer de forma ordenada, conforme explica Camacho (2017, p. 464) “essa variabilidade inerente não é aleatória nem casual, pois mantém uma correlação sistemática e regular com fatores linguísticos e sociais”

Ao dar destaque à relação que existe entre língua e sociedade, Labov percebeu as regularidades existentes entre a comunidade de fala e o uso da variação linguística por essa comunidade. Essa percepção pode ser atestada em dois de seus estudos que foram de extrema importância para o desenvolvimento de uma teoria que fosse capaz de dar conta da relação *língua x comunidade de fala*, a saber: a pesquisa sobre a realização dos ditongos /ay/ e /aw/ na ilha de Martha’s Vineyard e o estudo sobre a estratificação social do /R/ nas lojas de departamento de Nova Iorque.

Labov (1972) desenvolveu um estudo sobre a relação entre a centralização dos ditongos /ay/ e /aw/ e a comunidade de fala de Martha’s Vineyard e constatou que, na ilha, os

moradores locais apresentavam uma preferência pelo uso da variante não padrão, enquanto os turistas manifestavam uma tendência maior pela utilização da variante padrão. Com isso, o linguista concluiu que o uso mais frequente dessa variante é um traço local da comunidade de fala, como um aspecto que os identifica como pertencentes a um mesmo grupo social, ou seja, uma marca cultural dos moradores. Sobre essa questão, observamos em Tarallo (1999) que:

É evidente que a centralização do ditongo em Martha's Vineyard é somente um dos traços linguísticos que definem a língua falada na ilha. Os exemplos relatados sugerem, portanto, que a língua pode ser um fator extremamente importante na identificação de grupos, em sua configuração, como também uma possível maneira de demarcar diferenças sociais no seio de uma comunidade. (Tarallo, 1990, p.15).

Outro aspecto importante observado por Labov (1972) se refere ao estigma e ao prestígio que algumas variantes linguísticas possuem. Em sua pesquisa desenvolvida nas lojas de departamento de Nova Iorque sobre a realização do fonema /R/, o autor observou que a presença do fonema na fala dos nova-iorquinos é prestigiada, enquanto a sua ausência apresenta estigma social. Ainda, o trabalho de Labov constatou que a utilização do /R/ estava relacionada ao status social mais elevado. Com isso, o autor pôde comprovar a relação intrínseca entre língua e estrato social, além de atestar que a língua é essencialmente heterogênea.

Como resultado, em 1968, Uriel Weinreich, William Labov e Marvin Herzog publicaram a obra *Fundamentos Empíricos para uma Teoria da Mudança Linguística*, formulando uma nova teoria linguística. O objetivo deste livro, além de apresentar um novo arcabouço teórico, era orientar os procedimentos para o desenvolvimento de novas pesquisas, considerando a variação própria de cada língua, respaldando-se na ideia de heterogeneidade sistemática, e destacando, entre outras coisas, a variação inerente ao objeto. Assim, na perspectiva dos autores, não existe falante de estilo único, pois a variação está presente na língua de todos os falantes e esses possuem competência linguística para manejar a heterogeneidade, já que o caráter heterogêneo da língua não interfere no seu funcionamento. Ao contrário, a existência de um único estilo de fala, segundo os autores, comprometeria a funcionalidade da língua, conforme podemos observar:

A chave para uma concepção racional da mudança linguística – e mais, da própria língua – é a possibilidade de descrever a diferenciação ordenada numa língua que serve a uma comunidade. [...] Um dos corolários de nossa abordagem é que numa língua que serve a uma comunidade complexa, a ausência de heterogeneidade estruturada é que seria disfuncional. (Weinreich, Labov & Herzog, 2006, p. 36)

A partir disso, é importante elencarmos alguns pressupostos básicos que desenham a Teoria da Variação e Mudança Linguística. São eles:

- a) A língua é concebida como um sistema heterogêneo e ordenado;
- b) A variação linguística é inerente ao sistema da língua;
- c) Os falantes possuem competência linguística para comportar a heterogeneidade presente no sistema;
- d) Todos os falantes apresentam variação, seja ela de cunho fonológico, morfológico, sintático ou semântico, pois não existe falantes de estilo único;
- e) Fatores de ordem linguística e social exercem influência sobre a realização da variação;
- f) Nem toda variação corresponde a mudança linguística, mas toda mudança pressupõe variação (Coelho et al, 2015);

Considerando os pontos destacados acima, é importante também considerar que a teoria da variação e da mudança linguística apresenta dois modelos de análises dos fenômenos variáveis em uma comunidade de falar: as análises sincrônicas, denominadas de análise em tempo aparente; e as análises diacrônicas, nomeadas de análise em tempo real.

No que se refere a análise em tempo aparente, esse tipo de pesquisa traz uma abordagem sincrônica, que corresponde ao estudo realizado sobre o comportamento linguístico de falantes de uma determinada comunidade de fala em diferentes gerações, em um recorte no tempo. Com esse tipo de estudo, é possível identificar, através da variável faixa-etária, se o fenômeno variável estudado apresenta indícios de mudança. Conforme afirma Freitag (2005), a análise em tempo aparente “considera a distribuição das ocorrências do fenômeno em estudo em função das faixas etárias para caracterizar uma situação de estabilidade, mudança incipiente, mudança em progresso ou mudança completa” (Freitag, 2005, P. 110)

Já a análise em tempo real, em uma abordagem diacrônica, consiste na observação de uma mesma comunidade de fala em duas épocas diferentes. Desta forma, os dados sobre o fenômeno variável a ser estudado são coletados em duas décadas distintas e, com a ajuda de ferramentas computacionais que farão análises estatísticas, esses dados serão confrontados, fazendo uma comparação entre a frequência de uso nas duas décadas investigadas, com o objetivo de encontrar os sinais de mudanças. Assim, o estudo em tempo real é aquele que “compara amostras de fala dos mesmos indivíduos, ou amostras aleatórias da mesma comunidade de fala, mas com a mesma estratificação social, relativas a dois momentos diferentes” (Coelho; Görski; May; Souza, 2010, p. 164). Estudos em tempo real, ainda, podem ser realizados a partir de duas abordagens metodológicas, a saber, o *estudo em tendência* e *estudo em painel*.

Em um estudo em tempo real de tendência [*trend study*], os pesquisadores sociolinguistas coletam dados representativos de informantes aleatórios de uma mesma comunidade de fala, com a mesma estratificação social, em duas sincronias distintas. A análise de amostras de dois pontos temporais distintos permite a observação de aspectos não detectados pelo estudo de tempo aparente, pode fornecer estágios de mudança da comunidade de fala, a propagação e estabilização da mudança em progresso das variantes nas duas sincronias em estudo. O controle e o confronto entre sincronias distintas, com base em amostras representativas de fala ou escrita, permitem observar como as mudanças seguem uma tendência, uma propagação para o encaixamento e a implementação de formas inovadoras no sistema linguístico (Labov, 1994).

Por outro lado, o estudo em tempo real de painel [*painel study*] analisa dados de um mesmo grupo de falantes ao longo do tempo, com o objetivo de observar traços da mudança linguística em uma determinada comunidade e, para isso, o pesquisador precisa “entrevistar os mesmos informantes para proceder a uma análise comparativa dos dados” (Coelho et al, 2015, p. 128), tarefa demasiadamente difícil, visto a dificuldade de se contatar os mesmos informantes depois de intervalos de tempo relativamente longo. Os dados, nesse parâmetro metodológico, são coletados em intervalos regulares para monitorar e documentar as mudanças individuais e coletivas ocorridas em tempo real, ao longo de duas sincronias.

Em relação à mudança linguística, Labov (1994) estabelece uma correlação entre os estudos em tempo aparente e em tempo real, principalmente no que diz respeito à estabilidade ou instabilidade que determinado fenômeno manifesta em uma comunidade de fala. Para estabelecer essa relação, é observado o comportamento linguístico de uma comunidade como todo e do indivíduo em sua particularidade, confrontando o comportamento individual e coletivo da língua:

**Quadro 1:** Correlação entre indivíduo e comunidade em tempo real e aparente

	INDIVÍDUO	COMUNIDADE DE FALA
(1) Estabilidade	Estável	Estável
(2) Gradação etária	Instável	Estável
(3) Mudança geracional	Estável	Instável
(4) Mudança comunitária	Instável	Instável

**Fonte:** adaptado de Labov (1994, p. 83)

A teoria laboviana descreve quatro cenários diferentes que podem acomodar a variação e a mudança linguística no comportamento de uma comunidade ou de um indivíduo. No cenário (1), tanto o indivíduo quanto a comunidade de fala apresentam padrões linguísticos que permanecem contrastantes ao longo dos anos, o que pode refletir forte coesão social, além de normas muito bem definidas dentro da comunidade como um todo, refletindo estabilidade no comportamento linguístico. Nesse caso, como não há mudança no comportamento linguístico, já que este é caracterizado pela estabilidade, não há variação a se analisar.

O segundo cenário descreve uma gradação etária variável, tanto no indivíduo quanto na comunidade, que pode indicar que os padrões linguísticos do indivíduo podem mudar ao longo do tempo, principalmente entre os mais jovens, onde é possível encontrar indícios mais claros de mudança no comportamento linguístico (Labov, 1994). A análise linguística em função da faixa etária, em geral, sugere que os mais jovens adotam formas linguísticas inovadoras que estão em competição com formas mais conservadoras, e esse comportamento linguístico, embora possa acompanhar o falante para a vida, sofre pressões normativas naturais devido ao estilo de vida, ao ingresso no mercado de trabalho e na faculdade, o que pode acarretar mudança no comportamento individual de determinado falante. No entanto, analisando a comunidade de fala como um todo, percebe-se um padrão razoavelmente estável, o que sugere que as mudanças linguísticas individuais ainda não interferem nas normas linguísticas compartilhadas pela comunidade em geral.

Já em relação ao padrão (3), caracterizado pela mudança geracional, o comportamento linguístico é estável no indivíduo e instável na comunidade de fala. Esse cenário ocorre quando uma nova geração - geralmente impulsionada pelos mais jovens - adotam formas linguísticas diferentes da geração anterior, o que resulta em mudança na comunidade de fala como um todo, demonstrando a instabilidade característica na mudança geracional da comunidade. Por outro lado, os indivíduos tendem a manter seus padrões linguísticos, apesar das novas formas incorporadas e difundidas pelas novas gerações. Por exemplo, os membros mais velhos da comunidade tendem a manter comportamento linguístico mais tradicional, apresentando preferência por formas mais conservadoras.

Por fim, a mudança comunitária reflete um padrão no qual toda a comunidade desencadeia a mudança linguística, sendo instáveis tanto o indivíduo quanto a comunidade de fala. Este tipo de mudança ocorre geralmente quando uma forma inovadora é implementada, se torna popular e se expande rapidamente entre os membros da comunidade, geralmente impulsionada por um fator externo ao sistema linguístico que exerce forte influência sobre sua realização, como a globalização, o advento da tecnologia e as novas formas que a acompanha.

Outro fator que pode ocasionar esse tipo de mudança está relacionado a mudanças demográficas bruscas, como o ingresso de grande número de imigrantes em uma comunidade e, com isso, a introdução de novas formas de comunicação.

Esses cenários demonstram a complexidade da variação linguística e descrevem o estágio da implementação de novas formas linguísticas que passam a ser compartilhadas pelos membros da comunidade, sendo influenciadas por fatores de cunho social, cultural, individual e até ideológico. A observação em tempo real desses padrões individuais e coletivos pode indicar, entre outras coisas, o estágio de uma possível transição ou a instabilidade nas normas linguísticas de uma comunidade, onde formas linguísticas inovadoras podem ser adotadas, ao passo que formas mais antigas podem cair em desuso, ou podem coexistir com a forma mais nova no sistema linguístico por um longo tempo, em um processo de variação estável. Todos esses estágios da observação linguística geram explicações valiosas sobre os padrões linguísticos variáveis da comunidade em estudo.

Além das noções de tempo real e tempo aparente, Weinreich, Labov e Herzog (2006 [1968]) apresentam cinco questões, conhecidas como os cinco problemas empíricos que o linguista deve tratar ao realizar uma pesquisa sociolinguística, são eles: o problema da restrição ou fatores condicionantes; o problema da transição; da avaliação; do encaixamento e da implementação. Dentre esses problemas postulados pela teoria desenvolvida por Labov, nos detemos mais detalhadamente ao problema dos fatores condicionantes, uma vez que nos propomos a analisar traços linguísticos e sociais dos pronomes de primeira pessoa do plural em duas sincronias, com o objetivo de buscar indícios de mudanças nesses traços.

Para Weireich, Labov & Herzog (2006[1968]), o primeiro problema versa sobre as condições que podem favorecer ou restringir a mudança linguística, ou seja, analisar quais os fatores sociais e linguísticos atuam para a manutenção de uma forma ou para o seu desaparecimento. Para isso, é necessário um olhar atento do investigador para que este identifique em quais contextos a variação se torna mais produtiva e quais fatores atuam sobre esse tipo de variação, na busca de compreender como esses fatores podem impedir ou, até mesmo, acelerar o processo de mudança linguística dentro de determinada comunidade de fala.

O segundo problema atestado pela teoria, o problema da transição, trata da maneira como uma mudança ocorre em uma língua dentro de uma comunidade de fala determinada, pois o processo de mudança não acontece de forma repentina. Para Weireich, Labov & Herzog (2006, p. 122) “entre quaisquer dois estágios observados de uma mudança em progresso, normalmente se tentaria descobrir o estágio interveniente que define a trilha pela qual a estrutura A evoluiu para a estrutura B”. Em outras palavras, inicialmente existem duas formas

variáveis que disputam a preferência dentro da comunidade de fala, até que uma das formas linguísticas, por frequência de uso, assumo o lugar da outra e complete a mudança. É sobre esse processo de mudança linguística de uma forma para outra que trata o problema empírico da transição.

O problema da avaliação diz respeito a atitude linguística que o falante possui sobre as formas variáveis existentes na comunidade de fala e como esse falante julga o uso de uma ou de outra forma linguística, ou seja, o problema versa sobre o tratamento social que é dado às variantes, no sentido de serem estigmatizadas ou prestigiadas. Para Coelho *et al.* (2010) essa atitude pode se manifestar em dois níveis distintos: o primeiro está relacionado à avaliação linguística, e o segundo, à avaliação social. Esses níveis, vale pontuar, são inter-relacionados, já que os membros de uma comunidade atribuem significado sociais às formas linguísticas, partilhando atitudes em relação à língua.

Segundo Weireich, Labov & Herzog (2006, p. 124) os “correlatos subjetivos da mudança são por natureza mais categóricos do que os padrões cambiantes do comportamento”, pois a comunidade de fala nem sempre possui a consciência de como a variação ocorre na estrutura linguística. Dessa forma, conhecer a consciência social dos falantes de determinada comunidade é de extrema importância para se compreender essa questão, pois, quanto maior for a consciência de que uma variante é, por exemplo, estigmatizada, menor será sua ocorrência e, com isso, um estigma atrelado a uma forma variável em um processo de mudança pode retardar ou frear esse processo.

Além dos conceitos relatados cima, temos nos postulados labovianos o problema do encaixamento. Esse questiona sobre como a mudança é encaixada na estrutura social e linguística. Como algumas mudanças afetam, por exemplo, os terrenos da gramática e sobre como se dá a mudança, ou seja, como ela se encaixa no corpo social, considerando as variáveis sociais (sexo, idade, classe social, escolaridade). Por exemplo, podemos observar que uma dada variante linguística está fortemente associada a uma variável social em um determinado contexto, mas não se percebe associação desta variante em outros contextos. Em seu estudo clássico sobre a estratificação social do /R/ em Nova Iorque, Labov observou que a realização do som rótico estava correlacionado à classe social mais elevada, enquanto o não rótico era associado à classe social menos elevada, contexto em que a variante parecia bem encaixada. Essa observação pode levar a explicações sobre como as variáveis sociais favorecem ou inibem determinados aspectos da variação, além de sugerir detalhes que ajudam a compreender a dinâmica da variação e mudança linguística em uma comunidade de fala.

Por fim, Weinreich, Labov e Herzog (2006 [1968]) apresentaram o problema da implementação. Esse problema busca analisar como o processo de mudança se concretiza na estrutura social, em quais contextos estruturais e em quais estratos sociais essa mudança ocorreu. Vale ressaltar que a análise desse problema só é possível quando uma forma linguística deixa de ser tratada como uma variante e assume um traço constante na comunidade de fala.

Com tudo isso, a Teoria da Variação e da Mudança Linguística tem proporcionado um modelo teórico-metodológico que serviu de sustentação para uma enorme profusão de estudos sociolinguísticos desenvolvidos no Brasil e no mundo inteiro, incluindo as centenas de pesquisas variacionistas que analisam fenômenos fonéticos, morfológicos, sintáticos, semânticos e discursivos e que são realizadas no Brasil com dados advindos do português brasileiro, português europeu e do português africano.

Tecidas algumas considerações sobre o referencial teórico no qual se sustentará a realização desta pesquisa, a seção que segue tratará dos procedimentos metodológicos que foram utilizados para construir nossa análise.

### 3 O FENÔMENO DA PRIMEIRA PESSOA DO PLURAL

Neste capítulo, apresentamos alguns dos principais pressupostos produzidos sobre o nosso objeto de estudo. Inicialmente, tratamos de explicar o processo de implementação da forma *a gente* no quadro de pronomes do português, recorrendo a trabalhos que descrevem o percurso histórico da forma pronominal na língua. Em seguida, revisamos alguns dos principais trabalhos desenvolvidos sobre o comportamento variável da *alternância pronominal* e da *concordância verbal de primeira pessoa do plural*, focando nos resultados encontrados nesses estudos para, posteriormente, compararmos com os resultados desta tese. Destacamos, em cada trabalho, os percentuais de uso de cada variante, o *corpus* utilizado e as variáveis consideradas relevantes, tratando de observar nesses trabalhos tendências que se repetem.

#### 3.1 A inserção de *a gente* no quadro dos pronomes do português

É incontestável que a alternância no uso entre os pronomes *nós* e *a gente* permeia a língua falada e se manifesta em todas as esferas sociais e em ampla variedade de contextos, sejam aqueles que exigem um monitoramento maior da fala ou aqueles em que o falar é espontâneo.

Contudo, essa intercambialidade entre os pronomes de primeira pessoa do plural nem sempre foi característica do português brasileiro. Conforme evidenciam Menon (1995), Omena (1996) e Lopes (2003), a variante de primeira pessoa do plural *a gente*, antes de ser empregada na fala com valor de pronome pessoal, passou por um processo de mudança, tendo sua origem na palavra *gente*, usada como substantivo para se referir a um grupo indeterminado de pessoas. Com a evolução natural da língua e os diversos efeitos que sobre ela agem, ocorreu um processo de “gramaticalização” impulsionado por razões linguísticas e sociais, para que fosse possível o uso da forma inovadora *a gente* em competição com a forma canônica *nós*.

Em *A referência variável da primeira pessoa do discurso no Plural*, Omena (1986) afirma que, no Português Brasileiro, a expressão *a gente* passou por um processo que a autora denominou de “pronominalização”, para que hoje pudesse assumir a forma de pronome e competir com a forma *nós*. Por outro lado, para Menon (1996), a forma *a gente* teria passado por um processo de “cristalização”, a exemplo do que aconteceu com outras expressões, até que fosse possível seu emprego como pronome. No entendimento da autora, as formas nominais usadas para expressar um sujeito indeterminado, devido a frequência de uso, acabaram se

tornando expressões cristalizadas: *a gente*, o cara, a galera, o povo. Esses itens teriam passado por um processo de gramaticalização.

O processo de gramaticalização é tratado por diversos autores, que trazem abordagens teóricas distintas acerca desse fenômeno, conforme é possível observar em Traugott & Heine (1991), Hopper (1991), Heine et al. (1991), Hopper & Traugott (1993), Lehmann (1995), Heine (2003). No entanto, apesar dos enfoques diversos dados por esses trabalhos sobre a gramaticalização, devido a própria complexidade do fenômeno, alguns pontos de contato estão presentes em todos esses estudos, e nos interessa para a análise, como a noção de mudança linguística, unidirecionalidade do processo, o *continuum* da gramaticalização e a ideia de que um item, se lexical, passa a ser gramatical; se já gramatical se torna ainda mais gramatical (Hopper, 1991).

Apesar da maioria desses autores fornecer parâmetros que consideram os vários aspectos da mudança linguística, principalmente no que diz respeito às modificações na estrutura da língua, os conceitos por eles tratados são melhor aplicados às formas que estão em um estágio avançado do processo de gramaticalização (Lehmann, 1995). Neste estudo, por propor estudar os parâmetros que procuram destacar a forma gramaticalizada *a gente*, forma em avançado processo de gramaticalização, decidimos nos deter aos pressupostos tratados por Hopper (1991) no texto *On some principles of grammaticalization*, que assume uma concepção de língua emergente, pressupondo que novas formas podem surgir a partir de velhas, reconhecendo, portanto, o caráter fluido da língua formas estabelece. Hopper (1991) propõe cinco princípios básicos que caracterizam o processo de gramaticalização, a saber, *layering*, divergência, especialização, persistência e decategorização. Nesse sentido, a forma inovadora *a gente* percorreu todas as etapas descritas nestes princípios até se gramaticalizar e assumir o papel de pronome pessoal.

O primeiro princípio apresentado por Hopper (1991), o princípio do *layering*, afirma que, dentro de um domínio funcional, novas “camadas” surgem continuamente e passam a coexistir com as camadas mais antigas. Isso é possível porque o surgimento de uma nova camada (ou variante, nos estudos Sociolinguísticos) não implica na eliminação das camadas mais antigas, ao contrário, elas interagem e integram-se dentro de uma mesma área funcional. Em outras palavras, uma variante surge no seio da comunidade de fala e vai ganhando espaço em certos contextos. Em determinado momento, as duas variantes passam a coexistir e a competir o uso entre os falantes, no momento de transição, em que as duas formas convivem dentro dessa comunidade até que uma mudança se complete.

Para Omena e Braga (1996, pág. 78) “o comportamento da forma *a gente* manifesta-se como um legítimo caso de *layering* no português falado”. Segundo as autoras, as formas *nós* e *a gente* coexistem dentro do sistema linguístico e apresentam, em determinados contextos, o mesmo valor funcional, ou seja, apresentam formas distintas para o mesmo valor referencial que disputam a preferência do falante dentro da comunidade. Os excertos (7) e (8), extraídos da base de dados em análise, exemplificam o uso das duas formas distintas que apresentam o mesmo valor referencial:

- (7) as escolas públicas estão MUIto mais vioLEntas::... do que antigamente né... HOje ***a gente*** vê nas escolas professores sendo ameaçados... (Inq. 56 – F2 – D2, mulher, 50 anos)
- (8) encontram bons advogados que ... têm ... sempre aquelas saídas pra ele e a lei... DÁ essa condição... ***nós*** não devemos falar mal das leis brasileiras... que... imagino que elas sejam baseada em estudo científico né? (Inq. 11 – F1 – D2, homem, 33 anos)

Conforme é possível observar nos exemplos (7) e (8), o uso intercambiável entre as formas *nós* e *a gente* empregadas nas sentenças é possível, uma vez que nos dois casos a forma pronominal tem o mesmo valor referencial (*EU + EU AMPLIADO; TODOS*). Sendo assim, as duas camadas, a inovadora e a conservadora, coexistem dentro do mesmo domínio funcional, sendo utilizadas para fazer referência à primeira pessoa do plural.

O segundo princípio proposto por Hopper (1991), o princípio da divergência, diz respeito a conservação da forma lexical (chamada de forma fonte) que deu origem ao item gramaticalizado, coexistindo dentro do mesmo sistema linguístico. Assim, esse item lexical pode figurar como um item autônomo e permanecer com suas propriedades originais ou pode, ainda, sofrer novo processo de gramaticalização. Ou seja, mesmo que apresente funções distintas, existe um item lexical (forma fonte) convivendo de forma autônoma com a sua forma gramaticalizada.

Em relação ao pronome *a gente*, esse sofreu alterações ao se gramaticalizar, já que se estratificou com o artigo definido e cristalizou a relação determinante-determinado. No entanto, o substantivo *gente*, que deu origem ao pronome inovador, permaneceu como um item lexical autônomo, sem sofrer qualquer alteração morfológica ou fonológica, e convive dentro do mesmo sistema linguístico com a forma a que deu origem, o pronome *a gente*. No exemplo (9), é possível observar o pronome *a gente* e o substantivo *gente* coexistindo no sistema linguístico:

- (9) MUIta *gente* não se lembra em quem votou para vereador... muita *gente* não cobra... aí começa com vereador... aí vai para deputa:do estadual federal e presidente [...] mas também faz pouco tempo... que *a gente* começou a votar MESmo NÉ... (Inq. 56 – F2 – D2, mulher, 50 anos).

Já o princípio da especialização, o terceiro elencado por Hopper (1991), trata da limitação entre as escolhas possíveis dado o estreitamento da variedade de escolhas, o que tornaria o uso de uma forma, em determinados contextos, quase que categórico, e um número menor de formas atinge significados semânticos mais gerais, a medida em que uma ocupa mais espaço em detrimento da outra devido ao processo de gramaticalização. Ou seja, esse princípio estabelece os pontos em que o sistema pronominal apresenta mais fragilidade, tornando a escolha entre uma das formas quase que obrigatória em alguns contextos. Uma das consequências da especialização é o aumento da frequência de uso da forma mais inovadora no processo de gramaticalização, como aconteceu com a forma pronominal *a gente* (Tamanine, 2010).

Para exemplificar esse conceito, tomemos como base a pesquisa de Araujo (2016) sobre a variação *nós* e *a gente* na fala popular de Fortaleza. Os resultados demonstram um percentual geral para o uso de *a gente* de 66%, contra 34% para o uso de *nós*, indicando que a variante inovadora é expressivamente mais utilizada na comunidade de fala do que a variante padrão, mas ainda é um ambiente de disputa entre as formas. Porém, se observarmos a distribuição desses resultados entre as possíveis funções sintáticas controladas pelo autor, claramente percebemos contextos em que a forma inovadora predomina sobre a forma canônica quase que categoricamente, conforme é possível ver na tabela que segue:

**Tabela 1:** Atuação da função sintática sobre o pronome *a gente* em Araujo (2016)

Fatores	Aplica/Total	%	Peso Relativo
Adjunto	17/ 18	94,4%	0,722
Sujeito	657/ 999	65,8%	0,535
Objeto direto	23/ 31	74,2%	0,225
Predicativo do sujeito	3/ 9	33,3%	0,105
Objeto Indireto	16/ 35	45,7%	0,058

**Fonte:** Adaptado de Araujo (2016, p. 85)

Conforme exposto na tabela 1, alguns contextos favorecem intensamente o emprego pronome *a gente*: o uso do pronome com função de adjunto é visto em 94,4% das ocorrências

de primeira pessoa do plural, quase uma obrigatoriedade de uso na amostra pesquisada quando a função é adjunto. Além do expressivo percentual de uso nessa função sintática, a probabilidade de o pronome inovador ser usado nesse contexto é alta, com peso relativo de 0,722, de acordo com os dados da tabela 1, o que demonstra uso praticamente categórico do pronome *a gente* nesta função sintática.

O princípio da persistência remete a manutenção de alguns traços semânticos intrínsecos que a forma gramaticalizada absorve e conserva da forma original (forma fonte). Essa preservação dos traços intrínsecos é notória nas fases iniciais e intermediárias da gramaticalização, mas podem desaparecer se o processo estiver avançado. É importante ressaltar que esses traços intrínsecos podem explicar certas restrições que a forma gramaticalizada sofre, conforme explicam Omena & Braga (1996) sobre o traço de indeterminação associado ao pronome inovador:

No caso do uso de *a gente* a persistência do traço indeterminador provoca certas restrições em seu uso. Enquanto o pronome *nós* admite ser modificado por quantificadores, numerais, especificadores enfim, o mesmo não se dá com a forma *a gente*. *Todo, cada um, nenhum* podem modificar *nós*; mas não *a gente*. (Omena & Braga, 1996. p. 80).

A pronominalização de *gente* > *a gente* exemplifica bem o princípio da persistência. A forma *a gente*, ao se gramaticalizar, não perdeu todas as propriedades formais do substantivo *gente*, tampouco absorveu todas as propriedades dos pronomes pessoais. Assim, o pronome *a gente* mantém da sua forma original o traço formal de terceira pessoa, sendo associado a verbos conjugados no singular, porém aciona uma interpretação semântica de primeira pessoa do plural. Além disso, o pronome *a gente* manteve o caráter coletivo, indeterminador, herança do substantivo *gente* (forma fonte). Os excertos (10) e (11), retirados de nossa amostra da fala culta fortalezense, exemplificam a manutenção do traço generalizador da forma *a gente*, herança de sua forma fonte, que pode atingir o grau máximo de indeterminação:

(10) e muitas vezes ***a gente*** num precisa NEM comprar um jornal porque sai a notícias toda no rádio numa veLOCidade TÃO GRANde (Inq. 11 – F1 – D2, homem, 70 anos)

(11) é lamentável ***a gente*** chegar o índice de desemprego na região metropolitana de Fortaleza... (Inq. 56 – F2 – D2, mulher, 50 anos)

Nos dados (10) e (11), o pronome *a gente* é empregado com seu sentido máximo de indeterminação, que é característica da forma que deu origem à expressão, sendo possível inclusive a sua substituição pelo clítico *se*. No entanto, vale lembrar, várias pesquisas atestam

que esse traço intrínseco de indeterminação está em processo de mudança, uma vez que a forma *a gente* está cada vez mais sendo empregada em sentido específico (Omena, 2003; Tamanine, 2010; Araujo, 2016).

Por fim, o princípio da decategorização diz respeito à tendência que as formas em gramaticalização têm de perder os marcadores morfológicos e os privilégios sintáticos típicos de sua forma fonte, assumindo características de categorias secundárias. O substantivo *gente* possui certos privilégios que a forma *a gente*, ao passar pelo processo de gramaticalização, neutralizou, como a possibilidade de admitir quantificadores, marcadores, flexões de número e gênero, entre outros. Vale ressaltar que, no caso da alternância pronominal, o pronome padrão *nós* goza desses privilégios, ao contrário do pronome inovador *a gente*, que tem seu uso associado a ambientes de informalidade. No entanto, a forma gramaticalizada assumiu características próprias dos pronomes, como a possibilidade de se relacionar com adjetivos tanto masculinos quanto femininos em estruturas predicativas. As ilustrações (12) e (13), são ocorrências do pronome padrão e da forma inovadora em contextos relacionados a discussão:

(12) nenhuma vaiDAde em dizer isso mas acidentalmente... a a::... **nós TRÊS** que tínhamos vindo nessa circunstância... eh SAÍmo/ nas primeiras colocações... a a:: EU assim.. por exemplo eu tirei o primeiro lugar (Inq. 46 – F1 – DID, homem, 48 anos).

(13) se eu perguntar pra alguém aqui o que é que comeu antes de ontem talvez a pessoa não perceba ... porque **a gente** não tava **atento** ah o que é que eu comi antes de ontem no meu almoço né? (Inq. 06 – F2 – EF, homem, 30 anos).

No excerto (12), é possível observar a presença de um quantificador associado à forma padrão *nós*, o numeral “três”, que, além de quantificar, torna o referente mais determinado, pois limita a sua interpretação a um número específico de referentes. Essa característica de admitir certos tipos de quantificadores foi inibida na forma inovadora *a gente* ao passar pelo processo de gramaticalização. Nesse sentido, a forma padrão admite a possibilidade de usar expressões do tipo “nós três”, mas a forma inovadora neutraliza a possibilidade de se ligar a numerais, não admitindo, por exemplo, a expressão “a gente três”, sob risco de se tornar agramatical. Por outro lado, o pronome *a gente*, após se gramaticalizar, passou a receber características da classe de pronomes, como a possibilidade de ser associado a adjetivos em estruturas predicativas, como demonstra o exemplo (13), em que o adjetivo *atento* se refere a forma pronominal *a gente*.

O processo de gramaticalização da forma pronominal *a gente* no português falado no Brasil é destacado em pesquisas sobre a alternância pronominal de primeira pessoa do plural. Menon (1996), por exemplo, assumindo a noção proposta por Hopper (1991), e considerando que o pronome inovador percorreu todos os princípios da gramaticalização, propõe um quadro representativo em que descreve o percurso histórico percorrido pela expressão *a gente*, partindo de uma locução nominal plena (LNP), até se gramaticalizar (Menon, 1996, p. 626), passando a ocupar espaço no quadro pronominal do português brasileiro:

LNPlena > LNEspecífica > LNInvariável > Pron. Indet. > Pron. Pes. 1 (P>S)  
 ...gente... > a gente > [a gente] > a gente > a gente = nós, eu

Na fase de *locução nominal plena*, descrita pela autora como a fase de autonomia do substantivo *gente*, Menon (1996) observa que o nome *gente* admite flexões (especialmente de número) e pode constituir uma locução nominal se vier acompanhado de determinantes, como artigos, demonstrativos, numerais, por exemplo. Dessa forma, segundo a pesquisadora, o substantivo *gente* “poderia ser seguido de construção relativa ou receber a concordância de predicativo, no gênero feminino” (p. 397). Assim, nessa fase, seriam possíveis construções do tipo “a gente”, “as gentes”, “aquelas gentes” ou “essa gente bonita”.

Já na fase de *locução nominal específica*, embora não embora não tenhamos conhecimento até então de um estudo mais aprofundado sobre a forma *a gente* em locuções específicas, é nessa fase que se perdem a variação de número e a possibilidade de concordar, como predicativo do sujeito, com adjetivos femininos. Dessa maneira, se inicia o uso distinto - e, conseqüentemente, com significados distintos - para o substantivo *gente* e para a locução nominal *a gente*, dando início a um processo de gramaticalização.

Como *locução nominal invariável*, a expressão *a gente* passou a permitir a inclusão do falante na indeterminação, até então não possível. Menon (1996) hipotetiza que essa inclusão do falante só foi possível quando a locução nominal específica perdeu a capacidade de flexão de número, mas manteve seu sentido genérico e, com isso, deu à locução invariável *a gente* o status de primeira pessoa do plural.

Por fim, na fase que corresponde a *Pronome Indeterminado* > *Pronome Pessoal*, a autora enfatiza a possibilidade do “eu” ser incluído no que ela chama de ‘genérico

indeterminador’ (Menon, pág. 627), possibilitando a interpretação do [eu + tu], [eu + ele], [eu + vários] ou [eu + todos], passando a ser possível interpretar a forma *a gente* como *nós*. Ao assumir a função de pronome de primeira pessoa do plural, a forma *a gente* absorve características típicas dos pronomes, como a possibilidade de designar diretamente uma das pessoas do discurso, e perde certos “privilégios” do substantivo, como a possibilidade de admitir flexão de gênero e de número. Nesse sentido, ao perder o traço de gênero, a concordância passa a ser dada de acordo com fatores extralinguísticos, com elementos contextuais, tendo em vista que o pronome *a gente* não possui gênero inerente: “a gente está bonita” ou “a gente é honesto”.

Outros autores buscaram respostas que expliquem a inserção de *a gente* no quadro pronominal do português brasileiro. Para resgatar a etimologia da palavra *gente*, Lopes (2003, pag. 9), em seu estudo sobre o processo de gramaticalização da forma *a gente*, afirma que “substantivo *gente* se origina do substantivo latino *gēns, gēntis*: ‘raça’, ‘família’, ‘tribo’, ‘o povo de um país, comarca ou cidade’”. Origem que explica o caráter coletivo que a forma *gente* possui, referente a um grupo de pessoas, a coletividade. Essas características se mantêm em torno do léxico até os dias de hoje. Esse caráter generalizante pode ser observado em diversos documentos em língua portuguesa. Cândido Figueiredo (1913) traz em seu Novo Dicionário da Língua Portuguesa um verbete sobre a palavra *gente*:

Quantidade de pessoas: encontrei muita gente. População: a gente daquela terra. Habitantes de uma região. Humanidade: a gente começou em Adão? Pessoas, que têm a mesma natureza, a mesma profissão, as mesmas ideias, os mesmos hábitos: a gente das fábricas. Força armada: o commandante tinha pouca gente. Família. (Lat. *gens, gentis*) (Figueiredo, 1913, pág. 949).

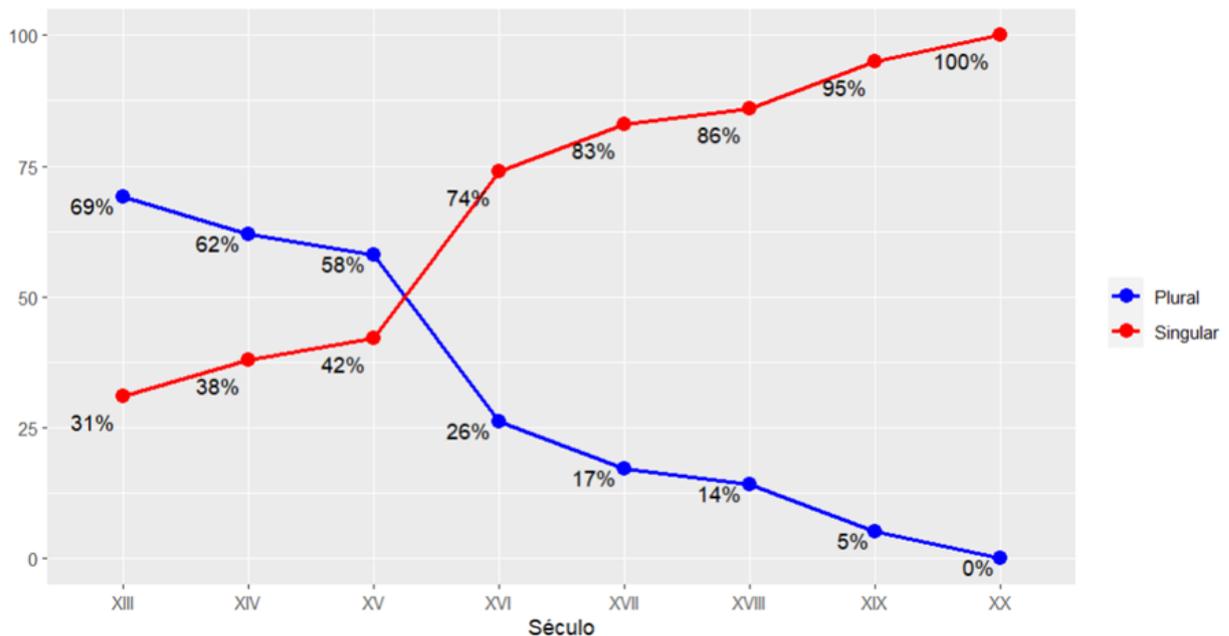
O autor lista uma série de significados para o verbete *gente*, mas em nenhum deles a expressão é empregada com o valor de pronome pessoal. De forma parecida, o Dicionário da Língua Portuguesa de Silva (1823) define o significado da palavra *gente* como “muitos indivíduos”, podendo até mesmo ser empregada no plural: “as *gentes* que os acompanhavam recebiam contia, ou soldos” (Silva, 1823, pág.34). Exemplos de emprego do substantivo, flexionado ou não, são fartos na literatura em língua portuguesa de maneira geral.

Em *Crestomatia Arcaica*, obra que traz excertos da literatura portuguesa desde o que de mais antigo se conhece até ao século XVI, Nunes (1906) dispõe diversos textos antigos que demonstram o uso do substantivo em abundância, tanto no plural quanto no singular, como é possível ver nesses trechos extraídos de textos que datam aos séculos XIII e XIV respectivamente:

- (III) Com ella vynham muytas *gentes* de desvayradas ações e de muytos e desvayrados apostamentos e ricos (**Da Igreja militante**, séc. XIII)
- (IV) grandes façanhas que a muy noble *gente* dos godos sempre fizeram (**Rei Bamba ou Vamba**, séc. XIV)

A partir do século XVI, a forma plural *gentes* começa a entrar gradativamente em desuso. Lopes (2003), em estudo sobre o processo de gramaticalização do pronome *a gente* através de uma análise em tempo real de longa duração, observou essa perda gradual dos traços sintáticos de plural do substantivo *gente*. O gráfico abaixo é uma representação do percurso histórico da queda dos traços de plural do substantivo *gente*, em análise de textos do século XIII ao século XX:

**Gráfico 1:** Queda do traço de plural do substantivo *gente* em tempo real de longa duração



Fonte: Elaboração própria com base nos dados de Lopes (2003)

Paralelamente, a autora verificou aumento cada vez mais acentuado de ocorrências do substantivo no singular a partir do mesmo século (séc. XVI), conforme demonstra o Gráfico 1, embora ainda fosse possível perceber o uso do plural “gentes” até o século XIX, como pode ser visto no trecho da peça *O Namorador ou a Noite de São João*, de Martins Pena:

- (V) O amor o centro ao redor do qual giram todas as afeições, intrigas, *gentes* e despesas (Pena, 1987).

Embora se encontre registros da forma *a gente* em uso com função de pronomes de primeira pessoa do plural desde o século XVI, é somente a partir do século XX que sua gramaticalização se delineia com mais clareza, em que se observava 100% de *a gente* com forma pronominalizada (Menon, 1995, 1996; Lopes, 2003). O processo de evolução do substantivo *gente* à forma *a gente* evidencia a perda gradual de alguns traços de pluralidade e a aquisição de peculiaridades que permeiam os pronomes, fatos que revelam como a gramaticalização altera a morfologia e molda a sintaxe da língua ao longo dos séculos. Sobre essas alterações na morfologia, sintaxe e semântica da língua, especialmente no que se refere aos traços intrínsecos envolvidos no processo de inserção de *a gente* no quadro pronominal do português, Lopes (2003) desenvolveu uma tabela, como um esquema, que sintetiza as propriedades de cada forma e as transformações ocorridas através do processo de gramaticalização de *gente* > *a gente*:

**Quadro 2:** Processo de gramaticalização de *a gente*

TRAÇOS		<i>GENTE</i>	<i>A GENTE</i>
NÚMERO	FORMAL	[αpl]	[-pl]
	SEMÂNTICO	[+ PL]	[+ PL]
GÊNERO	FORMAL	[+fem]	[φfem]
	SEMÂNTICO	[φFEM]	[αFEM]
PESSOA	FORMAL	[φeu]	[φeu]
	SEMÂNTICO	[φEU]	[+ EU]

Fonte: Adaptado de Lopes (2003, pág. 32)

Conforme é possível observar no quadro 1, o processo de gramaticalização de *gente* > *a gente* resultou na mudança de traços formais e semânticos de número, gênero e pessoa. Assim, em relação ao número, o traço formal do substantivo *gente* era variável [α plural], pois a forma *gentes* era possível; já o traço formal do pronome *a gente* não permite a utilização do plural [- plural], embora semanticamente tanto o substantivo quanto o pronome de primeira pessoa tenham traços de plural [+ plural], pois sempre se referem a mais de um referente, independente do “eu” estar incluído ou não. No que diz respeito ao gênero, podemos ver que o substantivo apresenta traços formais de feminino [+fem], enquanto o pronome é neutro [φfem],

pois seu gênero vai depender de fatores extralinguísticos, a saber, o sexo do referente (*a gente* está calmo / *a gente* está calma).

À luz do que foi discutido, é truísmo que o processo de gramaticalização que resultou na inserção da forma *a gente* no quadro dos pronomes do português é reflexo da capacidade da língua de se reinventar, de mudar e de se adaptar às demandas culturais e sociais de diferentes épocas, buscando atender a propósitos comunicativos específicos e que refletem processos intrínsecos na linguagem. Essas mudanças, especificamente em relação ao uso dos pronomes *nós* e *a gente* e suas respectivas concordâncias padrão e não padrão, são comprovadas por estudos sociolinguísticos.

Dito isto, a próxima seção será destinada à uma revisão da literatura sobre estudos que se dedicaram a investigar os padrões de *alternância pronominal* e de *concordância verbal* de primeira pessoa do plural, em que procuramos examinar os contextos linguísticos e sociais de cada estudo, bem como os principais resultados encontrados nas pesquisas desenvolvidas, a fim de comparar nossos resultados com os encontrados nesses estudos e replicar metodologias, buscando compreender como o fenômeno variável é tratado em análise de amostras de fala em situações reais de uso em diversas regiões do país.

### **3.2 Panorama dos estudos sobre *nós* e *a gente* no Brasil**

Esta seção trata da descrição de trabalhos variacionistas que analisam duas faces do mesmo fenômeno variável: o uso variável dos pronomes *nós* e *a gente* e as respectivas concordâncias verbais padrão e não padrão. Em um primeiro momento, tratamos de analisar trabalhos que abordam o fenômeno da alternância entre os pronomes *nós* e *a gente* à luz da teoria laboviana. Posteriormente, buscamos descrever as pesquisas que focalizam a concordância verbal com os pronomes de primeira pessoa *nós* e *a gente*, destacando os fatores condicionantes para cada variante, a base de dados utilizada e os resultados obtidos.

Algumas informações são pertinentes e necessárias para que possamos entender os resultados de cada pesquisa exposta nesta seção com clareza, dada as diferentes metodologias utilizadas na análise dos dados. Cabe destacar que, a maioria das pesquisas que tivemos acesso, utilizam como ferramenta estatística o pacote Varbrul, principalmente o programa Goldvarb X, que foi desenvolvido com o objetivo de auxiliar na análise de dados estatísticos, possibilitando acesso aos termos percentuais referentes a cada variável controlada, verificar os grupos de

fatores mais relevantes no processo de variação, e os pesos relativos referentes ao fenômeno estudado. Alguns desses estudos apresentam resultados por meio da frequência de uso das variantes. Outras pesquisas, além das frequências de uso, apresentam pesos relativos, que pondera a probabilidade de uma determinada variante ocorrer em um contexto de interação entre todas as variáveis elencadas (Guy e Zilles, 2007, p.158).

Isso se traduz em números que podem variar de 0,0 a 1,0. Então, se tratar de pesquisas que consideram os pesos relativos como medida de favorecimento, quanto mais próximos os pesos relativos são de 1,0, mais fortemente a aplicação a variante de referência é favorecida no contexto de análise e, ao contrário, quanto mais próximo de 0,0 for o peso relativo, menos favorecida será a aplicação da regra. Ademais, os pesos que se apresentarem próximo a 0,50 são considerados neutros em relação à regra aplicada.

Além do programa Varbrul, algumas pesquisas utilizaram a linguagem de programação R, através da interface *R Studio*, para o tratamento dos dados da variação (Teixeira, 2021; Freitas et al, 2021; Araujo e Carvalho, 2024). Esses estudos adotam como unidade de medida, em termos de favorecimento, a escala do logaritmo de chances (Log-odds). Essa escala é de fácil compreensão, tendo em vista que o ponto médio é zero, e valores positivos demonstram tendência ao favorecimento da variante de referência, enquanto valores negativos apontam desfavorecimento.

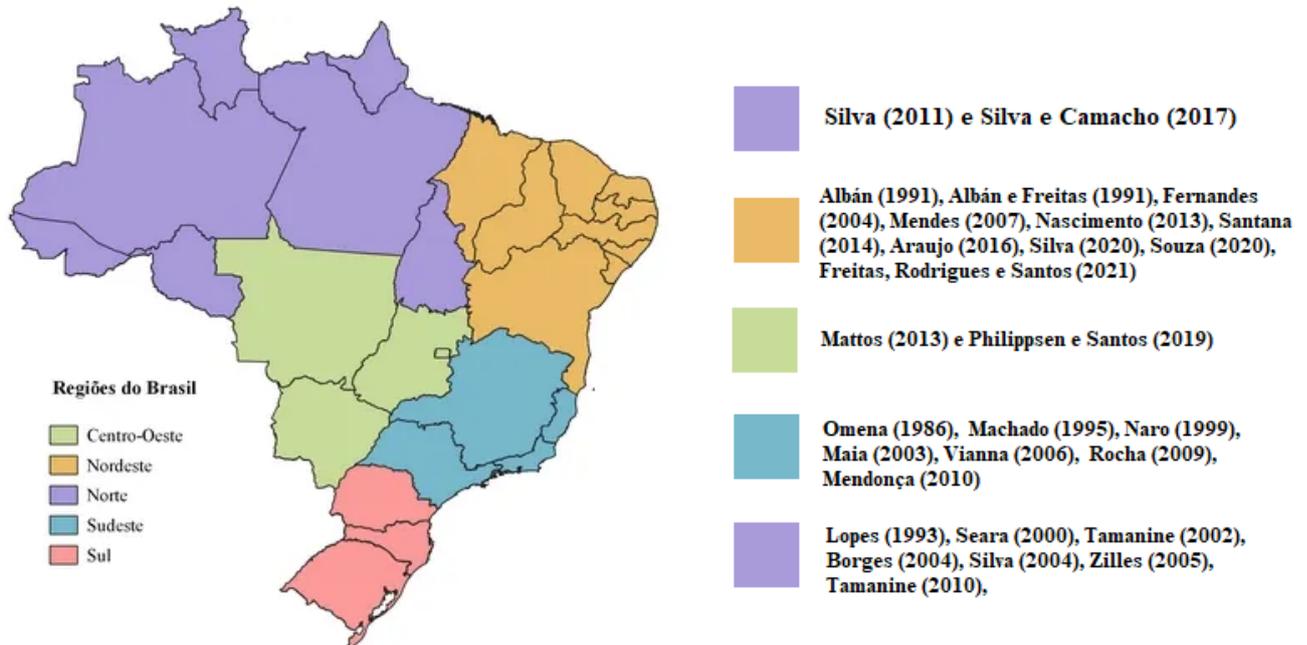
Apresentadas as considerações necessárias para um melhor entendimento dos resultados apresentados nas pesquisas consultadas, apresentamos a seguir uma revisão da literatura produzida sobre o fenômeno variável da primeira pessoa do plural à luz da teoria da variação e mudança linguística em suas duas faces: a *alternância pronominal* e a *concordância verbal de primeira pessoa do plural*.

### ***3.2.1 Panorama dos estudos variacionistas sobre alternância pronominal***

Após o levantamento da literatura, além dos trabalhos já explanados na Seção 2.1 (Omena, 2003; Lopes, 2003; e Araújo, 2016), encontramos uma vasta literatura da qual nos permite identificar que o objeto de estudo desta pesquisa vem sendo estudado há muito tempo e conta, atualmente, com trabalhos desenvolvidos em todas as regiões do Brasil, tanto no português falado quanto no escrito. Com o intuito de demonstrar graficamente a grande profusão de trabalhos desenvolvidos sobre a alternância entre os pronomes *nós* e *a gente* em abordagem variacionista no Brasil, a figura 1 demonstra que essas pesquisas foram realizadas

em todas as regiões do país. Destacamos que, apesar de o mapa representar os trabalhos divididos por região, a alternância entre os pronomes de primeira pessoa do plural não sofre pressões diatópicas, sendo comum em muitas variedades do português brasileiro:

**Figura 1:** Mapa do Brasil com pesquisas variacionistas em cada região



Fonte: Elaboração própria

Dentre as pesquisas que tivemos acesso, vale destacar, a maioria delas analisa dados extraídos da língua em uso, advindos de bancos de dados já constituídos (NURC, NORPOFOR, VALPAB, entre outros). No entanto, alguns trabalhos tratam de analisar dados extraídos de outros contextos, como as revistas em quadrinhos (Menon et al, 2003) e os telejornais da Rede Globo (Tavares, 2014), por exemplo. A seguir, faremos uma pequena descrição de cada uma dessas pesquisas e destacaremos, sucintamente, as variáveis mais relevantes em cada estudo.

### 3.2.1.1 Estudos sobre a alternância pronominal em tempo aparente

Sobre a variação *nós* e *a gente*, o trabalho de Omena (1986) merece destaque por ter sido precursora sobre o estudo desses pronomes no Brasil. Com o título “A referência variável da primeira pessoa do discurso no plural”, a pesquisadora analisou uma amostra do Projeto Censo Linguística do Rio de Janeiro - CENSO<sup>3</sup>, em 1978, investigando a fala de

<sup>3</sup> O Projeto Censo Linguística do Rio de Janeiro foi criado em 1979 com o objetivo de descrever a sistematicidade

sessenta e quatro informantes não cultos, socialmente estratificados de acordo com a faixa etária (I- 15 a 25 anos, II- 26 a 49 anos, e III- 50 anos ou mais) e com o grau de escolaridade (Primário – 1ª a 4ª série; Ginásio - 5ª a 8ª série; e Ensino Médio). As variáveis controladas no estudo foram, além das estratificadas socialmente, a *função sintática*, *sequência do discurso*, *saliência fônica*, *tempo verbal* e *grau de indeterminação do referente*. Ao todo, foram encontradas 3299 ocorrências dos pronomes *nós/a gente* nas amostras analisadas, sendo 2270 realizações da forma pronominal *a gente* (69%), e 1029 dados referentes à forma *nós*, correspondendo a 31% dos casos. A análise revelou alguns grupos de fatores que são significativos para o fenômeno analisado, a saber: *função sintática*, com maior ocorrência do pronome na função de adjunto adverbial (84%); e *saliência fônica*, com preferência para verbos de menor saliência (91%). Por fim, Omena (1986) acredita que a regra variável de primeira pessoa do plural consiste em um fenômeno em mudança, sustentada no fato dos informantes mais jovens serem mais propenso ao uso do pronome inovador.

Albán e Freitas (1991) analisaram a realização dos pronomes *nós* e *a gente* na função de sujeito em dados obtidos a partir do Projeto da Norma Urbana Oral Culta – NURC/Salvador, tendo sua amostra composta por inquéritos do tipo Diálogo entre dois informantes (D2) com informantes universitários do sexo masculino e estratificados de acordo com a idade: I- 25 a 35 anos e II- acima de 55 anos. O trabalho controlou as variáveis *idade*, *atitude assumida no diálogo* (se tensa ou se distensa) e o *momento da elocução* (primeiro, segundo ou terceiro segmento). Com o total de 264 dados obtidos na análise, os autores constataram que, na amostra selecionada, há uma disputa acirrada entre os pronomes *nós* (47,7%) e *a gente* (53,3%). As variáveis mais significativas foram a *idade* e o *momento da elocução*. Em relação à primeira variável relevante, constatou-se tendência maior entre os mais novos para o uso de *a gente* (79% dos casos), enquanto os mais velhos ainda preferem a forma pronominal *nós* (65% das ocorrências). Já no que se refere a segunda variável selecionada, o terceiro segmento da elocução parece favorecer o uso da variante não padrão, com 68% das ocorrências. Com os resultados, os autores concluíram que o uso da forma *a gente* possui alta vitalidade no português do Brasil, mesmo que ainda tratado de forma coloquial.

Os autores ainda realizaram outro estudo sobre os pronomes *nós* e *a gente* utilizando o mesmo banco de dados, porém com inquéritos do tipo DID (Diálogo entre Informante e Documentador). Para isso, foram controladas as variáveis sociais sexo (masculino e feminino),

---

da variação observada no português brasileiro, depreender mudanças em tempo aparente e identificar os correlatos estruturais, sociais e funcionais desses processos. Fonte: <http://www.lettras.ufrj.br/peul/historia.html>. Acesso em: 20/05/2022.

faixa etária (I- 15 a 25 anos, II- 26 a 59 anos e III- a partir de 50 anos) e escolaridade (A- 1º a 4ª série do EF, B- 5ª a 8ª série do EF e C- Ensino médio). Já as variáveis linguísticas testadas foram a *saliência fônica*, *tempo verbal*, *momento da elocução*, *referência do pronome*, *preenchimento do sujeito*, *estrutura do verbo*, *tipo de verbo* e *atitude de elocução*. Com um total de 779 ocorrências de primeira pessoa do plural, a pesquisa demonstrou uma utilização mais expressiva de *a gente* (73%) se comparado ao uso de *nós* (27%). Os fatores selecionados como relevantes para o condicionamento de *a gente* foram: *faixa etária* (I- 15 a 25 anos, 81%); *atitude do informante* (frases declarativas, 54%); e *momento da elocução* (segundo momento, 64%). Os autores afirmaram que existe uma tendência maior entre os mais jovens para o uso do pronome inovador, o que pode ser indício de uma mudança linguística em processo.

Machado (1995) pesquisou a realização de primeira pessoa do plural com dados do Projeto Atlas Etnolinguístico dos Pescadores do Estado do Rio de Janeiro (APERJ). Sua amostra foi constituída de 72 informantes, todos pescadores e homens, estratificados de acordo com a idade (I- 18 a 35 anos, II- 36 a 55 anos e III- 56 a 75 anos), escolaridade (analfabetos ou parcialmente escolarizados) e localidade de pesca fluminense (Itaocara, São Fidelis, Cambuci, São João da Barra, Itaperuna, Ponta Grossa dos Fidalgos, São Benedito, Barra de Itabapoana, Atafona, Guaxindiba, Gargaú e Farol de São Tomé). Foram observados 2971 dados, sendo 72% das ocorrências para *a gente* e 28% para *nós*. As variáveis controladas pelo estudo e que se mostraram significativas foram: *faixa etária* (os mais jovens favorecem o uso de *a gente*, 0.68); *grau de indeterminação do pronome* (referente indeterminado, 0.56); *paralelismo formal* (manutenção de *a gente*, 0.83) e *tempo verbal* (verbos no infinitivo, 0.75). Com os resultados, especialmente focados no favorecimento da forma *a gente* entre os mais jovens, a autora concluiu que o fenômeno variável apresenta indícios de mudança em curso.

O trabalho desenvolvido por Seara (2000) analisou uma amostra composta por 12 informantes de Florianópolis, advindos do Projeto Variação Linguística Urbana do Sul do País (VARSUL). Esses informantes foram divididos de acordo com sexo, a escolaridade e a idade. Após proceder a análise geral dos dados, o autor observou o uso mais expressivo da variante inovadora *a gente*, com 72% dos casos, e o uso menos frequente da forma *nós*, com 28%. Entre as variáveis mais significativas para a aplicação de *a gente*, o estudo apontou: *tempo verbal*, em que se destaca o pretérito imperfeito (0.68); *sexo*, sendo as mulheres favorecedoras da forma *a gente* (0.66); *traço semântico*, com o pronome no sentido genérico condicionando a forma inovadora (0.58); e a *faixa etária*, com os mais jovens preferindo o uso de *a gente* (0.68).

Outro trabalho realizado a partir de dados do Projeto Variação Linguística Urbana do Sul do País (VARSUL) foi a pesquisa feita por Tamanine (2002) sobre as formas de primeira

pessoa do plural no interior de Santa Catarina. Em sua análise, após realizar rodada geral através do programa computacional VARBRUL, a autora levantou 5235 dados referentes aos pronomes *nós* e *a gente*, em que identificou um uso discretamente mais frequente da forma *a gente*, tendo 55% das ocorrências, contra 45% para o uso de *nós*. Além de testar as mesmas variáveis sociais verificadas no estudo de Seara (2000), foram controladas as variáveis *localidade*, *concordância verbal*, *paralelismo*, *tipo de verbo* e *determinação do referente*. Diante dos resultados, a linguista atestou a preferência pelo uso de *a gente* entre os mais jovens nas três localidades pesquisadas: 52% em Lages; 64% em Chapecó e 77% em Blumenau.

Com dados de fala das comunidades de Pombal e de Belo Horizonte, Maia (2003) pesquisou a alternância entre os pronomes de primeira pessoa do plural no dialeto mineiro. Com o objetivo de comparar as variedades urbanas e rural, o estudo contou com 12 entrevistas com informantes estratificados de acordo com a faixa etária, o sexo e o grau de escolarização. Foram encontrados 672 dados do fenômeno variável, dos quais 53% são referentes ao pronome *a gente*, enquanto 47% são para a forma *nós*. Apesar da pouca diferença entre as frequências das duas formas pronominais, o autor chegou à conclusão de que há indícios de mudança em relação ao fenômeno, já que os jovens concentram a maioria das ocorrências da forma *a gente* (69%).

Outro trabalho importante foi o desenvolvido por Borges (2004). O autor selecionou amostras provenientes de 11 peças de teatro no período de 1896 a 1995, todas de autores regionais. Além disso, para complementar a base de dados da pesquisa, foram analisadas as falas de 60 informantes das cidades de Pelotas e Jaraguão, gravadas em áudio nos anos de 2000 e 2001. Obtendo o total de 1712 realizações de primeira pessoa do plural, o autor observou o uso mais acentuado do pronome *nós*, com 76% das ocorrências, e um uso menos frequente da forma *a gente*, com 24% dos dados em contextos de pronome expresso e não expresso.

Em artigo publicado com base em sua dissertação de mestrado de 1997, Fernandes (2004) investigou a primeira pessoa do plural na cidade de João Pessoa – Paraíba, com dados provenientes do Projeto Variação Linguística no Estado da Paraíba (VALPB)<sup>4</sup>. Sua amostra conta com 60 informantes que foram estratificados a partir da idade (I- 15 a 25 anos, II- 26 a 49 anos, e III- a partir de 50 anos), sexo (masculino e feminino) e grau de escolaridade (A- Nenhuma escolaridade, B- 5 a 8 anos de estudo e C- mais de 11 anos de estudo). Para analisar o fenômeno, foram controladas, além das variáveis sociais, as variáveis linguísticas *função sintática*, *tipo de discurso*, *referência semântica*, *estrutura do verbo*, *tempo verbal*, *posição do*

---

<sup>4</sup> Projeto Variação Linguística no Estado da Paraíba – VALPB apresenta um *corpus* constituído em 1993, com o objetivo de descrever o falar da Paraíba. Seu *corpus* é estratificado segundo o sexo, a faixa etária e os anos de escolarização.

*pronome em relação ao verbo e tipo de discurso*. O estudo analisou 2739 dados, confirmando um uso mais abundante de forma *a gente*, com 2153 ocorrências (79%), e o uso menos intenso da forma *nós*, com 586 ocorrências (ou 21% do total). Os fatores que se mostraram mais relevantes no favorecimento da regra foram: mais jovens (0.69), mais escolarizados (0.51), referência genérica do pronome (0.70) e o tempo verbal pretérito imperfeito (0.68).

Mendes (2007) apresentou, em sua dissertação de mestrado, um estudo em tempo aparente sobre a variação pronominal *nós/ a gente* com *corpus* constituídos de vinte e quatro entrevistas gravadas na cidade de Santo Antônio de Jesus, no Recôncavo Baiano. Os informantes da amostra foram estratificados de acordo com as variáveis sociais: sexo (homem e mulher), faixa etária (I- 20 a 40 anos, II- 41 a 60 anos e III- a partir de 61 anos), e grau de escolaridade (analfabetos e semianalfabetos). Entre as 1970 ocorrências de realização da primeira pessoa do plural, o autor verificou 93% de frequência para a variante *a gente*, e somente 7% do total correspondia a forma *nós*. Após análise realizada pelo programa computacional Varbrul, os grupos de fatores selecionados como relevantes para o uso de *a gente* foram: *faixa etária* (I- 20 a 40 anos, com peso relativo de 0.67); *saliência fônica* (redução dos ditongos, 0.78); *paralelismo formal* (antecedente, 0.98); e *grau de determinação do pronome* (indeterminado, 0.64). Com os resultados, o autor identificou um processo de mudança em curso na comunidade de Santo Antônio de Jesus, no sentido de o pronome *a gente* estar ganhando terreno na fala dos entrevistados em detrimento do pronome *nós*.

Assim como Maia (2003), outro trabalho que destaca o dialeto mineiro é o desenvolvido por Rocha (2009), que investigou a regra variável na comunidade de fala de Belo Horizonte. Com amostras provenientes do Projeto Descrição Sócio-histórica do Português de Belo Horizonte, o autor verificou a atuação das variáveis *tempo verbal*, *função sintática*, *localização na frase* e *referência do pronome*. Além disso, os informantes foram estratificados de acordo com a faixa etária, sexo e escolaridade. Após levantamento, foram encontrados 615 dados de primeira pessoa do plural, sendo 63% referentes ao pronome *a gente* e 37% referentes à forma *nós*. Dentre as variáveis selecionadas pelo programa, destacam-se: *referência do pronome*, *faixa etária* e *localização na frase*.

Mendonça (2010) analisou dados da fala capixaba referentes ao Projeto Português Falado de Vitória (PortVix). Para sua amostra, o autor selecionou 40 informantes distribuídos entre homens e mulheres de três faixas etárias distintas e de três níveis de escolaridade: fundamental, médio e superior. O levantamento de dados encontrou 1745 realizações dos pronomes estudados, averiguando uso bem mais produtivo da forma *a gente*, com 71% dos casos, e um uso mais discreto da forma *nós*, com 29%. Os grupos de fatores considerados

relevantes para a pesquisa foram o *paralelismo formal*, o *preenchimento do pronome* e o *grau de referencialidade do pronome*.

Na região Sul do país, Tamanine (2010) apresentou tese de doutorado sobre a variação pronominal *nós/ a gente* na cidade de Curitiba. Para tanto, foi utilizada uma amostra constituída por 32 entrevistas provenientes do Projeto VARSUL, com informantes estratificados conforme o sexo, a faixa etária e o grau de escolaridade. A autora controlou as seguintes variáveis linguísticas: *preenchimento do pronome*, *concordância verbal*, *tempo verbal*, *grau de determinação do sujeito*, *tipos de texto* e *tipo de verbo*. Após submeter os dados ao programa Goldvarb, a rodada obteve um total geral de 2084 dados referentes à primeira pessoa do plural, sendo destes 54% de *a gente* e 46% de *nós*. Entre os apontamentos importantes revelados na pesquisa, a autora identificou tendência maior para o uso da forma inovadora entre os mais jovens (0.67), e, quanto ao *tipo de verbo*, o verbo *saber* se mostrou favorecedor do pronome *a gente* (0.96).

Na região Norte, o trabalho realizado por Silva (2011) analisou a variação na alternância pronominal entre *nós* e *a gente* em amostra de fala dos moradores de Nazaré, bairro da classe média-alta da cidade de Belém – Pará. Para isso, a autora analisou entrevistas de 35 informantes provenientes do Projeto Variação Linguística Urbana na Região Norte – VALUNORTE, distribuídos de acordo com o sexo (masculino e feminino), a faixa etária (I- 18 a 25 anos; II- 26 a 45 anos; e III- a partir de 46 anos) e escolaridade (ensino médio, ensino fundamental e não alfabetizado). Os resultados gerais do estudo apontaram para um uso mais frequente da forma inovadora *a gente*, com 54,5% das ocorrências, em relação a forma canônica *nós*, com 45,4%. As variáveis mais relevantes que condicionam o uso dos pronomes, de acordo com o estudo, foram *a faixa etária* e *a escolaridade*, conforme demonstrou o resultado de 66,67% de uso de *a gente* entre os mais jovens. Já entre os mais escolarizados, a preferência pelo uso do pronome padrão *nós* se mostrou acentuada, com 60% das ocorrências.

A tese de doutorado de Mattos (2013) versa sobre a alternância pronominal e a concordância verbal de primeira pessoa do plural na fala de Goiás. Para compor sua amostra, a linguista contou com gravações em áudio de 55 informantes cultos com, no mínimo, 10 anos de estudo, de onde foram extraídas 2412 ocorrências do fenômeno variável, distribuídos da seguinte forma: 77% de *a gente*, e 23% de *nós*. Esses informantes foram estratificados conforme o sexo, a faixa etária e a escolarização. O estudo selecionou como relevante as seguintes variáveis: *tempo verbal*, *ritmo*, *expressão do sujeito* e *faixa etária*. De posse dos resultados, o autor identificou uma tendência maior ao uso da forma inovadora *a gente* entre os mais jovens, o que pode indicar início de um processo de mudança.

Tavares (2014) tratou de analisar os pronomes de primeira pessoa do plural na fala de âncoras e repórteres dos telejornais da Rede Globo. Para isso, o autor utilizou gravações em áudio de cinco programas de cada um dos seguintes telejornais: Bom Dia Brasil, Jornal Hoje, Jornal Nacional e Jornal da Globo. Após submeter os dados ao programa computacional Goldvarb, foram totalizadas 221 ocorrências dos pronomes *nós* e *a gente*, sendo 54% delas correspondente ao uso de *nós* (112 ocorrências) e 46% referentes a forma pronominal *a gente* (98 ocorrências). Apesar da pesquisa atestar o uso mais frequente da forma *nós*, o autor concluiu que o pronome inovador *a gente* está ganhando terreno mesmo em ambientes mais formais, como é o caso dos telejornais.

Outro importante trabalho que retrata a variação *nós* e *a gente* é a dissertação de mestrado defendida por Santana (2014), com dados fala popular de Salvador, Bahia. O banco de dados utilizado pelo pesquisador foi o Programa de Estudos Sobre o Português Falado de Salvador (PEPP)<sup>5</sup>, constituído entre 1998 e 2000. Para tanto, foi selecionada uma amostra composta por dose informantes estratificados de acordo com o sexo (homem e mulher), faixa etária (I- 15 a 24 anos, II- 25 a 35 anos, III- 45 a 55 anos e IV- acima de 65 anos), e escolaridade (A-fundamental, B- médio e C- superior). Além dessas variáveis sociais, foram controladas as variáveis linguísticas saliência fônica, paralelismo, tempo verbal, preenchimento do sujeito e posição em relação ao verbo. Ao todo, foram encontradas 453 ocorrências do fenômeno variável, dividida da seguinte maneira: 76% para *a gente* e 24% para *nós*. Após análise realizada pelo programa GoldvarbX, os fatores considerados mais significativos entre as variáveis controladas foram: indeterminação parcial (0.67); redução do ditongo (0.68); idade acima de 65 anos (0.67); ensino fundamental (0.58) e paralelismo antecedente (0.98). Com a análise dos resultados, o autor constatou que, mesmo uma frequência de uso bem mais expressiva para a forma *a gente*, o fenômeno encontra-se em estado de variação estável na comunidade de fala pesquisada.

Sobre a fala dos fortalezenses, comunidade investigada também nesta pesquisa, destacamos o trabalho de Araujo (2016). Neste estudo, o autor utilizou uma amostra do *corpus* do Projeto Norma Oral do Português Popular de Fortaleza (NORPOFOR), composto por gravações em áudio de entrevistas sociolinguísticas. A amostra escolhida é constituída por 53

---

<sup>5</sup> O PEPP é formado por quarenta e oito gravações de entrevista de aproximadamente quarenta minutos, com o seguinte perfil dos informantes, que estão distribuídos em quatro faixas etárias: de 15 a 24 anos, de 25 a 35 anos, de 45 a 55 anos e de 65 anos em diante. Fonte: [https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwiHtO7GuNTLAhXDjJAKHUB8DLIQFggkMAI&url=http%3A%2F%2Fwww.quartetoeditora.com.br%2Fcatalogo-livro-scolhido.php%3Flivro%3D43&usq=AFQjCNHmQLt71muPOG1YcF5\\_WFIXiSrS3Q&sig2=d5Ry38sXib5hEluLXxtoWQ](https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwiHtO7GuNTLAhXDjJAKHUB8DLIQFggkMAI&url=http%3A%2F%2Fwww.quartetoeditora.com.br%2Fcatalogo-livro-scolhido.php%3Flivro%3D43&usq=AFQjCNHmQLt71muPOG1YcF5_WFIXiSrS3Q&sig2=d5Ry38sXib5hEluLXxtoWQ). Acesso em: 22/03/2022.

informantes: 26 mulheres e 27 homens; 17 indivíduos da faixa etária I (15 a 25 anos), 18 da faixa etária II (26 a 49 anos) e 18 da faixa etária III (a partir dos 50 anos); 17 informantes com nível de escolaridade A (0 a 4 anos), 18 com nível B (5 a 8 anos) e 18 com nível C (9 a 11 anos). Foram utilizados os inquéritos do tipo D2 (diálogo entre dois informantes), por ser constituído de falas mais espontâneas. A análise quantitativa dos dados foi realizada com o auxílio do programa computacional GoldVarb X. Na rodada geral, foram levantadas 1.092 ocorrências da variação pronominal *nós* e *a gente* no falar popular de Fortaleza, das quais a forma *a gente* contou com 716 dados, apresentando uma frequência de uso de 65,6%, enquanto a forma *nós* aparece em 34,4% do total de ocorrências, totalizando 376 dados. De maneira geral, a análise do *corpus* apresentou frequência maior da forma inovadora *a gente* em detrimento da forma conservadora *nós*, no que se refere à realização da primeira pessoa do plural. Os dados da pesquisa revelaram que a maior proporção das ocorrências de *a gente* se dá na função de sujeito, embora o seu peso relativo encontre-se próximo do ponto neutro, com probabilidade de aplicação de *a gente* de .535. A função de adjunto favorece de forma expressiva o pronome *a gente* (.722), apesar de ter apresentado um baixo número de ocorrências (17 dados). As funções sintáticas objeto direto (.225), predicativo do sujeito (.105) e objeto indireto (.058) desfavorecem significativamente a forma *a gente*.

Silva e Camacho (2017) conduziram uma pesquisa em tempo aparente sobre a variação pronominal de primeira pessoa do plural com dados de fala de Rio Branco – Acre. Para isso, foi utilizado um *corpus* constituído por 40 gravação em áudio realizadas nos anos de 2011 e 2012, com informantes estratificados de acordo com o sexo/gênero, idade e escolaridade. Após submeter análise ao programa computacional Goldvarb, foram encontradas 1061 ocorrências dos pronomes *nós/a gente*, das quais 76,7% eram para a variante inovadora, enquanto 23,3% correspondiam à variante padrão. Entre os fatores controlados, os que se mostraram relevantes para a análise foram: *sexo*, com 68,1% das ocorrências para os homens e peso relativo de 0.617; *idade*, obtendo 74,4% das ocorrências entre os mais jovens (peso relativo: 0.551); e a *referência do pronome*, demonstrando que o pronome no sentido específico condiciona o uso de *a gente*, com 74% de frequência de uso e 0.810 de peso relativo. Com a análise dos dados, os autores concluíram que a forma pronominal *a gente* está incorporada na gramática da língua falada em Rio Branco, mesmo não sendo reconhecida por grande parte das gramáticas.

Philippesen e Santos (2019) estudaram o uso das formas pronominais *nós* e *a gente* em cinco comunidades rurais de Mato Grosso. Para a análise, foram observados os fatores: escolaridade, faixa etária e o sexo na fala de vinte informantes. A análise geral das comunidades

estudadas concluiu uma maior preferência pela variante conservadora, com 83% dos casos para *nós* e somente 17% para *a gente*. No entanto, com os resultados, o autor percebeu uma preferência pelo uso de *a gente* com maior frequência entre os mais escolarizados (57%), entre os mais jovens (57%) e pelos homens (65%).

Mais recentemente, Silva (2020) realizou uma investigação sobre o uso intercambiável de *nós* e *a gente* na fala culta cearense em amostra proveniente do Projeto Dialetos Sociais Cearenses - DSC, cujos dados foram obtidos por meio de entrevistas, conversas espontâneas e interação médico-paciente, com cerca de 18 horas de gravação realizadas entre 1986 e 1987 (Aragão; Soares, 1996). Ao todo, foi analisada a fala de 18 informantes selecionados de 11 bairros de Fortaleza e distribuídos de acordo com as seguintes variáveis estratificadas, organizadas conforme a metodologia variacionista: quatro faixas etárias (10-11, 14-15, 18-25 e 37-43), quatro graus de escolaridade (analfabeto, primário, ginásial e segundo grau) e sexo (masculino e feminino). Os resultados indicaram uma frequência maior da forma inovadora *a gente* (200 do total de 288 ocorrências - 69,4% dos casos) em detrimento do pronome canônico *nós* (88 do total de 288 ocorrências - 30,6% dos casos), em termos de percentual de uso. A rodada estatística, com peso relativo, selecionou três grupos de fatores como significativos para o fenômeno em estudo, na seguinte ordem de significância (Input: .803 e 64 Significância: .010): *determinação do referente* (indeterminação universal .574, e indeterminação circunscrita .323), *sexo* (masculino .666, e feminino .375) e *paralelismo formal* (o fator que mais favoreceu o uso de *a gente* foi a referência anterior *a gente*, independente do referente, com percentual de 85,7% e peso relativo de .715)

Por fim, na região Nordeste, devemos citar a pesquisa de Souza (2020). A pesquisadora utilizou amostra composta por 20 informantes provenientes do projeto A Língua Usada no Sertão Alagoano (LUSA), todos da comunidade Quilombola Serra das Viúvas, que foram estratificados socialmente de acordo com as variáveis sexo (masculino/ feminino) e faixa etária (F1 – 25 a 50 anos e F2 – 60 anos em diante). Seguindo a tendência das pesquisas sobre os pronomes de primeira pessoa do plural no Nordeste, o estudo apresentou uma tendência maior para a utilização da variável inovadora. Assim, das 429 ocorrências do fenômeno pesquisado, 56% foram para a forma *a gente* (241 ocorrências), e 44% delas foram para o pronome canônico *nós* (188 ocorrências). Com o auxílio do programa computacional GoldVarb X, foram selecionadas quatro variáveis significativas no condicionamento do uso de *nós/ a gente*, a saber, *paralelismo formal*, *marca morfêmica*, *faixa etária* e *sexo*. Sobre o *paralelismo*, a autora evidenciou que o uso de *a gente* foi maior quando antecedido da mesma forma pronominal, com percentual de 90% e peso relativo de 0.95. Em relação à *marca morfêmica*,

Souza destacou com a forma *a gente* é favorecida pela marca de terceira pessoa do singular, com peso relativo de 0.60 e percentual de uso de 63%. No que diz respeito à *faixa etária*, o trabalho confirmou que a forma pronominal *a gente* é preferida pelos mais jovens, com 0.59 de peso relativo e 62% de frequência de uso. Por fim, os homens tendem a utilizar mais a forma *a gente* do que as mulheres (frequência: 54%. Pero relativo: .66). Com isso, a autora concluiu que, na comunidade Serra das Viúvas, de onde os informantes são provenientes, existe uma expressiva utilização tanto de *nós* quanto de *a gente*, porém não foi encontrada indícios que indiquem que a variação pronominal *nós* e *a gente* esteja em um processo de mudança, mas parecem estar em um estado de variação estável, mesmo que o pronome *a gente* seja favorecido em diversas circunstâncias.

A predominância da forma pronominal *a gente* em diversas regiões e estratos sociais pode sugerir indícios de uma mudança linguística significativa no português brasileiro. Esta mudança pode ser vista como parte de um processo mais amplo de simplificação do paradigma pronominal, onde as formas menos marcadas ganham mais espaço. No entanto, ainda se observa resistência significativa do pronome padrão *nós*, como demonstram os estudos analisados, o que pode indicar que essa mudança não é uniforme e pode estar sujeita a variações de outras naturezas, como pressão normativa, por exemplo e, até mesmo, a elementos culturais (Mattos, 2013).

A observação desses resultados destaca a importância de considerar múltiplos fatores ao investigar mudanças linguísticas. A alternância entre os pronomes *nós* e *a gente* reflete a realidade linguística da comunidade de fala e como isso afeta sua dinâmica social. Estudos futuros devem continuar a explorar a língua em situações de uso, utilizando tanto análises em tempo aparente quanto em tempo real para obter uma compreensão mais completa das mudanças em curso no português falado no Brasil e no mundo.

### 3.2.1.2 Estudos sobre a alternância pronominal em tempo real

Nesta subseção apresentamos os trabalhos que desenvolvem análises em tempo real sobre a alternância pronominal entre *nós* e *a gente* à luz da teoria da variação e mudança linguística, ou seja, que analisaram o comportamento dos pronomes em duas décadas distintas, destacando os aspectos da variação em cada uma das sincronias, em uma análise comparativa. Nesses trabalhos, buscamos averiguar as mudanças nos traços e no uso dos pronomes em cada época analisada, a fim de encontrar indícios de mudança linguística nos padrões de alternância

pronominal. Trabalhos que analisam dados diacrônicos ainda são raros no Brasil, dada a escassez de bancos de dados com amostras representativas do falar de duas épocas diferentes. Dentre esses trabalhos, destacamos Omena (2003), Lopes (2003), Vianna, (2006) e Nascimento (2013).

Omena (2003) investigou o comportamento dos pronomes *nós* e *a gente* na fala popular carioca em amostras representativas das décadas de 1980 e 2000, provenientes do Projeto Censo Linguística do Rio de Janeiro – CENSO. Os informantes foram socialmente estratificados de acordo com a faixa etária (7 a 14 anos, 15 a 25 anos, 26 a 49 anos, e 50 anos ou mais) e o grau de escolaridade<sup>6</sup> (1º segmento EF, 2º segmento EF e Ensino Médio). Com o objetivo de investigar os contextos em que o pronome inovador está tomando o lugar antes ocupado pelo pronome conservador, a autora analisou o uso dos pronomes em função das seguintes variáveis linguísticas: traço do *referente*, *saliência fônica*, *tempo verbal*, *aspecto verbal*, e *paralelismo*.

Os resultados evidenciam expressivo percentual de uso do pronome inovador *a gente* e pouca recorrência do pronome *nós* nas duas décadas investigadas. A amostra representativa dos anos 1980 demonstrou frequência de 78% para o pronome inovador, em detrimento do pronome padrão, que apresentou 22% das ocorrências. De maneira parecida, os resultados da amostra da década de 2000 apontou alta frequência do pronome inovador, com 79% das realizações, e baixa recorrência do pronome padrão (21% das realizações).

Entre grupos de fatores que se mostraram relevantes para o condicionamento do pronome *a gente* na década de 1980 se destacaram a *faixa etária* (sendo os mais jovens franco favorecedores do pronome inovador, com frequência de uso de 89% e peso relativo de 0.79), *grau de escolaridade* (informantes com Ensino Médio concluído - 81% e peso relativo igual a 0.56), *saliência fônica* (verbos menos salientes na relação singular e plural ‘falava/falávamos’, frequência de uso 71% e peso de 0.72), *tempo verbal* (formas nominais, 96% de uso e P.R. 0.87), *aspecto verbal* (processo imperfeito, com frequência de uso de 79% e peso relativo de 0.67) e *paralelismo* (*a gente* com referência igual a anterior: 91% e 0.69). A variável *traço do referente* não se mostrou relevante para o uso do pronome inovador na análise com dados da década de 1980.

Em relação a análise dos dados da década de 2000, a autora destacou forte influência das seguintes variáveis sobre a realização do pronome *a gente*: *faixa etária* (mais jovens, tendo frequência de uso de 94% e peso relativo de 0.84), *grau de escolaridade*

---

<sup>6</sup> De acordo com a autora, 1º segmento EF corresponde ao primário corresponde ao Primário – 1ª a 4ª série, e 2º segmento EF corresponde ao Ginásio - 5ª a 8ª série.

(informantes com ensino médio: 79% e 0.73), *traço do referente* (referência determinada, tendo 80% das ocorrências e P.R. igual a 0.61) *saliência fônica* (verbos menos saliente, com frequência de 81% e peso igual a 0.72), *tempo verbal* (formas nominais: 96% e 0.72), *aspecto verbal* (imperfectivo, apresentando frequência de uso de 83% e peso relativo de 0.65) e *paralelismo* (referência igual a anterior: 91% e 0.64).

Os resultados obtidos pela autora evidenciam uma tendência no uso dos pronomes nas duas décadas analisadas: “a forma inovadora continua a predominar e a direção da mudança é a mesma” (Omena, 2003, p. 79). Nesse sentido, ao considerar a atuação das forças linguísticas e sociais sobre a alternância pronominal, a autora concluiu que o indivíduo, ao adquirir a língua, já está submetido a um ambiente linguístico de maior ocorrência da variante inovadora, em detrimento da forma mais antiga e, com o passar dos anos, embora falante amplie seus contatos sociais e, conseqüentemente, vai sendo submetido à forças mais conservadoras, o padrão linguístico adquirido enquanto jovem ainda predomina em sua vida adulta.

Outro estudo importante foi o conduzido por Lopes (2003), que trata a inserção de *a gente* no quadro pronominal do português através de análises em tempo real de longa e curta duração. Na análise linguística de longa duração, objetivou-se investigar o processo de gramaticalização da expressão pronominal *a gente* através da análise de textos escritos, divididos em duas grandes amostras. A primeira é composta por uma documentação referente ao período entre o século XIII e XVI, chamada de Amostra do Português Arcaico. A segunda delas corresponde às origens do português como língua escrita, com amostras do português europeu, brasileiro e africano, num período que compreende ao recorte temporal entre os séculos XIII e XX, com o objetivo de identificar propriedades intrínsecas do substantivo *gente* que foram inibidas ou que persistiram ao processo de pronominalização. Os resultados da pesquisa demonstram que o processo de gramaticalização da forma *a gente* iniciou no século XVI, quando foi identificado primeiro uso da forma inovadora com função de pronome, sendo intensificado no século XX, quando a forma pronominal se expandiu e passou a ocupar contextos diversificados no português, disputando com o pronome *nós* a referência à primeira pessoa do discurso. Entre os achados da análise, a autora identificou maior tendência ao uso de *a gente* com formas na terceira pessoa do singular e com função pronominal em de sujeito genérico, tendo uso gradualmente intensificado com o passar dos séculos.

Para a análise linguística em tempo real de curta duração, Lopes (2003) contou com amostras constituídas por informantes cariocas com ensino superior concluído, a partir de gravações do Projeto Norma Urbana Oral Culta do Rio de Janeiro - NURC/ RJ, com gravações realizadas nas décadas de 1970 e 1990. Neste estudo, a pesquisadora apresentou análise de curta

duração através de dois parâmetros distintos, a análise em painel e a análise em tendência (Labov, 1977). Nesse sentido, os mesmos informantes que participaram das gravações na década de 70 foram contatados na década de 90 para novas entrevistas sociolinguísticas (Amostra Recontato), com o intuito de comparar o comportamento do fenômeno na fala dos mesmos informantes vinte anos depois. Além disso, uma nova coleta com informantes diferentes foi realizada, para comparar os dados em tendência na comunidade de fala.

Os resultados gerais apontam para a expansão no uso do pronome *a gente* na comunidade de fala em vinte anos. Na década de 1970, a frequência de uso do pronome inovador era de 42%, menos recorrente que o pronome *nós*, que apresentou 58% do total de realizações de primeira pessoa do plural. Na Amostra Recontato, gravada nos anos 1990, o pronome inovador ganha espaço na fala dos mesmos informantes da década anterior, tendo percentual de uso de 54%, enquanto o pronome *nós* figura em 46% dos dados. Já os resultados da fala de novos informantes na década de 1990 apontou para uso expressivo do pronome *a gente* (75%) em detrimento de *nós* (25%), o que representa uma tendência no uso dos pronomes pessoais de primeira pessoa do plural na fala culta carioca dos anos 1990.

Na análise da amostra dos anos 70, os fatores se mostraram relevantes para a aplicação do pronome *a gente* foram: 1) *paralelismo com a quarta pessoa*, sendo a ausência da desinência *-mos* franca aliada da forma inovadora, com 86% e P.R. de 0.89; 2) *tipologia semântica no sujeito*, tendo maior probabilidade de ocorrer em uso com referente genérico (76% e 0.89); 3) *saliência fônica*, 67% das ocorrências para verbos menos saliente e *pero* relativo de 0.78; 4) *tempo verbal*, com favorecimento do pronome *a gente* em verbos no pretérito imperfeito (53% e 0.67); 5) *sexo/faixa etária*, sendo os homens mais jovens responsável por 98% das ocorrências do pronome inovador, apresentando peso relativo de 0.98, o que indica alto favorecimento da forma nesse contexto.

Os resultados da análise em tempo real com dados da Amostra Recontato demonstram comportamento parecido em relação ao uso dos pronomes pelos mesmos informantes após 20 anos. Os fatores selecionados como favorecedores da forma pronominal *a gente* foram: *paralelismo com a quarta pessoa* (ausência da da desinência *-mos*: 0.76), *tempo verbal* (pretérito imperfeito, com peso relativo de 0.64), *tipologia semântica do sujeito* (uso genérico favorecendo a forma inovadora: 0.75), e *sexo/faixa etária* (mulheres mais jovens: P.R. 0.67).

Por fim, a análise realizada com dados da nova amostra de 1990, com informantes diferentes da década anterior, identificou grande expansão da forma inovadora *a gente* em diversos contextos linguísticos e sociais, sendo os fatores mais significativos para a realização

da forma gramaticalizada o *paralelismo com a quarta pessoa* (o pronome *a gente* está fortemente associado à ausência da desinência *-mos*, com peso relativo de 0.83), *tempo verbal* (pretérito imperfeito, 0.79), *sexo/faixa etária* (homens mais jovens sendo favorecedores de *a gente*: P.R. 0.92), *tipologia semântica do sujeito* (contextos em que o referente é genérico são quase categóricos ao uso de *a gente*: P.R. 0.96) e, enfim, *saliência fônica* (verbos menos saliente ‘falava/falávamos’, sendo fortes favorecedores da forma gramaticalizada: 0.71).

A autora destaca que, após a realização das três análises, identificou duas tendências diferentes associadas ao parâmetro pesquisado, ou seja, se a análise comparativa investigou os mesmos informantes, e também informantes distintos. Em relação ao estudo em painel, a investigação com os mesmos indivíduos concluiu que os informantes mais jovens e os idosos apresentaram comportamento estável em relação ao uso dos pronomes. De maneira parecida, as mulheres mais jovens apresentaram comportamento estável mesmo depois de vinte anos, enquanto as mulheres idosas demonstram comportamento linguístico parecido de quando eram mais jovens. Além disso, houve aumento no uso de *a gente* entre os homens e retração na substituição do pronome padrão pelo pronome inovador entre as mulheres.

Por outro lado, na análise do comportamento da comunidade em tendência revelou estabilidade no comportamento linguístico após vinte anos, mas há diferenças etárias significativas. Desta forma, na década de 1990, há estabilidade entre as mulheres mais velhas, com discreta retração na substituição de *nós* por *a gente*. Por outro lado, as gerações mais jovens demonstram comportamento que caminha para a mudança linguística, no sentido de *nós* está sendo substituído por *a gente* de forma acelerada, a ponto de ser categórico na amostra de 1990 com novos informantes, sendo utilizado com percentual de 100% entre os mais jovens.

Além destes trabalhos, Vianna (2006) investigou o uso de *nós* e *a gente* e as suas concordâncias de gênero e número em estruturas predicativas na fala carioca. Para tanto, a autora utilizou dois *corpora* com amostras representativas do falar carioca. O primeiro é constituído por entrevistas realizadas no início da década de 1980, extraídas do Projeto CENSO (Censo da variação linguística no estado do Rio de Janeiro), enquanto o segundo é composto por entrevistas provenientes do Projeto Peul (Programa de Estudos do Uso da Língua), gravadas na década de 2000. Em ambos os *corpora*, os informantes foram estratificados de acordo com sexo (homem e mulher), faixa etária (15 a 25 anos, 26 a 49 anos, e mais de 50 anos), e escolaridade (1º ou º grau). O objetivo da linguista foi realizar uma análise em tempo real de curta duração em tendência, comparando o comportamento do fenômeno linguístico na mesma comunidade de fala em dois momentos.

As variáveis linguísticas testadas para confrontar os dados de duas décadas foram

a *extensão semântica do referente*, *saliência fônica do verbo*, *concordância verbal*, *tempo verbal* e *concordância de gênero e número* com as formas pronominais. Os fatores extralinguísticos controlados na pesquisa consideram a *faixa etária*, o *sexo* e a *escolaridade* do entrevistado. Os resultados iniciais da pesquisa atestam maior recorrência do pronome conservador na amostra de 1990, com dados do Projeto Censo, com um total de 105 realizações das variantes de primeira pessoa do plural, sendo 58% de ocorrências do pronome *nós*, e 42% de *a gente*. Tendo em vista a segunda amostra, extraída de entrevistas sociolinguísticas representativa do falar dos anos 2000 provenientes do Projeto Peul, foram observadas 82 ocorrências da primeira pessoa do plural em estruturas predicativas, sendo 48% referentes ao pronome padrão *nós*, e 52% referentes a forma inovadora *a gente*.

A análise das variáveis revelou maior uso da forma *a gente* associado a verbos na terceira pessoa do singular, tanto na década de 1980 (84%), quanto na década de 2000 (91%), sendo a concordância verbal uma variável significativa na realização da primeira pessoa do plural. Da mesma maneira, a variável concordância de gênero e número demonstrou que, tanto nos anos de 1990 quanto 2000, o uso do *a gente* com singular é categórico, enquanto o pronome *nós* demonstra variação de número, sendo observado maior recorrência de uso com masculino singular (10 ocorrências) na amostra de 1980, e na amostra de 2000, o plural (16 ocorrências) passa a ser mais recorrente que o singular. Em relação ao sexo dos informantes, a autora constatou que, tanto entre os homens quanto entre as mulheres o pronome *a gente* é favorecido com referentes masculinos no singular, enquanto o pronome *nós* é beneficiado com referente masculino plural.

Além dos trabalhos que investigam o fenômeno em amostras do português oral, há estudos que investigam o comportamento do pronome em outras modalidades da língua. Menon, Lambach e Landarin (2003) desenvolveram uma análise em tempo real sobre a variação pronominal entre *nós* e *a gente* em 156 edições da revista Pato Donalds, publicadas nas décadas 50, 60, 70, 80 e 90 do século XIX. Com o auxílio da ferramenta computacional VARBRUL, responsável pela análise estatística, os autores trabalharam com 2059 dados dos pronomes de primeira pessoa do plural, das quais 1840 ocorrências correspondiam ao pronome *nós* (89%), enquanto 219 dados se referiam ao pronome *a gente* (11%). Os fatores selecionados pelo pacote como significativamente relevantes foram a *década de publicação*, *faixa etária*, *classe social* e *faixa etária*.

Os resultados da análise apontaram os anos 50 como fator que desfavorece quase que categoricamente o uso do pronome inovador, tendo peso relativo de 0.10, sendo esta década favorável ao uso do pronome padrão. Por outro lado, os anos de 1990 privilegiam o uso da

forma pronominal *a gente* (0.90). Em relação a classe social, os resultados obtidos pelos autores evidenciam maior probabilidade ao uso do pronome *a gente* quando o personagem pertence a classe social baixa (0.64), e maior preferência ao pronome *nós* quando o personagem pertence a classe social mais alta (0.82). E no que se refere a faixa etária, as crianças demonstraram comportamento linguístico que favorece, ainda que de maneira discreta, o pronome *a gente* (0.58), enquanto os adultos se inclinam ao uso do pronome *nós* (0.44).

Os autores ainda trataram de observar a concordância verbal com a primeira pessoa do plural, ou seja, a ausência ou a presença da desinência verbal *-mos* associada aos pronomes. Nesse sentido, foram investigadas ocorrências de *a gente + -mos*, *a gente + 3ª pessoa, nós + -mos*, e *nós + 3ª pessoa*. No entanto, somente uma única ocorrência de concordância não padrão foi encontrada, fato que impediu o prosseguimento da análise em função da concordância verbal.

Destacamos também a pesquisa de Nascimento (2013) sobre a variação na alternância entre os pronomes *nós* e *a gente* em amostras representativas do falar culto soteropolitano das décadas de 1970 e 1990 em amostras provenientes do Projeto Norma Linguística Urbana Culta - NURC/SSA. A estratificação do banco de dados leva em conta a faixa etária (25 a 35 anos; 36 a 55 anos; e 56 anos em diante), o sexo do informante (homem e mulher), a década de gravação (1970 e 1990) e a situação comunicativa: aulas e conferências (Elocução Formal - EF), diálogos menos formais (Diálogo entre dois informantes - D2), e entrevistas com os informantes (Diálogo entre informante e documentador - DID). Com o objetivo de analisar a mudança linguística em tempo real, as duas décadas foram confrontadas, tendo em vista as variáveis linguísticas nível de referencialidade do pronome, paralelismo discursivo, tempo e modo verbal, tipo de oração, tipo de verbo, tipo de texto, polaridade da frase, e as variáveis sociais estratificadas no próprio *corpus*.

Com o auxílio da ferramenta estatística GoldVarb/2001, a autora observou 554 ocorrências dos pronomes de primeira pessoa do plural na amostra analisada, sendo a década de 1970 mais favorável ao uso do pronome padrão *nós* (72%), em detrimento do pronome inovador *a gente* (28%), e a década de 1990 mais propensa ao uso do pronome *a gente* (69%), e o pronome *nós* menos recorrente nessa década (31%).

Ao confrontar as duas décadas em rodada, foram estatisticamente significativas para a realização do pronome padrão *nós* as variáveis *a década de gravação*, *sexo/gênero* e *faixa etária*. Em relação a década de gravação, os resultados apontam para favorecimento do pronome *nós* nos dados da década de 1970, com peso relativo 0.67, e forte desfavorecimento na década de 1990, com peso relativo de 0.33. Os resultados observados da variável *sexo/gênero* demonstram maior probabilidade do uso do pronome *a gente* entre os falantes do sexo

masculino (P.R. 0.87) na década de 70, enquanto a década de 1990 apresenta favorecimento da forma pronominal *a gente* tanto entre os homens (0.74) quanto entre as mulheres (0.77), demonstrando expansão do uso do pronome entre os homens. A faixa etária condiciona o uso das formas pronominais em alternância, no sentido de apresentar uma tendência em ambas as décadas: os mais jovens apresentam tendência a favorecer o pronome inovador (0.64 e 0.73, respectivamente), e os mais velhos tendem ao favorecimento do pronome padrão *nós* nas duas sincronias analisadas, sendo mais fortemente favorecido na década de 1970 (0.76), e de forma menos acentuada na década de 1990 (0.66).

Em pesquisa mais recente, Freitas, Rodrigues e Santos (2022) investigaram a variação na alternância entre os pronomes *nós* e *a gente* na fala culta de Fortaleza em duas sincronias. Para tanto, foi utilizada amostra composta por 24 informantes do PORFUCORT – Fase I (1993-1996) e outros 24 extraídos da Fase II (2018 – 2022), estratificados por sexo/gênero (homens e mulheres), faixa etária (I - 22 a 35 anos; II - 36 a 55 anos; III - 56 em diante) e fase (década de 1990 e década de 2010).

Com o auxílio da linguagem de programação *R*, por meio da ferramenta *R Studio*, os autores observaram a atuação das variáveis linguísticas *grau de referência do pronome e tempo e tipo de paradigma verbal*, e das variáveis de ordem extralinguística *sexo/gênero, faixa etária e fase de gravação*.

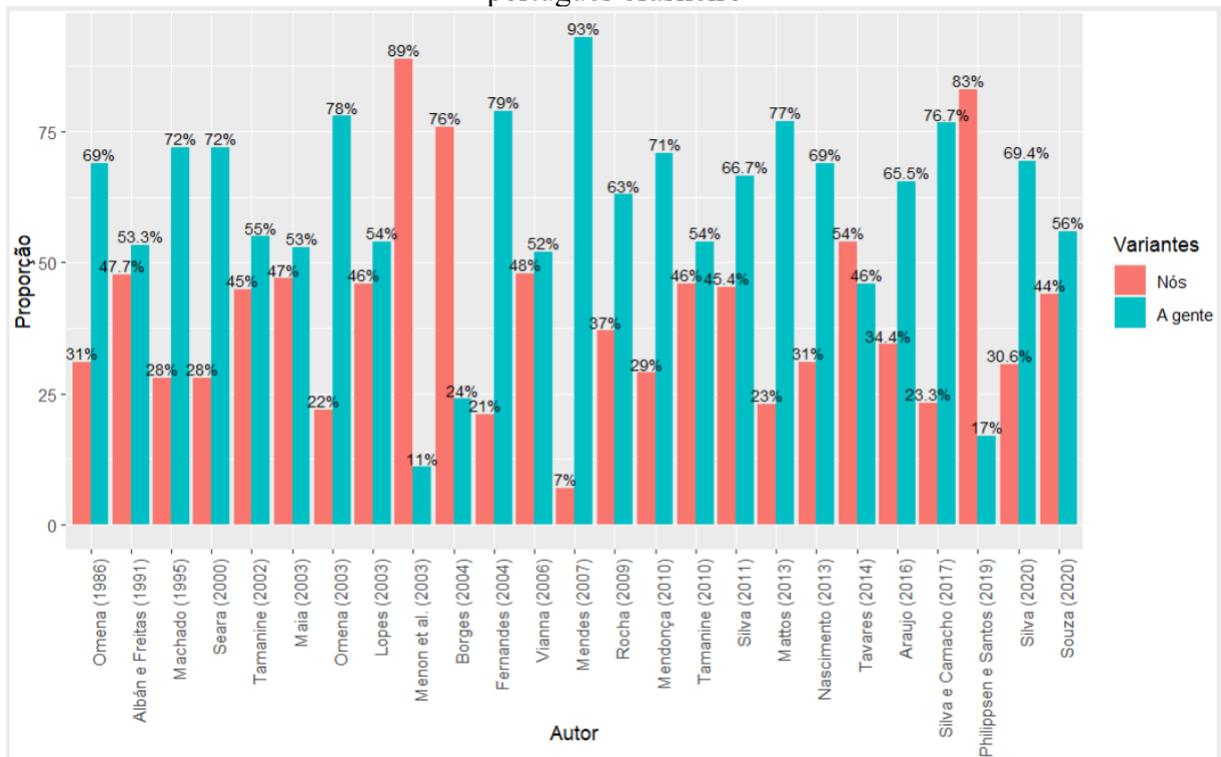
Os resultados revelam 1002 dados dos pronomes de primeira pessoa do plural na Fase I do projeto, sendo 32,82% do pronome *nós* e 62,18% de *a gente*. Além disso, o uso do pronome inovador, na década de 1990, está associado às faixas etárias mais jovens (88,29%), ao uso com referências genérica (65,69%) e ao presente do indicativo com forma igual para o passado (72,25%). Em relação a Fase II (década de 2010/2020), foram observadas 1199 ocorrências de *nós* e *a gente*, com retração no uso do pronome canônico *nós* (17,85%), e aumento significativo na proporção de uso do pronome *a gente* (82,15%). Os dados da década mais atual revelam que o pronome inovador mantém a tendência observada da década anterior: pronome *a gente* preferencialmente usado entre os mais jovens (89,79%), associado mais intensamente ao uso genérico (91,07%), e a verbos no presente do indicativo com a mesma forma para o passado (94,31%), o que parece atuar no sentido de desfazer a ambiguidade temporal dessas formas verbais.

Entre as conclusões chegadas, os autores identificaram uma tendência padrão de mudança linguística em que há um crescente uso da forma inovadora ao longo do tempo, sendo em maior proporção nas faixas etárias mais jovens da amostra. No que se refere aos dados da

Fase II, percebemos a implementação do pronome *a gente* em todas as faixas etárias, merecendo destaque o aumento significativo na proporção de uso da forma *a gente* em um intervalo de tempo de uma geração.

Ao proceder o levantamento bibliográfico dos estudos disponíveis sobre o fenômeno que pretendemos analisar neste trabalho, constatamos a presença abundante da forma *a gente*, mesmo em ambientes de fala culta, como revistas e telejornais, que exigem uso da língua mais próxima da norma padrão. Neste sentido, é possível perceber que a forma inovadora *a gente* é mais frequente que a forma pronominal padrão *nós* na maioria das pesquisas que tivemos acesso, conforme exposto nesta seção. Para observar mais claramente a comparação entre uso de *nós* e *a gente* na literatura por nós revisada, apresentamos o gráfico que demonstra as frequências de uso de cada uma das variantes nas pesquisas realizadas no Brasil:

**Gráfico 2:** Comparação da proporção de uso de *nós* e *a gente* nos estudos variacionistas do português brasileiro



Fonte: elaboração própria

Com a comparação entre as formas pronominais variáveis, podemos identificar que a variante inovadora (*a gente*) vem sendo usada com mais frequência na maioria dos estudos analisados (Alban, Freitas, 1991; Machado, 1995; Omena, 1986; Tamanine, 2002; Fernandes, 2004; Silva, 2004; Borges, 2004; Vianna, 2006; Mendes, 2007; Rocha, 2009; Mendonça, 2010; Tamanine, 2010; Silva, 2011; Mattos, 2013; Santana, 2014; Araujo, 2016; Silva e Camacho,

2017; Philipsen e Santos, 2019; Silva, 2020; Souza, 2020). Esses dados são importantes, pois fornece resultados que serão comparados com os de nossa pesquisa com base na fala culta fortalezense. Além disso, o levantamento bibliográfico aponta para uma tendência no uso dos pronomes, no sentido de a forma pronominal *a gente* ser expressivamente mais utilizada que o pronome *nós* em amostras de fala de diversas regiões do país. Buscamos verificar se a tendência observada nesses trabalhos se repete nos dados de fala culta de Fortaleza.

### 3.2.2 Panorama dos estudos sobre concordância verbal

Ainda em relação ao uso dos pronomes *nós* e *a gente*, além da alternância pronominal, o fenômeno variável manifesta outra face que não pode ser compreendida senão em função desses pronomes: a *concordância verbal de primeira pessoa do plural*, padrão e não padrão. Nesta seção, tratamos de resumir sucintamente algumas das pesquisas que tratam da concordância verbal com os pronomes *nós* e *a gente*, destacando os *corpora* utilizados em cada estudo, as variáveis testadas e os principais resultados obtidos.

Entre as pesquisas que tratam da concordância verbal com os pronomes *nós* e *a gente*, destacamos a investigação de Rubio (2012), ao analisar dados de fala do português brasileiro (PB) e do português europeu (PE). O estudo utiliza *corpus* de fala do português brasileiro composto por 64 entrevistas advindas do projeto Amostra Linguística do Interior Paulista – ALIP, mais precisamente do banco de dados da Região de São José do Rio Preto (Amostra IBORUNA), sendo 32 informantes do sexo masculino e 32 do sexo feminino, estratificados de acordo com a idade (16 a 25 anos, 26 a 35 anos, 36 a 55 anos e mais de 55 anos) e o grau de escolaridade (1º ciclo - Ensino Fundamental, 2º ciclo - Ensino Fundamental e Ensino Médio). A amostra representativa do português europeu faz parte do projeto Corpus de Referência do Português Contemporâneo (CRPC), da Universidade de Lisboa, tendo os informantes a mesma estratificação social do ALIP.

Os grupos de fatores analisados na pesquisa foram a *explicitude do sujeito*, *paralelismo*, *grau de determinação do pronome*, *tempo verbal* e *saliência fônica*. Os resultados gerais obtidos indicam, no português brasileiro, predominância do pronome *nós* associado a desinência verbal de primeira pessoa do plural *-mos*, com 85,5% das ocorrências, e preferência pelo uso da forma *a gente* com verbos na terceira pessoa do singular, tendo 94% das realizações. Já no português europeu, há emprego categórico do pronome *nós* com concordância padrão

(100%), enquanto a concordância com o pronome *a gente* apresenta comportamento variável, sendo mais recorrência a forma *a gente* sem *-mos* (75,5%), e em menor proporção a forma *a gente* com *-mos* (24,5%). O programa computacional GoldVarb X foi usado no tratamento dos dados e selecionou como estatisticamente significativo para a realização da concordância verbal com o pronome *nós* os seguintes fatores: *saliência fônica*, com verbos de saliência mínima inibindo a desinência *-mos* (0.271); *paralelismo*, sendo o uso da marca presente em verbos anteriores favorecedores da desinência *-mos* (0.816); explicitude do sujeito, tendo a forma implícita inclinação ao emprego de *-mos* (0.710); *escolaridade*, os mais velhos favorecendo o uso da concordância padrão (0.852).

O autor concluiu que a variação na concordância verbal com os pronomes *nós* e *a gente* se mostrou um fenômeno estável no português brasileiro nas amostras do interior de São Paulo, especialmente a concordância com o pronome *nós*, diferentemente do português europeu que não apresenta variação nesse contexto, sendo o uso da concordância padrão categórico na amostra analisada do português europeu.

Outro trabalho de cunho variacionista sobre a primeira pessoa do plural é a tese de doutorado de Mattos (2013) que, além de tratar da alternância pronominal, trata de pesquisar a concordância verbal padrão e não padrão com os pronomes *nós* e *a gente* na fala de Goiás. O corpus analisado neste estudo é constituído por 55 entrevistas sociolinguísticas gravadas na área urbana de Goiás, nas cidades de Anápolis, Goiânia, Cromínia, Brasília, Goianésia, Formosa, Uruaçu, Goianápolis, Hidrolina, Taguatinga, Carmo do Rio Verde, Abadiânia, Pirenópolis, Ouro Verde, Corumbá de Goiás, Ceres, Jaraguá, Niquelândia e Piracanjuba. Do total de informantes que participaram da pesquisa, 27 eram homens e 28 mulheres, divididos em função de três faixas (16 a 24 anos, 25 a 40 anos e 41 a 86 anos) e dois graus de escolaridade (ensino médio completo ou incompleto e ensino superior completo ou incompleto). As entrevistas aconteceram no ano de 2018.

As variáveis linguísticas testadas na pesquisa foram: *tipos de sujeito*, *tempo verbal*, *roteiro rítmico na forma verbal (ritmo)*, *tipo de estrutura sintática* e *tipo da fala*; as variáveis sociais foram: *sexo/gênero do falante*, *nível de escolarização* e *faixa etária*. Com o auxílio do programa estatístico GoldVarb X, foram observados 2412 dados de *nós* e *a gente* com suas respectivas concordâncias. Em relação ao pronome *a gente*, os resultados apontam forte associação com a ausência da desinência verbal *-mos* (*a gente* sem *-mos*), com 97% das ocorrências, enquanto o seu uso com verbos na primeira pessoa do plural, *a gente* com *-mos*, é pouco recorrente (3%). De maneira parecida, o pronome *nós* está inclinado ao uso da concordância verbal padrão (*nós* com *-mos*), tendo 75% das realizações com verbos de primeira

pessoa do plural, e a forma *nós* sem a desinência *-mos* aparece em 25% dos casos. Entre os fatores selecionados pelo programa estatístico, se destacam a *faixa etária*, com os mais jovens favorecendo a concordância verbal não padrão (0.82), a escolaridade, sendo os menos escolarizados mais propensos ao uso de *nós* sem a desinência *-mos* (0.80), assim como o *sexo/gênero*, com as mulheres à frente na concordância não padrão (0.69), e *roteiro rítmico na forma verbal (ritmo)*, com a manutenção da paroxítona favorecendo a queda da desinência de primeira pessoa do plural (0.58).

A pesquisadora aponta que os resultados obtidos na pesquisa demonstram uma particularidade no comportamento linguístico da comunidade de fala: o uso expressivo do pronome *nós* com verbos na terceira pessoa do singular imprime uma característica linguística da fala goiana, uma identidade que é cultural e remete às raízes rurais do povo goiano, que a praticam sem estigmatização.

Foeger (2014) conduziu estudo com objetivo de analisar a concordância verbal de primeira pessoa do plural na fala dos moradores da área rural de Santa Leopoldina. A base de dados utilizada na pesquisa é constituída por trinta e duas entrevistas sociolinguísticas provenientes do Projeto Amostra do Português Falada na Zona Rural de Santa Leopoldina, coletadas no período de novembro de 2011 a janeiro de 2023. Os informantes foram estratificados em função do gênero/sexo (feminino e masculino), da idade (07 a 14 anos, 15 a 25 anos, 26 a 49 anos e a partir de 60 anos), e escolaridade (fundamental I ou II).

As variáveis linguísticas observadas no estudo foram a *explicitude do sujeito*, o *paralelismo*, a *referencialidade*, a *função sintática*, o *tempo verbal* e a *saliência fônica*. Com o auxílio da ferramenta estatística GoldVarb X, foram encontrados 817 dados do pronome *nós*, sendo 52,5% das ocorrências com concordância verbal não padrão (*nós comeu*, *nós comemo*), e 47,5% das realizações com concordância verbal padrão (*nós comemos*). Os grupos de fatores considerados relevantes para a aplicação da concordância padrão foram: a *saliência fônica*, sendo os verbos mais salientes favorecedores da concordância padrão (*nós com -mos*) com peso relativo de 0.96; *tempo verbal*, com concordância padrão favorecida em contextos de pretérito perfeito, P.R. 0.98; *faixa etária*, sendo os informantes com mais de 49 anos aliados à concordância verbal *nós* com *-mos*, conforme demonstra peso relativo de 0.79; e *referencialidade*, estando a concordância padrão associada a referência genérica/indefinida (.96).

Os resultados obtidos na pesquisa, de um modo geral, apontam a ausência da desinência verbal *-mos* como uma marca na variedade rural de Santa Leopoldina, influenciada pela variável estilística, sendo mais propício à interação com os membros da comunidade. De

acordo com a autora, quando a entrevistadora era natural da comunidade pesquisada, a probabilidade de queda da desinência *-mos* era maior, tornando mais fértil o terreno para o uso de *nós* sem *-mos*.

Benfica (2016) traz contribuições para o conhecimento sobre os padrões de concordância verbal no português falado no Brasil, ao analisar a concordância verbal de primeira e de terceira pessoa na fala de Vitória - ES. A autora investigou as ocorrências de concordância padrão (*nós* + *-mos*) e de concordância não padrão (*nós* tava, fez, fizemo, etc) em amostras de fala extraídas de duas bases de dados distintas. A primeira delas é o projeto Português Falado Na Cidade de Vitória - PORTVIX, que conta com 46 entrevistas, gravadas entre os anos de 2001 e 2003, na capital do Espírito Santo. A segunda é uma amostra de fala casual composta por três entrevistas (Calmon, 2010). Os informantes nas duas amostras foram estratificados de acordo com o sexo/gênero (homem e mulher), faixa etária (07 a 14 anos, 15 a 25 anos, 26 a 49 anos e 50 anos ou mais) e escolarização (ensino fundamental - 1 a 8 anos, ensino médio - 9 a 11 anos e Ensino Superior - mais de 11 anos).

Com o objetivo de identificar os contextos mais propícios a cada padrão de concordância, a autora testou as variáveis *tempo e paradigma verbal, escolarização, paralelismo discursivo, faixa etária, sexo/gênero e explicitude do sujeito*. O tratamento dos dados foi realizado através da ferramenta GoldVarb X, que identificou 521 ocorrências de primeira pessoa do plural, sendo 471 realizações de *nós* com concordância padrão (90,4%), e apenas 50 manifestações do pronome *nós* sem marcação de desinência *-mos* (9,6%). Os resultados apontam favorecimento de *nós* com desinência *-mos* em verbos no pretérito perfeito com forma diferente do presente (0,797), em realização de sujeito oculto (0,710) e entre os informantes de nível universitário (0,900).

O programa computacional estatístico, no entanto, excluiu a variável *sexo/gênero*, pois não apontou diferenças significativas na concordância verbal de primeira pessoa do plural entre homens e mulheres. A autora sugere a necessidade de investigações que destaquem dinâmica social que envolve os gêneros na comunidade de fala de Vitória, a fim de entender melhor como se dá o encaixamento das formas em variação entre os gêneros.

## 4 METODOLOGIA

Sob o arcabouço teórico metodológico da Teoria da Variação e Mudança Linguística (Weinreich; Labov; Herzog, 2006), o presente trabalho objetiva analisar, em tempo real de curta duração, a realização dos fenômenos da *alternância pronominal* e da *concordância verbal de primeira pessoa do plural* e os efeitos das variáveis linguísticas e sociais sob o uso de um ou de outro pronome, tendo foco na comunidade de fala de Fortaleza – CE, em duas décadas distintas, a partir de amostras advindas do Projeto PORCUFORT nas décadas de 1990 e de 2020. Os dados obtidos a partir das amostras analisadas serão submetidos a análises estatísticas por meio da linguagem de programação *R*, através do *software R Studio*, que nos fornecerá resultados quantitativos dos quais mostraremos em forma de gráficos e tabelas, sempre com discussões sustentadas na teoria desenvolvida por Labov.

Assim sendo, nesta seção, serão explicados os métodos que utilizamos nesta pesquisa, ou seja, o modelo metodológico da Sociolinguística Variacionistas (Labov, 2008; Guy; Zilles, 2007), além do modelo de análise qualiquantitativa que dê conta de analisar os padrões de alternância pronominal e de concordância verbal de primeira pessoa do plural. Além disso, trataremos de algumas questões pertinentes a esta pesquisa, como a caracterização da comunidade de fala pesquisada, o perfil socioeconômico dos informantes envolvidos na análise de nossa amostra, a ferramenta estatística empregada no tratamento dos dados e as técnicas de coleta de dados no PORCUFORT Fase I (década de 1990) e Fase II (década de 2020).

### 4.1 Método de pesquisa empregado

Labov (2008) desenvolveu um método de análise linguística que considera o caráter heterogêneo e sistemático da variação linguística, baseado na relação intrínseca entre língua e sociedade. A partir desse modelo, foi possível explicar como a estrutura de determinada língua sofre alterações dentro de uma comunidade de fala ao longo tempo sem que a língua e a comunidade necessariamente se modifiquem, mas adquiram novas formas e novos usos. Com isso, tanto a língua como a comunidade de fala passaram a ser analisadas em situações reais de uso.

Diante disso, por estarmos analisando a língua em situações de uso real, optamos por adotar esse modelo, denominado de Teoria da Variação e Mudança Linguística (Labov, 2008), também conhecido por sociolinguística variacionista ou, ainda, sociolinguística

quantitativa. Essa denominação se deve ao fato de que, em uma pesquisa que faz uso deste modelo, opera-se com números e tratamento estatístico, nos termos de Tarallo (2001). A partir dos resultados obtidos com o emprego desse modelo metodológico, é possível depreender informações importantes para a análise, como as motivações linguísticas e sociais que podem favorecer uma ou outra regra variável, com tratamento de dados linguísticos em termos probabilísticos.

Naro (2004), por exemplo, afirma que esse modelo teórico-metodológico se propõe a analisar a força de atuação de determinados fatores sobre a realização das formas variantes que disputam lugar na fala da comunidade. No entanto, o autor pondera que, em contextos de uso real da língua, esses fatores não agem separadamente, mas em conjunto, sendo necessário isolar e medir o efeito de cada fator separadamente e, posteriormente, observar o seu efeito em interação com outros fatores.

Para Labov (2008), o ponto de partida de uma investigação linguística é a comunidade de fala. Para o sociolinguista, “a comunidade de fala não é definida por nenhum acordo marcado quanto ao uso dos elementos da língua, mas, sobretudo, pela participação em um conjunto de normas compartilhadas” (Labov, 2008, p. 120). Guy (2001), pautado no que postula Labov, elenca três características básicas que definem uma comunidade linguística: (1) os falantes precisam compartilhar traços linguísticos próprios; (2) eles devem se comunicar entre si frequentemente; (3) devem apresentar as mesmas normas e atitudes em relação ao uso da língua.

Além disso, Labov (2008) propõe que as investigações linguísticas sejam feitas a partir de dois parâmetros: a pesquisa em tempo aparente, ou seja, um estudo realizado a partir de um recorte temporal em que se analisa a amostra de uma sincronia; e a pesquisa em tempo real, isto é, a análise de duas sincronias diferentes no intervalo que corresponde a uma geração (Paiva & Duarte, 2004). Os estudos em tempo real, ainda, podem ser realizados com base em dois modelos metodológicos, a saber, os estudos em tendências e os estudos em painel, parâmetros já detalhados na fundamentação teórica (Ver capítulo 2).

Para a investigação sociolinguística dos falantes cultos<sup>7</sup> da comunidade de fala de Fortaleza-CE, utilizaremos o método de pesquisa em tempo real em tendência, já que nesse modelo de investigação é possível fazer “o rastreamento do processo histórico de mudança em diferentes épocas da língua, valendo-se o pesquisador de amostras orais” (Tavares, 2011, p.

---

<sup>7</sup> O conceito de falante culto adotado neste estudo se refere a informantes com nível superior completo.

397), o que nos permite observar, entre outras questões, se o fenômeno variável *nós* e *a gente* com suas respectivas concordâncias se encontra em processo de mudança linguística.

Além do mais, nos valem de um poderoso banco de dados de registro oral dos falantes cultos de Fortaleza, o Projeto Português Oral Culto de Fortaleza (PORCUFORT), que, atendendo ao modelo teórico e metodológico da Sociolinguística Variacionista, nos permitiu reunir amostras de duas sincronias distintas com falares representativos da década de 1990 e da década de 2020.

Ainda, a partir das gravações realizadas no projeto PORCUFORT, foi possível coletar dados de fala em dois momentos distintos (década de 1990 e décadas de 2020) para confrontar os resultados das duas amostras. A realização de uma análise comparativa em tempo real de curta duração é importante para fornecer dados que demonstram o atual estágio da variação e mudança linguística, em quais contextos a forma inovadora ganhou terreno e em quais ambientes ainda há resistência à sua implementação na fala culta fortalezense.

#### **4.2 Descrição da comunidade de fala: os falantes cultos de Fortaleza**

Esta seção dedica-se a contextualizar a comunidade da fala estudada nesta pesquisa com o estudo sociolinguístico que pretendemos desenvolver. Por tratarmos somente dos falantes cultos de Fortaleza, necessitamos definir o conceito de “português culto” que adotaremos aqui. De acordo com Bagno (2003) a definição de norma culta se refere à linguagem concretamente empregada pelas pessoas pertencentes aos segmentos mais favorecidos da população, tendo seus falantes cultos definidos em função de dois critérios básicos, a saber: possuir ensino superior básico e ter antecedentes culturais urbanos. Essa definição, vale lembrar, vem sendo empregada em diversos *corpora* do português brasileiro como, por exemplo, o Projeto Norma Urbana Culta – NURC. Portanto, quando tratamos do falar culto de Fortaleza, estamos nos referindo à fala dos fortalezenses que possuem nível superior completo.

As amostras que analisamos constituem um recorte do Projeto Norma Urbana Culta de Fortaleza – PORCUFORT, composto por entrevistas sociolinguísticas que foram gravadas nas décadas de 1990 (Fase I) e de 2020 (Fase II), respectivamente. Por isso, consideramos importante tratarmos, mesmo que de forma breve, de alguns dados referentes à população com acesso ao nível superior em cada uma das décadas, para entendermos, ainda que de forma pouco aprofundada, como as questões sociais, políticas e culturais podem exercer força sobre a fala dos informantes. Para tanto, recorreremos, sempre que possível, a informações advindas do

Censos demográficos de 1991, 2000 e 2010, produzidos pelo Instituto Brasileiro de Geografia e Estatística - IBGE, além de outros trabalhos que tratam do assunto.

Fortaleza é um município brasileiro, capital do estado do Ceará, localizado no nordeste do país, a uma distância 2.285 km de Brasília, a capital federal. A cidade se desenvolveu às margens do Riacho Pajeú e tem seu marco inicial em alusão ao Forte Schoonenborch, de onde foi criado o município edificado pelos holandeses durante sua segunda estada no local, entre 1649 e 1654.

A cidade está localizada na costa do oceano Atlântico, a uma altitude média de dezesseis metros acima do nível do mar, com 34 km de praia, além de possuir uma área de 313.140 km<sup>2</sup> e uma população estimada de 2.686.612 habitantes em 2022.

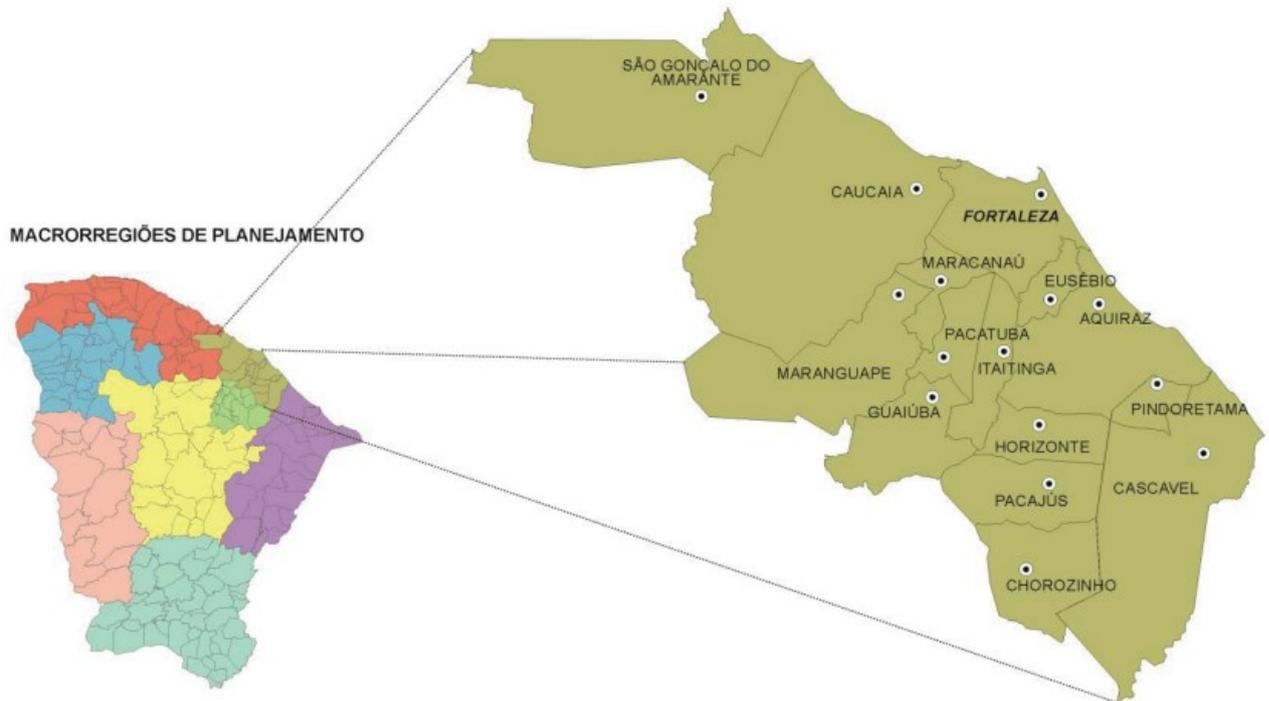
**Figura 2:** Imagem aérea do litoral de Fortaleza-CE



Fonte: Acervo Fortaleza em foto

Além disso, Fortaleza também possui a maior densidade populacional das capitais do país com 7.786,44 habitantes/quilômetro quadrado. Em termos de população, é a maior cidade-estado do Ceará e a quinta maior cidade-estado do país. A região metropolitana de Fortaleza é a sexta mais populosa do Brasil e a primeira das regiões Norte e Nordeste, com 4.167.996 habitantes em 2021. É a cidade-estado nordestina com maior influência regional e a terceira maior rede urbana do Brasil, em termos de população, atrás apenas de São Paulo e Rio de Janeiro.

**Figura 3:** Mapa da região metropolitana de Fortaleza



Fonte: Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE)

Em 2018, Fortaleza foi a nona cidade mais rica do país em termos de PIB e a primeira mais rica do Nordeste com 67 bilhões de reais. Também é classificada como a terceira metrópole mais rica do Norte-Nordeste. É um importante centro industrial e comercial do nosso país, tendo a população o oitavo maior poder aquisitivo municipal brasileiro.

Em termos de turismo, a cidade é o segundo destino mais popular do país e a quarta cidade brasileira que mais recebe turistas, segundo o Ministério do Turismo. Além disso, Fortaleza é a sede do Banco do Nordeste, da Transnordestina Logística e do Departamento Nacional de Obras Contra as Secas - DNOCS. Ainda, vale lembrar, a BR-116, a rodovia mais importante do país, começa em nosso município.

Para que seja possível explicar o perfil dos falantes cultos de Fortaleza, é necessário conhecermos o contexto sócio-histórico que o Brasil vivenciava a partir do início da década de 90, e como esse contexto refletiu na capital cearense.

#### ***4.2.1 Um breve contexto sócio-histórico***

No início dos anos de 1990, a economia do nosso país passava por um momento

delicado, marcado por instabilidade política, má distribuição de renda, inflação exagerada, índices alarmantes de crescimento e uma enorme dívida externa herdada dos anos 80. Foi nesse cenário de muita insegurança que, em março de 1990, Fernando Collor iniciou seu governo, adotando medidas econômicas impopulares, na tentativa de frear a enorme taxa de inflação, que à época chegava aos 1700%.

Como parte das medidas implementadas pelo Plano Brasil, o governo Collor confiscou a poupança dos brasileiros por 18 meses na tentativa de amenizar a situação econômica na qual o país estava mergulhado. No entanto, ao invés da economia reagir positivamente às intervenções feitas pelo governo, tivemos como resultado uma enorme recessão econômica, dando início a abertura indiscriminada do comércio e do capital financeiro ao exterior. De acordo com Antunes (2001, p. 64), várias medidas foram tomadas para “facilitar a entrada de capitais estrangeiros tais como: a abertura de contas correntes por não-residentes, permissão para os investimentos em bolsas de valores, títulos públicos e renda fixa”.

Diante desse cenário, associado a falta de habilidade política, ao pouco apoio partidário e aos diversos escândalos de corrupção que assolavam seu governo, em 1992 o então presidente Collor passou por um processo de impedimento, conhecido como *impeachment*, sendo retirado da presidência do Brasil e dando lugar ao seu sucessor: Itamar Franco, que permaneceu na presidência da república até 1994. O Governo de Itamar se situou em um período de transição e, além de não cumprir com o que havia prometido (a melhoria da economia), demorou para tomar qualquer ação mais enérgica em relação à política econômica no país, substituindo três vezes o ministro da Fazenda em apenas seis meses. O quarto ministro da fazenda do governo transitório de Franco, Fernando Henrique Cardoso, apresentou vários planos diretos de ações com o intuito de promover a economia e controlar a inflação, entre eles está o Plano de Ação Imediata, que consistia basicamente em cortar os gastos do governo. Apesar das ações executadas pelo novo ministro, a inflação chegava na casa dos 20% ao mês, situação que só foi controlada com o Plano Real.

Como esperado, essa situação de insegurança que pairava sobre o Brasil refletia na cidade de Fortaleza. Guimarães (2019, p. 107), ao se referir a situação política do Brasil nos anos 90, assevera que “toda essa turbulência de acontecimentos no Brasil causava impactos em todos os estados, inclusive no Ceará”. Esses impactos foram sentidos principalmente no processo de urbanização da cidade e no acesso ao ensino superior.

Em relação ao processo de infraestrutura e urbanização, muitas ações foram realizadas durante o período para proporcionar um desenvolvimento urbanístico para Fortaleza.

Dentre essas ações podemos destacar as grandes obras<sup>8</sup> realizadas durante a gestão do prefeito Juraci Magalhães, que esteve à frente da Prefeitura de Fortaleza de 1990 a 1992, e depois de 1997 a 2004. As ações realizadas pelo prefeito eram divulgadas com frequência na imprensa e incluíam:

crescimento horizontal da cidade, mapeando as principais obras de mobilidade urbana, tendo como ação a abertura, duplicação, alargamento, prolongamento e recuperação de várias avenidas. A construção de viadutos, giradores e alargamentos de acesso foram os meios que a gestão utilizou para interligar estas vias (Saldanha & Benevides, 2013, p. 19).

No entanto, cabe aqui pontuar que, apesar das diversas realizações urbanísticas, a população estava descontente com a gestão do prefeito, devido, principalmente, à morosidade das obras e a destruição de parte da flora local. Além disso, as ações do prefeito contribuíram diretamente para a periferação da população, favoreceu a especulação imobiliária e estimulou o uso do transporte individual.

**Figura 4:** Av. Domingos Olímpio antes do alargamento (08/01/1992)



**Fonte:** Acervo Fortaleza em foto.

<sup>8</sup> De acordo com Benevides e Saldanha (2013) as principais obras realizadas foram a construção da Avenida Raul Barbosa, duplicação da Avenida Padaria Espiritual, Perimetral, alargamento da Avenida Leste Oeste e do acesso à Avenida Frei Cirilo, Avenida Godofredo Maciel, abertura da Av. Rogaciano Leite, implantação da Av. Osório de Paiva, construção da Avenida da Urucutuba, Avenida da Independência, Avenida Jardim Fluminense, prolongamento da Avenida Cônego de Castro, Avenida Maria Gomes de Sá, Avenida Thompson Bulcão, Avenida Borges de Melo, Avenida Paulino da Rocha, alargamento da Av. Expedicionários, continuação da Justiniano de Serpa até a Av. Duque de Caxias, Avenida Parsifal Barroso, recuperação da Av. Eduardo Girão alargamento da Avenida Castro e Silva. continuação da Avenida Padre Antonio Thomaz, Avenida Pontes Vieira, Via Expressa, Avenida do Canal e Aguanambi. Construção do girador ligando a BR 116, Ponte ligando bairro Conjunto Ceará ao bairro Benibaú, Viadutos da Avenida 13 de maio, Santana Junior e Mister Hull, Drenagens da Avenida Santos Dumont com Avenida Desembargador Moreira, Avenida Barão de Studart com Rua Pinto Madeira e da Rua Padre Mororó com Avenida Duque de Caxias.

Se de um lado a capital cearense vivenciava um momento de grandes investimentos nas áreas de mobilidade urbana, saneamento e urbanização geral, por outro percebemos uma realidade muito diferente no que se refere ao acesso ao ensino superior em Fortaleza e em todo Brasil. Devido aos cortes nos gastos públicos promovidos pelo Plano de Ação Imediata, as universidades públicas passaram por uma grande onda de sucateamento, oferecendo números limitados de vagas e apresentando dificuldades com a impossibilidade de preencher o quadro total de professores, tornando cada vez mais difícil o acesso da população ao ensino superior. Por conta disso, na década de 1990, “o ritmo de crescimento da matrícula na esfera privada mostra-se bem mais intenso que o do conjunto das redes públicas” (Corbucci, 2001, p 105).

De acordo com Corbucci (2001), esse crescimento do número de matrículas em universidades privadas se dava, além da retração municipal, por conta de diversas políticas públicas desenvolvidas pelo MEC para incentivar a expansão das instituições privadas no país. Todo esse cenário trouxe como reflexo, em 1999, um total de 2,38 milhões de alunos matriculados no ensino superior, sendo 65% das matrículas realizadas somente em instituições de ensino privadas. A expansão de instituições privadas no Brasil, entre os anos de 1990 e 1998, pode ser atestada no quadro a seguir:

**Quadro 3:** Evolução do Número de Instituições de Ensino Superior no Brasil - 1980-1998

<b>Ano</b>	<b>Total</b>	<b>Federal</b>	<b>Estadual</b>	<b>Municipal</b>	<b>Privada</b>
<b>1990</b>	<b>918</b>	55	83	84	696
<b>1991</b>	<b>893</b>	56	82	84	671
<b>1992</b>	<b>893</b>	57	82	88	666
<b>1993</b>	<b>873</b>	57	77	87	652
<b>1994</b>	<b>851</b>	57	73	88	633
<b>1995</b>	<b>894</b>	57	76	77	684
<b>1996</b>	<b>922</b>	57	74	80	711
<b>1997</b>	<b>900</b>	56	74	81	689
<b>1998</b>	<b>973</b>	57	74	78	764

Fonte: Inep (2000)

Conforme exposto no gráfico 2, é possível observar que o número de instituições privadas no Brasil se mantém, durante toda a década, expressivamente acima do número de

instituições de ensino públicas. Nos primeiros anos, é possível constatar que a quantidade de instituições privadas se manteve relativamente estável e que, a partir de 1994, houve um crescimento desse número devido a incentivos do governo federal para a expansão de universidades, faculdades e centros universitários privados.

No que se refere a demanda das instituições superiores nos anos 90, o autor identificou duas tendências distintas: o aumento na procura por vagas nas instituições públicas (tendo em média 8,3 candidatos por vaga disputada) e uma diminuição na procura por vagas nas faculdades privadas (cerca de 2,3 candidatos por vaga disputada). Corbucci (2001) acredita que essas tendências refletiram diretamente, ao longo da década de 90, a perda do poder aquisitivo da classe média, principalmente nos primeiros anos da década.

Foi somente a partir de 1994 que se notou um crescimento de matrículas no ensino superior, com taxa de crescimento anual de 4%. No entanto, segundo informações divulgadas pelo Instituto Brasileiro de Geografia e Estatística (1997), mesmo com esse crescimento nas matrículas do ensino superior na década de 1990, a quantidade de vagas era insuficiente para atender a demanda da população em todo país. A situação era ainda mais complicada no Nordeste, onde se concentrava um grande volume de jovens (cerca de 25,7%), fazendo com que as vagas fossem ainda mais escassas.

Todo esse contexto, obviamente, refletia no estado do Ceará que, seguindo a tendência do Brasil, não possuía oferta de vagas suficientes em suas instituições de ensino superior para atender a demanda da população. Abaixo, é possível observar a comparação entre a o número de vagas no ensino superior ofertadas no Brasil, no Nordeste e no Ceará, e o número de inscritos no vestibular em cada ano, conforme exposto na tabela que segue:

**Quadro 4:** Comparação entre o número de vagas ofertadas e o número de inscrições no vestibular entre os anos de 1990 e 1998

Ano	Vagas ofertadas por Federação			Número de inscritos no vestibular		
	Brasil	Nordeste	Ceará	Brasil	Nordeste	Ceará
1990	502.784	62.088	9.405	1.905.498	304.151	42.313
1991	516.663	64.618	8.765	1.985.825	328.403	52.915
1992	534.847	65.807	8.890	1.836.859	279.331	42.125
1993	548.678	66.170	8.500	2.029.523	320.137	46.117
1994	574.135	68.152	9.100	2.237.023	358.569	58.747
1995	610.355	73.652	9.560	2.653.853	457.077	66.865

<b>1996</b>	634.236	76.055	9.462	2.548.077	429.567	75.261
<b>1997</b>	699.198	81.819	10.121	2.711.776	421.109	68.882
<b>1998</b>	776.031	90.556	11.224	2.858.016	474.844	78.471

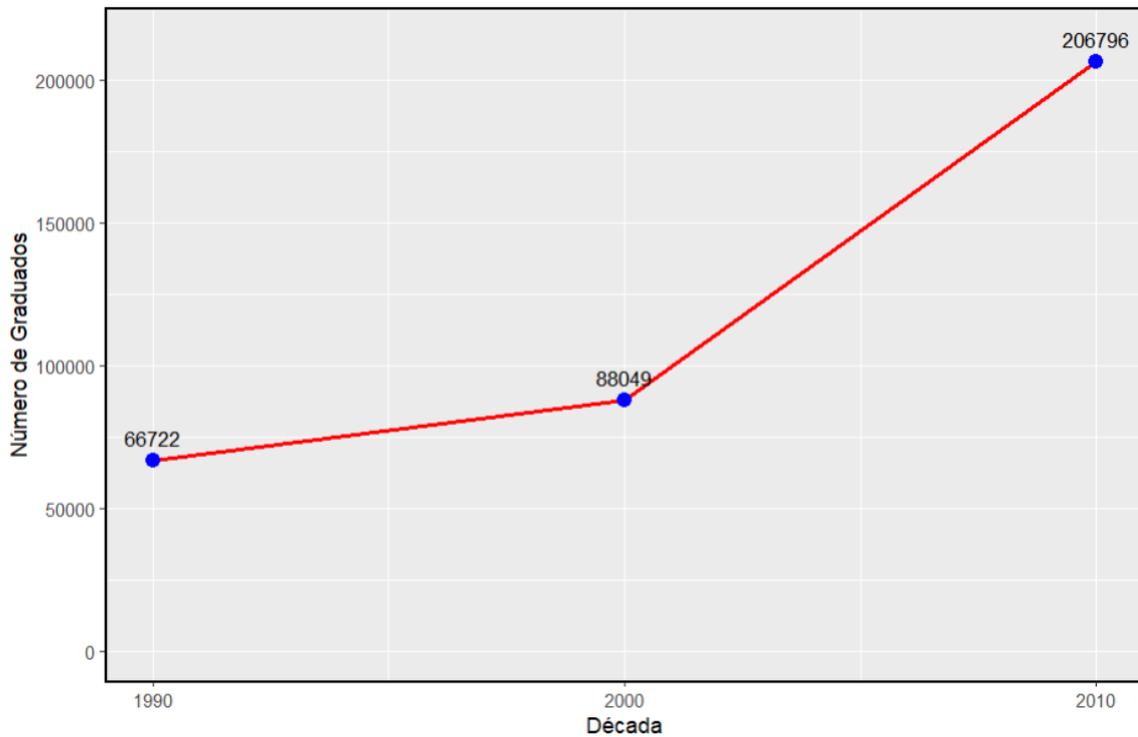
**Fonte:** elaboração própria baseada em Inep (2000)

De acordo com as informações divulgados nos Censos do Ensino Superior da década de 1990, e resumidos no quadro acima, a quantidade de vagas ofertadas no Brasil e, conseqüentemente, no Ceará eram muito aquém da demanda apresentada pela sociedade à época. Nos primeiros anos, a realidade educacional do Brasil se traduz em um número de inscritos no vestibular mais de quatro vezes maior do que a quantidade de vagas oferecidas. A mesma realidade pode ser vista no Ceará, que em 1990 teve uma oferta de 9.405 vagas pelas instituições de ensino superior e uma demanda de mais de 40 mil inscritos nos exames vestibulares no mesmo ano. Por conta dessas questões, compartilhamos do pensamento de Araújo, Viana e Pereira (2018, p. 164) quando as autoras afirmam que os indivíduos que compõem o projeto PORCUFORT Fase I são, de alguma maneira, “pessoas socioeconomicamente privilegiadas, pois, na década de 90, ter o ensino superior completo era, como afirma Bagno (2003) e Serraine (1984), um forte indicativo de nível sociocultural elevado”.

Este era o cenário em que estavam inseridos os falantes cultos na época das gravações da primeira fase do projeto PORCUFORT, ou seja, uma realidade bastante distinta da que vivenciavam os falantes cultos da segunda fase do projeto, iniciada em 2018.

Segundo dados do Censo de 2010, a população de graduados em Fortaleza-CE chegou a 206.796 pessoas. Apesar de ser um número aparentemente alto, comparada a população total de 2.452.185 de pessoas, fica nítido que a cidade de Fortaleza ainda carece de instituições de ensino para atender a população. Para melhor compreensão sobre a realidade da população com curso superior completo na capital cearense, o Gráfico 3 traz dados que nos mostra o crescimento do número de pessoas formadas no ensino superior no município, de acordo com o IBGE, nos censos realizados entre as décadas em que foram realizadas as gravações de nossa amostra:

**Gráfico 3:** Crescimento da população de graduados nas décadas de 1990, 2000 e 2010



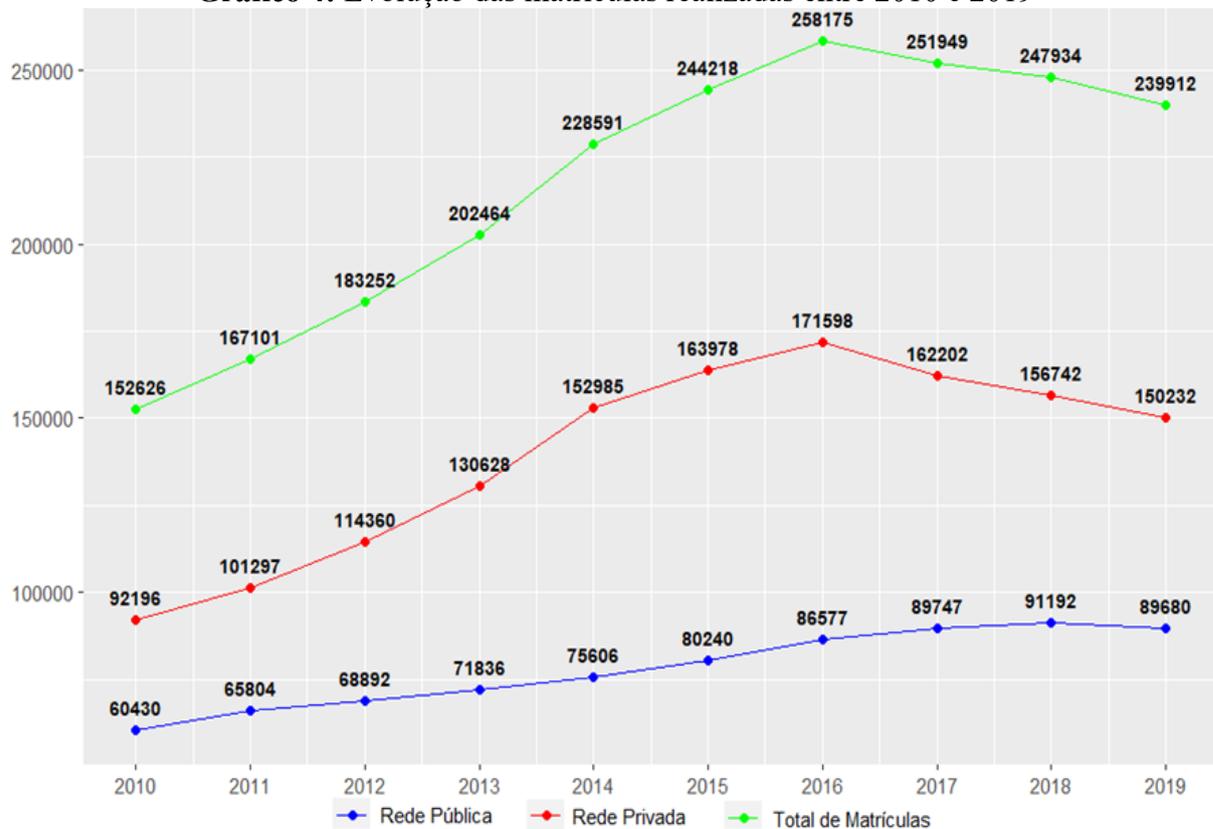
Fonte: elaboração própria, com dados do IBGE

Conforme observamos no gráfico acima, de 1990 para 2010, houve um crescimento muito expressivo na população com nível superior completo, um aumento que chega a 68%. Esse número reflete o contexto educacional vivenciado pelo Brasil entre os anos 2000 e 2010, principalmente devido a políticas públicas voltadas para o acesso da população à universidade, implementadas pelo governo do presidente Luiz Inácio Lula da Silva, tendo exercido o cargo de 1 de janeiro de 2003 a 1 de janeiro de 2011. Programas instituídos na gestão do ex-presidente, como o programa Universidade para Todos (ProUni), o Programa de Financiamento Estudantil (Fies), o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (Reuni), além do aumento da oferta de cursos superiores a distância (EAD), contribuíram efetivamente para facilitar o acesso da população às universidades e faculdades.

O Mapa do Ensino Superior, documento desenvolvido pela Secretaria de Modalidades Especializadas de Educação (Semesp), apresenta dados sobre a educação superior no Ceará em 2018, ano em que foram iniciadas as gravações para compor o banco de dados PORCUFORT - Fase II. De acordo com o documento, o Ceará registrou em 2018 mais de 300 mil matrículas, distribuídas em 83 instituições de ensino superior (IES) que oferecem ensino presencial, e outras 65 na modalidade EAD, sendo essas instituições concentradas na capital e

região metropolitana: 63,8% na modalidade presencial e 36,2% na modalidade à distância. Além disso, é importante pontuarmos que, no que se refere a modalidade de ensino a distância (EAD), a grande maioria dos inscritos se concentram na rede privada de ensino, com um percentual que chega a impressionantes 98%. Se compararmos a evolução do total das matrículas realizadas durante a década de 2010, fica claro que o ensino superior privado se destaca por atender a maior parte da demanda, conforme indica o gráfico abaixo:

**Gráfico 4:** Evolução das matrículas realizadas entre 2010 e 2019



Fonte: elaboração própria com dados extraídos da Semesp (2019)

Apesar de todos esses mecanismos que facilitam o acesso ao ensino superior no Ceará, o número de vagas ofertadas ainda se mostra insuficiente se comparado a demanda da população. De acordo com as informações divulgadas pelo Censo do Ensino Superior de 2019, as vagas oferecidas pelas IES no Ceará, e em todo o Nordeste, representam apenas um terço dos inscritos nos vestibulares em 2018, conforme exposto no quadro abaixo:

**Quadro 5:** Comparativo entre vagas ofertadas, inscritos no vestibular e ingressos no ano de 2018 no estado do Ceará

	Vagas Ofertadas	Inscritos no Vestibular	Ingressos
<b>Ceará</b>	140.531	516.891	60.837
<b>Pública</b>	22.992	347.793	20.506
<i>Federal</i>	14.446	276.957	13.276
<i>Estadual</i>	8.546	70.836	7.230
<b>Privada</b>	117.539	169.098	40.331

Fonte: Elaboração própria baseado em Inep (2020)

Com a discussão realizada e apresentação dos dados nesta seção, foi possível compreender melhor o cenário educacional que pairava sobre a comunidade de fala que propomos estudar, contextualizando-a com o panorama político e educacional do que ocorria em cada uma das décadas, para melhor entender como todos os acontecimentos refletiram na formação dos informantes cultos de Fortaleza. A seção que segue apresentará as considerações pertinentes sobre o banco de dados e a amostra selecionada para este estudo.

### 4.3 Os corpora e as amostras

Bancos de dados representativos da fala têm sido fonte frutífera para investigações sociolinguísticas desenvolvidas em todo mundo. No Brasil, a partir dos anos 1970, acervos de perfil sociolinguístico começaram a ser constituídos, dentre eles temos o projeto NURC – Norma Urbana Culta que, em 1968, que iniciou trabalho para captar e descrever a fala de cinco capitais. A partir no ano de 1973, o projeto NURC deu início a formação de amostra da fala culta do Rio de Janeiro, composta por três tipos de inquérito: diálogo entre dois informantes (D2), diálogo entre informante e documentador (DID) e elocuições formais (EF). Essa amostra era formada somente por informantes de nível superior, sendo homens e mulheres, de idades diversificadas.

Além do NURC, merece destaque o Programa de Estudo sobre o Uso da Língua (PEUL), que, coordenado pelo professor Anthony Naro, em 1979 iniciou coleta de amostra com base na fala popular carioca. Com um total de 64 informantes, o projeto possibilitou a realização de diversas pesquisas sociolinguísticas com base na Teoria da Variação e da Mudança Linguística.

Outros *corpora* ainda se formaram ao longo dos anos, como podemos citar o Projeto

do Atlas Etnolinguístico dos Pescadores do Estado do Rio de Janeiro (APERJ), o Projeto Análise Contrastiva de Variedades do Português (VARPORT), o Atlas Linguístico do Brasil (ALIB), o Português Não-Padrão no Ceará (ELOC), o Dialetos Sociais Cearenses (DSC)<sup>9</sup>, e o Projeto Português Oral Culto de Fortaleza (PORCUFORT), que mais recentemente constituiu uma segunda amostra com dados representativos do falar culto fortalezense dos anos finais da década de 2010 e dos anos iniciais da década de 2020. Esses bancos de dados se figuram como um importante aparato metodológico para o desenvolvimento de pesquisas, tanto em tempo aparente, quanto em tempo real.

Para esta pesquisa, como já explanado anteriormente, dispomos de dois bancos de dados de fala culta. Um deles, o PORCUFORT Fase I, que captou amostra representativa da fala fortalezense dos anos de 1990, encontra-se totalmente digitalizado e disponível para realização de pesquisas sincrônicas. O segundo deles, o PORCUFORT Fase II, teve as entrevistas iniciadas em 2018 e concluídas em 2022, tendo gravações realizadas, inclusive, durante a época de isolamento social provocado pela pandemia de Covid-19.

Em relação ao primeiro projeto, o PORCUFORT Fase I, de acordo com Araújo, Viana e Pereira (2018), este foi constituído no período de 1993 a 1995, coordenado pelo professor José Lemos Monteiro, da Universidade Estadual do Ceará, seguindo os mesmos procedimentos técnicos e metodológicos empregados no Projeto Norma Urbana Culta do Brasil (NURC). Assim, o PORCUFORT Fase I é um único banco de dados disponível atualmente que retrata a fala culta dos fortalezenses da década de 1990 (Araújo; Viana; Pereira, 2018a). Este banco de dados teve a seleção de seus informantes realizada de acordo com os pressupostos da teoria desenvolvida por Labov (1972), que estabelece critérios que devem ser levados em conta ao selecionar o informante que irá participar das gravações para o banco de dados, com o objetivo de relatar de maneira fiel o falar de uma comunidade de fala. Sobre esses critérios, as autoras afirmam:

Os informantes do PORCUFORT apresentam, ainda, as seguintes características: são fortalezenses natos; filhos de pais cearenses; mantêm residência fixa em Fortaleza; aqueles que realizaram viagens para fora de Fortaleza só o fizeram por um curto período; possuem nível superior completo. (Araújo; Viana; Pereira, 2018<sup>a</sup>, p. 183).

A Fase I do projeto, que está sob a tutela da professora Aluiza Alves de Araújo – UECE, é composta por gravações em áudio de 73 informantes, estratificados de acordo com

---

<sup>9</sup> Os inquéritos descritos, bem como o áudio das entrevistas dos projetos ELOC e DSC estão disponíveis no site do Projeto PROFALA, que conta ainda com banco de dados representativo do Português Africano. Acesse o link: <https://profala.ufc.br/pt/>

sexo (homem e mulher), faixa etária (I - 22 a 35 anos, II - 36 a 55 anos e III - 56 anos em diante) e tipo de registro (Diálogo entre Informante e Documentador– DID, Diálogos entre dois Informantes – D2 e Elocuções Formais – EF) e, atualmente, está totalmente transcrito e digitalizado, disponível para realizações de pesquisas sociolinguísticas (Araújo, Viana e Pereira, 2018). Assim, o banco de dados está distribuído, de acordo com as variáveis sociais, da seguinte maneira:

**Quadro 6:** Estratificação dos informantes do PORCUFORT Fase I (década de 1990)

<b>Fase I (1990)</b>						
<b>Sexo</b>						
<b>Masculino</b>			<b>Feminino</b>			
<b>Tipo de Registro → Faixa etária ↓</b>	<b>D2</b>	<b>DID</b>	<b>EF</b>	<b>D2</b>	<b>DID</b>	<b>EF</b>
<b>I (22 a 35 anos)</b>	6	4	3	4	8	4
<b>II (36 a 55 anos)</b>	6	4	5	5	4	3
<b>III (56 anos acima)</b>	6	3	2	3	3	0
<b>Total</b>	<b>18</b>	<b>12</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>7</b>
<b>Total Geral</b>	<b>73 informantes</b>					

Fonte: Adaptado de Araújo, Viana e Pereira (2018)

A ideia inicial do projeto, conforme explicam Araújo, Viana e Pereira (2018) era alocar, pelo menos, cinco informantes em cada célula para obter uma distribuição equitativa desses informantes em relação às variáveis sociais. No entanto, como explicado na seção anterior, o acesso da população fortalezense ao ensino superior era bastante precário, o que dificultou o trabalho dos pesquisadores envolvidos na coleta de dados para encontrar informantes que preenchessem o perfil necessário para participar do *corpus*. Por esse motivo, percebemos que algumas células apresentam apenas dois ou três informantes e, ainda, existe uma célula que não possui qualquer informante com o perfil que o tornasse elegível para a gravação: tratasse de uma célula do registro do tipo Elocução Formal (EF), que deveria ser preenchida por informantes do sexo feminino com mais de 55 anos.

No entanto, a ausência desse informante na célula não prejudicou a composição do *corpus* e nem tampouco a realização de pesquisas com dados advindo deles, já que o projeto

PORCUFORT vem sendo usado como fonte de dados da língua falada em diversas pesquisas sociolinguística sob a luz da Teoria da Variação e da Mudança Linguística, entre livros publicados, artigos científicos, dissertações de mestrado e teses de doutorado.

Entre esses, para citar alguns, podemos destacar o livro *Estudos sobre a língua culta falada em Fortaleza: explorando dados do Porcufort* (Pinheiro, 2010), que traz uma coletânea de 9 textos com estudos sobre fenômenos de variação linguística sob a ótica da Teoria da Variação da Variação e Mudança Linguística. Além deste, destacamos as dissertações defendidas por Araújo (2000), que estudou a monotongação dos ditongos [aj], [ej] e [ow]; e por Almeida (2017), que trata da realização das vogais médias pretônicas /e/ e /o/. Em relação às teses de doutorado que utilizaram o *corpus* como fonte de dados, destacamos as pesquisas de Guimarães (2019) sobre a variação entre os pronomes *tu* e *você* na fala culta fortalezense e o trabalho de Rodrigues (2018) sobre a realização das africadas. Mais recentemente, temos a pesquisa de Pereira (2023) sobre o uso dos verbos existenciais, sendo estas teses contempladas com a realização de testes de atitudes linguísticas.

Conforme é possível observar, o projeto PORCUFORT tem sido um poderoso instrumento para produção de pesquisas variacionistas em tempo aparente, pois atende os requisitos para o desenvolvimento de banco de dados para que pesquisas sociolinguísticas sejam realizadas. Além disso, vale lembrar que o banco de dados passou por submissão e aprovação do Comitê de Ética da Universidade Estadual do Ceará - UECE, facilitando o trabalho do pesquisador que necessita de um *corpus* organizado com maestria e que possua credibilidade acadêmica.

No que diz respeito à segunda fase do PORCUFORT, o projeto foi idealizado pela professora Dra. Aluiza Alves de Araújo, com o intuito de criar um banco de dados representativo do falar culto do início do século XXI e, além disso, servir como *corpus* para análises em tempo real de fenômenos variáveis. *Corpora* com dados disponíveis com falares representativos de duas décadas, para realização de estudos diacrônicos, ainda são raros no Brasil. Com isso, a Fase II do projeto surge como aparato linguístico para que análises comparativas de duas décadas possam ser desenvolvidas. Em 2018, as gravações para o novo banco de dados foram iniciadas, seguindo os procedimentos metodológicos estabelecidos pela teoria variacionista (Labov, 2006 [1966]), os mesmos procedimentos empregados na composição da Fase I do projeto.

Um ponto importante de lembrar é que o PORCUFORT Fase II, com dados representativos do falar culto fortalezense da década de 2020, já possui entrevistas finalizadas, transcritas e revisadas, e gravações que se distribuem nos três tipos de inquérito: D2, DID e EF.

O quadro 7 apresenta a distribuição dos informantes que compõem a base de dados, em função da estratificação social adotado no *corpus*:

**Quadro 7:** Estratificação dos informantes do PORCUFORT Fase II (década de 2020)

	Fase II (2020)					
	Sexo					
	<i>Masculino</i>			<i>Feminino</i>		
<b>Tipo de Registro</b> → <b>Faixa etária ↓</b>	<b>D2</b>	<b>DID</b>	<b>EF</b>	<b>D2</b>	<b>DID</b>	<b>EF</b>
<b>I (22 - 35 anos)</b>	6	4	6	6	6	6
<b>II (36 - 55 anos)</b>	6	6	6	6	7	6
<b>III (56 anos em diante)</b>	7	6	6	6	7	3
<b>Total</b>	<b>19</b>	<b>16</b>	<b>18</b>	<b>18</b>	<b>20</b>	<b>14</b>
<b>Total geral</b>	<b>105 informantes</b>					

Fonte: Adaptado de Araújo, Viana e Pereira (2018)

Conforme mencionamos, o *corpus* é dividido em três tipos de inquéritos, a saber, D2, DID e EF. Desses, os mais formais, ou seja, aqueles que apresentam um alto nível de monitoramento da própria fala, são os inquéritos do tipo EF (Elocução Formal), pois esses foram gravados em situações de aulas em escolas, palestras em universidades, sermões na igreja. Desta forma, temos a hipótese de que o uso das variantes padrão ocorre mais frequentemente em contextos formais, pois esse tipo de inquérito transmite “a ideia de que o maior monitoramento tende a aproximar a fala da variante de prestígio (desde que o falante a domine)” (Butzge, 2009, p.13).

Já os inquéritos do tipo D2 (diálogo entre dois informantes) apresentam gravações mais espontâneas, com baixo nível de monitoramento, e há pouca, ou quase nenhuma, interferência do documentador. Dessa forma, dentre os inquéritos, o tipo D2 é “o menos formal de todos, já que, neste tipo de inquérito, os informantes, necessariamente, são familiares ou amigos” (Araújo, 2011, p. 9).

O terceiro deles, os inquéritos do tipo DID (Diálogo entre Informante e Documentador) apresentam nível intermediário de formalidade, haja vista que, nesse tipo de inquérito, as entrevistas eram sempre conduzidas pelos documentadores, que encaminhava o diálogo para que o informante falasse de experiências pessoais, “de forma que este possa

esquecer de que está em uma situação de entrevista, deixando, se possível, de prestar atenção à própria fala” (Araújo; Viana; Pereira, 2018<sup>a</sup>, p. 185).

Além de estratificado pelo tipo de inquérito, o *corpus* também foi estratificado de acordo com o sexo (masculino e feminino) e nas três faixas etárias (I - 22 a 35 anos, II - 36 a 55 anos, e III - 56 anos em diante), ou seja, o PORCUFORT Fase II foi organizado seguindo as mesmas metodologias empregadas na composição da primeira fase do projeto.

No entanto, nossa amostra deve contemplar todos os perfis sociolinguísticos presentes nos *corpora*, a fim de distribuir as células de forma homogênea e, com isso, alcançar dados mais representativos do falar culto de Fortaleza. Para tanto, foi necessário proceder um ajuste nas faixas etárias, a fim de preencher todas as células da amostra. Esse ajuste foi necessário devido à falta de informantes femininos da faixa etária III em inquéritos do tipo EF (pois estas células contam com zero informantes na Amostra 1990). Por conta disso, adotamos uma nova caracterização para a faixa etária (I - 22 a 35 anos, II - 36 a 50 anos, e III - 51 anos em diante), a fim de que cada célula seja alocada.

Com isso, nos propomos a utilizar as amostras parciais das duas fases do projeto PORCUFORT com informantes estratificados de acordo com as variáveis sociais apresentadas nos *corpora*, e uma adaptação nas faixas etárias das amostras, conforme podemos observar no quadro 1:

**Quadro 8:** Distribuição dos informantes nas amostras do PORCUFORT Fase I (década de 1990) e Fase II (década de 2020)

	Fase I (1990)						Fase II (2020)					
	Sexo						Sexo					
	<i>Masculino</i>			<i>Feminino</i>			<i>Masculino</i>			<i>Feminino</i>		
Tipo de Registro → Faixa etária ↓	D2	DID	EF	D2	DID	EF	D2	DID	EF	D2	DID	EF
I (22 a 35 anos)	3	3	3	3	3	3	3	3	3	3	3	3
II (36 a 50 anos)	3	3	3	3	2	2	3	3	3	3	3	3
III (51 anos acima)	3	3	2	3	3	1	3	3	3	3	3	3
<b>Total</b>	<b>9</b>	<b>9</b>	<b>8</b>	<b>9</b>	<b>9</b>	<b>6</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>	<b>9</b>
<b>Total geral</b>	<b>50</b>						<b>54</b>					
	<b>104</b>											

Fonte: elaboração própria

Conforme observado no Quadro 1, nossa amostra conta com 104 informantes

distribuídos em dois sexos diferentes, três faixas etárias, três tipos de registro e duas décadas de gravação, o que totaliza um montante de 104 perfis sociolinguísticos para analisar. Vale lembrar, no entanto, que a falta de simetria causada pela célula EF, ocupada por informante feminino de faixa etária III (apenas 1 informante) não compromete a análise estatística da amostra, pois o programa estatístico *R Studio* permite observar os dados tanto em sua totalidade quanto considerar o informante em sua individualidade (Levshina, 2015). Dito isto, o quadro abaixo apresenta as características sociais de cada um dos informantes selecionados para compor nossa amostra do PORCUFORT Fase I (década de 1990):

**Quadro 9 - Detalhamento dos informantes do PORCUFORT Fase I (Amostra 1990) selecionados para esta pesquisa**

Tipo/Número do inquérito	Sexo/Gênero	Faixa etária	Profissão	Bairro
D2 02	Mulher	Faixa I (23 anos)	Arquiteta	Aldeota
D2 02	Mulher	Faixa I (24 anos)	Professora	D. Torres
D2 07	Mulher	Faixa I (26 anos)	Psicóloga	Cajazeiras
D2 07	Mulher	Faixa II (47 anos)	Pedagoga	Centro
D2 11	Homem	Faixa I (33 anos)	Comerciante/veterinário	Aldeota
D2 11	Homem	Faixa III (70 anos)	Advogado	Aldeota
D2 16	Mulher	Faixa II (37 anos)	Professora de Ensino Básico	Centro
D2 26	Mulher	Faixa II (36 anos)	Professora	Jardim Iracema
D2 28	Homem	Faixa I (35 anos)	Professor Universitário	Benfica
D2 30	Homem	Faixa II (40 anos)	Técnico Operacional	Papicu
D2 33	Mulher	Faixa III (69 anos)	Aposentada/Professora de Ensino Básico	São João do Tauape
D2 34	Homem	Faixa II (41 anos)	Professor de Ensino Básico	São Gerardo
D2 39	Mulher	Faixa III (66 anos)	Professora	Jacarecanga
D2 39	Mulher	Faixa III (62 anos)	Funcionária Pública	Monte Castelo
D2 45	Homem	Faixa II (50 anos)	Engenheiro Químico	Meireles
D2 47	Homem	Faixa I (27 anos)	Professor	Aldeota
D2 47	Homem	Faixa III (64 anos)	Advogado	Fátima
D2 48	Homem	Faixa III (71 anos)	Aposentado/Professor	Benfica
D2 80	Mulher	Faixa III (68 anos)	Bacharelado em Direito	Aldeota
DID 08	Homem	Faixa III (40 anos)	Advogado, Promotor, Professor de Ensino Superior aposentado	Fátima
DID 09	Mulher	Faixa II (38 anos)	Professora do Ensino Básico	Álvaro Weyne
DID 10	Homem	Faixa II (45 anos)	Professor Universitário	Fátima
DID 12	Mulher	Faixa II (41 anos)	Professora do Ensino Básico	Aldeota
DID 13	Mulher	Faixa III (58 anos)	Aposentada/Funcionária pública federal	Parquelândia
DID 15	Homem	Faixa III (58 anos)	Administrador	Dionísio Torres
DID 20	Mulher	Faixa I (31 anos)	Professora do Ensino Básico	Carlito Pamplona
DID 21	Homem	Faixa I (25 anos)	Terapeuta Ocupacional	Centro
DID 22	Homem	Faixa III (67 anos)	Professor de Ensino Básico e Diretor aposentado	Parquelândia

DID 23	Mulher	Faixa III (67 anos)	Professora de Ensino Básico aposentada	Messejana
DID 24	Mulher	Faixa III (61 anos)	Professora Universitária	Aldeota
DID 31	Mulher	Faixa I (25 anos)	Dentista	
DID 32	Mulher	Faixa II (51 anos)	Professora de Ensino Básico aposentada	Quintino Cunha
DID 44	Homem	Faixa III (70 anos)	Professor de Ensino Básico	Dionísio Torres
DID 46	Homem	Faixa II (48 anos)	Engenheiro Químico e Coordenador Administrativo	Varjota
DID 50	Homem	Faixa I (29 anos)	Médico	Papicu
DID 106	Mulher	Faixa I (25 anos)	Comerciante	Parquelândia
DID 150	Homem	Faixa I (34 anos)	Advogado	São José do Tauape
EF 17	Homem	Faixa I (32 anos)	Professor de Ensino Básico	Cidade dos Funcionários
EF 18	Mulher	Faixa II (42 anos)	Professora Universitária	Meireles
EF 19	Homem	Faixa II (41 anos)	Médico Psiquiatra	Aldeota
EF 25	Mulher	Faixa I (32 anos)	Professora Universitária	Aldeota
EF 35	Mulher	Faixa I (30 anos)	Professora de Ensino Básico	Parque São José
EF 36	Homem	Faixa I (24 anos)	Professor de Ensino Básico	Barra do Ceará
EF 52	Homem	Faixa II (39 anos)	Professor de Ensino Básico	Piedade
EF 54	Homem	Faixa I (28 anos)	Professor Universitário/Engenheiro Mecânico	Fátima
EF 55	Mulher	Faixa II (45 anos)	Professora Universitária/Estatística	Papicu
EF 56	Mulher	Faixa II (54 anos)	Professora Universitária	Aldeota
EF 114	Homem	Faixa III (62 anos)	Professor Universitário	Parquelândia
EF 138	Mulher	Faixa I (31 anos)	Professora de Ensino Básico	Monte Castelo
EF 152	Homem	Faixa II (54 anos)	Professor / Engenheiro Civil	Aldeota
EF 214	Homem	Faixa III (56 anos)	Professor Universitário	Messejana
<b>Total de informantes</b>		<b>50 informantes</b>		

Fonte: elaboração própria.

Em relação às informações presentes no quadro, a primeira coluna corresponde aos tipos e números dos inquéritos do PORCUFORT que utilizamos para compor nossa amostra. Como é possível ver, trabalhamos com os três tipos de inquéritos, sendo 16 inquéritos do tipo D2, 18 inquéritos do tipo DID e outros 15 inquéritos do tipo EF, distribuídos em função do sexo/gênero e da faixa etária.

Conforme informações presentes na segunda coluna do quadro, nós trabalhamos, na primeira fase do PORCUFORT, com 26 informantes do sexo masculino e 24 informantes do sexo feminino. Cabe ressaltar neste momento que, na estratificação dos participantes em função da variável *sexo*, foram consideradas apenas questões que correspondem as diferenças biológicas que caracterizam as pessoas, assim, como já mencionado, os sujeitos foram estratificados em homens e mulheres. Ou seja, não tratamos de trazer considerações acerca da

construção social das questões que envolvem gênero, questões que estão cada vez mais presentes em trabalhos variacionistas, especialmente aqueles de terceira onda sociolinguística (Eckert, 2003, 2005, 2012).

Na coluna quatro, como observamos no quadro 9, foi destinada a descrição das profissões ou ocupações que cada um dos sujeitos participantes ocupava na época em que foram realizadas as entrevistas para o PORCUFORT. Apesar de não incluirmos a profissão como uma variável controlada neste estudo, consideramos um aspecto muito importante que caracteriza a identidade social dos informantes cultos de Fortaleza. Ou seja, a partir da análise da profissão dos informantes que compõem nossa amostra, é possível perceber que “grande maioria dos informantes selecionados exerciam profissões que gozam de notório prestígio social na sociedade brasileira, caso de [professores], advogados, terapeuta ocupacional, arquitetos, médicos dentre outros” (Pereira, 2021, p. 131).

Por fim, na última coluna, apresentamos os bairros em que cada informante de nossa amostra residia na época em que foram entrevistados. As informações referentes a localidade de moradia dos informantes são importantes para demonstrar que utilizamos amostra que reflete o falar dos informantes distribuídos por todas as áreas de Fortaleza. Conforme destaca o quadro, nossos informantes são provenientes de bairros como Aldeota, Dionísio Torres, Cajazeiras, Centro, Jardim Iracema, Benfica, Papicu, São João do Tauape, São Gerardo, Monte Castelo, Jacarecanga, Fátima, Álvaro Weyne, Parquelândia, Carlito Pamplona, Messejana, Quintino Cunha, Varjota, Cidades dos Funcionários, Parque São José, enfim, uma distribuição que representa a totalidade da cidade de Fortaleza.

De igual maneira, procuramos, em nossos dados representativa da década de 2020, ou seja, do PORCUFORT Fase II, selecionar perfis sociolinguísticos variados, de forma que, em nossa amostra, fosse contemplado o maior número de bairros possível para termos uma representação do falar de Fortaleza em sua totalidade. As informações socioeconômicas dos informantes que compõem nossa amostra podem ser vistas detalhadamente no quadro 10:

**Quadro 10** - Detalhamento dos informantes do PORCUFORT Fase II (Amostra 2020) selecionados para esta pesquisa

Tipo/Número do inquérito	Sexo/Gênero	Faixa etária	Profissão	Bairro
D2 08	Homem	Faixa I (27 anos)	Psicólogo	Granja Lisboa
D2 85	Homem	Faixa I (26 anos)	Administrador/ desempregado	Parangaba
D2 83	Homem	Faixa I (34 anos)	Professor de Ensino Básico	Monte Castelo
D2 02	Mulher	Faixa I (26 anos)	Química	Barra do Ceará

D2	02	Mulher	Faixa I (26 anos)	Professora	Pirambu
D2	08	Mulher	Faixa I (30 anos)	Psicólogo	Granja Lisboa
D2	83	Homem	Faixa II (38 anos)	Professor de Ensino Básico	São João do Tauape
D2	76	Homem	Faixa II (38 anos)	Servidor Público	Guararapes
D2	76	Homem	Faixa II (39 anos)	Oficial de Justiça	Centro
D2	55	Mulher	Faixa II (46 anos)	Professora	Vila Ellery
D2	55	Mulher	Faixa II (49 anos)	Professora	Farias Brito
D2	56	Mulher	Faixa II (50 anos)	Professora de Ensino Básico	Salinas
D2	86	Mulher	Faixa III (58 anos)	Professora de Ensino Básico	Parquelândia
D2	62	Homem	Faixa III (60 anos)	Perito Criminalista	Farias Brito
D2	62	Homem	Faixa III (69 anos)	Aposentado	Farias Brito
D2	56	Homem	Faixa III (62 anos)	Servidor Público e Radialista	Salinas
DID	03	Homem	Faixa I (26 anos)	Economista	Mondubim
DID	14	Homem	Faixa I (29 anos)	Professor de Ensino Básico	Parangaba
DID	19	Homem	Faixa I (34 anos)	Administrador	Paupina
DID	04	Mulher	Faixa I (31 anos)	Contadora	Jardim das Oliveiras
DID	07	Mulher	Faixa I (24 anos)	Administradora	Granja Lisboa
DID	12	Mulher	Faixa I (27 anos)	Costureira	Antônio Bezerra
DID	16	Homem	Faixa II (38 anos)	Educador Físico	Jardim Cearense
DID	37	Homem	Faixa II (36 anos)	Personal Trainer	Meiros
DID	48	Homem	Faixa II (36 anos)	Professor de Ensino Básico	Antônio Bezerra
DID	25	Mulher	Faixa II (52 anos)	Contadora	Parangaba
DID	28	Mulher	Faixa II (44 anos)	Secretária Escolar	Parangaba
DID	32	Mulher	Faixa II (53 anos)	Professora de Ensino Básico	Henrique Jorge
DID	13	Homem	Faixa III (59 anos)	Professor de Ensino Básico	Monte Castelo
DID	60	Homem	Faixa III (63 anos)	Engenheiro Mecânico	Parreão
DID	50	Homem	Faixa III (56 anos)	Consultor de Empresas	Guararapes
DID	24	Mulher	Faixa III (63 anos)	Professora de Ensino Infantil	Jardim América
DID	29	Mulher	Faixa III (62 anos)	Supervisora Escolar	Montese
DID	34	Mulher	Faixa III (61 anos)	Servidora Pública	Quintino Cunha
EF	06	Homem	Faixa I (30 anos)	Professor Universitário	Meiros
EF	20	Homem	Faixa I (28 anos)	Professor Universitário	São João do Tauape
EF	30	Homem	Faixa I (33 anos)	Professor de Ensino Básico	Demócrito Rocha
EF	39	Homem	Faixa I (33 anos)	Professora de Ensino Superior	Jacarecanga
EF	46	Mulher	Faixa I (32 anos)	Professora de Ensino Básico	Benfica
EF	18	Mulher	Faixa I (33 anos)	Professora Universitária e Advogada	Maraponga
EF	44	Mulher	Faixa I (32 anos)	Professora Universitária	Passaré
EF	21	Homem	Faixa II (39 anos)	Professor Universitário	Alto da Balança
EF	42	Homem	Faixa II (44 anos)	Professor e psicólogo	Papicu
EF	66	Homem	Faixa II (49 anos)	Professor e Procurador	Cambeba
EF	17	Mulher	Faixa II (46 anos)	Professora Universitária	Jacarecanga
EF	39	Mulher	Faixa II (36 anos)	Professora Universitária	Curió
EF	79	Mulher	Faixa II (36 anos)	Professora de Ensino Básico	Conjunto Ceará
EF	69	Homem	Faixa III (58 anos)	Escrivão de Polícia	Presidente Kennedy

				aposentado(a)	
EF	81	Homem	Faixa III (58 anos)	Professor de Ensino Básico	Praia de Iracema
EF	87	Homem	Faixa III (63 anos)	Engenheiro Civil	Fátima
EF	53	Mulher	Faixa III (61 anos)	Professora de Ensino Superior	Cocó
EF	89	Mulher	Faixa III (60 anos)	Professora do ensino básico	Benfica
EF	63	Mulher	Faixa III (61 anos)	Professora de Ensino Básico	Luciano Cavalcante
<b>Total de informantes</b>			<b>54 informantes</b>		

Fonte: elaboração própria

Tal qual o quadro representativo de nossa amostra do PORCUFORT Fase I, o quadro 10 demonstra o detalhamento de cada perfil social da segunda fase do PORCUFORT que utilizamos como base de dados nesta pesquisa. Assim, trabalhamos com os três tipos de inquiridos, sendo 18 do tipo DID, 18 do tipo D2 e 18 inquiridos do tipo EF, conforme é possível observar no quadro.

Além do mais, do PORCUFORT Fase II, trabalhamos com um total de 54 informantes, tendo uma amostra composta por 27 homens e 27 mulheres, distribuídos uniformemente entre os três tipos de inquirido. Quanto à faixa etária, nossa amostra dos anos 2020 conta com 18 sujeitos da faixa I (22-35 anos); 18 informantes da faixa II (36-50) e 18 da faixa III (a partir de 51 anos).

Ademais, destacamos que as informações referentes ao detalhamento de cada informante foram obtidas através das fichas pessoais preenchidas pelo informante após a gravação do inquirido e que, para preservar a identidade dos participantes desta pesquisa, não divulgaremos os nomes ou as fichas individuais. No entanto, todas as fichas estão sob posse da professora Dra. Aluiza Alves de Araújo, que gentilmente as disponibilizou para que esta pesquisa fosse realizada.

#### 4.4 Variáveis

As pesquisas quantitativas em geral e, especialmente, as que tomam por base a Sociolinguística Variacionista, por conta da ferramenta estatística utilizada para o tratamento dos dados, usam o termo variável dependente para expressar o lugar na língua onde ocorre a variação. Para Mollica e Braga (2012), o uso do termo ‘dependente’ se dá pelo fato das formas variantes não ocorrerem aleatoriamente, mas sim sistematicamente, por influência de outras variáveis (de ordem linguística ou extralinguística) que condicionam a escolha de determinada

variante. No entanto, em análises quantitativas que tomam como base o programa *R Studio*, o termo variável dependente é comumente chamado de variável resposta, motivo pelo qual adotaremos esse termo.

Este estudo revela o comportamento de duas faces distintas do mesmo fenômeno variável: a realização variável da primeira pessoa do plural. Contamos então, para efeito de análise e descrição do fenômeno linguístico nesta tese, com duas variáveis resposta (ou variáveis dependentes): a variação entre os pronomes *nós* e *a gente*, e a variação na concordância verbal com esses pronomes. Tendo isso em vista, apresentamos inicialmente a descrição das variáveis previsoras (ou preditoras) testadas na análise da alternância pronominal de primeira pessoa do plural, sendo esta uma variável resposta é binária, isto é, comporta duas variantes linguísticas, uma tida como padrão e outra inovadora, a saber, *nós* e *a gente*, assim como fizeram outros autores (Menon, 1995; Omena, 1996; Lopes, 2003; Zilles, 2005; Araújo, 2016; Souza, 2020). Em um segundo momento, tratamos de descrever os preditores elencados para analisar os padrões de concordância verbal com os pronomes *nós* e *a gente*, sendo esta uma variável resposta constituída de quatro variantes: *nós* com *-mos* (concordância padrão), *nós* sem *-mos* (concordância não padrão), *a gente* sem *-mos* (padrão emergente) e *a gente* com *-mos* (concordância não padrão). Para cada uma das variáveis resposta, observamos grupo de variáveis previsoras diferentes.

#### ***4.4.1 Variáveis previsoras: o fenômeno da alternância pronominal***

Com base na literatura sobre o fenômeno, como Omena (1979, 1986, 1996); Albán e Freitas (1991); Lopes (1993, 2003); Duarte (1995); Menon (1995, 2003); Naro (1999); Tamanine (2002, 2010); Seara (2000); Faraco e Moura (2000); Maya e Silva (2000) Brusolin (2009); Zilles (2003, 2005); Silva (2004); Menon, Lambach e Landarin (2003); Borges (2004); Fernandes (2004); Zilles (2005); Cunha (1993); Cavalcante (1997, 2001, 2002); Callou e Lopes (2003); Machado (1995); Ferreira (2002); Vianna (2006) e Tavares (2014), para citar alguns, decidimos testar algumas variáveis previsoras (também nomeadas de variáveis preditoras e, em estudos mais clássicos da sociolinguística variacionista, grupo de fatores), ou seja, as variáveis de ordem linguística e as de ordem extralinguística (variáveis sociais) que acreditamos atuar significativamente sobre a realização do fenômeno de alternância pronominal entre *nós* e *a gente*. A escolha das variáveis testadas neste estudo tomou como base a literatura consultada de estudos sobre o fenômeno, bem como embasam as hipóteses sobre a realização dos pronomes

*nós* e *a gente* nas amostras analisadas. A seguir, detalhamos cada um dos grupos de fatores elencados para a análise do fenômeno variável e ilustramos, sempre que possível, com ocorrências extraídas de nossa amostra de fala culta.

#### 4.4.1.1 Variáveis previsoras de ordem linguísticas

As variáveis linguísticas analisadas neste estudo sobre a alternância entre os pronomes de primeira pessoa do plural foram: *grau de referencialidade do pronome*, *tempo verbal*, *tipo de verbo*, *paralelismo formal* e *saliência fônica*, conforme detalhamos abaixo. As variáveis linguísticas correspondem a grupos de fatores de ordem estrutural, que atuam no interior do próprio sistema linguístico e condicionam o uso da determinada variante.

##### I) **Grau de referencialidade dos pronomes *nós/a gente***

O grau de referencialidade dos pronomes de primeira pessoa do plural foi considerado relevante em diversas pesquisas (Omena, 1996; Lopes, 2003; Araujo, 2016, entre outros). Para Silva (2004), o uso dos pronomes pode ser feito pelos falantes de uma comunidade de fala como estratégia para resignar o grau de referência em diversas situações, tendo inclusive a função de indeterminação do sujeito, deixando claro a forte interação dessa variável sobre o uso desses pronomes de primeira pessoa do plural.

Com base na literatura, acreditamos na ação direta desta variável sobre a alternância pronominal em nossa amostra de fala culta, nas duas sincronias investigadas. As ocorrências apresentadas abaixo, retiradas de nossa base de dados, ilustram os fatores/níveis/categorias que compõem esta variável.

a) *Sentido genérico* - quando a realização dos pronomes faz uma retomada circunscrita a um grupo (Lopes, 2003), ou quando ocorre indeterminação universal (*EU+ EU AMPLIADO; TODOS*):

(14) Inf.: e muitas vezes *a gente* num precisa NEM comprar um jornal porque sai a notícias toda no rádio numa veLOCidade TÃO GRANde (F2 - D2 – Inq. 11)  
(Inq. 11 – F2 – D2, homem, 70 anos)

(15) Inf.: E e eu acredito que é um trabalho muito BEM:: muito bem disfarçado muito bem planejado... eh ah eh:: *nós* que fazemos universiDAde... eh eu /tô

dizendo **NÓS** que fazemo/ universidade professores alunos e funcioNÁrios...  
(Inq. 28 – F1 – D2, homem, 35 anos)

b) *Sentido específico* - a realização dos pronomes retoma o referente (EU + VOCÊ; EU + ELES), especificado no contexto (Lopes, 2003):

(16) Inf.: ...na adolescência dela catorze anos QUINze anos ela já faLava isso que aqui não tinha nada pra ela...já eu e minha outra irmã a que é advogada eh **a gente** não pensava assim eu ir pra fora morar fora mas ainda irmã pesquisava MUIto sobre EuRO::pa (Inq. 16 – F2 – DID, homem, 38 anos)

(17) Inf.: você disse que não vai morar com a minha mãe... e eu... com certeza não vou morar com a sua... porque **nós** não temos problemas com nenhuma família... MAIS:: se é pra casar a gente tem que ter aonde morar (Inq. 04 – F2 – DID, mulher, 31 anos)

## II) Tempo verbal e formas nominais do verbo:

A variável *tempo verbal e formas nominais do verbo* tem se mostrado relevante para a análise da alternância pronominal de primeira pessoa do plural, visto que é corriqueiramente selecionada como estatisticamente significativa para a realização dos pronomes *nós* e *a gente*, conforme demonstram estudos sociolinguísticos sobre o tema (Omena, 1986, 1998; Lopes, 1993; Machado, 1995; Tavares, 2014; Araujo, 2016; Silva, 2022). Em vista disso, em nossos dados, decidimos testar a relevância dos seguintes tempos verbais/formais nominais sobre a realização do fenômeno variável em análise:

a) *Presente do indicativo*

(18) Inf.: Isso está na nossa compeTÊNcia de falante da língua... **nós sabemos** quais são os verbos que exigem tais e tais preposições... QUAIS os verbos que perMitem ser substituídos por OUtras preposições **nós fazemos** isso... né? (Inq. 25 – F1 – EF, mulher, 32 anos)

(19) Inf.: a pessoa bota aquilo ali e pesa e ela tenta entender peso que aquilo tem e o que ela poderia ter se ela mudasse comportamento então a técnica **a gente aplica** no paciente pra:: descobrir a motivação e fazer a pessoa se conscientizar também (Inq. 06 – F2 – EF, homem, 30 anos)

b) *Pretérito perfeito do indicativo*

- (20) Inf.: então... esse processo é chamado de/de so-lu-ção fracionada... Né? fracionada porQUÊ você usa mais de um MÉtodo... Né? O que foi que **nós fizemos?** tinha minha amostra de areia/... areia e SAL... Né? adicionei Á::gua... (Inq. 87 – F2 – EF, homem, 63 anos)
- (21) Inf.: POSto de saúde TUdo... tudo tudo tudo tudo atrativos natuRAIS tudo... então NEM Isso... **a gente conseguiu...** verba da UNIFOR Nada foi tudo pelo bolso da gente tudo foi **a gente** que **fez...** tudo... (Inq. 106 – F2 – EF, mulher, 25 anos)

c) *Pretérito imperfeito do indicativo*

- (22) Inf.: as Obras de Guerra Junqueiro... enfim... um exemplar de cada obra DEle para a nossa futura biblioTEca... uma coisa assim... todos **nós tínhamos** entusiASmo pela gloriosa constelação portuGUÊsa... recolhemo-nos para meditar frases ao Eça ao Nobre... (Inq. 156 – F1 – EF, mulher, 54 anos)
- (23) Inf.: eu exigia MUIto dos meus alunos exige MUIto porQUE... eu fui de uma fase de escola... quando que::... **a gente ensinava** o aluno a entrar e a sair numa sala de aula... não admitia mesmo não que o aluno s::a/... chegasse... depois que eu entrasse na sala de aula (Inq. 39 – F1 – D2, mulher, 66 anos)

d) *Futuro do indicativo*

- (24) Inf.: ...o CLIma... em que foi estipul::/ estipulado o... ReaLISmo... em Portugal... e... outro autor que **nós iremos ver...** que é::... o Antero de Quental... aí tem aí Antero de Quental na página cento e quarenta e três "o poeta fiLÓsofo" (Inq. 36 – F1 – EF, homem, 24 anos)
- (25) Inf.: aí do mesmo jeito vocês vão localiZAR... aonde é que esses conceitos FORam trabalhados lá...aí **a gente terá** dois casos de emPREsas sendo::... analisados a luz... das teorias que a gente tá estudando... (Inq. 46 – F2 – EF, homem, 44 anos)

Além do futuro simples, conforme exemplificados com as ocorrências (24) e (25), também atribuímos o status de *futuro* aos verbos formados pelo auxiliar IR + verbo principal infinitivo, estrutura que compõe o futuro perifrástico. Nesse sentido, em nossa análise,

não há distinção entre futuro perifrástico e futuro simples, devido às poucas ocorrências desse tempo verbal em nossa amostra, fato que nos fez amalgamá-lo já no momento da codificação. Com isso, as realizações como as observadas em (26) e (27) foram classificadas, em nossa análise, simplesmente como futuro.

(26) Inf.: temos os protozoários... que são os esporozoários... causador de uma doença chamada maleica... uma/e tem outros casos dela que **nós vamos ver** mais a frente... quem são as maiores doenças desses protozoários?... (Inq. 36 – F1 – EF, mulher, 61 anos)

(27) Inf.: mas (tem um pai que ainda não descobri) --até o final do semestre **a gente vai descobrir** é tudo e tem gente que viajou pra fazer concurso foi? pra trabalho foi isso que tava no grupo? essa disciplina transforma vidas. (Inq. 17 – F2 – EF, mulher, 46 anos)

e) *Presente do subjuntivo*

(28) Inf.: nenhum HOMem capaz de ser presidente da República... porque TODos são... politicamente... certo?... contaminados... certo?... pela política atual... POSSA SER que **nós tenhamos**... uma pessoa capaz...e (Inq. 17 – F1 – EF, homem, 33 anos)

(29) Inf.: É bixim... Os bixim... num pode escutar um não... e não pode ser chamado atenção que fica COM RAiva... **A gente que fique** com raiva e/ e/ e pronto... (Inq. 83 – F2 – D2, homem, 34 anos)

f) *Pretérito imperfeito do subjuntivo*

(30) Inf.: nós queremos muito comprar um carro... estamos nos organizando pra isso... mas se **nós ganhássemos** um de presente seria bem melhor né? que comprar... (Inq. 09 – F2 – DID, mulher, 24 anos)

(31) Inf.: então é um prejuízo que uma locomotiva financiava esse projeto no Brasil inteiro...só uma locomotiva...se **a gente vendesse** uma locomotiva dava pra financiar esse projeto no Brasil inteiro... (Inq. 45 – F1 – D2, homem, 50 anos)

g) *Futuro do subjuntivo*

(32) Inf.: a questão do contexto histórico isso influencia... então quando **nós nos detivermos** a um autor... é necessário que a gente observe em que Época... esse autor... eh:: produziu... sua obra... /tá?... (Inq. 35 – F1 – EF, mulher, 30 anos)

(33) Inf.: vão atuar dentro dessa instituição de forma negativa... éh::... de acordo com seus interEsSES... realmente... provocando malefícios de TOdos os... de todos os... de todos os jeitos que **a gente puder imaginar**... então éh::... não é DIzer que ela não foi tão má assim... (Inq. 48 – F2 – DID, homem, 36 anos)

#### h) Gerúndio

(34) Inf.: só que nós não tínhamos percebido que nós tínhamos parado na calçada em frente a faixa de pedestre... e **nós conversando** aqui quando nós olhamos... tinha uma fila de carro... que a/a/o pessoal tava esperando a gente passar... na faixa de pedestre... (Inq. 03 – F2 – DID, homem, 26 anos)

(35) Inf.: o J. diz que eu tenho né? raiva da:: mas num É Isso né? bem o caso... o que ele quer dizer ((falou rindo)) a história do J. todim começou... PORQUE **a gente assistindo** a novela no sítio o:: esse mesmo que a gente fez aqui todo mundo SINE... (Inq. 30 – F1 – D2, mulher, 31 anos)

#### i) Infinitivo

(36) Inf.: então o que é que nós vamos aprender aqui nesse tema pessoal... vamos aprender aqui... saber... **nós saber** aqui quem é o autor viu?... quem é o Coautor... e quem é o partícipe... o partícipe certo... então vamos aprender esses três figurantes... (Inq. 69 – F2 – EF, homem, 69 anos)

(37) Inf.: esse boletim... bem é:: uma coisa muito difícil **a gente compreender** dentro dos nossos parâmetros... não seria (a questão) desse EsTAdo ser democrático é possível o Estado ser democrático? (Inq. 18 – F1 – EF, mulher, 42 anos)

Pontuamos aqui, o fato de, em nossa amostra, não haver ocorrências dos pronomes de primeira pessoa do plural associadas a verbos no particípio, devido a própria natureza dos dados e ao modo de análise, visto que trabalhamos unicamente com realizações dos pronomes *nós* e *a gente* com função de sujeito, o que cria um ambiente pouco propício a esse tipo de ocorrência. Dessa forma, os verbos na forma nominal, em nossa amostra, apenas

se realizaram nas formas de gerúndio e de infinitivo.

j) *Futuro do pretérito*

(38) Inf.: dizer que no caso aqui do PIDGIN o crioulo *nós teríamos* o que nós chamamos de BAfileto... depois *nós teríamos* o conTInuum... certo? (Inq. 138 – F1 – EF, mulher, 31 anos)

(39) Inf.: NÓS tivemos essa vantagem né?... de:... bem:: mostrar que através do esFORço próprio né? do estudo tal *a gente chegaria* mais além né? (Inq. 12 – F1 – DID., mulher, 31 anos)

### III) Tipo de verbo

Tamanine (2002) realizou pesquisa em que comprovou a influência dessa variável linguística sobre a realização dos pronomes *nós* e *a gente*. Baseando-se em sua pesquisa, e fazendo as adaptações necessárias para esse estudo, resolvemos por controlar os seguintes fatores/níveis:

a) *Verbo epistêmico* (verbos que expressam atividade mental)

(40) Inf.: na Escola de Engenharia que trabalhava/ na IBM... *nós sabíamos* que nós tínhamos sido... a/ a:: aproVAdo/... todos os todos os cinco... inclusive a a::... é acho... o GRUpo que tinha tinha ti/ obtido o melhor resultado (Inq. 46 – F1 – DID, homem, 48 anos)

(41) Inf.: também tem uma/uns peixinhos umas sardinhas não é aquelas sardinhas que a *gente pensa* aqui que vem na latinha não...é uma sardinha maior e eles fazem no forno lá que só eles sabem fazer DESmancha na boca (Inq. 30 – F2 – DID, homem, 56 anos)

b) *Verbo dicendi* (verbos que expressam elocução/declaração)

(42) Inf.: nós vamos ver esse deSÂnimo e apatia esse tedium vitae tudo isso e a morte como libertação... de de que *nós falamos* eh baseado em José Carlos Seabra Pereira... o:: soneto anterior ...é em versos alexandrinos... (Inq. 03 – F1)

– EF, homem, 55 anos)

- (43) Inf.: DANÇAS e folGUEDOS então... o Ceará é riquíssimo... e isso TUDO faz PARte... do que **a gente CHAmA**... de... patrimônio cultural em seu sentido assim mais AMPlo... uma lenda...um ditado popuLAR... uma comida típica... (Inq. 24 – F1 – DID, mulher, 61 anos)

c) *Verbo de estado*

- (44) Inf.: eu consiDERo por exemplo a a conVERsa de que **nós somos** o único planeta habitado do uniVERso... iGUAL a um ratinho ( )... que começa a namorar uma ratinha... (Inq. 53 – F1 – EF, homem, 53 anos)
- (45) Inf.: ela cresceu (pancada) e./e... começou a trabalhar por lá e meu pai... meu avô comprou uma casa e aí **a gente ficou** lá:... porque a gente não tinha uma casa fixa a gente sempre morava ou de aluGUEL (Inq. 04 – F2 – DID, mulher, 31 anos)

d) *verbo de ação*

- (46) Inf.: então esse sistema que **NÓS implantamos** aqui...ele evita um tipo de acidente desse tipo né?...porque...o maquinista /tá se comunicando com as estações vizinha...se comunica com a Central(Inq. 45 – F1 – D2, homem, 50 anos)
- (47) Inf.: lá a A. fez/ foram logo fazer o mercantil... aí o café da manhã **a gente fazia** lá né... comprou TANta coisa... TANta coisa... que sobrou foi coisa... agora o almoço a gente ia almoçar fora... (Inq. 24 – F2 – DID, mulher, 63 anos)

e) *Verbo ter*

- (48) Inf.: **nós não tínhamos** o DESprazer... de ter essa INsegurança que **nós temos** hoje... momentos DELiciosos vamos dizer... a gente era feliz e NÃO sabia... HOje hoje vamos dizer... (Inq. 56 – F2 – D2, mulher, 63 anos)
- (49) Inf.: sistema em coordenadas esFÉricas... a gente vai... integrar duas vezes... daí **a gente tem** a equação da/ **a gente tem** a distribuição de temperatura... COM a distribuição de temperatura... aplicando Fourier... **a gente tem** equação da taxa de transferência de calor isso aí **a gente tem**... (Inq. 54 – F1 – EF, homem, 28 anos)

IV) **Paralelismo formal**

Vários autores, em seus estudos, consideraram a variável paralelismo formal como um preditor linguístico de grande relevância para os estudos relacionados à variação dos pronomes de primeira pessoa do plural. Essa variável designa a manutenção de determinada forma pronominal, representando o mesmo referente, ou a mudança do pronome mantendo a mesma referência formal nas estruturas linguísticas. Dessa forma, com esse grupo de fatores, buscamos verificar a manutenção ou a mudança, pelo falante, da forma pronominal em uso em uma série de realizações no discurso. Dentre as pesquisas que testaram essa variável linguística, podemos destacar: Lopes (1993), Machado (1995), Tamanine (2002), Borges (2004), Muniz (2008), Mendonça (2010), Nascimento (2013), Souza (2020) e Fernandes (2021). Tendo por base as pesquisas citadas, decidimos controlar o paralelismo formal em dados de nossas amostras. Abaixo, ilustramos cada fator/nível/categoria controlado nesta variável previsora com ocorrências extraídas do PORCUFORT:

*a) Oração Isolada*

- (50) Inf.: aquela coisa que BRAsileiro geralmente é o melhor né? **nós** somos os melhores... Éramos os melhores... na Fórmula UM... e agora recentemente no VÔlei né? (Inq. 20 – F1 – DID, mulher, 32 anos)
- (51) Inf.: na aula anterior... **a gente** já havia discutido... sobre a diferença entre esgoto secundário e esgoto primário né?... (Inq. 153 – F1 – EF, homem, 41 anos)

*b) Primeira referência*

- (52) Inf.: mas **a gente** acompanha... é um painel ainda muito modesto certo? a gente acompanha manuAL já existe painéis eletrônicos e coisa e tal (Inq. 42 – F1 – DID, homem, 42 anos)
- (53) Inf.: os candidatos cada qual cheio de projetismos... e **nós** num temos ainda... uma pessoa capaz de representar o povo... o que nós temos... são interesses (Inq. 17 – F1 – EF, homem, 53 anos)

*c) Segunda ou terceira realização da série com referência anterior feita pelo mesmo pronome (manutenção do paralelismo)*

- (54) Inf.: no estudo que **nós** vamos fazer daqui pra frente... eh:: no nosso caso **nós** vamos nos deter PRIoritariamente na Literatura tá? (Inq. 35 – F1 – EF, mulher, 30 anos)
- (55) Inf.: então **a gente** num tem assim:: o poDER de dizer o que é que o EsTAdo vai fazer... **a gente** pode DAR sugestões trabalhar (Inq. 24 – F1 – DID, mulher, 61 anos)

d) *Segunda ou terceira realização da série com referência diferente em relação a imediatamente anterior (quebra do paralelismo)*

(56) Inf.: **nós** não planejamos e aí ela aí apareceu a gravidez né? a seGUNda foi arriscada de noite... e a terceira... (ela) também vamos **a gente** que queria TRÊS filhos (Inq. 08 – F1 – DID, homem, 40 anos)

(57) Inf.: pode partir desse ponto agora dizer que os canhotos vivem menos isso é puro folclore /tá? simplesmente **a gente** diz que **nós** trabalhamos um lado só do cérebro o lado direito ou o lado esquerdo... (Inq. 21 – F1 – DID, homem, 25 anos)

## V) Saliência fônica

A variável saliência fônica se refere ao nível de diferenciação no material fônico entre as formas singular e plural do verbo associado aos pronomes *nós* e *a gente* (Lopes, 1993; Machado, 1995; Borges, 2004; Mendes, 2007; Rocha, 2009; Brustolin, 2009; Silva, 2013; Foeger, 2014; Santana, 2014; Souza, 2020; Lima, 2020; Nardelli, 2021; Silva, 2022). Essa variável previsora tem demonstrado que, quanto menor a diferença fônica entre a 3<sup>o</sup> pessoa do singular e a 1<sup>a</sup> pessoa do plural, maior a possibilidade da forma pronome *a gente* ser empregada. Ao contrário, quanto maior a diferença fônica, maior a probabilidade de uso do pronome *nós*. Estudos sobre a alternância pronominal têm apresentado diferentes abordagens para o tratamento dos dados em função da diferença fônica na relação singular/plural. Por exemplo, Guy (1981) divide a escala de saliência fônica em quatro níveis, que vão do menor grau de diferença fônica até o grau mais acentuado. Por outro lado, Nicolau (1995) propõe doze níveis de saliência em uma escala de diferença fônica. Por sua vez, Rocha (2009), em sua dissertação de mestrado sobre a alternância entre os pronomes pessoais no português falado em Belo Horizonte, atribuiu seis níveis à variável saliência fônica.

Percebemos as distintas abordagens dadas a essa variável previsora, que tem se mostrado relevante para a explicação do fenômeno variável em questão, com diferentes escalas que medem o grau de diferença entre os verbos no singular e no plural. Para o tratamos de nossos dados de fala culta fortalezense, no entanto, replicamos a escala de saliência fônica adotada por Borges (2004) e Francischini (2011), pois estes apresentam uma escala que, além de dar conta da classificação de todos os verbos, têm o enfoque na diferenciação fônica entre as formas singular/plural, assim como pretendemos tratar nossos dados à luz desta variável.

Sendo assim, consideramos os seguintes níveis de saliência, que parte de verbos sem diferenciação fônica entre as formas até os que apresentam grande diferenciação fônica, sendo uma escala constituída por 7 níveis:

a) *Nível 1 (cantando) - Mesma forma para 3ª pessoa do singular e 1ª pessoa do plural*

- (58) Inf.: ou quando ele tá assim afetado a gente procura ficar tanto é que tava eu o M.H conversando com ele **a gente conversando** DAR atenção a ele... pra ele se sentir como pessoa pra se sentir valorizado (Inq. 13 – F2 – DID, homem, 59 anos)
- (59) Inf.: não tínhamos percebido que nós tínhamos parado na calçada em frente a faixa de pedestre... e **nós conversando** aqui quando nós olhamos... tinha uma fila de carro (Inq. 03 – F2 – DID, homem, 26 anos)

b) *Nível 2 (cantar/cantarmos) - Infinitivo com acréscimo da desinência de plural -mos*

- (60) Inf.: pois é... claro... mas a... o negócio é **a gente saber** o que deve falar e como falar:: e... a QUEM deve falar... num é? (Inq. 39 – F1 – D2, mulher, 66 anos)
- (61) Inf.: tendo em vista **nós sermos** um dos grandes produtores de petróleo...o Estado do Rio Grande do Norte talv/ também é um GRANde produtor de petróleo aqui vizinho... (Inq. 45 – F1 – D2, homem, 52 anos)

c) *Nível 3 (falava/falávamos) - Conservação da sílaba tônica e acréscimo da desinência -mos*

- (62) Inf.: tu se lembra quando **a gente morava** ali atrás que agora hoje é o Frei Lauro... naquela ruazinha que/ que é atualmente é:: Rubens Sampaio e antigamente era a Capitu dos Santos... **a gente morava** ali. (Inq. 62 – F2 – D2, homem, 60 anos)
- (63) Inf.: **nós morávamos** ali ali na Gomes de Matos HOje... RAramente vai encontrar um domicílio residencial você vai encontrar muitas casas comerciais... (Inq. 56 – F2 – D2, mulher, 50 anos)

d) *Nível 4 (fala/falamos) - Deslocamento do acento tônico e acréscimo da desinência -mos*

- (64) Inf.: muito BEM... estudo... comPARAdo das LÍNGuas românicas... quando **a gente fala** em filologia roMÂNica... a PALAVRA... MÉTODO HISTÓRICO-comparaTivo... deve estar presente (Inq. 214 – F1 – EF, homem, 60 anos)
- (65) Inf.: a cidade não é violenta eu tô falando que há um exaGEro... principalmente pela aquela parte política que **nós falamos**... falamos no início de tenTAR... éh::... de qualquer forma... ((barulho de buzina de carro)) fazer com que a gente pense que a Única solução pra violência seja o armamento

(Inq. 37 – F2 – DID, homem, 36 anos)

e) *Nível 5 (tem/temos) - Monossílabos tônicos ou oxítonos que se tornam paroxítonos no plural*

(66) Inf.: pro meu irmão que foi traumática mas::... mas:: **a gente tem** mais lembranças positivas quando a gente morava na minha vó (Inq. 04 – F2 – DID, mulher, 31 anos)

(67) Inf.: estado deve ser apenas para garantir essas liberdades e não ir além disso e aí nessa perspectiva **nós temos** a liberdade propriedade como um direito absoluto (Inq. 39 – F2 – EF, homem, 36 anos)

f) *Nível 6 (cantou/cantamos) - Redução dos ditongos finais em vogais e acréscimo de -mos*

(68) Inf.: :: e a gente sabia porque eles faLaram:: isso pra gente... quando **a gente falou** que eles não iam fazer a prova porque eles faltaram porque foram pra um casamento (Inq. 79 – F2 – EF, mulher, 36 anos)

(69) Inf.: foi criada a Guarda Vermelha então como **nós falamos**... a Guarda Vermelha... era:: praticaMENTe... ela mantinha uma identidade de um sistema (Inq. 17 – F1 – EF, homem, 33 anos)

g) *Nível 7 (é/somos) - Diferenciação fônica muito acentuada*

(70) Inf.: tinha um tempo disponível né?... ocioso e... num ia deixar de pegar porque... eh **a gente é** professor de Educação... oh... técnico de E/ de basquetebol somos educador né? (Inq. 34 – F1 – D2, homem, 25 anos)

(71) Inf.: eu sempre fui muito ativa... mas... mas não é isso... o nosso corpo mesmo... **nós somos** uma máquina... nosso coração é uma máquina... que funciona durante algum... tantos anos (Inq. 08 – F2 – D2, mulher, 30 anos)

#### 4.4.1.2 Variáveis previsoras de ordem extralinguística

Estas variáveis levam em conta os aspectos socioculturais e estilísticos que envolvem o indivíduo. Assim, controlamos, em nossos dados de fala culta de Fortaleza, os preditores *sexo/gênero, faixa-etária, tipo de registro e década de gravação*, conforme detalhado abaixo:

## I) Sexo/ gênero

A variável *sexo/gênero* torna-se relevante nas pesquisas realizadas sobre a variação pronominal *nós* e *a gente*, na medida em que as mulheres tendem a utilizar mais a formação gramaticalmente correta, enquanto os homens favorecem a utilização da forma inovadora, como vemos em Omena (1979,1986, 1996), Albán e Freitas (1991, 1991a, 1991b), Almeida (1992), Lopes (1993, 1998), Duarte (1995), Menon (1994, 1995, 2003), Tamanine (2002, 2010), entre outros. Nosso objetivo é analisar se existe diferença estatisticamente significativa entre homens e mulheres no que se refere aos pronomes de primeira pessoa do plural.

Para Labov (2003), homens e mulheres apresentam comportamento sociolinguístico distinto, no sentido de as mulheres apresentarem uma sensibilidade maior que a apresentada pelo homem no que diz respeito às correções sociais e que, por isso, tendem a optar pelas variantes linguísticas consideradas socialmente prestigiadas.

Labov (1972), em seu estudo nas lojas de Nova Iorque, conclui que as mulheres usam com mais frequência a variante de prestígio e ainda respondem de forma mais radical a teste de reações subjetivas, sendo mais tendenciosas que os homens a estigmatizar a variante não-padrão. Sendo assim a atuação global da mulher se encaixa perfeitamente no princípio sociolinguístico postulado por Labov (2003) de que aqueles que usam mais as formas não padrão no seu discurso casual serão mais sensíveis a essas formas no discurso dos outros.

No que diz respeito ao uso de *nós* e *a gente*, na fala popular, Omena (1986) mostra que há uma tendência das mulheres em empregar o pronome *nós*, enquanto os homens tendem ao uso de *a gente*. Já na norma culta, Lopes (1993) conclui que as mulheres favorecem mais a variante inovadora, enquanto os homens favorecem mais o pronome *nós*. Dessa forma, controlaremos essa variável previsora a partir de dois fatores:

- a) *Masculino*
- b) *Feminino*

## II) Faixa-etária

A variável *faixa etária* é de fundamental importância para observar o comportamento linguístico dos falantes e atestar, entre outras questões, se o fenômeno aqui estudado se encontra em variação estável ou se demonstram indícios de uma mudança em curso no sentido da forma inovadora, *a gente*, estar substituindo o pronome *nós*. É importante levar

em consideração que cada faixa etária possui um comportamento linguístico diferenciado, já que, no geral, os mais jovens buscam ser mais inovadores e isso é expresso em sua fala. O mesmo ocorre com os falantes mais velhos que, no geral, tendem a usar mais a forma padrão, o que é explicado por exigências sociais, com o ingresso no mercado de trabalho, por exemplo.

Estudos, como os de Omena (1986) e Lopes (1993), apontam para uma substituição do pronome *nós* por *a gente*, em todas as faixas etárias, porém isso se dá de forma bem mais expressiva entre os mais jovens. Dessa forma, neste trabalho, buscamos fazer uma análise comparativa das diferentes faixas etárias, assim como fizeram Omena (2003), Lopes (2003) e Nascimento (2013), no que diz respeito à utilização da forma *nós* ou *a gente*, em tempo real de curta duração, pois o banco de dados analisado nos oferece a possibilidade metodológica para este tipo de pesquisa. Assim, testamos as seguintes faixas etárias:

- a) 22-35 anos
- b) 36-50 anos
- c) 51 anos em diante

### III) Tipo de Registro

Para Labov (2008), os falantes de uma língua demonstram uma tendência a apresentar comportamento linguístico distintos a depender do grau de atenção dado a fala no momento da elocução. Essa atenção por parte do falante pode ser medida pelo grau de formalidade no ambiente de comunicação, ou seja, é possível observar a influência do nível de formalidade por meio do controle do *tipo de inquérito*.

Assim, o controle da variável *tipo de inquérito*, embora raro em pesquisas sociolinguísticas, se faz importante nesta pesquisa pois trabalhamos com um pronome tipo como padrão e outro por vezes caracterizado como não padrão. Assim, tendo em vista o grau de formalidade que representa os inquéritos do tipo DID, D2 e EF, conforme já discutido, determinada variante pode ser favorecida ou, em outros casos, inibida em função do tipo de registro. Por isso, decidimos testar, nesta variável preditora, os seguintes níveis:

- a) *Diálogo entre Informante e Documentador – DID*
- b) *Diálogos entre dois Informantes – D2*
- c) *Elocuções Formais – EF.*

#### IV) Década de gravação

Por se tratar de uma pesquisa em tempo real, que consideram duas sincronias distintas, a década de gravação se configura como uma variável previsoras imprescindível neste estudo, como já demonstraram algumas pesquisas sobre a variação pronominal de primeira pessoa do plural (Omena, 2003; Lopes, 2003; Nascimento, 2013). As análises foram realizadas individualmente para cada década. No entanto, na codificação dos nossos dados, incluímos esta variável para que fosse possível observar a tendência geral no uso de *nós* e *a gente* e a tendência em cada sincronia analisada. Com isso, controlamos nesta pesquisa as seguintes décadas:

- a) *Anos 1990*
- b) *Anos 2020*

#### 4.4.2 Variáveis previsoras: o fenômeno da concordância verbal

A concordância verbal é a segunda face que se manifesta no fenômeno variável da primeira pessoa do plural, estando diretamente relacionada à alternância entre os pronomes *nós* e *a gente* (Scherre, Yacovenco & Naro, 2018). Partimos da hipótese que o fenômeno de concordância verbal de primeira pessoa do plural está sob efeito de variáveis de ordem linguística e social, que determinam os contextos mais propícios a determinado padrão de concordância com o verbo. Dentre os diversos condicionamentos, decidimos testar as mesmas variáveis de natureza social apresentadas para verificar os padrões de *alternância pronominal* (ver 5.4.1.2), acreditando ser fundamental para entender o processo de variação e mudança linguística entre as formas de concordância verbal de primeira pessoa. Além das variáveis previsoras de cunha social, decidimos testar as variáveis linguísticas *tempo e tipo de paradigma verbal do modo indicativo, estrutura do verbo, conjugação verbal e tonicidade verbal*. As variáveis sociais próprias da estratificação de nossa base de dados, com a adaptação feito na variável *faixa etária*, também foram adotadas como parâmetro de análise para concordância com os pronomes *nós* e *a gente*.

#### I) Tempo e paradigma verbal do modo indicativo

Naro, Görski e Fernandes (1999), ao analisarem o *tempo verbal* e a *saliência fônica*

dos verbos, apontam para uma tendência relacionada ao uso da desinência de plural *-mos*. Segundo os autores, a partícula *-mos* tem se especializado como morfema que marca o tempo presente, enquanto a ausência dessa desinência (morfema zero) tem se associado ao tempo pretérito. Essa tendência é evidenciada quando identificamos que alguns “verbos regulares do português brasileiro na 1PP no presente e no pretérito perfeito apresentam a mesma forma” (Benfica, 2016, p.78.). Em nossa base de dados, verbos com essas características são abundantes, conforme demonstram os exemplos:

(72) Inf.: os mais diversos animais que habitam variados ambientes::... certo?... com vida LIVre... e quando *nós falamos* em ambiente esses animais têm vida livre (Inq. 63 - F2 - EF, homem, 61 anos)

(73) Inf.: não tô falando que a cidade não é violenta eu tô falando que há um exaGero... principalmente pela aquela parte política que *nós falamos*... falamos no início da aula... tenTAR... éh::... de qualquer forma..... (Inq. 37 - F2 - DID, homem, 36 anos)

Conforme podemos observar nos excertos (72) e (73), o verbo “falar” na primeira pessoa do plural possui o mesmo paradigma, tendo em vista que não há distinção formal entre os tempos passado e futuro. No exemplo (72), o contexto é o responsável por recuperar a interpretação de que o verbo está no presente do indicativo, diferentemente do exemplo (73), em que se atribui o pretérito perfeito do indicativo ao verbo. Essa diferença na relação presente/passado do verbo “falar” é recuperada por outros elementos do texto. Esses são exemplos de ambiguidade temporal causada por verbos com o mesmo paradigma.

Scherre, Yacovenco e Naro (2018), inspirados no estudo de Naro, Görski e Fernandes (1999), testaram a variável *tempo e tipo de paradigma verbal do modo indicativo*, que coloca como foco a análise da *saliência fônica* e do *tempo verbal* em conjunto, tendo em vista a forte correlação entre esses dois preditores. Uma das hipóteses que justificam a análise dessa variável é a que alguns verbos no presente do indicativo, assim como no pretérito perfeito do indicativo podem gerar ambiguidade temporal, pois apresentam formas iguais para presente/passado, a exemplo do verbo cantar, que não há distinção entre os tempos em “nós cantamos”, por exemplo, sendo necessário recorrer a advérbios ou ao contexto para inferir o seu tempo verbal. Nesse sentido, Scherre, Yacovenco e Naro (2018) acreditam que a forma inovadora *a gente*, ao se implementar na língua, pode desfazer a ambiguidade potencial existente entre esses tempos verbos. Com a inserção do pronome *a gente*, criou-se um padrão emergente de concordância, com o pronome associado a verbos na terceira pessoa do singular

(*a gente* sem *-mos*), que apresenta formas distintas para o presente e para o passado. Com isso, não seria mais preciso recorrer a outros elementos do texto para determinar o tempo verbal de verbos com o mesmo paradigma, resolvendo assim a questão da ambiguidade potencial. Em nossa amostra, observamos exemplos que ilustram esses verbos:

(74) Inf.: então é um negócio que me dá muito praZER éh:... lá eu encontro vários amigos éh.../ tem **a gente conversa** muito SObre o TIme não só sobre o time sobre o resultado em campo mas tudo que que acontece dentro do clube (Inq. 19 - F2 - DID, homem, 27 anos)

(75) Inf.: aí ela perguntou aí que **a gente conversou** e tudo ela disse que ia fazer um teste comigo... porque ela ia viajar daqui a... no dia eu que eu fui lá uns vinte dias depois ela ia viajar. (Inq. 02 - F1 - D2, mulher, 23 anos)

Nos exemplos (74) e (75), o tempo verbal já é marcado na própria desinência do verbal, não sendo mais necessário recorrer a elementos do discurso para inferir o tempo expresso pelo verbo “conversar”. Esse é um dos aspectos funcionais que desempenha o pronome *a gente* no português brasileiro, tendo o pronome inovador se especializado como forma pretérita, enquanto a forma *nós* parece mais associado a formas verbais no presente, conforme hipotetizaram Scherre, Yacovenco e Naro (2018)

Com objetivo de testar essa hipótese em nossos dados de fala culta fortalezense a partir de amostras representativas das duas décadas, decidimos replicar, com as devidas adaptações, a análise desta variável sobre nossos dados, tendo em vista que “reproduzir estudos originais, testar grupos de fatores já adotados em outros estudos e usar métodos similares têm fornecido uma base sólida para a teorização científica da Sociolinguística” (Carvalho, Freitas e Favacho, 2020, p. 32). Para isso, ao analisar a concordância verbal de primeira pessoa do plural, focamos nossa análise nos dados que representam o modo indicativo, uma vez que verbos com esse tempo verbal são os mais abundantes em nossa base de dados, conforme indicou análise prévia. Destacamos que exemplos trazidos para ilustrar cada variável foram extraídos de nossa base de dados, de acordo com cada nível das variáveis previsoras. Neste caso, há fatores para qual não encontramos ocorrência de determinada variante de concordância verbal, motivo pelo qual não apresentaremos exemplos para cada um dos 4 padrões de concordância na descrição de todas as variáveis predictoras. Dito isto, observamos as seguintes categorias que compõem a variável *tempo e tipo de paradigma verbal do modo indicativo*:

a) *Presente com formas iguais na relação passado/presente*

(76) Inf.: a minha preferência... sempre foi alfabetização... e depois então o::

primeiro grau:: maior como **nós chamamos** aqui né? o primeiro grau maior que é quinta à:: oitava (Inq. 36– F1 – D2, mulher, 66 anos)

- (77) Inf.: eu me lembrei que **a gente fala** muito de Europa **a gente fala** de culinária aqui na América do Sul inclusive era preconceituoso e eu acabei com o preconceito quando eu fui pro Peru (Inq. 50 – F2 – DID, homem, 56 anos)
- (78) Inf.: ... então nós tivemos... com isso... aquilo que **nós chama** de SOvietes... DESde o sistema militar... então... um grupo de miliTares... SE... REBELOU... contra oficiais (Inq. 50 – F1 – DID, homem, 33 anos)

b) *Presente com formas diferentes na relação passado/presente*

- (79) Inf.: vou falar de exemplos um pouco mais palpáveis né tenho certeza que todos nós aqui né que todos **nós estamos** hoje aqui porque **nós temos** uma certa rotina ... seguimos uma rotina diária né (Inq. 21 – F2 – EF, homem, 39 anos)
- (80) Inf.: comprar uma lata de leite NInho setecentos cruzeiro... TODas as semanas... então a **gente tem** uma uma alimentação re::gular... DENtro de um padrão de classe MÉdia... (Inq. 09 – F1 – DID, mulher, 38 anos)
- (81) Inf.: e aí eu sozin::ha não não conSIgo colocar essa vonTAdé né pra caminhar:: e aí eu tô/ eu também queria MUIto ser ser uma:: uma empreSÁria mas no momento ainda não... **nós não tem** condições:: é pode ser um sonho futuro (Inq. 04 – F2 – DID, mulher, 31 anos)
- (82) Inf.: hoje em dia no caso nós viramos melhores amigas... no caso **a gente somos** muito unidas até porque... esto::u na minha casa... quando vejo elas... é às vezes... então a gente só para pra conversar (Inq. 07 – F2 – DID, mulher, 24 anos)

c) *Pretérito perfeito com formas diferentes na relação passado/presente*

- (83) Inf.: pra você ter uma idéia **nós tivemos** uma economia de dez mil dólar...só pra seis meses...e e só uma determinada quantidade de óleo...por aí você vê (Inq. 45 – F1 – D2, homem, 52 anos)
- (84) Inf.: a gente passou:: alguns anos lá estudando e depois fui pro:: Rui BarBOsa ali::... na::... na:: Imperador ai da/do Rui Barbo::sa que **a gente foi** pro Lourenço Filho eu o seu pai (Inq. 85 – F2 – D2, mulher, 56 anos)
- (85) Inf.: porque quando nosso tio morreu **nós não foi** mais a praia... não gostava de praia não (Inq. 62 – F2 – D2, homem, 60 anos)
- (86) Inf.: brincava no meio da rua corren::do brincava de PEdra brincava de (guizado)... era a gente sentava na calça::da tudo isso a gente fazia tudo tudo tudo isso **a gente fizemos** aí quando chegava o terço... (Inq. 29 – F2 – DID,

mulher, 62 anos)

d) *Pretérito perfeito com formas iguais na relação passado/presente*

- (87) Inf.: as palavras... quando **nós estudaMos**... no nosso tempo quando **a gente estudou**... eh::... a colocaÇÃO... a única coisa que vem ... cabeça é a colocação dos pronomes... (Inq. 114 – F1 – EF, homem, 62 anos)
- (88) Inf.: não queria mais morar com ele nem (ele)nem minha mãe mas aí depois de uns meses foram conversar::... aí **a gente voltou** pra casa... pra pro/ pro apartamento de novo (Inq. 04 – F2 – EF, mulher, 31 anos)
- (89) Inf.: parece que eu dei um chute que embarquei a bola... eles foram pegar a bola e **a gente corremos** né e fomos embora... mas foi dezenove a nove né... (Inq. 62 – F2 – D2, homrm, 69 anos)

e) *Pretérito imperfeito*

- (90) Inf.: quer dizer eram... VINte trinta raPAzes sabe?... **nós fazíamos** festas constantemente **nós tínhamos** um clube... então nesse clube **nós chamávamos** PAdre pra dar conferência psiquiAtra psicólogo... (Inq. 12 – F1 – DID, mulher, 41 anos)
- (91) Inf.: eu me lembro que até quando **a gente ficava** muito tarde na rua a mãe da gente fazia medo que a POLÍcia ia peGAR **a gente tinha** nem ladrão era que a polícia ia pegar a gente (Inq. 16 – F2 – DID, homem, 38 anos)
- (92) Inf.: os rapazes só entrava com de paletó e graVA::ta né... e assim era feito no ano que tivesse essas festas **nós tava** lá né SEMpre tinha um amigo de um conhecido (Inq. 13 – F2 – DID, homem, 59 anos)
- (93) Inf.: a gente chegou lá uma vez a gente tava **a gente estávamos** em quatro na época dois casais aí chegamos numa cidadezinha que não me lembro bem o nome AÍ eu acostumado com as pizzas daqui achava que a pizza grande seria pouco (Inq. 50 – F2 – DID, homem, 56 anos)

## II) Estrutura do verbo:

Trabalhos, como o de Zilles, Maya e Silva (2000), atestaram que essa variável exerce poder sobre a realização dos pronomes *nós/ a gente*. Para os autores, os verbos simples condicionam a realização da variável inovadora com ausência do morfema plural (*a gente sem -mos*). Com o objetivo de testar essa hipótese descrita pelos autores, controlamos os seguintes fatores:

a) *Verbo simples*

- (94) Inf.: porque é uma linGUAgem metonímica tão próxima da realidade que **nós conhecemos**... né?... não é NAda de fantasia... num é alguma coisa distante daquilo que conhecemos (Inq. 56 – F1 – EF, mulher, 54 anos)
- (95) Inf.: **a gente se reunia** constantemente então **a gente vestia** quase que... a mesma coisa assim enTÃO... no/ eh:: as ROUpas na época era/ o QUE aparecia como MO::da... (Inq. 12 – F1 – DID, mulher, 41 anos)
- (96) Inf.: nós assistimos ((ruído)) a tantos filmes que:: principalmente cinema de ARte... quase todo sábado **nós ia**... mas eu me lembro de UM que é Jhonny vai à GUerra... (Inq. 08 – F1 – DID, homem, 40 anos)
- (97) Inf.: muitas casas aqui caíram... acho que tu não lembra... mas mas **a gente botamos** foi uma forquilha pra segurar... parece que quem botou foi aquele nosso primo/ primo do pai... era o Z. né... (Inq. 62 – F1 – D2, homem, 60 anos)

b) *Verbo composto (verbo ter/haver + participio)*

- (98) Inf.: ele realmente necessitava/ tínhamos que ser polivalente... com a farmácia do meu esposo **nós tínhamos formado** no começo... e:: nós atingimos não só a parte de saúde mas a parte social (Inq. 23 – F1 – DID, mulher, 61 anos)
- (99) Inf.: ESSA visualização essa menTAlização... da população em si também /tá? então os distúrbios mais freqüentes que **a gente tem pego** nesses últimos anos é os distúrbios de relacionados à área educacional (Inq. 21 – F1 – DID, homem, 25 anos)

c) *Locução verbal (verbo principal conjugado + verbo auxiliar em sua forma nominal)*

- (100) Inf.: fly debly quer dizer... objetos voadores sombras voadoras né? já no caso que **nós estamos falando** aqui não... seriam SObras que vão formar OUtros astros menores... (Inq. 53 – F1 – EF, homem, 53 anos)
- (101) Inf.: trabalhar para o reino de Deus gratuitamente como eu sempre digo tem que ser gratuitamente... e CURsos só mesmo é de:: aprimoraMENto de como:: **a gente pode levar** o reino de Deus aos nossos irmãos:: porque:: eles tão precisando MUIto (Inq. 39 – F1 – D2, mulher, 66 anos)
- (102) Inf. PARA... eh:: a oraÇÃO... ou para a nossa composiÇÃO... **nós num precisava escrever** mais nada porque o dicionário traz todas as palavras... então num seria... o que falta... é a nossa... criatividade (Inq. 114 – F1 – EF, homem, 62 anos)
- (103) Inf.: aí te/ tinha um local lá que **a gente vamos conhecer** aqui... é um local parecia assim tipo a costa mendes daqui... cheio de DOce doce... cada coisa/ pra gente degu/ comer né... comprar... (Inq. 24 – F2 – DID, mulher, 63 anos)

### III) Conjugação verbal

A variável conjugação verbal é considerada em alguns trabalhos sobre a concordância verbal de primeira pessoa do plural em amostras do português falado no Brasil. Zilles, Maya e Silva (2000), Zilles e Batista (2006), Almeida (2006), Agostinho (2013) e Silvano (2016) apontam maior tendência dos verbos de terceira conjugação ao uso da marca de plural *-mos*, que está associada ao uso do pronome padrão *nós*. Foeger (2014), em estudos sobre os padrões de concordância verbal, identificou uso mais acentuado da concordância não padrão com verbos na primeira conjugação, que atribui ao fato de, nesses verbos, ocorrer troca da vogal temática na diferenciação entre singular e plural.

Pautados na literatura consultada, acreditamos que os resultados desta pesquisa vão ao encontro dos resultados obtidos nas pesquisas sobre o tema, os confirmando. Com isso, a título de análise, observaremos a variável conjugação verbal distribuída nos seguintes níveis:

#### a) Primeira conjugação

- (104) Inf.: um di::a... que tem vinte quatro ho::ras... ***nós precisamos*** dormir... oito horas por di::a... né... ***nós precisamos*** fazer as nossas refeições mais nós também temos... o nosso trabalho... (Inq. 48 – F2 – DID, homem, 36 anos)
- (105) Inf.: outra caracteRÍStica... SERIA... pelo menos o... o MAIS geral que ***a gente encontra*** nos LIVros... seria QUE ele tem um sisTEma ... fácil de ser aprendido... certo? (Inq. 138 – F1 – EF, mulher, 31 anos)
- (106) Inf.: perto da minha vó era um terreno SÓ... né? ***nós morava*** num terreno grande... na frente ficar na casa da minha avó e atrás ficava a casa dos meus pais e aí eu passei minha inFÂNcia... (Inq. 29 – F2 – DID, mulher, 62 anos)
- (107) Inf.: as memórias vai falhando né? vai falhando... ***a gente já falamos*** dos carro... ah... também uma coisa que:: que:: num existe hoje...

#### b) Segunda conjugação

- (108) Inf.: na sociedade capitalista... ***nós vimos***... recentemente... todo dia ***nós vemos*** esse desperdício... o BraSIL... é o país do desperdício... mas no nos paÍ-ses comunistas... eles começaram... a CONtrolar... o abastecimento (Inq. 17 – F1 – EF, homem, 33 anos)
- (109) Inf.: por que o Deus dele HOje... HOje no mundo que ***a gente vive*** hoje é o dinheiro que na minha concepção de vida isso NINguém tira da minha cabeça que o dinheiro é o diabo em forma de papel (Inq. 25 – F2 – EF, mulher, 52 anos)

- (110) Inf.: todo o desTIIno... desse homem... quer dizer ele quesTIO na exatamente tudo isso... na poeSIa... **nós tem**:: aqui nós temos apenas um fragmento vamos ler... (Inq. 56 – F1 – EF, mulher, 54 anos)
- (111) Inf.: eu tenho vinte e quatro... aí com cinco anos de namoro a gente se casou e hoje **a gente já temos**... no caso dia seis vamos completar três anos de casados... (Inq. 07 – F2 – DID, mulher, 24 anos)

c) *Terceira conjugação*

- (112) Inf.: ligaram pra cá de São Paulo mas eu num /tava aqui /tava viajando... aí nada né?... **nós conseguimos** com o:: o governo né? o governador Tasso Jereissati ele deu nos deu... através do serviço sociAL ele deu nos deu um::... fogão (Inq. 13 – F1 – DID, mulher, 58 anos)
- (113) Inf.: TOdo dia a gente tinha que ir pra vó... e LÁ tamBÉM **a gente construiu** aSSIM uma inFÂNcia né? que tinham os meus tios que na época era tudo solTE::ro... até hoje tem muitas músicas que eu escuto que eu gosto (Inq. 28 – F2 – DID, mulher, 44 anos)
- (114) Inf.: aí eu só sei que nós/ eles resolveram passar o final do ano... **nós só ia** pro natal... aí resolveram... tá certo vamos passar o final do ano... passamos... aí o pai ficou chateado porque nós não passamos aqui em casa (Inq. 24 – F2 – DID, mulher, 63 anos)
- (115) Inf.: vamos dizer a gente tinha um... vamos dizer... **a gente saíamos** pra::... vamos dizer para as festas em Parangaba tranquilamente íamos e voltávamos a qualquer hora da noite... (Inq. 56 – F2 – D2, homem, 63 anos)

#### IV) Tonicidade do verbo

Estudos sociolinguísticos que versam sobre a concordância verbal da primeira pessoa do plural têm demonstrado efeito significativo da variável *tonicidade* sobre a escolha entre os pronomes *nós* e *a gente* com as concordâncias padrão e não padrão (Borges, 2004; Tamanine, 2010; Franceschini, 2011). Borges (2004), em estudo sobre a gramaticalização do pronome *a gente* nas cidades de Jaguarão e Pelotas, identificou os monossílabos tônicos e átonos associados a forma inovadora. Por outro lado, nas duas amostras, o autor percebeu que paroxítonos favorecem o uso do pronome padrão *nós*, evidenciando a tendência entre os informantes de evitar as formas proparoxítonas. Franceschini (2011) encontrou resultados muito próximos aos encontrados por Borges (2004), sendo os monossílabos tônicos e átonos categóricos ao uso do pronome *a gente*, enquanto os paroxítonos são fortes aliados da forma padrão *nós* na comunidade de Concórdia -SC.

Baseados nesses estudos, acreditamos que, na comunidade de fala de Fortaleza, encontraremos as mesmas tendências referentes ao uso dos pronomes em função da tonicidade. Nossa hipótese é a de que haverá maior favorecimento da forma pronominal *a gente* sem *-mos* com monossílabos átonos e tônicos. Sendo assim, as categorias testadas nesta variável, em nosso estudo, são as seguintes:

a) *Monossílabos tônicos ou átonos*

- (116) Inf.: eu tenho até uma concepção... eh:: acho que durante a esCOla... faculdade... sempre/ sempre **nós é** lecioNAdo que essa questão do de um conflito entre empregaDOR e empregado... né? (Inq. 76 – F2 – D2, homem, 39 anos)
- (117) Inf.: HOje **a gente vê** nas escolas professores... sendo ameaçados... constantemente as escolas sendo assaltadas... né principalmente nos bairros da periferia... AÍ:: **a gente vê** na história... um professor::... com cinquenta e cinco sessenta anos numa sala de aula (Inq. 56 – F2 – D2, mulher, 50 anos)

b) *Oxítonos*

- (118) Inf.: aí depois você POde juntar aqui coMIgo ou com Eri-NALdo a Tânia uma pessoa mais esclarecida... pra **nós fazer** aquela reunião com os professores (Inq. 16 – F1 – D2, mulher, 30 anos)
- (119) Inf.: ... é a mesma coisa do... dessa aposta aí que eu falei... e ontem **a gente ganhou** de oito ponto lá dentro do Náutico... fomo/ campeão do primeiro turno mas o jogo (Inq. 34 – F1 – D2, homem, 25 anos)

c) *Paroxítonos*

- (120) Inf.: ...então esse sistema que **NÓS implantamos** aqui...ele evita um tipo de acidente desse tipo né?...porque...o maquinista /tá se comunicando com as estações vizinha... (Inq. 45 – F1 – D2, homem, 50 anos)
- (121) Inf.: tinha::...aí tinha que **a gente botava** os números ai muita gente marcava um encontro ai se juntava uma TURma ai ia se encontrar nas praças... ai eu nunca tive coragem por que eu achava que era MUIta gente desconhecida... (Inq. 08 – F2 – D2, mulher, 30 anos)
- (122) Inf.: não era só Farias Brito que tinha era o mais famoso **nós ficava** resolvendo os TDS... o vestibular no ar e nos passava o final de semana o sábado e domingo no/numa no terceiro ano (Inq. 13 – F2 – DID, homem, 59 anos)
- (123) Inf.: muitas casas aqui caíram... acho que tu não lembra... mas mas **a gente botamos** foi uma forquilha pra segurar... parece que quem botou foi aquele nosso primo/ primo do pai... era o Z. né... Inq. 62 – F2 – D2, homem, 60 anos)

d) *Proparoxítonos*

- (124) Inf.: e hoje NEM se brinca mas dessa::... no meio da rua... vamos dizer HOje... **nós morávamos** ali/ ali na Gomes de Matos HOje... RAraramente vai encontrar um domicílio residencial (Inq. 56 – F2 – D2, mulher, 50 anos)
- (125) Inf.: procurou prestar todo esse serviço... que ele realmente necessitava... lá **a gente tínhamos** que ser polivalente... com a farmácia ((ruído)) do meu esposo **nós tínhamos** formado... no começo....

#### 4.5 Procedimentos de coleta e análise dos dados

Os dados deste estudo foram coletados através da audição dos inquiridos das amostras selecionadas, advindas do PORCUFORT Fases I (década de 1990) e II (década de 2020). Trata-se de entrevistas sociolinguísticas, a partir das quais tabulamos as ocorrências do fenômeno variável de primeira pessoa do plural: a *alternância pronominal* e a *concordância verbal* com os pronomes de primeira pessoa *nós* e *a gente*. As amostras coletadas do banco de dados usado como base desta pesquisa já possuem transcrições finalizadas, em documento editável no formato .docx. Sendo assim, à medida em que ocorriam as audições, acompanhamos as transcrições dos inquiridos, nos certificando dos contextos de realização de cada ocorrência do fenômeno para, enfim, coletarmos os dados em planilha do excel.

Após a audição e coleta desses dados, nós procedemos a atribuição de características a cada ocorrência em função das variáveis previsoras elencados, sejam elas linguísticas ou sociais. Por exemplo, em relação ao *sexo/gênero*, atribuímos a categoria “masculino” ou “feminino”, de acordo com o sexo/gênero do informante. De igual maneira, classificamos as ocorrências em relação à faixa etária: 22-35 anos (1a), 36-50 anos (2a), a partir de 51 anos (3a). O mesmo foi feito em relação as variáveis previsoras de ordem linguística. Desta forma, fizemos com que cada ocorrência fosse codificada em função das variáveis, linguísticas e sociais, testadas neste estudo. A figura abaixo ilustra a planilha na qual os dados foram inseridos e classificados:

**Figura 5:** Coleta dos dados de nossa pesquisa

	A	B	C	D	E	F	G	H	K	L	P
1	VR	SEXO.GENERO	FAIXA.ETARIA	DECADA	TIPO.INQUERITO	REFERENCIALIDADE	TEMPO.VERBAL	TIPO.VERBO	PARALELISMO	CONCORDANCIA	INFORMANTE
2	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	primeiro	agente0	F1D202.1
3	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	manutencao	agente0	F1D202.1
4	agente	mulher	1a	1990	D2	generica	presente.ind	acao	isolada	agente0	F1D202.1
5	agente	mulher	1a	1990	D2	especifica	pret.imper.sub	ter	primeiro	agente0	F1D202.1
6	agente	mulher	1a	1990	D2	especifica	pret.imper.sub	acao	manutencao	agente0	F1D202.1
7	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	isolada	agente0	F1D202.1
8	agente	mulher	1a	1990	D2	especifica	pret.imper	acao	manutencao	agente0	F1D202.1
9	agente	mulher	1a	1990	D2	especifica	pret.imper	acao	isolada	agente0	F1D202.1
10	agente	mulher	1a	1990	D2	especifica	presente.ind	acao	isolada	agente0	F1D202.1
11	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	primeiro	agente0	F1D202.2
12	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	manutencao	agente0	F1D202.2
13	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	manutencao	agente0	F1D202.2
14	agente	mulher	1a	1990	D2	especifica	fut.pret	acao	isolada	agente0	F1D202.2
15	agente	mulher	1a	1990	D2	especifica	pret.perf	dicendi	isolada	agente0	F1D202.2
16	agente	mulher	1a	1990	D2	generica	presente.ind	acao	primeiro	agente0	F1D202.2
17	agente	mulher	1a	1990	D2	generica	presente.ind	epistemico	manutencao	agente0	F1D202.2
18	agente	mulher	1a	1990	D2	especifica	futuro	acao	primeiro	agente0	F1D202.2
19	agente	mulher	1a	1990	D2	especifica	futuro	acao	manutencao	agente0	F1D202.2
20	agente	mulher	1a	1990	D2	especifica	presente.ind	dicendi	isolada	agente0	F1D230.1
21	agente	mulher	1a	1990	D2	especifica	pret.perf	acao	manutencao	agente0	F1D230.1
22	agente	mulher	1a	1990	D2	generica	presente.ind	dicendi	isolada	agente0	F1D230.1
23	agente	mulher	1a	1990	D2	especifica	pret.imper	acao	primeiro	agente0	F1D230.1
24	agente	mulher	1a	1990	D2	especifica	pret.imper	acao	manutencao	agente0	F1D230.1

Fonte: elaboração própria

Conforme observamos na figura 5, foi atribuída uma característica para cada dado em função das variáveis controladas, de modo que fosse contemplada todas as nossas ocorrências de primeira pessoa do plural. Salientamos que procedemos do mesmo método de codificação detalhado acima para todos os 4574 dados da realização dos pronomes de primeira pessoa do plural com suas respectivas concordâncias em nossa amostra. Em seguida, transportamos o arquivo em uma planilha no formato .csv, para que pudesse ser submetida à análise estatística por meio do software *RStudio*. A esse respeito, a próxima seção é dedicada a informações sobre o modelo quantitativo e a ferramenta estatística.

#### 4.6 Análise estatística: o modelo quantitativo

Um modelo quantitativo é um dispositivo metodológico que visa explicar a realização de fenômenos variáveis nas línguas e compreender os padrões quantitativos de uso desses fenômenos por meio de um modelo matemático (Guy; Zilles, 2007). De acordo com Labov (1972), um modelo quantitativo deve envolver várias etapas: (I) definir a variável resposta (variável dependente) e as possíveis variáveis preditoras (grupo de fatores de ordem linguística e social que podem condicionar o uso da variável resposta); (II) escolher a comunidade de fala e selecionar dos informantes; (III) coletar os dados de fala, (IV) proceder uma análise quantitativa dos dados; e (V) apresentar a interpretação dos resultados e indicar quais fatores que influenciam o uso de uma das variantes da variável resposta. Tendo as três

primeiras etapas desenvolvidas, enfocaremos aqui no tratamento dos dados em termos de estatística básica e inferencial.

A análise quantitativa de determinado fenômeno variável possibilita o estudo mais aprofundado da variação e mudança linguística, permitindo ao pesquisador compreender sua sistematicidade, os contextos sociais e linguísticos que atuam a variação, bem como possíveis indícios de mudança linguística (Guy; Zilles, 2007). Para isso, realiza-se uma análise distribucional dos dados linguísticos, utilizando métodos quantitativos, além de calcular o efeito das variáveis (tanto linguísticas quanto sociais) na seleção de uma das formas linguísticas em competição na amostra analisada (Tagliamonte, 2007). O objetivo da análise distribucional é resumir os dados de forma a minimizar detalhes irrelevantes e apresentar uma visão geral do fenômeno, para proporcionar ao pesquisador um panorama dos dados originais a serem tratados.

Vale destacar que a escolha do método para tratamento inicial dos dados de uma pesquisa depende da natureza da variável resposta: se numérica (como a altura das vogais médias pretônicas, por exemplo) ou nominal (também chamada de categórica, como a realização dos pronomes *nós* e *a gente*).

Na análise de uma variável resposta nominal, como a alternância ou a concordância verbal entre os pronomes *nós* e *a gente*, verifica-se inicialmente a proporção geral de cada uma das variantes nos dados da amostra e em cada um dos contextos pesquisados. Isso inclui examinar os níveis que compõem as variáveis preditoras, recorrentemente chamada de variáveis independentes nos estudos sociolinguísticos (Tagliamonte, 2012). Calcula-se então a proporção ou percentual correspondente aos valores de frequência encontrados. A porcentagem varia no intervalo de 0% a 100% e oferece uma maneira de resumir a proporção de resultados, possibilitando uma visão global sobre o comportamento do fenômeno em análise.

Ou seja, essa análise inicial, em termos de frequência e proporção, permite que o pesquisador visualize a distribuição das variantes. Isso envolve verificar e demonstrar as tendências encontradas, apresentando os dados por meio de tabelas ou gráficos para facilitar a compreensão do fenômeno estudado (Tagliamonte, 2012). Essa análise distribucional também possibilita a formulação de hipóteses não elaboradas antes da análise distribucional, não observadas em forma de variáveis, que serão posteriormente testadas por meio de um modelo estatístico.

Segundo Guy e Zilles (2007), a análise de fenômenos variáveis envolve a contagem das ocorrências das variantes, a descrição de tendências e da extensão da variabilidade, bem como das restrições ou fatores que as influenciam, por meio de métodos estatísticos. Essa

análise multivariada, desenvolvida na linguística, visa compreender a variação estruturada, governada por regras, no uso da língua. Seu objetivo é separar, quantificar e testar a significância dos efeitos de fatores contextuais, sociais e linguísticos na escolha de uma das variantes do fenômeno linguístico analisado.

Em outras palavras, objetivo da análise quantitativa é verificar o quanto cada variável previsora, seja ela linguística ou social, contribui para a realização de uma ou outra forma variante que constitui a variável resposta. Considerando que, na língua em uso, essas variáveis independentes sempre aparecem associadas, a atuação de uma regra variável ocorre em consonância com o efeito simultâneo de mais de uma variável independente. Portanto, é necessário calcular o efeito de todas as variáveis predictoras presentes em determinado contexto, atuando em conjunto, em interação.

Assim, uma análise quantitativa da variação busca entender o comportamento de uma variável resposta de acordo com um conjunto de variáveis sociais e estruturais que coocorrem com essa variável resposta (Guy; Zilles, 2007; Tagliamonte, 2012). Para alcançar esse objetivo, desde a década de 1960, com as publicações de Willian Labov, foram propostos alguns modelos matemáticos para análise de dados linguísticos, até o surgimento do modelo logístico, que é adotado nos trabalhos mais recentes em sociolinguística.

Labov (1969) discutiu suas teorias e métodos principalmente em suas obras mais conhecidas, como *Contraction, Deletion, and Inherent Variability of the English Copula* (1969). Nesta obra, o autor descreve suas observações sobre a variação linguística e os fatores que a influenciam, mas não apresenta explicitamente uma formulação matemática que representa seu primeiro modelo de análise, o Modelo Aditivo. Para tornar a compreensão mais didática, no entanto, apresentaremos a seguir uma simplificação conceitual para expressar a ideia central do modelo de análise proposto pelo linguista: a ideia de que vários fatores contribuem independentemente para a variação linguística.

$$Y = X_0 + X_1 + X_2 + \dots + X_n$$

Neste modelo, Y é a frequência geral de aplicação da regra, ou seja, o total da frequência dos fatores.  $X_0$  pode ser entendido como *input*, o valor inicial comum para todos os contextos.  $X_1$  e  $X_2$  são a frequência do efeito individual de cada fator, e  $X_n$  representa o número total de fatores que atuam sobre a variável resposta. Para Cedergren e Sankoff (1974) esse modelo não se mostrou eficiente, pois permitia que dois fatores altamente favoráveis à aplicação da regra tivessem o mesmo efeito, sem distinção. Além disso, o modelo aditivo

apresentou problemas na escala de valores, pois sua fórmula não seria capaz de prever probabilidades de aplicação em muitos contextos diferentes, fora do intervalo entre zero e um. Não havia, ainda, garantia de que seu resultado não pudesse ultrapassar 100% ou ficar abaixo de 0%, e, com isso, o modelo foi “abandonado por causa dos problemas de natureza técnica considerados insuperáveis na época” (Naro, 2002, p. 20).

Desenvolvidos como uma alternativa ao modelo aditivo de Labov, os modelos multiplicativos propostos por Henrietta Cedergren e David Sankoff, em 1974, são uma interpretação probabilística que substituiu frequências por probabilidades na análise de regras variáveis. Mollica e Braga (2015) explicam que os modelos multiplicativos consideram que a probabilidade de aplicação de uma variante é o produto das probabilidades de cada fator contextual de aplicação e, para isso, passam a adotar probabilidades no lugar das frequências na análise de regras variáveis.

O primeiro modelo proposto pelos autores, o Modelo Multiplicativo de Aplicação (Cedergren e Sankoff, 1974, p. 339), apresenta a seguinte formulação:

$$P = p_0 \times p_1 \times p_2 \times \dots \times p_n$$

Em que P é a probabilidade de aplicação da regra, considerando os diversos contextos gerados pela interação entre as variáveis,  $p_0$  é a probabilidade do *input* comum em todos os ambientes,  $p_1$  e  $p_2$  são valores para cada variável. Nesse modelo, todos esses fatores individuais são multiplicados para obter o produto total, ou seja, é levado em conta a interação de fatores que coexistem no processo de realização de determinada regra variável. Apesar de considerar a interação entre os diversos fatores, Mollica e Braga (2015, p.21) explicam que o “modelo multiplicativo de aplicação parecia até menos satisfatório do que o aditivo, por postular que o efeito de dois fatores altamente favoráveis a aplicação da regra é menor que o efeito individual de cada um deles separadamente”.

Vamos imaginar o efeito de duas variáveis altamente favorecedoras de determinada variante, as variáveis A e B, com valores 0,9 e 0,8 respectivamente. De acordo com o modelo multiplicativo de aplicação, é necessário multiplicar os valores de todas as variáveis para se chegar ao produto da probabilidade, considerando a atuação desses dois fatores em conjunto. O produto de A x B é 0,72, ou seja, a atuação conjunta dessas duas variáveis (altamente favorecedoras da regra variável) resulta em um valor menor que a atuação individual de cada uma delas na aplicação da regra, motivo que leva os autores a criticar esse modelo.

Cedergren e Sankoff (1974), partindo da ideia de que a probabilidade é uma escala que vai de 0 até 1, em que  $P$  é a probabilidade para aplicação da regra, desenvolveram um novo modelo – Modelo Multiplicativo de Não Aplicação - que parte do pressuposto de que  $(1-P)^{10}$  é a probabilidade de não aplicação de uma variante. Com isso a probabilidade de não aplicação é calculada multiplicando a probabilidade complementar de cada fator, conforme podemos ver em Cedergren e Sankoff (1974, p. 337):

$$(1-P) = (1 - p_0) \times (1 - p_1) \times (1-p_2) \times \dots \times (1 - p_n)$$

Neste modelo,  $(1-P)$  é probabilidade de não aplicação da variante;  $p_0, p_1, p_2$  e  $p_n$  são os fatores individuais que contribuem para a aplicação da variante. Cada  $p$  representa a probabilidade de um fator específico contribuir para a aplicação da variante de referência;  $(1 - p_0), (1 - p_1), (1 - p_2)$  e  $(1 - p_n)$  correspondem a probabilidade complementar de cada fator. Ou seja, considerando que  $p$  é a probabilidade de um fator favorecer determinada variante, então  $(1-p)$  é a probabilidade de o fator não favorecer.

Multiplicar todas essas probabilidades complementares fornece a probabilidade de não aplicação da variante como um todo. Essa expressão é comumente usada em contextos em que vários fatores independentes podem afetar a ocorrência de um evento - caso da variação entre as formas pronominais *nós* e *a gente* e suas devidas concordâncias verbais -, e a probabilidade de o evento ocorrer é calculada considerando as probabilidades complementares de cada fator.

Inicialmente, o desenvolvimento de um modelo de não aplicação resolveria a questão dos dois fatores altamente favorecedores, uma vez que o valor da interação entre os fatores ( $P$ ), em termos de probabilidade, seria maior que a atuação de cada fator isoladamente. No entanto, espera-se que a interação entre dois fatores altamente desfavorecedores desfavoreça ainda mais a atuação de determinada variante, fato que não se confirma no modelo proposto.

Por exemplo, considerando dois fatores altamente desfavorecedores, na realização de determinada variável linguística, C (0,1) e D (0,3), espera-se que a ação conjunta dessas variáveis resulte em um valor muito baixo, próximo de zero, ainda mais desfavorecedor que a ação individual de cada fator. Entretanto, no modelo multiplicativo de não aplicação, a interação de C x D gera uma probabilidade de  $P = 0,37$ , ou seja, menos desfavorecedora que a atuação

---

<sup>10</sup> Lê se: “um menos a probabilidade de aplicação da regra”.

isolada de cada fator. Naro (2015) reconhece o problema de ambos os modelos e ressalta que a utilização de determinado método deve atender ao propósito do pesquisador:

o modelo multiplicativo de aplicação é apropriado à coatuação de fatores desfavorecedores, enquanto o modelo multiplicativo de não aplicação é apropriado para fatores favorecedores. Em qualquer caso, estamos longe da função que procuramos. Em face da massa de dados linguísticos reais, não há nenhum meio rigoroso de decidir qual modelo utilizar (Naro, 2015, p.22).

Com o objetivo de superar as limitações dos modelos anteriores, principalmente a inadequação no tratamento de múltiplos fatores, Rousseau & Sankoff (1978) desenvolveram o Modelo Logístico como uma alternativa, tendo como conceito principal a ideia de que a probabilidade de aplicação de uma variante segue uma curva logística, ou seja, não é linear, mas leva em conta fatores contextuais e suas interações, e a curva logística representa como a probabilidade varia em diferentes contextos. O modelo apresenta a seguinte sintaxe, conforme exemplifica Naro (2015, p. 23):

$$\frac{P_t}{(1 - P_t)} = \frac{P_o}{(1 - P_o)} \times \frac{P_1}{(1 - P_1)} \times \frac{P_2}{(1 - P_2)} \dots \frac{P_n}{(1 - P_n)}$$

Com o desenvolvimento do modelo de regressão logística, foi possível analisar de forma equilibrada as questões dos fatores desfavorecedores e favorecedores, de forma conjunta, englobando características dos dois modelos anteriores. O Modelo Logístico de Rousseau & Sankoff revolucionou a análise da variação linguística, permitindo uma compreensão mais refinada dos padrões de uso das variantes em diferentes contextos.

Para Klunk (2007, p. 35), esse modelo “é mais usado na análise de dados binários, pois é simétrico no que diz respeito a probabilidades de aplicação e não-aplicação”. Essa simetria se dá pela adoção de pesos relativos no lugar de probabilidade, sendo uma escala que vai de 0 a 1, tendo o ponto de neutralidade definido em 0,5. Logo, pesos relativos acima de 0,5 indicam favorecimento de determinado fator e pesos menores que o ponto neutro indicam desfavorecimento.

A partir do século XXI, ocorreram avanços significativos nas técnicas estatísticas utilizadas pela Sociolinguística para o tratamento de dados de variação linguística. Tagliamonte (2012), assevera que um dos avanços mais importantes foi o desenvolvimento de modelos de efeitos mistos para tratamento dos dados. Esses modelos, aplicado ao modelo logístico, inclui, além das variáveis independentes fixas, variáveis aleatórias. Para Oushiro (2017, p. 177), de maneira geral, uma variável aleatória é aquela que não pode ser reproduzida em diferentes

amostras, e tem por objetivo “checar num modelo de efeitos mistos se os mesmos efeitos fixos continuam a ser relevantes após a inclusão das variáveis aleatórias”. O *Informante*, por exemplo, é um tipo de variável aleatória, pois o mesmo indivíduo não pode fornecer dados para amostra de fala culta e de fala popular ao mesmo tempo, já que o perfil social muda de acordo com a amostra. Deferentemente do *Sexo/gênero* que, em ambas as amostras, pode ser controlado em função dos fatores *masculino* e *feminino*.

Trabalhos recentes em Sociolinguística quantitativa têm demonstrado resultados de análises realizadas a partir de modelos de regressão (Teixeira, 2021; Oushiro, 2022; Freitas, Rodrigues e Santos, 2022). Os modelos de regressão são uma ferramenta poderosa para analisar dados sociolinguísticos, especialmente quando lidamos com variáveis categóricas – caso da variação na alternância pronominal e na concordância verbal de primeira pessoa do plural – pois fornecem dados importante sobre a atuação das variáveis previsoras (independentes) sobre a realização da variável resposta.

No entanto, a escolha do modelo de regressão deve ser definida em função do fenômeno a ser analisado. Se considerarmos como variável resposta à altura da vogal /e/ (F1), medida em Hz, estamos diante de uma *variável contínua* (ou numérica), e, portanto, o modelo adequado para a análise dos dados é o de *regressão linear*. Por outro lado, ao adotarmos como variável resposta a alternância pronominal entre *nós* e *a gente*, temos uma *variável categórica* (ou nominal), logo o modelo de análise adequado é o *modelo de regressão logística*. Por estarmos analisando os padrões de *alternância pronominal* e de *concordância verbal*, ou seja, variáveis categóricas, enfocaremos no modelo de regressão logística, dada a sua adequação para tratar este tipo de variável

#### **4.7 Estatística básica e inferencial para o tratamento de dados linguísticos**

Antes de detalharmos o modelo de regressão logística adotado para análise dos dados de alternância pronominal e de concordância verbal de primeira pessoa do plural, é necessário dar luz a alguns princípios básicos da estatística que guiaram a análise e a interpretação dos dados, como os conceitos de hipótese nula e hipótese alternativa, *loggods*, entre outros. Esses conceitos são fundamentais para o pleno entendimento do modelo de regressão logística, bem como a interpretação dos resultados de nossa análise.

#### 4.7.1 Alguns conceitos básicos da estatística inferencial

A estatística inferencial nos ajuda a entender tendências em grupos maiores com base em dados de determinada amostra. Desta forma, com base em amostras representativas, podemos projetar padrões linguísticos, em termos de tendência, adotados por toda a comunidade de fala. Esses padrões são analisados a partir da formulação de hipótese, essencial a qualquer pesquisa científica. Em termos de estatística inferencial, o conceito de hipótese se remodela, e passa a ser associado ao efeito de uma variável que pode ser testada em amostras representativas. Para análise dos dados, é preciso considerar a hipótese nula e a hipótese alternativa.

A hipótese nula (representada por  $H_0$ ) é uma afirmação que declara que não há relação significativa entre o fenômeno analisado e determinado fator. Por exemplo, ao analisar a alternância entre os pronomes *nós* e *a gente* na fala de informantes do sexo feminino e do sexo masculino, a hipótese nula assume que não há relação entre o sexo do informante e a realização da primeira pessoa do plural. Em outras palavras, a hipótese nula é a de que homens e mulheres apresentam o mesmo comportamento no uso de determinada variante (ou diferenças estatisticamente não significativas) para os pronomes.

A hipótese alternativa (representada por  $H_1$ ) é a proposição de que há interação estatisticamente significativa entre dois fatores. Assim, ao analisar a alternância pronominal entre homens e mulheres, a hipótese alternativa assume a ideia de que há sim diferenças significativas no emprego dos pronomes entre homens e mulheres. Os indicativos para acatar ou não a hipótese nula são o nível de significância e, principalmente, o valor de  $p$  (ou *p-value*) (Oushiro, 2017).

O nível de significância é um parâmetro que determina quando é possível acatar ou rejeitar a hipótese nula ou alternativa e, por isso, é essencial verificar essa medida em um estudo com dados estatísticos. Oushiro (2017, p. 83) explica que não existe um número exato que determine o limite fixo para considerar determinado resultado estatisticamente significativo, mas “convencionalmente, a comunidade científica costuma usar o limite de 5% para considerar algo como muito pouco provável de acontecer ao acaso. Isso é chamado de nível  $\alpha$  (*nível alfa*): o limite estabelecido pelo pesquisador para rejeitar a hipótese nula”. Nesta pesquisa, adotamos o nível alfa convencionado pela comunidade acadêmica de  $p < 0,05$ .

O *valor p*, por sua vez, é a probabilidade de se obter determinado resultado em caso de a hipótese nula ser verdadeira, ou seja, indica se as diferenças observadas em determinado fator são estatisticamente significativas ou se são fruto do acaso. Com isso, se o valor de  $p$  é

menor que o nível de significância ( $p < 0,05$ ), podemos rejeitar a hipótese nula e assumir que a hipótese alternativa é verdadeira. Ao contrário, se o valor de  $p$  é maior que a significância ( $p > 0,05$ ), não é possível rejeitar a hipótese nula, e assumimos que há pouca ou nenhuma interação entre as variáveis.

Além da significância e do *valor p*, outros conceitos pertinentes para a análise em estatística inferencial são as definições de *odds*, *loggods* e *odds ratio*. No entanto, para entender cada um desses princípios, é necessário que se tenha claro a ideia de probabilidade, uma vez que esses elementos se pautam, essencialmente, no conceito de probabilidade. Para ilustrar esse conceito, vamos observar os resultados de Araujo (2018) sobre a variação entre os pronomes sujeitos *nós* e *a gente* em função da faixa etária:

**Tabela 2** - Atuação da *faixa etária* sobre o pronome *a gente* em Araujo (2018)

<b>Fatores</b>	<b>Nós</b>	<b>A gente</b>	<b>Total</b>
15 a 25 anos	50	191	241
26 a 49 anos	136	185	321
A partir de 50 anos	156	281	437
<b>Total</b>	<b>342</b>	<b>657</b>	<b>999</b>

Fonte: adaptado de Araujo (2018)

A probabilidade é a unidade de medida que indica a possibilidade de ocorrência de determinada variante em um dado contexto, sendo obtida pela razão entre o número de ocorrências dessa variante e o total geral de ocorrências de uma amostra pesquisada, em uma escala de vai de 0 a 1. Assim sendo, a probabilidade geral de realização da forma pronominal inovadora *a gente* (Tabela 2) obedece a seguinte formulação:

$$P_{\text{geral}} = \frac{\text{TOTAL } a \text{ gente}}{\text{TOTAL geral}} \quad \rightarrow \quad P_{\text{geral}} = \frac{657}{999} = 0,65$$

Podemos dizer que, de forma geral, na pesquisa de Araujo (2018), há maior probabilidade de realização do pronome *a gente* do que de *nós* na amostra pesquisada, com valor de 0,65, indicando favorecimento da forma inovadora de forma geral em toda a amostra. É possível também checar a probabilidade de realização da variável de referência em contextos específicos. Ao considerar uma faixa etária específica, por exemplo, a probabilidade de realização da forma *a gente* é dada pela razão entre o número de ocorrências da variante em uma faixa etária específica e o número total de realizações do fenômeno variável dessa faixa

etária. Podemos, então, checar a probabilidade de aplicação da regra entre os mais jovens e entre os mais velhos da amostra, partindo da formulação abaixo:

$$P(\text{mais jovens}) = \frac{\text{TOTAL } a \text{ gente (mais jovens)}}{\text{TOTAL geral (mais jovens)}} \Rightarrow P = \frac{191}{241} = 0,79$$

$$P(\text{mais velhos}) = \frac{\text{TOTAL } a \text{ gente (mais velhos)}}{\text{TOTAL geral (mais velhos)}} \Rightarrow P = \frac{281}{437} = 0,64$$

O resultado do cálculo apontou a probabilidade de 0,79 para a realização da forma pronominal inovadora *a gente* entre os informantes de 15 a 25 anos, e probabilidade de 0,64 entre os mais velhos da amostra, indicando que há maior probabilidade de realização da forma inovadora, em comparação com a forma canônica, em ambas as faixas etárias, sendo mais expressiva entre os mais jovens.

Tendo claro o conceito de probabilidade, básico para interpretação de outras medidas estatísticas, nos deteremos na definição de *odds* e *logodds*. Em relação ao primeiro, *odd* é a razão de chance da realização de um evento, ou seja, é a medida estatística que estabelece uma relação entre a probabilidade de um evento acontecer e a probabilidade de não acontecer. Em trabalhos sociolinguísticos, as *odds* são frequentemente usadas na interpretação de resultados com variável resposta binária, como a realização variável dos pronomes *nós* e *a gente*, devido ao resultado, que indica a possibilidade entre “realização” ou “não realização”, “sucesso” ou “fracasso”, e “sim” ou “não”, por exemplo. A formulação abaixo representa o cálculo para se checar as chances de realização de um evento:

$$O = \frac{P}{(1-P)}$$

Por exemplo, com dados da tabela 2, foi possível calcular a probabilidade de realização da variável inovadora entre os informantes mais jovens (0.79) e os mais velhos (0.64) em amostra de fala popular de Fortaleza. Tendo em mãos as probabilidades, é possível chegar ao valor das *odds* em cada contexto:

$$O = \frac{P}{(1-P)} \Rightarrow O_{(\text{jovens})} = \frac{0.79}{(1-0.79)} \Rightarrow O_{(\text{jovens})} = \frac{0.79}{0.21} \Rightarrow O_{(\text{jovens})} = 3,76$$

$$O_{(\text{velhos})} = \frac{0.64}{(1-0.64)} \Rightarrow O_{(\text{velhos})} = \frac{0.64}{0.36} \Rightarrow O_{(\text{velhos})} = 1,77$$

Através dos cálculos, verificamos que o valor em *odds* de realização da variante inovadora entre os mais jovens é de 3,76, e entre os mais velhos 1,77. Ou seja, é possível constatar, a partir da interpretação do valor de *Odd*, que as chances de realização de *a gente* entre os mais jovens é de 3,76 para 1, ou seja, há 3,76 vezes mais chances de os mais jovens usarem a forma *a gente* ao invés de *nós*. Já entre os mais velhos, o valor de *odds* indica que há 1,77 vezes mais chances de realização do pronome inovador.

Para Levshina (2015), é possível ainda comparar as chances de um evento ocorrer em dois contextos distintos, possibilitando identificar a relação entre esses dois fatores de uma mesma variável. Isso é possível através do cálculo de *odds ratio*, ou razão das chances, uma medida estatística que compara as chances de um evento ocorrer em um grupo com as chances de ocorrer em outro grupo. A formulação segue a seguinte sintaxe:

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} \quad \Rightarrow \quad OR = \frac{Odds(ab)}{Odds(cd)}$$

Com base nessa fórmula, é possível identificar a razão das chances, ou odds ratio, para a aplicação da regra entre os mais jovens e os de mais idade. Substituindo cada um dos termos, temos então os seguintes resultados:

$$Odds(faixa I) = \frac{Odds(faixa I)}{Odds(faixa III)} \quad \Rightarrow \quad Odds(faixa I) = \frac{3,76}{1,77} \quad \Rightarrow \quad Odds(faixa I) = 2,12$$

$$Odds(faixa III) = \frac{Odds(faixa III)}{Odds(faixa I)} \quad \Rightarrow \quad Odds(faixa III) = \frac{1,77}{3,76} \quad \Rightarrow \quad Odds(faixa III) = 0,47$$

Se a *Odds Ratio* for maior que 1, há indícios de que o contexto em análise apresenta maior chance de aplicação da regra. Se a *odds ratio* for menor que 1, há menor chance de uso da variante de referência. Os resultados indicam, então, que, entre os mais jovens, há tendência ao uso da forma inovadora, enquanto entre os informantes da faixa etária III, há menos chances de que o pronome inovador ocorra.

As *Odds*, ainda, são frequentemente utilizadas em análise de estatística inferencial, através de modelos de regressão, onde os seus valores são convertidos em *logodds*, para ajudar na modelagem dos dados. Para Oushiro (2017), *logodds* constitui uma medida estatística que representa o logaritmo natural da razão entre a probabilidade de um evento ocorrer e a probabilidade de não ocorrer. Em outras palavras, se a probabilidade de um evento ocorrer é *p*, então a probabilidade de não ocorrer é  $1 - p$ . O *logodds* é então calculado conforme a seguinte fórmula:

$$LO = Ln . \left( \frac{P}{1 - p} \right)$$

Onde LO é o valor em *logodds*, Ln é o logaritmo natural, e P é a probabilidade de aplicação da regra. Para Oushiro (2021, p. 325), os *logodds* constituem uma medida de estatística inferencial em que, de modo geral, “os valores positivos indicam tendência ao favorecimento (em relação a outro nível da mesma variável previsor); e valores negativos indicam tendência a desfavorecimento (em relação a outro nível da mesma variável previsor)”. Além disso, valores iguais a zero indicam que o fator ou a variável analisada não tem efeito sobre a regra de aplicação (Gries, 2013). Vale destacar que não é necessário calcular o valor de estimativa, dado em *logodds*, manualmente, pois os valores são fornecidos de forma automatizada através de uma ferramenta estatística.

#### 4.7.2 O modelo de regressão logística

O modelo de regressão é uma ferramenta estatística capaz de prever a probabilidade de realização de uma variável resposta (dependente) em função de uma ou mais variáveis preditoras (independentes), ou seja, a regressão estabelece correlação entre as variáveis, permitindo compreender como o conjunto de fatores sociais e linguísticos condicionam determinada variante (Levshina, 2015).

Segundo a autora, o objetivo da regressão é estimar os valores dos coeficientes de regressão que melhor se ajustam aos dados. Isso é feito usando um algoritmo de otimização que minimiza a diferença entre as probabilidades previstas pelo modelo e as probabilidades reais observadas nos dados. Uma vez que os coeficientes de regressão são estimados, eles podem ser usados para prever a probabilidade de um evento ocorrer em função de novos fatores das variáveis predictoras. Os modelos de regressão podem ser lineares, logísticos ou multinomiais, a depender da natureza da variável resposta.

O *modelo de regressão linear* é um modelo estatístico usado para prever o valor da variável resposta com base nos valores das variáveis predictoras, pressupondo que a relação entre as variáveis pode ser representada por uma equação linear, ou seja, por uma linha reta quando retratada em um gráfico e, para isso, a variável resposta analisada deve ser necessariamente numérica e contínua. Introduzidos por Nelder e Wedderburn em 1972 como uma extensão do Modelo linear, Os *Generalized Linear Models (GLM)* são uma classe de modelos estatísticos

que generalizam a regressão linear tradicional para acomodar diferentes tipos de variáveis de resposta, como variáveis resposta nominais, através de um modelo logístico.

Já o *modelo de regressão logística* é usado para prever o valor da variável resposta quando esta for nominal, categórica, e binária, com dois níveis (Oushiro, 2017), como o caso da alternância entre os pronomes *nós* e *a gente*. Por fim, o *modelo multinomial* é adequado quando a variável resposta for de natureza nominal, e possuir três ou mais níveis, ou seja, resultados possíveis que não podem ser ordenados, como a concordância verbal com os pronomes de primeira pessoa do plural, que possui quatro níveis (*nós* com *-mos*, *nós* sem *-mos*, *a gente* sem *-mos* e *a gente* com *-mos*). A regressão logística é frequentemente usada em análises de dados linguísticos, por dar conta da interação entre as diversas variáveis na explicação do fenômeno variável.

Devido à natureza dos dados, nos deteremos mais detalhadamente ao *modelo de regressão logística*. A equação da regressão logística binária é expressa em termos do *logit* da probabilidade. O *logit* é a transformação logarítmica da razão de chances, ou seja, é a probabilidade de um evento ocorrer dividida pela probabilidade de não ocorrer, gerando assim um valor de estimativa para realização da regra variável, podendo expressar favorecimento ou desfavorecimento. Gries (2013, p. 259) apresenta a equação da regressão logística binária:

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

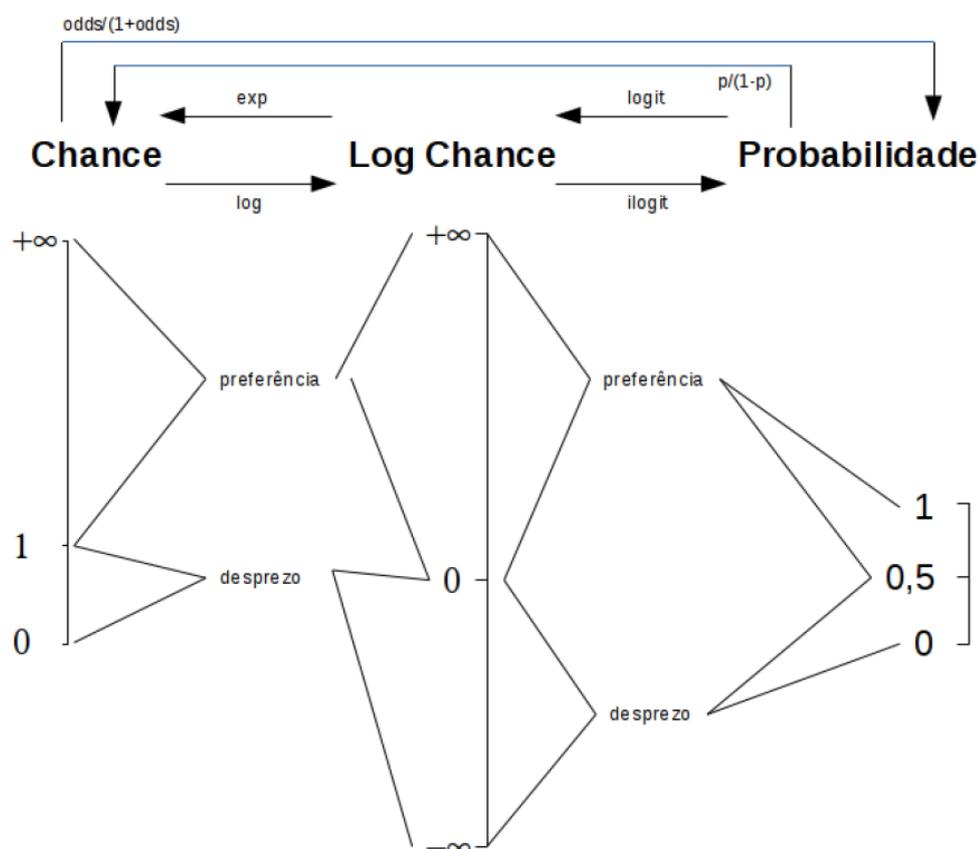
Na fórmula,  $P$  é a probabilidade de realização da variável analisada;  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$  são os coeficientes de regressão, ou seja, as estimativas da mudança média na variável resposta para uma unidade de mudança na variável preditora, para prever a presença ou ausência de uma característica ou resultado com base em valores de um conjunto de variáveis preditoras em interação;  $x_1, x_2, x_3, \dots, x_k$  são as estimativas das variáveis preditoras; e  $\log$  é o logaritmo natural, usado para transformar a probabilidade em uma escala linear e assim obter os valores previstos para a aplicação da regra variável.

Neste tipo de modelo, as estimativas de realização da variável resposta em função de cada variável previsoras são apresentadas em *logodds*, assim como o valor de intercepto (*intercept*). Como vimos, os resultados em *logodds* podem assumir valores positivos ou negativos, a depender do nível de favorecimento da aplicação da regra (Gries, 2013). Para Oushiro (2017), a análise de dados linguísticos por *logodds* permite identificar padrões e

tendências do comportamento linguístico em grandes conjuntos de dados linguísticos, como banco de dados de língua falada ou escrita.

A principal vantagem da análise de dados linguísticos a partir de *logodds* é que ela permite identificar esses padrões, que seriam difíceis de detectar por meio de outras técnicas de análise de dados. Além disso, é uma unidade de fácil interpretação, tendo em vista que sua escala vai do menos infinito ( $-\infty$ ) até o mais infinito ( $+\infty$ ), tendo zero como ponto de neutralidade (Oushiro, 2017), diferentemente da escala de probabilidade, que é uma escala que vai de 0 a 1 e tem em 0.5 o seu ponto neutro. Apesar de serem escalas de medida distintas, ambas apresentam a mesma informação em relação à análise dos dados, sendo possível determinar os valores correspondentes dos *logodds* em probabilidade, conforme demonstra figura proposta por Gries (2013, p.265).

**Figura 6:** Correspondência entre os valores em *odds*, *logodds* e probabilidade



Fonte: Gries (2013, p. 265)

Gries (2013) explica que essas transformações são úteis pois ajudam a expressar os valores em probabilidade em uma escala contínua entre 0 e 1, que são muitas vezes mais interpretáveis em contextos estatísticos. No entanto, o autor explica que a escolha da unidade

de valor empregada para análise dos dados depende da preferência do pesquisador, da ferramenta estatística e da natureza da análise, tendo em vista que ambas as medidas carregam a mesma informação. Assim, para transformar *odds* (chances) em *logodds* (logaritmo da chance), é necessário usar função logarítmica *logit* da probabilidade, por meio da seguinte formulação:

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right)$$

É possível também transformar as medidas de *logodds* em probabilidade através da função logarítmica *ilogit*. A função *ilogit* é a função inversa do *logit* comumente empregada na conversão de valores de *logodds* em probabilidade. Oushiro (2017, p. 182) apresenta a fórmula para calcular o *ilogit* e se chegar ao valor de probabilidade:

$$\text{ilogit}(\text{logodds}) = \frac{e^{\text{logodds}}}{1 + e^{\text{logodds}}}$$

Na fórmula, *logodds* é o valor de *logodds* que você deseja converter em probabilidade, *e* é a constante matemática conhecida como número de Euler, que é aproximadamente 2,71828. Em trabalhos sociolinguísticos, o número de Euler é usado para modelar a distribuição de características linguísticas em uma comunidade de fala e, com isso, entender como as variantes linguísticas se espalham e se encaixam ao longo do tempo em uma comunidade.

É possível ainda verificar a correspondência dos valores de *logodds* em probabilidade, ou seja, a equivalência dos valores apresentados em *logodds* em pesos relativos, sem a necessidade de cálculos através de equações matemáticas. Natalia Levshina, na obra o livro "How to do Linguistics with R", apresenta uma tabela de correspondência entre os valores de probabilidade, *odds* e *logodds*, bem como o ponto de neutralidade em cada escala, instrumento facilitador para auxiliar na análise e modelagem dos dados linguísticos, uma vez que, dependendo do propósito do pesquisador, é possível converter os valores das estimativas de uso em pesos relativos.

**Tabela 3:** Valores correspondentes entre probabilidades, *odds* e *logodds*

Probability	Odds	Logit (log odds)
0.001	0.001	-6.91
0.01	0.01	-4.6
0.05	0.05	-2.94
0.1	0.11	-2.2
0.25	0.33	-1.1
<b>0.5</b>	<b>1</b>	<b>0</b>
0.75	3	1.1
0.9	9	2.2
0.95	19	2.94
0.99	99	4.6
0.999	999	6.91

Fonte: Levshina (2015, p. 265)

Em estudos sociolinguísticos, a regressão logística é aplicada quando se deseja investigar a influência de determinados fatores sociais - como idade, gênero, etnia, classe social, nível educacional – e linguísticos, sobre a ocorrência ou não da variante linguística. O modelo de regressão logística é geralmente aplicado para modelar a probabilidade de ocorrência da variante linguística analisada em função das variáveis explicativas (Miranda et al., 2017).

Através do modelo de regressão logística, é possível estimar os coeficientes de cada variável previsor, gerados em *logodds*, e verificar o seu impacto na probabilidade de ocorrência da variante linguística. Para realizar essa análise, esse tipo de modelo oferece ao pesquisador uma visão dos dados com máxima verossimilhança para estimar os parâmetros linguísticos. Com isso, a análise de regressão logística permite obter estimativas dos coeficientes das variáveis explicativas e, com base nessas estimativas, interpretar como cada variável influencia a probabilidade de ocorrência da variante linguística (Picanço, 2007).

Levshina (2015) afirma que a interpretação de um modelo de regressão logística – seja em *logodds* ou em probabilidade – envolve a análise dos coeficientes estimados para cada variável previsor e a interpretação dos resultados em termos de probabilidade de ocorrência do evento binário, observando os *insights* sobre características sociais e linguísticas de uma comunidade. No entanto, antes de observar os coeficientes, é necessário verificar se os dados são estatisticamente significativos. Isso pode ser feito verificando se o valor *p* (*p-value*) associado a cada variável é menor que um nível de significância (geralmente arbitrado em 0,05).

Coefficientes com valores  $p$  maiores que o nível de significância não são considerados estatisticamente relevantes e podem ser ignorados.

Desde a década de 1990, linguistas têm usado o auxílio de ferramentas estatísticas para realizar a análise de regressão logística, especialmente com o desenvolvimento do programa VARBRUL, também conhecido como *Variable Rule Analyses*, um modelo logístico de análise de dados que foi proposto inicialmente por David Sankoff em 1978 e, posteriormente, com a criação do GoldVarb X, por David Sankoff, Sali A. Tagliamonte e Eric Smith em 2005.

As pesquisas sociolinguísticas atuais estão adotando linguagens de programação para o tratamento dos dados da variação, tendo disponíveis várias ferramentas estatísticas que podem ser usadas para elaborar e ajustar um modelo de regressão logística, incluindo o Python, o SPSS e a linguagem de programação *R*. Neste estudo, para o tratamento dos dados, adotamos a Linguagem de programação *R*, por meio do ambiente *R Studio*, devido à sua interface amigável, recursos para visualização de dados, gráficos e resultados de análises de forma interativa e uma ampla gama de ferramentas para análise de dados, o que facilita a escrita, execução e depuração dos dados (Oushiro, 2015).

#### **4.7.3 A linguagem de programação *R* e a interface *R Studio***

*R* é uma linguagem de programação e um ambiente de desenvolvimento frequentemente utilizado por pesquisadores para análise estatística, visualização de dados e modelagem estatística, distribuído sob a licença GNU GPL (Levshina, 2015). Seu desenvolvimento ocorreu na década de 1990, na Nova Zelândia, a partir das pesquisas de Ross Ihaka e Robert Gentleman na Universidade de Auckland, com objetivo de proporcionar uma alternativa à *linguagem S*, sendo um software livre e gratuito, mantido por uma comunidade global de desenvolvedores e usuários que contribuem com pacotes e recursos para aprimorar suas funções.

Por esse motivo, a linguagem *R* é referência entre as ferramentas estatísticas, dada a sua extensa coleção de bibliotecas e pacotes, que oferece grande variedade de funções estatísticas e ferramentas para tratamento e visualização dos dados, sendo útil desde o gerenciamento dos dados até a análise estatística e apresentação dos resultados. As bibliotecas, por exemplo, possibilitam que os usuários do *R* procedam todos os tipos de análise estatística, desde a descrição de um fenômeno em termos de proporção e frequência, a análises mais complexas, como o desenvolvimento de um modelo de regressão logística (Sen, 2021).

A *linguagem R* oferece ainda poderosas ferramentas para desenvolvimento de gráficos e visualizações dos resultados, por meio de seus pacotes, como o *ggplot*, sendo possível a representação visual dos dados com eficácia. Esses pacotes permitem a criação de gráficos de barras, gráfico de linha, histogramas, gráficos de dispersão. Além disso, é possível gerar tabelas e realizar testes estatísticos, o que facilita na interpretação dos resultados obtidos nas análises estatísticas. (Gries, 2013; Everitt e Hothorn, 2014; Levshina, 2015).

Integrado a *linguagem R*, o *software RStudio* é um ambiente de desenvolvimento integrado (IDE), responsável por fornecer uma interface gráfica do usuário (GUI) para o *R*, o que na prática torna mais fácil e eficiente a escrita de código, a execução de análises de dados e a criação de visualizações (Oushiro, 2015). Devido a sua interface amigável e a recursos estatísticos, o *RStudio* permite ao investigador linguista o processamento e a análise de dados linguísticos de forma mais precisa. Além disso, o *software* também auxilia na organização e no gerenciamento dos dados linguísticos, permitindo a importação e exportação de dados em diferentes formatos, como planilhas do Excel (.xlsx), *comma-separated values* (.csv), arquivos XML (.xlm), arquivo de texto simples (.txt), arquivo de dados do *R* (.Rdata), entre outros.

Estudos sociolinguísticos, especialmente na última década, têm recorrido ao ambiente *RStudio* para identificar os padrões e tendências linguísticas em diferentes comunidades de fala, através de modelos estatísticos desenvolvidos com o objetivo de investigar como fatores sociais e linguísticos atuam no favorecimento de determinada variante, proporcionando maior confiabilidade aos resultados obtidos nas análises. Esses modelos são desenvolvidos com o auxílio de pacotes, e executados por comandos.

No *RStudio*, os comandos são frequentemente executados por meio de *scripts*. Um *script* é um arquivo de texto que contém uma sequência de comandos que podem ser executados em ordem. É através de um *script* que podemos elaborar um modelo de regressão logística para análise de uma variável. Para ilustrar, recorreremos ao estudo de Freitas, Rodrigues e Santos (2022), que, em estudo sobre a variação entre os pronomes *nós* e *a gente* com informantes de nível superior em Fortaleza, elaboraram, através do pacote *glm*<sup>11</sup>, modelo incluindo as variáveis *grau de referência do pronome*, *tempo verbal* e *tipo de verbo*, a fim de testar a interação dessas variáveis em função da variável resposta. O modelo elaborado pelas autoras tem a seguinte estrutura:

---

<sup>11</sup> O *glm* (generalized linear model) é um modelo estatístico que generaliza a regressão linear para acomodar diferentes tipos de variáveis respostas, como variáveis do tipo nominal categórica, para possibilitar a regressão logística.

(glm = VD ~ referencia + tempoverbal + tipodeverbo)

A execução do *script* acima gerou como *output* o resultado do modelo, com dados estatísticos referente a atuação de cada variável previsoras incluída no modelo. De acordo com Oushiro (2022, p. 271), “como se trata de um modelo pouco usual nos estudos sociolinguísticos brasileiro”, a análise dos resultados da regressão logística obtidos por meio do *R Studio* pode ser um desafio para o linguista e, por isso, é necessário conhecer bem cada informação presente no modelo. Vejamos os resultados obtidos por Freitas, Rodrigues e Santos (2022):

**Tabela 4:** Resultados do modelo de regressão em Freitas, Rodrigues e Santos (2022)

<b>Coefficients</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>	
<b>(Intercept)</b>	3.71898	0.88607	4.197	4.197	
<b>Referência</b>					
Específica	-0.95380	0.25486	-3.742	0.000182	***
<b>Tempo verbal</b>					
Presente dif.	-1.27102	0.37873	-3.356	0.000791	***
Pretérito igual	-1.72369	0.37727	-4.569	4.90e-06	***
Imperfeito	0.20587	0.39699	0.519	0.604051	
Pretérito dif.	-1.39641	0.39255	-3.557	0.000375	***
<b>Tipo de verbo</b>					
Ter	-1.23886	0.82341	-1.505	0.132442	
Ação	-0.42252	0.79315	-0.533	0.594239	
Epistêmico	-0.83888	0.90888	-0.923	0.356017	
Estado	1.51077	0.82031	1.842	0.065516	
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
<b>glm = VD ~ referencia + tempoverbal + tipodeverbo</b>					

Fonte: Adaptado de Freitas, Rodrigues e Santos (2022, p. 173)

Como é possível observar na tabela 4, as informações presentes no resultado do modelo são apresentadas em colunas, sendo cada uma delas parâmetros para a interpretação da regressão logística. Oushiro (2022) explica que a primeira coluna, *Coefficients*, apresenta os níveis das variáveis previsoras incluídas no modelo. Cada uma delas tem um coeficiente associado que indica o efeito das variáveis previsoras sobre as chances de realização da variante de referência. A segunda coluna apresenta os valores das estimativas – em termos de *logodds*,

que é a transformação logarítmica da razão de chances (*odds ratio*) – que representam o efeito das variáveis predictoras sobre a probabilidade de ocorrência da variante de referência.

A terceira coluna exibe os valores do erro padrão, ou *standard error*, uma medida que indica a precisão dos valores estimados para os coeficientes, ou seja, aponta o quanto os valores das estimativas podem variar a depender da amostra e da aleatoriedade dos dados. Dessa forma, quanto menor for o erro padrão, mais precisa e confiável será a representação das estimativas para cada uma das variáveis em análise. Na coluna seguinte, o *valor z* representa uma medida usada para avaliar a significância das estimativas, indicando o quão distante de zero estão seus valores. O *valor z* é calculado pela razão entre a estimativa dos coeficientes e o seu erro padrão.

A quinta coluna demonstra a medida estatística tradicional para avaliar a significância dos coeficientes já explanada na seção anterior (ver 3.7.2). O valor *p* é importante pois possibilita testar a hipótese nula, ao observar se os coeficientes são significativamente diferentes de zero. Como nos estudos sociolinguísticos recentes, adotamos como nível de significância o valor  $p > 0,05$ , o que implica dizer que se o valor *p* for menor que 0,05, a hipótese nula pode ser rejeitada e a hipótese alternativa deve ser acatada, pois o coeficiente é significativamente distante de zero, indicando um efeito estatisticamente relevante sobre a aplicação da regra.

Além das informações apresentadas nas colunas do modelo de regressão (ver tabela 4), é de fundamental importância compreender o valor de *intercept* e o que ele indica. O *intercept* é o valor estimado, dado em *logodds*, quando todas as variáveis predictoras do modelo são iguais a zero, ou seja, quando não estão exercendo efeito sobre o fenômeno estudado. O valor de *intercept* é constituído pela estimativa de uso do segundo nível da variável resposta sobre o primeiro nível de cada variável predictorora.

No modelo apresentado por Freitas, Rodrigues e Santos (2022, p. 173), na tabela 4, o valor de *intercept* corresponde, então, a estimativa de uso do pronome *a gente* (segundo nível da variável resposta), com referente genérico, no presente do indicativo com mesmo paradigma relação passado/presente, e acompanhado de verbos *dicendi* (os primeiros níveis das variáveis predictororas).

Além disso, para atribuir uma estimativa a determinado fator de uma variável predictorora, é necessário que o valor da estimativa desta variável seja somado ao valor de *intercept* do modelo, pois este é o valor de referência do modelo. Por exemplo, na tabela 4, para se chegar ao valor da estimativa de uso de *a gente* com referente específico, é necessário somar o valor do coeficiente *específica* (-0.95380) ao valor do *intercept* (3.71898). O resultado dessa

soma é de 2,76518, indicando que, em relação ao uso no sentido genérico (*intercept*), as realizações com referência específica tendem a favorecer menos o uso do pronome inovador, já que seu valor é menor que o do *intercept*. Oushiro (2022) alerta para a importância de se verificar as estimativas dos coeficientes em relação ao *intercept*:

A apresentação de resultados na forma de diferenças em relação ao *intercept* em vez dos valores estimados em si permite verificar, rapidamente, o quanto tal valor difere de zero: se a estimativa for nula (ou próxima dela), isso significa que é praticamente idêntica à do *intercept*, ou, visto de outro modo, que provavelmente não há diferença significativa entre os níveis da variável (Oushiro, 2022, p. 82).

Por fim, os fatores apontados como significativos estão indicados por asteriscos na última coluna da tabela. Essa informação é importante, pois nos permite ter um panorama dos fatores considerados estatisticamente significativos, bem como o nível de significância de cada um deles. No final da tabela 4, é possível observar os códigos de significância, sendo ‘\*\*\*\*’ correspondente a 0, representando alta significância estatística; ‘\*\*\*’ correspondendo a 0.001; ‘\*\*’ equivalente a 0.01, ainda apontando relevância estatística; ‘.’ representa 0.05, o limite do valor alfa de 5%; e ‘ ’ equivale a 1, não demonstrando qualquer interação sobre a realização da variável resposta, considerado estatisticamente irrelevante para o modelo.

Feitas algumas considerações sobre a interpretação dos dados e o uso da ferramenta *RStudio*, a próxima seção tratará da descrição dos dados e das análises dos resultados, por meio de testes estatísticos, como o qui-quadrado, e de modelos de regressão logística adequados para o tratamento do fenômeno variável.

## 5 DESCRIÇÃO DOS DADOS E ANÁLISE DOS RESULTADOS: O FENÔMENO DA ALTERNÂNCIA PRONOMINAL

Neste capítulo, apresentamos os resultados referentes aos testes estatísticos e ao modelo de regressão logística desenvolvido para o tratamento da regra variável. Por se tratar de uma pesquisa em tempo real, analisamos dados de duas sincronias da mesma comunidade de fala, com amostras provenientes do projeto PORCUFORT - Amostra 1990 e Amostra 2020. Durante a análise, tratamos de resgatar, sempre que possível, as hipóteses iniciais levantadas sobre a atuação de cada variável previsorora sobre a realização da variável resposta deste estudo, em cada uma das amostras da fala culta fortalezense.

Com o intuito de tornar mais didática e organizada a apresentação dos resultados, optamos por dividir este capítulo em algumas seções e subseções. Inicialmente, na seção 5.1, apresentamos uma análise dos padrões de alternância pronominal entre *nós* e *a gente* na amostra da década de 1990. Em seguida, na seção 5.2 tratamos de analisar a alternância pronominal de primeira pessoa do plural em amostra da década de 2020.

Ao analisar a alternância entre os pronomes *nós* e *a gente*, primeiramente tratamos, na subseção 5.1.1, da atuação individual de cada preditor sobre a realização dos pronomes - ou seja, a distribuição das frequências e proporções de uso das variantes em função de cada variável previsorora, bem como o resultado dos testes de qui-quadrado -, a fim de responder a uma das questões que deram norte a essa pesquisa: quais as variáveis previsororas, de ordem linguística e social, que favorecem e condicionam o uso dos pronomes *nós* e *a gente* no falar culto de Fortaleza na década de 1990?

Em um segundo momento, na subseção 5.2.1, discutimos os resultados da análise com dados da década de 2020, testando as mesmas variáveis previsororas, a fim de identificar tendências que se repetem no comportamento da variação entre os pronomes pessoais de primeira pessoa do plural. Nesta subseção, buscamos analisar os resultados dos testes estatísticos e do modelo de regressão, com o objetivo de responder a seguinte questão norteadora: quais e como as variáveis linguísticas e sociais motivam a realização dos pronomes de primeira pessoa do plural na fala culta de Fortaleza na Amostra 2020?

Por fim, na seção 5.3, procedemos uma análise comparativa em tempo real de curta duração do tipo tendência com os resultados das análises com as amostras das duas décadas (1990 e 2020), buscando indícios de mudança linguística e de expansão da variante inovadora nos contextos controlados neste estudo, além de averiguar as características convergentes e divergentes no comportamento do fenômeno nas duas décadas analisadas, como os contextos

mais produtivos para a realização dos pronomes, as variáveis que deixaram de ter efeito significativo sobre o uso das variantes e aquelas que passaram a atuar de forma relevante na alternância entre os pronomes. Neste ponto, nos guiamos pelas seguintes questões de pesquisa: a proporção de uso do pronome inovador *a gente* aumentou ou diminuiu no decorrer de duas décadas na fala culta de Fortaleza? Existem indícios de mudança linguística em progresso ou o fenômeno figura como um caso de variação estável?

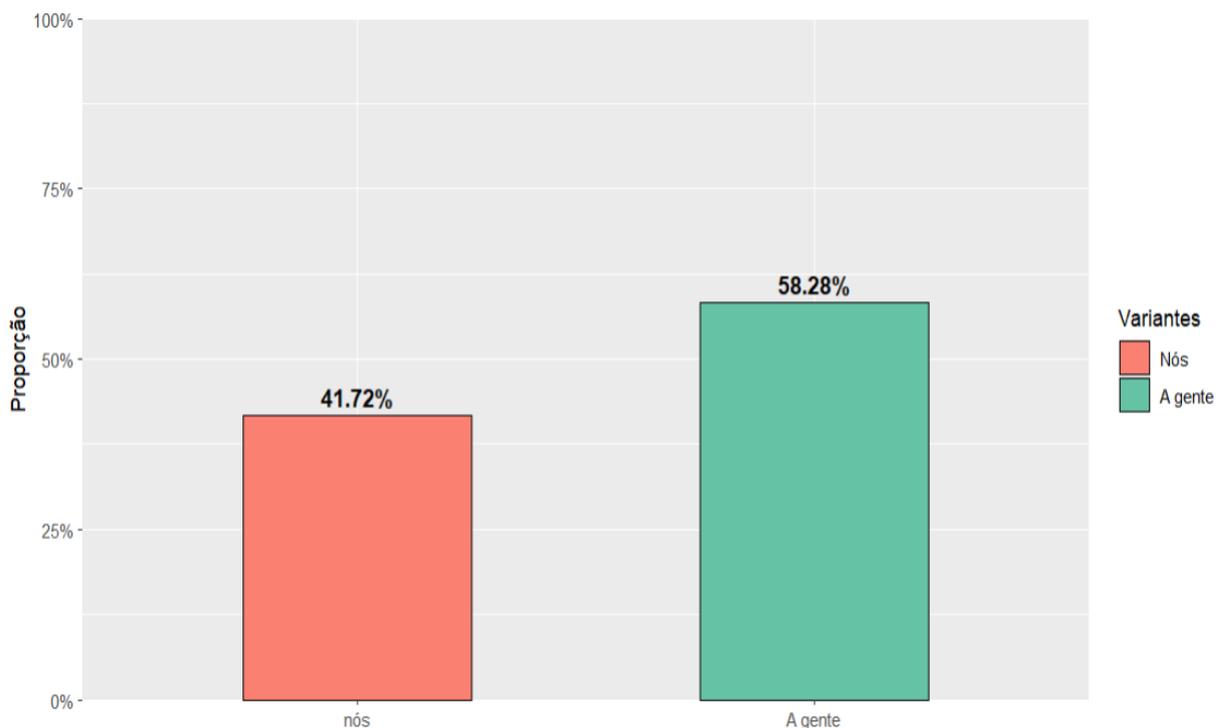
Tendo apresentado cada uma das subseções, passamos para a descrição dos resultados dos testes estatísticos realizados com os dados de fala coletados, e para a análise multivariada, através do modelo de regressão logística. Os resultados obtidos nesta análise serão discutidos à luz da teoria variacionista e comparados, sempre que possível, a outros trabalhos que tratam da alternância entre os pronomes *nós* e *a gente*.

### **5.1 A alternância pronominal de primeira pessoa do plural no falar culto de fortaleza: análise descritiva dos dados de fala da década de 1990**

Para o processamento dos nossos dados linguísticos, foi adotada a linguagem de programação *R* (R Core Team, 2020), por meio da interface *RStudio*, em que foi possível separar, quantificar e testar a significância dos efeitos das variáveis previsoras sobre nossa variável resposta (*nós* x *a gente*), bem como plotar gráficos de barras e de linhas, além de tabelas com a apresentação dos resultados. Neste sentido, o *RStudio* nos forneceu frequências e proporções, assim como resultados de modelo de regressão logística, referentes à atuação de todos os fatores que elencamos como possíveis condicionadores da variação entre *nós* e *a gente* no falar culto de Fortaleza, possibilitando uma análise estatística dos dados coletados em nossa amostra.

Ressaltamos que, ao todo, em amostra representativa da década de 1990, trabalhamos com um total de 1957 observações de primeira pessoa do plural, sendo 1140 ocorrências do pronome inovador *a gente*, e 817 dados do pronome canônico *nós*. O gráfico 5 nos revela as proporções de uso em cada uma das variantes na amostra de fala culta de Fortaleza da década de 1990:

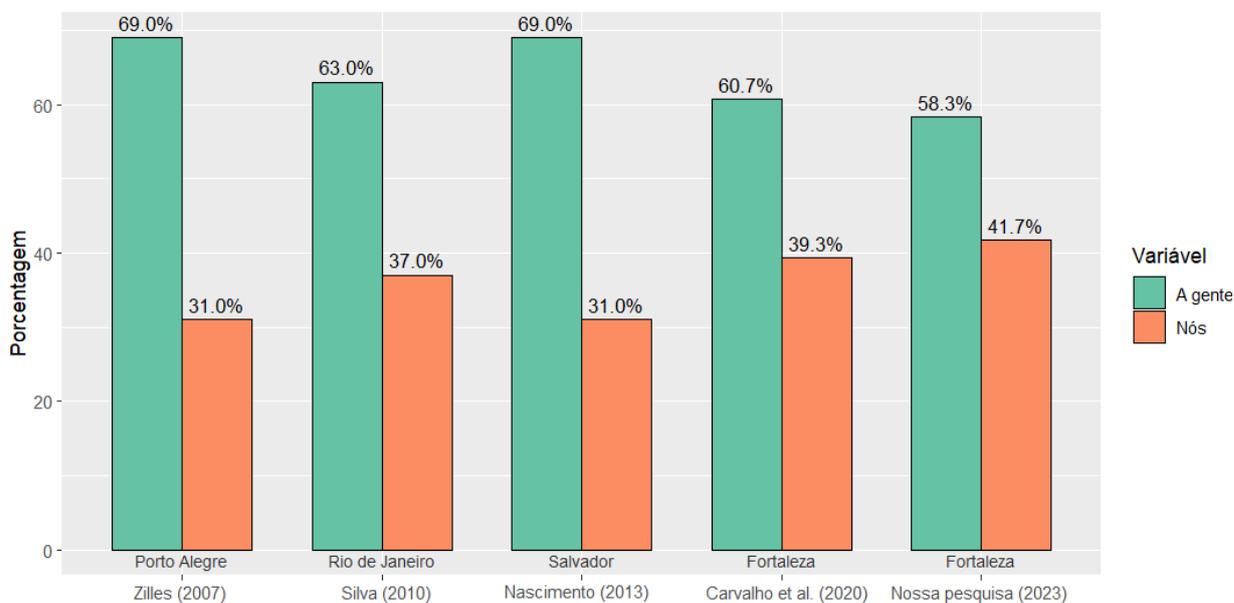
**Gráfico 5:** Proporção de uso das variantes *nós* e *a gente* na amostra da década de 1990



Fonte: elaboração própria

Conforme podemos observar, os dados do gráfico 5 mostram que, de um modo geral, na fala culta de Fortaleza dos anos 1990, a variante tida como inovadora é mais utilizada entre os informantes, com 58,28% das observações, ao passo que a variante padrão apresenta menor proporção de uso, sendo 41,72% das ocorrências. Esse resultado inicial confirma nossa hipótese sobre o comportamento das formas variantes de primeira pessoa do plural na década de 1990, pois esperávamos que o pronome *a gente* fosse usado com mais frequência que o pronome *nós*, assim como ocorre em diversos estudos sobre o tema em *corpora* de fala culta e de fala popular (Lopes, 1999; Oliveira, 2006; Brustolin, 2009; Tavares, 2014; Araujo, 2016; Fernandes, 2021).

Pesquisas variacionistas sobre os padrões de alternância pronominal de primeira pessoa do plural, em diferentes partes do Brasil, indicam uma tendência no comportamento das variantes em competição na década de 1990, no sentido de que, em amostras de fala culta, assim como na fala popular, há maior preferência ao uso da variante inovadora em diversas regiões do Brasil (Araujo, 2021). Essa tendência se torna evidente ao compararmos nosso resultado com os de trabalhos sobre a variação entre os pronomes *nós* e *a gente*, também em amostras de fala culta representativas da década de 1990 de diversas localidades, como em Zilles (2007), Silva (2010), Nascimento (2013), Carvalho et al. (2020):

**Gráfico 6:** Comparação de nossos resultados com outras pesquisas com amostras de fala culta

Fonte: Elaboração própria

O que se percebe no gráfico 6 é uma tendência no que concerne ao uso das variantes em diversas comunidades de fala no Brasil na década de 1990: o levantamento realizado demonstra que, em todas as pesquisas, os informantes de nível superior apresentam preferência para o pronome inovador *a gente*. Zilles (2007) identificou, em amostras no projeto NURC - Porto Alegre, maior uso do pronome inovador (69%) e menor proporção de uso do pronome canônico (31%). Cabe pontuar, inicialmente, que, embora as pesquisas analisem amostras representativas da fala de informantes com nível superior completo, foram incluídos, no entanto, na amostra de Porto Alegre da década de 1990, “quatro falantes com educação secundária, supondo não haver diferença significativa entre eles, com base nos resultados do estudo de tempo aparente” (Zilles, 2007, p.35). Segundo o autor, a inclusão desses informantes não acarretaria prejuízo na análise dos dados, já que não haveria, aparentemente, diferença expressiva no uso dos pronomes entre os informantes.

Seguindo essa tendência, Silva (2010), em amostras do NURC - Rio de Janeiro, identificou maior ocorrência de *a gente* (63%) em detrimento do *nós* (37%). De maneira parecida, com dados de fala do projeto NURC Salvador, Nascimento (2013) verificou maior incidência de *a gente* (69%) e menor uso de *nós* (31%), apontando para um padrão linguístico no falar culto de diversas regiões do país na década de 1990: maior uso da forma pronominal *a gente* em detrimento do pronome *nós*. Nesse sentido, o trabalho de Carvalho et al. (2020) constatou a mesma tendência apontada pelos autores. Com base na análise de dados de fala

culta de Fortaleza, houve maior proporção de uso do pronome *a gente* (60,7%), diferente do pronome *nós* (39,3%), que apresentou menor ocorrência.

Por fim, nossa pesquisa sobre o falar culto de Fortaleza reafirma a tendência que vem sendo apontada no português brasileiro: maior proporção de uso do pronome *a gente* (58,3%), e uso menos recorrente do pronome conservador *nós* (41,7%). Com esse resultado, entendemos que o comportamento da variação entre os pronomes *nós* e *a gente*, na fala culta de Fortaleza, se caracteriza como uma regra variável, nos termos de Labov (1969).

Uma regra variável, para Labov (1969), é um fenômeno linguístico variável que apresenta formas em competição em determinada comunidade de fala, tendo proporções de uso das variantes maiores que 5% ou menores que 95%. Em outras palavras, as proporções de uso dos pronomes revelam que, na amostra do PORCUFORT da década de 1990, a variação entre as formas pronominais *a gente* e *nós* se configura como um caso de regra variável, com percentuais de 58,28% e 41,72%, estando, portanto, dentro do intervalo estabelecido por Labov. Neste caso, as duas formas pronominais coexistem dentro da comunidade de fala, disputando espaço como referência de primeira pessoa do plural, sendo intercambiáveis, assim como nos exemplos (116) e (117):

(116) Inf.: tô me atualizando muito CEdo sabe? pa-pai?... e muitas vezes ***a gente*** num precisa NEM comprar um jornal porque sai a notícias toda no rádio numa veLOCidade TÃO GRANDe (Inq. 11 – D2, homem, 33 anos).

(117) Inf.: ***nós*** ouvíamos aquela ladainha TO::da (menina)... de um... e de outra parte e tínhamos que:: procurar apaziguar uma e outra parte... (Inq. 23 – DID, mulher, 61 anos)

Nos termos de Labov (1969), a realização da primeira pessoa do plural apresenta duas formas variantes distintas que podem ser empregadas no mesmo contexto, sem prejuízo em seu significado ou em sua função. Nesse sentido, tanto o pronome *nós* quanto o pronome *a gente* representam o mesmo referente nas pesquisas sobre o tema. Sendo assim, pode-se dizer que o estágio de mudança iniciado no século XVIII (Lopes, 2003), ainda não se completou, apresentando as duas formas em competição ainda acirrada na década de 1990 na comunidade de fala de Fortaleza, que podem variar de acordo com pressões de ordem linguística e social. Essas pressões podem explicar a escolha entre uma das variantes em detrimento da outra e o comportamento da variação. Vale destacar que regra variável é um dos principais conceitos epistemológicos difundidos pela Teoria da variação e mudança linguística, sendo base para o desenvolvimento de estudos variacionistas.

Tendo, então, apresentado e discutido os termos percentuais das variantes obtidos a partir da primeira análise realizada através do ambiente *RStudio*, passamos a apresentação e descrição dos testes estatísticos realizados para averiguar a interação entre as variáveis predictoras e a variável resposta, a realização dos pronomes *nós/a gente*. Dessa forma, a medida em que fomos apresentando os resultados, em tabelas, discutiremos os dados em termos de estatística descritiva básica. Mais à frente, na subseção 5.1.2, apresentaremos os dados em termos de estatística inferencial, por meio de um modelo de regressão logística.

### 5.1.1 Atuação das variáveis predictoras em termos de frequência e proporção de uso: a alternância pronominal de primeira pessoa do plural na Amostra 1990

Com o objetivo de investigar os contextos sociais e linguísticos que atuam sobre a realização dos pronomes de primeira pessoa do plural, aplicamos testes de qui-quadrado a cada variável predictoras deste estudo. O teste de qui-quadrado de Pearson ajuda a inferir se existe algum tipo de associação entre a variável predictoras e a variável resposta e, para isso, o teste “compara valores observados com valores esperados de acordo com a hipótese nula” (Oushiro, 2021, p. 192).

Para Gries (2013), a diferença entre os valores esperados, que são estimados estatisticamente por meio da *linguagem R*, e os valores observados, que são os resultados reais dos dados de fala, gera um valor residual. É a partir da soma desses valores residuais que se dá o valor do  $\chi^2$ . Com isso, quanto mais distante do 0 (zero) for o resultado do teste, maior a probabilidade da variável predictoras exercer influência sobre a realização da variável resposta.

Os resultados com as proporções de uso de cada variável predictoras na amostra do PORCUFORT com dados de fala representativos da década de 1990, bem como os valores de qui-quadrado, estão dispostos a seguir em forma de tabelas. A variável *sexo/gênero* foi a primeira elencada por nós, conforme podemos observar na tabela 5:

**Tabela 5:** Distribuição de *nós* e *a gente* por *sexo/gênero*: Amostra 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
<i>Sexo/Gênero</i>	<i>Nós</i>	%	<i>A gente</i>	%
Homem	488	44,1%	619	55,9%
Mulher	329	38,7%	521	61,3%
$\chi^2 (1) = 5.6466, p = 0.01749$				

Fonte: elaboração própria.

As informações apresentadas na tabela 5 sugerem uma diferença significativa na utilização desses pronomes de acordo com o *sexo/gênero* do informante. Esta diferença é estatisticamente relevante, conforme indicam os dados. Em relação ao pronome padrão, percebe-se que os homens (44,1%) tendem a fazer uso de *nós* em maior proporção se comparados com as mulheres (38,7%). No entanto, há maior preferência pelo pronome inovador em ambos os *sexos/gêneros*, sendo mais expressivo entre as mulheres (61,3%), que costumam empregar o pronome *a gente* com maior proporção, do que os homens (55,9%).

Para a variável *sexo/gênero*, o teste de qui-quadrado chegou ao resultado de  $\chi^2 = 5.6466$ , com um grau de liberdade. Graus de liberdade (1) indicam que o teste é univariado, relacionado a uma única variável categórica, e se refere a possibilidade de associação entre os fatores dessa variável categórica com a variável resposta, sendo calculado pelo produto do número de fatores menos 1 (Levshina, 2015). Por exemplo, a variável *sexo/gênero* apresenta dois fatores, homem e mulher, sendo esta variável categórica. Ao subtrair o número um da soma dos fatores (dois fatores), temos que o número de graus de liberdade na variável *sexo/gênero* é 1. Assim, devemos “entender o valor de graus de liberdade como o número de células de que você precisa, junto com os valores totais de linhas e colunas, para conseguir deduzir os demais valores” (Oushiro, 2022, p. 194).

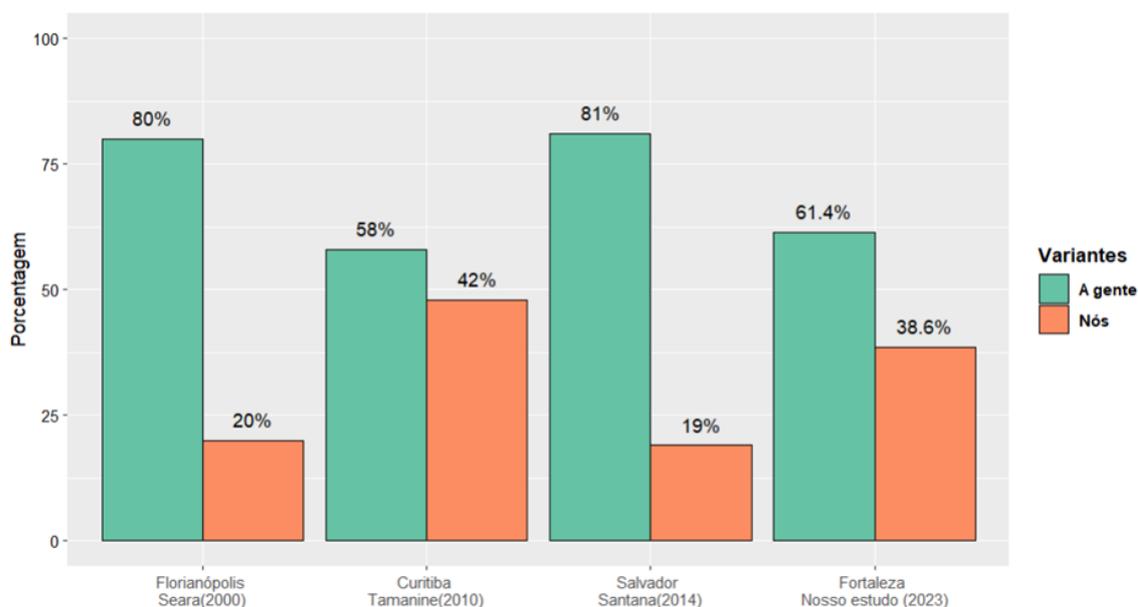
O resultado do teste de qui-quadrado apontou para uma correlação entre a realização dos pronomes de primeira pessoa do plural e o *sexo/gênero* dos informantes da Amostra 1990. O valor de  $p = 0.01749$  é menor que o nível de significância adotado nesta pesquisa, que é  $p < 0.05$ . Portanto, há evidências estatísticas para rejeitar a hipótese nula de que não há associação entre a variável *sexo/gênero* e a realização dos pronomes de primeira pessoa do plural, e acatar assim a hipótese alternativa.

A sociolinguística variacionista procura com frequência compreender de que modo aspectos sociais, como *sexo/gênero*, influenciam as variações linguísticas. Estes resultados podem sugerir que há uma variação sistemática no emprego desses pronomes associado ao *sexo/gênero* dos informantes na comunidade de fala investigada (Araujo, 2016). Estudos sobre o fenômeno variável *nós* e *a gente* confirmam a relação entre o *sexo/gênero* dos informantes e a escolha entre o uso dos pronomes, conforme demonstram Seara (2000), Tamanine (2010) e Santana (2014).

Ao compararmos os resultados do nosso estudo, que tem como base dados do falar de Fortaleza, e os resultados de trabalhos que analisaram o falar de outras capitais do Brasil, em amostras de fala representativas dos anos de 1990, é possível identificar, em termos de proporção, uma tendência no uso dos pronomes *nós* e *a gente* entre homens e mulheres em

diversas capitais do Brasil: os informantes do sexo feminino, em geral, tendem a usar mais o pronome inovador *a gente* que os homens. Para uma melhor visualização dessa tendência, o gráfico 7 apresenta a comparação no uso dos pronomes de primeira pessoa, somente entre as mulheres, em diversas capitais brasileiras.

**Gráfico 7** – Comparação no uso de *nós* e *a gente* entre mulheres em estudos variacionistas



Fonte: Elaboração própria

O gráfico 7 demonstra que as mulheres, de maneira geral, apresentam comportamento semelhante em relação ao uso dos pronomes nas diversas capitais pesquisadas, ou seja, em todas os estudos há maior preferência pela forma pronominal *a gente* entre as informantes do sexo feminino. Em Florianópolis, por exemplo, a proporção de uso da forma *a gente* chega a 80%, enquanto o pronome *nós* apresenta 20% de proporção de uso. De maneira similar, as mulheres de Salvador utilizam mais o pronome *a gente* (81%) em detrimento de *nós* (19%). Em Curitiba, embora a diferença entre o uso de *nós* e *a gente* seja mais discreto, as mulheres fazem uso do pronome *a gente* em 58% dos casos, enquanto o pronome *nós* é usado em 42% das ocorrências. Fortaleza tem comportamento semelhante a Curitiba, ou seja, a diferença entre os pronomes, embora expressiva (*a gente*: 61,4%, *nós*: 38,6%), não é tão atenuada quanto em Curitiba e Salvador.

Seguindo com a apresentação dos resultados dos testes estatísticos, a segunda variável de natureza social testada neste estudo foi a *faixa etária*. A tabela 6 apresenta as frequências e as proporções de uso dos pronomes de primeira pessoa do plural distribuídos em função das três faixas etárias dos informantes, bem como o resultado do teste de qui-quadrado:

**Tabela 6:** Distribuição de *nós* e *a gente* por faixa etária: Amostra de 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
Faixa Etária	Nós	%	A gente	%
22 a 35 anos	269	36,2%	475	63,8%
36 a 50 anos	234	33,4%	466	66,6%
51 anos ou mais	314	61,2%	199	38,8%
$\chi^2 (2) = 108.12, p < 0,001$				

Fonte: elaboração própria

Em termos de proporção, os dados da década de 1990 demonstram maior preferência pela variante *a gente* entre os informantes de faixas etárias mais jovens, com 63,8% das realizações para os informantes de 22 a 35 anos, e 66,6% para os de 36 a 50 anos. Por outro lado, os mais velhos tendem a usar mais o pronome canônico *nós*, com percentual de 61,2%. Esse resultado parcial aponta para um possível indício de mudança linguística, tendo em vista que a mudança linguística é constantemente associada a variável *faixa etária*, nos termos de Labov (2008 [1972]).

A relação entre a *faixa etária* e os aspectos de variação e mudança linguística é um dos pressupostos mais importantes da teoria variacionista. Para Labov (2008 [1972]), os mais jovens de uma comunidade de fala são os principais agentes da mudança linguística, pois são mais propensos a aderir a formas linguísticas inovadoras. Isso se dá, de acordo com a Teoria da Variação e Mudança Linguística, porque os mais jovens estão mais suscetíveis a pressões externas, como a influência da mídia e da cultura popular, o que os tornam mais receptivos às novas formas de expressão na língua. Nesse sentido, o pronome inovador *a gente* é preferível entre os mais jovens, enquanto os mais velhos tendem a manter o comportamento linguístico mais conservador, adotando o pronome mais antigo na língua.

Essa interação entre a variável *faixa etária* e a realização dos pronomes de primeira pessoa do plural foi observada no resultado do teste estatístico. O valor do teste de qui-quadrado é 109,68 com 2 graus de liberdade e valor  $p < 0,001$ , o que significa que há uma diferença estatisticamente significativa no uso de *nós* e *a gente* entre as diferentes faixas etárias. Nesse caso, a hipótese nula é a de que não existe diferenciação entre o uso dos pronomes *nós* e *a gente* e as faixas etárias testadas, ou seja, a escolha entre uma das variantes é independente da idade do informante. No entanto, tendo o valor  $p$  menor que 0,05 (nível de confiança adotado neste estudo), possuímos dados suficientes para rejeitar a hipótese nula e acatar a hipótese alternativa de que há diferenças no emprego dos pronomes entre as distintas faixas etárias.

Essa distinção de uso dos pronomes de primeira pessoa do plural entre diversas faixas etárias é constatada por estudos sociolinguísticos realizados sobre o comportamento do

fenômeno. Com resultados parecidos aos encontrados em nossa pesquisa, Silva (2014) observou, em amostra de dados de fala culta de Blumenau – SC, que a variante inovadora *a gente* é usada em maior proporção pelos informantes mais jovens, com 64%, enquanto o pronome padrão *nós* é usado por apenas 36% dos falantes dessa faixa etária. Já em relação aos mais velhos, estes apresentam comportamento inverso ao dos jovens, a saber, menor uso da forma pronominal *a gente* (31%), e uso mais frequente do pronome canônico *nós* (69%). Esses resultados apontam para resistência à mudança linguística nas faixas etárias mais elevadas, tendo em vista que os mais velhos tendem a manter uso mais tradicional e conservador da língua, nos termos da teoria laboviana (Labov, 2008 [1972]).

Além da *faixa etária*, a variação entre os pronomes *nós* e *a gente* pode ser explicada por outros fatores de ordem social, como o prestígio dado a determinada variante, a localidade, o grau de formalidade da amostra, entre outros, que podem condicionar o falante ao uso de determinada variante a depender da situação comunicativa. Labov (1972), por exemplo, propõe uma metodologia para captar a variação em vários estilos, baseado no princípio de que, quanto mais formal e monitorada for a fala, mais próxima ela será da fala padrão e, ou contrário, quanto menos formal e menos monitorada for a fala, mais propício será a realização de variantes inovadoras.

Neste sentido, analisamos três tipos de inquéritos que classificamos em uma escala que vai do menos monitorado ao mais monitorado: os inquéritos do tipo *diálogo entre dois informantes* (D2) são os que apresentam menor nível de monitoramento da fala; os inquéritos do tipo *diálogo entre informante e documentador* (DID) foram definidos com monitoramento intermediário da fala; e os inquéritos do tipo *elocuções formais* (EF) são os que apresentam maior nível de monitoramento da fala. Os resultados referentes ao tipo de inquérito estão expostos na tabela 7:

**Tabela 7:** Distribuição de *nós* e *a gente* por tipo de inquérito: Amostra 1990 (N=1956)

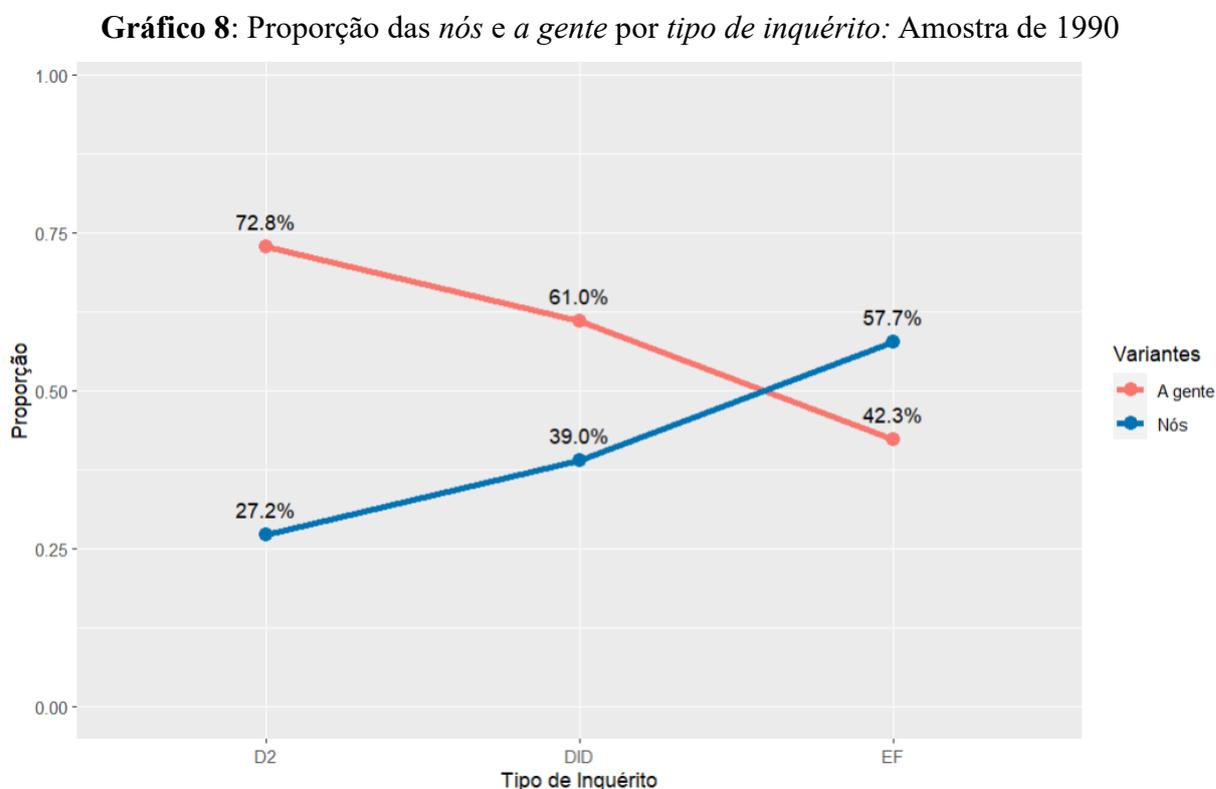
PORCUFORT I - Anos 1990 (N = 1956)				
<i>Tipo de Inquérito</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
D2	145	27,2%	386	72,8%
DID	315	39%	492	61%
EF	357	57,7%	262	42,3%
$\chi^2 (2) = 113.34, p < 0,001$				

Fonte: Elaboração própria

Confirmando nossa hipótese inicial sobre a atuação da variável *tipo de inquérito* sobre o uso dos pronomes *nós* e *a gente*, a forma inovadora é usada mais frequentemente em

inquéritos do tipo D2, que apresentam menor grau de monitoramento da fala, com percentuais de 72,8% para *a gente*, enquanto o pronome *nós* é usado por apenas 27,2% dos falantes que compõem os inquéritos dessa modalidade. Em inquéritos do tipo DID, que possui grau moderado de monitoramento (Araujo, 2011), há também maior proporção de uso do pronome *a gente*, embora em menor intensidade se comparado aos inquéritos do tipo D2, sendo usado em 61% dos casos, em detrimento do pronome *nós*, que é usado em apenas 39% das observações. Em relação aos inquéritos do tipo EF, há maior proporção de uso da forma padrão *nós* (57,7%), sendo a forma pronominal *a gente* menos recorrente nestes inquéritos (42,3%). Esses resultados também confirmam nossa hipótese sobre o uso das variantes em *elocuições formais*, pois esperávamos que os informantes dos inquéritos do tipo EF optassem pela forma mais antiga na língua.

O valor do teste estatístico de qui-quadrado de  $\chi^2 = 113.34$ , com dois graus de liberdade e valor de  $p < 0,001$ , nos permite acatar a hipótese alternativa de que há diferença significativa de uso dos pronomes *nós* e *a gente* entre os tipos de inquéritos testados neste estudo. O gráfico 8 demonstra o uso pronomes *nós* e *a gente* nos três tipos de inquéritos (D2, DID e EF):



Fonte: Elaboração própria

A partir dos dados presentes no Gráfico 8, é possível observar que a forma inovadora na comunidade pesquisada é mais frequente entre os informantes nos inquéritos do tipo D2. De maneira geral, esses informantes se conhecem e têm algum tipo de relação entre si, fato que contribuiria para menor monitoramento na fala. Na amostra, forma canônica *nós*, por outro lado, é mais usada em *elocuições formais* (EF), ou seja, em ambientes de palestras, sala de aula, ou pregações religiosas que exigem alto monitoramento da fala. Uma das explicações é que a forma inovadora, por ser avaliada como informal (Carvalho et al, 2020) emerge em situações de menor monitoramento, enquanto a forma conservadora é preferencial em ambientes de maior formalidade, o que exige maior monitoramento da língua (Coelho et al, 2015).

Eckert (2018), por exemplo, em estudo sobre a variação fonológica em diferentes contextos comunicativos, como conversa entre amigos e apresentações orais na escola, identificou que as formas inovadoras são usadas preferencialmente em estilos mais informais, enquanto a forma padrão é preferida em estilos mais formais. De acordo com a autora, isso ocorre devido a indexação das formas variantes, pois a escolha entre uma forma ou outra reflete identidades, relações sociais, questões sociais e ideológicas.

O que se percebe é que a relação entre o *tipo de inquérito* e o uso das variantes é atestada não somente pelos testes estatísticos realizados nesta pesquisa, mas também são consistentes com outras pesquisas que apresentaram resultados semelhantes. Lima (2017, p. 4), ao pesquisar os padrões de *alternância pronominal* em Luziânia – GO, observou que “em situações de monitoramento de fala os falantes optam pela variante que indica mais prestígio, *nós*; em contrapartida, os falantes preferem em sua maioria o uso da variante *a gente* em discurso menos formal e com fala sem monitoramento”. Sendo atestado a interação entre o *tipo de inquérito* e o processo de realização dos pronomes de primeira pessoa do plural, essa variável deve ser incluída no modelo de regressão logística para testar sua interação em um contexto multivariado.

Além das variáveis de ordem extralinguística observadas neste estudo, consideramos testar variáveis de ordem linguística, tendo em vista sua importância descrita no levantamento da literatura. Uma das variáveis linguísticas mais importantes em estudos sobre a variação pronominal *nós/a gente* é o *grau de referencialidade do pronome*, grupo de fator já exaustivamente analisado em diversas pesquisas sobre a variação pronominal de primeira pessoa do plural, conforme demonstram as pesquisas de Machado (1995), Tamanine (2002, 2010), Francischini (2011), Araujo (2016), Mendes (2019), Souza (2020), entre outros. Conforme tratado no capítulo de metodologia (ver subseção 4.4.1), essa variável se divide em

dois níveis: genérico e específico. Os resultados referentes à atuação desta variável podem ser observados na tabela 8.

**Tabela 8:** Distribuição de *nós* e *a gente* por grau de referencialidade: Amostra 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
Referencialidade	Nós	%	A gente	%
Genérico	339	28,7%	841	71,3%
Específico	478	61,5%	299	38,5%
$\chi^2 (1) = 205.04, p < 0,001$				

Fonte: elaboração própria.

Os dados da tabela 8 demonstram alta proporção de uso do pronome *a gente* no sentido genérico/indeterminado, apresentando-se em 71,3% das ocorrências, enquanto a forma canônica *nós* tem uso menos frequente em contextos de indeterminação, apresentando o percentual de 28,7%. Em contrapartida, em contexto de emprego da primeira pessoa com sentido específico/determinado, o pronome padrão *nós* se destaca, sendo responsável por 61,5% das ocorrências, enquanto o pronome *a gente* é menos utilizado em referência específica, apresentando 38,5% das ocorrências.

Mendonça e Nascimento (2015), quando pesquisaram as estratégias de indeterminação do sujeito, observaram que o pronome *a gente* também pode ser usado pelo falante como estratégia para indeterminar o sujeito, generalizando-o. Além disso, como discutido na seção de fundamentação teórica (ver seção 3.1), vimos que o percurso de gramaticalização do pronome *a gente* partiu, inicialmente, do substantivo genérico *gente*, que tinha um caráter generalizador, para a forma *a gente* (pronome indefinido), até que, enfim, fosse possível seu emprego como pronome pessoal de primeira pessoa do plural, concorrendo com o pronome *nós* (Lopes, 2003).

Baseado no processo de gramaticalização que o item sofreu ao longo do tempo, nossa hipótese alternativa é a de que o pronome inovador *a gente* seria favorecido quando empregado em sentido mais genérico, indeterminado, já que, de acordo com o princípio da persistência de Hopper (1991), em um processo de gramaticalização, “alguns traços do significado lexical original de um item tendem a aderir à nova forma gramatical, e detalhes de sua história lexical podem refletir-se na sua distribuição gramatical” (Hopper, 1991, p. 124). Nesse sentido, acreditávamos que o traço semântico de indeterminação estava associado ao pronome *a gente*, traço que persistiu do substantivo *gente* mesmo após a gramaticalização. Em termos de proporção de uso, nossa hipótese se confirma, uma vez que o pronome *a gente* se

mostrou fortemente associado ao traço genérico, conforme demonstra a tabela 8.

Além disso, o teste estatístico de qui-quadrado apresentou como resultado  $\chi^2 = 205.04$ , com um grau de liberdade e valor de  $p < 0,001$ , sendo significativamente diferente de zero. Com esses dados, podemos descartar a hipótese nula de que não existe influência do *grau de referencialidade* sobre a realização dos pronomes *nós* e *a gente*, e podemos acatar a hipótese alternativa, de que há interação estatisticamente relevante entre as variáveis. Com isso, os dados revelam que a variável *grau de referencialidade do pronome* apresenta forte interação com a variável resposta, motivo pelo qual deve ser incluída no modelo logístico.

A variável *tipo de verbo* foi a segunda variável previsora de ordem linguística que testamos neste estudo. Para Tamanine (2010) o *tipo de verbo* se figura como motivador de diversos contextos que podem ser favorecedores ou inibidores do uso das formas pronominais *nós* e *a gente*. Com o objetivo de analisar os contextos que favorecem os pronomes, os dados da tabela 9 demonstram, em termos percentuais, os tipos de verbo que mais propiciam o uso de um pronome em detrimento do outro.

**Tabela 9:** Distribuição de *nós* e *a gente* segundo o *tipo de verbo*: Amostra 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
<i>Tipo de Verbo</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
Ação	334	31,1%	739	68,9%
Estado	93	57,4%	69	42,6%
<i>Dicendi</i>	53	31,4%	116	68,6%
Epstêmico	37	29,4%	89	70,6%
Ter	301	70,3%	127	29,7%
$\chi^2 (4) = 224.85, p < 0,001$				

Fonte: elaboração própria

Os resultados expostos na tabela 9 apresentam as proporções de uso dos pronomes em cada tipo de verbo. De um lado, observamos tipos de verbos que são mais propensos a realização do pronome inovador *a gente*, como verbos de ação, *dicendi* e epistêmico. Do outro lado, identificamos tipos verbais que aparentam ser ambientes mais férteis a realização do pronome *nós*, como é o caso dos verbos de estado e do verbo ter, sendo este último o tipo de verbo com maior proporção de uso do pronome padrão.

Em relação aos verbos de ação, o resultado da análise estatística descritiva demonstrou maior proporção do uso do pronome *a gente*, com 68,9% das observações, em detrimento de *nós*, que apresentou 31,1% das realizações. Esse resultado vai ao encontro do obtido por Tamanine (2002), que verificou os pronomes *nós* e *a gente* em função do *tipo de verbo* e constatou que as realizações do pronome *a gente* se concentram, em proporção de uso,

nos verbos que expressam ação.

No mesmo sentido, verbos *dicendi* constituem contextos verbais com maior proporção de uso de *a gente*, com percentual de uso de 68,6%, sendo o pronome conservador *nós* menos recorrente nesses contextos (Ver tabela 9). Resultados semelhantes foram obtidos por Francischini (2011), que obteve em sua análise 78% de proporção de uso para a variante inovadora com verbos *dicendi*. Segundo o autor, a explicação dos altos índices de uso atribuídos ao pronome *a gente* pode estar relacionada ao fato de que, nos inquéritos analisados, o falante recorria com frequência ao pronome inovador como uma estratégia de generalizar o sujeito, com construções do tipo “a gente fala”, “a gente diz” ou “a gente chama”, no mais alto grau de generalização, podendo inclusive substituir o pronome pelo clítico *se*. As ilustrações (118) (119) e (120) demonstram esse tipo de emprego do pronome inovador em nossa amostra.

(118) Inf.: o PAPEL do antropólogo é muito imporTANTE... (...) Ele tratava mais do que *a gente chama* do patrimônio de pedra e cal... dos bens iMÓveis... BENS históricos (Inq. 24 – F1 – DID, homem, 61 anos)

(119) a memória... a idade... PESA *a gente diz* "AH NÃO:: ficou nela num tem esse negócio idade... a idade está no espírito está num sei em que:: mas tá também no físico está também no psíquico tem está aSSIM" (Inq. 39 – F1 – D2, mulher, 25 anos)

(120) no CAso o... E\* se... se fosse:: ele colocou ele fechou o E mas aqui *a gente não fala* muito o Ê... é porque ele quis aqui a gente só usa aberto (Inq. 26 – F1 – D2, mulher, 33 anos)

De maneira parecida, os verbos epistêmicos agem no sentido de proporcionar um ambiente mais fértil à realização do pronome inovador. Os dados descrevem alta proporção de uso da forma pronominal *a gente* com verbos epistêmicos, com 70,6% das observações, e uso de *nós* menos recorrente com esse tipo verbal, tendo 29,4% das realizações. Araujo (2018), ao pesquisar os pronomes *nós* e *a gente* na fala popular de Fortaleza, identificou 75,5% de *a gente* e 24,5% de *nós* com verbos epistêmicos, o que demonstra que o uso de *a gente* está associado a verbos que expressam atividade mental, tanto no falar culto quanto no falar popular de Fortaleza.

Em relação aos verbos que expressam estado, há uso mais frequente do pronome padrão *nós*, assim como demonstram outras pesquisas (Tamanine, 2002; Francischini, 2011; Araujo, 2016). Os dados da tabela 9 demonstram 57,4% de uso do pronome *nós*, e 42,6% de uso do pronome inovador *a gente*. Tamanine (2002), que analisou amostras de fala do interior de Santa Catarina, obteve resultados próximos aos nossos. A análise da autora constatou que o

pronome *a gente* (45%) ocorre em menor proporção quando associado a verbos de estados, sendo estes mais produtivos para a realização do pronome *nós* (55%). Na mesma direção, Francischini (2011) observou que, em contextos de uso de verbos de estado, a utilização da variante *a gente* é menor (41% das ocorrências), enquanto a variante padrão *nós* predomina nesse contexto (59% das realizações).

Além dos verbos de estado, outro contexto que parece beneficiar o pronome *nós* é o uso do verbo *ter*, apresentando 70,3% dos casos para o pronome *nós*, enquanto o pronome *a gente* figura em 29,7% das realizações de primeira pessoa. Isso demonstra que o pronome canônico *nós* é favorecido em contextos de utilização do verbo *ter*, sendo o principal tipo de verbo, em proporção de uso, a beneficiar o pronome canônico em nossa amostra. Resultado semelhante é encontrado em Tamanine (2010, pág. 131), quando a autora atesta que “verbo *ter* apresentou o maior número de ocorrências no âmbito dos verbos utilizados junto ao pronome *nós*, percentualmente sendo o verbo mais freqüente da amostra”.

Além disso, o teste qui-quadrado de Pearson evidenciou uma associação estatisticamente significativa, dentro do limite estabelecido de  $p < 0,05$  entre o *tipo de verbo* e a realização de uma das variantes ( $\chi^2 = 224.85$ ,  $df = 4$ ,  $p\text{-value} < 0,001$ ), sugerindo que a variabilidade observada não ocorre aleatoriamente, mas é motivada pelo tipo de verbo, razão pelo qual devemos incluir essa variável no modelo de análise multivariada.

Seguindo a análise dos fatores condicionantes, temos o *paralelismo formal*, que é um conceito que se refere a manutenção ou a alternância de formas linguísticas variáveis em um mesmo turno de fala, em um discurso, uma frase ou um período. Para Tamanine (2002), o paralelismo formal pode exercer distintas funções na língua, como criar ritmo, dar destaque, e até mesmo, ênfase no discurso, razão pela qual testamos a influência dessa variável sobre a realização dos pronomes pessoais de primeira pessoa do plural em nossa amostra. Os resultados estão dispostos na tabela 10:

**Tabela 10:** Distribuição de *nós* e *a gente* segundo o *paralelismo*: Amostra de 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
<i>Paralelismo formal</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
Primeira da série	193	44,2%	244	55,8%
Segundo ou terceiro da série com manutenção da forma anterior	328	40,3%	485	59,7%
Ocorrência isolada	231	39,4%	356	60,6%
Forma pronominal diferente em relação à anterior	65	54,2%	55	45,8%
$\chi^2 (3) = 10.594$ , $p = 0.01414$				

Fonte: elaboração própria

De acordo com os dados apresentados na tabela 10, podemos observar que a forma *a gente* é mais frequente nos casos de ocorrências isoladas (60,6%), de primeira ocorrência da série (55,8%) e nos casos em que o pronome se realiza como segundo ou terceiro da série com a manutenção da mesma forma anterior (59,7%) do paralelismo formal, enquanto a forma *nós* ocorre em maior proporção quando há quebra no paralelismo, ou seja, quando a realização ocorre com forma diferente em relação a imediatamente anterior na mesma série (54,2%). Esses resultados estão de acordo com os obtidos em Teixeira (2021), ao analisar dados de fala de Porto Alegre, que identificou maior proporção de uso do pronome inovador *a gente* em contexto de primeira ocorrência da série e de ocorrências isoladas.

Além do estudo de Teixeira (2021), destacamos os resultados que foram obtidos por Borges (2004), ao investigar a variação pronominal *nós* e *a gente* nas cidades de Jaguarão e Pelotas, que, em relação a variável paralelismo formal, nos resultados de sua pesquisa, assim como nos nossos, observou forte favorecimento do pronome inovador *a gente* como primeira ocorrência da série, com frequência de uso de 82% e 83% nas cidades de Jaguarão e Pelotas, respectivamente.

Além disso, conforme podemos observar nos dados expostos na tabela 10, o teste qui-quadrado de Pearson confirmou uma associação estatisticamente significativa entre a variável previsora paralelismo formal e o uso de *nós* e *a gente*, com  $\chi^2 = 10.594$ , três graus de liberdade e  $valor\ p = 0.01414$ , indicando que a variação entre as formas pronominais de primeira pessoa do plural não ocorre ao acaso, mas é influenciada por fatores como o paralelismo formal. Resultados mais apurados sobre a influência dessa variável na realização dos pronomes *nós* e *a gente* serão apresentados na análise multivariada, através do modelo de regressão logística.

Outra variável que desempenha papel importante nos estudos de cunho variacionista é o *tempo verbal*. Essa variável foi exaustivamente controlada em diversos estudos sobre a variação pronominal *nós* e *a gente*, conforme demonstram as pesquisas de Machado (1995), Tamanine (2002, 2010), Lopes (1993, 2003), Francischini (2011), Araujo (2016), Mendes (2019), Souza (2020), entre outros. Os resultados referentes a essa variável preditora podem ser observados na tabela 11:

**Tabela 11:** Distribuição de *nós* e *a gente* segundo o *Tempo verbal* e *formas nominais do verbo*: Amostra 1990 (N=1956)

PORCUFORT I - Anos 1990 (N = 1956)				
<i>Tempo Verbal</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
Pres. Indicativo	430	44,4%	538	55,6%
Pret. Perfeito	181	57,1%	136	42,9%
Pret. Imperfeito	86	27,5%	227	72,5%
Futuro	75	51,4%	71	48,6%
Pres. Subjuntivo	4	22,2%	14	77,8%
Imp. Subjuntivo	9	36%	16	64%
Fut. Subjuntivo	3	12,5%	21	87,5%
Fut. do pretérito	11	36,7%	19	63,3%
Gerúndio	1	5,3%	18	94,7%
Infinitivo	17	17,5%	80	82,5%
$\chi^2 (9) = 114.68, p < 0,001$				

Fonte: elaboração própria

Em relação aos dados da fala culta fortalezense provenientes do projeto PORCUFORT Fase 1 (década de 1990), os resultados da tabela 11 demonstram maior proporção de uso da forma inovadora *a gente* em quase todos os tempos verbais, sendo o *pretérito perfeito do indicativo* e o *futuro do indicativo* os únicos fatores a apresentar maior percentual de uso para o pronome padrão *nós*, com percentuais de 57,1% e 51,4%, respectivamente. O tempo verbal *presente do indicativo*, embora apresente maior proporção de uso para *a gente* (55,6%), ainda representa um contexto de disputa acirrada entre as formas pronominais em variação, tendo em vista sua taxa próxima do ponto médio de 50%. De forma parecida, os resultados apontam uso mais recorrente do pronome inovador *a gente* nos tempos *pretérito imperfeito do indicativo* (72,5%), *presente do subjuntivo* (77,8%), *futuro do pretérito* (63,3%) e nas *formas nominais*.

Tínhamos a hipótese de que o pronome *a gente* ocorreria em maior proporção acompanhados de verbos no pretérito imperfeito, tendo em vista que as entrevistas sociolinguísticas utilizadas como base de dados deste estudo frequentemente tratam de experiências pessoais em que informantes relatam atividades que aconteciam corriqueiramente no passado, característica de verbos no tempo pretérito imperfeito, configurando um ambiente fértil a realização de variantes inovadoras (Labov, 2008[1972]). Essa hipótese, em termos de estatística descritiva, pode ser confirmada. Conforme vemos na tabela 11, o pronome *a gente* apresenta proporção de uso de 72,5% quando acompanhados de verbos no *pretérito imperfeito*, enquanto o pronome padrão *nós* aparece em 27,1% das ocorrências.

O modo *subjuntivo*, de forma geral, beneficia o pronome inovador, em maior proporção no *futuro do subjuntivo* (87,5%), seguido de verbos no *presente do subjuntivo*

(77,8%), e em menor proporção em verbos no *imperfecto do subjuntivo* (64%). Apontando para esta mesma direção, estão os verbos no *futuro do pretérito*, que indicam uma inclinação para o uso do pronome inovador (63,3% de ocorrências), e menor uso do pronome padrão (36,7% do total de realizações de primeira pessoa do plural).

Em relação às *formas nominais* do verbo, destacamos o fato de não haver, na amostra pesquisada, ocorrências dos pronomes *nós/a gente* com formas no particípio isolado, tendo apenas ocorrências de verbos no *infinitivo* e no *gerúndio*. Em contextos de verbo no *infinitivo*, o pronome *a gente* é usado em maior proporção, tendo 82,5% das ocorrências, em detrimento do pronome padrão *nós*, que ocorreu em apenas 17,5% das realizações. Por fim, no que diz respeito a verbos no *gerúndio*, localizamos apenas 19 ocorrências dos pronomes de primeira pessoa do plural, 18 realizações do pronome *a gente* (94,7%), e apenas um dado relacionado ao pronome *nós* (5,3%). Mais à frente, na subseção 5.1.2, procedemos ao teste para checar multicolinearidade dos dados e identificar se a falta de balanceamento gerado pela desproporcionalidade dos dados no *gerúndio* pode comprometer o resultado, para então proceder com os devidos ajustes.

Ainda em relação a variável *tempo verbal*, obtivemos, através de teste qui-quadrado de Pearson, o valor de  $\chi^2 = 114.68$ , com oito graus de liberdade, e valor de  $p < 0,001$ . Tendo como valor de referência estatística  $p = 0,05$ , podemos dizer que a variável *tempo verbal* exerce influência estatisticamente significativa sobre os pronomes de primeira pessoa *nós* e *a gente*, sendo um indicador de que essa variável deve ser incluída no modelo multivariado.

Por fim, a última variável previsorora que testamos neste estudo sobre a alternância pronominal de primeira pessoa do plural é a *saliência fônica*, que já se mostrou relevante em pesquisas anteriores, como Lopes (1993), Borges (2004), Seara (2000), Lopes (2003), Silva (2011), Souza (2020). O resultado desta variável está distribuído na tabela 12:

**Tabela 12:** Distribuição de *nós* e *a gente* segundo a *saliência fônica*: Amostra 1990 (N=1956)

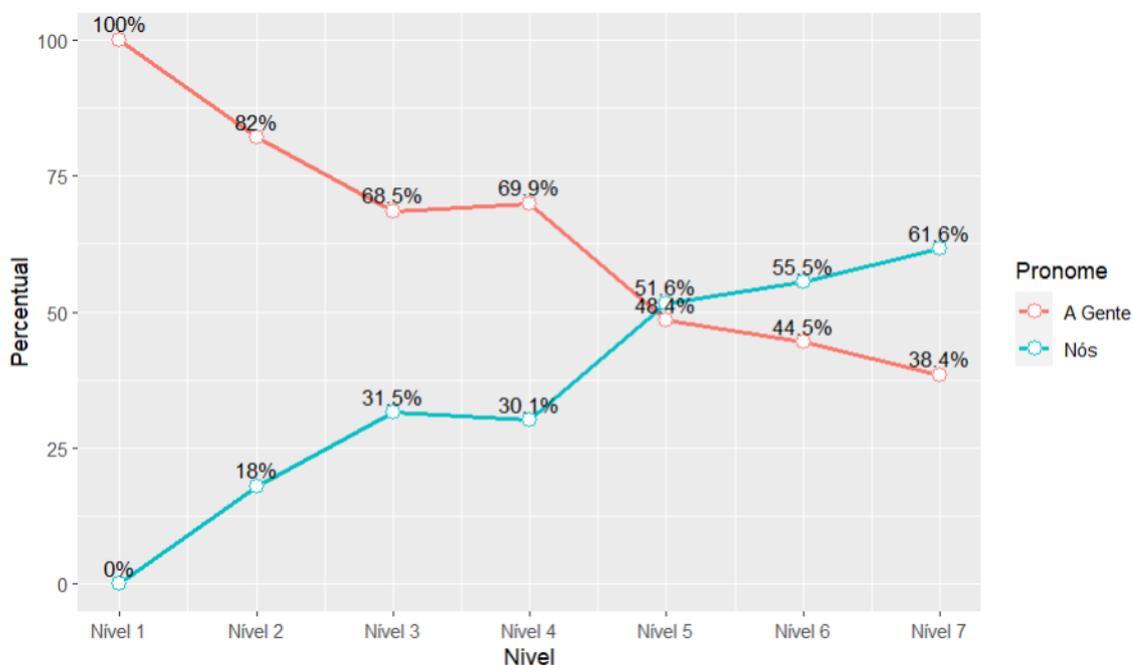
PORCUFORT I - Anos 1990 (N = 1956)				
<i>Saliência Fônica</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
Nível 1 - (cantando)	1	5,3%	18	94,7%
Nível 2 - (cantar/cantamos)	20	18%	91	82%
Nível 3 - (falava/ falávamos)	121	31,5%	263	68,5%
Nível 4 - (fala/ falamos)	134	30,1%	311	69,9%
Nível 5 - (tem/ temos)	241	51,6%	226	48,4%
Nível 6 - (cantou/cantamos)	247	55,5%	198	44,5%
Nível 7 - (é/ somos)	53	61,6%	33	38,4%
$\chi^2 (6) = 147.23, p < 0,001$				

Fonte: elaboração própria

Considerando a saliência fônica um conceito que se refere ao grau de destaque que uma variante tem em relação a outra, em uma escala de sete níveis, que vai do nível 1 (mesma forma para singular e plural), o menos saliente, ao nível 7 (diferenças fônicas acentuadas), o mais saliente, tínhamos como hipótese inicial a ideia de que, quanto menos saliente fosse o verbo na relação entre as formas de primeira pessoa do plural e terceira pessoa do singular, mais o pronome inovador seria favorecido. Em termos de proporções de uso entre os pronomes *nós* e *a gente*, nossa hipótese pode ser comprovada, tendo em vista que os níveis mais baixos de saliência apresentaram as maiores proporções de uso do pronome inovador *a gente*, e os níveis mais altos apresentam maiores percentuais de uso para a forma padrão *nós*.

O que se percebe, baseado nos dados da tabela 12, é que a distribuição dos pronomes *nós* e *a gente* em função dos níveis de saliência fônica sugere um padrão geral no comportamento das variantes: à medida em que a saliência do verbo diminui, a proporção de uso do pronome inovador *a gente* aumenta, e, ao contrário, a proporção de uso do pronome padrão *nós* tende a diminuir. Para ter uma interface gráfica desse comportamento, o gráfico 9 demonstra a tendência no uso dos pronomes *nós* e *a gente* em função da saliência fônica.

**Gráfico 9:** Proporção de uso de *nós* e *a gente* em função do grau de saliência fônica



Fonte: elaboração própria

Conforme observamos no gráfico 9, o nível 1 de saliência do verbo (mesma forma verbal para os pronomes *nós* e *a gente*), em nossa amostra, é um ambiente praticamente

categórico para a realização do pronome inovador *a gente*, com 18 dados (94,7%). Cabe ressaltar neste ponto que os verbos de saliência *nível 1* correspondem aos verbos no *gerúndio* (ver *tabela 11*), que também possuem 18 ocorrências para a forma *a gente* e apenas 1 dado do pronome *nós*, tendo demonstrado exatamente o mesmo resultado. Isso é um indicativo de que haverá multicolinearidade na amostra, tendo em vista que estamos diante de fatores que são redundantes (Oushiro, 2017), ou seja, todas as realizações de verbos no *gerúndio* têm nível de saliência fônica 1. Com isso, para o modelo de regressão convergir, possivelmente será necessário proceder a retirada desses dados, uma vez que a ausência de colinearidade é um dos princípios do modelo logístico (Levshina, 2015).

O *nível 2* de saliência fônica (cantar/cantamos), que representa os verbos no infinitivo com acréscimo de *-mos* na distinção entre *nós* e *a gente*, é um contexto fônico preferido, em termos de proporção de uso, para a realização do pronome *a gente*, dispondo de 82% das ocorrências. O pronome *nós*, por outro lado, é preterido nas realizações com verbos de saliência *nível 2*, apresentando apenas 18% das observações, conforme demonstra o gráfico 9.

A variação entre os pronomes *nós* e *a gente* tem comportamento parecido nos níveis 3 e 4 de saliência fônica. O *nível 3* (falava/falávamos), quando há conservação da sílaba tônica e acréscimo de *-mos* na distinção entre os pronomes, apresentou 68,5% de ocorrências para o pronome inovador *a gente* e 31,5% para o pronome padrão *nós*. Já o *nível 4* (fala/falamos), quando há deslocamento da sílaba tônica e acréscimo de *-mos* na distinção entre os pronomes, a forma *a gente* é preferida em 69,9% das ocorrências, enquanto o pronome *nós* se realiza em apenas 30,1% dos casos.

Verbos de saliência fônica *nível 5* (tem/temos), monossílabos tônicos ou oxítonos que passam a paroxítonos na relação singular/plural, são ambientes de disputa acirrada entre o pronome inovador e o pronome mais antigo na língua. É exatamente nesse contexto que o pronome padrão passa a ser usado em maior proporção que o pronome inovador, mesmo que a diferença ainda seja discreta. Neste caso, o pronome *nós* ocorre em 51,6% dos casos, superando a proporção de uso da forma pronominal *a gente*, que se realizou em 48,4% das ocorrências.

Em relação aos verbos de *nível 6* (cantou/cantamos), quando há redução do ditongo e acréscimo de *-mos* na distinção entre singular e plural, a expansão no uso de *nós* se torna mais marcada, tendo ocorrido em 55,5% das vezes, enquanto o pronome *a gente* apareceu em 44,5% dos casos. De igual maneira, os verbos de saliência *nível 7* (é/somos), que apresentam diferenças fônicas acentuadas, tendem a ser mais produtivos para a realização de *nós* (61,6%), em detrimento de *a gente* (38,4%), sendo o fator que mais beneficia, em proporção de uso, o

pronome padrão. Esse fato confirma nossa hipótese sobre o uso dos pronomes em função do nível de saliência do verbo, uma vez que prevíamos que os verbos mais salientes seriam mais favoráveis ao uso do pronome padrão *nós*.

O teste de estatístico de qui-quadrado realizado para testar a influência desta variável sobre a realização dos pronomes pessoais de primeira pessoa do plural apontou o valor de  $\chi^2 = 147.23$ , com seis graus de liberdade e valor de  $p < 0,001$ , o que indica interação significativa em termos estatísticos entre a variável previsoras e a variável resposta. Neste ponto, destacamos que o qui-quadrado de Pearson é um teste univariado, que verifica o efeito individual da variável previsoras sobre a variável resposta. É importante, no entanto, observar o comportamento das variáveis previsoras em interação, em um modelo de regressão.

A próxima seção se dedica à construção de um modelo de regressão logística que explica os dados em interação com as outras variáveis, ou seja, em um contexto multivariado.

### ***5.1.2 Análise multivariada da atuação das variáveis linguísticas e sociais sobre a variável resposta: o modelo de regressão logística – Amostra 1990***

Tendo em vista os resultados dos testes estatísticos, em que todas as variáveis previsoras (*sexo/gênero, faixa etária, tipo de inquérito, grau de referencialidade do pronome, tempo verbal, tipo de verbo, paralelismo formal e saliência fônica*) apresentaram interação significativa com a variável resposta, construímos um modelo de regressão logística que incluiu todas as variáveis previsoras testadas neste estudo sobre os padrões de alternância pronominal. Como vimos na subseção 4.7.2, a regressão logística é um dispositivo estatístico que ajuda a modelar os dados linguísticos da variável resposta em função da relação com as variáveis previsoras, possibilitando prever a probabilidade de um evento linguístico ocorrer em diferentes contextos (Gries, 2013).

Para criar um modelo de regressão logística na interface *RStudio*, podemos utilizar a função *glm* (Generalized Linear Model), que constitui o pacote base de instalação do *R*. A função *glm* permite, entre outras coisas, ajustar os diversos tipos de modelos lineares generalizados, ou seja, modelos lineares para variáveis respostas que não seguem uma distribuição normal, mas sim exponencial, como a distribuição binomial, por exemplo. Para efeitos mistos, uma extensão da função *glm*, permite a inclusão de variáveis aleatórias: trata-se da função *glmer* (Generalized Linear Mixed-Effects Models). Através do ambiente *RStudio*, por meio da função *glmer*, criamos um modelo de regressão logística que incluiu as variáveis

previsoras *sexo/gênero, faixa etária, tipo de inquérito, grau de referencialidade dos pronomes, paralelismo formal, tempo verbal, saliência fônica* e a variável aleatória *informante*. O modelo por nós constituído apresenta a seguinte sintaxe:

```
glmer(VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO +
REFERENCIALIDADE + TIPO.VERBO + TEMPO.VERBAL + PARALELISMO +
SALIENCIA + (1|INFORMANTE), data = Dados_1990, family = binomial)
```

Para verificar a presença de colinearidade em nossos dados, submetemos o modelo de regressão criado à análise através da função *vif* (Variance Inflation Factor). *Vif* é uma função que integra o pacote *CAR*, que permite calcular a inflação da variância nas variáveis predictoras do modelo logístico, indicando o grau de multicolinearidade presente entre as variáveis previsoras. Ou seja, essa função indica se há ou não presença de variáveis redundantes, correlacionadas entre si, o que pode prejudicar a análise dos dados, tornando-os imprecisos. Nosso modelo logístico, no entanto, não convergiu, apresentando o seguinte erro:

“singular information matrix in lrm.fit (rank= 28). Offending variable(s):SALIENCIA=n1”

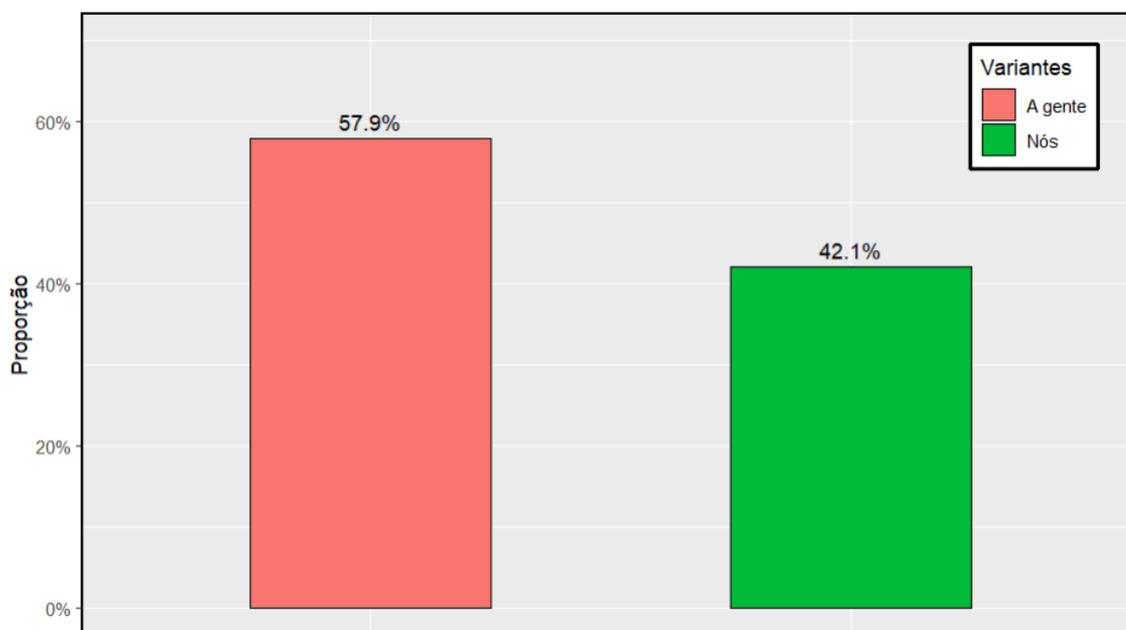
A informação do erro indica que o modelo logístico não pode ser ajustado, pois há alta correlação entre as variáveis predictoras. A mensagem "singular information matrix" constata que existe uma colinearidade perfeita entre algumas variáveis do modelo, isso quer dizer que uma ou mais variáveis do modelo testam a mesma informação, o que pode causar problemas durante o ajuste do modelo de regressão logística. Além disso, o aviso de “Offending variable(s):SALIENCIA=n1” indica que a colinearidade é causada pela variável *saliência fônica*, e que o problema está no *nível 1* dessa variável.

Isso ocorre, como já constatamos, porque todas as ocorrências de *saliência fônica nível 1* são com verbos no *gerúndio*, ou seja, o *nível 1* de *saliência* e o fator *gerúndio* da variável *tempo verbal e formas nominais do verbo* são redundantes, pois testam exatamente a mesma informação. Como prevíamos, para que o modelo pudesse convergir e gerar um *output* com resultados da interação entre todas as variáveis, foi necessário retirar os dados de *nós e a gente* com verbos de *saliência nível 1*.

Após a retirada dos dados que causaram colinearidade em nosso modelo, nos restaram 1938 observações dos pronomes de primeira pessoa do plural em amostra do PORCUFORT da década de 1990. Desse total de realizações, 1122 ocorrências foram do pronome inovador *a gente*, o que corresponde a 57,9% dos dados coletados. Por outro lado,

localizamos 816 ocorrências do pronome canônico *nós*, com percentual de 42,1% das realizações de primeira pessoa, conforme é possível observar no gráfico 10:

**Gráfico 10:** Proporção de uso das variantes *nós* e *a gente* após remoção de dados colineares



Fonte: elaboração própria

Neste ponto da análise, destacamos que o total de dados descartados em nossa amostra é de 19 ocorrências, sendo todas realizações da forma pronominal *a gente*. No entanto, a remoção desses dados não impactou expressivamente a análise final dos dados, conforme aponta o gráfico 10, tendo em vista que, em um universo de 1957 realizações dos pronomes de primeira pessoa do plural, 19 dados correspondem a menos de 1% do total. O impacto da retirada desses dados pode ser medido se compararmos as informações do gráfico 5 e do gráfico 10, que demonstram uma diferença no percentual de uso dos pronomes *nós* e *a gente* de apenas 0,43%.

Após a retirada dos dados colineares, ou seja, os dados referentes ao nível 1 da variável preditora *saliência fônica*, submetemos novamente o modelo de regressão com todas as variáveis preditoras à função *vif*, do pacote *CAR*, para observar se a remoção dos dados sanou o problema da multicolinearidade causado pela variável *saliência fônica*. Através da ferramenta computacional *R Studio*, que computou a inflação da variância dos fatores, obtivemos o seguinte *output*:

**Tabela 13:** Resultado do teste de colinearidade das variáveis preditoras através da função *vif*

Resultado do teste de inflação da variância			
	<b>GVIF</b>	<b>Df</b>	<b>GVIF<sup>^(1/(2*Df))</sup></b>
Sexo/gênero	1.139543	1	1.067494
Faixa etária	1.294888	2	1.066739
Tipo de inquérito	1.416741	2	1.090995
Grau de referencialidade	1.372493	1	1.171535
Tipo de verbo	2.435793	4	1.117712
Paralelismo	1.085255	3	1.013729
Tempo Verbal	944.888300	8	1.534480
Saliência Fônica	1389.812672	5	2.062033
car::vif(modelo)			

Fonte: Elaboração própria.

Os resultados gerados com a aplicação da função *vif* estão apresentados na tabela 13 em colunas. A primeira coluna representa o valor de *GVIF*, uma medida que indica o grau de multicolinearidade entre as variáveis preditoras, e é calculado pela razão entre a variância do coeficiente estimado pelo conjunto de variáveis incluídas no modelo e a variância no coeficiente esperado para uma única variável previsora, desprezando os graus de liberdade de cada variável. A segunda coluna apresenta os graus de liberdade, já explicados na seção 4.7.3.

A última coluna representa um valor de *GVIF* ajustado ao modelo, levando em consideração os graus de liberdade de cada variável. A interpretação do *GVIF* ajustado corresponde a uma análise mais equilibrada da multicolinearidade, tendo em vista que essa medida considera a complexidade da variável previsora em relação a cada nível dessa variável, levando em conta os graus de liberdade. Por esse motivo, adotamos os valores de *GVIF* ajustado para definir se há ou não multicolinearidade em nosso modelo.

É necessário, entretanto, considerar algumas informações sobre essa medida. Inicialmente, quanto maior for o valor de *GVIF* ajustado, maior será a multicolinearidade presente na amostra. De acordo com Levshina (2015) existem várias regras práticas, ou valores arbitrados, para determinar quando uma variável apresenta ou não colinearidade, sendo algumas delas mais rigorosas (as pontuações *vif* não devem exceder 5) e outras menos rigorosas (as pontuações *vif* não devem ser superiores a 10). Além disso, quanto maior o número de graus de liberdade (Df), maior será o ajuste feito no modelo, considerando a complexidade da variável e os valores atribuídos a cada fator desse preditor.

De posse dessas informações, os dados da tabela 13 revelam que, depois de ajustado, nosso modelo não apresenta colinearidade, sendo o maior valor de *GVIF* atribuído a variável *saliência fônica* (2.062033), valor que, mesmo se aderirmos ao valor de referência de pesquisas mais rigorosas (que não podem ultrapassar o 5), ainda assim estaria longe de indicar presença

de colinearidade em nosso modelo de regressão logística. Podemos, então, prosseguir com a análise dos dados.

Após checarmos a interação das variáveis isoladamente, através de testes de qui-quadrado, e de verificarmos a ausência de colinearidade nas variáveis controladas neste estudo, procedemos a análise do modelo completo através da função *step()*. Essa função, parte integrante do pacote base de instalação do *R*, é utilizada, essencialmente, para selecionar automaticamente as variáveis que devem ou não ser incluídas no modelo de regressão final.

A seleção das variáveis que devem compor o modelo de regressão final é automatizada e pode ser feita por três métodos distintos, a saber, *step forward*, *step backward* e *step both*, que objetivam selecionar as variáveis mais relevantes para o modelo, usando como parâmetro o Critério de Informação de Akaike (AIC). De acordo com Oushiro (2017), o valor de AIC é uma medida que determina a qualidade do ajuste no modelo de regressão, desenvolvida em 1974 pelo estatístico japonês Hirotugu Akaike, baseado na tese de que o modelo ideal deve explicar os dados com o menor número de parâmetros possíveis, ou seja, quanto menor for o valor de AIC, melhor será o modelo em relação aos outros analisados. Percebe-se que o critério de Akaike aborda uma questão de parcimônia, no sentido de que, entre um modelo maior e um modelo menor que tenha a mesma capacidade explicativa, é preferível que se adote um modelo menos complexo para o tratamento dos dados. Nesse sentido, para construir o modelo ideal para explicar nossos dados linguísticos, consideramos aqueles que apresentaram menores índices de AIC.

Em relação aos métodos de seleção das variáveis, a função *step forward* analisa, inicialmente, um modelo vazio, a partir do qual as variáveis são adicionadas uma a uma, verificando se a inclusão de determinada variável melhora o modelo, considerando o índice AIC. As variáveis são adicionadas até que a inclusão de qualquer outra variável previsora não seja capaz de melhorar o modelo (Levshina, 2015). Por outro lado, a função *step backward* parte do modelo completo, que contém todas as variáveis controladas no estudo, e remove as variáveis predictoras uma a uma, observando se o modelo apresenta melhor índice de AIC à medida em que determinada variável é removida. Essa remoção ocorre continuamente até que não haja mais melhora significativa do modelo com a remoção de qualquer outra variável.

Além dos métodos apresentados, é possível analisar a seleção automática das variáveis por meio da função *step both*. Na abordagem *both*, há uma combinação entre os dois métodos, *forward* e *backward*, através do qual a função *step* adiciona e remove variáveis predictoras, uma a uma, a cada etapa da análise. À medida em que esse processo ocorre, são selecionadas as variáveis que otimizam o ajuste do modelo, de acordo com o AIC (Levshina,

2015). Cabe destacar que todas as abordagens objetivam equilibrar a capacidade explicativa do modelo de regressão e a simplicidade, buscando evitar *overfitting*<sup>12</sup>, sendo que a escolha entre os métodos depende da natureza dos dados.

A natureza dos nossos dados, a alternância pronominal de primeira pessoa do plural, permitiu que o modelo fosse submetido a análise através dos três métodos, *farward*, *backward* e *both*. Em todas as abordagens de análise, a função *step* recomendou a retirada das variáveis *sexo/gênero* e *paralelismo formal* do modelo final. Apesar do teste estatístico de qui-quadrado demonstrar relação entre o *sexo/gênero* do informante e a realização dos pronomes *nós* e *a gente* (ver tabela 5), quando em interação com todas as variáveis predictoras testadas no modelo, esta variável parece perder o efeito significativo. A seguir, apresentaremos, a título de ilustração, o *output* gerado pela função *step farward*<sup>13</sup>:

**Tabela 14:** Resultado do teste de seleção automática das variáveis por meio da função *step*

Resultado Step farward (AIC = 782.3)			
	Df	Df Deviance	AIC
<none>		720.30	782.30
+ Sexo/gênero	1	720.30	784.30
+ Paralelismo	3	720.30	784.30
Modelo sugerido: VR ~ SALIENCIA + TEMPO.VERBAL + REFERENCIALIDADE + TIPO.INQUERITO + TIPO.VERBO + FAIXA.ETARIA			

Fonte: elaboração própria

A tabela 14 apresenta os resultados de AIC de dois modelos. O primeiro não inclui variáveis revisoras (modelo nulo), e o segundo inclui as variáveis *sexo/gênero* e *paralelismo formal*. Os resultados demonstram que um modelo nulo apresenta valor de AIC menor que o modelo que inclui estas duas variáveis. Isso significa que a inclusão destas variáveis não trouxe melhorias para o modelo, pois até mesmo um modelo que não possui variáveis (modelo nulo) tem melhor poder explicativo que um modelo que contenha as variáveis predictoras *sexo/gênero* e *paralelismo formal*. Sendo assim, o resultado da análise através da função *step farward* recomendou um modelo que incluísse apenas as variáveis *saliência fônica*, *tempo verbal*, *grau de referencialidade do pronome*, *tipo de inquerito*, e *faixa etária*. Com isso, acatamos a recomendação da análise automática e removemos as variáveis *sexo/gênero* e *paralelismo formal* do modelo de regressão.

<sup>12</sup> Para Oushiro (2017, p.218), *overfitting* se refere a “a inclusão de mais variáveis do que necessário, algo que viola o Princípio da Navalha de Occam”.

<sup>13</sup> Apesar das variáveis *sexo/gênero* e *paralelismo formal* terem sido excluídas nos três métodos de análise, apresentamos somente os resultados de *step farward* para exemplificação, tendo em vista que todos os métodos apresentam o mesmo resultado.

A próxima etapa da análise foi submeter o modelo de regressão logística à função *lrm* (Logistic Regression Model), que constitui o pacote *rms* (Regression Modeling Strategies), da linguagem *R*. Através do método da verossimilhança, a função *lrm* estima uma série de medidas estatísticas que são usadas para avaliar a qualidade e o poder de explicação do modelo de regressão logística, como o teste de Wald, o teste de razão de verossimilhança, o índice de concordância C, a curva ROC<sup>14</sup>, entre outras. Na tabela 15, é possível ver o resultado do *output*, com as medidas estatísticas atribuídas após ser submetido a função *lrm*:

**Tabela 15:** Resultado do teste multivariado de *nós* e *a gente* através da função *lrm*

	Model Likelihood Ratio Test		Discrimination Indexes		Rank Discrim. Indexes		
<b>Obs</b>	1938	<b>LR chi2</b>	873.78	<b>R2</b>	0.488	<b>C</b>	0.861
<b>A gente</b>	1122	<b>d.f.</b>	26	<b>R2(26,1938)</b>	0.354	<b>Dxy</b>	0.721
<b>Nós</b>	816	<b>Pr(&gt;chi2)</b>	>0.0001	<b>R2(26,1417.3)</b>	0.450	<b>Gamma</b>	0.722
<b>Max reriv</b>	1e-07			<b>Brier</b>	0.149	<b>Tau-a</b>	0.352

Fonte: Elaboração própria.

Para Levishna (2015) as estatísticas mais frequentemente relatadas para determinar o poder explicativo de uma regressão logística são o índice de concordância C, também conhecido como área sob a curva ROC, e o pseudo-R2 de Nagelkerke. Nesta pesquisa, adotamos o índice de concordância C como indicador para determinar a capacidade de explicação do modelo. Em relação a esse índice, a autora explica que valores de C que são iguais a 0,5 não apresentam valor discriminatório. Valores entre 0,7 e 0,8 demonstram poder de discriminação aceitável na descrição do fenômeno. Resultados do índice de concordância C superiores 0,8 indicam um excelente poder de discriminação do modelo de regressão. Desta forma, conforme exposto na tabela 15, foi atribuído ao nosso modelo um índice de concordância C igual a 0.861, tendo assim um excelente poder discriminatório dos dados.

Com o auxílio da interface *Rstudio*, por meio da função *glmer*, procedemos a análise do modelo de regressão para testar a atuação das variáveis predictoras controladas neste estudo sobre a variável resposta. Antes de apresentar a análise, é preciso considerar que o resultado do modelo de regressão logística gerou um valor de *intercept*<sup>15</sup>. O valor de *intercept* atribuído ao nosso modelo logístico corresponde à estimativa de uso da forma inovadora *a gente* nos

<sup>14</sup> Para Levshina (2015) A curva ROC (Receiver Operating Characteristic) é uma representação gráfica que representa o desempenho de um modelo com variáveis binárias. Ela estabelece a relação entre a taxa de sensibilidade (verdadeiros positivos) e a taxa de especificidade (falsos positivos) para diferentes pontos do modelo. Sendo assim, um modelo ideal apresentaria uma curva ROC que subiria na vertical e, em seguida, apontaria na horizontal para a direita.

<sup>15</sup> Oushiro (2017), explica que o valor de *intercept* equivale a estimativa de uso do segundo nível da variável resposta sobre o primeiro nível de cada variável predictor. O *intercept* corresponde ao valor de referência do modelo logístico.

seguintes contextos: (1) 22 a 35 anos, (2) inquérito do tipo DID, (3) pronome com referência genérica, (4) verbos de ação, (5) presente do indicativo, e (6) saliência fônica nível 7<sup>16</sup>. O resultado da análise dos dados da amostra dos anos 1990 está discriminado na tabela 16:

**Tabela 16:** Estimativas de uso de *a gente* modelo de regressão logística: 1990 (N = 1938)

Coefficients	Estimate	Std. Error	z value	Pr(> z )	
<b>(Intercept)</b>	1.52171	0.25207	6.830	8.48e-12	***
<b>Faixa etária</b>					
36 a 50 anos	-0.09866	0.14475	-0.682	0.49552	
A partir de 50 anos	-1.59842	0.15868	-10.073	< 2e-16	***
<b>Tipo de Inquérito</b>					
D2	0.19933	0.15378	1.296	0.19490	
EF	-1.36684	0.15543	-8.794	< 2e-16	***
<b>Referencialidade</b>					
Específica	-2.04715	0.13820	-14.813	< 2e-16	***
<b>Tipo de Verbo</b>					
Estado	-1.53241	0.23667	-6.475	9.49e-11	***
Ter	-2.46290	0.19082	-12.907	< 2e-16	***
<i>Dicendi</i>	0.03386	0.22493	0.151	0.88035	
Epistêmico	-0.32980	0.25620	-1.287	0.19799	
<b>Tempo verbal</b>					
Pretérito perfeito	0.10113	0.21674	0.467	0.64079	
Pretérito imperfeito	1.76996	0.64690	2.736	0.00622	**
Futuro do indicativo	0.66278	0.28472	2.328	0.01992	*
Pres. do subjuntivo	0.80117	0.66565	1.204	0.22875	
Pret. imp. subjuntivo	0.93777	0.79311	1.182	0.23705	
Fut. do subjuntivo	2.63566	1.08427	2.431	0.16959	
Futuro do pretérito	1.07639	0.78367	1.374	0.01506	*
Infinitivo	3.27407	1.35861	2.410	0.01596	*
<b>Saliência Fônica</b>					
Nível 2 (cantar/cantamos)	-1.21491	1.33240	-0.912	0.36186	
Nível 3 (falava/falávamos)	0.18340	0.65658	0.279	0.78000	
Nível 4 (fala/falamos)	1.01549	0.22196	4.575	4.76e-06	***
Nível 5 (tem/temos)	-1.21027	0.27059	4.473	7.72e-06	***
Nível 6 (cantou/cantamos)	0.41901	0.33167	1.263	0.20646	
Modelo: glm(VR ~ FAIXA.ETARIA + TIPO.INQUERITO + REFERENCIALIDADE + TIPO.VERBO + TEMPO.VERBAL + SALIENCIA)					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fonte: elaboração própria.

<sup>16</sup> Optamos por incluir o nível 7 da saliência fônica, ao invés do nível 2, no *intercept* para contrastar os resultados da atuação dos verbos mais salientes com as estimativas para verbos menos salientes. Para isso, tivemos que reordenar os fatores através da função *fct\_relevel()*.

O resultado da regressão logística demonstrou que, na amostra representativa da década de 1990, todas as variáveis incluídas no modelo apresentam interação estatisticamente significativa com a variável resposta, indicada pelo asterisco (\*) no final de cada fator na tabela 16. Além disso, o *output* apresenta informações relevantes para a análise em termos de estatística inferencial, como estimativa, erro padrão, valor de z e valor de p. Dentre essas unidades de medida, enfocaremos nos resultados das estimativas, pois indica o efeito da variável em termos de favorecimento, e no valor de p (*p-value*), que aponta o nível de interação entre as variáveis.

Neste ponto da análise, precisamos lembrar uma informação importante para a interpretação dos dados: os valores de estimativas apresentados na tabela 16 não correspondem ao valor final da atuação de uma variável. Para que se atribua uma estimativa a determinado fator, é necessário que o valor da estimativa deste nível seja somado ao valor de *intercept* do modelo logístico. O resultado desta soma corresponde ao valor, em *logodds*, da atuação da variável previsora analisada. Por conta disso, nesta análise, serão apontados os valores finais da atuação de cada variável, correspondente ao resultado de sua soma ao valor do *intercept* (ver seção 4.7.3).

#### a. Atuação da faixa etária sobre alternância pronominal de primeira pessoa do plural

A primeira variável apresentada no modelo logístico é a *faixa etária*. Conforme podemos observar na tabela 16, o valor de *intercept*, que apresenta estimativa de uso do pronome *a gente* entre os *mais jovens* (22 a 35 anos de idade), aponta estimativa de 1.52171 e *p-value* < 8.48e-12, demonstrando que, nos contextos que constituem o *intercept*, os informantes da *faixa a etária I* tendem a favorecer expressivamente a forma *a gente*, inibindo o uso do pronome padrão *nós*.

No que se refere aos informantes da segunda faixa etária (36 a 50 anos de idade), em comparação com os *mais jovens*, esses tendem a desfavorecer discretamente o pronome inovador *a gente*, com estimativa de 1.42305. Nota-se que esse valor de estimativa é menor que o valor de *intercept*. Isso indica que os informantes de *faixa etária I* favorecem mais o uso do pronome *a gente* do que os informantes da *faixa etária II*, tendo em vista o valor da estimativa, que é maior entre os *mais jovens* (*intercept*). No entanto, o valor de p estimado para esta faixa etária está fora do intervalo de confiança, ou seja, o valor de p = 0.49552 indica que os

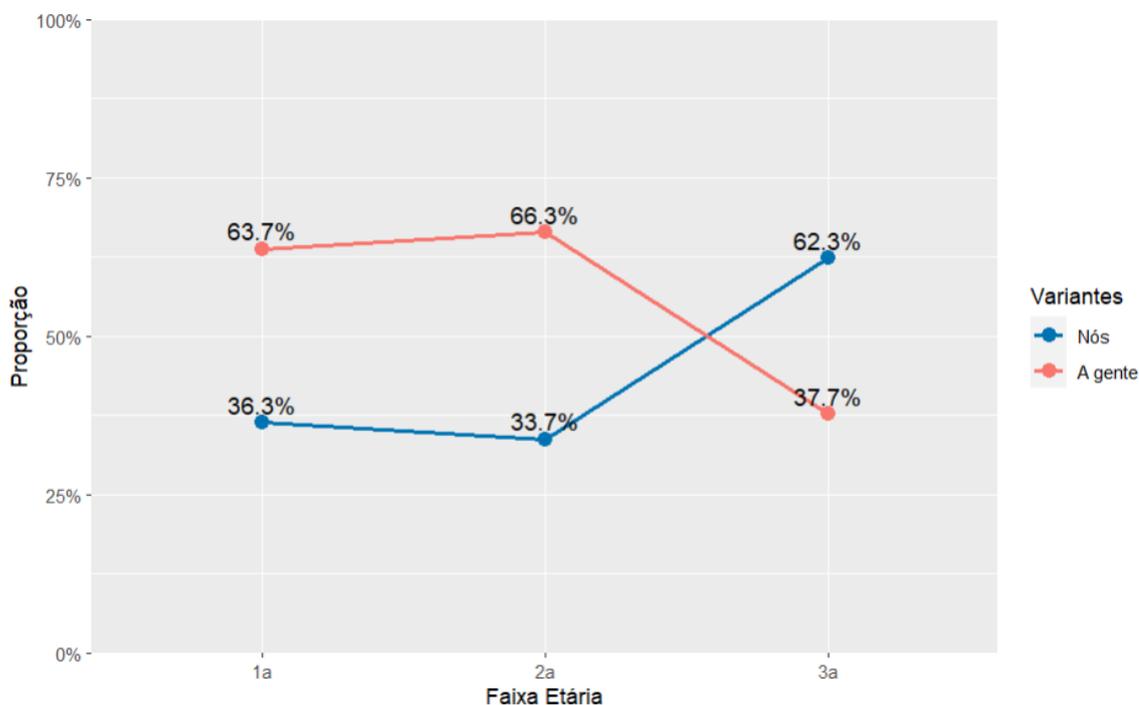
informantes de 36 a 50 anos não demonstram, em nosso modelo, associação estatisticamente significativa sobre a realização dos pronomes de primeira pessoa do plural.

Por outro lado, para os informantes *mais velhos*, com idade a partir de 51 anos, o valor de  $p$  estimado em nosso modelo ( $p > 2e-16$ ) indica forte associação entre o uso dos pronomes *nós* e *a gente* e os informantes desta faixa etária. Além disso, a estimativa para o uso de pronome *a gente* entre os informantes da terceira faixa etária apresentou resultado negativo (-0,07671), demonstrando que os falantes *mais velhos* da Amostra 1990 tendem a inibir fortemente o uso do pronome inovador, sendo estes favorecedores do pronome *nós*. Esses resultados confirmam nossa hipótese inicial sobre o comportamento linguístico dos informantes em relação a sua *faixa etária*, pois prevíamos que os *mais jovens* favoreceriam o pronome inovador, enquanto os *mais velhos* se inclinariam ao uso do pronome padrão, como demonstram várias pesquisas (Seara, 2000; Tamanine, 2002; Nascimento, 2013; Araujo, 2016; Vitorio, 2015, entre outros).

De maneira geral, os estudos sociolinguísticos de orientação laboviana costumam correlacionar o fator *faixa etária* a determinada mudança linguística, ou seja, frequentemente pesquisadores costumam analisar o comportamento do fenômeno em função das faixas etárias controladas, em busca de indícios que associe esta variável à mudança linguística. Sobre essa questão, Dallemole, Osorio e Patatas (2018) discorrem:

Quando a correlação da variante em questão ao fator idade for constatada, identifica-se alguma tendência, seja ela inovadora ou conservadora, de acordo com a distribuição nos grupos etários. Por exemplo, se o uso da variante inovadora for mais frequente entre os mais jovens, decrescendo em relação aos informantes mais velhos, identifica-se uma tendência de mudança em progresso, uma vez que aqueles de maior idade tendem a empregar a variante conservadora na medida em que o percentual decresce para a variante inovadora. (Dallemole, Osorio e Patatas, 2018, p. 112).

A distribuição das proporções de uso de cada variável de nossa amostra em função da *faixa etária* (ver tabela 6) demonstra uma tendência dos *mais jovens* ao uso da variante *a gente*, enquanto os *mais velhos* se inclinam ao uso do pronome padrão. Em nossos dados, a linha estatística da proporção de uso dos pronomes apresenta uma curva, isso indica que a relação entre a variável resposta e a variável previsora *faixa etária* não é estritamente linear, podendo haver efeitos não lineares que influenciam esse resultado (Santos; Branches, 2018). Para uma visualização gráfica da curva estatística que representa o uso dos pronomes *nós* e *a gente* em nossa amostra, apresentamos o gráfico 11:

**Gráfico 11:** Proporção de uso de *nós* e *a gente* por faixa etária (N = 1938)

Fonte: elaboração própria

Como podemos observar no gráfico 11, a proporção de uso de cada variante em relação à *faixa etária* não representa uma tendência linear, sugerindo uma mudança na taxa de variação nas proporções de uso ao longo do *eixo x*. Em outras palavras, baseado em nossos dados, não se pode dizer exatamente que, à medida em que a idade do informante aumenta, a proporção de uso da variante inovadora diminui, conforme teorizam Dallemole, Osorio e Patatas (2018). Em nossa amostra, isso acontece porque a proporção de uso do pronome *a gente* entre os informantes de 36 a 50 anos é discretamente maior que a proporção de uso entre os informantes de 22 a 35 anos (66,3% e 63,7%, respectivamente). No entanto, há de se considerar que a linha do gráfico representa uma curva suave, o que pode indicar que contextos mais complexos podem estar atuando sobre os dados, tendo em vista que não há diferenças significativas entre o comportamento linguístico dos informantes das faixas etárias I e II.

Porém, ao comparar o comportamento linguístico dos informantes *mais jovens* com os *mais velhos* de nossa amostra, há uma queda significativa no percentual de uso do pronome inovador e, em contrapartida, um aumento expressivo na proporção de uso do pronome padrão, apontando para uma tendência no que diz respeito ao uso dos pronomes de primeira pessoa do plural entre os informantes *mais velhos*, sendo estes os que mais utilizam o pronome conservador. Nesse sentido, de acordo com o que teorizam Dallemole, Osorio e Patatas (2018)

e a própria teoria variacionista (Labov, 2008 [1972]), quando há, entre os mais jovens, preferência pelo uso da variante inovadora e, entre os mais velhos, maior tendência ao uso do pronome conservador, como é o caso dos pronomes *nós* e *a gente* em nossa amostra, há indícios de uma mudança linguística em progresso. Essa tendência pode ser observada em nossos dados, tendo em vista que a proporção de uso do pronome padrão *nós* entre os informantes *mais velhos* é quase o dobro da observada entre os *mais jovens*.

Freitag (2005), no entanto, pondera sobre a atuação da *faixa etária* no processo de mudança, principalmente em relação a interpretação dos resultados com base nessa variável. Para a autora, os resultados levantados a partir do controle da variável *faixa etária* devem ser analisados com total cuidado, “pois nem todo indício de mudança em curso apresentado pela distribuição dos resultados em função da faixa etária é reflexo somente da gradação etária dos falantes que constituem a amostra do estudo” (Freitag, 2005, p. 106).

Ainda de acordo com a autora, a análise de determinado fenômeno variável em função da *faixa etária* é realizada para caracterizar situações de estabilidade, mudança incipiente, mudança em progresso ou mudança completa. No entanto, atribuir características às fases de uma mudança (ou mesmo o *status* da mudança) baseado na distribuição do fenômeno em função da idade é uma questão complicada, pois:

não só a comunidade como também o indivíduo podem influir no fenômeno da variação e da mudança. Por isso, é necessário controlar o comportamento linguístico da comunidade como um todo e o comportamento linguístico de cada indivíduo da amostra, especificamente (Freitag, 2005, p.110).

Neste ponto, vale pontuar que, tendo a autora destacado o efeito individual de cada informante da comunidade de fala, incluímos no modelo logístico a variável aleatória *informante*, que balanceia os dados da amostra, considerando que cada indivíduo pode apresentar padrões distintos de uso dos pronomes *nós* e *a gente*. Com isso, é possível verificar a variabilidade entre cada informante e modelar os dados de modo a evitar que o efeito individual do informante prevaleça sobre o efeito geral na amostra. A partir da inclusão da variável aleatória, o modelo passa a considerar tanto os efeitos gerais dos dados da amostra quanto a heterogeneidade entre os informantes individualmente, o que torna possível uma análise balanceada e realista dos dados de fala. Isso permite atribuir veracidade a distribuição da variável resposta em função das faixas etárias, sendo possível a interpretação de que a alternância entre os pronomes de primeira pessoa do plural apresenta indícios de mudança em curso, uma vez que as faixas mais jovens utilizam em maior proporção a forma inovadora *a gente*, e esse comportamento linguístico tende a acompanhá-los para a vida (Lucchesi, 2015).

Na mesma direção, Naro (2008) apresenta um posicionamento teórico que nos fornece ferramentas para analisar a atuação dessa variável previsora sobre o fenômeno linguístico da variação entre os pronomes *nós* e *a gente*. O pressuposto estabelecido pelo autor é baseado na crença de que os padrões linguísticos são adquiridos até a puberdade, por volta dos 15 anos de idade, e que, a partir dessa idade, o vernáculo permanece basicamente estável. Ou seja, um adulto hoje estaria reproduzindo o comportamento linguístico de quando tinha, aproximadamente, 15 anos de idade. Além disso, a mudança linguística pode ser atestada ao observar o comportamento linguístico da comunidade, comparando as faixas etárias *mais jovens* e as *mais velhas* (Labov, 2008), e não a fala de um mesmo indivíduo isoladamente. Nesses termos, ao comparar a distribuição das variantes entre as faixas etárias, se observam indícios de mudança quanto ao uso dos pronomes em nossa amostra, no sentido de *a gente* estar ganhando espaço como referência à primeira pessoa do plural, demonstrando que o “processo de mudança está sendo liderado pela geração *mais jovem* que vem aumentando seu repertório e sua consciência linguística” (Dália, 2021, p 44).

*b. Atuação do tipo de inquérito sobre alternância pronominal de primeira pessoa do plural*

A segunda variável testada em nosso modelo de regressão logística foi o *tipo de inquérito*. No capítulo de metodologia (ver subseção 4.4.1.2), atribuímos o status de extralinguística, ou social, à essa variável. No entanto, é importante considerar que, diferente da outra variável social incluída neste modelo logístico (*a faixa etária*), o *tipo de inquérito* não reflete qualquer traço social do informante, mas apenas o contexto de gravação e de interação entre os indivíduos que participaram na coleta de dados, sendo uma estratificação metodológica proveniente do próprio projeto PORCUFORT, que adotamos como variável preditora nesta pesquisa.

Desta forma, é importante salientar que cada tipo de inquérito presente em nossa amostra corresponde a situações comunicativas distintas, com níveis de atenção à fala igualmente distintos. Nessa perspectiva, os inquéritos do tipo *diálogo entre dois informantes* (D2) são os que apresentam menor nível de formalidade, os inquéritos do tipo *diálogo entre informante e documentador* (DID) são os que possuem nível intermediário de formalidade, e as *elocuições formais* (EF) são os inquéritos com maior nível de formalidade entre os três (Araújo, 2011). Em relação a nossa hipótese sobre variável, acreditávamos que, nos inquéritos do tipo D2, haveria maior favorecimento do pronome *a gente*, devido à natureza menos formal do

inquérito, enquanto nos inquéritos do tipo EF, o pronome *nós* seria beneficiado, devido ao alto grau de monitoramento.

Os resultados da tabela 16 confirmam nossa hipótese inicial sobre a atuação dessa variável em relação à realização dos pronomes de primeira pessoa do plural. Inicialmente, precisamos considerar que o valor de referência desta variável corresponde a estimativa de uso do pronome *a gente* em inquéritos do tipo DID, no mesmo contexto que constitui o *intercept*, indicando favorecimento do pronome inovador nesse contexto, com valor de estimativa de 1.52171 e valor de  $p < 0,001$ . Decidimos usar como referência os inquéritos do tipo DID, que consideramos possuir grau intermediário de formalidade, para que fosse possível contrastar os resultados dos inquéritos do tipo D2 (menos formal), com as Elocuções Formais (mais formal), em termos de favorecimento do pronome *a gente*.

Inquéritos com diálogos entre dois informantes (D2), como prevíamos, tendem a favorecer o uso do pronome inovador, com estimativa de uso de 1.72104, valor maior que o intercepto do nosso modelo, indicando que esse tipo de inquérito é ambiente ainda mais propenso ao uso de *a gente* em comparação com os inquéritos do tipo DID. No entanto, apesar da estimativa apontar para um expressivo favorecimento da forma pronominal inovadora nesse tipo de inquérito, o valor de  $p$  é igual a 0.19490, resultado que aponta ausência de interação significativa entre esse fator e a realização dos pronomes de primeira pessoa, estando seu resultado fora do intervalo de confiança. Ou seja, os inquéritos do tipo D2, em nosso modelo logístico, não demonstram interação estatisticamente relevante com a variável resposta.

Por outro lado, os inquéritos do tipo EF, em termos de favorecimento, demonstram uma tendência à utilização do pronome padrão *nós*, sendo a forma pronominal *a gente*, em relação ao nível de referência, inibida nesse tipo de inquérito, com estimativa de uso de 0,15487, valor expressivamente menor que a estimativa atribuída ao *intercept*. Isso aponta para o alto desfavorecimento da forma pronominal *a gente* em contextos de maior monitoramento da fala. Além disso, o  $p\text{-value} < 0,001$  indica forte interação desta categoria sobre a realização dos pronomes *nós* e *a gente*. Esse resultado também confirma nossa hipótese em relação ao uso do pronome padrão nesse tipo de inquérito, estando o uso dos pronomes associado ao tipo de inquérito e ao nível de formalidade empregado em cada um deles.

A análise dos dados de fala em nossa amostra demonstra comportamentos distintos do fenômeno variável em cada tipo de inquérito, apontando para o pressuposto de que há “usos variados conforme a situação – mais formal ou menos formal – em que se está falando” (Coelho et al, 2015, p. 82). Ou seja, há adequação do comportamento dos pronomes ao contexto da

situação comunicativa a depender do grau de formalidade atribuído a cada situação. Nessa perspectiva, são pertinentes as palavras dos autores:

Certamente, em situações mais formais usamos uma linguagem mais monitorada, ou seja, prestamos mais atenção à forma como falamos, enquanto que em situações mais informais usamos uma fala mais coloquial. Essas duas linguagens são chamadas, respectivamente, de registro formal e registro informal. (Coelho et al, 2015, p. 83).

O que se espera de qualquer registro formal é que se atenda às normas prescritas pela gramática normativa, ou pelo menos que se aproxime delas, diferentemente dos registros menos formais, em que há preocupação menos acentuada com o monitoramento da fala. No entanto, os autores alertam para a importância de não tratar a questão da formalidade categorizando-a em dois polos distintos – formal e informal – mas pensar num *continuum* que perpassa em situação de maior e de menor formalidade, como uma escala gradativa, que parte de um menor grau, passando por contextos intermediários até atingir o maior grau de formalidade, a depender a situação comunicativa. Sobre essa questão, Leite (2005) discorre:

há variáveis próprias do falante, que são sua origem geográfica e sua classe social, o que confirma o que se pode entender por dialeto. Outros dados, no entanto, são típicos dos diversos contextos de comunicação em que se integra o usuário ao longo de seu dia. Esses são, então, chamados registros ou níveis de fala, e se configuram pelo maior ou menor grau de formalidade ou informalidade nos contratos sociais. (Leite, 2005, p. 185).

Nesse sentido, a variabilidade, de acordo com o autor, está correlacionada ao grau de formalidade atribuído a determinada situação comunicativa. Na mesma direção, Labov (2008) explica que, quanto maior o grau de formalidade em determinados contextos, maior o monitoramento que os falantes fazem de sua própria língua. Assim, a maneira como o usuário utiliza a língua – e conseqüentemente a variabilidade intrínseca a ela – está diretamente ligada à atenção dada à fala no momento da elocução, ou seja, ao monitoramento da língua, nos termos do autor. Os dados de nossa pesquisa estão em consonância com os autores, tendo em vista que atribuímos menor grau de monitoramento aos registros do tipo D2, pois nesses inquéritos os informantes “mantêm entre si um certo grau de intimidade, como amigos ou familiares, o que torna esse tipo de inquérito o mais espontâneo dentre os três já citados” (Araújo; Viana; Pereira, 2018, p.184), sendo estes favorecedores do pronome *a gente* (ver tabela 16).

Os inquéritos com *entrevista entre informante e documentador* (DID), apresentam características diferentes das encontradas nos inquéritos D2, pois não há relação de intimidade entre os envolvidos na gravação. Há se se ponderar, de acordo com Votre (2003), que outros fatores, como de presença de uma pessoa culta, de equipamentos para gravação e, até mesmo,

o assunto da conversa podem interferir na realização linguística, acionando o monitoramento do discurso e, conseqüentemente, induzindo o falante a utilizar-se de um discurso mais formal. Isso poderia interferir na gravação do vernáculo, no sentido de conduzir o informante ao uso da forma padrão.

Para amenizar o impacto do monitoramento, no entanto, em registros do tipo DID e D2, foi adotado o método da narrativa de experiência pessoal, que reduz “o efeito negativo causado pela presença do pesquisador na naturalidade da situação de coleta de dados” (Tarallo, 2001, p. 21) e, assim, possibilitando a captação do dado próximo ao vernáculo. O Inquérito do tipo DID “apresenta um nível intermediário de formalidade, já que, em geral, não havia intimidade entre informante e documentador” (Araujo, 2000, p. 09). Nesses inquéritos, o favorecimento da variante inovadora reduz se comparado ao D2, possivelmente motivada pelo grau intermediário de formalidade atribuído a esse tipo de registro, sendo mais formal que os inquéritos D2 e menos formal que os EF.

Já os inquéritos do tipo EF, como já mencionado, apresentam alto grau de formalidade, pois os falantes estão expostos a ambiente de fala que exige elevado grau de monitoramento. Nesses ambientes, a variante conservadora é favorecida, confirmando novamente a teoria de Labov. Com isso, temos a descrição clara do *continuum* de formalidade que cada tipo de inquérito representa no comportamento do fenômeno variável, que perpassa do menor grau (D2), passando pelos níveis intermediários (DID), até o maior nível de monitoramento de nossa amostra (EF).

Em nossos dados, a representação gráfica desse *continuum* na atuação do grau de formalidade sobre a realização dos pronomes de primeira pessoa (ver o Gráfico 8), aponta em direção à redução nos percentuais de uso do pronome inovador a medida em que o grau de formalidade aumenta. Ao avaliar o resultado desta pesquisa em função do status atribuído a cada tipo de inquérito, percebemos com clareza o *continuum* que permeia essa variável (Coelho et al., 2015). Nos inquéritos menos formais, por exemplo, o uso da forma pronominal *a gente* chega a 72,8%, proporção bastante elevada se considerarmos que os dados são provenientes de falantes cultos, de quem se espera maior uso da variante tida como padrão. Esse percentual diminui conforme o nível de monitoramento aumenta, sendo a variante inovadora responsável por 61% das ocorrências de primeira pessoa do plural em inquéritos do tipo DID, que possui nível intermediário de formalidade. Por fim, a proporção de uso do pronome inovador diminui significativamente em contextos mais formais, sendo realizado em 42,3% dos dados em registros de Elocuções Formais gravadas na década de 1990. Esses resultados comprovam a relação direta entre a realização dos pronomes *nós* e *a gente* e o grau de formalidade do registro,

como demonstra Leite (2005).

Dada a discussão sobre essa variável, os resultados da tabela 16 confirmam que a variável tida como inovadora (*a gente*) é sensível aos contextos em que a atenção da fala apresenta menor grau de monitoramento, sendo ambientes férteis à sua realização. Já a variante que recebe status de padrão (*nós*), parece estar associada a contextos de maior atenção à fala, sendo esses contextos favorecedores ao uso deste pronome.

*c. Atuação do grau de referencialidade sobre alternância pronominal de primeira pessoa do plural*

A próxima variável selecionada como estatisticamente relevante através da função *step*, foi o *grau de referencialidade do pronome*. Nossa hipótese era a de que o pronome inovador *a gente* seria favorecido quando empregado em sentido mais genérico, indeterminado, pois, nos termos de Hopper (1991, p.124), ao passar por um processo de gramaticalização, “alguns traços do significado lexical original de um item tendem a aderir à nova forma gramatical e detalhes de sua história lexical podem refletir-se na sua distribuição gramatical”. Nesse sentido, o traço generalizador proveniente de sua forma original acompanharia o pronome *a gente*, estando tão aderido à forma inovadora a ponto de permitir sua utilização como estratégia de indeterminação (Oliveira, 2006).

O princípio da persistência remete a manutenção de alguns traços semânticos intrínsecos que a forma gramaticalizada absorve e conserva da forma original. Essa preservação dos traços intrínsecos é notória nas fases iniciais e intermediárias da gramaticalização, mas podem desaparecer se o processo estiver avançado. É importante ressaltar que esses traços intrínsecos podem explicar certas restrições que a forma gramaticalizada *a gente* sofre, conforme explicam Omena & Braga (1996):

No caso do uso de *a gente* a persistência do traço indeterminador provoca certas restrições em seu uso. Enquanto o pronome *nós* admite ser modificado por quantificadores, numerais, especificadores enfim, o mesmo não se dá com a forma *a gente*. *Todo, cada um, nenhum* podem modificar *nós*; mas não *a gente*. (Omena & Braga, 1996. p. 80).

Essas restrições são resultado do traço generalizador que integra o pronome, admitindo que a inclusão de quantificadores comprometeria o caráter genérico que persistiu na forma após o processo de gramaticalização. A pronominalização de *gente* > *a gente* exemplifica bem o princípio da persistência. A forma *a gente*, ao se pronominalizar, não perdeu todas as

propriedades formais do substantivo *gente*, tampouco absorveu todas as propriedades dos pronomes pessoais. Assim, o pronome *a gente* mantém da sua forma original o traço formal de terceira pessoa, porém aciona uma interpretação semântica de primeira pessoa do plural. Além disso, o pronome *a gente* manteve o caráter coletivo, indeterminador, típico do substantivo *gente*.

No mesmo caminho, Mendonça e Nascimento (2015), quando pesquisaram as estratégias de indeterminação, observaram que o pronome *a gente*, em geral, também pode ser usado pelo falante para indeterminar o sujeito, generalizando-o. O autor atribui essa característica ao percurso de pronominalização do substantivo *gente*, que passou por um processo de gramaticalização, chegando à forma *a gente*, empregada como pronome indefinido, até que, enfim, fosse possível seu emprego como pronome pessoal, concorrendo com o pronome *nós* (Lopes, 2003).

Dito isto, o resultado do modelo logístico exposto na tabela 16 confirma nossa expectativa em torno dessa variável. Como podemos ver na tabela 16, a forma inovadora é favorecida em contextos de uso com referente genérico (*intercept*), e expressivamente desfavorecida quando empregada com sentido específico (-0,52544). Além disso, o modelo demonstrou forte interação entre realização dos pronomes de primeira pessoa do plural e o grau de referencialidade específica, apresentando valor de *p* significativamente diferente de zero (*p-value* < 0,001), atuando de modo a favorecer o pronome padrão *nós* em realizações com referentes específico, mantendo assim sua função prototípica (Lopes, 2003). Relacionando esses resultados com a proporção de uso dos pronomes em função do grau de referencialidade, vemos que as informações se confirmam. O pronome *a gente* apresenta alto percentual de uso em contextos de referencialidade genérica (71,3%), enquanto em contextos de referencialidade específica há menor recorrência do pronome inovador (38,5%). Desta forma, ocorrências presentes em nossa amostra, como as (121), (122) e (123), configuram ambientes férteis, em termos de favorecimento e de proporção de uso, para a realização do pronome *a gente*:

- (121) Inf.: com o crescimento da burguesia e com o fortalecimento de ideias ideais alinhados com os interesses burgueses *a gente* percebe um rompimento desse modelo de estado absolutista graças as chamadas revoluções burguesas (F2 EF 39 mulher, 36 anos)
- (122) Inf.: ... COM com a pulga atrás d'orelha mais sair com um certo orgulho né? Pow *a gente* tem uma moeda forte... a nossa moeda ta mais forte que o dólar... (F1 DID 29, homem 28 anos)
- (123) Inf.: que isso aqui é diferente das ideias da teoria da reminiscência de Platão

Platão num dizia que ah... *a gente* nasce com todas as ideias do mundo das ideias aí quando NASce aí dá um BRAN::co (F2 EF 46, mulher 33 anos)

Nos excertos (121), (122) e (123), extraídos de nossa base de dados, o pronome *a gente* resgata uma referência genérica, indeterminada. Em alguns casos, como na ilustração (123), por exemplo, a referencialidade do pronome atinge o grau mais elevado de indeterminação, sendo possível sua substituição pelo clítico *se*. Cabe aqui ponderar, no entanto, que várias pesquisas atestam que esse traço intrínseco de indeterminação que o pronome *a gente* possui está em processo de mudança linguística, uma vez que a forma *a gente* está cada vez mais sendo empregado em sentido específico (Omena, 2003; Tamanine, 2010; Araujo, 2016, entre outros).

Seguindo a mesma tendência dos trabalhos citados, os resultados dessa pesquisa demonstram que o pronome inovador é também empregado com referente específico, mesmo que o resultado do modelo tenha indicado esse contexto como desfavorável à realização do pronome inovador (ver tabela 16). Assim, na amostra da comunidade de fala de Fortaleza, embora a referencialidade específica favoreça o uso do pronome *nós*, o uso do pronome inovador *a gente* neste contexto corresponde a 38,5% das realizações de primeira pessoa do plural, demonstrando que a forma pronominal é também empregada com frequência em contextos de referencialidade específica. Construções como as encontradas em (124) e (125) exemplificam o uso da forma inovadora no sentido de especificidade:

(124) Inf.: enTÃO é o sonho dele é morar em casa aí eu já disse C. se *a gente* tivesse condição de fazer a casa que a gente quiSEsse... EU claro precisa... a minha a a o meu QUARto que eu acho assim que eu acho assim o MÁximo um QUARto pra mim (Inq. 02 – F1 – D2, mulher, 23 anos)

(125) Inf.: na eduCAção física nera de manhã que era o professor P. era P. nera? gêmeos nera os gêmeos? aí a R. fazia hora com a cara dele... depois *a gente* ia até tomar BANho naquele banheiro bem grande... (Inq. 55 – F2 – D2, mulher, 49 anos)

O comportamento apresentado pelo pronome padrão *nós* em nossa amostra, por outro lado, demonstra que há maior probabilidade de uso do pronome canônico quando o referente é específico, seguindo sua função prototípica. Lopes (1993, p. 129) explica que “o falante utiliza preferencialmente o pronome *nós* para se referir a ele mesmo e mais o interlocutor (não-eu), ou a não-pessoa: referente [+perceptível] e [+determinado]”. Os dados presentes em (126) e (127) exemplificam esse fator:

- (126) Inf.: por onde passei fui um bom profissional fui um bom promotor... um bom profesSOR... e a vida continua e a vida continua e *nós* estamos é aqui conversando ah ah ah ah a vida da gente e foi bom demais essa conversa (Inq. 80 – F2 – D2, homem, 77 anos)
- (127) Inf.: eu era subdiretor do Náutico... subdiretor social... e quando eu ia dar planTÃO... que eu ia com a M. ... a M. ia comigo e o A. pra só *nós* dois ir né? (Inq. 80 – F2 – D2, homem, 77 anos)

Esses resultados estão em consonância com os obtidos por Lopes (1993), Seara (2000), Tamanine (2002), Borges (2004), Araujo (2016) e Souza (2020), no sentido de a forma pronominal inovadora *a gente* ser favorecida quando a referência do pronome é genérica, enquanto o pronome padrão *nós* tende a se destacar em contextos em que o referente é específico, o que revela uma tendência em relação à referencialidade dos pronomes no português brasileiro, confirmado por diversas pesquisas.

*d. Atuação do tipo de verbo sobre alternância pronominal de primeira pessoa do plural*

O *tipo de verbo* foi a segunda variável linguística incluída no modelo de regressão logística desenvolvido para análise dos nossos dados. Esse grupo de fatores tem demonstrado relevância em alguns estudos sobre a variação pronominal *nós* e *a gente* (Oliveira, 2006; Tamanine, 2010; Nascimento, 2013; Araujo, 2016). Para Tamanine (2010) o tipo de verbo se figura como motivador de diversos contextos que podem ser favorecedores ou inibidores do uso das formas pronominais *nós* e *a gente*. A análise dessa variável tem objetivo de verificar se, em nossa amostra, encontramos tendências no comportamento dos verbos que já foram atestadas em outras pesquisas. Por exemplo, os verbos de estado apresentam tendência ao favorecimento de *nós* (Tamanine, 2010), e os verbos *dicendi* tendem a beneficiar a forma pronominal inovadora *a gente* (Araujo, 2016). A tabela 16 apresenta os ambientes mais propícias, em termos de favorecimento, à realização da forma pronominal *a gente*.

Os verbos de ação constituem o valor de *intercept* do modelo, ou seja, o valor de referência que é utilizado como parâmetro para análise. O valor de estimativa positivo indica que intercepto, do qual faz parte os verbos de ação, beneficiam a realização do pronome *a gente* (1.52171), e valor de p igual a 8.48e-12 atesta correlação entre este fator e a variável resposta. De modo a ilustrar, a utilização do pronome *a gente* parece ser favorecida quando associados a

formas verbais que expressam ação, como nas ocorrências (128) e (129):

- (128) Inf.: a função social da propriedade então é só pra vocês verem como é algo AMplo **a gente coloca** o olhar só no imóvel até por força do que tá na constituição (Inq. 39 – F2 – EF, mulher, 36 anos)
- (129) Inf.: ... mas só fez ferir um pouco seu braço... **a gente faz** um curativo isso não tem importância... (Inq. 23 – F2 – EF, mulher, 36 anos)

Ainda em relação aos verbos que expressam ação, podemos atribuir esse favorecimento da forma *a gente* ao elevado número de ocorrências dos pronomes de primeira pessoa do plural com verbos dessa natureza, representando a maioria dos verbos de nossa amostra. Para dimensionar a proporção de verbos de ação em nossos dados, estamos trabalhando com um total de 1938 realizações de primeira pessoa do plural, das quais 1073 estão diretamente associadas à presença de verbos de ação, sendo 739 realizações do pronome *a gente*, e apenas 334 ocorrências do pronome padrão *nós*. Essa expressiva diferença entre a frequência de uso das formas pronominais de primeira pessoa do plural atua diretamente no favorecimento de um pronome em detrimento do outro, como apontam os resultados.

Por outro lado, os *verbos de estado* estão associados ao favorecimento do pronome padrão *nós*. Conforme exposto na tabela 16, as estimativas atribuídas aos verbos que expressam estado demonstram que, em relação ao *intercept*, há desfavorecimento do pronome *a gente* (-0,0107). Nesse sentido, quando o membro da comunidade de fala usa o verbo de estado, há maior probabilidade para o emprego de *nós*. Outros autores, a saber, Tamanine (2002) e Francischini (2011), encontraram resultados que se aproximam dos obtidos nesta pesquisa sobre a atuação dos *verbos de estado* na realização dos pronomes.

Tamanine (2002), em pesquisa com dados de fala dos catarinenses das cidades de Chapecó, Blumenau e Lages, coletados a partir do projeto VARSUL, observou que os *verbos de estado*, além recorrentes na amostra (55%), são os que apresentam menor probabilidade para o uso de *a gente*, com peso relativo de 0.390, indicando alto favorecimento do pronome *nós* com verbos desse tipo. Já Francischini (2011) observou que, em contextos de uso de *verbo de estado*, a probabilidade para a aplicação de *a gente* é menor, tanto em frequência de uso (41% das ocorrências com verbos de estado), quanto em peso relativo (0.400). Esse resultado, de acordo Tamanine (2002), pode estar associado a alta frequência de verbos com a estrutura *presente do indicativo* + desinência *-mos*, que, segundo a autora, pode inibir o uso do pronome *a gente*. Para testar essa hipótese, analisamos a distribuição, em nossa amostra, dos verbos de estado que se expressam no presente do indicativo com desinência *-mos*. O levantamento dos

verbos neste contexto demonstrou alta proporção de uso dos pronomes com os verbos *somos* (49%) e *estamos* (37%), sendo verbos de estados. A hipótese da autora prescreve que verbos com a desinência -mos estão associados ao uso de *nós* e, sendo os verbos de estados, a expressiva proporção de uso age de modo a favorecer o emprego do pronome inovador. Os exemplos (130), (131), (132) e (133) ilustram esse tipo de ocorrência:

- (130) Inf.: ***nós somos*** a primeira geração que temos pais e nascemos em Fortaleza...que nasceram em Fortaleza (Inq. 02 – F2 – D2, mulher, 26 anos)
- (131) Inf.: *nós* num ***somos nós somos*** a maiorIA... mas *nós* num TEemos o uso da força... QUEM DETÉM a força num país é o sisTEma... político... e o sistema militar.. (Inq. 17 – F1 – EF, homem, 33 anos)
- (132) Inf.: aqui ***nós já estamos*** no Rio Grande do Sul... frio gosTO::so::...outro OUTro local que parece que você não tá no Brasil (Inq. 51– F2 – D2, mulher, 36 anos)
- (133) Inf.: e::... no futeBOL já fomos... né? ***nós estamos*** com esPERanças aí nesse ano (Inq. 20 – F1 – DID, mulher, 32 anos)

Outro tipo de verbo que parece favorecer o pronome padrão é o verbo *ter*, como confirmam estudos sociolinguístico sobre a variação pronominal (Tamanine, 2010; Araujo, 2016). Nos dados de fala culta representativo da década de 1990, há maior proporção de uso do pronome *nós* junto a verbos do tipo *ter* (57,4%), sendo o pronome inovador menos recorrente neste contexto. Em termos de favorecimento, análise do modelo de regressão logística apontou estimativa de uso de -0,94119 para a atuação do verbo *ter*, sendo o fator que mais desfavorece a aplicação do pronome inovador *a gente*. Resultado semelhante é encontrado em Tamanine (2010, pág. 131), quando a autora atesta que “verbo *ter* apresentou o maior número de ocorrências no âmbito dos verbos utilizados junto ao pronome *nós*, percentualmente sendo o verbo mais freqüente da amostra”. Abaixo é possível observar algumas das ocorrências que ilustram a realização dos pronomes associados ao verbo *ter* (134), (135) e (136):

- (134) Inf.: não tem saneamento... então aqui ***nós*** não ***tínhamos*** aqui na escola mas já fizeram né... no início da/do ano (Inq. 63 – F2 – EF, mulher, 61 anos)
- (135) Inf.: Além do Brasil... ***nós temos*** a ITÁlia... temos a::... a Colômbia... né? que:: num sei se a Bolívia acho que é a CoLÔMbia que... que es/... que surpreENdeu está um time muito bom... (Inq. 08 – F1 – DID, homem, 40 anos)
- (136) Inf.: .. quer dizer... ***NÓS tivemos*** educação dos nossos pais e a gente passoua

para os nossos meninos... os meninos respeitam... os meninos respeitam o magro o feio esquelético o aleijado (Inq. 86 – F2 – D2, mulher, 58 anos)

Além disso, é importante destacar que a atuação do *verbo ter* gerou um valor de  $p > 0,001$ , sendo este estatisticamente relevante, pois se encontra dentro do limite estabelecido pelo valor alfa. Isso indica forte correlação entre o *verbo ter* e a escolha entre os pronomes *nós* e *a gente*. Os demais verbos, *discendi* e epistêmico, não demonstraram efeito significativo sobre a variável resposta, sendo o valor de  $p$  atribuído a esses verbos maior que 5%, limite arbitrado nesta pesquisa para determinar interação significativa.

*e. Atuação do tempo verbal sobre alternância pronominal de primeira pessoa do plural*

A variável *tempo verbal* foi apontada na regressão logística como estatisticamente significativa. Nossa hipótese sobre essa variável previa maior probabilidade de *a gente* no tempo verbal *pretérito imperfeito* (Mendonça, 2010; Araujo, 2016; Fernandes, 2021; Silva (2022)). Os dados confirmam nossa hipótese a respeito da influência dessa variável sobre a realização dos pronomes de primeira pessoa do plural, já que os valores em *logodds* dispostos na tabela 16 apontam o *pretérito imperfeito* como contexto favorecedor para a realização da forma pronominal *a gente*, com estimativa de 3,29167 e valor  $p > 6.53e-10$ , sendo até este momento da análise o fator que mais beneficia o uso do pronome inovador. O valor de estimativa para esta variável é ainda maior que o valor do intercepto do modelo (1.52171), indicando alto favorecimento do pronome inovador neste contexto.

Com o objetivo de explicar a motivação de verbos no *pretérito imperfeito* se inclinarem ao uso do pronome *a gente*, analisamos características da nossa amostra, buscando padrões que pudessem ser atribuídos ao resultado, considerando nossa hipótese. Nos inquéritos analisados, provenientes de gravações realizadas nos anos de 1990, observamos que os informantes frequentemente narram suas experiências vivenciadas e que, em sua maioria, aconteciam repetidas vezes no passado, expressando uma ideia de recorrência, características que estão associadas aos verbos no *pretérito imperfeito*. Isso ocorre principalmente em inquéritos do tipo DID e D2, como nas amostras (137), (138) e (139):

(137) Inf.: eu queria que cê visse ...as coisa que *a gente fazia* NA ESCOLA mesmo porque *a gente* num tinha condição de alugar clube né? (Inq. 07 – F1 – D2, mulher, 26 anos)

- (138) Inf.: tu se lembra que *a gente brincava* de carimba também lá no Joaquim Nogueira.. eu gostava daquele tempo da nossa esco::la era bom (Inq. 55 – F1 – D2, mulher, 46 anos)
- (139) Inf.: já em noventa e oito *a gente chegava* nos barzinho e lá é mulher com mulher homem com homem era normal (Inq. 50 – F2 – DID, homem, 56anos)

O próprio projeto PORCUFORT, baseados nos pressupostos variacionista, adotou o método das experiências pessoais como instrumento metodológico na coleta de dados. Isso porque, para Labov (2001), os dados mais interessantes para análise linguística provem de narrativa de experiências pessoais, pois o informante desvia a atenção de sua própria fala e deixa o seu vernáculo emergir, evitando o policiamento à fala e dando espaço a utilização das variantes não padrão. Em concordância a Labov, Tarallo (1999) afirma que:

o estudo das narrativas de experiência pessoal tem demonstrado que, ao relatá-las, o informante está tão envolvido emocionalmente com o que relata que presta o mínimo de atenção ao como. E é precisamente esta a situação natural de comunicação almejada pelo pesquisador-sociolinguista (Tarallo, 1999, p. 22).

Devido à natureza dos inquéritos usados como base de dados nesta pesquisa, especialmente os do tipo DID e D2, é normal que haja alta frequência de verbos relacionados a narrativa. Nesse sentido, as narrativas pessoais, com frequência, se utilizam de verbos no pretérito, que não expressam um limite temporal, ou seja, que indicam ações que estavam em andamento no passado, sem um tempo final, sendo estas características do *pretérito imperfeito*. Neves (2000) discute a relação entre os verbos imperfeitos e o tipo textual narrativo, pontuando que esses verbos são utilizados, no geral, para “situar o leitor no tempo e no espaço da narrativa, apresentando-lhe o quadro em que se desenrolam os fatos” (Neves, 2000, p. 549). Nesse sentido, uma das explicações para o fato da variante *a gente* ser favorecida com verbos no tempo *pretérito imperfeito* está relacionada, diretamente, à natureza da base de dados de onde extraímos nossa amostra, que, utilizando-se de narrativas de experiências pessoais, criam ambientes férteis para este tipo de tempo verbal, em que há predomínio da variante inovadora.

Nossos resultados<sup>17</sup> se comparam aos obtidos por Seara (2000), Lopes (2003) e Mattos (2013), que demonstram favorecimento do pronome *a gente*, tanto em frequência de uso quanto em peso relativo, diante de verbos no pretérito imperfeito. Em todos os estudos, os pesos relativos referentes ao uso da forma *a gente* junto a verbos no passado imperfeito

<sup>17</sup> Em Seara (2000), Lopes (2003) e Mattos (2013), a unidade de medida adotada como referência para determinar favorecimento, diferente desta pesquisa, é o Peso Relativo (P.R.), resultado da análise multivariada realizada através do programa computacional GoldVarb.

apresentam peso relativo acima de 0.500 (ponto de neutralidade), o que aponta favorecimento da forma pronominal inovadora. Seara (2000), obteve resultados que indicam maior frequência de uso de *a gente* entre os verbos no pretérito imperfeito, totalizando 86% das ocorrências e 0,680 de peso relativo. De forma ainda mais acentuada, Lopes (2003) percebeu alta probabilidade do uso de *a gente* para esse tempo verbal, frequência de 88% e peso relativo de 0.760 para uso de *a gente*. Por fim, Mattos (2013), verificou que o pronome inovador em sua amostra é favorecido quando o verbo está no pretérito imperfeito, com 88% das ocorrências e peso relativo de 0,640. A tendência apresentada nesses trabalhos se repete em nossos dados, de maneira que o *pretérito imperfeito* favorece expressivamente o uso da variante inovadora, e inibe significativamente o emprego da variável padrão.

A outra explicação que justifica esse resultado pode ser a proposta por Souza (2022), que acredita na associação entre o emprego do *pretérito imperfeito* e contextos de indeterminação. Como já demonstramos durante a análise dos pronomes em função do grau de referencialidade (ver tabela 8), em nossa amostra, encontramos expressivo favorecimento do pronome inovador *a gente* em contextos de indeterminação do referente. Considerando que o *pretérito imperfeito* pode estar associado à indeterminação (Souza, 2022), pode ser justificável o favorecimento do pronome inovador em contextos de pretérito imperfeito.

O *futuro do indicativo* é outro tempo verbal que apresenta comportamento que beneficia o uso da forma pronominal inovadora *a gente*, atuando de forma discreta em termos de favorecimento. Conforme apresentado na tabela 16, o valor estimado para os verbos no futuro é de 2,18983, indicando que, em relação ao *intercept* (1,52171), há maior probabilidade de uso do pronome inovador *a gente*. O *p-value* que o modelo logístico atribuiu a esse fator foi de  $p = 0.01992$ , valor que está dentro do intervalo de confiança e, portanto, demonstra associação estatisticamente significativa entre a realização dos pronomes e os verbos que expressam futuro. Em termos percentuais, nossos resultados estão de acordo com os de Ribeiro (2020), que, em amostras do português contemporânea falado em Goiás, detectou distribuição equilibrada entre as formas *nós* e *a gente* no futuro. Em relação ao favorecimento, apesar do equilíbrio, o autor identificou leve inclinação ao uso da forma *a gente* associada a verbos no futuro, com peso relativo de 0,580, confirmando nosso resultado sobre atuação desse fator. Abaixo as ocorrências (140) e (141) exemplificam este nível:

- (140) Inf.: e... outro autor que *nós iremos* ver... que é::... o Antero de Quental... aí tem aí Antero de Quental na página cento e quarenta e três. (Inq. 36 – F1 – EF, homem, 24 anos)

- (141) Inf.: não... *a gente vai usar* o carro mas depois pode mandar procurar que *a gente vai deixar* num canto fácil prá você achar... Tá? (Inq. 76 – F2 – D2, homem, 38 anos)

No entanto, diferente dos nossos, os resultados de Omena (1998, 2003), ao analisar a variação entre os pronomes *nós* e *a gente* no falar carioca, apontam os verbos no *futuro* como inibidores do pronome *a gente*, sendo estes francos favorecedores do pronome padrão. Para a autora, o tempo verbal pode estar associado à mudança de *nós* por *a gente*, tendo em vista que os tempos verbais “mais marcados (como passado e futuro) tendem a refrear a mudança; os menos marcados (formas nominais e presente) a impulsionam” (Omena, 2003, p. 70). Nesse sentido, o presente (0.59) e as formas nominais (0.72), condicionam o pronome inovador *a gente*, enquanto tempos *passados* (0.31) e *futuros* (0.38), apresentam tendência ao favorecimento do pronome *nós*.

Em relação às formas nominais dos verbos, após tratamento dos dados, com a remoção das observações conflituosas e as de multicolineariedade, nos restaram apenas ocorrências da forma nominal *infinitivo*. Como é possível observar na tabela 16, os resultados apontam para interação significativamente relevante entre verbos desta natureza e a realização dos pronomes pessoais de primeira pessoa do plural ( $p\text{-value} = 0.01596$ ). Além disso, o valor de estimativa atribuído a atuação desse fator em nosso modelo de regressão (4.79578) indica que a forma pronominal *a gente* é fortemente favorecida em contextos de emprego de verbos no *infinitivo*. Esse resultado confirma a hipótese de Omena (2003), de que verbos menos marcados se inclinam à maior realização da variável inovadora.

Lopes (2003), em estudo sobre a alternância pronominal *nós* e *a gente* com amostra de fala culta do Rio de Janeiro, apresenta resultados parecidos aos nossos a respeito da atuação de verbos no *infinitivo* sobre a realização dos pronomes de primeira pessoa do plural. Assim como ocorreu em nossos resultados, a autora identificou intenso favorecimento do pronome inovador junto a formas nominais (gerúndio: 0.75, *infinitivo*: 0.65), sendo este o favor que mais condiciona o uso de *a gente* entre os tempos verbais. Em nosso caso, a estimativa de uso do pronome *a gente* com verbos no *infinitivo*, assim como Lopes, apontou esse fator como maior favorecedor da forma *a gente* em nossa amostra. Desta forma, quando se trata da variação pronominal *nós/a gente* em relação ao uso do *infinitivo*, prevalecem construções do tipo:

- (142) Inf.: é o que eu me lembro mais dele ((som do vento)) me lembro muito *a gente ir* pro um navio...quando o navio da::va Fortaleza (Inq. 16 – F2 – DID, homem, 28 anos)

- (143) a... expressão do Barroco... exemplo museus e bibliotecas... existem vão:: sempre ter... o que **a gente pesquisar** sobre a obra barroca... e ainda a questão do pré-barroco né?... (Inq. 35 – F1 – DID, mulher, 30 anos)

Por fim, outro fator do tempo verbal que atua no favorecimento do pronome *a gente* é o *futuro do pretérito* (ver tabela 16). A análise do modelo logístico realizada pelo programa estatístico identificou estimativa de uso para esse fator de 2,5981, apontando para um alto favorecimento, em relação ao *intercept* (1,52171), da variante inovadora. Além disso, o *p-value* = 0.01506 demonstra interação estatisticamente significativa entre esse tempo verbal e a variável resposta, o que significa que os verbos no futuro do pretérito interferem na realização dos pronomes, no sentido de beneficiar a forma *a gente*. Há de se considerar, no entanto, que o alto favorecimento do pronome inovador nesse contexto pode ser explicado em função da proporção de uso na distribuição das ocorrências de verbos no futuro do pretérito entre os pronomes, tendo em vista que, em nossa amostra, identificamos apenas 30 observações dos pronomes de primeira pessoa do plural associados a verbos no futuro do pretérito, das quais 19 ocorrências são do pronome *a gente* (63,3%), e 11 correspondem a realizações do pronome *nós* (36,7%). O excerto (144) e (145) exemplificam a atuação deste tipo de verbo.

- (144) Inf.: tivemos essa vantagem né?... de:... bem:: mostrar que através do esFORço próprio né? do estudo tal **a gente chegaria** ((ruído)) mais além né? (Inq. 12 – F1 – DID, mulher, 41 anos)
- (145) aqui nós vamos ter o conTI-nuum que se desenVOLve e o contínuo reestruturador... NEssa primeira coLUNA... **nós teríamos** o... o:: JARGÃO né?... o PIDGIN estável... que passaria depois que a gente já viu né? (Inq. 138 – F1 – EF, mulher, 31 anos)

#### f. Atuação da *saliência fônica* sobre alternância pronominal de primeira pessoa do plural

O princípio da *saliência fônica* diz respeito ao grau de diferenciação no material fônico do verbo na relação singular/plural (Lopes, 2003). Trabalhos que analisam os pronomes *nós* e *a gente* à luz da Teoria da variação e mudança linguística apontam para uma restrição no uso dos pronomes relacionada ao grau de facilidade na percepção entre as formas verbais com e sem a desinência formal de primeira pessoa do plural (*-mos*). De acordo com Lopes (2003), quanto mais acentuada for a diferença entre as formas singular e plural, maior a probabilidade de utilização do pronome padrão *nós*. Da mesma forma, Borges (2004), ao analisar o uso dos

pronomes de primeira pessoa do plural, verificou que a forma *a gente* é preferida nos menores níveis de saliência (quando há menor diferenciação fônica), enquanto a forma padrão *nós* é favorecida nos níveis mais altos de saliência (quando a diferença fônica é acentuada).

Nesse mesmo caminho, seguindo o proposto pelos autores, tínhamos a hipótese de que o pronome inovador seria favorecido em contextos de diferenciação fônica pouco acentuada. Destacamos, neste ponto, que atribuímos ao nível 7 de saliência fônica (é/somos) o valor de referência (*intercept*) para que fosse possível procedermos a comparação entre os níveis mais altas e mais baixos de saliência e verificar os ambientes mais propensos a realização do pronome *a gente*. Cabe lembrar também que os dados que representavam o nível 1 de saliência, dados de gerúndio, foram retirados, devido a problemas com multicolineariedade.

Dito isto, conforme exposto na tabela 16, os níveis mais baixos de saliência fônica não apresentam associação estatisticamente significativa com a realização dos pronomes de primeira pessoa, sendo os valores de *p* atribuídos aos níveis 2 e 3 maiores que 5%, ou seja, estão fora do intervalo de confiança atribuído a esta pesquisa. Sendo assim, mesmo que as estimativas revelem, em relação ao *intercept*, desfavorecimento da variante inovadora no nível 2 (cantar/cantarmos) e favorecimento desta variante no nível 3 (falava/falávamos), não podemos considerar esses resultados relevantes na nossa análise. Sendo assim, em termos de favorecimento, não há como confirmar nossas hipóteses sobre a atuação dos pronomes *nós* e *a gente* em realizações com verbos de menor saliência.

Em relação a verbos com nível 4 (fala/falamos) de saliência fônica, ou seja, verbos em que a diferenciação entre as formas singular e plural se faz a partir do deslocamento da sílaba tônica e acréscimo da desinência *-mos*, esses tendem a favorecer o pronome inovador *a gente*, com estimativa de uso de 2,5372, demonstrando maior atuação sobre a realização do pronome inovador. O *p-value* atribuído a este fator também demonstra forte associação entre o uso das variantes e verbos com saliência nível 4. Borges (2004), em pesquisa sobre o comportamento da variação pronominal *nós/ a gente* nas cidades de Pelotas e Jaguarão, concluiu que “a forma *a gente* é mais favorecido nos níveis em que não há, ou há menor diferenciação de material fônico, e nos níveis nos quais a oposição não é tônica ou é tônica em apenas uma das formas” (Borges, 2004, p. 139). No caso dos verbos com saliência nível 4, apenas uma das formas apresentam oposição tônica (canta x cantamos), ambiente favorecedor do pronome inovador em nossos dados, fato que corrobora a hipótese do autor.

Um ponto importante de destacar sobre essa variável é que os verbos de saliência nível 4, em maioria, são verbos no presente do indicativo, o que atesta correlação entre a saliência fônica e o tempo verbal. O presente do indicativo, assim como o pretérito perfeito do

indicativo, pode gerar ambiguidade temporal nos verbos, tendo em vista que existem verbos que apresentam a mesma forma na relação presente/passado (Scherre et al, 2018), sendo necessário recorrer a advérbios ou ao contexto para inferir o seu tempo verbal, pois a marca formal junto ao verbo não é suficiente para definir o seu tempo. Para ilustrar a realização dos verbos com mesmo paradigma, segue as ocorrências (146) e (147), extraídas de nossa amostra de fala culta, proveniente do projeto PORCUFORT Fase I (década de 1990):

- (146) Inf.: ao adentrar lá no reCINto... que eu cheguei eh:: *nós entramos* tinha uma rede armada uma criança que tinha uns quatro cinco meses era... o bracinho engesAado... (Inq. 56 – F2 – D2, homem, 50 anos)
- (147) Inf.: foi criada a Guarda Vermelha então como *nós falamos*... a Guarda Vermelha... era:... praticaMENte... ela mantinha... uma identidade de um sistema (Inq. 17 – F1 – EF, homem, 33 anos)

As ilustrações (146) e (147) demonstram que os verbos empregadores apresentam ambiguidade temporal, no sentido de não haver clareza quanto ao seu tempo (se presente ou se passado), pois possuem a mesma forma em ambos os tempos. Nesse sentido, Scherre, Yacovenco & Naro (2018) acreditam que a forma inovadora, ao se implementar na língua, pode desfazer a ambiguidade potencial existente entre esses dois tempos verbais, criando ambiente fértil a utilização do pronome inovador. Isso explicaria o fato deste nível 4 de saliência fônica agir no sentido de favorecer o pronome *a gente*, conforme demonstra o resultado da análise do nosso modelo de regressão logística (ver tabela 16). Para confirmar a hipótese dos autores, decidimos observar o paradigma verbal do modo indicativo em nossos dados e averiguar se, em termos percentuais, há maior inclinação ao uso da forma *a gente* para verbos com mesmo paradigma, ou seja, se o uso de *a gente*, em nossa amostra, pode ser uma estratégia para evitar a ambiguidade nesses verbos. O resultado está disposto na tabela 17:

**Tabela 17:** Distribuição de *nós* e *a gente* por paradigma verbal do modo indicativo

Anos 1990 (N = 1603)				
Paradigma verbal	Nós	%	A gente	%
Presente igual	93	27,15%	245	72,5%
Presente diferente	334	53,2%	294	46,8%
Pretérito igual	87	49,7%	88	50,3%
Pretérito diferente	97	66%	50	34%
Imperfeito	86	27,3%	229	72,7%
$\chi^2 (4) = 125.74, p < 0,001$				

Fonte: Elaboração própria

Os resultados obtidos através do levantamento dos dados de paradigma verbal do modo indicativo em nossa amostra (ver tabela 17) parecem confirmar o prescrito por Scherre et al (2018). A respeito dos verbos com mesmo paradigma na relação presente/passado, os dados da tabela 17 apontam para maior proporção uso do pronome inovador *a gente*, com 72,5% das ocorrências no presente, e 50,3% no pretérito perfeito com verbos do mesmo paradigma. Esse resultado, de acordo com os autores, pode apontar o uso do pronome *a gente*, nesse contexto, como estratégia para desfazer a ambiguidade temporal atrelada a esses verbos, ou seja, em verbos que tem o mesmo paradigma para o presente e o passado, o pronome inovador agiria como mecanismo para marcar o tempo verbal, tendo em vista que essa ambiguidade está relacionada à marca morfológica de plural -mos, sendo esta desinência fortemente associada ao pronome *nós*, enquanto a marca morfológica  $\emptyset$  (ausência da desinência -mos) está relacionada, de forma quase categórica, ao pronome *a gente*. Com isso, o uso da forma pronominal inovadora, nesses contextos, condicionaria a queda a desinência formal de primeira pessoa do plural -mos, evitando assim a ambiguidade potencial desses verbos.

Em relação aos verbos monossílabos tônicos ou oxítonos que, no plural, passam a ser paroxítonos (saliência fônica nível 5), nossos resultados confirmam os encontrados em Lopes (2003) que, em amostras da fala culta do Rio de Janeiro, identificou os monossílabos com acréscimo da desinência -mos como contextos favorecedores da forma padrão da língua, tendo apresentado peso relativo de 0.65 para uso do pronome *nós*. A autora atribui o resultado ao grau de diferenciação fônica acentuado entre as formas desse tipo de verbo, uma vez que a desinência -mos, estando associada ao pronome padrão, favoreceria seu uso. Em concordância com Lopes, Scherre et al (2018), reconhecendo que a marca morfológica -mos condiciona o uso do pronome padrão, identificou tendência ao uso preferencial de *nós* com a desinência -mos em verbos que apresentam formas distintas para o passado e o presente, ou seja, os verbos “sem possibilidade de ambiguidade, majoritariamente de oposição singular/plural mais saliente (*nós foi/nós fomos, nós tem/nós temos; nós vai/nós vamos; nós é/nós somos*)” (Scherre et al, 2018, p. 14).

Em nossa amostra representativa da fala culta de Fortaleza dos anos de 1990, a tendência identificada pelos autores em relação aos verbos que apresentam paradigmas verbais diferentes para presente/passado se confirma. Os dados da tabela 17 demonstram maior proporção de uso do pronome *nós* em verbos com paradigmas diferentes, sendo mais intensa em verbos no pretérito perfeito (66% das ocorrências), e um pouco mais discreto entre os verbos no presente do indicativo (53,2%). Em termos de favorecimento, em nosso modelo de regressão, os verbos com saliência fônica nível 5 (monossílabos tônicos ou oxítono que passam a

paroxítonos) desfavorecem o uso do pronome *a gente* em comparação com o *intercept*, apontando favorecimento do pronome padrão neste contexto. Na mesma direção, Franchescini (2011, p. 149), que adotou a mesma escala de saliência fônica adotada nesta pesquisa, encontrou resultados que confirmam os nossos, tendo em vista que, ao analisar a fala de Concórdia – SC, “os resultados apresentam os níveis 3 e 4 de saliência favorecendo *a gente* e os níveis 5 e 6 favorecendo *nós*”.

Em nossos dados, os verbos com saliência nível 5 (tem/temos) correspondem, em sua maioria, ao verbo *ter* (*a gente tem/ nós temos*), que representa 22,03% do total de realizações dos pronomes, e verbos de estado (*a gente está/ nós estamos*), com proporção de 8,36% do geral. Os dados da tabela 16, como já discutido quando tratamos da variável tipo de verbo, apontam os verbos dos tipos *ter* e de estado como os mais propícios a ocorrência do pronome padrão em nossa amostra, fato que também explicaria a atuação deste nível de saliência fônica no sentido de favorecer o pronome *nós*. Os níveis 2, 3 e 6 dessa variável não apresentaram interação estatisticamente significativa com a realização dos pronomes, tendo em vista que o valor de *p* atribuído a esses níveis estão fora do intervalo de confiança arbitrado como 5%.

### 5.1.3 Síntese dos resultados – A alternância pronominal na Amostra 1990

Em termos gerais, identificamos em amostra de fala culta dos anos 1990, maior proporção de uso do pronome inovador *a gente* (58,3%) e uso menos recorrente do pronome padrão *nós* (41,7%). Entre as variáveis que foram testadas neste estudo, se mostraram estatisticamente significativas para a realização dos pronomes a *faixa etária*, o *tipo de inquérito*, o *grau de referencialidade do pronome*, o *tipo de verbo*, o *tempo verbal* e o *grau de saliência fônica*.

Em relação à *faixa etária*, houve maior preferência pela variante *a gente* entre os informantes de faixas etárias mais jovens, com 63,9% das realizações para os informantes de 22 a 35 anos, e 66,6% para os de 36 a 50 anos. Já os informantes com mais de 51 anos apresentaram tendência ao uso pronome canônico *nós*, com percentual de 61%. Esse resultado parcial aponta para um possível indício de mudança linguística.

Na variável *tipo de registro*, a forma *a gente* teve maior proporção de uso em inquéritos do tipo D2 (72,8%), o menos monitorado entre os registros. Em seguida, em inquéritos do tipo DID, o pronome inovador atingiu 61% das ocorrências da década de 1990, sendo os inquéritos do tipo EF os únicos a favorecerem a forma canônica *nós*, com 57,7% das

ocorrências. Esses resultados confirmam relação direta entre o uso dos pronomes e o grau de monitoramento da fala.

Os resultados referentes à variável *grau de referencialidade do pronome* demonstram que, na década de 1990, o uso dos pronomes mantém suas funções prototípicas. O pronome *a gente* no sentido genérico/indeterminado, apresentando-se em 71,3% das ocorrências, enquanto a forma canônica *nós* com uso menos frequente em contextos de indeterminação, apresentando o percentual de 28,7%. Já em uso com sentido específico/determinado, o pronome padrão *nós* teve uso mais recorrente, sendo responsável por 61,5% das ocorrências, enquanto o pronome *a gente* foi menos utilizado em referência específica, apresentando 38,5% das ocorrências.

Na variável *tipo de verbo*, apenas os verbos de estado e os verbos do tipo *ter* demonstraram interação estatisticamente significativa com a realização dos pronomes de primeira pessoa do plural. Em relação aos verbos que expressam estado, houve uso mais frequente do pronome padrão, 57,4% de uso do pronome *nós*, e 42,6% de uso do pronome inovador *a gente*. Além dos verbos de estado, outro contexto que parece beneficiar o pronome padrão é o *verbo ter*, apresentando 70,3% dos casos para o pronome *nós*, enquanto o pronome *a gente* figura em 29,7% das realizações de primeira pessoa. Isso demonstra que o pronome canônico *nós* é favorecido em contextos de utilização do verbo *ter*, sendo o principal fator, em proporção de uso, a beneficiar o pronome canônico em nossa amostra.

Em relação ao *tempo verbal e formas nominais do verbo*, estas variáveis se mostraram relevantes para a análise, no sentido de favorecer o pronome inovador *a gente*, os verbos no *pretérito imperfeito* (72,9%), *futuro do subjuntivo* (87,5%) e na forma nominal *infinitiva* (82,5%). Ao contrário, agindo significativamente a favorecer o pronome *nós*, a análise estatística apontou o *futuro do indicativo* como aliado da forma pronominal padrão (51,4%);

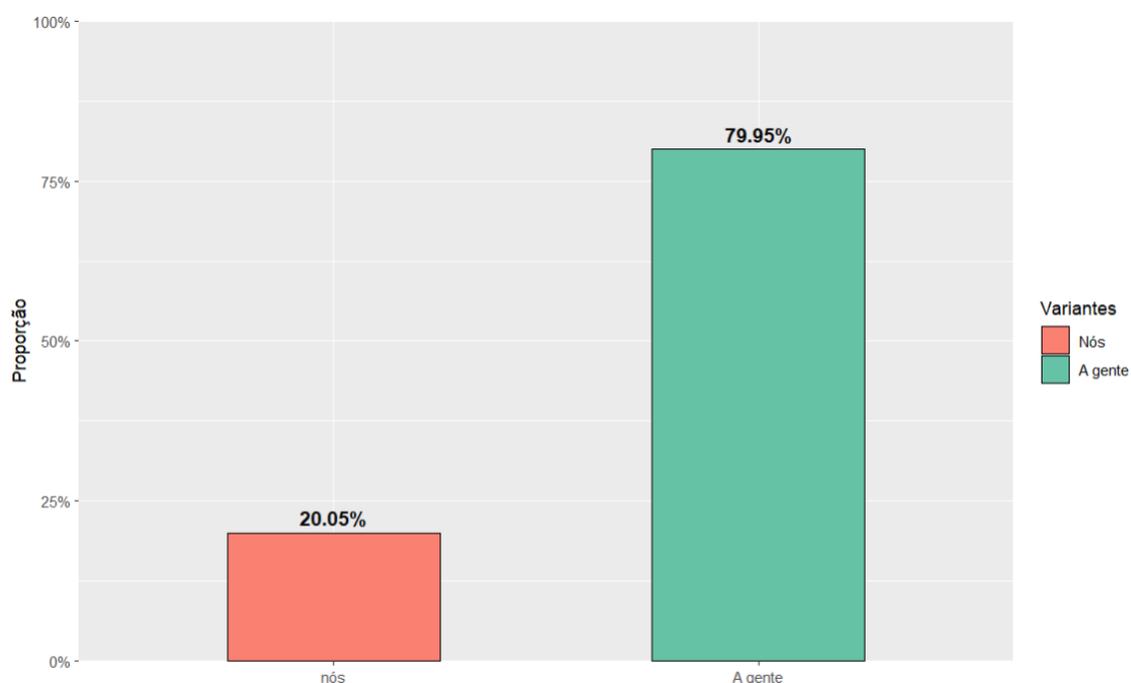
Por fim, sobre a variável *saliência fônica*, os verbos de saliência Nível 4 (fala/falamos) e Nível 5 (tem/temos) foram os únicos níveis que demonstraram atuação relevante com a realização dos pronomes *nós* e *a gente*. No nível 4, a forma *a gente* foi preferida em 69,9% das ocorrências, enquanto o pronome *nós* se realizou em apenas 30,1% dos casos. Já nos verbos de saliência fônica *nível 5*, o pronome *nós* ocorreu em 51,6% dos casos, superando a proporção de uso da forma pronominal *a gente*, que se realizou em 48,4% das ocorrências. É exatamente nesse contexto que o pronome padrão passa a ser usado em maior proporção que o pronome inovador, mesmo que a diferença ainda seja discreta.

## 5.2 A alternância pronominal de primeira pessoa do plural no falar culto de fortaleza: análise descritiva dos dados de fala da década de 2020

Nesta seção, apresentaremos a análise descritiva dos dados de fala referente a década de 2020. Como explicado na seção que trata da metodologia (ver seção 4.5), os dados referentes a Amostra 2020 foram coletados com o auxílio do pacote *dmsocio* (Oushiro, 2018), através do programa computacional *RStudio*, que gerou uma planilha editável no formato .csv com as variáveis sociais elencadas neste estudo. Posteriormente, procedemos à atribuição das variáveis linguísticas de forma manual para o tratamento dos dados coletados.

Após análise dos dados conduzida por meio do *software RStudio*, nossa amostra representativa da década de 2020 revelou 2618 observações dos pronomes de primeira pessoa do plural. Desse total, 525 dados correspondem à realização do pronome *nós*, enquanto o pronome *a gente* foi registrado em 2093 observações. Notavelmente, a variante inovadora foi expressivamente mais utilizada que a variante padrão em nossa amostra do PORCUFORT da década de 2020. O gráfico 12 ilustra visualmente a proporção de uso de cada pronome na amostra pesquisada.

**Gráfico 12:** Proporção de uso dos pronomes *nós* e *a gente*: Amostra 2020 (N=2618)



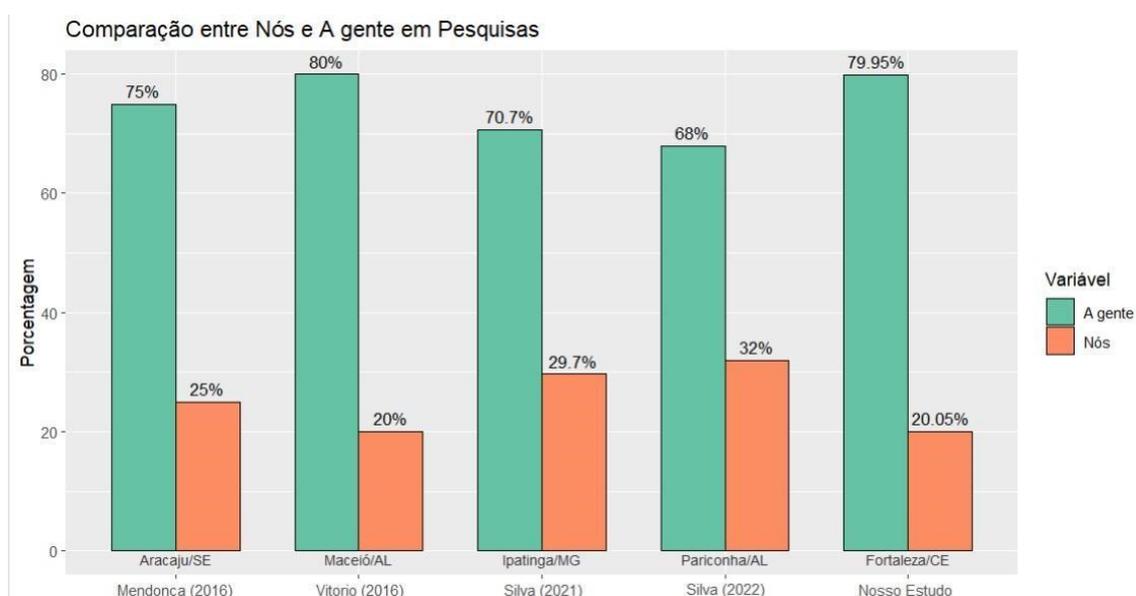
Fonte: elaboração própria.

Os resultados gerais da análise, em termos percentuais, revelam significativa

preferência pelo pronome inovador na língua, o pronome *a gente*, entre os fortalezenses de ensino superior completo, com 79,95% do total de dados. Em relação ao pronome mais antigo na língua, o pronome *nós*, considerado padrão, este se mostrou menos recorrente em nossa amostra, tendo se realizado em 20,05% das ocorrências de primeira pessoa do plural. Os resultados gerais com as proporções dos pronomes *nós* e *a gente* em dados da década de 2020 reafirmam nossa hipótese inicial sobre o comportamento geral dos pronomes entre os falantes cultos de Fortaleza, tendo em vista que prevíamos maior proporção de uso para o pronome inovador e menor proporção do pronome padrão.

Essa inclinação ao maior uso do pronome inovador em detrimento do pronome padrão demonstra uma tendência no uso dos pronomes nas décadas de 2010 e de 2020 que vem sendo atestadas por outras pesquisas que envolvem amostras de língua oral. Isso se torna evidente ao compararmos o comportamento do fenômeno em amostras de diversas comunidades de fala do Brasil. Trabalhos sociolinguísticos que analisam dados de fala representativos das décadas de 2010 e 2020 (Mendonça, 2016; Fernandes, 2021; Silva, 2021; Rodrigues, 2022; Silva, 2022), ressaltando que os *corpora* utilizados possuem critérios<sup>18</sup> diferentes, apontam maior uso de *a gente* e uso menos recorrente de *nós* em diversas partes do país, conforme podemos observar no gráfico 13:

**Gráfico 13:** Comparação entre proporções de uso de *nós* e *a gente* em estudos variacionistas



<sup>18</sup> O projeto Norma Urbana Culta – NURC, por exemplo, estabelece como critério para seleção dos informantes que estes tenham ensino superior completo. Por outro lado, o projeto Censo Linguístico - CENSO realizou coleta de dados com informantes de ensino fundamental e de médio completos, mas não com informantes de nível superior.

Fonte: Elaboração própria.

Conforme podemos observar no gráfico 13, os dados de fala de várias regiões no Brasil apontam uso expressivo do pronome *a gente* nos anos de 2010/2020 e, em contrapartida, baixa ocorrência do pronome *nós*. Esse padrão na alternância pronominal traduz uma característica no comportamento dos pronomes na década mais atual, tendo em vista que os percentuais de cada variante se assemelham em todas as pesquisas consultadas. Nesse sentido, os resultados obtidos neste estudo corroboram com os de Mendonça (2016), Fernandes (2021), Silva (2021) e Silva (2022).

Mendonça (2016), em seu estudo sobre a variação na expressão da primeira pessoa do plural com dados provenientes da amostra de dados de fala de estudantes do Atheneu Sergipense, identificou alta proporção de uso do pronome *a gente* entre os estudantes do ensino médio, com percentual de uso de 75%. O pronome *nós*, por exemplo, se realiza em apenas 25% dos dados, demonstrando grande preferência dos estudantes pelo pronome inovador. De forma parecida, Vitorio (2021) analisou a alternância entre os pronomes de primeira pessoa do plural a partir de onze entrevistas realizadas durante o ano de 2010 com informantes de nível superior completo residentes de Maceió, capital do estado de Alagoas, e observou maior recorrência da forma pronominal *a gente* (80%) e baixa ocorrência do pronome *nós*, tido como padrão (20%). Com a análise dos dados de fala culta maceioense, o autor concluiu que “gradativamente o pronome *a gente* tem ocupado o lugar do pronome *nós*” na fala culta de Maceió (Vitorio, 2016, p. 170).

A mesma tendência no uso dos pronomes é observada na pesquisa desenvolvida por Silva (2021), ao investigar a variação entre os pronomes de primeira pessoa na fala de 12 informantes, sendo seis deles com formação superior completa. A análise apontou maior uso de *a gente*, com 70,7% dos dados de fala, enquanto o pronome *nós* se realizou apenas em 29,3% dos casos. Por fim, o estudo de Silva (2022), utilizando base de dados composta por 45 entrevistas de falantes da zona rural de Pariconha, em Alagoas, demonstra que a mesma tendência no que concerne ao uso dos pronomes encontrada na fala dos informantes mais escolarizados também é encontrada na fala dos informantes menos escolarizados. Em sua pesquisa, a linguista identificou maior proporção de uso de *a gente* (68%), em detrimento do pronome canônico *nós* (32%), destacando que “o comportamento da comunidade se assemelha ao comportamento de outras comunidades” (Silva, 2022, p. 85).

Em termos de tendência, o uso expressivo da forma inovadora em todas as

comunidades de fala das quais fizemos levantamento biográfico demonstram que o fenômeno em estudo, na década de 2020, ainda se configura como uma regra variável, nos termos de Labov (1969), ou seja, nestas comunidades, não há evidências de mudança linguística completa, estando a referência à primeira pessoa do plural ainda em disputa entre os pronomes *nós* e *a gente* pela preferência dos falantes. O mesmo ocorre entre os falantes cultos fortalezenses, que, apesar da preferência pela forma pronominal *a gente* (79,95%), estes ainda utilizam o pronome padrão *nós* de forma relativamente expressiva (20,05%), o que configura um ambiente de disputa entre as formas. Os *supertokens* (148) e (149), extraídos de nossa amostra representativa da década de 2020 do Projeto PORCUFORT, atestam o uso intercambiável entre os pronomes ainda presente na fala culta fortalezense:

(148) Inf.: *nós* fomos visitar o rio Jaguaribe mostrá-la... eu e as internas *a gente* tinha passeios passeios da gente era no Jaguaribe... né e aí nisso *a gente* pegou... *nós* pegamos um barco e atravessamos e ficamos numa croinha... (Inq. 29 – F2 – DID, mulher, 62 anos)

(149) Inf.: e na hora de escrever... que é para poder atingir as pessoas que estão fora do contexto onde *nós* onde *nós* estamos entendendo entende?... porque *a gente* começa a fortalecer na nossa forma de falar em forma de desabafo (Inq. 37 – F2 – DID, homem, 37 anos)

Nos excertos (148) e (149), é possível perceber o uso dos pronomes *nós* e *a gente* com o mesmo referente, demonstrando ser esse ambiente de disputa entre as variantes. Na ocorrência exposta no exemplo (148), tanto o pronome *nós* quanto o pronome *a gente*, por exemplo, se referem a “eu e as internas”, o que torna possível a intercambialidade entre esses pronomes. Já na ocorrência (149), as duas formas pronominais em disputa são empregadas com grau moderado de indeterminação, generalizando o referente, ambiente igualmente propício à intercambialidade entre os pronomes.

Tendo observado o comportamento geral do fenômeno variável e discutido a proporção de uso dos pronomes em nossa amostra, nos detemos na próxima seção a investigar a atuação das variáveis previsoras sobre a realização da variável resposta, através de testes estatísticos e de um modelo de regressão logística, que nos forneceu informação sobre como cada variável favorece ou desfavorece a utilização dos pronomes na fala culta de Fortaleza da década de 2020. Relembramos, neste ponto, que os dados de fala da Amostra 2020 foram observados em função das mesmas variáveis preditoras adotadas para a análise dos dados de fala da década de 1990.

### 5.2.1 Atuação das variáveis predictoras em termos de frequência e proporção de uso: a alternância pronominal de primeira pessoa do plural na Amostra 2020

Com o propósito de investigar o papel da atuação de cada variável predictor, de caráter linguístico e social, sobre a realização da variável resposta, ou seja, sobre a realização dos pronomes de primeira pessoa do plural *nós* e *a gente*, conduzimos testes estatísticos de qui-quadrado para testar a relação entre a *alternância pronominal* e cada predictor por nós elencado. Como já tratado na seção anterior (ver 5.1.1), o teste de qui-quadrado de Pearson auxilia na identificação da relação entre as variáveis, determinando o grau de interação entre elas (Oushiro, 2021).

Tratamos os dados referentes à amostra representativa da fala culta de Fortaleza nos anos de 2020 do PORCUFORT com o auxílio da ferramenta estatística *R Studio*, que nos forneceu valores de qui-quadrado, proporções e frequências de uso correspondentes a atuação de cada variável predictor sobre a alternância entre os pronomes *nós* e *a gente*. Dentre as variáveis de ordem social, a primeira que testamos foi *sexo/gênero*, conforme podemos observar na tabela 18:

**Tabela 18:** Proporção de nós e a gente por *sexo/gênero*: Amostra 2020 (N= 2618)

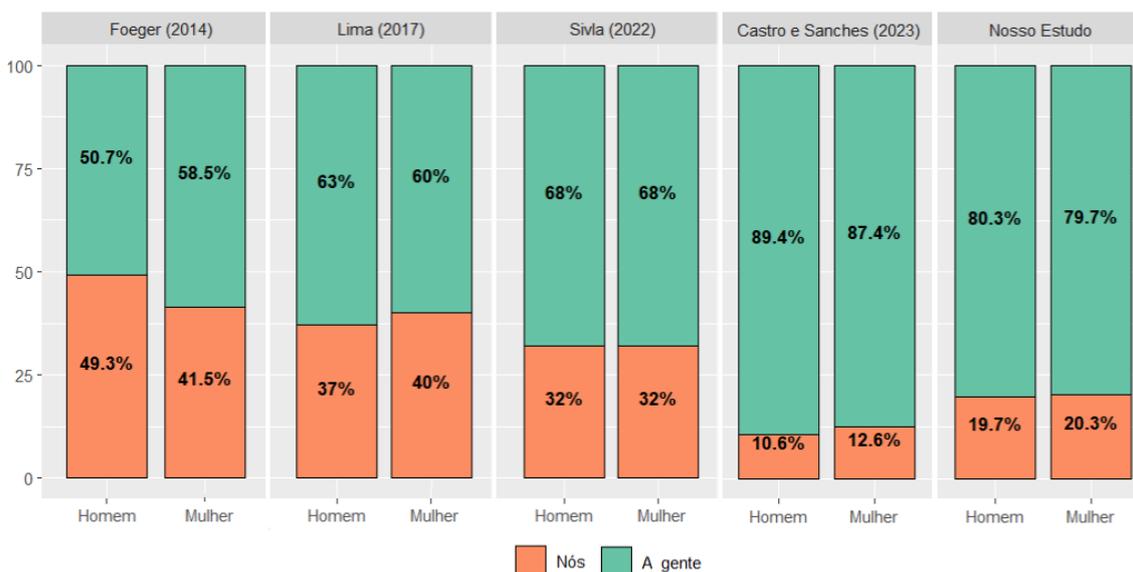
PORCUFORT II - Anos 2020 (N = 2618)				
<i>Sexo/Gênero</i>	<i>Nós</i>	%	<i>A gente</i>	%
<b>Homem</b>	214	19,7%	871	80,3%
<b>Mulher</b>	311	20,3%	1222	79,7%
$\chi^2 (1) = 0.093147, p < 0,7602$				

Fonte: elaboração própria.

As informações apresentadas na tabela 18 indicam uma certa simetria na proporção de uso dos pronomes de primeira pessoa do plural, independente do *sexo/gênero* do informante que, nos dados do PORCUFORT Fase II, apresentam comportamentos semelhantes, ou seja, nos informantes de ambos os sexos o pronome inovador *a gente* se destaca. Entre homens, há maior proporção de uso do pronome *a gente* (80,3%), e uso pouco acentuado do pronome *nós* (19,7%). De forma parecida, as mulheres tendem a usar mais o pronome inovador (79,9%) que o pronome padrão (20,3%). Resultados próximos ao nosso foram encontrados por Foeger (2014), Lima (2021), Silva (2022), e Castro e Sanches (2023) quando identificaram comportamentos linguísticos parecidos entre homens e mulheres no uso dos pronomes de primeira pessoa do plural, a partir da análise de amostras representativas da década de 2010/2020. Nesse sentido, “homens e mulheres apresentam percentuais semelhantes em relação

ao uso de *a gente* – 68%, logo não existem diferenças relevantes entre o sexo masculino e feminino” (Silva, 2022, p. 65). O gráfico abaixo compara os resultados desta investigação com as pesquisas supracitadas:

**Gráfico 14** – Comparação no uso de *nós* e *a gente* entre homens e mulheres em outros estudos



Fonte: elaboração própria.

Conforme exposto no gráfico 14, o uso dos pronomes em amostras de fala das décadas de 2010/2020 segue uma tendência no sentido de homens e mulheres apresentarem o comportamento parecidos no que se refere a proporção de uso da *alternância pronominal de primeira pessoa do plural* em diversas localidades do Brasil. Por exemplo, Foeger (2014), em investigação sobre o uso dos pronomes *nós* e *a gente* na fala de Santa Leopoldina, através de entrevistas gravadas entre 2011 e 2013, indica maior uso da forma inovadora em ambos os sexo, sendo um pouco mais intenso entre as mulheres (58,5%) se comparado com os homens (50,7%).

De maneira parecida, o estudo de Lima (2017) sobre os padrões de alternância entre os pronomes *nós* e *a gente* em entrevistas sociolinguísticas gravadas na cidade de Luziania/GO, com dados representativos da década de 2010, aponta preferência pelo pronome *a gente*, tanto entre os homens (50,7%), quanto entre as mulheres (58,5%). Segundo o autor, na comunidade de fala investigada, há equilíbrio no uso das variantes de primeira pessoa do plural em informantes de ambos os sexos, não existindo, portanto, diferença significativa entre os sexos, levando o linguista a concluir que “o fator sexo não é determinante na escolha de uma das variantes” (Lima, 2017, p. 13).

Seguindo a tendência observada entre os informantes do sexo masculino e feminino, o trabalho de Silva (2022) sobre a alternância pronominal na comunidade de fala de Periconha, em Alagoas, demonstra uso acentuado de forma pronominal *a gente*, em detrimento do pronome considerado padrão, independente do sexo do informante. Em seu estudo, o autor encontrou comportamento linguístico relacionado ao uso dos pronomes de primeira pessoa do plural semelhante entre os entrevistados do sexo masculinos e feminino, sendo a forma *a gente* responsável por 68% das ocorrências, enquanto o pronome canônico *nós* se realizou em apenas 32% dos dados coletados. Através da análise, o autor concluiu que não existe diferença estatisticamente significativa no uso dos pronomes em relação ao sexo do informante, uma vez que o teste estatístico realizado sugeriu falta de interação entre a variável *sexo/gênero* e a realização dos pronomes ( $p = 0.935$ ).

Por fim, Castro e Sanches (2023) procederam a análise dos padrões de alternância pronominal de primeira pessoa do plural na comunidade de Mazagão Velho/AP, com dados provenientes de entrevistas sociolinguísticas gravadas no ano de 2021, mesma época em que foi gravada parte de nossa amostra. Os resultados apontam para a mesma direção dos resultados encontrados em nosso estudo, ou seja, expressiva utilização da variante inovadora *a gente*, e baixa ocorrência da variante padrão *nós*. Entre os homens, a proporção de uso da forma pronominal inovadora é de 89,4%, expressivamente mais utilizada que o pronome padrão, que obteve apenas 10,6%. As mulheres apresentam comportamento parecido quanto ao uso dos pronomes de primeira pessoa, tendo em vista sua expressiva preferência pelo pronome *a gente* (87,4%) em relação ao pronome *nós* (12,6%). Assim como as outras pesquisas abordadas, na comunidade de fala de Mazagão Velho “o fator sexo parece não condicionar o uso dos pronomes investigados” (Castro e Sanches, 2023, p. 50).

Os nossos dados da Amostra 2020 revelam também não haver diferenciação significativa no uso dos pronomes a depender do *sexo/gênero* do informante, tendo em vista que, tanto entre homens quanto entre as mulheres, há maior inclinação ao uso do pronome inovador *a gente*. Essa proposição é confirmada através do teste estatístico de qui-quadrado, que apontou valor de 0.093147, com um grau de liberdade, e nível de significância  $p$  igual a 0,7602, estando acima do valor alfa (intervalo de confiabilidade) adotado nesta pesquisa de  $p > 0,05$ , ou 5%. Este resultado sugere que não há evidências estatisticamente significativas para rejeitar a hipótese nula de que não há associação entre o *sexo/gênero* do informante e a realização dos pronomes de primeira pessoa do plural. Isso é um indicativo de que, possivelmente, esta variável será excluída do modelo de regressão logística, durante a análise através da função *step* ().

Tendo apresentado os resultados do teste de estatísticos sobre a atuação da variável social *sexo/gênero*, seguimos com a exposição dos resultados da próxima variável previsor de natureza social testada neste estudo. Trata-se da variável *faixa etária*, que vem sendo apontada como forte condicionante na escolha entre os pronomes *nós* e *a gente* em estudos sociolinguísticos. A tabela 19 apresenta os dados de frequência e proporção de uso das variantes de primeira pessoa do plural em função das três faixas de idade controladas neste estudo, assim como o resultado do teste de qui-quadrado de Pearson, que aponta a significância da interação entre a variável previsor e a variável resposta.

**Tabela 19:** Distribuição de *nós* e *a gente* por *faixa etária*: Amostra 2020 (N=2618)

PORCUFORT II - Anos 2020 (N = 2618)				
Faixa Etária	Nós	%	A gente	%
22 a 35 anos	77	11,5%	591	88,5%
36 a 50 anos	116	13,6%	739	86,4%
51 anos ou mais	332	30,3%	763	69,7%
<b><math>\chi^2 (2) = 124.17, p &lt; 0,001</math></b>				

Fonte: elaboração própria

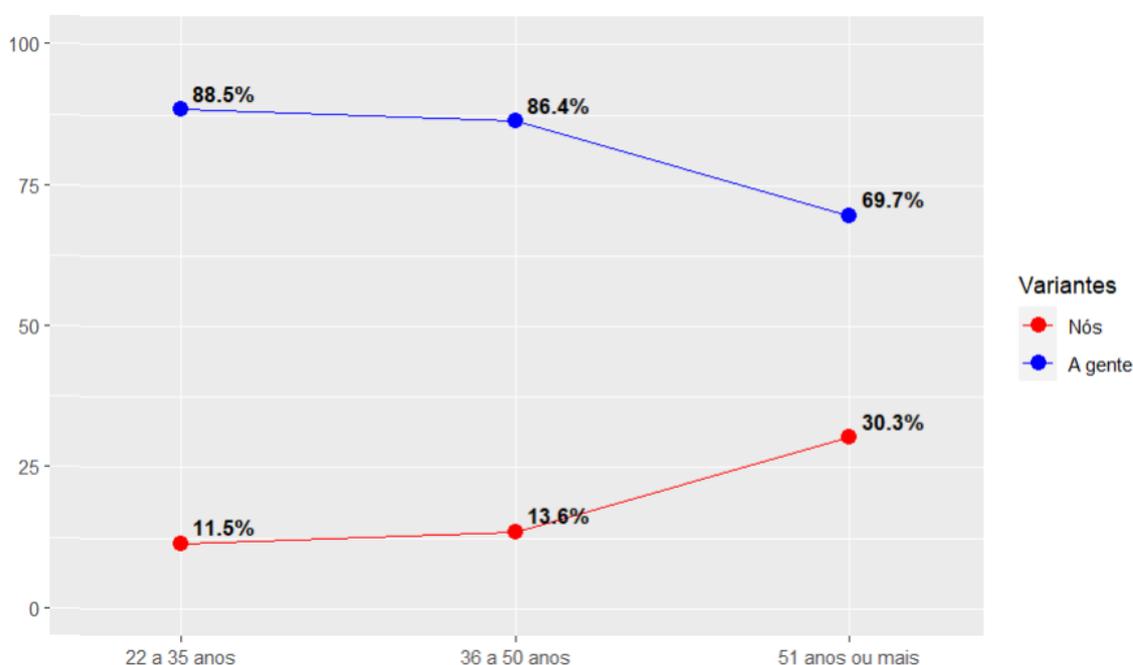
Em termos gerais, conforme apresentado na tabela 19, a análise com os dados representativos da fala da década de 2020 aponta forte inclinação dos mais jovens pelo uso da variante inovadora *a gente*, o que confirma a nossa hipótese inicial sobre a atuação dos pronomes na Amostra 2020. Como vemos, os informantes de 22 a 35 anos, os mais jovens da amostra, apresentam alta proporção de uso do pronome inovador *a gente*, tendo obtido 88,5% do total de dados de primeira pessoa do plural, sendo o pronome *nós* pouco utilizando entre esses informantes, com apenas 11,5% das observações.

Os informantes de faixa etária intermediária, de 36 a 50 anos, apresentam comportamento linguístico, em relação ao uso dos pronomes de primeira pessoa, parecido com o comportamento dos mais jovens, sendo a variante inovadora usada com maior proporção. Em termos percentuais, o levantamento dos dados de fala da década de 2020 apontou forte inclinação dos informantes com essa faixa etária ao uso de *a gente*, sendo responsáveis por 86,4% das realizações de primeira pessoa do plural, enquanto o pronome padrão *nós* tem baixa proporção de uso entre esses informantes, com percentual de uso de 13,6% das ocorrências, número parecido ao apresentado pelos mais jovens.

O comportamento do fenômeno entre os mais velhos, os informantes com mais de 50 anos de idade, indica favorecimento do pronome *a gente*, resultado que também confirma nossa hipótese, pois esperávamos que o pronome inovador seria utilizado em maior proporção

em todas as faixas etárias representativas do falar culto fortalezense de 2020. Nesse sentido, o pronome *a gente* é utilizado em 69,7% das observações entre os informantes mais velhos, ao passo que o pronome *nós* é utilizado em 30,3% dos casos. Um ponto importante de destacar é que, entre os mais velhos, a proporção de uso do pronome canônico *nós* é quase o triplo da proporção apresentada pelos mais jovens, isso sugere forte interação entre a faixa etária e a realização dos pronomes. Os resultados com a proporção de cada pronome em função das três faixas etárias podem ser conferidos no gráfico 15:

**Gráfico 15:** Proporção das variantes por *faixa etária* na Amostra 2020



Fonte: elaboração própria

Conforme é possível perceber no gráfico 15, a medida em que a faixa etária aumenta, o uso do pronome *a gente* diminui, sendo esta uma diminuição discreta se compararmos os informantes de faixa I e da faixa II, e um pouco mais expressiva se comparadas as proporções de uso entre os informantes mais velhos da amostra. Uma tendência inversa é percebida ao analisar o comportamento do pronome tido como padrão, uma vez que, à medida que a faixa etária aumenta, também aumenta a proporção de uso do pronome *nós*, conforme demonstrou a análise em nossa base de dados. Esses dados sugerem indícios de mudança linguística em progresso, nos termos da teoria da variação e mudança linguística (Labov, 2008 [1977]).

De acordo com a teoria laboviana, uma mudança em progresso pode ser observada quando há diferença significativa no uso das formas linguística entre as gerações. Neste caso,

a forma pronominal *a gente* é usado em maior proporção entre os mais jovens, e em menor proporção entre os informantes mais velhos. Esses resultados sugerem que a forma padrão está se tornando cada vez menos popular na fala dos mais jovens, que destacam o uso da forma inovadora, enquanto os mais velhos parecem apresentar maior resistência à implementação de *a gente*, mesmo com a significativa proporção de uso (69,7%). Esses resultados estão de acordo com a teoria variacionista, que prevê que as mudanças linguísticas, muitas vezes, são impulsionadas pelos mais jovens da comunidade de fala (Labov, 2008 [1977]).

Além disso, a análise estatística realizada por meio do teste de qui-quadrado de Pearson apontou para interação significativamente relevante entre o uso dos pronomes de primeira pessoa do plural e a faixa etária dos informantes da amostra investigada. Nesse caso, o valor de qui-quadrado obtido no teste é  $\chi^2 = 124.17$ , com dois graus de liberdade e valor de  $p < 0,001$ , indicando forte associação entre as variáveis, uma vez que o valor de  $p$  é inferior ao nível de significância previamente estabelecido nesta pesquisa ( $p > 0,05$ ). Com o resultado deste teste estatístico, temos evidências suficientes para rejeitar a hipótese nula de que não há interação entre a variável faixa etária e a realização dos pronomes *nós* e *a gente*, e acatar a hipótese alternativa de que há associação relevante entre as variáveis. Ou seja, os resultados demonstram que a escolha entre os pronomes pode estar associada a idade do falante, o que destaca a importância de se considerar fatores de natureza social na análise da variação linguística.

A fim de investigar a influência do grau de formalidade sobre a alternância entre os pronomes pessoais de primeira pessoa do plural, decidimos testar o *tipo de inquérito*, a última variável de ordem extralinguística por nós elencada. Os resultados referentes a atuação dessa variável previsora, em termos de frequência e proporção de uso, bem como o valor de qui-quadrado podem ser conferidos na tabela 20:

**Tabela 20:** Distribuição de *nós* e *a gente* por *tipo de inquérito*: Amostra 2020 (N=2618)

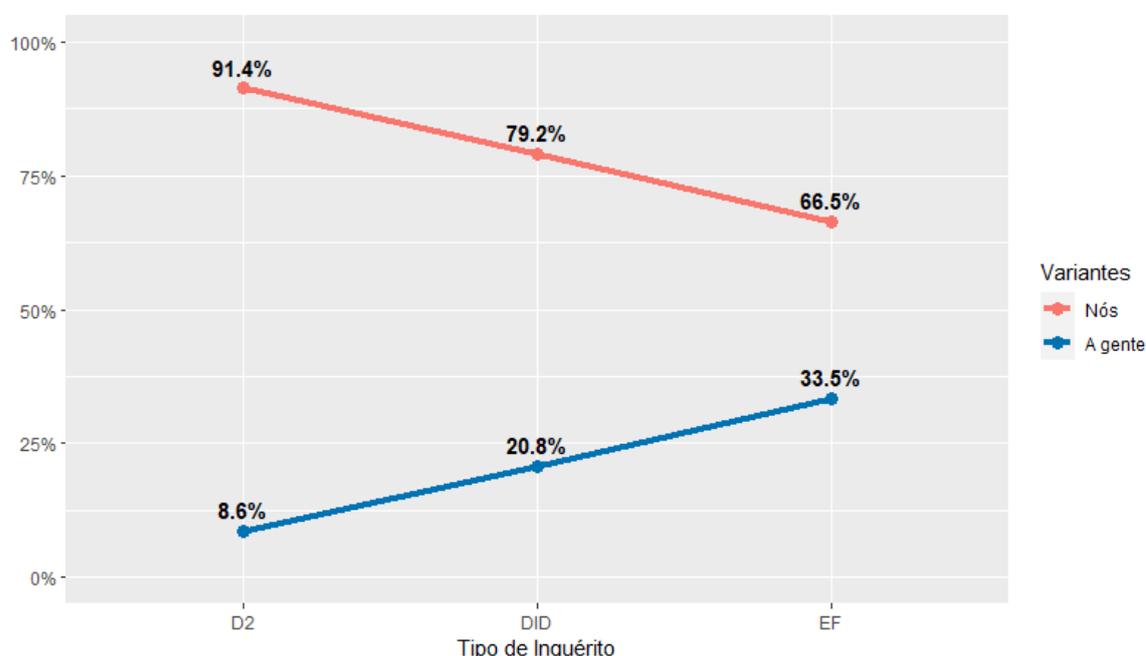
<b>PORCUFORT II - Anos 2020 (N = 2618)</b>				
<b><i>Tipo de Inquérito</i></b>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
<b>D2</b>	75	8,6%	792	91,4%
<b>DID</b>	225	20,8%	855	79,2%
<b>EF</b>	225	33,5%	446	66,5%
<b><math>\chi^2 (2) = 146.76, p &lt; 0,001</math></b>				

Fonte: Elaboração própria

Conforme demonstram os dados presentes na tabela 20, o uso do pronome *a gente* é mais abundante que o uso de *nós* em amostra do PORCUFORT Fase II (anos 2020). Essa

abundância é observada em todos os tipos de inquéritos, sendo mais intensa nos inquéritos do tipo Diálogo entre dois informantes (D2), e menos acentuada entre os inquéritos classificados como Elocução Formal (EF). Nesse sentido, nossa hipótese sobre a atuação do tipo de inquérito prescrevia que a proporção de uso da variante inovadora diminuiria conforme o nível de formalidade aumentasse, seguindo a seguinte direção D2 => DID => EF. Nossa hipótese pode ser confirmada através dos resultados, que estão dispostos visualmente através do gráfico abaixo:

**Gráfico 16:** Proporção de uso de *nós* e *a gente* por tipo de inquérito na Amostra 2020



Fonte: elaboração própria.

Os resultados apontam o uso do pronome *a gente* próximo do categórico em inquéritos do tipo D2, tendo 91,4% do total de ocorrências, enquanto o pronome *nós* apresenta proporção de uso de apenas 8,6%. Esse resultado confirma nossa hipótese de que os inquéritos categorizados como D2 estão mais propensos a realização do pronome inovador, tendo em vista a menor formalidade presente nestes registros (Araújo; Viana; Pereira, 2018). De maneira geral, as variantes inovadoras são usadas com maior proporção em estilos menos formais, enquanto as conservadoras são mais produtivas em estilos mais formais (Labov, 1972).

Em relação aos inquéritos do tipo DID, estes também parecem ser francos aliados no uso da forma pronominal *a gente*, tendo ocorrido em 79,2% do total de realizações de primeira pessoa do plural em inquéritos deste tipo. Em contraste, o pronome padrão *nós*

apresenta percentual de uso consideravelmente menor, que corresponde a apenas 20,8% das observações em nossos dados. Esse achado evidencia uma tendência no que se refere a preferência pelo uso da variante inovadora em detrimento do pronome padrão na amostra de fala representativa de 2020.

Embora ocorra de forma menos acentuada, os dados presentes nos inquiridos classificados como Elocução Formal (EF) indicam maior proporção de uso do pronome inovador, com percentual de uso de 66,5% do total de observações. Por outro lado, o pronome padrão *nós* ocorre em menor proporção em elocuições formais, sendo responsável por 33,5% dos dados, fato que contraria nossa hipótese sobre a atuação dessa variável, tendo em vista que esperávamos que o pronome padrão fosse mais recorrente em elocuições formais devido à natureza mais formal desse tipo de inquirido.

A primeira variável de natureza linguística controlada nesta pesquisa é o *grau de referencialidade dos pronomes*. Trabalhos realizados (Lopes, 2003; Tamanine, 2010; Araujo, 2016; Souza, 2022) à luz da teoria da variação e mudança linguística revelam que a escolha entre os pronomes *nós* e *a gente* é motivada por uma série de variáveis, incluído o grau de referencialidade, que exerce um papel importante nesse processo de escolha entre os pronomes.

O uso intercambiável entre as formas *nós* e *a gente* só é possível devido à capacidade desses pronomes de assumirem o mesmo significado referencial, ou seja, ambos fazem referência a primeira pessoa do plural no discurso. No entanto, o emprego dessas variantes pode apresentar diferenças semânticas quanto ao grau de referencialidade que elas representam. Isso ocorre porque os pronomes de primeira pessoa do plural podem assumir diferentes referências, podendo ser específicas ou genéricas/indeterminadas (Omena, 1996).

Lopes (2003), por exemplo, em seu estudo sobre a alternância pronominal em dados de fala culta carioca, descreveu quatro possibilidades de interpretação semântica atribuídas aos pronomes de primeira pessoa: a) o falante e o seu (s) interlocutor (es); b) o falante e outros indivíduos, excluindo o interlocutor; c) o próprio falante isoladamente, sendo possível o emprego como *eu*; e d) uma referência genérica, interpretada como *todos* ou *qualquer um*. Nos resultados de sua análise, a linguista observou referência [+ genérica] associada ao pronome *a gente*, enquanto o pronome *nós* era preferencialmente usado com referência [+ específica]. Com o objetivo de observar essa tendência sobre nossos dados, a tabela 21 apresenta os percentuais e proporções de uso dos pronomes *nós* e *a gente* distribuídos em função do *grau de referencialidade do pronome*:

**Tabela 21:** Distribuição de *nós* e *a gente* segundo a referencialidade: Amostra 2020 (N=2618)

PORCUFORT II - Anos 2020 (N = 2618)				
Referencialidade	Nós	%	A gente	%
Genérico	225	16,6%	1127	83,4%
Específico	300	23,7%	966	76,3%
$\chi^2 (1) = 19.858, p < 0,001$				

Fonte: elaboração própria.

Os resultados apresentados na tabela 21 se aproximam, em partes, dos encontrados por Lopes (2003), tendo em vista a alta proporção de uso do pronome inovador *a gente*, empregado com sentido genérico, correspondendo a 83,4%. Esse resultado indica a manutenção da referencialidade prototípica da forma pronominal. Por outro lado, a proporção de uso do pronome padrão *nós*, em contextos de referência genérica, corresponde a apenas 16,6% dos dados. Esse fato confirma nossa hipótese sobre a atuação dos pronomes em relação ao grau de referencialidade, uma vez que esperávamos poucas ocorrências do pronome padrão pouco associado às referencialidades genéricas ou indeterminadas.

Em relação ao emprego de pronomes com referencialidade específica, em nossa amostra representativa da fala culta fortalezense da década de 2020, esperávamos maior uso do pronome *nós*, conforme demonstram estudos sociolinguísticos sobre a *alternância pronominal de primeira pessoa do plural* (Lopes, 1993; Omena, 2003; Silva, 2013; Araujo, 2016). No entanto, a tabela 21 revela baixa proporção de uso do pronome padrão com referente específico, representando apenas 23,7% do total de realizações. Em contraste, o pronome *a gente*, mesmo com referente específico, é responsável por 76,3% dos dados obtidos em nossa análise. Esse resultado sugere uma ampla expansão do pronome inovador na década de 2020 em contextos de referencialidade específica, que antes figurava como um ambiente linguístico de resistência à entrada do pronome inovador, mas que agora é preferida entre os falantes, independente do grau de referência.

Para Seara (2000), inicialmente, a forma pronominal *a gente* era considerada menos específica, sendo empregada mais efetivamente em asserções genéricas. Entretanto, essa característica está gradativamente se dissipando, tendo em vista o uso cada vez mais frequente da forma em contextos de especificação. O autor aponta uma mudança envolvendo o pronome *a gente* que, de acordo com os dados obtidos em sua investigação sobre o uso dos pronomes na fala de Florianópolis, explicaria a expansão do pronome inovador em diversos contextos, como os de referencialidade específica:

Ao que parece, o significado dessa variante (*a gente*), quer semântica, quer gramaticalmente, se modificou, pois acresceu-se ao significado originalmente

indeterminado a referência à pessoa que fala, passando assim à forma determinada e, quanto à visão gramatical, ela deixa de ser substantivo para integrar o sistema de pronomes pessoais, apesar de sua não inclusão, até o presente, no quadro de pronomes nas gramáticas tradicionais. (Seara, 2000, p. 184).

Os resultados da tabela 21, em termos de proporção de uso, sugerem pouca influência dessa variável sobre a escolha entre as formas pronominais de primeira pessoa do plural, já que há expressiva preferência por *a gente* em todos os graus de referencialidades testados neste estudo. No entanto, o teste estatístico de qui-quadrado de Pearson resultou em um valor de  $\chi^2 = 19.858$ , com um grau de liberdade e *p-value*  $< 0,001$ , apontando interação estatisticamente significativa desta variável previsora sobre a realização da variável resposta. Com base nesse resultado, é possível descartar a hipótese nula de que não há influência da referencialidade sobre a escolha entre os pronomes *nós* e *a gente*. Em vez disso, devemos aceitar a hipótese alternativa de que há associação relevante entre as variáveis e, por isso, a variável previsora *grau de referencialidade* deve ser incluída no modelo de regressão logística.

A próxima variável linguística apresentada é o *tipo de verbo*. O controle desta variável tem como referência a pesquisa de Tamnine (2002), que analisou dados de fala das cidades de Lages, Blumenau e Chapecó, e de Araujo (2016), que investigou o comportamento dos pronomes em função do tipo de verbo em amostra da fala popular fortalezense. Os resultados podem ser conferidos na tabela 22.

**Tabela 22:** Distribuição de *nós* e *a gente* por *tipo de verbo*: Amostra 2020 (N=2618)

<b>PORCUFORT II - Anos 2020 (N = 2618)</b>				
<i>Tipo de Verbo</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
<b>Ação</b>	283	16,2%	1465	83,8%
<b>Estado</b>	63	26,5%	175	73,5%
<b>Dicendi</b>	18	13,5%	115	86,5%
<b>Epistêmico</b>	13	10,3%	113	89,7%
<b>Ter</b>	148	39,7%	225	60,3%
<b><math>\chi^2 (4) = 122.97 p &lt; 0,001</math></b>				

Fonte: Elaboração própria

Como é possível observar na tabela 22, o pronome inovador é mais produtivo que o pronome padrão em todos os tipos de verbo. Em relação aos verbos de ação, foram levantados 1465 dados, sendo este o tipo de verbo o mais produtivo da Amostra 2020. Quando associado ao verbo de ação, o pronome *a gente* é usado com alta proporção de uso, correspondendo a 83,8% das ocorrências, enquanto o pronome *nós* corresponde a apenas 16,2% das realizações. Esses resultados estão em conformidade com Tamnine (2002), que verificou maior probabilidade do uso de *a gente* com verbos que expressam ação.

Em relação aos verbos que expressam estado, os resultados encontrados em Tamanine (2002) e Franceschini (2011), apontam maior proporção de uso para o pronome padrão *nós*, contrariando os nossos resultados. A tabela 22 indica baixa proporção de uso do pronome *nós*, que corresponde a 26,5% dos dados de fala culta da década de 2020. Já o pronome *a gente*, quando associado a verbos de estado, apresenta elevada proporção de uso, sendo responsável por 73.5% das realizações de primeira pessoa. Esse resultado é explicado pelo comportamento geral da amostra analisada, tendo em vista que os dados gerais (ver gráfico 12) apontam para uso preferencial do pronome inovador.

Seguindo a tendência pela preferência ao uso do pronome *a gente*, os dados da tabela 22 apontam proporção de uso igual a 86,5% para o pronome inovador associado a verbos *dicendi*. Esses resultados corroboram com a pesquisa Franceschini (2011), que analisou a alternância entre os pronomes pessoais na fala da cidade de Concórdia/SC. Segundo o autor, a elevada frequência de uso atribuída a forma pronominal *a gente* está relacionada ao fato de que, nas entrevistas sociolinguísticas analisadas, os informantes frequentemente recorriam a esse pronome como estratégia para generalizar o sujeito, utilizando-se de verbos de fala. Por outro lado, o uso do pronome padrão apresenta baixa proporção de uso em nossa amostra, correspondendo apenas a 13,5% dos casos, refletindo uma tendência linguística que se manifesta em vários contextos de uso.

Os resultados da tabela 22 apontam os verbos epistêmicos como principais favorecedores da forma pronominal *a gente*, com percentual de uso de 89,7%. Em contraste, nossa amostra apontou baixa proporção de uso do pronome *nós*, tendo apenas 10,3% do total de realizações. Esse resultado se assemelha ao obtido na pesquisa de Franceschini (2011), que observou que a maior incidência do pronome inovador estava associada a realizações com verbos epistêmicos (75%), enquanto o pronome padrão apresentou pouca recorrência com este tipo de verbo (25%). Uma das explicações para a forte associação do pronome inovador aos verbos epistêmicos é o fato de que, em geral, esses apresentam ambiguidade temporal. Muitos verbos que indicam atividade mental possuem a mesma forma para o passado e o presente, como “achamos”, “pensamos” e “imaginamos”, por exemplo, e apresentam a mesma forma para o passado e para o presente. Nesse sentido, o pronome inovador agiria de forma a sanar a ambiguidade temporal, motivo pelo qual, possivelmente, há maior proporção de uso de *a gente* com verbos epistêmicos.

Por fim, ao analisar a variabilidade linguística entre os pronomes pessoais de primeira pessoa do plural e sua distribuição em função do verbo do tipo *ter*, esses são os mais recorrentes da amostra, pois, somente ocorrências do verbo *ter* somam 373 dados. As

informações presentes na tabela 22 revelam que a forma pronominal *a gente* apresenta o menor índice em comparação com os outros tipos de verbos controlados nesta pesquisa, sendo responsável por 60,3% do total de ocorrências. Em contrapartida, o pronome padrão *nós* em ocorrências com verbos do tipo *ter*, ainda que apresente proporção de uso menor que 50%, possui expressiva proporção entre os verbos investigados, com 39,7% dos dados de nossa amostra. Esse resultado confirma o de Tamanine (2010), que observou que os verbos do tipo *ter* são os mais frequentes em sua amostra.

Além do mais, o teste de qui-quadrado determinou valor de  $\chi^2 = 122.97$ , apresentando 5 graus de liberdade, e valor de  $p < 0,001$ , estando essa variável dentro do valor alfa. Isso significa que a atuação univariada deste favor é relevante sobre a escolha entre os pronomes. Mais à frente apresentaremos uma análise multivariada que poderá confirmar a relevância estatística desta variável em um contexto de interação entre todas as variáveis.

Conduzimos a análise sobre a alternância entre os pronomes *nós* e *a gente* com o objetivo de identificar possíveis associações entre os padrões de uso dos pronomes e o *paralelismo formal*. De acordo com Scherre (1998), a ocorrência de uma forma pronominal tende a condicionar a realização da mesma forma no contexto subsequente dentro da mesma série de realizações. Tomando essa proposição como base, nossa hipótese inicial parte do pressuposto de que a forma pronominal *a gente* é favorecida quando antecedida pelo mesmo pronome. Com o intuito de testar esta hipótese, realizamos análise quantitativa dos dados, que podem ser conferidos na tabela 23.

**Tabela 23:** Distribuição de *nós* e *a gente* segundo o *paralelismo*: Amostra 2020 (N=2618)

<b>PORCUFORT I - Anos 1990 (N = 1956)</b>				
<b><i>Paralelismo formal</i></b>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
<b>Ocorrência isolada</b>	184	25,7%	532	74,3%
<b>Primeira ocorrência da série</b>	111	19,1%	469	80,9%
<b>Segundo ou terceiro da série com manutenção da forma anterior</b>	158	13,4%	1025	86,6%
<b>Forma pronominal diferente em relação à anterior</b>	72	51,8%	67	48,2%
<b><math>\chi^2 (3) = 135.01, p &lt; 0,001</math></b>				

Fonte: elaboração própria

Os dados apresentados na tabela 23 apontam maior proporção de uso do pronome inovador em quase todos os contextos. Ao todo, trabalhamos com um total de 2618 dados de primeira pessoa do plural, sendo as ocorrências isoladas responsáveis por 716 realizações dos pronomes, das quais se verificou maior ocorrência do pronome *a gente*, sendo utilizado em 532

dados de primeira pessoa (74,3%), enquanto o pronome canônico *nós* foi utilizado em 184 observações (25,7%). Estes resultados confirmam os de Vitorio (2016) que, ao analisar a fala culta de Maceió, observou percentual de uso de 80% para a forma *a gente* e de apenas 20% para o pronome *nós* em contextos de ocorrência isolada.

Em relação *a primeira ocorrência da série*, esse fator também está associado ao uso do pronome inovador. Cabe destacar que o comportamento geral da amostra aponta maior inclinação ao uso do pronome *a gente*, o que explica a expressiva proporção de uso dessa forma pronominal empregada como primeiro em uma série de repetições. Conforme demonstra a tabela 23, o uso de *a gente* como primeira referência da série é de 80,9%, percentual significativamente maior que o uso do pronome inovador, que detém 19,1% do total de realizações de primeira pessoa do plural.

O emprego dos pronomes como *segunda ou terceira realização de uma série com a manutenção da mesma forma* usada imediatamente anterior também age de forma a beneficiar o uso do pronome inovador *a gente*, conforme tendência constatada na amostra de fala pesquisada. Os resultados apontam relevância da ação da variável *paralelismo formal* quanto a escolha entre as formas pronominais em variação. Nesse sentido, pesquisas demonstram maior tendência a repetição do mesmo pronome quando realizado mais de uma vez na série, conforme aponta Vitorio (2016):

tanto para o pronome *nós* quanto para o pronome *a gente*, a escolha da primeira forma pronominal condiciona a realização da forma subsequente, desencadeando, assim, uma série de repetições da mesma forma linguística, o que confirma a hipótese de que a preferência por determinada forma pronominal tende a exercer influência sobre as demais formas numa dada sequência discursiva. (Vitorio, 2016, p. 165).

Essa afirmação é confirmada em nossos dados, tendo em vista que o pronome *a gente* tende a se realizar em 86,6% quando a referência anterior é *a gente*, o que confirma a ideia de que marcas levam a marcas e zeros levam a zeros (Scherre, 1988). No que diz respeito à forma padrão *nós*, este segue a tendência geral da amostra representativa da década de 2020, sendo pouco recorrente, inclusive como segundo e terceira da série, mantendo o paralelismo, com percentual de uso igual a 13,4%.

Por fim, em relação *a forma pronominal diferente em relação à anterior*, ou seja, quando há quebra do paralelismo, este é o único contexto a não privilegiar a forma pronominal *a gente*. Nesse sentido, quando a referência a primeira pessoa do plural é antecedida por *nós*, pronome diferente do subsequente, observamos baixa realização do pronome inovador *a gente*. As ocorrências de *forma pronominal diferente em relação à anterior* são escassas, totalizando apenas 139 observações. Desse total, 67 ocorrências são do pronome *a gente* (48,2%) e 67

dados são do pronome *nós* (51,8%), apontando para uma quebra na tendência no que concerne ao uso dos pronomes pessoais de primeira pessoa, sendo o pronome padrão privilegiado nesses ambientes.

Além disso, análise estatística realizada por meio do teste de qui-quadrado de Pearson revelou associação estatisticamente significativa entre o uso dos pronomes de primeira pessoa do plural e a variável paralelismo formal, tendo apresentado resultado de  $\chi^2 = 135.01$ , três graus de liberdade e valor  $p < 0,001$ , indicando que a escolha entre as formas pronominais é motivada por fatores como o *paralelismo formal*. Resultados mais detalhados serão apresentados mais à frente por meio de análise multivariada, através do modelo de regressão logística.

O *tempo verbal* é outra variável previsora testada com frequência em estudos sociolinguísticos sobre a variação na alternância pronominal entre *nós* e *a gente* (Omena, 1996; Lopes, 1998; Araújo, 2016; Vitorio, 2016; Silva, 2020; Souza, 2020; Silva, 2022) . Nossa hipótese, assim como Lopes (1993), prevê maior proporção de uso do pronome em tempos verbais menos marcados, como presente. A fim de confirmar nossa hipótese, procedemos testes estatísticos que revelam a atuação isolada dessa variável sobre as escolhas dos pronomes. A tabela 24 apresenta os dados que demonstram a interação desta variável previsora.

**Tabela 24:** Distribuição de *nós* e *a gente* segundo o *Tempo verbal* e *formas nominais do verbo*: Amostra 2020 (N=2618)

<b>PORCUFORT II - Anos 2020 (N = 2618)</b>				
<i>Tempo Verbal</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
<b>Pres. Indicativo</b>	232	21,1%	869	78,9%
<b>Pret. Perfeito</b>	166	32,3%	348	67,7%
<b>Pret. Imperfeito</b>	54	8,4%	586	91,6%
<b>Futuro</b>	59	36%	105	64%
<b>Pres. Subjuntivo</b>	2	8,3%	22	91,7%
<b>Imp. Subjuntivo</b>	3	20%	12	80%
<b>Fut. Subjuntivo</b>	0	0,0%	16	100%
<b>Fut. do pretérito</b>	1	8,3%	11	91,7%
<b>Gerúndio</b>	2	9,1%	20	90,9%
<b>Infinitivo</b>	6	5,5%	104	94,5%
<b><math>\chi^2 (9) = 146.96 p &lt; 0,001</math></b>				

Fonte: elaboração própria

A tabela acima apresenta a distribuição da frequência e proporção dos pronomes *nós* e *a gente* em função do *tempo verbal*. Destacamos que, do total de 2618 dados de primeira pessoa, 2418 são referentes a verbos no modo indicativo, o que representa 92,47% do total de dados, representando a grande maioria das ocorrências de primeira pessoa em nossa amostra.

Nesse sentido, o pronome inovador *a gente* é favorecido, em termos de proporção de uso, em todos os tempos verbais do modo indicativo, sendo mais intensamente no pretérito imperfeito, com proporção de uso da forma *a gente* de 91,6%, enquanto o pronome *nós*, nesse contexto, se realizou apenas em 8,4% dos dados. Esse resultado se assemelha ao de Tamanine (2010) e ao de Fernandes (2021), quando observaram o pretérito imperfeito do indicativo como favorecedor do uso de *a gente*.

Em relação ao presente do indicativo, se observa na tabela 24, proporção de uso mais elevada para o pronome *a gente* (78,9%), em detrimento do pronome *nós* (21,1%). Além desses tempos, se observou maior uso do pronome no pretérito perfeito (67,7%) e no futuro do presente (64%). Apesar da porcentagem de uso mais elevada da forma pronominal *a gente* nestes dois tempos verbais, é importante destacar que, entre todos os fatores testados nesta variável, os tempos pretérito perfeito e futuro do presente foram os que apresentaram menor proporção de uso do pronome inovador *a gente*.

O modo subjuntivo também se mostrou ambiente fértil para a ocorrência da forma *a gente* na Amostra 2020, sendo o futuro do subjuntivo categórico para o uso do pronome inovador, correspondendo a 100% das ocorrências neste contexto linguístico. De maneira parecida, nos tempos verbais presente do subjuntivo e pretérito imperfeito do subjuntivo há uso mais recorrente da forma pronominal inovadora, apresentando proporções de uso iguais 91,7% e 80%, conforme demonstram os dados apresentados na tabela 24.

Apontando para a mesma direção, os resultados referentes à atuação do futuro do pretérito sobre os pronomes pessoais de primeira pessoa revelam uso próximo do categórico para a realização do pronome *a gente* (91,7%), e pouca ocorrência do pronome *nós* (8,3%), seguindo a tendência geral da amostra, que apontou uso significativo do pronome inovador em detrimento do pronome padrão. Esse resultado também é observado em Borges (2004) que, ao investigar a inserção de *a gente* na comunidade de fala de Pelotas, identificou o futuro do pretérito fortemente associado ao pronome inovador.

Em relação aos verbos em suas formas nominais, percebemos a mesma propensão ao uso do pronome inovador encontrada nos resultados referentes aos outros tempos verbais. Verbos no gerúndio apresentam proporção de uso elevada para o pronome *a gente*, sendo responsável por 90,9% das ocorrências nestes contextos. Já o pronome *nós* figura discretamente, tendo percentual de uso igual a 9,1%. No mesmo sentido, os verbos no infinitivo são francos aliados, em termos de proporção de uso, da forma inovadora, obtendo 94,5% das ocorrências associadas a essa forma verbal. Esses resultados corroboram os achados de Vianna e Lopes (2015, p. 112), ao evidenciarem que, nestes contextos, “é mais provável o emprego de ‘a gente’

– sendo categórico com o gerúndio e altamente favorecido com o infinitivo”. Realizações dos pronomes associados a verbos no particípio não foram encontradas na amostra de fala do PORCUFORT da década de 2020.

A análise do teste estatístico aplicado para testar a relevância da variável *tempo verbal* sobre a alternância pronominal, realizado por meio do qui-quadrado de Pearson, apontou valor  $\chi^2 = 146.96$ , com 9 graus de liberdade e valor de  $p < 0,001$ . Destacamos que o *p-value*, estando dentro do intervalo de confiança adotado nesta pesquisa, destaca forte interação entre a variável *tempo verbal* e a escolha entre os pronomes. No entanto, é necessário considerar que esse resultado corresponde à atuação isolada desta variável previsora. Mais à frente, realizaremos testes multivariados para testar a influência desta variável em interação com as outras variáveis testadas nesta pesquisa, a fim de construir um modelo de regressão logística que explique a variação entre os pronomes *nós* e *a gente*.

A última variável de natureza linguística testada nesta pesquisa foi a *saliência fônica*, variável amplamente utilizada em estudos sobre a alternância pronominal. Nossa hipótese sobre a atuação dessa variável é a de que, quanto menos saliente for o verbo, maior seria a proporção dos pronomes *a gente*. Em oposição, quanto maior for a saliência fônica do verbo, mais favorável seria a utilização do pronome padrão *nós*, tendo em vista que “formas mais salientes, por provocarem maior grau de percepção, tendem a favorecer a ocorrência de plural no verbo” (Mattos, 2013, p. 75). A tabela 25 apresenta a distribuição dos pronomes em cada nível de saliência fônica:

**Tabela 25:** Proporção de *nós* e *a gente* segundo a *saliência fônica*: Amostra 2020 (N=2618)

<b>PORCUFORT II - Anos 2020 (N = 2618)</b>				
<i>Saliência Fônica</i>	<b>Nós</b>	<b>%</b>	<b>A gente</b>	<b>%</b>
<b>Nível 1 - (cantando)</b>	2	9,1%	20	90,9%
<b>Nível 2 - (cantar/cantamos)</b>	6	5,3%	108	94,7%
<b>Nível 3 - (falava/ falávamos)</b>	58	8,5%	623	91,5%
<b>Nível 4 - (fala/ falamos)</b>	70	12,5%	492	87,5%
<b>Nível 5 - (tem/ temos)</b>	132	27,6%	347	72,4%
<b>Nível 6 - (cantou/cantamos)</b>	227	34,3%	435	65,7%
<b>Nível 7 - (é/ somos)</b>	30	30,6%	68	69,4%
<b>(6) = 201.31, p &lt; 0,001</b>				

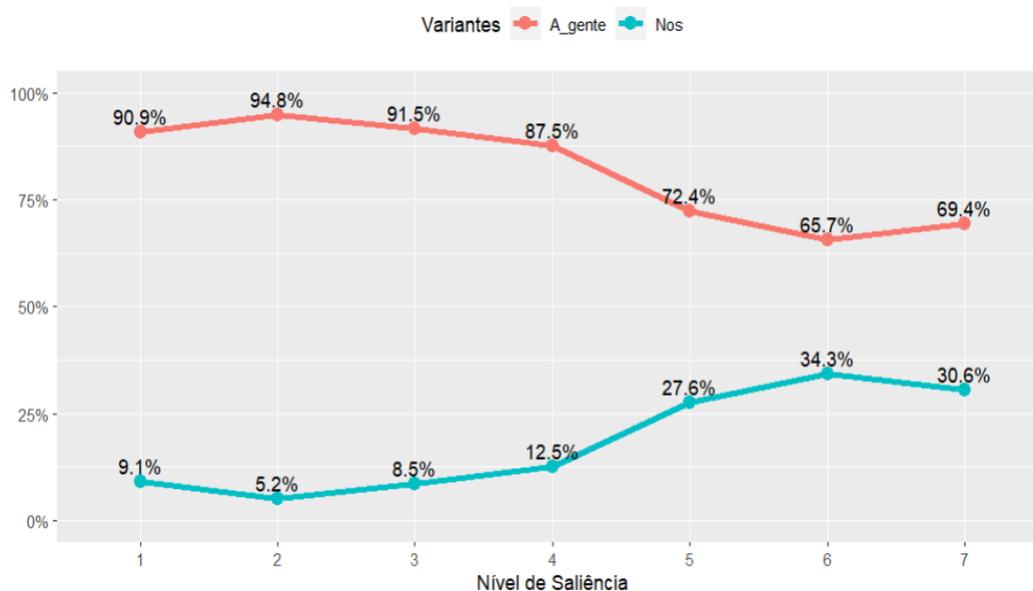
Fonte: elaboração própria

Em relação às proporções de uso das formas pronominais *nós* e *a gente* distribuídas em função dos sete níveis de *saliência fônica*, nossa hipótese é parcialmente confirmada. Há de se considerar que os dados de fala culta fortalezense representativos da década de 2020, de

maneira geral, apresentam maior inclinação à realização do pronome *a gente*, e essa tendência se reflete na distribuição das ocorrências em cada nível de saliência, tendo em vista que, em todos os fatores, há maior proporção de uso do pronome inovador, conforme demonstram os resultados da tabela 25.

Por outro lado, apesar dos termos percentuais privilegiarem o uso de *a gente*, esse privilégio se manifesta com intensidades diferentes em cada nível de saliência testado. Nesse sentido, o pronome inovador tem proporções de uso mais elevadas em verbos com menor diferenciação entre o material fônico, ou seja, os verbos de menor saliência, enquanto o pronome padrão, embora não se sobressaia ao pronome inovador em nenhum dos fatores, figura com proporção pouco mais elevada nos verbos de maior saliência. Sendo assim, à medida que a saliência do verbo aumenta, o uso do pronome *a gente* diminui e, por outro lado, o uso do pronome padrão *nós* aumenta, mas não ultrapassa o ponto médio (50%). No entanto, é necessário destacar que isso não ocorre de forma simétrica, uma vez que a curva estatística não é totalmente decrescente, mas sinuosa. Para uma representação visual desse comportamento, o gráfico abaixo ilustra a tendência no uso desses pronomes em relação ao nível de diferenciação fônica:

**Gráfico 17:** Proporção de *nós* e *a gente* em função do nível de *saliência fônica*



Fonte: Elaboração própria.

O nível de saliência 1 (que corresponde à mesma forma verbal para os pronomes *nós* e *a gente*) é contexto praticamente categórico para a realização do pronome inovador. Em nossa amostra do PORCUFORT Fase II, encontramos apenas 22 ocorrências dos pronomes de

primeira pessoa associados a verbos de saliência nível 1, sendo 20 realizações do pronome *a gente* (90,9%), e apenas 2 manifestações do pronome *nós* (9,1%). Vale ressaltar que todos os dados associados ao nível 1 de *saliência fônica* correspondem a verbos na forma nominal gerúndio, sendo este tipo de verbo um fator também controlado na variável *tempo verbal*. Essa redundância nas variáveis pode indicar presença de multicolinearidade (Oushiro, 2017). Sendo a ausência de multicolinearidade um dos princípios do modelo logístico, possivelmente será necessária a exclusão desses dados para que o modelo possa convergir adequadamente (Levshina, 2015).

Nos níveis de saliência 2 e 3, é possível identificar comportamentos de certo modo parecidos na alternância entre os pronomes *nós* e *a gente*, demonstrando certa estabilidade quanto às escolhas dos pronomes. Em relação aos verbos de saliência nível 2, no contexto fônico, são verbos que se apresentam na forma nominal infinitiva, tendo sua diferenciação entre singular e plural determinada pelo acréscimo da desinência de plural *-mos*. Esse contexto, assim como os outros, é o forte favorecedor do pronome *a gente*, representando 94,7% das ocorrências, em detrimento do pronome *nós*, que se realizou em apenas 5,3% dos dados neste nível. A partir do nível 3 de saliência fônica (conservação da sílaba tônica e acréscimo de *-mos*), o percentual de uso começa cair, mesmo ainda estando expressivamente associado ao pronome inovador (91,5%), enquanto o pronome padrão tem baixa proporção de uso (8,5%), conforme demonstram as informações do gráfico 17.

No nível 4 de saliência, quando ocorre o deslocamento da sílaba tônica com acréscimo da desinência *-mos* na distinção entre singular e plural, é notório a tendência na redução no percentual de uso da forma pronominal inovadora conforme aumenta o grau de diferenciação fônica. Mesmo com a redução no uso, o pronome *a gente* continua preferido entre os informantes com verbos neste nível, o que corresponde a 87,5% dos dados. Por outro lado, a medida em que aumenta o grau de saliência fônica, aumenta também a proporção de uso do pronome padrão *nós*, que, no caso dos verbos de nível 4, se manifesta em 12,5% dos casos, uma diferença de 4 pontos percentuais em relação ao nível anterior.

A partir do nível 5, a diferenciação fônica entre as formas verbais na relação singular e plural se torna mais evidente e, conseqüentemente, há um aumento pouco mais expressivo do pronome *nós*, ao passo que se observa redução no percentual de uso da forma *a gente*, mesmo que isso ocorra de maneira discreta. Conforme demonstra a tabela 25, monossílabos tônicos ou oxítonos que passam a paroxítonos quando plural, ainda tendem a privilegiar a forma pronominal inovadora (72,4%), enquanto o pronome padrão, apesar de ter o percentual de uso aumentado, continua a apresentar menor proporção (27,6%).

Por fim, apresentamos os resultados referentes à atuação dos níveis 6 e 7 da saliência fônica. Apesar da baixa proporção de uso do pronome *nós* em todos os graus de saliência, é nestes níveis que a variante padrão apresenta maior expressividade entre os contextos fônicos analisados, mesmo que o pronome inovador ainda seja preferido. Em relação aos verbos de nível 6, nos quais ocorre a redução do ditongo associado ao acréscimo da desinência *-mos*, observa-se maior proporção de uso do pronome *a gente* (65,7%), e uso mais restrito do pronome *nós* (34,3%). Ainda que haja pouca proporção de uso, há um aumento de 6,7 pontos percentuais do pronome padrão, dado importante, tendo em vista que esses tipos de verbo são altamente frequentes em nossa amostra (662 dados observados), o que representa 25,3% de todos os dados. Com isso, podemos concluir que os verbos de saliência nível 6 apresentam um contexto fértil à produção da variante padrão, sendo o ambiente com maior frequência de uso desta variante, com 227 dados de fala. No que se refere aos verbos no nível 7 (diferenças fônicas acentuadas entre as formas singular e plural), foram encontrados 98 dados de primeira pessoa, tendo maior proporção de uso do pronome *a gente* (69,4%) e menor uso do pronome *nós* (30,6%), conforme demonstra o gráfico 17.

O teste de qui-quadrado de Pearson revelou um valor de  $\chi^2 = 201.31$ , possuindo 6 graus de liberdade e valor de  $p < 0,001$ . Isso indica que esta variável previsora apresenta interação estatisticamente significativa com o uso dos pronomes. Neste ponto, reafirmamos a necessidade de se entender o qui-quadrado como um teste univariado, que avalia apenas o efeito individual de cada variável previsora sobre a variável resposta. Sendo assim, é crucial observar o comportamento de todas as variáveis em interação, através de um modelo de regressão. Dedicamos a próxima seção à descrição do modelo de regressão logística, desde sua construção até os resultados em termos de favorecimento.

### ***5.2.2 Análise multivariada da atuação das variáveis linguísticas e sociais sobre a variável resposta: o modelo de regressão logística - Amostra 2020***

Para a construção de um modelo logístico capaz de explicar o fenômeno da variação entre pronomes *nós* e *a gente*, primeiramente levamos em consideração os resultados dos testes estatísticos apresentados na subseção anterior (Oushiro, 2017). Esses resultados apontaram relevância estatística em quase todas as variáveis predictoras testadas, sendo a variável *sexo/gênero* a única não considerada estatisticamente relevante na atuação sobre a variável resposta, a alternância pronominal de primeira pessoa do plural. Isso é indicativo de que,

possivelmente, essa variável será excluída do modelo sugerido nos testes de *step forward*, *step backward*, e *step both*, que analisará o contexto multivariado e apresentará um modelo que melhor explique os dados de nossa amostra de fala culta dos anos 2020. Para confirmar a exclusão ou a manutenção da variável *sexo/gênero*, é necessário incluí-la no modelo inicial e observar se o teste de seleção automática recomendará sua permanência ou retirada.

Utilizando a linguagem de programação *R*, por meio da *interface R Studio*, elaboramos um modelo logístico para testar a relevância de cada variável em um contexto multivariado, com todos os fatores em interação. Através da função *glm* (Generalized Linear Model), elaboramos um modelo de regressão logística incluindo todas as variáveis predictoras testadas neste estudo, a saber, *sexo/gênero*, *faixa etária*, *tipo de inquerito*, *grau de referencialidade dos pronomes*, *paralelismo formal*, *tempo verbal*, e *saliência fônica*. O modelo elaborado apresenta a seguinte sintaxe:

```
Modelo_2020 <- glmer (VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO +
REFERENCIALIDADE + TIPO.VERBO + TEMPO.VERBAL + PARALELISMO +
SALIENCIA + (INFORMANTE), data = Dados_2020, family = binomial)
```

No entanto, após a execução no *software R Studio*, o modelo não convergiu. Quando um modelo de regressão logística não converge, significa que o algoritmo de otimização usado para estimar os parâmetros das variáveis não foi suficientemente significativo para atestar verossimilhança no modelo. Isso se deve a vários fatores, como problemas nos parâmetros, presença de multicolinearidade, dados inconsistentes, problemas relacionados à curva da função de verossimilhança, entre outros (Levshina, 2015). Após tentar convergir o modelo, o *software* tratou de apontar o motivo da não convergência:

**Figura 7:** Mensagem de erro ao convergir o modelo de regressão

Error in vif.default(Modelo\_2020):  
there are aliased coefficients in the model

Fonte: elaboração própria

A mensagem descrita no erro apresentado, em tradução livre, indica que ‘existem coeficientes aliados no modelo’ e significa que há multicolinearidade nos dados processados. Como vimos na subseção 5.1.1.1, a multicolinearidade ocorre quando os dados “não são

totalmente ortogonais” (Oushiro, 2017, p. 202), pois apresentam variáveis redundantes. Diante deste erro, decidimos criar um modelo logístico usando outra abordagem, a função *lrm*, que constitui o pacote *rms* (Regression Modeling Strategies). Esta função tem finalidade de ajustar os modelos de regressão logística binária, através do método da verossimilhança. Tentamos, então, convergir este modelo, que inclui todas as variáveis testadas. Novamente, o modelo não convergiu, apresentando o seguinte erro:

“singular information matrix in lrm.fit (rank= 28). Offending variable(s): SALIENCIA=n1”

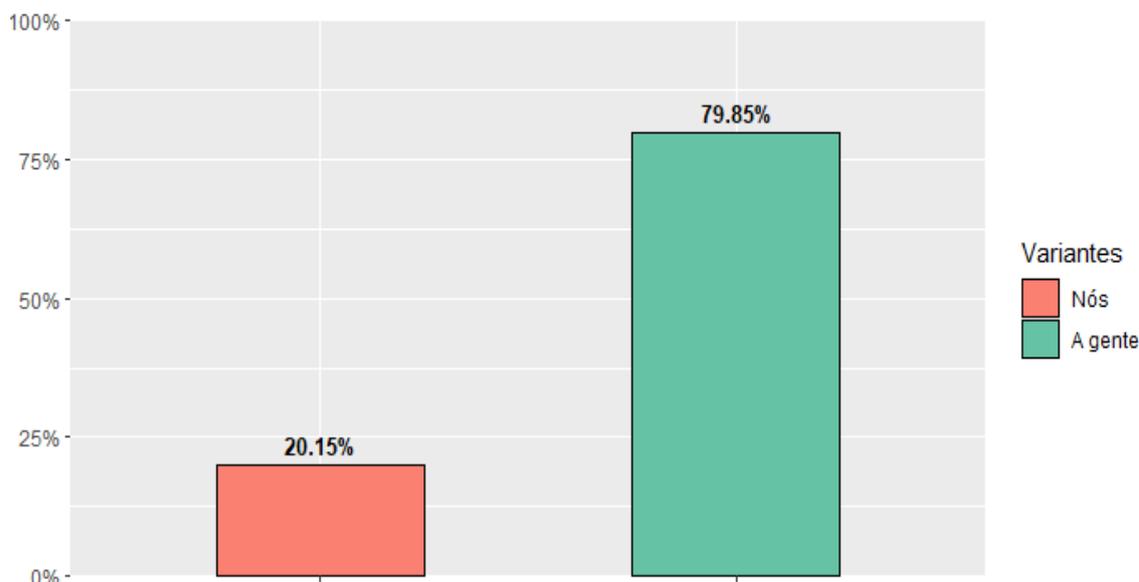
Como vimos anteriormente (ver subseção 5.1.1.1), esse erro indica presença de multicolinearidade, sendo dois ou mais fatores altamente relacionados. O erro aponta onde existe a redundância e, como prevíamos, os dados de *nível 1* da saliência fônica estão simetricamente relacionados com os dados de *gerúndio*, da variável *Tempo verbal e formas nominais do verbo*. Por esta razão, para que o modelo pudesse convergir, foi necessário proceder com a exclusão desses dados que, ao todo, representam 22 ocorrências, sendo duas realizações do pronome *nós* e vinte de *a gente*. As ocorrências (150) e (151) apresentam os dois dados do pronome padrão removidos da análise, e os excertos (152) e (153) demonstram exemplos das ocorrências do pronome inovador neste contexto, tendo em vista que são 20 ocorrências, e tornaria inviável apresentar todas.

- (150) Inf.: mas hoje nós estamos aqui contando a nossa história... conversando **nós botando**... a/ as PREcúndia de fora... e:... também com a felicidade de ter os nossos neto né? (Inq. 80 – F2 – D2, mulher, 61 anos)
- (151) Inf.: não tínhamos percebido que nós tínhamos parado na calçada em frente a faixa de pedestre... e **nós conversando** aqui quando nós olhamos... tinha uma fila de carro... (Inq. 03 – F2 – DID, homem, 26 anos)
- (152) Inf.: a minha irmã G. nasCEU né... eu lembro aTÉ assim:: **a gente brinCANdo** eu com a minha outra irmã (Inq. 28 – F2 – DID, mulher, 44 anos)
- (153) Inf.: e eu e lembro MUIto bem que ele chegou na academia e **a gente tendo** uma conVERsa faZENdo a anamnese dele foi onde ele se emocionou pediu pra eu ajudar ele (Inq. 16 – F2 – DID, homem, 38 anos)

Com a exclusão dos dados que impediam a convergência do modelo de regressão, ficaram 2596 dados de primeira pessoa do plural em nossa amostra representativa da década de

2020 (PORCUFORT Fase II), sendo 2073 observações da forma pronominal *a gente*, que corresponde a 79,85% do total de dados, ampla maioria, e outras 523 ocorrências do pronome inovador *nós* 523, representando 20,15% do total de ocorrência na amostra. O gráfico abaixo demonstra a proporção de uso de cada pronome após a retirada dos dados colineares.

**Gráfico 18:** Proporção de uso de *nós* e *a gente* após remoção de dados colineares (N=2596)



Fonte: elaboração própria

Conforme podemos observar nas informações do gráfico 18, a remoção dos dados não causou alterações significativas nas proporções de uso dos pronomes de primeira pessoa, tendo em vista que apenas 22 dados foram desconsiderados. Essa frequência de dados, por não ser significativa, não causará interferência ou implicações nos resultados do modelo logístico, uma vez que a diferença entre as proporções de uso de cada pronome antes e depois da retirada desses dados é de apenas 0,10 pontos percentuais.

Após sanar o problema da multicolinearidade, procedemos novamente a criação de um modelo de regressão envolvendo todas as variáveis. O modelo convergiu com sucesso, indicando que o problema, de fato, estava na questão da ortogonalidade dos dados. Submetemos, então, o modelo criado à função *vif* (Variance Inflation Factor - Fator de Inflação da Variância), do pacote *CAR* (Companion to Applied Regression), que mede a estimativa de correlação com as variáveis do modelo, podendo confirmar que a questão da multicolinearidade foi resolvida. Após aplicar a função, o *R Studio* forneceu o seguinte *output*:

**Tabela 26:** Resultado do teste de colinearidade das variáveis predictoras através da função *vif*

<b>Resultado do teste de inflação da variância</b>			
	<b>GVIF</b>	<b>Df</b>	<b>GVIF<sup>1/(2*Df)</sup></b>
<b>Sexo/gênero</b>	1.053656e+00	1	1.026477
<b>Faixa etária</b>	1.125058e+00	2	1.029897
<b>Tipo de inquérito</b>	1.640029e+00	2	1.131652
<b>Grau de referencialidade</b>	1.617001e+00	1	1.271614
<b>Tipo de verbo</b>	2.457155e+00	4	1.118933
<b>Tempo Verbal</b>	1.066998e+09	8	3.666572
<b>Paralelismo</b>	1.074373e+00	3	1.012028
<b>Saliência Fônica</b>	1.568809e+09	5	8.309158
<b>car::vif(modelo)</b>			

Fonte: Elaboração própria

Conforme destacamos, os menores valores de *Gvif* ajustado indicam menor presença de colinearidade entre as variáveis. Analisando os dados da tabela 26, observamos as variáveis *saliência fônica*, com alto valor *Gvif* (8.309158), e *tempo verbal*, também com valor expressivo (3.666572), sendo as variáveis que obtiveram maior índice de *Gvif* entre as testadas. Vale destacar que, embora ambos os valores estejam dentro do considerado aceitável, isso aponta para uma forte correlação entre essas duas variáveis.

Comprovada a ausência de correlação significativa entre as variáveis, procedemos a análise do modelo através da função *step*, por meio da interface *RStudio*. Essa função faz parte da instalação base do *R*, e é utilizada para selecionar de forma automática as variáveis que, em um ambiente multivariado, apresentam maior interação com a realização da variável resposta. Como parâmetro usado para adotar um modelo ideal para os nossos dados, adotaremos os menores valores descritos no Critério de Informação Akaike (AIC), já detalhado na subseção 5.1. Neste ponto, procedemos a análise a partir de três funções, a saber, *step forward*, *step backward* e *step both*:

**Tabela 27:** Resultado do teste de seleção automática das variáveis por meio da função *step*

<b>Resultado Step forward (AIC = 1907.8)</b>			
	<b>Df</b>	<b>Df Deviance</b>	<b>AIC</b>
<none>		1859.8	1907.8
+ Sexo/gênero	1	1859.3	1909.3
+Tempo verbal	8	1859.2	1911.2
<b>Modelo sugerido: VR ~ SALIENCIA + TEMPO.VERBAL + REFERENCIALIDADE + TIPO.INQUERITO + TIPO.VERBO + FAIXA.ETARIA</b>			

Fonte: Elaboração própria

Os resultados dos três testes de *step* apontaram o mesmo resultado, ou seja, a exclusão da variável social *sexo/gênero*, e da variável linguística *tempo verbal*, designando o AIC = 1907.8 como valor limite para determinar se uma variável contribui significativamente para o modelo. Conforme observamos na Tabela 27, até mesmo um modelo vazio, chamado de *<none>*, tem maior relevância explicativa que um modelo que incluísse essas duas variáveis. Vale destacar que a exclusão da variável *tempo verbal*, importante preditor observado em estudos sobre a alternância pronominal, pode ter sido motivada pela sobreposição de outra variável significativa para a explicação do fenômeno, a saber, a *saliência fônica*. Naro e Scherre (2010), explicam que essas duas variáveis estão correlacionadas e que “saliência e tempo verbal são variáveis fortemente significativas em análises distintas, mas saliência sempre supera tempo/modo quando são colocadas em uma mesma análise como variáveis separadas” (Naro e Scherre, 2010, p. 74). Nesse sentido, a atuação da *saliência fônica* podem estar se sobressaindo aos do *tempo verbal*, motivo que pode ter levado essa variável a ser excluída no teste de seleção automatizada.

Além disso, os resultados dos três testes indicaram que um modelo logístico ideal para explicar os nossos dados deve incluir as variáveis *saliência fônica*, *grau de referencialidade do pronome*, *tipo de inquerito*, e *faixa etária*. Considerando os resultados dos testes, excluímos as variáveis *sexo/gênero* e *tempo verbal* do nosso modelo final.

Tendo formado o modelo ideal para tratamento dos dados em termos de estatística referencial, procedemos com a análise através da função *lrm*, que fornece uma série de medidas estatísticas importantes para entender a eficiência do modelo criado. Entre essas medidas, está o índice de discriminação C, que adotamos como parâmetro para classificar o poder explicativo que o conjunto de variáveis selecionadas possuem para explicar os dados da variação. O ambiente *R Studio*, após submeter o modelo logístico a função *lrm*, gerou o seguinte *output*:

**Tabela 28:** Resultado do teste multivariado de *nós* e *a gente* através da função *lrm*

	Model Likelihood Ratio Test			Discrimination Indexes		Rank Discrim. Indexes	
<b>Obs</b>	2596	<b>LR chi2</b>	736.75	<b>R2</b>	0.390	<b>C</b>	0.853
<b>A gente</b>	523	<b>d.f.</b>	17	<b>R2(17,2596)</b>	0.242	<b>Dxy</b>	0.706
<b>Nós</b>	2073	<b>Pr(&gt;chi2)</b>	<0.0001	<b>R2(17,1252.9)</b>	0.437	<b>Gamma</b>	0.708
<b>Max reriv</b>	2e-13			<b>Brier</b>	0.114	<b>Tau-a</b>	0.227

Fonte: Elaboração própria.

A tabela 28 apresenta dados importantes sobre o modelo que ajudam a determinar sua eficiência no tratamento da variação. O primeiro valor que vamos considerar é o

determinado pelo Teste de Razão de Verossimilhança (LR chi2), que compara o modelo atual (com as variáveis predictoras selecionadas) com um modelo nulo (sem variáveis predictoras), para determinar a eficácia na explicação dos dados. O resultado do teste apontou LR chi2 igual a 736.75, valor relativamente alto se comparado ao ponto médio (500), o que indica que nosso modelo atual possui capacidade explicativa melhor que a de um modelo nulo, razão pela qual devemos insistir no modelo. Além disso, o valor do Índice de Concordância C foi de 0.853, sendo classificado como um modelo que apresenta excelente poder discriminatório dos dados da amostra, conforme explica Levshina (2015).

Em posse de todas essas informações sobre o modelo, iniciamos a análise multivariada da atuação das variáveis predictoras, por meio da estatística inferencial. Desta forma, através da análise do programa computacional *R Studio*, submetemos este modelo logístico a função *glm*, a fim de gerar um resultado com as estimativas da atuação de cada variável elencada sobre a realização dos pronomes. Neste ponto, é necessário destacar que o modelo de regressão logística gerou um *intercept*, valor de referência usado para comparar as estimativas geradas. Em nosso modelo de regressão logística, o valor de *intercept* corresponde a estimativa de uso da variante considerada inovadora, *a gente*, nos seguintes contextos: (1) 22 a 35 anos, (2) inquérito do tipo DID, (3) pronome com referência genérica, (4) verbos de ação, (5) segunda ou terceira ocorrência da série com quebra de paralelismo, e (6) saliência fônica nível 7. Os resultados com as estimativas de uso de cada variável estão dispostos na tabela 29:

**Tabela 29:** Estimativas de uso de *a gente* no modelo de regressão logística: Amostra 2020 (N = 2596)

<b>Coefficients</b>	<b>Estimate</b>	<b>Std. Error</b>	<b>z value</b>	<b>Pr(&gt; z )</b>	
<b>(Intercept)</b>	1.1090	0.2795	3.968	7.26e-05	***
<b>Faixa etária</b>					
36 a 50 anos	-0.5514	0.1797	-3.069	0.00215	**
A partir de 50 anos	-1.5087	0.1600	-9.430	< 2e-16	**
<b>Tipo de Inquérito</b>					
D2	1.1513	0.1654	6.962	3.36e-12	***
EF	-1.2678	0.1487	-8.524	< 2e-16	***
<b>Referencialidade</b>					
Específica	-1.1747	0.1432	-8.205	2.31e-16	***
<b>Tipo de Verbo</b>					
Estado	-1.0600	0.2068	-5.125	2.97e-07	***
Ter	-1.8270	0.1912	-9.555	< 2e-16	***
<i>Dicendi</i>	0.1850	0.2958	0.625	0.53180	
Epistêmico	0.2558	0.2558	0.3474	0.736	
<b>Paralelismo Formal</b>					
Ocorrência isolada	1.2186	0.2285	5.334	9.61e-08	***
Primeira ocorrência da série	1.6105	0.2387	6.747	1.50e-11	***
Manutenção da forma anterior	2.0390	0.2291	8.899	< 2e-16	***
<b>Saliência Fônica</b>					
Nível 2 (cantar/cantamos)	2.2570	0.4577	4.931	8.18e-07	***
Nível 3 (falava/falávamos)	1.6258	0.1922	8.459	< 2e-16	***
Nível 4 (fala/falamos)	1.0584	0.1823	5.806	6.42e-09	***
Nível 5 (cantou/cantamos)	0.9466	0.2032	4.658	3.20e-06	***
Nível 6 (cantou/cantamos)	0.7572	0.2929	2.585	0.00973	**
<b>Modelo: glm(VR ~ FAIXA.ETARIA + TIPO.INQUERITO + REFERENCIALIDADE + TIPO.VERBO + TEMPO.VERBAL + SALIENCIA)</b>					
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					

Fonte: elaboração própria.

A análise do modelo logístico apontou interação estatisticamente significativa em todas as variáveis previsoras testadas nos dados representativos da década de 2020 (PORCUFORT Fase II), tendo apenas dois fatores estatisticamente irrelevantes no modelo, a saber, os tipos de verbo *dicendi* e epistêmico, que constituem a variável *tipo de verbo*. De forma geral, os resultados do modelo logístico confirmam uma tendência já atestada nas análises da atuação individual de cada variável (ver subseção 5.2.1). Ao analisarmos os percentuais de uso de cada pronome distribuídos em função das variáveis previsoras, percebemos maior emprego

do pronome *a gente*, o que é confirmado no resultado deste modelo multivariado, conforme demonstra a tabela acima, a variante inovadora *a gente* é favorecida em praticamente todos os contextos linguísticos e sociais.

Para analisar a atual situação das variáveis em um contexto multivariado, é essencial destacar que o modelo de regressão logística gera coeficientes que são apresentados em *logodds* (*log-odds-ratio*). No entanto, os valores das estimativas de cada fator apresentado no modelo não correspondem ao valor final da atuação desse fator sobre a realização da variável resposta, uma vez que é necessário considerar o valor de *intercept* para determinar o real grau de atuação da variável. Assim, para se chegar ao valor da estimativa de determinado nível de uma variável previsora, é preciso que o valor atribuído a este nível seja somado ao valor de referência do modelo, ou seja, ao valor do *intercept*. O resultado dessa soma, expresso em *logodds*, determina o valor real de estimativa de uso de determinada variável, sendo interpretado da seguinte forma: valores positivos, de modo geral, indicam favorecimento, e valores negativos apontam desfavorecimento em relação ao nível de referência (Oushiro, 2017). Os valores descritos a seguir retratam o valor final da atuação das variáveis, resultado da soma com o *intercept*.

A seguir, descreveremos os resultados de cada uma das variáveis testadas em um contexto multivariado, obtidos através de um modelo de regressão logística. A primeira variável a ser descrita é a variável *faixa etária*.

#### a) Atuação da faixa etária sobre alternância pronominal de primeira pessoa do plural

A tabela 29 apresenta o resultado da atuação multivariada dos fatores sobre a realização dos pronomes de primeira pessoa. A primeira variável apresentada no modelo é a *faixa etária*. Os resultados da tabela 29 apontam o valor de *intercept*, que se refere a atuação dos mais jovens (22 a 35 anos), igual a 1.1090 e valor  $p > 0,001$ . Esses resultados demonstram que, no contexto que compõe o *intercept*, os informantes mais jovens favorecem significativamente o uso do pronome *a gente*, contexto social que, de acordo com nosso dispositivo estatístico, inibe a atuação do pronome padrão. A observação da faixa etária é essencial para o estudo de variação, tendo em vista que, em tempo aparente, essa variável aponta indícios de mudança em progresso, conforme aponta Monteiro (2000):

O tempo aparente refere-se pois ao padrão de distribuição do comportamento linguístico através de vários grupos etários num determinado momento do tempo, ou

seja, se o uso de uma variante inovadora for mais frequente entre os jovens decrescendo em relação à idade dos grupos mais idosos tudo indica que se trata de uma situação de mudança em progresso (Monteiro, 2000, p.132).

Considerando esse pressuposto, a análise das proporções de uso dos pronomes de primeira pessoa em relação à faixa etária apresenta possíveis indícios de uma mudança em progresso avançada do pronome *nós* pelo pronome *a gente*, uma vez que, na década de 2020, o pronome *a gente* é preferido em todas as faixas etárias, em maior proporção entre os mais jovens e em menor proporção pelos mais velhos. Em termos de favorecimento, os dados da tabela 29 demonstram que essa tendência também é confirmada em nosso modelo de regressão, uma vez que os informantes de 36 a 50 anos, em relação aos mais jovens, tendem a desfavorecer o uso de *a gente*, com estimativa de uso de 0,5503, estando abaixo do *intercept*, porém ainda com valor positivo. Além do mais, atuação dessa faixa etária apresenta interação significativamente relevante, com *p-value* = 0,002, estando dentro do intervalo de confiança. Por outro lado, Araújo (2007), acredita que os pronomes *nós* e *a gente* configuram uma variação estável, pois, com o passar do tempo, informantes tendem a adquirir a forma padrão:

Um padrão encontrado confirma a hipótese de variação estável: a variável *a gente* seria mais frequente na fala dos jovens e que ao alcançar a idade adulta esses falantes entrariam em contato com a forma *nós* (já conhecida, mas pouco percebida) sendo submetidos a forças mais conservadoras e conseqüentemente aumentariam a freqüência da variante de prestígio. (Araújo, 2007, p. 393).

Esse padrão é confirmado em nosso modelo logístico, pois os falantes mais jovens tendem a favorecer o uso do pronome inovador, e, à medida que a faixa etária aumenta, o nível de favorecimento da forma pronominal *a gente* vai diminuindo. Isso pode ser comprovado ao analisarmos as estimativas de uso referentes aos informantes mais velhos da amostra que, em relação aos mais jovens, desfavorecem expressivamente o uso da variante inovadora, passando a privilegiar o uso do pronome padrão *nós* (-0,3997), sendo a única faixa etária a apresentar coeficiente negativo (o que indica favorecimento expresso da forma padrão *nós*). Além do mais, o valor de  $p < 0,001$  comprova que o resultado do teste é estatisticamente significativo, o que pode apontar para uma mudança em tempo aparente (Coelho et al, 2012)

Silva (2022), ao investigar a alternância pronominal entre *nós* e *a gente* na comunidade de Pariconha, em Alagoas, chegou a resultados sobre a atuação da variável faixa etária que se aproxima dos nossos. Utilizando a ferramenta *R Studio* para o tratamento dos dados, o autor identificou favorecimento do pronome inovador *a gente* nas três faixas etárias pesquisadas, sendo mais intensa entre os mais jovens (82%) e menos intensa entre os mais velhos (63%), apontando relevância estatisticamente significativa para a atuação da variável

faixa etária ( $\chi^2 = 21.14$ , e valor de  $p = 2.56e-05$ ). Essa mesma tendência é observada em outras pesquisas que investigam a fala representativa de décadas mais atuais, como a década de 2010 e de 2020. Nesse sentido, Vitória (2016), ao investigar falantes de nível superior de escolaridade na cidade de Maceió, e Souza (2020), que investigou a fala de informantes quilombolas não escolarizados, constataram maior favorecimento da variante inovadora em todas as faixas etárias, sendo os mais jovens os que privilegiam mais intensamente essa forma pronominal.

De acordo com Labov (1972), para que seja possível identificar indícios de uma mudança em curso, é necessário correlacionar a distribuição das variantes em função das faixas etárias investigadas, para observar seus níveis de incidência em cada intervalo etário. Dessa forma, se a variante inovadora demonstra forte favorecimento entre os mais jovens e esse favorecimento diminui conforme a idade dos grupos mais velhos, isso pode sugerir uma mudança em progresso. No caso específico da fala culta de Fortaleza-CE, os resultados apontam fortes indícios de que a referência a primeira pessoa do discurso está passando por uma mudança em curso no sentido de *a gente* estar gradativamente substituindo o *nós*, tendo em vista que, tanto os termos percentuais quanto as estimativas apresentadas no modelo, apontam maior favorecimento da forma linguística inovadora entre os mais jovens, o que pode sugerir uma transformação linguística na comunidade. No entanto, é importante ressaltar que não há mudança completada, uma vez que os pronomes *nós* e *a gente* ainda disputam a preferência entre os informantes de todas as faixas etárias, sendo então necessário levar em consideração outros fatores de natureza social e linguística para entender a dinâmica dessa variação.

b. *Atuação do tipo de inquérito sobre alternância pronominal de primeira pessoa do plural*

A variável previsor *tipo de inquérito* reflete o contexto de interação entre os informantes nas entrevistas gravadas, sendo esta uma estratificação metodológica prescrita pela própria base de dados adotada nesta pesquisa, o projeto PORCUFORT. Como tratado na metodologia (ver seção 4.4), essa variável corresponde a situações comunicativas com diferentes estilos, sendo os inquéritos do tipo D2 os que apresentam estilo menos formal, os DID possuem nível intermediário de formalidade e os inquéritos EF são aqueles que assumem um estilo mais formal, tendo em vista que estes são provenientes de gravações de palestras, aulas, cerimônias religiosas e outros tipos de registro que exigem maior monitoramento da fala (Araujo, 2011).

Nossa hipótese inicial sobre a atuação dessa variável prescreve maior favorecimento do pronome inovador em inquéritos D2, por ser menos formal e propenso a uso

de formas não padrão. Para Azevedo (2012), por exemplo, a forma pronominal *a gente* está associada ao uso da fala em ambientes menos formais, o que também pode significar intimidade entre as pessoas do discurso. Por outro lado, esperávamos que os inquéritos do tipo EF favoreceriam o uso do pronome padrão, já que possuem alto nível de monitoramento da fala, o que torna o ambiente pouco produtivo para formas inovadoras. Em relação a alternância entre os pronomes de primeira pessoa do plural, Cezário e Votre (2008) asseveram que, apesar de a forma *a gente* não apresentar marcas de estigma social, o pronome padrão *nós* é usado em situações de fala mais formais, enquanto a forma *a gente* tem preferência em contextos mais coloquiais.

Os resultados apresentados na tabela 29 confirmam nossa hipótese inicial sobre a atuação dessa variável. Destacamos que o nível de referência dessa variável corresponde a atuação de inquéritos do tipo DID (diálogo entre informante e documentador), que, no contexto de realização do *intercept*, tendem a favorecer o uso da variante inovadora *a gente* (1.1090), confirmando o que já foi atestado na análise em termos de proporção de uso, que apontou maior proporção de uso para o pronome inovador (79,1%). Considerando que este tipo de inquérito possui grau intermediário de monitoramento da fala, tê-lo como nível de referência desta variável nos permite contrastar com clareza os contextos menos formais (D2) e mais formais (EF) da amostra.

Inquéritos do tipo D2 (diálogo entre dois informantes), conforme hipotetizamos, tendem a favorecer expressamente o uso da forma pronominal *a gente* que, em relação ao nível de referência, apresenta expressiva estimativa de uso, com valor igual a 2,2603, o dobro do que representa o *intercept*, e  $p\text{-value} < 0,001$ , o que indica relevância estatística. Uma das possíveis explicações para este resultado é que a forma *a gente* é usada em situações de comunicação menos monitorada, em que o informante deixa de se preocupar com a própria fala, tornando-a mais espontânea e natural, o que a faz se aproximar de seu vernáculo. Essa explicação é sustentada por Labov (1972), que analisou situações comunicativas na coleta de dados, que vão desde entrevistas sociolinguísticas até leitura de textos, e observou que nos estilos menos formais há maior recorrência de variantes inovadoras, enquanto as situações mais formais estão mais propensas a variantes mais antigas na língua.

Por outro lado, em termos de favorecimento, os inquéritos do tipo EF, demonstram maior tendência ao uso da variante conservadora *nós*, conforme demonstram os dados presentes na tabela 29. Em relação aos inquéritos do tipo DID (*intercept*), os inquéritos que representam as elocuições formais (EF) tendem a desfavorecer intensamente o pronome inovador, com estimativa de uso de -0,3997, valor significativamente menor que o nível de referência (1,1090),

além de ser negativo, o que significa que, nos inquéritos do tipo EF, o pronome padrão *nós* é privilegiado, conforme deduziu nossa hipótese. Neste ponto, destacamos que, embora a amostra demonstrou maior proporção de uso da forma inovadora (ver gráfico 20), em termos de favorecimento ou desfavorecimento, há maior propensão ao uso da forma padrão *nós* em ambientes mais formais, como é o caso das elocuções formais (EF).

A influência do grau de formalidade sobre o uso de formas variantes foi observada no estudo de Eckert (2008) em pesquisa sobre o significado social da variação, com base em um estudo da prática estilística. Para a autora as variáveis não possuem significados estáticos, mas sim significados gerais que se tornam mais específicos no contexto dos estilos. Ao analisar situações de fala em diferentes níveis de formalidade, como conversa entre amigos e apresentação oral em escolas, a linguista verificou maior ocorrência de variantes conservadoras em estilos mais formais, ao passo que formas inovadoras estão associadas a ambientes de menor monitoramento da fala, explicando que o processo variável é uma constelação de significados ideologicamente relacionados, qualquer um dos quais pode ser ativado na situação de uso da variável. Essa associação entre o uso das formas variantes e o estilo também é tratada por Oliveira (2022), que explica:

Pensar a língua de um ponto de vista social por meio de estudos sociolinguísticos variacionistas implica necessariamente considerar que a língua só existe inserida em seus contextos de uso – ela não é usada de forma idêntica por pessoas diferentes, em momentos e lugares diferentes. Ela está sujeita a variações e mudanças que, aos poucos, modificam a estrutura original. (Oliveira, 2022, p. 12).

Para a autora, as motivações das escolhas entre os pronomes *nós* e *a gente*, que apresentam o mesmo valor referencial, não depende apenas de fatores de natureza social e linguística, pois parte do processo que envolve a alternância entre os pronomes é determinado por situações comunicativas, tipos de registros e o estilo empregado no discurso. Ou seja, o grau de atenção dado a fala, no momento da enunciação, é franco motivador para a seleção das formas em variação. Dessa maneira, segundo assevera a autora, “quanto mais atento ao modo como fala, menos traços vernaculares o indivíduo emprega, isto é, menos marcas de fala casual e despreocupada emergem em sua fala” (Oliveira, 2022, p. 13). Nesse sentido, o estilo empregado em cada tipo de inquérito explicaria os resultados do nosso modelo de regressão logística, uma vez que a forma mais antiga na língua é geralmente mais valorizada, fato que torna o registro mais formal mais inclinado ao seu uso, enquanto a forma mais inovadora, como é o caso do pronome *a gente* tende a ser mais expressivamente vinculado a situações comunicativas menos formais.

c. *Atuação do grau de referencialidade sobre alternância pronominal de primeira pessoa do plural*

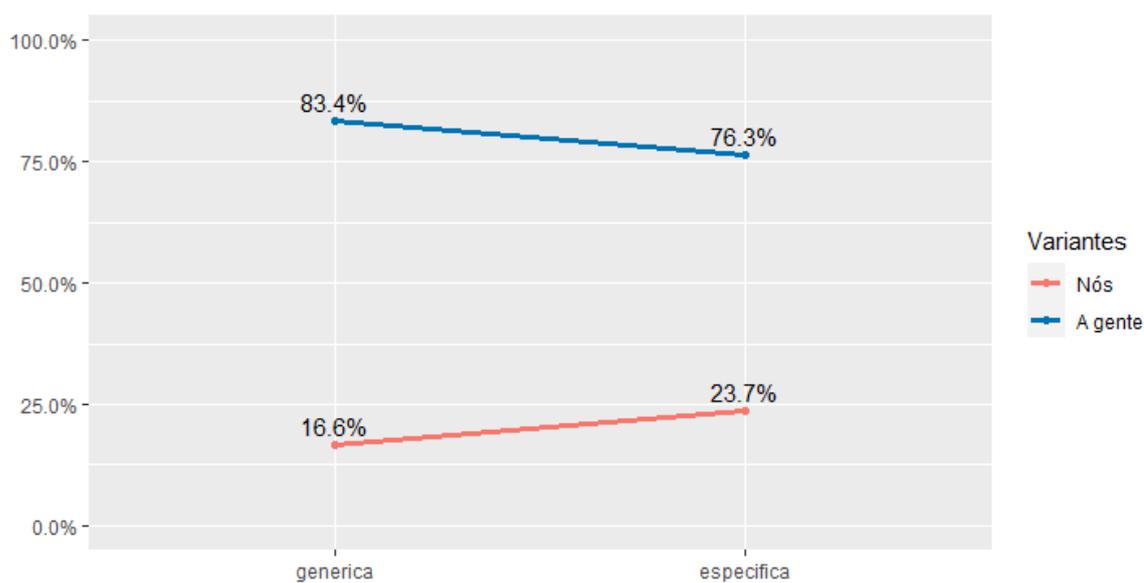
A primeira variável previsora de natureza linguística foi o grau de referencialidade do pronome, variável muito importante para explicar a variação entre os pronomes de primeira pessoa do plural. Como explicado anteriormente (ver seção 3.1), o pronome *a gente* tem origem no substantivo *gente*, usado para se referir a um grupo de pessoas “de forma coletiva, indeterminadora, mais ou menos geral” (Omena, 2003, p. 64). Por esse motivo, estudos sociolinguísticos testaram exaustivamente esta variável e, de modo geral, observaram a forma pronominal *a gente* associada à referentes genéricos - fato que atribuíram à sua origem -, e pronome *nós* correlacionado à referência específica (Machado, 1995; Lopes, 1998; Omena, 2003; Borges, 2004; Tamanine, 2010; Mattos, 2013; Araujo, 2016; Souza, 2020; Silva, 2022). Com base na literatura, nossa hipótese previa que o pronome inovador seria favorecido em contextos de referência genérica, enquanto o pronome *nós* seria privilegiado em contextos de referência específica.

Em relação a atuação dessa variável no modelo de regressão, o valor de referência corresponde ao emprego do pronome *a gente* associado a referencialidade genérica, sendo assim, no contexto social e linguístico que constituem o *intercept*, o pronome *a gente* é favorecido quando a referência é genérica, tendo estimativa de uso de 1,1090 (intercepto) e valor de  $p > 0,001$ . Esse resultado aponta para forte associação desta variável com a realização dos pronomes de primeira pessoa em um contexto multivariado, em interação com as outras variáveis. Sendo assim, em nossa amostra de fala culta representativa da década de 2020, ocorrências como as encontradas em (153) e (154) são ambientes favorecedores do uso do pronome inovador *a gente*:

- (153) Inf.: igreja é é de Deus a única na terra se dividisse né essa es/ era o maior temor dele quando ela se universalizasse né... bom mais é isso ***a gente*** vive ainda hoje principalmente no Brasil né conflitos de/ desse tipo (Inq. 20 – F2 – EF, homem, 28 anos)
- (154) Inf.: não tem UM psicólogo que não faça terapia... quem bem souBESse todas as pessoas faziam terapia que faz BEM... ***a gente*** Vive numa sociedade... que nós temos milHÕES de seguidores... (Inq. 78 – F2 – EF, mulher, 43 anos)

Por outro lado, o uso dos pronomes de primeira pessoa associados a referências específicas apresenta forte tendência a desfavorecer a forma pronominal inovadora, dispondo de uma estimativa de uso que, em relação ao emprego dos pronomes com referência genérica (*intercept*), é igual a  $-0,0657$ . Considerando o número negativo, é possível interpretar que a referencialidade específica tende a favorecer o uso do pronome padrão *nós*, mantendo sua função prototípica. É necessário considerar, entretanto, que os valores em *logodds* representam uma escala que vai de menos infinito até mais infinito, tendo o zero como ponto médio (Oushiro, 2015). Nesse sentido, estando as estimativas referentes ao uso dos pronomes com referencialidade específica próximas do ponto neutro ( $-0,0657$ ), averiguamos que o favorecimento do pronome canônico em contextos de referência específica ocorre de forma bastante discreta na amostra representativa dos anos 2020, tendo em vista que sua estimativa é próxima de zero. Isso pode ser explicado pela alta proporção do uso do pronome *a gente* na amostra em ambas as situações de referencialidades, como demonstra o gráfico 19:

**Gráfico 19:** Proporção de uso de *nós* e *a gente* em função do grau de referencialidade



Fonte: elaboração própria.

As informações apresentadas no gráfico 19 apontam alta proporção de uso do pronome inovador tanto em contextos de referencialidade genérica (83,4%), quanto específica (76,3%), atestando que, em termos de proporção de uso, considerando a atuação individual desse fator na escolha entre os pronomes, não há diferenciação expressiva entre as formas de acordo com a referência semântica. No entanto, ao considerar um modelo multivariado, em que

se observa a atuação em conjunto das variáveis através de um modelo de regressão logística, percebe-se diferenças quanto ao favorecimento das formas pronominais a depender do grau de referencialidade que o pronome expressa, sendo a forma *a gente* favorecida em referências genéricas (1.1090), e discretamente desfavorecida em referências específicas (-1.1747). Essa característica também foi observada em Lopes (1993).

O falante utiliza preferencialmente o pronome *nós* para se referir a ele mesmo e mais o interlocutor (não-eu), ou a não-pessoa: referente [+ perceptível] e [+ determinado]. No momento em que o falante amplia a referência, indeterminando-a, há maior favorecimento para a forma *a gente* (Lopes, 1993. p. 130).

Assumindo essa tendência em nossos dados, o valor da estimativa de uso dos pronomes em referência específica indica um ambiente que privilegia a forma conservadora, no entanto o faz de forma pouca intensa. Essa baixa intensidade no favorecimento de *nós* pode apontar para uma mudança no traço intrínseco de referencialidade que está associado ao pronome *a gente*. As informações do gráfico 19 apontam alto percentual de uso do pronome *a gente* também em contextos de referência específica, o que demonstra que, na fala culta fortalezense da década de 2020, o pronome *a gente* adentrou nos mais diversos ambientes linguísticos, inclusive se expandindo em contextos de referencialidade específica. Esses resultados apresentam indícios de que o pronome inovador *a gente*, na amostra pesquisada, está começando a assumir outros traços, além da função prototípica de pronome indeterminador, como vem sendo atestado em diversos estudos sobre a alternância pronominal (Omena, 2003; Tamanine, 2010; Mendonça e Nascimento, 2015; Araujo, 2016, entre outros).

A próxima variável preditora de ordem linguística testada no modelo de regressão logística trata-se do *tipo de verbo*. A seguir, trataremos de apresentar os resultados da análise multivariada que observa a atuação desta variável, em interação com os demais preditores, sobre a realização dos pronomes de primeira pessoa do plural.

#### d. *Atuação do tipo de verbo sobre alternância pronominal de primeira pessoa do plural*

A análise através do modelo de regressão logística apontou ação estatisticamente significativa da variável *tipo de verbo* sobre a realização dos pronomes. Está variável, que considera a classificação do verbo quanto ao sentido empregado, busca analisar quais tipos de verbos são formal e semanticamente associados às formas pronominais. Para isso, analisamos os verbos de ação, de estado, verbo *ter*, verbos *dicendi* e verbos epistêmicos, a fim de buscar

indícios de favorecimento da forma inovadora em função dos tipos semânticos de verbos testados.

Inicialmente, há de se considerar que os verbos de ação constituem o nível de referência dessa variável, através do qual os outros tipos de verbo serão comparados quanto a sua estimativa de uso. O valor de estimativa atribuído ao intercepto, do qual engloba os verbos de ação, é igual a 1,1090. Isso significa que, em realizações com o mesmo contexto social e linguístico que se realiza o *intercept*, os verbos de ação têm tendência a favorecer o uso do pronome inovador. Essa tendência é atestada pelo valor de  $p < 0,001$ , o que demonstra interação estatisticamente relevante. Os verbos de ação, vale destacar, são os verbos mais abundantes de nossa amostra, sendo observadas 1465 realizações destes verbos associados ao uso dos pronomes de primeira pessoa, o que corresponde a uma frequência de uso de 83,8% (ver tabela 16). Sendo assim, podemos dizer que, tanto em termos de proporção de uso quanto em favorecimento, há maior disposição para o uso do pronome inovador *a gente* junto a verbos de ação.

Em relação aos verbos que expressam estado, o resultado da análise através do modelo de regressão logística indica que, em relação ao *intercept* (verbos que expressam ação) os verbos de estado tendem a desfavorecer o uso do pronome inovador, tendo estimativa de uso de 0,049 *logodds*. Esse resultado, apesar de apresentar valor de estimativa expressivamente menor que a do intercepto, está próximo do ponto de neutralidade, indicando pouco favorecimento dos pronomes em relação ao verbo de estado. Estudos sociolinguísticos, no entanto, têm demonstrado forte favorecimento do pronome padrão quando associados a verbos de estado (Tamanine, 2002; Francischini, 2011; Araujo, 2016). Esses trabalhos têm apontado que “os verbos de estado podem inibir o uso da forma *a gente* por estarem associados à variante conservadora e não à forma inovadora, contribuindo assim para esse favorecimento pelo uso de *nós*” (Araujo, 2016, p. 85). Essa tendência, em relação ao *intercept*, se confirma em nossos dados, tendo em vista que, em relação aos verbos de ação, os de estado tendem a desfavorecer o pronome *a gente*. Vale ressaltar, ainda, que o resultado da atuação desse tipo de verbo gerou um valor de  $p < 0,001$ , o que aponta forte interação entre esses verbos e o uso dos pronomes de primeira pessoa do plural. Os exemplos abaixo, ilustram ocorrências dos pronomes de primeira pessoa junto a verbos que expressam estado em nossa amostra representativa da década de 2020:

- (155) Inf.: na verdade meus pais viraram amigos... mais do que pais assim eles viraram amigos minha mãe também ***nós somos*** muito pró::ximas nos vemos todo di::a... sempre que posso passo bastante tempo lá (Inq. 07 – F2 – DID, mulher, 24 anos)

- (156) Inf.: eu vou falar de exemplos um pouco mais palpáveis né tenho certeza que todos nós aqui né que todos **nós estamos** hoje aqui porque nós temos uma certa rotina (Inq. 21 – F2 – EF, homem, 39 anos)

Seguindo a análise do modelo de regressão apresentado na tabela 29, verbos do tipo *ter*, em um contexto multivariado, apresentam forte tendência a desfavorecer a variante inovadora *a gente*. Os resultados do modelo logístico apontam, em relação aos verbos de ação (valor de *intercept*), estimativa de uso igual a -0,718, indicando que, quando associado a verbos do tipo *ter*, há maior favorecimento da variante padrão *nós*. Esses resultados também foram observados em análise da fala popular de Fortaleza da década de 2000, que apontou os verbos do tipo *ter* como inibidores da forma pronominal *a gente* (Araujo, 2016). Dessa forma, contextos de emprego dos pronomes pessoais de primeira pessoa do plural acompanhados de verbos do tipo *ter*, como os exemplificados nos excertos (157) e (158), são ambientes que parecem favorecer o uso da variante mais antiga na língua, o pronome *nós*.

- (157) Inf.: os cantores... os cantores surgiam através das RÁdio né? aqui por eXEMplo na Ceará Rádio CLUbe tinha:: tinha:: **nós tínhamos** canTOres naquela Época como Gilberto MilFON Mário ALves:: (Inq. 22 – F2 – DID, homem, 67 anos)
- (158) Inf.: antes quando eu era solteira meu deus saco ir pro cinema ficar lá sentada... eu tinha essa mentalidade mas aí quando eu fui com ele **nós temos** cinco anos de relacionamento ele foi/ eu fui aprendendo a gostar... (Inq. 07 – F2 – DID, mulher, 31 anos)

Ademais, é importante ressaltar que a influência dos verbos do tipo *ter*, em nosso modelo de regressão logística, resultou em um valor de  $p < 0,001$ , demonstrando que esses verbos se revelam estatisticamente significativos, tendo em vista que o resultado o *p-value* está dentro do valor de significância adotado nesta pesquisa (valor alfa). Essa constatação indica correlação relevante entre o uso do verbo *ter* e a seleção entre os pronomes sujeitos a *nós* e *a gente*. Por outro lado, os outros tipos de verbos testado no modelo logístico, os verbos do tipo *dicendi* e epistêmico, não apresentaram efeito estatisticamente significativo sobre a realização da variável resposta, visto que os valores de *p* atribuídos a cada um desses verbos estão fora do limite do valor alfa adotado nesta pesquisa (5%).

Dito isto, tratamos de descrever a próxima variável analisada no modelo de regressão logística. A seguir, apresentaremos os resultados, em termos de favorecimento, da atuação da variável *paralelismo formal* sobre a realização dos pronomes de primeira pessoa do

plural.

e. *Atuação do paralelismo formal sobre alternância pronominal de primeira pessoa do plural*

A variável *paralelismo formal* é compreendida como a tendência que o falante possui de repetir a mesma forma empregada dentro de uma sequência discursiva. Essa variável tem se mostrado relevante em diversas pesquisas sobre o uso dos pronomes de primeira pessoa do plural que, assim como esta pesquisa, investigam amostras representativas da década de 2010/2020, como pode ser observado em Vitorio (2017) Souza (2020), Silva (2022), entre outros.

Para Schere e Naro (1993), o princípio do paralelismo estabelece que os falantes tendem a repetir suas escolhas linguísticas em uma sequência de fala, no sentido de que “marcas levam a marcas, e zeros levam a zeros” (1993, p. 03). Ou seja, nessa perspectiva, em uma sequência discursiva, a presença do pronome *a gente* favorece o emprego da mesma forma pronominal posteriormente dentro da mesma sequência, pois “marcas explícitas e de mesma natureza favorecem as mesmas marcas no sentido de ocorrerem paralelamente” (Pacheco, 2014, p. 207)

Em nosso modelo de regressão logística, que descreve a atuação de cada variável em um contexto multivariado, a variável *paralelismo formal* demonstrou ser um preditor significativo para a realização do pronome inovador *a gente* (nível de referência), com diferentes níveis atuando sobre a variável resposta. A significância estatística, apontada pelo *p-value* e pelos asteriscos (\*) sugere que todos os fatores desta variável são relevantes para predizer a variável dependente. Desta forma, o modelo, de maneira geral, parece se ajustar bem aos dados.

O valor de *intercept* neste contexto se refere ao emprego do pronome *a gente* antecedido pelo pronome diferente, em que há quebra de paralelismo formal. Assim, considerando todos os fatores que compõem o intercepto, quando todas as variáveis têm valor zero, há maiores chances de realização do evento observado (uso do pronome inovador *a gente*). Optamos por inserir esse contexto no valor de referência para comparar, em termos de favorecimento, com outras três categorias: ocorrências isoladas, primeira realização da série e segunda ocorrência com emprego do mesmo pronome imediatamente anterior (manutenção do paralelismo). Com isso, poderemos observar o efeito médio de cada nível dessa variável e sua

importância na modelagem dos dados através do modelo logísticos apresentado na tabela 29.

Nos dados da Amostra 2020, todos os ambientes analisados demonstram favorecimento da forma inovadora, em maior ou menor grau. Quando o emprego de *a gente* é o primeiro da série, em relação ao *intercept*, há favorecimento da variante inovadora. Nesse caso, o valor positivo da estimativa desta categoria, expresso em *logodds*, sugere que a probabilidade de ocorrência do nível de referência da variável resposta (*a gente*), se comparado com o valor de intercepto do modelo, aumenta consideravelmente (2.3276). Esse resultado imprime o comportamento geral da amostra analisada, uma vez que a variante inovadora tem sido empregada em maior proporção em praticamente todos os contextos analisados. Resultados parecidos foram encontrados nas pesquisas de Vitória (2016), Souza (2020) e Alves e Souza (2020), quando identificaram tendência ao favorecimento do pronome *a gente* em contextos de realização isolada.

Em relação ao próximo nível da variável *paralelismo formal*, quando se trata do uso do pronome *a gente* como primeira ocorrência em uma série de realizações, esse contexto favorece intensamente o pronome inovador, apresentando estimativa de uso de 2,7195, valor bem acima do estimado para o *intercept* (1,1090). Em outras palavras, quando comparado com contextos em que o pronome aparece isoladamente, empregá-lo como a primeira ocorrência em uma série de realizações é um fator que favorece o uso da forma pronominal *a gente*. Esse resultado corrobora com os achados de Souza (2020), que pesquisou os pronomes *nós* e *a gente* em amostra composta por entrevistas gravadas em 2016 (década de 2010). Os resultados encontrados pelo autor apontaram, em relação a variável *paralelismo formal*, que “o fator primeiro da série também se destaca como relevante, uma vez que apresenta percentual de 60% de aplicação e peso relativo .54” (Souza, 2020. p 75). Além disso, o valor de *p* atribuído a este fator indica associação estatisticamente significativa com a realização do pronome *a gente* ( $p > 0,001$ ), fato que reafirma o favorecimento da forma inovadora nesses contextos.

Quanto ao uso do pronome com manutenção da forma imediatamente anterior, o modelo de regressão logística apontou forte favorecimento para o pronome *a gente* quando antecedido pelo mesmo pronome, conforme apresentam os resultados da tabela 29. A estimativa de uso para esse nível, em relação ao *intercept*, indica forte vantagem do pronome *a gente*, com valor que chega a 3,148, número expressivamente maior que o valor de referência dessa variável. Resultados próximos ao nosso foram encontrados por Omena (2003, p; 72), ao pesquisar a fala carioca, e identificar que “a primeira referência é influenciada pela forma predominantemente utilizada”.

Esse resultado também foi observado no estudo conduzido por Mendonça (2010),

com dados da fala popular de Vitória - ES, a partir de amostras extraídas do Projeto Português Falado na Cidade de Vitória – PortVix. Após levantamento dos dados e análise da variável, o autor percebeu forte influência da variável *paralelismo formal* na fala de Vitória, sendo a variante inovadora fortemente privilegiada quando antecedida pelo *a gente* explícito (0.71), e ainda mais favorecido quando a forma anterior é o pronome *a gente* implícito (0.96). De maneira parecida, o pronome padrão *nós* é favorecido quando empregado como segundo ou terceiro da série, mantendo a mesma forma na série de repetições.

Na mesma direção, Lima (2020), quando investigou a alternância entre os pronomes *nós* e *a gente* em amostra de fala coletada na cidade de São Luís - MA, colhidos entre os anos de 2011 e 2015 (década de 2010), chegou a resultados semelhantes ao encontrados nesta pesquisa. Segundo o autor, há maior favorecimento do pronome *a gente* quando antecedido pela mesma forma pronominal, com frequência de uso de 91,1% das ocorrências e peso relativo igual a 0.80, demonstrando forte tendência ao uso do pronome inovador nesses contextos. Nesse sentido, para o linguista, “a realização de uma forma pronominal, implícita ou explícita, seja de *nós* ou *a gente*, desencadeará sucessivas repetições do mesmo pronome na sequência discursiva, constituindo, portanto, alto índice de aplicação do paralelismo linguístico” (Lima, 2020, p. 90).

O valor de *p*, que representa a significância da atuação deste fator sobre a realização do pronome *a gente*, apontou  $p > 0,001$ , o que indica que o evento observado é fortemente motivado por este fator. Além do mais, cabe destacar que, dentre as categorias observadas nesta variável, a manutenção do paralelismo formal é o fator que mais favorece o emprego da variante inovadora *a gente*.

#### *f. Atuação da saliência fônica sobre alternância pronominal de primeira pessoa do plural*

O princípio da saliência fônica se refere à diferenciação do material fônico entre as formas de terceira pessoa do singular e de primeira pessoa do plural. Esse preditor tem se mostrado significativo na explicação da variação entre os pronomes *nós* e *a gente*, tendo sido exaustivamente testado em diversas variedades de língua falada, como demonstram estudos sociolinguísticos (Scherre, 1988; Lopes, 1993; Borges, 2004; Brustolin, 2009; Lima, 2020; Silva, 2022). Estas pesquisas apontam para maior favorecimento do pronome inovador nos níveis mais baixos de saliência fônica e, ao contrário, maior propensão ao uso de *nós* nos níveis mais elevados de saliência. Santana (2014), ao analisar a comunidade de fala de Salvador/BA,

em amostras extraídas do Programa de Estudos sobre o Português Popular Falado de Salvador (PEPP), identificou a mesma tendência. Segundo prescreve o autor.

A saliência fônica entre as formas verbais de 3ª pessoa do singular e 1ª do plural também influem no aparecimento de *a gente*, e a menor diferença entre as formas verbais pode favorecer a ocorrência de *a gente*, ao passo que a maior diferença entre elas pode favorecer a forma *nós* (Santana, 2014, p. 38).

Essa tendência, em partes, é atestada pelo resultado do nosso modelo de regressão (ver tabela 29). No entanto, vale ressaltar que, diferentemente do atestado nestas pesquisas, não há favorecimento do pronome padrão *nós* em nossos dados, de acordo com o resultado do modelo. Em nossa amostra de fala culta da década de 2020, o pronome *a gente* é beneficiado em todos os níveis que compõem a variável saliência fônica, em uma escala de intensidade que vai do nível mais baixo ao mais alto, seguindo o seguinte esquema: Nível 2 (cantar/cantarmos) > Nível 3 (falava/falávamos) > Nível 4 (fala/falamos) > Nível 5 (tem/temos) > Nível 6 (cantou/cantamos) > Nível 7 (é/somos). Neste ponto, é importante lembrar que, para que o modelo logístico pudesse convergir, foi necessário a retirada dos dados referentes ao *nível 1* de saliência fônica. Ademais, o *nível 7* (grau máximo de saliência), compõe o contexto de realização do *intercept*, razão pelo qual não aparece explicitamente em nosso modelo.

Dito isto, em relação ao Nível 2 (cantar/cantarmos), verbos no infinitivo com acréscimo da desinência *-mos*, os resultados da tabela 27 discriminam alta associação com a forma pronominal *a gente* que, em relação ao Nível 7 (é/somos), demonstram estimativa de 3,366, representando o contexto de diferenciação fônica mais fortemente inclinado ao uso da forma inovadora entre todos os níveis. Silva (2010), em pesquisa que toma como base de dados amostras de fala culta carioca, explica que a impossibilidade de o pronome *a gente* ser acompanhando por determinantes pode contribuir para esse resultado, tendo em vista:

o fato de [a gente] ser uma forma mais neutra do que *nós*, posto que mantém vínculo com o núcleo do SN coletivo, do qual se originou. Isso geraria a maior ocorrência de *a gente* em ambientes menos marcados, seja quanto ao tempo do verbo, seja quanto à saliência fônica. (Silva, 2010, p.67).

O resultado do modelo logístico, ainda, gerou um valor de  $p > 0,001$ . Esse número sugere interação estatisticamente significativa deste fator com a realização da variável resposta, o que significa que o uso de verbos com Nível 2 (cantar/cantarmos) pode condicionar o uso do pronome inovador.

Alguns verbos em nossa amostra, apresentam diferenciação fônica entre as formas singular e plural marcada pela conservação da sílaba tônica e acréscimo da desinência *-mos* (falava/falávamos). Esse tipo de verbo (Nível 3), quanto a diferenciação do material fônico,

representa o segundo ambiente, em uma hierarquia do maior para o menor, mais favorecedor do pronome *inovador a gente*. Em relação aos verbos com Nível 7 (*é/somos*), os de Nível 3 (*falava/falávamos*) estão mais propensos ao uso da variante inovadora *a gente*, com valor de estimativa de 2,7348. Essa categoria se mostrou estatisticamente relevante para o uso do pronome, com  $p\text{-value} > 0,001$ .

Já os verbos com saliência fônica Nível 4 (*fala/falamos*) são aqueles que, na diferenciação entre as formas singular e plural, há o deslocamento do acento tônico e acréscimo da desinência *-mos*. Verbos com essas características apresentam comportamento parecido com os verbos do nível anterior. Ou seja, mesmo que o valor de estimativa diminua conforme o grau de diferenciação fônica baixe, o Nível 4 é um fator que, em relação ao *intercept* (1,1090), beneficia fortemente o uso da forma pronominal *a gente*, tendo proporção igual a 87,5% e estimativa de uso de 2,1674, valor maior que o atribuído ao nível de referência desta variável (Nível 7). Ainda, o valor de  $p > 0,001$  indica forte associação deste fator com a variante inovadora.

Nos verbos de Nível 5 (*tem/temos*), monossílabos tônicos ou oxítonos que passam a ser paroxítonos na diferenciação entre singular e plural, há uma queda na proporção de uso de *a gente* (72,4%). Diferente dos resultados encontrados em alguns estudos de referência (Lopes, 2003; Franceschini, 2011; Scherre et al, 2018), em nossa amostra, verbos com essas características tendem a favorecer o uso do pronome *a gente*, mesmo que em menor intensidade se comparado aos verbos de saliência nível 4 (*fala/falamos*). Esse resultado pode ser explicado pelo comportamento geral da amostra da década de 2020, que demonstrou maior propensão ao uso da variante inovadora em praticamente todos os contextos de fala pesquisados, apresentando estimativa de uso igual a 2,0556, além de  $p\text{-value} > 0,001$ , atestando interação significativa deste verbo sobre a realização da variante de referência.

Por fim, o Nível 6 (*cantou/cantamos*) foi o que apresentou menor favorecimento ao uso da forma *a gente* em nosso modelo de regressão. Santana (2014), ao observar a atuação a atuação da saliência fônica sobre o uso dos pronomes de primeira pessoa do plural constatou que “O fator redução dos ditongos com acréscimo de desinência *-mos* (*cantou/cantamos*) foi o que teve maior probabilidade de favorecimento de *nós* (.78)” (Santana, 2014, p.71) De forma parecida, Lopes (1993), em estudo sobre a alternância pronominal na fala culta do Rio de Janeiro, identificou que verbos com este grau de diferenciação fônica são mais inclinados ao uso do pronome inovador *nós* (.77). Em nossos dados, esse pronome apresenta menor grau de favorecimento, embora ainda esteja fortemente associado ao uso do pronome *a gente*, o que demonstra que, na fala culta de Fortaleza da década de 2020, o pronome *a gente* é privilegiado

em todos os contextos fônicos, como demonstra a tabela 29.

Em termos de significância, os verbos que, na diferenciação entre as formas singular e plural, acontece a redução dos ditongos finais em vogais com o acréscimo da desinência *-mos* (Nível 6), esses se mostraram significativamente relevantes para a utilização da forma pronominal inovadora *a gente*, conforme indica o valor de  $p < 0,001$ . Com isso, as proporções de uso e as estimativas se confirmam, uma vez que, tanto o percentual de uso quando o resultado da atuação de cada nível de saliência em nosso modelo de regressão logística, apontaram para forte favorecimento do pronome *a gente* em dados de fala culta de Fortaleza da década de 2020.

### 5.2.3 Síntese dos resultados – A alternância pronominal na Amostra 2020

Na década de 2020, a análise estatística realizada por meio do *R Studio* apontou expressiva proporção de uso do pronome *a gente* (79,95%) e uso menos recorrente do pronome padrão *nós* (20,05%). Dentre os preditores que se mostraram relevantes para a análise da Amostra 2020, se destacam como estatisticamente significativos para a análise a *faixa etária*, *o tipo de inquérito*, *o grau de referencialidade do pronome*, *tipo de verbo*, *paralelismo formal* e *saliência fônica*.

Em termos gerais, os dados da Amostra 2020 apontam forte inclinação dos mais jovens pelo uso da variante inovadora *a gente*, tendo obtido 88,5% do total de ocorrências, sendo o pronome *nós* pouco utilizado entre esses informantes, com apenas 11,5% das observações. De igual maneira, em termos de proporção de uso, o comportamento do fenômeno entre os informantes com mais de 50 anos de idade indica favorecimento do pronome *a gente*, sendo utilizado em 69,7% das observações, ao passo que o pronome *nós* é utilizado em 30,3% dos casos. Um ponto importante de destacar é que, entre os mais velhos, a proporção de uso do pronome canônico *nós* é quase o triplo da proporção apresentada pelos mais jovens, isso sugere forte interação entre a *faixa etária* e a realização dos pronomes.

Em relação ao *tipo de inquérito*, a análise indicou proporção de uso do pronome *a gente* próximo da categórica em inquéritos do tipo D2, tendo 91,4% do total de ocorrências, enquanto o pronome *nós* apresenta proporção de uso de apenas 8,6%. Inquéritos do tipo DID, na mesma direção, parecem ser francos aliados da forma pronominal *a gente*, tendo ocorrido em 79,2% do total de realizações, enquanto o pronome padrão *nós* apresenta percentual de uso de 20,8% das observações em nossos dados. De forma menos acentuada, os dados de Elocução Formal (EF) indicam maior proporção de uso do pronome inovador, com percentual de uso de

66,5% do total de observações. Por outro lado, o pronome padrão *nós* ocorre em menor proporção em elocuições formais, sendo responsável por 33,5% dos dados, fato que contraria nossa hipótese sobre a atuação dessa variável, tendo em vista que esperávamos que o pronome padrão fosse mais recorrente em elocuições formais devido à natureza mais formal desse tipo de inquérito.

Na Amostra 2020, quando se trata do *grau de referencialidade dos pronomes*, houve alta proporção de uso do pronome inovador *a gente*, empregado com sentido genérico (83,4%), diferente do pronome *nós*, que foi pouco recorrente neste contexto (16,6%). Por outro lado, os dados revelaram baixa proporção de uso do pronome padrão *nós* com referente específico (23,7%) e alta proporção de *a gente*, mesmo com referente específico (76,3%). Esse resultado sugere uma ampla expansão do pronome inovador na década de 2020 em contextos de referencialidade específica, que antes figurava como um ambiente linguístico de resistência à entrada do pronome inovador, mas que agora é preferido entre os falantes cultos de fortaleza, independente do grau de referência.

Os resultados referentes ao preditor *tipo de verbo* destacam que o pronome inovador é mais produtivo que o pronome padrão em todos os tipos de verbo, porém somente os verbos de estado e os verbos do tipo *ter* apresentam relevância estatística ( $p < 0,05$ ). Em relação aos verbos que expressam estado, os resultados apontam menor proporção de uso para o pronome padrão *nós* (26,5%) nos dados de fala culta da década de 2020 e elevada proporção de uso do pronome inovador *a gente* (73,5%). Esse resultado é explicado pelo comportamento geral da amostra analisada, tendo em vista que os dados gerais apontam para uso preferencial do pronome inovador.

Outra variável preditora que se mostrou significativa na análise da Amostra 2020 trata-se do *paralelismo formal*. Em relação *a primeira ocorrência da série*, esse fator está associado ao uso do pronome inovador *a gente* (80,9%), percentual significativamente maior que o uso do pronome padrão, que detém 19,1% do total de realizações. O emprego dos pronomes como *segunda ou terceira realização de uma série com a manutenção da mesma forma* usada imediatamente anterior também age de forma a beneficiar o pronome inovador *a gente* (86,6%), e desfavorecer o uso de *nós* (13,4%). Por fim, em relação *a forma pronominal diferente em relação à anterior*, ou seja, quando há quebra do paralelismo, este é o único contexto a não privilegiar a forma pronominal *a gente* (48,2%); destacando o pronome *nós* (51,8%), apontando para uma quebra na tendência no que concerne ao uso dos pronomes pessoais de primeira pessoa, sendo o pronome padrão privilegiado nesse ambiente.

Sobre a variável *saliência fônica*, apesar da baixa proporção de uso do pronome *nós* em todos os graus de saliência, é nos níveis mais altos que a variante padrão apresenta maior expressividade entre os contextos fônicos analisados, mesmo que o pronome inovador ainda seja preferido. Já nos níveis mais baixos de saliência fônica, confirmando os estudos prévios, há maior favorecimento do pronome inovador *a gente*, conforme comprova nossa análise estatística.

### **5.3 Análise comparativa das décadas de 1990 e 2020: um estudo em tempo real de curta duração de tendência**

A natureza da atividade humana da linguagem é caracterizada por dois pontos que, embora pareçam divergentes, estão completamente correlacionados, a saber, a estabilidade e a constante mudança linguística, natural das línguas, sendo processos contínuos que ocorrem tanto no indivíduo quanto na comunidade de fala (Paiva e Duarte, 2003). A interação entre esses dois processos tem sido questão central nos trabalhos desenvolvidos à luz da teoria da variação e mudança linguística, que considera a língua uma entidade em constante mudança. Mas especificamente, trabalhos que trazem análises em tempo real de fenômenos variáveis demonstram que “um estado de língua é a face sincrônica da mudança linguística, ou seja, é caracterizado pela coexistência de formas representativas de diferentes estágios do sistema” (Paiva e Duarte, 2003, p. 13).

Nesse sentido, é importante considerar que os padrões linguísticos atuais podem prever a mudança futura, assim como o comportamento linguístico do passado anunciava o estágio atual da língua. Ao analisar um fenômeno variável considerando duas sincronias distintas, em um intervalo de mais de vinte anos, podemos observar quatro possíveis padrões distintos no comportamento linguístico de uma comunidade de fala: I) as proporções de uso das variantes podem se manter constantes (variação estável); II) a variável inovadora pode se expandir e aumentar sua proporção de uso, ao passo que o uso da variante mais antiga na língua diminui (mudança em progresso); III) a mudança linguística pode se retrair, no sentido de a variante padrão ter seu uso alavancado, à medida que as ocorrências da variante inovadora diminuem; e IV) a mudança linguística pode se completar, ou seja, a variante inovadora assume o lugar que antes era ocupado pela variante padrão, tendo esta se tornado obsoleta (mudança completa).

Labov (1994) afirma que os processos linguísticos ocorridos no passado são os mesmos que ocorrem no estágio atual da língua e que continuarão a ocorrer nos estágios futuros. Nesse sentido, as forças estruturais que motivaram a mudança linguística no passado são as mesmas que motivam a mudança no presente, sendo estas consistentes ao longo do tempo. De acordo com o Princípio do Uniformitarismo, prescrito por Labov (1994), as mudanças observadas hoje são reflexos dos processos que operaram sobre a língua no passado e as mudanças já concluídas na língua podem ser entendidas ao observar a dinâmica dos processos linguísticos ocorridos em sincronias passadas, o que torna essencial a análise de fenômenos variáveis em tempo real.

Neste estudo, ao confrontar duas sincronias da mesma comunidade de fala, buscamos identificar os processos estruturais e sociais responsáveis por impulsionar ou restringir a variação e mudança linguística, considerando que as forças que agem sobre o comportamento linguístico da comunidade de fala na sincronia mais atual eram as mesmas que atuavam na sincronia passada (Labov, 1994). Com isso, podemos observar quais os contextos linguísticos e sociais em que houve expansão ou retração no uso do pronome inovador *a gente* entre os falantes cultos de Fortaleza nos anos 1990 e 2020.

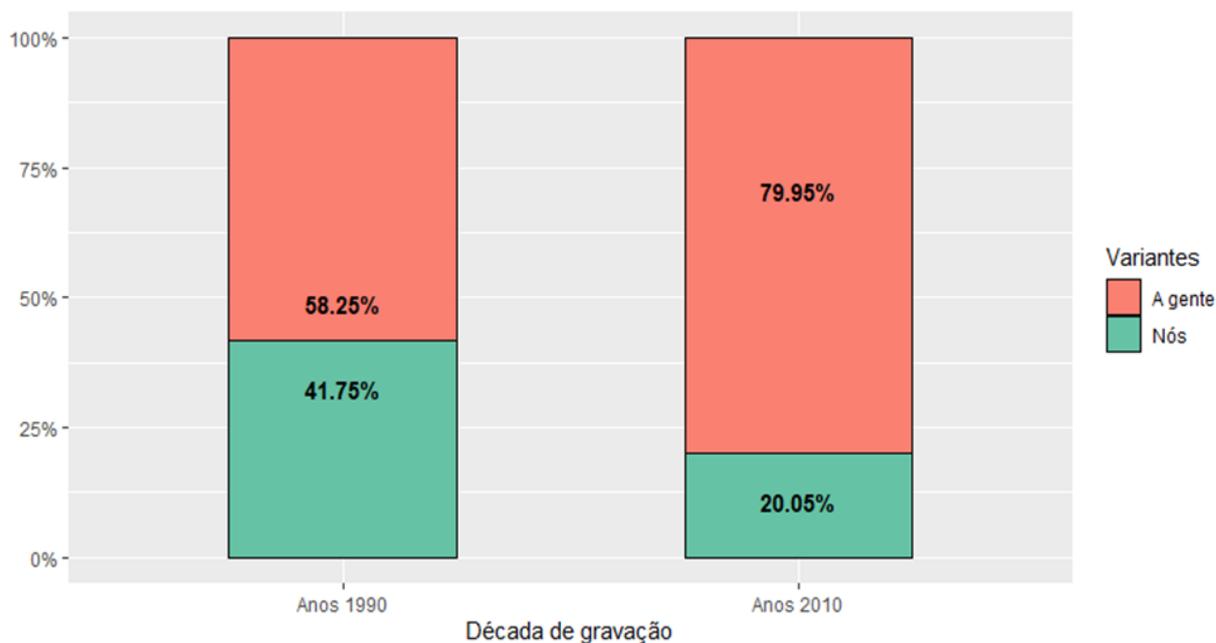
Ao considerar o intervalo de mais de duas décadas, assumimos que a análise, em termos de aumento ou retração na proporção de uso das variantes, segue uma tendência linear. Ou seja, por não acompanhar o comportamento da comunidade a cada ano neste intervalo, enfocaremos nossa análise comparativa nos contextos em que houve aumento ou diminuição, em termos de proporção e de favorecimento, no uso das formas *nós* e *a gente*. Para o processamento dos dados linguísticos, foi adotada a linguagem de programação *R* (R Core Team, 2020), por meio do *software R Studio*, para quantificar e testar o efeito das variáveis predictoras sobre a variável resposta (*a gente* e *nós*), plotar gráficos e realizar testes estatísticos e de regressão logística.

O estudo em questão compara as estatísticas referentes à comunidade através de duas amostras, Amostra 1990 (PORCUFORT – Fase I) e Amostra 2020 (PORCUFORT – Fase II). As análises estatísticas dos pronomes sujeitos *a gente* vs. *nós* em função das variáveis predictoras foram realizadas separadamente para cada uma das décadas. Neste ponto, buscamos comparar as tendências relativas ao uso das variantes, com base nos resultados obtidos nas análises de cada uma das décadas (ver subseções 5.1 e 5.2)

Em termos totais, foram contabilizadas 4575 observações de *a gente* e *nós* na base de dados pesquisada, sendo 1957 ocorrências na amostra de 1990, e 2618 ocorrências na de

2020. O gráfico abaixo indica as proporções de uso de *a gente* e *nós* em cada uma das décadas pesquisadas:

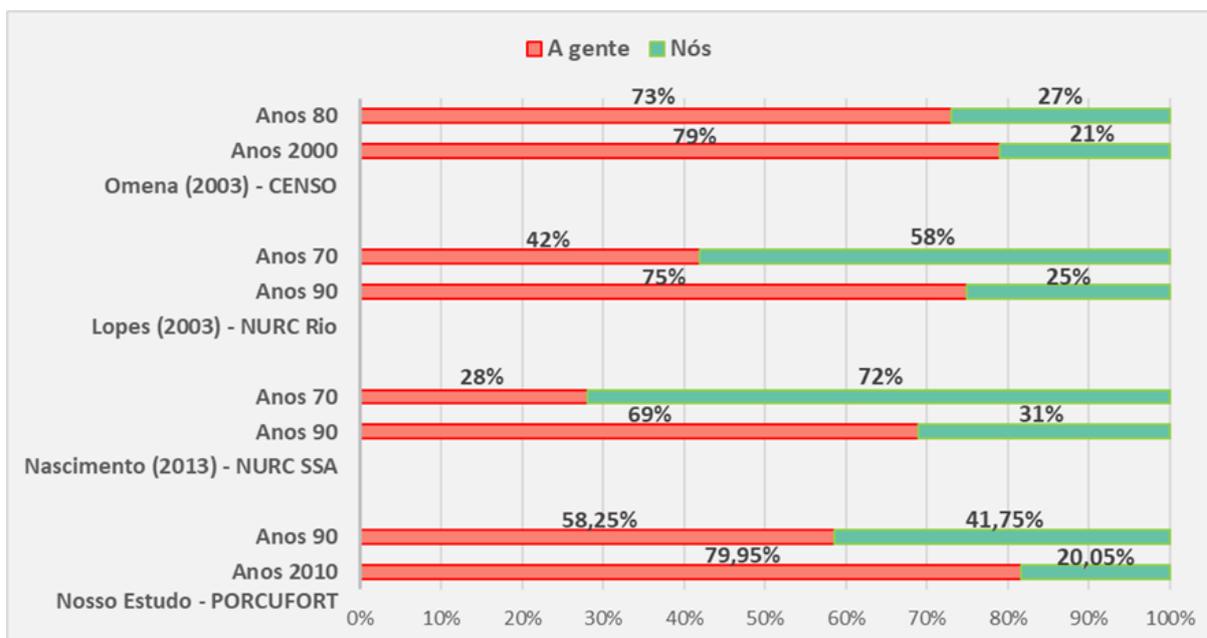
**Gráfico 20:** Proporção de uso de *a gente* e *nós* em tempo real: décadas de 1990 e 2020



Fonte: Elaboração própria

Conforme exposto no Gráfico 20, o pronome *a gente* é mais produtivo que *nós* em ambas as décadas. Na amostra de 1990, foram identificados 58,25% de *a gente* (1140 obs.) e 41,75% do pronome *nós* (817 obs.). No que se refere à amostra de 2020, foram observadas 79,95% das ocorrências (2093 obs.) de *a gente*, um aumento significativo na proporção de uso se comparada à década anterior, e apenas 20,05% de *nós* (525 obs.), demonstrando que o pronome *a gente* se destaca entre os informantes cultos como referência à primeira pessoa do plural.

A maior proporção para a forma inovadora *a gente* em detrimento do pronome *nós* confirma uma tendência de uso já apontada por outros estudos com dados em tempo real de curta duração, como os de Omena (2003), Lopes (2003) e Nascimento (2013). A esse respeito, vejamos o Gráfico 21:

**Gráfico 21:** Comparação do resultado de nosso estudo com outras pesquisas em tempo real

Fonte: Elaboração própria

O Gráfico 21 aponta para uma tendência em tempo real sobre o uso do pronome *a gente* em diversas comunidades de fala no Brasil: o uso do pronome inovador se torna mais expressivo a cada década. Omena (2003) investigou a variação no uso dos pronomes *nós* e *a gente* em tempo real com dados de fala popular carioca a partir de amostras do projeto Censo Linguístico. A primeira sincronia analisada (anos 1980), apontou uso expressivo do pronome *a gente* (73%) e uso pouco recorrente do pronome *nós* (27%). Em relação a análise da segunda sincronia (década de 2000), a autora identificou comportamento linguístico parecido ao da década anterior, com aumento no uso do pronome inovador *a gente*, sendo este pronome ainda mais utilizado (79%) em comparação com o pronome padrão, que ocorreu em menor proporção na sincronia mais recente analisada pela autora (21%).

Essa tendência no que concerne ao uso do pronome inovador também foi observada na fala culta do Rio de Janeiro, em pesquisa conduzida por Lopes (2003), com dados da fala culta carioca provenientes do projeto NURC de duas décadas. A análise da primeira década observada, anos 1970, apontou para o maior uso do pronome padrão *nós* (58%) em detrimento do pronome inovador, que foi menos recorrente na amostra (42%). Já na segunda sincronia analisada, anos 1990, o comportamento linguístico dos informantes mudou, tendo estes demonstrado preferência pelo pronome inovador na mesma comunidade de fala após vinte anos, com a expressiva proporção de 75% de uso do pronome *a gente*, enquanto o pronome padrão

*nós* passou a ser menos utilizado, tendo se realizado em apenas 25% dos dados, uma mudança significativa no emprego dos pronomes de primeira pessoa do plural na comunidade.

De forma parecida, os resultados encontrados por Nascimento (2013) demonstram mudança linguística expressiva no uso dos pronomes pessoais de primeira pessoa do plural entre os falantes cultos de Salvador em vinte anos. A análise da amostra referente à década de 1970 indicou uso significativo do pronome *nós*, sendo responsável por 72% das realizações de primeira pessoa do plural, com o pronome *a gente* figurando em apenas 28% dos casos. Vinte anos depois, na década de 1990, o comportamento da mesma comunidade de fala se inverte, tendo os falantes cultos se inclinado ao uso em maior proporção do pronome inovador *a gente* (69%), enquanto o pronome *nós* sofreu uma queda considerável em sua proporção de uso (31%), apontando para uma substituição no uso das formas pronominais na cidade de Salvador.

Ao compararmos os resultados da nossa pesquisa com os resultados dos estudos sobre a fala do Rio de Janeiro (Omena, 2003; Lopes, 2003) e de Salvador (Nascimento, 2013), observamos que a flutuação entre *nós* e *a gente* permanece nessas amostras, compostas por informantes graduados e do Ensino Médio (Amostra Censo). Esta tendência demonstra a implementação e a franca expansão da variante *a gente* no português brasileiro, tendo em vista que estudos comprovam que o pronome inovador está ganhando cada vez mais espaço como expressão de primeira pessoa do plural a cada década que passa.

Pontuamos que as pesquisas descritas no Gráfico 21 apresentam algumas diferenças metodológicas quanto à amostra utilizada. Omena (2003), por exemplo, coletou dados de informantes que não possuíam o nível superior, diferente de *nós*, que analisamos somente a fala de informantes com nível superior completo. Lopes (2003) e Nascimento (2013), assim como *nós*, utilizaram amostras de fala culta, porém as amostras desses estudos foram compostas somente por inquéritos do tipo DID (diálogo entre informante e documentador), enquanto nossa pesquisa conta com inquéritos dos três tipos (DID, D2 e EF). No entanto, apesar dessas diferenças nas amostras analisadas, as pesquisas se assemelham em termos percentuais, no sentido de o pronome *a gente* estar cada vez mais se expandindo no português do Brasil, em um processo que envolve indícios de mudança linguística (Lopes, 2003).

Cabe ponderar, no entanto que, apesar das altas taxas de uso que o pronome inovador *a gente* apresenta nas comunidades de fala estudadas em tempo real (ver gráfico 21), as variantes de primeira pessoa do plural ainda coexistem dentro do sistema linguístico dessa comunidade, em maior ou menor grau, e continuam sendo usadas em diferentes contextos de comunicação, seja em ambientes que exigem maior monitoramento da fala, como é o caso dos registros do tipo EF, seja aqueles em que o falar é mais espontâneo, como os inquéritos do tipo

D2. Desta forma, podemos constatar que, apesar da alta proporção de uso da forma inovadora, o fenômeno ainda constitui uma regra variável, conforme evidencia Labov (2003). Ou seja, o estágio de variação e mudança iniciado no século XVIII (Lopes, 2003), ainda não se completou no português falado em diferentes regiões do Brasil, estando as duas formas ainda em competição.

O confronto entre as duas décadas possibilita o reconhecimento dos ambientes linguísticos mais propensos à expansão da forma inovadora, considerando que os fatores que agiam sobre a escolha entre os pronomes no passado podem ser os mesmos que agem na década atual, conforme descreve o Princípio do Uniformitarismo. Para além disso, pretendemos testar a Hipótese da Taxa Constante (Constant Rate Hypothesis), também conhecida como Efeito de Taxa Constante, difundida por Kroch (1989), que prescreve que, em um processo de mudança linguística em que se observa duas variantes linguísticas em competição entre si, a substituição de uma forma pela outra segue uma taxa constante, e “embora a taxa de utilização de opções gramaticais na competição seja geralmente diferente entre contextos e em cada período de tempo, a taxa de mudança será a mesma entre contextos” (Kroch, 1989, p. 6). Para o autor, técnicas estatísticas podem ser empregadas para comparar dados de diferentes épocas e, a partir de um modelo matemático, estimar parâmetros que são capazes de responder às questões referentes à variação e mudança linguística ao longo do tempo, conforme descreve:

Uma taxa constante de mudança entre contextos é matematicamente equivalente à fixidez dos efeitos contextuais, em direção e tamanho, ao longo dos períodos de tempo. Assim, se um estudo reportar uma série de análises multivariadas para diferentes períodos de tempo, e os efeitos contextuais forem constantes ao longo destas análises, a taxa de mudança de cada contexto medida separadamente seria necessariamente a mesma. Esta equivalência é válida porque, em termos estatísticos, a hipótese da taxa constante é a afirmação de que a taxa global de utilização de um formulário é independente dos efeitos contextuais sobre a sua utilização. Esta independência estatística pode ser expressa modelando o uso das formas antigas e novas em substituição em um conjunto de equações. (Kroch, 1989, p. 6)

Testar a hipótese da taxa constante pode ser importante para explicar o fenômeno variável, pois pode fornecer dados que ajudam a compreender algumas questões que são centrais nos estudos sociolinguísticos, como o problema da transição e da implementação. Durante a transição de uma forma linguística variante para a outra, por exemplo, a hipótese defendida por Koch sugere que essa taxa de substituição entre as formas é constante em todos os contextos de mudança linguística, sendo a velocidade com que os membros de determinada comunidade adotam a nova variante igualmente considerada uniforme. Isso, na prática, sugere que é possível prever a direção da mudança e os contextos mais favoráveis à expansão de

determinada variante durante o período de transição. A implementação de uma nova forma variante, considerando a dinâmica temporal da mudança, pode ocorrer de forma gradual ou rapidamente, a depender de fatores, como o prestígio social de determinada forma variável, a resistência individual à mudança, a aceitação de novas variantes entre os diferentes grupos sociais, entre outros. Nesse sentido, a hipótese da taxa constante ajuda a compreender como determinada variante se implementa e se espalha no seio da comunidade de fala, até que seja incorporada ao sistema linguístico.

Lopes (2003), na busca pelos ambientes mais favoráveis ao uso da forma pronominal *a gente*, verificou tendências que são constantes em cada amostra analisada. Primeiramente, a autora procedeu uma análise geral através de ferramenta estatística, englobando todos os dados de seu estudo. Em seguida, procedeu uma rodada específica para cada amostra separadamente, com o objetivo de verificar se a tendência geral observada na comunidade se repete nas análises individuais. Ao confrontar o resultado da análise envolvendo todos os dados com o resultado das amostras individualmente, observando os contextos mais propícios ao uso do pronome *a gente*, constatou-se “uma coincidência quase total entre os grupos de fatores selecionados na rodada geral” (Lopes, 2003, p. 123). Esse resultado pode sugerir que as forças linguísticas e sociais que agiam sobre o comportamento geral da amostra analisada são as mesmas que influenciavam a realização do fenômeno na década mais antiga e que seguiram influenciando na década mais atual.

O trabalho conduzido por Lopes (2003) é considerado referência na análise de primeira pessoa do plural em tempo real, motivo pelo qual decidimos replicar esse procedimento metodológico em nossa análise, com devidas adaptações. Com o auxílio do ambiente *RStudio*, por meio de um modelo de regressão logística, realizamos uma análise geral englobando todos os dados de nossa pesquisa, ou seja, os dados das amostras representativas das décadas de 1990 e 2020. Nesse ponto, por analisar as amostras das duas sincronias em uma mesma rodada, passamos a observar a variável *década de gravação*, constituída por dois níveis: Anos 1990 e Anos 2020. Submetemos o modelo geral com os dados das duas décadas à seleção automática das variáveis, por meio da função *step*, que forneceu informações sobre o comportamento geral da amostra no que se refere ao uso dos pronomes de primeira pessoa do plural. Em seguida, analisamos separadamente cada amostra. Por fim, atendendo ao prescrito por Labov (1994), para uma análise em tempo real de curta duração, comparamos a relevância de cada variável previsoras nas duas sincronias (décadas de 1990 e 2020). O quadro abaixo apresenta os ambientes mais favorecedores ao uso do pronome *a gente* na rodada geral (com todos os dados) e nas rodadas específicas de cada década analisada:

**Quadro 11:** Comparação entre a tendência geral na amostra e nas análises de 1990 e de 2020

<b>Análise geral (1190 e 2020)</b>	<b>Década 1990 (Fase I)</b>	<b>Década 2020 (Fase II)</b>
1. Década de gravação	1. Referencialidade	1. Tipo de inquérito
2. Tipo de verbo	2. Tipo de verbo	2. Faixa etária
3. Referencialidade	3. Faixa etária	3. Tipo de verbo
4. Tipo de inquérito	4. Tipo de inquérito	4. Saliência fônica
5. Faixa etária	5. Saliência fônica	5. Paralelismo formal
6. Paralelismo formal	6. Tempo verbal	6. Referencialidade
7. Saliência fônica		
8. Tempo verbal		

Fonte: elaboração própria.

Como visto, a hipótese da taxa constante prevê que a tendência da amostra que engloba todos os dados seja a mesma observada nas análises das amostras específicas. Nesse caso, espera-se que os fatores selecionados na análise em que engloba todos os dados das amostras sejam os mesmos selecionados na análise individual de cada sincronia. O resultado da análise demonstra que, após submeter o modelo que inclui as duas décadas à seleção a partir da função *step* no ambiente *RStudio*, quase todas as variáveis predictoras testadas no modelo foram selecionadas como significativamente relevante, sendo apenas a variável *sexo/gênero* removida do modelo estatístico através da seleção automática.

Conforme demonstra o panorama das análises estatísticas realizadas, representada no Quadro 11, a variável *década de gravação* foi a primeira selecionada na análise que engloba todos os dados de nosso estudo, demonstrando que o uso dos pronomes de primeira pessoa do plural apresenta comportamento estatisticamente diferente a depender da década de gravação, como já atestado nas análises individuais de cada sincronia (ver subseções 5.1 e 5.2). Além disso, na rodada geral, as variáveis predictoras de natureza linguística demonstraram significância estatística, sendo todas selecionadas. Já em relação às variáveis sociais, se mostraram relevantes o *tipo de inquérito* e a *faixa etária*, sendo a variável predictoras *sexo/gênero* a única que teve sua exclusão recomendada pelo programa estatístico. Em termos de tendência, se observa que quase todas as variáveis predictoras selecionadas na análise geral se repetem na análise de cada amostra separadamente. A exclusão da variável *sexo/gênero* nas três análises também demonstra uma tendência no uso dos pronomes que se repete em cada década analisada.

Por outro lado, a análise de cada uma das sincronias aponta comportamento muito parecido quanto a influência dos preditores. Cabe destacar que os dados apresentados nesta análise comparativa correspondem aos dados incluídos em cada modelo de regressão logística, envolvendo as duas amostras. Sendo assim, relembramos que as ocorrências referentes ao nível 1 de saliência fônica foram removidos da análise para que o modelo pudesse convergir e então ser submetido à seleção automática através da função *step*, restando, portanto, um total de 4534 observações dos pronomes *nós* e *a gente* nas amostras provenientes do projeto PORCUFORT Fase I (Amostra 1990) e Fase II (Análise 2020).

Em relação a análise dos dados representativos da década de 1990 (PORCUFORT Fase I), a função *step* indicou um modelo ideal para tratamento dos nossos dados que inclui as variáveis *grau de referencialidade do pronome*, *tipo de verbo*, *faixa etária*, *tipo de inquérito*, *saliência fônica* e *tempo verbal*. Além disso, a ferramenta recomendou a remoção das variáveis *sexo/gênero*, como dito, e *paralelismo formal*, que não apresentaram relevância estatística na análise ( $p\text{-value} > 0.05$ ). Já a análise da amostra representativa da fala culta de Fortaleza da década de 2020 apontou relevância estatística das variáveis preditoras *tipo de inquérito*, *faixa etária*, *tipo de verbo*, *saliência fônica*, *paralelismo formal* e *grau de referencialidade do pronome*. A seleção automática das variáveis recomendou a retirada dos preditores *sexo/gênero* e *tempo verbal*. Com isso, dos sete grupos de fatores analisados nos modelos separadamente (o que exclui a variável *década de gravação*), houve coincidência de pelo menos cinco variáveis selecionadas em todas as análises, demonstrando que os contextos favoráveis ao uso do pronome inovador na década mais antiga continuam a favorecer o pronome *a gente* na década mais atual. A seguir, detalhamos o comportamento do fenômeno variável em tempo real de curta duração de tendência nas duas décadas analisadas em relação aos preditores incluídos no modelo.

#### a) *A variável faixa etária*

Entre essas variáveis preditoras, a *faixa etária* exerce uma função muito importante na observação do fenômeno variável. Como vimos, a análise em tempo aparente de determinada sincronia leva em conta a distribuição das ocorrências do fenômeno variável em função de cada faixa etária e permite inferir se o processo de variação apresenta estabilidade, mudança em progresso ou mudança completa. Já na análise em tempo real é possível observar, em vez de inferir, os estágios da mudança com base nos dados de fala de diferentes grupos etários que

compõem a amostra (Labov, 2008 [1972]). Análise em tempo real do tipo tendência [*tend study*] permite confrontar o comportamento da comunidade de fala em duas sincronias, comparando a distribuição das ocorrências entre as faixas etárias em diferentes pontos do tempo, a fim de constatar se o comportamento dessa comunidade continua o mesmo, o que pode caracterizar uma gradação etária, ou se há indícios de mudança em progresso, impulsionado pelos mais jovens. A tabela abaixo apresenta o resultado da análise comparativa do uso do pronome inovador *a gente* em relação à *faixa etária* dos informantes nas duas sincronias.

**Tabela 30:** Distribuição de *nós* e *a gente* por *faixa etária* em tempo real

Faixa etária	Anos 1990			Anos 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
22 a 35 anos	63,7%	1.52171	***	88,6%	1.1090	***
36 a 50 anos	66,3%	-0.09866		86,3%	-0.5514	**
A partir de 51 anos	37,7%	-1.59842	***	69,4%	-1.5087	**
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

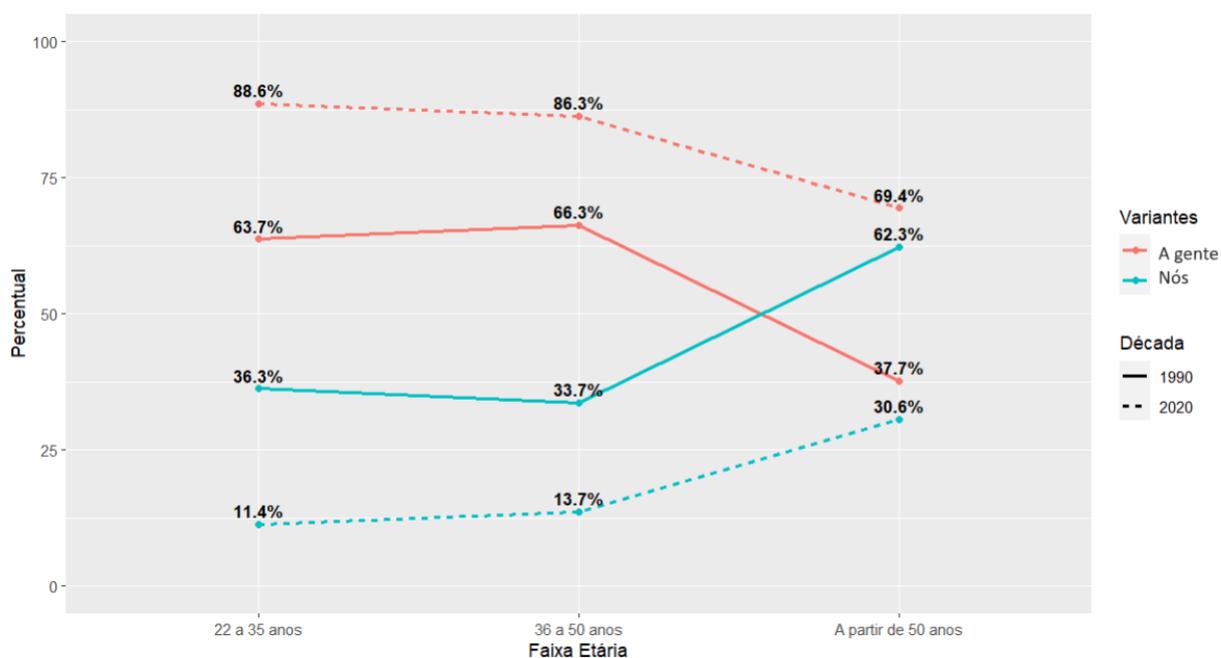
Os dados presentes na Tabela 30 ilustram a evolução da forma pronominal *a gente* em tempo real de curta duração. Conforme é possível observar, há aumento significativo no percentual de uso da variante em todas as faixas etárias no intervalo de mais de vinte anos. Entre os mais jovens da amostra (informantes de 22 a 35 anos), a diferença no uso de uma década para outra representa um aumento de 24,9 pontos percentuais, indicando uma tendência crescente no uso dessa variante entre os jovens. Em termos de favorecimento, as estimativas apontam os informantes mais jovens como fortes aliados da forma *a gente* em ambas as décadas (1.52171 e 1.1090, respectivamente). Além disso, a atuação dessa categoria se mostrou estatisticamente significativa em ambas as décadas (\*\*\*). Esses resultados evidenciam uma expressiva preferência pela variante inovadora entre os mais jovens, apontando para uma possível mudança em curso.

A mesma tendência é observada entre os informantes de 36 a 50 anos, que, assim como os mais jovens, apresentaram evolução na proporção de uso do pronome *a gente* de uma década para outra. A diferença no percentual entre as sincronias é de 20 pontos percentuais. Esse aumento na proporção pode sugerir uma mudança na preferência linguística dos informantes, sendo o pronome inovador mais evidente a cada década. As estimativas de uso entre os informantes desta faixa etária demonstram discreto desfavorecimento no uso deste pronome, se comparado aos mais jovens (*intercept*). No entanto, na década de 1990, não há

significância estatística atribuída a informantes desta faixa etária, diferentemente da década de 2020, em que este fator se mostrou relevante (\*\*). Essa diferença na relevância estatística sugere uma mudança no comportamento dos informantes em relação ao uso dos pronomes, sendo todos os informantes na década mais atual mais propensos a interagir significativamente para a realização do pronome *a gente*.

Em relação aos informantes mais velhos da amostra (a partir de 51 anos), os dados da Tabela 30 destacam a progressão do pronome *a gente* entre dois pontos do tempo. A comparação em tempo real entre as duas sincronias revela uma diferença no uso que chega a 31,7 pontos percentuais. Em relação às estimativas de uso, a análise identificou forte desfavorecimento da forma pronominal inovadora em ambas as décadas entre os mais velhos, ainda que os resultados referentes a amostra da década 2020 demonstrem maior proporção de uso deste pronome. Em ambas as décadas, essa *faixa etária* apresentou interação estatisticamente significativa, conforme aponta a Tabela 30. Para uma representação visual dos resultados, o Gráfico 22 apresenta a tendência no comportamento das variantes em tempo real nas três faixas:

**Gráfico 22:** Proporção uso de *nós* e *a gente* por *faixa etária* em tempo real (N = 4534)



Fonte: elaboração própria.

Conforme é possível observar no Gráfico 22, na década de 1990, os percentuais indicam maior proporção de uso da forma pronominal *a gente* entre os informantes de 36 a 50 anos (66,6%), sendo esta a faixa etária mais favorável ao uso do pronome inovador. Por outro

lado, vê-se menor proporção de uso do pronome *nós* (36,2%). Entre os informantes de 22 a 35 anos, os mais jovens da amostra, há alta proporção de uso de *a gente* (63,8%), e uso menos recorrente do pronome *nós* (36,2%). No que se refere aos informantes mais velhos, acima de 50 anos de idade, esses utilizam preferencialmente o pronome padrão *nós* (61,2%), em detrimento do pronome inovador *a gente* (38,8%) nos anos 90, sendo a única faixa etária pesquisada na amostra a apresentar maior uso para o pronome canônico.

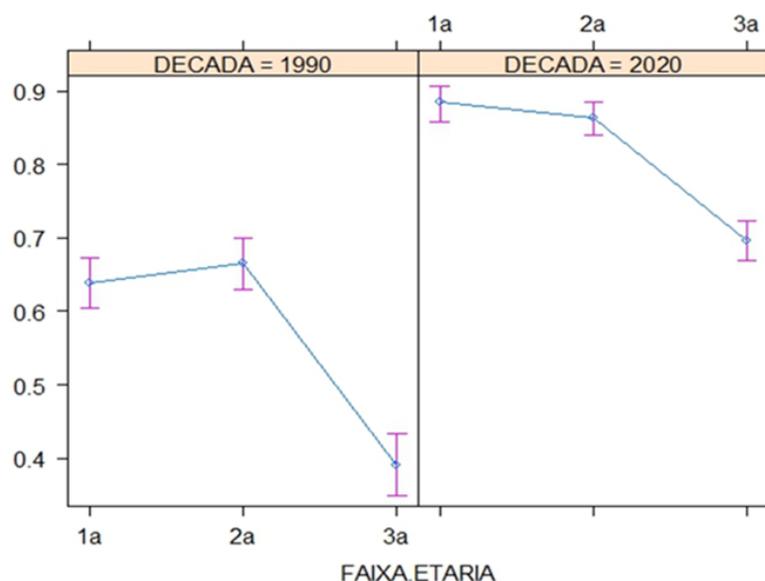
A análise em tempo real apontou aumento significativo da forma inovadora de uma década em relação a outra. Como prevíamos, a forma pronominal *a gente* é mais frutífera que o pronome *nós* em todas as faixas etárias representativas da década de 2020, sendo utilizada com em maior proporção entres os mais jovens (22 a 35 anos), com 88,5% das ocorrências, e menos intensamente entre os mais velhos (a partir de 51 anos), embora estes ainda prefiram a forma inovadora, apresentando proporção de uso que corresponde 69,7% do total de ocorrências. Destacamos que, na amostra referentes à década de 2022, nenhuma das faixas etárias pesquisadas apresentou maior proporção de uso para o pronome conservador *nós*, conforme é possível ver no Gráfico 22. Esse crescimento no percentual de uso da forma pronominal *a gente* em todos os níveis desta variável previsora pode apontar para indícios de que estamos diante um fenômeno em mudança linguística em progresso. Esses indícios podem ser observados tanto em tempo aparente quanto em tempo real, ao observar a distribuição dos dados de fala em função da *faixa etária* e, mais atentamente, a curva estatística que representa a variação em dois pontos no tempo. Sobre essa questão, são pertinentes as palavras de Maia (2009):

Uma variação pode manter-se estável ou evoluir para uma mudança em progresso. No segundo caso, a variação tem maiores índices de realização nos grupos sociais centrais, resultando em representações gráficas de padrão curvilíneo. Além disso, os falantes mais jovens tendem a realizar a variante inovadora mais frequentemente que os falantes mais velhos (evidência de tempo aparente). A essas duas evidências, Faixa Etária e Padrão Curvilíneo, associa-se a observação das transformações ocorridas ao longo do tempo, denominada evidência de tempo real. Se há variação estável, uma variante ocorre mais nas classes mais altas, e a outra variante ocorre mais nas classes mais baixas. A representação gráfica desse processo resulta em uma linha com vários picos, não havendo hierarquia entre as faixas etárias. (Maia, 2009, p. 49).

Como é possível observar na representação exposta no Gráfico 21, o comportamento linguístico da comunidade, no que diz respeito ao uso dos pronomes de primeira pessoa do plural, não se manteve completamente estável. As transformações que ocorreram referentes ao uso das variantes ao longo dos anos na comunidade de fala são visíveis e apontam uma hierarquização entre as faixas etárias. Além disso, o padrão curvilíneo que representa a atuação de cada *faixa etária* sobre a realização dos pronomes pode sugerir uma

mudança significativa, sendo mais visível entre os informantes mais velhos, em que a curva apresenta uma interseção. Esse ponto de interseção na curva estatística indica mudança no comportamento linguístico da comunidade, principalmente entre os falantes mais velhos que, na década de 1990, preferiam o uso do pronome padrão *e*, duas décadas depois, apresentam preferência pelo uso da variante inovadora (ver Gráfico 21). Essa constatação demonstra que o uso do pronome *a gente* se expandiu em todos os contextos ao longo do tempo, inclusive em ambientes em que antes sofria restrição, como era o caso dos informantes mais velhos, que inverteram o comportamento linguístico. O Gráfico 23 demonstra mais efetivamente a curva estatística que representa o uso do pronome *a gente* em tempo real, em análise comparativa de duas décadas.

**Gráfico 23:** Tendência de uso de *a gente* por faixa etária em tempo real



Fonte: elaboração própria

A comparação em tempo real de curta duração (Amostra 1990 e Amostra 2020 - Gráfico 23) apontou uma tendência para maior uso da forma inovadora entre os falantes da primeira faixa etária (22 - 35 anos). Por outro lado, entre os informantes da terceira faixa etária (a partir de 51 anos), há uma queda significativa do uso de *a gente* nas duas sincronias analisadas. Para Labov (2001[1974]), as diferenças linguísticas sincrônicas de uma comunidade de fala refletem desenvolvimentos diacrônicos da língua, ou seja, em tempo real. Dessa forma, buscar tendências que se repetem na mesma comunidade de fala em amostras com intervalo de tempo é importante para revelar possíveis estágios da variação e mudança linguística. Sobre essa questão, são pertinentes as palavras de Tavares (2011):

O quadro de inter-relações linguísticas delineado hoje é reflexo dos usos anteriores dados à língua por seus falantes e é a base dos usos futuros, em um contínuo de pequenos incrementos inovadores levando a grandes mudanças. Sendo assim, os indícios de mudança linguística podem ser buscados em estudos que envolvem dados de tempo real (Tavares, 2011, p. 397).

Utilizando-se dessa óptica sobre nosso estudo, o que se observa no Gráfico 23 é um padrão de uso do pronome inovador na fala culta de Fortaleza, ou seja, em ambas as amostras e nas três faixas etárias houve um aumento significativo nas proporções de uso do pronome *a gente* se comparada a década mais atual à mais antiga, sendo maior entre os mais jovens (22 a 35 anos) e menor entre os mais velhos (acima de 50 anos). Além disso, na amostra representativa dos anos 90, percebe-se queda acentuada no percentual de uso do pronome inovador entre os mais velhos, conforme é possível contemplar no Gráfico 23. Essa tendência se repete na amostra dos anos 2020, ou seja, entre os informantes com mais de 50 anos, observa-se uma queda significativa na proporção de uso do pronome *a gente* em comparação às outras faixas etárias analisadas, mesmo que o pronome inovador passe a ser preferido entre os mais velhos da Amostra 2020, demonstrando que a tendência no uso da variante se manteve, mesmo que as proporções de uso tenham aumentado. A mesma constatação também foi feita por Lopes (2003), quando observou, em um estudo em tempo real, que os falantes da década de 1970 empregavam o pronome *a gente* com mais frequência entre os mais jovens, com uma queda no percentual de uso dos mais velhos, e esse tendência se repetiu na década de 1990, em que houve aumento da frequência de uso do pronome inovador, revelando que o uso de *a gente* se torna mais evidente de uma década para outra, mas a tendência do uso dos pronomes pode se manter nas duas sincronias.

#### *b) Tipo de inquérito*

A segunda variável de natureza extralinguística que foi incluída no modelo de regressão nas análises com os dados das duas sincronias foi o *tipo de inquérito*. Como dito, esse preditor não caracteriza qualquer traço social dos informantes, mas apenas descreve a modalidade de gravação e o nível de atenção dada a fala em cada tipo de registro que constitui o projeto PORCUFORT Fase I (Amostra 1990) e Fase II (Amostra 2020). Nesse sentido, no modelo laboviano, o conceito de estilo está intimamente ligado ao grau de monitoramento ou atenção dada à fala. Conforme discutimos, a atenção dada a fala representa um *continuum*, que

vai do menos formal, com pouco monitoramento da linguagem empregada, até o mais formal, em que o monitoramento é maior.

Em nossas amostras, o estilo empregado pelos informantes cultos de Fortaleza pode ser analisado através de dois parâmetros metodológicos, a saber, em tempo aparente e em tempo real. Sincronicamente (em tempo aparente), a análise do tipo de inquérito pode revelar tendências no uso das variantes e prever ambientes mais favoráveis à implementação da forma inovadora. A teoria variacionista prescreve que o uso de variantes inovadoras pode ocorrer em maior proporção em registros mais casuais, enquanto as variantes padrões tendem a ser usadas com mais frequência em contextos mais formais (Labov, 2008[1972]). Em tempo real, pode-se observar se o fenômeno apresenta indícios de que está em variação estável ou em uma mudança em progresso na comunidade de fala. Em caso de mudança no comportamento linguístico, a análise desta variável permite observar a direção e o estágio dessa mudança, os contextos que ainda atuam de forma a restringir ou não a implementação da forma inovadora no intervalo de mais de vinte anos, e quais registros se tornaram mais favoráveis. O confronto entre os resultados da atuação desse preditor em tempo real pode ser observado na Tabela 31.

**Tabela 31:** Distribuição de *nós* e *a gente* por tipo de inquérito em tempo real

Tipo de inquérito	Anos 1990			Anos 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
DID	60,8%	1.52171	***	79,1%	1.1090	***
D2	72,3%	0.19933		91,4%	1.1513	***
EF	42%	-1.36684	***	66,4%	-1.2678	***
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

Como explanado, os tipos de inquérito representam um *continuum* de formalidade que vai do menor ao maior grau de formalidade. Eckert (2004) defende a ideia de que essa escala de monitoramento da fala representa o que denominou de *eixo de formalidade*, e define as categorias de estilo num gradiente. Em relação aos nossos dados, cabe lembrar, os inquéritos do tipo diálogo entre dois informantes (D2) estão na base desta escala, sendo o menos formal de nossa amostra. As gravações de diálogos entre informante e documentador (DID), representam o ponto médio da escala, com grau intermediário de atenção dada à fala. Por fim, os registros do tipo elocução formal (EF), apresentam alto grau de monitoramento, sendo o tipo de registro mais formal de nossa amostra.

Tendo isso em vista, o confronto entre as duas sincronias revela tendências que se

repetem nas duas décadas analisadas. Essa tendência é observada com maior intensidade na década de 2020. Em termos de favorecimento, inquéritos do tipo DID (*intercept*) se mostraram propensos à realização do pronome *a gente*, tanto na década de 1990, tendo estimativa de 1.52171, quanto na de 2020, com estimativa de uso de 1.1090. Observa-se, na comparação entre os resultados das duas amostras, expansão no uso do pronome inovador em inquéritos deste tipo, representando um aumento de 18,3 pontos percentuais. Essa tendência também é observada na amostra geral, que demonstrou aumento significativo do pronome inovador no decorrer de vinte anos. Além disso, inquéritos do tipo DID se mostraram significativamente relevantes para o uso dos pronomes nas duas análises (\*\*\*)

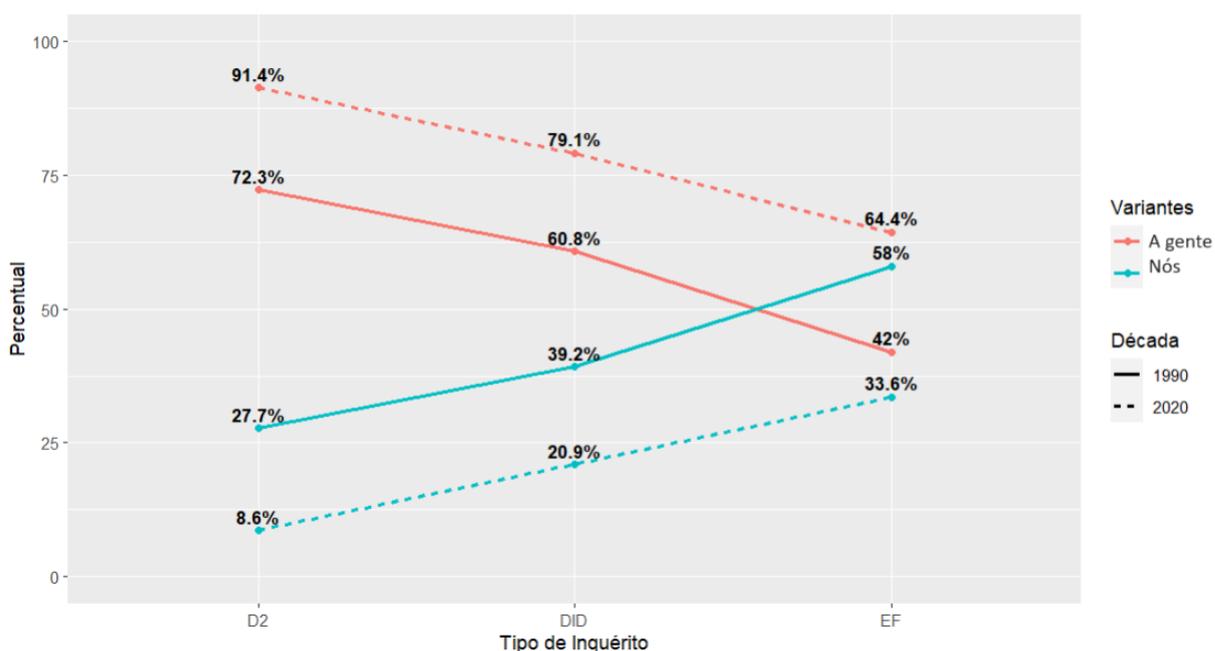
Registros do tipo D2, em ambas as amostras, apresentam tendência ao expressivo favorecimento do pronome *a gente* inovador, conforme apontam as estimativas de uso presentes na Tabela 31. Em dados do PORCUFORT Fase I (Amostra 1990), há significativa tendência ao uso de *a gente* que, em relação ao *intercept*, apresenta estimativa igual a 1.72104. A mesma tendência é constatada nos dados de fala culta provenientes do PORCUFORT Fase II (Amostra 2020) que, em relação aos inquéritos DID, demonstraram intenso favorecimento ao uso da variante inovadora, com estimativa de 2.2603. Ou seja, os inquéritos do tipo D2 representam um ambiente ainda mais propenso ao uso de *a gente* na década de 2020, indicando um contexto de mudança linguística, em que, aos poucos, parece estar ocorrendo a substituição da variante padrão pela inovadora. Apesar desses valores indicarem favorecimento, os dados deste tipo de registro da Amostra 1990 não se mostraram estatisticamente relevantes para a análise do fenômeno, diferente da Amostra 2020, que apresentou forte interação sobre a realização da variável resposta (\*\*\*)

Os mais formais dos registros, as elocuições formais (EF), indicam desfavorecimento da variante inovadora nas análises das duas décadas. Em relação ao *intercept* do modelo, nota-se maior probabilidade de uso da variante padrão *nós* em registros do tipo EF em ambas as amostras. As estimativas, ainda que sejam parecidas nas duas amostras, apresentam diferenças, mesmo que discretas. Os dados da década de 1990 demonstram maior disposição ao uso do pronome canônico *nós*, se comparados aos dados da década de 2020. A análise sugere a implementação de forma inovadora, mesmo que de forma menos intensa, em ambientes mais formais na década atual. Esses resultados são consistentes com a teoria da variação de mudança linguística, que aponta o estilo de fala mais formal como o menos favorável ao uso de variantes inovadoras.

Em termos de favorecimento, as estimativas de uso das variantes nos forneceram uma fotografia sociolinguística que revelou a atuação do tipo de registro sobre a realização dos

pronomes, que constatou uma tendência ao maior favorecimento do pronome inovador nos registros de menor formalidade, e do pronome padrão nos registros de maior formalidade nas duas décadas, demonstrando que, embora haja diferença nos percentuais, a tendência no uso dos pronomes se mantém na década mais atual. Em termos percentuais, a mesma tendência é observada, o Gráfico 24 ilustra a mudança na proporção de uso dos pronomes *nós* e *a gente*, distribuídos em função da variável *tipo de inquérito*, em duas décadas.

**Gráfico 24:** Tendência no uso de *nós* e *a gente* por *tipo de inquérito* em tempo real (N=4573)



Fonte: elaboração própria

O Gráfico 24 aponta para uma tendência da comunidade de fala pesquisada relacionadas ao uso dos pronomes e o grau de atenção dado à fala. Conforme podemos observar, na década de 1990, a forma pronominal inovadora era predominante utilizada nos registros do tipo D2 (72,3%), demonstrando ser o registro mais favorável ao uso de *a gente* em nossa amostra. Por outro lado, o pronome padrão era pouco utilizado (27,7%). Nos inquéritos do tipo DID, identificamos a mesma tendência, de forma menos intensa, no que se refere ao uso dos pronomes na década de 1990, ou seja, maior ocorrência de *a gente* (60,8%) e uso menos recorrente do pronome *nós* (39,2%). As elocuições formais (EF) são os únicos registros que se apresentam mais favoráveis ao uso do pronome padrão *nós* (58%) na década de 1990, sendo a variante inovadora menos utilizada nesta modalidade (42%). Cabe destacar que, entre todos os tipos de inquéritos analisados na Amostra 1990, os tipos EF são os únicos mais favoráveis ao uso da forma pronominal padrão *nós*.

Na década de 2020, a análise em tempo real apontou aumento circunstancial nas proporções de uso da forma inovadora, ao passo que o uso do pronome *nós* vem diminuindo em cada tipo de inquérito. Esse dado pode demonstrar um processo de mudança linguística, em que o uso de *a gente* se torna mais frequente a cada década, independentemente do estilo de fala empregado. Em relação aos inquéritos do tipo D2, as realizações do pronome *a gente* chegam a 91,4%, índice próximo do categórico, enquanto o pronome *nós* ocorreu apenas em 8,6% dos dados. Esses resultados demonstram que, na fala menos monitorada, o fenômeno pode abrir caminho para uma substituição no uso de *nós* por *a gente*.

Comportamento parecido se observa ao analisar os registros do tipo DID, em que houve crescimento significativo nos percentuais de uso da forma inovadora, indicando ampla expansão da forma em contextos de formalidade moderada, saltando de 60,8% para 79,1%, uma diferença de 18,3 pontos percentuais. O pronome *nós*, no entanto, obteve retração no uso, despencando de 39,2% para 20,9%. Já nos inquéritos mais formais da amostra (EF), houve inversão na preferência dos informantes pelas variantes, elevando significativamente a proporção de uso do pronome *a gente* de 40,2% para 66,4% e, com a mesma intensidade, houve retração do pronome padrão em neste tipo de registro. A progressão no uso do pronome *a gente* em tempo real confirma a forte expansão da variante inovadora entre os falantes cultos de Fortaleza, inclusive em ambientes em que, na década de 1990, sofria resistência, como é o caso dos contextos de maior formalidade. Vale destacar que, apesar da evolução no uso do pronome *a gente* em todos os tipos de inquérito, em termos de tendência, a comunidade se manteve estável: uso mais intenso do pronome *a gente* nos inquéritos do tipo D2, e menos intensos em inquéritos do tipo EF nas duas décadas.

Para Labov (2001), o estilo empregado na fala remete a noções de prestígio ou de estigmas que determinadas variantes podem receber através da avaliação social. A implementação de *a gente* em duas décadas sugere que a expansão do pronome inovador está acelerada na comunidade de fala de Fortaleza. No entanto, consideramos importante correlacionar esses resultados a estudos de percepção, crença, avaliação linguística em que o *a gente* é considerado informal e o *nós* goza de prestígio, formalidade e de uso preferencial da escrita e da fala mais monitorada (Vitório, 2015; Freitas, Favacho e Carvalho, 2020).

Nesse sentido, uma possível mudança linguística de *nós* por *a gente* pode não se completar facilmente em ambientes mais formais por questões normativas ou por pressões sociais. No entanto, os resultados da comparação em tempo real parecem sugerir que a forma *a gente* não sofre estigma, essencialmente na década de 2020, uma vez que se tornou a referência

à primeira pessoa do plural preferida entre os falantes de nível superior completo, inclusive em ambientes com alto nível de formalidade e de monitoramento da fala.

c) *Grau de referencialidade do pronome*

A variável *grau de referência do pronome*, também nomeada de *referência do sujeito*, *grau de referência dos pronomes* e *tipologia semântica do sujeito*, tem se destacado como relevante para a análise do fenômeno variável de primeira pessoa do plural, conforme demonstram as pesquisas de Machado (1995), Tamanine (2002, 2010), Francischini (2011), Araujo (2016), Mendes (2019), Souza (2020), entre outros. Como explicado, a forma pronominal *a gente* originou-se do substantivo *gente*, que levava consigo uma carga de generalização, uma vez que o substantivo era utilizado para se referir a um grupo indeterminado de pessoas. Nesse sentido, ao se pronominalizar, a forma pronominal *a gente* manteve, do substantivo que a originou, o traço generalizador que carrega até os dias atuais.

Estudos sociolinguísticos têm observado que o pronome inovador está fortemente associado ao traço de generalização do referente, enquanto o pronome padrão é favorável em contextos de referencialidade específica. Essa característica no uso do pronome *a gente* quanto ao grau de referencialidade é mais marcante em sincronias mais antigas. Lopes (1993, p 130), em estudo sobre a fala carioca dos anos 1970, verificou que “há a diferenciação no emprego de *nós* e *a gente* em relação a um uso mais restrito ou mais genérico”. Segundo a autora, o falante usa o pronome *nós* para se referir, preferencialmente, a referentes específicos, enquanto o pronome *a gente* passa a ser preferido quando o informante amplia a referência, de modo a determiná-la. No entanto, estudos sobre a alternância pronominal entre *nós* e *a gente* que analisam sincronias mais recentes, têm demonstrado que o pronome *a gente*, embora permaneça o traço generalizador, se implementou e se expandiu em contextos de referencialidade específica, ganhando ainda mais espaço no uso entre os falantes no português brasileiro.

Mendonça (2022) conduziu estudo sobre a mudança nos traços de referencialidade associados aos pronomes *nós* e *a gente*, tendo como base de dados amostra da fala sergipana em dois recortes temporais, com a primeira amostra gravada no ano de 2010, e a segunda, em 2018. Dentre os resultados, a autora identificou que “as variantes de primeira pessoa do plural estão passando por um processo de mudança em função dos traços semânticos, com aumento da frequência de uso de *a gente* em contextos referenciais determinados” (Mendonça, 2023, p. 95). Neste ponto, o confronto dos resultados referentes a essa variável nas duas sincronias nos

permite analisar se, na fala culta de Fortaleza, há indícios de mudança no traço semântico genérico da forma *a gente* ao longo de vinte anos, observando se houve ou não perda da distinção semântica entre o uso dos pronomes. Os dados da análise comparativa estão dispostos na tabela abaixo:

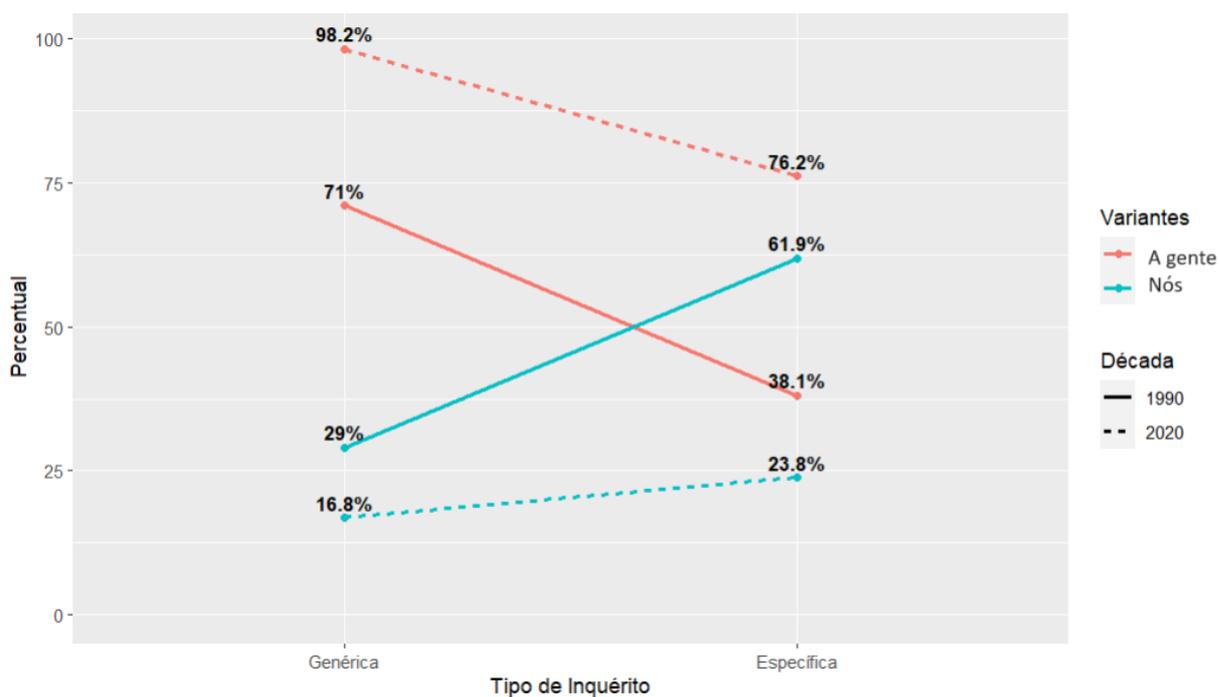
**Tabela 32:** Distribuição de *nós* e *a gente* por grau de referencialidade em tempo real

Referencialidade	Anos 1990			Anos 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
Genérica	71%	1.52171	***	83,2%	1.1090	***
Específica	38,1%	-2.04715	***	76,2%	-1.1747	***
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

Conforme podemos ver nos dados da Tabela 32, a tendência geral no uso dos pronomes também é observada ao analisarmos as ocorrências em função do grau de referencialidade do pronome, ou seja, há ampla expansão na proporção de uso da forma pronominal *a gente* em ambos os níveis que contemplam esta variável. Em termos de favorecimento, as estimativas de uso revelam tendências que se repetem nas duas sincronias analisadas: o emprego do pronome com referente genérico está fortemente associado ao pronome inovador nas duas amostras, sendo este favorecimento mais evidente na década de 1990 (1.52171) se comparado à década de 2020 (1.1090). Cabe destacar que os dados referentes a referencialidade genérica correspondem ao *intercept* em ambos os modelos. Isso quer dizer que, quando usado no contexto de realização que constitui o *intercept*, a forma inovadora é expressivamente favorecida no sentido indeterminado/genérico.

Por outro lado, quando usado com referência específica, nota-se um expressivo desfavorecimento do pronome inovador nas décadas de 1990 e de 2020, criando contextos mais prósperos ao uso do pronome padrão *nós*, conforme apontam as estimativas em ambos os modelos (-2.04715 e -1.1747, respectivamente). Em termos de proporção de uso, percebemos mudança acentuada no comportamento do fenômeno em tempo real, tendo em vista que se evidencia um aumento significativo do pronome *a gente*, inclusive em contextos de especificidade, ultrapassando o uso de *nós*, que detinha a referencialidade específica como função prototípica. Os resultados podem ser conferidos no Gráfico 25:

**Gráfico 25:** Tendência de uso de *nós* e *a gente* por grau de referencialidade em tempo real

Fonte: elaboração própria

O Gráfico 25 demonstra a evolução no uso dos pronomes em dois marcos temporais distintos. Em ambas as sincronias, mostra-se evidente evolução nos percentuais de uso da variante inovadora e, em contrapartida, uma redução significativa no uso da variante padrão. A atuação dos pronomes com referente genérico nas duas décadas evidencia uma redução significativa do pronome padrão *nós*, passando de 29% para 16,8%, característica que reflete o padrão geral das amostras. Por outro lado, observa-se aumento expressivo no uso do pronome inovador *a gente*, que passa de 71% para 98,2%. Em relação ao emprego dos pronomes com referencialidade específica na Amostra 2020, a mudança no uso dos pronomes é ainda mais acentuada. Na década de 1990, era notória a associação do pronome *nós* a contextos de especificidade, enquanto o pronome *a gente* estava claramente associado a realizações em contextos de indeterminação. Esse comportamento sofreu alteração na análise em tempo real, ou seja, em contextos de sentido específico, na década de 2020, houve significativa redução do pronome *nós* (61,9% para 23,8%), e aumento expressivo no uso do pronome *a gente* (de 38,1% para 76,2%).

De acordo com a hipótese da taxa constante de Kroch (1989), uma mudança linguística pode iniciar em determinada variedade da língua, ou em determinado contexto linguístico que pode beneficiar a expansão da nova forma e que, com o passar do tempo, essa mudança se espalha a outras variedades ou em outros contextos, se tornando universal, em uma

taxa constante de mudança. Ao observar os dados da Tabela 32, notamos alta recorrência do pronome inovador *a gente* em contextos de referencialidade genérica na década de 1990, assumindo sua função prototípica, enquanto em contextos de referencialidade específica se observa baixa proporção de uso dessa forma pronominal. Ao analisar a década mais atual, percebe-se que o pronome se espalhou em contextos que, na década passada, era pouco usual. Esse comportamento pode manifestar uma tendência universal da comunidade após quase 3 décadas, sendo o pronome inovador preferido em todos os contextos.

Além disso, podemos observar nos dados do Gráfico 25 que o uso da variante *a gente* evoluiu significativamente nas duas categorias de referencialidade analisadas em tempo real. Isso poderia ser interpretado como uma mudança linguística que está ocorrendo nos termos da hipótese de Kroch. No entanto, embora os resultados indiquem para uma possível mudança no comportamento dos pronomes *nós* e *a gente* entre as duas décadas, essas mudanças não demonstram seguir uma taxa de constância, ou seja, não se percebe um padrão constante na mudança de *nós* por *a gente*. Ao contrário, as mudanças observadas são substanciais, não uniformes e relativamente rápidas nas preferências pelo uso de um pronome em detrimento do outro, o que contradiz a hipótese da taxa constante. Essa constatação sugere que existem outros fatores atuando diretamente no comportamento dos pronomes, que vão além da ação do tempo e que estão exercendo forças sobre a rápida expansão do pronome *a gente* na comunidade analisada.

Labov (1972) defende que, se uma mudança linguística se inicia em determinada variedade da língua (digamos, na variedade popular ou na culta), com o passar do tempo essa mudança pode passar a se espalhar a outras variedades da mesma comunidade de fala. Entre os falantes com nível superior completo, percebe-se uma característica que pode estar associada mudança nos traços intrínsecos associados aos pronomes de primeira pessoa do plural, como uma possível mudança na função prototípica de especificidade do pronome *nós* na década 2020, reflexo do uso expressivo do pronome inovador. A hipótese de Labov prevê que essas mudanças podem também atingir outros tipos de variedade da língua. Nesse sentido, é pertinente também analisar como os traços semânticos dos pronomes se manifestam em outras variedades da língua, como a falada por falantes que não possuem ensino superior, comumente chamada de fala popular em estudos sociolinguísticos.

Para isso, recorreremos aos trabalhos de Silva (2020), que observou os pronomes *nós* e *a gente* na fala de Fortaleza a partir de amostras do Projeto Dialetos Sociais Cearenses - DSC, gravadas na década de 1980; e Araujo (2016), que analisou a fala popular da capital cearense com dados provenientes de Projeto Norma Oral do Português Popular de Fortaleza -

NORPORFOR, com gravações representativas do falar da década de 2000. Em ambos os trabalhos, a variável *referencialidade do pronome* foi selecionada como estatisticamente relevante. Apesar dos dois trabalhos adotarem metodologia distintas, a descrição desta variável nestas pesquisas é importante por se tratar da mesma comunidade de fala analisada nesta tese.

Silva (2020) observou, entre os falantes que não possuem nível superior em Fortaleza, maior proporção de uso do pronome inovador *a gente* em contextos de referencialidade genérica (83,3%) e uso pouco recorrente do pronome *nós* (16,7%) nos anos 1980. Ao analisar a mesma comunidade de fala após vinte anos, Araujo (2016) identificou a mesma tendência em amostra de fala popular da década de 2000: *a gente* preferencialmente usado no sentido genérico (72,6%) e uso menos frequência de *nós* com essa referencialidade (27,4%). À medida em que o uso do pronome *a gente* com emprego genérico diminui de uma década para outra, percebe-se na comparação um aumento do pronome *nós* em contexto de referencialidade genérica, contexto antes resistente ao uso do pronome padrão. Mendonça (2022) chegou a resultados próximos a esse em sua pesquisa com dados de fala de Sergipe:

Os resultados da análise em tempo real da frequência de uso de *nós* em função da referência evidenciam que o pronome canônico está aumentando sua taxa de aplicação em contextos de referência genérica. Os testes estatísticos mostram que a associação entre a frequência de uso de *nós* em cada amostra e a referência é estatisticamente significativa, mas com associação fraca entre as variáveis. (Mendonça, 2022, p. 95).

De acordo com a autora, esse resultado reflete uma possível mudança linguística nos traços de referencialidade semântica, ou seja, “a análise do traço semântico de referencialidade do pronome *nós* a partir da oposição binária determinado/genérico dá indícios de que o pronome canônico está aumentando sua frequência de uso em contextos com referentes genéricos” (Mendonça, 2022, p. 95). Esse comportamento, no entanto, é observado apenas entre os falantes não cultos nas décadas de 1980 e de 2000, tendo em vista que os falantes com nível superior completo apresentaram diminuição no uso de *nós* em realizações com referentes indeterminados/genéricos, conforme demonstra o Gráfico 25.

Em nossos dados, quando se trata da realização dos pronomes com sentido específico, o comportamento do fenômeno variável segue uma tendência. Os dados do Gráfico 25 demonstram uma interseção na linha estatística que representa o uso dos pronomes de primeira pessoa em tempo real. Isso revela inversão na tendência de uso dos pronomes de uma década para outra, ou seja, na década de 1990, o pronome inovador era menos usado em contextos de referencialidade específica, e na década de 2020, este pronome expandiu seu uso e se tornou preferido para a expressão de primeira pessoa em contextos que antes era menos

recorrente. Essa tendência expressa pode indicar uma possível mudança no traço semântico genérico associado a forma inovadora, conforme já atestou Mendonça (2022):

Embora a forma *a gente* tenha sido gramaticalizada a partir de um substantivo coletivo, com noção semântica de genericidade implícita, resultados de estudos variacionistas mostram que o pronome inovador apresenta estimativas de uso positivas em contextos referenciais determinados, dando indícios de que *a gente* está perdendo sua restrição atrelada ao valor genérico (Mendonça, 2022, p. 93).

Diante do exposto, os resultados das análises das Amostras de 1990 e 2020 em relação ao grau de presencialidade do pronome (Gráfico 25) indicam:

- a) nos anos 1990, o uso da forma *nós* com 29% para a referência genérica e 61,9% para a referência específica, mantendo sua função de referência prototípica;
- b) ainda nos anos 1990, a forma *a gente* mostrou-se associada à referência genérica (71%), mantendo resquícios de sua forma substantiva *gente*, remetendo à ideia de coletividade, indeterminação (Lopes, 2003);
- c) já na Amostra de 2020, em termos de tendência, o pronome *nós* mantém a maior proporção de uso associada à referência específica com 23,8% das observações;
- d) também na segunda Amostra, o pronome *a gente* apresenta proporção de 83,2% para referência genérica e 76,2% para referência específica, expandido seu uso para a referência específica;
- e) o uso de *a gente* em sentido específico, em 1990 e 2020 (Tabela 32), apresentou um aumento expressivo (de 38,1% na Amostra 1990 para 76,2% na Amostra 2020), indicando que a forma inovadora se implementou nos diversos contextos, especialmente nos de referencialidade específica.

#### d) Tipo de verbo

O controle desse preditor, como explanado, se justifica pela hipótese de que alguns verbos podem ser mais receptivos ao pronome inovador. Tamanine (2010) verificou que o tipo de verbo associado aos pronomes de primeira pessoa do plural pode condicionar ou, em alguns contextos, restringir o uso de determinada variante. A análise em tempo real preocupa-se em buscar indícios de aumento ou retração nos percentuais de uso de cada pronome, verificando se determinado tipo de verbo se tornou mais receptivo a forma inovadora, passando a favorecer o pronome, ou se o comportamento dos pronomes em função do tipo de verbo se manteve o mesmo no intervalo de mais de vinte anos na mesma comunidade de fala. Os resultados comparativos podem ser conferidos na Tabela 33:

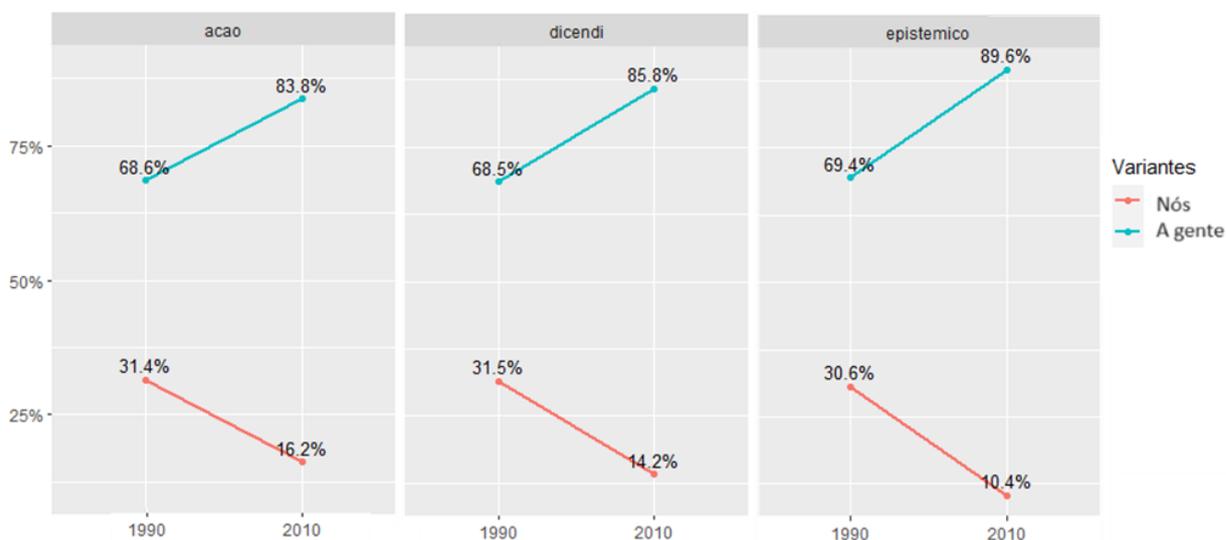
**Tabela 33:** Distribuição de *nós* e *a gente* por *tipo de verbo* em tempo real

Tipo de verbo	Anos 1990			Anos 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
Ação	68,6%	1.52171	***	83,8%	1.1090	***
<i>Dicendi</i>	68,5%	0.03386		85,8%	0.1850	
Epistêmico	69,4%	-0.32980		89,6%	0.2558	
Estado	42,6%	-1.53241	***	73,5%	-1.0600	***
Ter	29,7	-2.46290	***	60,3%	-1.8270	***

**Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1**

Fonte: elaboração própria.

Conforme podemos observar na Tabela 33, a proporção de uso do pronome *a gente* aumentou em todos os tipos de verbos observados na comparação entre duas décadas. A análise em tempo real revela, em termos de proporção de uso, dois comportamentos distintos relacionados ao *tipo de verbo* expresso junto aos pronomes. De um lado, existem verbos que eram favoráveis o uso da forma inovadora na década de 1990 e se tornaram ainda mais favoráveis ao uso desta forma pronominal na década de 2020 (*ação*, *dicendi* e epistêmico) e, por outro lado, identificamos verbos que inibiam o uso da variante inovadora na década de 1990 e passaram a ser aliados dessa variante na década atual (*estado* e *ter*), destacando a franca expansão da variante inovadora em todos os contextos relacionados ao *tipo de verbo*. Essa é a mesma tendência observada na amostra geral do estudo. A distribuição dos pronomes em relação ao *tipo de verbo* pode ser conferida no gráfico que segue.

**Gráfico 26:** Uso de *nós* e *a gente* por verbos de *ação*, *dicendi* e *epistêmico* em tempo real

Fonte: Elaboração própria

A análise dos dados apresenta os padrões de uso dos pronomes de primeira pessoa do plural distribuídos de acordo com o *tipo de verbo*. Quando realizados juntos a verbos de ação, *dicendi* e epistêmico, os resultados revelam um padrão de uso compartilhado entre esses tipos de verbos, ou seja, a evolução no uso do pronome *a gente* de uma década para outro representa uma linha estatística crescente, com aumento significativo na proporção de uso, ao passo que o uso do pronome *nós*, de uma década para outra, é representado por uma linha decrescente, indicando que o uso do pronome padrão está diminuindo significativamente entre esses verbos na análise em tempo real.

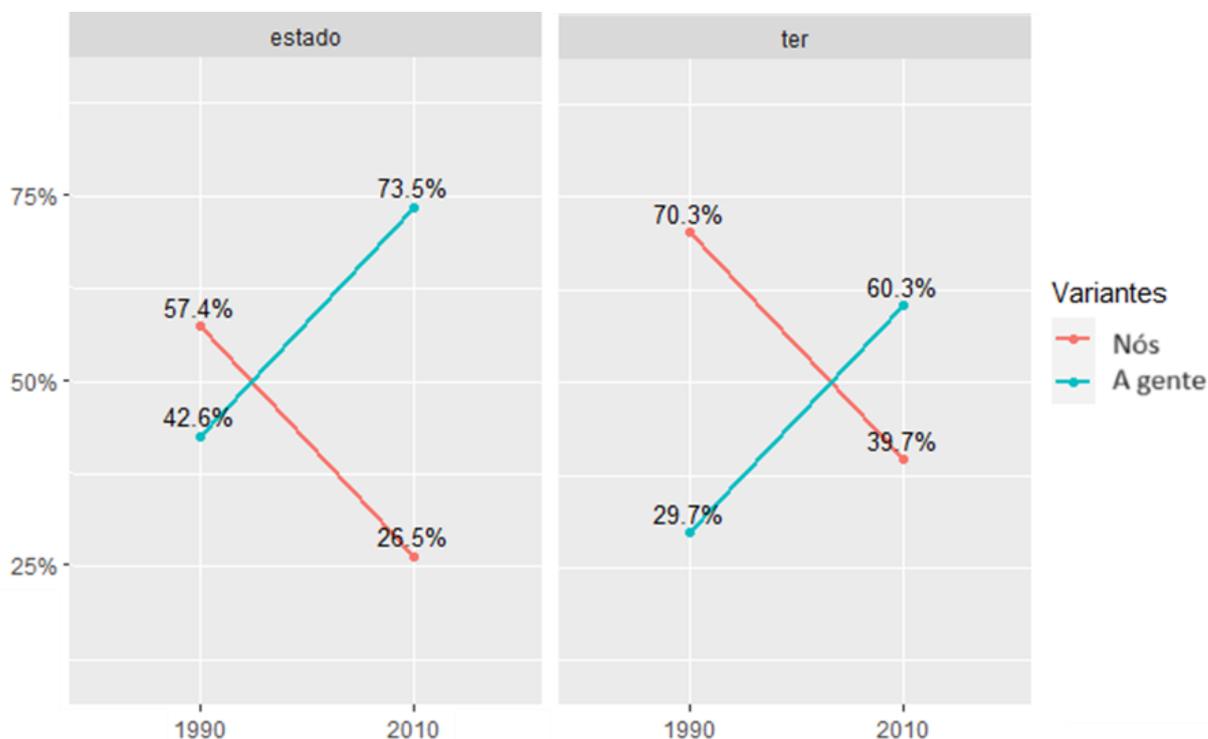
Na década de 1990, podemos identificar uma predominância do pronome inovador nestes contextos linguísticos analisados. Em relação aos verbos de ação, nota-se preferência pelo pronome *a gente*, sendo este utilizado em 68,6% dos dados de fala culta fortalezense, enquanto o pronome *nós* foi empregado em 31,4% das realizações de primeira pessoa do plural. Esse padrão também é observado na análise das observações com verbos *dicendi* e epistêmico, em que a variante inovadora é usada em maior proporção entre esses verbos (68,5% a 69,4%, respectivamente), sendo o pronome canônico menos utilizado nesses contextos (14,2% e 10,4%, respectivamente).

Na década de 2020, no entanto, nota-se uma mudança significativa neste padrão de uso dos pronomes de primeira pessoa nos contextos verbais supracitados. As ocorrências de *a gente* relacionadas a verbos de ação, por exemplo, aumentaram para 83,8% na sincronia de 2020, reafirmando a preferência pelo pronome inovador neste contexto verbal, enquanto o uso do pronome padrão *nós* diminuiu consideravelmente (de 31,4% para 16,2%). Já os verbos *dicendi* também apresentaram aumento na proporção de uso junto a forma inovadora, saltando de 68,5% para 85,8% das realizações, demonstrando serem verbos receptivos ao pronome *a gente*. Essa tendência relacionada ao aumento na proporção de uso do pronome inovador pode ser constatada também entre os verbos que exprimem ideia de atividade mental, os verbos epistêmicos, sendo estes ainda mais favoráveis ao uso da variante inovadora, tendo a maior proporção entre os tipos de verbos pesquisados na década mais atual (89,6%).

O aumento na proporção de uso do pronome inovador ao longo de duas décadas sugere sua expansão na comunidade de fala pesquisada nestes contextos verbais, que na década passada já eram favoráveis ao uso da forma *a gente*. Porém, na década de 1990, alguns tipos de verbo destacavam o uso da forma padrão em nossa amostra, com maior proporção de uso para o pronome *nós*, como é o caso dos verbos de estado e do verbo *ter*. Na década de 2020, esse

comportamento se inverteu, tornando esses contextos fortemente favoráveis ao uso do pronome *a gente*, como demonstra o gráfico abaixo.

**Gráfico 27:** Uso de *nós* e *a gente* por verbos de *estado* e *verbo ter* em tempo real



Fonte: elaboração própria

As informações apresentadas no Gráfico 27, sugerem uma mudança no comportamento dos pronomes relacionada aos verbos do tipo *ter* e de *estado*. Diferentemente do que revelou a análise dos verbos de ação, *dicendi* e epistêmico, os verbos de estado e *ter* apontam para uma inversão na preferência no uso dos pronomes pelos falantes cultos de Fortaleza. Essa inversão é claramente visível na observação da linha estatística que representa o uso dos pronomes em dois tempos distintos, havendo uma interseção entre as linhas. Esse ponto de interseção sugere que, no decorrer de duas décadas, a disputa pela preferência dos pronomes se tornou mais acirrada até que, em determinado momento, o uso entre as formas representava o mesmo percentual (ponto de interseção entre as linhas). Em um momento posterior, a preferência pela variante inovadora se mostrou mais expressiva, apontando uma mudança na preferência entre os pronomes de primeira pessoa em duas décadas na fala culta fortalezense relacionada a esses verbos.

Nos anos 1990, o pronome padrão era usado preferencialmente para se referir a primeira pessoa do discurso nos contextos de verbos de estado, representando um total de 57,4%,

enquanto o pronome inovador era menos utilizado, com 42,6% das realizações neste contexto linguístico. Ao longo de duas décadas, no entanto, o pronome inovador se tornou predominante nas realizações associadas a verbos de estado, atingindo 73,5% na década de 2020, conforme demonstra o Gráfico 27. Por outro lado, a variante padrão seguiu direção oposta, reduzindo drasticamente a sua proporção de uso até chegar a 26,5% das ocorrências, indicando uma inversão na escolha entre os pronomes ao longo do tempo.

Nos contextos de verbos do tipo *ter*, é possível observar um padrão parecido aos verbos de estado. Esse tipo de verbo representava o contexto verbal mais perceptível ao uso do pronome *nós* na década de 1990, apontando uma expressiva preferência pelo uso do pronome canônico, com 70,3% das ocorrências, ao passo que o pronome inovador sofria restrição nesse contexto, representando apenas 29,7% dos dados de primeira pessoa do plural. Contudo, na década de 2020, o comportamento dos pronomes também se inverteu nestes contextos verbais, ou seja, o pronome padrão *nós* associado a esse verbo perdeu a preferência entre os falantes, tendo sua proporção de uso reduzida expressivamente (de 70,3% para 39,7%). Já o pronome *a gente* se expandiu em realizações com verbos do tipo *ter*, se tornando preferido, com um salto na proporção de uso de 39,7% para 60,3% dos dados.

Os dados em tempo real indicam que a forma pronominal *a gente* está em ascensão em todos os contextos verbais investigados, se tornando expressivamente mais evidente a cada década na comunidade de fala, enquanto o pronome canônico *nós* parece estar em declínio, sendo cada vez menos usado com o passar do tempo. A diminuição expressiva no uso do pronome *nós* aponta para uma mudança na preferência dos falantes em duas décadas, indicando uma tendência maior da comunidade a aderir a formas inovadoras. A ampliação no uso da variante inovadora indica que esta se implementou em todos os contextos linguísticos, o que pode expressar que o tipo de verbo já não é mais um fator determinante na escolha entre os pronomes de primeira pessoa do plural, uma vez que, em termos de proporção de uso, a análise demonstrou que todos os contextos verbais se tornaram mais receptíveis ao uso de *a gente*.

#### e) *Saliência fônica*

O princípio da saliência fônica observa a diferenciação no material fônico que apresenta as formas verbais na oposição entre singular e plural. Fernando e Gorski (1986) pontuam a relação direta entre as variáveis *saliência fônica* e *tempo verbal*, destacando que alguns tempos verbais, devido ao acréscimo da desinência modo temporal, tendem a se inclinar

ao uso do pronome *nós*, enquanto outros verbos podem ser favoráveis ao uso do pronome *a gente*. Estudos sociolinguísticos (Omena, 1986; Lopes 2003; Brustolin, 2009; Silva, 2013; Lima, 2020; entre outros) têm destacado que a forma *a gente* é usada preferencialmente com verbos que apresentam menor saliência fônica, enquanto o pronome *nós* está associado às formas verbais de maior saliência. Neste estudo em tempo real, a *saliência fônica* foi uma variável selecionada na análise multivariada das duas amostras, enquanto o *tempo verbal*, por exemplo, somente foi selecionado como estatisticamente significativo nos dados referentes à década de 1990, sendo excluído da análise com dados da amostra de 2020. Cabe pontuar que esta variável apresentou multicolinearidade na análise das duas décadas, uma vez que os dados de saliência nível 1 (cantando), que apresentam a mesma forma para singular e plural, correspondem exatamente aos verbos na forma nominal gerúndio, uma das categorias da variável *tempo verbal e formas nominais do verbal*. Sendo assim, para que a análise fosse possível, retiramos os dados referentes ao nível 1 de saliência fônica. O comparativo da atuação dessa variável nas duas amostras pode ser conferido na tabela 34:

**Tabela 34:** Distribuição de *nós* e *a gente* segundo o grau de *saliência fônica* em tempo real

Saliência fônica	Amostra 1990			Amostra 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
Nível 1 - (cantando)	82%	-1.21491		94,7%	2.2570	***
Nível 2 - (cantar/cantarmos)	68,5%	0.18340		91,5%	1.6258	***
Nível 3 - (falava/ falávamos)	69,9%	1.01549	***	87,5%	1.0584	***
Nível 4 - (fala/ falamos)	48,4%	-1.21027	***	72,4%	0.9466	***
Nível 5 - (tem/ temos)	44,5%	0.41901		65,7%	0.7572	**
Nível 6 - (cantou/cantamos)	38,4%	1.52171	***	69,4%	1.1090	***
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

Os dados da Tabela 34 revelam significativo aumento no uso do pronome *a gente* entre os falantes cultos fortalezenses, em todos os níveis da *saliência fônica*. Neste ponto, cabe lembrar que, em nossos modelos de regressão logística, o valor do intercepto desta variável se refere a ocorrências *a gente* junto a verbos com nível máximo de saliência, Nível 7 (é/somos), motivo pelo qual daremos luz aos resultados sempre os relacionando com os resultados deste nível. Em termos de estimativa, a atuação do nível 2 (cantar/cantarmos) de saliência (infinitivo com acréscimo da desinência *-mos*), em nossa amostra, aponta para o desfavorecimento no uso do pronome inovador (-1.21491) na década de 1990, e um expressivo favorecimento na década

de 2020 (2.2570). No entanto, o valor de  $p$  atribuído a este nível na Amostra 1990 está fora do intervalo de confiança, demonstrando que os verbos de menor saliência não apresentam atuação estatisticamente significativa sobre a realização dos pronomes nesta sincronia. Diferentemente da década de 2020, que se mostrou relevante para a análise ( $p < 0,005$ ).

Os verbos que conservam a sílaba tônica com acréscimo da desinência *-mos*, Nível 3 (falava/ falávamos), demonstram comportamento semelhante aos verbos de Nível 2, conforme indicam os dados da Tabela 34. Desta forma, a atuação deste nível não se mostrou significativa estatisticamente na Amostra 1990, apresentando valor de  $p$  fora do intervalo de confiança de 5%, mas apresentou relevância estatística na Amostra 2020 ( $p\text{-value} < 0,005$ ). Além disso, os resultados da análise da atuação desses verbos sobre a realização dos pronomes na década de 2020 indica forte favorecimento do pronome *a gente* quando associado a verbos com este nível de saliência. Em relação ao valor de referência da variável (*intercept*), os verbos com esse perfil apresentam estimativa de uso igual a 1.6258, demonstrando atuação de modo a favorecer a realização de forma expressiva do pronome *a gente* na década mais atual.

Já nos verbos que, na relação singular/plural, apresentam deslocamento do acentoônico associado ao acréscimo da desinência *-mos* (Nível 4 – fala/falamos), os resultados da Tabela 34 apontam para favorecimento da forma pronominal *a gente* em ambas as amostras. O nível 4 de saliência fônica se destacou como estatisticamente significativo para a realização dos pronomes nas análises das duas décadas. Estes verbos representam, em nossa amostra, a maior proporção de *nós* e *a gente* entre os níveis realizados (ver tabelas 12 e 25), motivo que poderia explicar a relevância estatística destes verbos em ambas as análises. Além disso, são nesses verbos que se concentram as formas verbais com ambiguidade temporal, ou seja, que apresentam a mesma forma para a primeira pessoa do plural na relação presente/passado (falamos, cantamos, sorrimos). Nesse sentido, Yolenco, Schere e Naro (2018), acreditam que a inserção do pronome *a gente* no quadro pronominal seria responsável por desfazer essa ambiguidade, motivo pelo qual esses níveis podem atuar tão fortemente na realização dos pronomes nas duas sincronias, no sentido de favorecer o pronome inovador. Ademais, análise da atuação desse nível nas duas amostras apontaram valor de  $p < 0,005$ , o que indica forte interação com a realização da variável resposta.

Por outro lado, os monossílabos tônicos ou átonos que passam a ser paroxítona no plural, Nível 5 (tem/temos), também se mostraram relevantes na realização dos pronomes em nossas amostras, com algumas diferenças na atuação em cada década. Nos anos de 1990, por exemplo, os resultados da Tabela 34 destacam desfavorecimento do pronome *a gente* que, em relação ao *intercept* do modelo, apresentou estimativa igual a -1.21491, sendo este um contexto

de maior resistência à entrada do pronome inovador. Após mais de vinte anos, a atuação desses verbos na comunidade de fala se inverteu. Ou seja, os verbos com saliência nível 5, que antes eram desfavoráveis ao uso de *a gente*, agora passam a favorecer fortemente seu uso, com estimativa de 0.9466 em relação ao valor de *intercept*. Esses resultados apontam para a significativa expansão do pronome *a gente* em diversos contextos sociais e linguísticos.

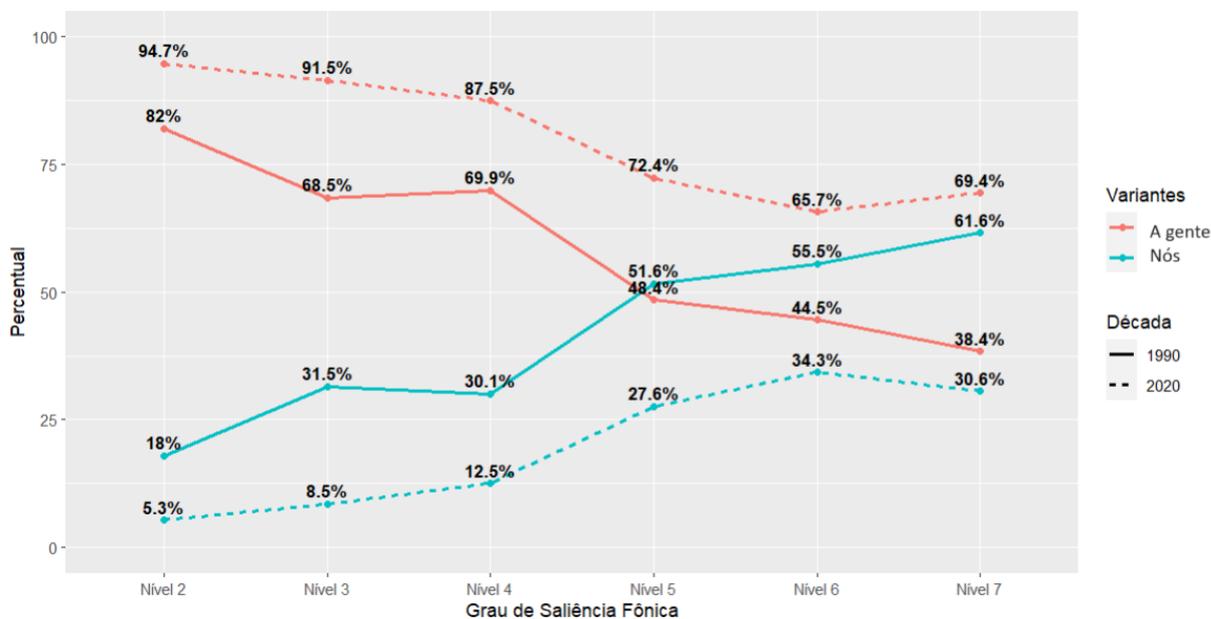
A saliência fônica Nível 6 (cantou/cantamos) corresponde a realizações dos pronomes com verbos que, na relação singular e plural, reduzem os ditongos finais em vogais, com acréscimo da desinência *-mos*. Esses verbos também apresentam ambiguidade temporal, tendo em vista que, em sua maioria, apresentam a mesma forma para o passado e para o presente. No entanto, diferentemente dos verbos que apresentam essa mesma característica (nível 4), os verbos de nível 6 não apresentaram atuação significativamente relevante na década de 1990, estando o valor de *p* atribuído a esse nível acima do intervalo de confiança. Por outro lado, a análise da Amostra 2020 revela atuação estatisticamente significativa desse nível de saliência, com *p-value* < 0,005. Além disso, a estimativa deste nível aponta para o favorecimento da forma inovadora *a gente* que, em relação ao intercepto do modelo, apresenta valor igual a 0.7572.

Por fim, como dito, o nível 7 (é/somos) corresponde ao valor de referência da variável. Isso significa que, nos mesmos contextos de realização do *intercept*, os verbos com diferenças fônicas muito acentuadas (nível 7), favorecem o pronome *a gente* em ambas as Amostras. O resultado em tempo real referente a atuação da variável *saliência fônica* indica implementação acelerada da forma pronominal inovadora na década de 2020. Em todos os níveis, o pronome *a gente* adentrou e se expandiu. Lopes (2003), ao investigar a atuação dos pronomes em função do nível de saliência fônica em tempo real, observou a expansão da forma *a gente* em todos os níveis de saliência na amostra mais atual de sua análise. Para a autora, essa implementação da forma inovadora em diversos contextos “talvez tenha ocorrido pelo fato de *a gente* estar se tornando cada vez mais usual, e os espaços funcionais se tornam cada vez mais restritos” (Lopes, 2003, p. 134)

Em termos percentuais, a tendência no que concerne ao uso dos pronomes também é observada: expressiva expansão no uso da forma pronominal *a gente* e retração no uso do pronome *nós*. No entanto, percebe-se dois padrões relacionados ao nível de saliência fônica em nossa amostra. O primeiro deles revela que os níveis mais baixos de saliência fônica (níveis 1, 2 e 3) já beneficiaram a forma *a gente* na década de 1990 e passaram a beneficiar ainda mais na década 2020. O segundo demonstra que os verbos de maior saliência fônica (níveis 4, 5 e 6), que inibiam o uso do pronome *a gente* na década de 1990, passaram a ser favoráveis ao uso da

forma inovadora na década de 2020. A tendência no uso das formas pronominais em função do grau de saliência fônica, pode ser conferida no gráfico 28:

**Gráfico 28:** Tendência de uso de *nós* e *a gente* por saliência fônica em tempo real



Fonte: Elaboração própria.

Como exposto no Gráfico 28, em contextos de menor saliência do verbo, há maior proporção de uso do pronome inovador *a gente* nas duas décadas, representando um aumento no percentual de ocorrências ao longo do tempo. A possível substituição de *nós* por *a gente* segue uma direção, representada pelas linhas estatísticas no gráfico, ou seja, o aumento ou diminuição no uso dos pronomes parecem seguir tendências parecidas em todos os níveis de saliência fônica. Nesse sentido, percebemos que a proporção de uso da variante inovadora aumenta com o passar do tempo, ao passo que a proporção da variante padrão diminui, com diferença de, pelo menos, dez pontos percentuais entre 1990 e 2020. Esse padrão semelhante nas taxas de uso de *nós* e *a gente* pode ser interpretada como um elemento que sustenta a hipótese da taxa constante de Kroch (1989), que prescreve que as mudanças linguísticas numa comunidade seguem uma taxa constante e uniforme. Por outro lado, os resultados podem sugerir uma mudança na distribuição dos pronomes em tempo real, mas não se caracteriza como mudança drástica no uso dos pronomes na comunidade de fala, tendo em mente que o pronome inovador já era favorito entre os informantes na década passada nos níveis mais baixos de saliência (níveis 2, 3 e 4).

Os verbos com nível 2 (cantar/cantaremos) são os verbos mais favoráveis ao uso do pronome *a gente*, com proporção de uso de 82% na década de 1990, saltando para 94,7% na década de 2020, representando uso próximo do categórico nos termos de Labov (2003). Em relação ao pronome canônico, os verbos de nível 2 representam contexto de inibição, sofrendo retração no uso, saindo de 18% para 5,3%, uma diminuição expressiva. Esses resultados estão em conformidade com as hipóteses levantadas sobre a atuação dessa variável, pois esperávamos maior uso do pronome *a gente* associado a verbos de baixa saliência fônica.

Os verbos de saliência fônica nível 3 (falava/falávamos) apresentaram comportamento parecido aos de nível 2. A inclinação da linha estatística que representa a atuação dos pronomes com estes níveis sugere que os verbos do nível 3 foram os que mais se tornaram receptivos ao uso da forma inovadora, com uma diferença que corresponde a 23 pontos percentuais. A análise da Amostra 1990 revela proporção de uso do pronome *a gente* igual a 68,5% dos dados, expandindo seu uso na Amostra 2020 para 91,5% do total de observações em nossos dados. Por outro lado, o uso do pronome padrão caiu consideravelmente entre as duas décadas, saindo de 31,5% na década de 1990 para 8,5% das ocorrências na década de 2020. Esse comportamento, vale ressaltar, manifesta a tendência geral da amostra, com acelerada implementação da variante inovadora nos mais diversos contextos.

Já o nível 4 (fala/falamos) de saliência fônica atua de forma parecida aos níveis anteriores, ou seja, percebe-se aumento significativo no uso do pronome *a gente* entre as décadas. Esse aumento na proporção da variante inovadora é proporcional a diminuição no emprego do pronome *nós* nas duas sincronias. Nos anos 90, o uso de *a gente* junto a verbos com saliência nível 4 representava 69,9% das ocorrências neste contexto, saltando para 87,5% nos anos 2020, conforme demonstra o Gráfico 28, uma evolução que corresponde a 17,6 pontos percentuais. Em contrapartida, nota-se a expressiva retração no uso do pronome canônico nestes contextos verbais, de 30,1% das realizações na Amostra 1990, para somente 12,5% na Amostra 2020. Essas primeiras categorias seguem o mesmo padrão no comportamento dos pronomes, com franca expansão de *a gente* e expressiva retração de *nós*.

Outros níveis dessa variável apresentam um padrão de uso dos pronomes diferente. Os verbos de maior saliência, como dito, tendem a se inclinar ao uso do pronome padrão na década de 1990, com maior proporção de uso para o pronome *nós*. Com o passar do tempo, essa tendência se inverte e o pronome *a gente* passa a ser preferido pelos falantes cultos de Fortaleza, inclusive nos níveis mais altos de saliência fônica.

Essa mudança na preferência dos informantes pelo pronome de primeira pessoa do plural, sendo o pronome *nós* preferido na década de 1990, e o pronome *a gente* preferido na

década de 2020, segue a tendência dos dados. A inversão no comportamento dos pronomes pode ser atestada ao observar a linha estatística que representa a proporção de uso dos pronomes em tempo real (ver Gráfico 28), especialmente pelo ponto de interseção entre as linhas na década de 1990. É exatamente no ponto em que as linhas se tocam que se inicia a mudança na tendência referente ao uso dos pronomes. Na década de 1990, por exemplo, os verbos mais marcados representavam contextos de resistência ao pronome inovador, sendo estes ambientes mais favoráveis ao uso de *nós*. Ao se implementar na língua, o pronome *a gente* buscou os ambientes mais receptíveis a sua inserção, como afirma Omena (2003) em pesquisa sobre *nós e a gente* em tempo real:

As formas mais marcadas, uma vez adquiridas, tendem a uma maior estabilidade que as não marcadas, possivelmente devido a fatores de memória. O pronome de primeira pessoa do plural *nós* (ou *a gente*) pode combinar-se com o núcleo verbal que tem a forma mais ou menos saliente. A nova forma se insere no ponto mais frágil, o menos marcado. Assim, há maior probabilidade de ocorrência de *a gente*. (Omena, 2003, p. 69).

Os nossos dados estão em consonância com a proposição da autora, tendo em vista que, quanto mais marcada for a diferença entre singular e plural, maior a probabilidade no emprego do pronome *nós*. Isso se confirma na análise da Amostra 1990, em que o pronome padrão apresentou maior proporção de uso nos três maiores níveis de saliência fônica, de forma menos intensa no nível 5 (tem/temos), estando as proporções de uso próximas ao ponto neutro, e mais intensa no nível 7 (é/somos), sendo este o que representa a maior proporção no uso do pronome canônico. Estes resultados se assemelham aos encontrados por Omena (2003, p.70), ao identificar que “o contexto para a referência à primeira pessoa do plural em que há menos diferença entre a terceira pessoa do singular e a primeira pessoa do plural favorece o uso de *a gente*. O inverso favorece *nós*”.

Esse cenário se inverte na análise dos dados representativos da década de 2020, com expressiva expansão do pronome inovador até em contextos de maior diferenciação fônica. Em relação aos dados de saliência nível 5 (tem/temos), conforme é possível observar no gráfico 28, houve um salto na proporção de uso do pronome *a gente*, de 48,4% para 72,4% em pouco mais de vinte anos, enquanto o pronome *nós* passou a ser menos recorrente, caindo de 51,6% para 27,6%, uma diferença que representa 24 pontos percentuais no uso dos pronomes. De forma parecida atuam as ocorrências referentes ao nível 6 de saliência (cantou/cantamos), em que houve aumento em tempo real na proporção de uso do pronome inovador (de 44,5% para 65,7%), e retração no uso do pronome padrão (de 55,5% para 34,3%).

Por fim, o nível 7 de saliência (é/somos) apresentou a mudança mais expressiva no comportamento dos pronomes pessoais. Verbos com diferenças acentuadas no material fônico entre as formas singular e plural (ex. é/somos, veio/viemos) eram significativamente favoráveis a variante padrão nos anos 1990, com percentual de uso de 61,6%, e inibidores da variante inovadora, com 38,4% das realizações neste contexto. A análise dos anos 2020 revela mudança expressiva no comportamento do fenômeno entre os falantes cultos fortalezenses, com aumento da forma *a gente* de 38,4% para 69,4% no nível 7 de saliência, enquanto o pronome padrão apresentou diminuição na proporção, de 61,6% para 30,6%. Cabe destacar que esses verbos representam a mudança mais perceptível no comportamento dos pronomes dentre os níveis de saliência fônica analisados, com uma diferença de 31 pontos percentuais.

Lopes (2003) analisou o uso dos pronomes *nós* e *a gente* com dados de fala coletados em 1970 e comparados com dados de 1990. As amostras da década mais recente incluíram os mesmos informantes contatados na década de 70 e também uma nova coleta realizada na comunidade de fala. O resultado encontrado pela autora se assemelha ao nosso, com destaque na amplificação da forma inovadora na década mais recente, principalmente entre os primeiros níveis da variável saliência fônica:

Na comparação das amostras, vê-se que, se for analisado o comportamento dos mesmos indivíduos (década de 70 e recontato), as freqüências/pesos relativos apresentam valores aproximados (indivíduo mantém o seu comportamento), com um ligeiro acréscimo, em todos os níveis, na amostra recontato. Interessante observar que os valores tornam-se ainda maiores na amostra com indivíduos diferentes (nova amostra 90) com freqüências ainda mais altas nos três primeiros níveis. (Lopes, 2003, p. 132).

Esses resultados demonstram que a forma pronominal *a gente* tem ganhado espaço entre os falantes cultos fortalezenses na preferência a referência à primeira pessoa do plural, especialmente em contextos que representavam maior restrição à forma inovadora, como nos níveis mais altos de saliência fônica. Confrontando os resultados da análise das duas amostras, no estudo em tendência (Labov, 1994), nota-se indícios de substituição do pronome *nós* por *a gente*, estando perfeitamente implementado no português falado na comunidade de fala analisada, e amplamente adotado pelos informantes como referência à primeira pessoa do discurso.

Por fim, é importante salientar as variáveis que deixaram de ser significativas na análise em tempo real de curta duração. A variável *sexo/gênero* não se mostrou relevante em nenhuma das análises. Por outro lado, a variável *tempo verbal* apresentou relevância estatística na década de 1990, enquanto na década de 2020, a seleção automática recomendou a sua

remoção, por não apresentar interação estatisticamente significativa com a realização dos pronomes. Nesse ponto, ao considerar a proporção geral de *a gente* na Amostra 1990 (58,3%), nota-se que a forma já havia se implementado e já era preferida entre os informantes. Na Amostra 2020, a proporção geral de uso saltou para 79,95%, destacando ainda mais a preferência pela variante inovadora. Dada a preferência pela forma *a gente*, intensificada na sincronia mais atual, é normal que alguns fatores deixem de exercer efeito sobre o uso das variantes, pois a forma inovadora está consolidada na fala nos diversos contextos. Sobre esta questão, Weinreich et al. (2006, pág. 123) afirmam que “no desenvolvimento da mudança linguística, encontramos estruturas linguísticas encaixadas desigualmente na estrutura social; e nos estágios iniciais e finais de uma mudança, pode haver pouca correlação com fatores sociais”. Nesse caso, estando o pronome *a gente* tão amplamente implementado na comunidade de fala, é normal que alguns fatores deixem de exercer influência significativa sobre a escolha entre os pronomes de primeira pessoa do plural.

### **5.3.1 Síntese dos resultados – A alternância pronominal em tempo real de curta duração**

A análise em tempo real de curta duração contabilizou 4575 observações de *a gente* e *nós* na fala culta de Fortaleza, com 1957 ocorrências na amostra de 1990, e 2618 ocorrências na de 2020, sendo o pronome *a gente* é mais produtivo que *nós* em ambas as décadas. Na amostra de 1990, foram identificados 58,25% de *a gente* (1140 obs.) e 41,75% do pronome *nós* (817 obs.). No que se refere à amostra de 2020, foram observadas 79,95% das ocorrências (2093 obs.) de *a gente*, um aumento significativo na proporção de uso se comparada à década de 1990, e apenas 20,05% de *nós* (525 obs.), demonstrando que o pronome *a gente* se destaca entre os informantes cultos como referência à primeira pessoa do plural.

Houve aumento significativo no percentual de uso da variante em todas as faixas etárias no intervalo de duas décadas. Entre os mais jovens da amostra (informantes de 22 a 35 anos), a diferença no uso de uma década para outra representa um aumento de 24,9 pontos percentuais, indicando uma tendência crescente no uso dessa variante entre os jovens. A mesma tendência é observada entre os informantes de 36 a 50 anos, que, assim como os mais jovens, apresentaram evolução na proporção de uso de uma década para outra. A diferença no percentual entre as sincronias é de 20 pontos percentuais. Em relação aos informantes mais velhos da amostra (a partir de 51 anos), houve igualmente progressão do pronome *a gente* entre

dois pontos do tempo. A comparação em tempo real entre as duas sincronias revela uma diferença no uso que chega a 31,7 pontos percentuais.

Em relação ao *tipo de inquérito*, houve expansão significativa de *a gente* em todos os tipos de registros. Os inquéritos do tipo D2 saltaram de 72,3% na Amostra 1990 para 91,4%, índice próximo do categórico, na Amostra 2020. Comportamento parecido se observa ao analisar os registros do tipo DID, em que houve crescimento significativo nos percentuais de uso da forma inovadora, indicando ampla expansão da forma *a gente* em contextos de formalidade moderada, saltando de 60,8% para 79,1%, uma diferença de 18,3 pontos percentuais. Já nos inquéritos mais formais da amostra (EF), houve inversão na preferência dos informantes pelas variantes, elevando significativamente a proporção de uso do pronome *a gente* de 40,2% para 66,4% e, com a mesma intensidade, houve retração do pronome padrão em neste tipo de registro.

Sobre a variável *grau de referencialidade do pronome*, na década de 1990, era notória a associação do pronome *nós* a contextos de especificidade, enquanto o pronome *a gente* estava claramente associado a realizações em contextos de indeterminação. Esse comportamento sofreu alteração na análise em tempo real, ou seja, em contextos de sentido específico, na década de 2020, houve significativa redução do pronome *nós* (61,9% para 23,8%), e aumento expressivo no uso do pronome *a gente* (de 38,1% para 76,2%).

Por fim, os resultados da *saliência fônica* demonstram que, na análise da Amostra 1990, o pronome padrão apresentou maior proporção de uso nos três maiores níveis de saliência fônica, de forma menos intensa no nível 5 (tem/temos), estando as proporções de uso próximas ao ponto neutro, e mais intensa no nível 7 (é/somos), sendo este o que representa a maior proporção no uso do pronome canônico. Esse cenário se inverte na análise dos dados representativos da década de 2020, com expressiva expansão do pronome inovador até em contextos de maior diferenciação fônica. Em reação aos dados de saliência nível 5 (tem/temos), houve um salto na proporção de uso do pronome *a gente*, de 48,4% para 72,4% em mais de duas décadas, enquanto o pronome *nós* passou a ser menos recorrente, caindo de 51,6% para 27,6%, uma diferença que representa 24 pontos percentuais no uso dos pronomes.

## 6 DESCRIÇÃO DOS DADOS E ANÁLISE DOS RESULTADOS: O FENÔMENO DA CONCORDÂNCIA VERBAL

O uso variável da *alternância pronominal* e da *concordância verbal de primeira pessoa do plural* são fenômenos variáveis que se correlacionam, tendo em vista que a escolha entre o uso dos pronomes *nós* ou *a gente* influencia diretamente a marcação do verbo subsequente, com presença ou ausência da desinência de primeira pessoa do plural *-mos* (Scherre, Yacovenco e Naro, 2018). Em relação a realização da concordância verbal de primeira pessoa do plural, entendemos que a língua é constantemente avaliada socialmente, sendo as variantes que constituem o fenômeno variável da concordância verbal passíveis de estigma ou de prestígio (Freitas, Favacho e Carvalho, 2021). De acordo com as autoras, a concordância emergente *a gente* sem *-mos* é avaliada pelos informantes escolarizados como informal e periférico, enquanto a forma *nós* com *-mos* goza de prestígio. Questões de percepção e avaliação linguística são importantes, tendo em vista que a avaliação negativa feita pelo informante sobre determinada variante pode restringir seu uso, ao passo que a avaliação positiva pode motivá-lo (Moreno Fernández, 1998).

A norma gramatical difundida pela escola e presente nas instituições de ensino de modo geral classificam a concordância verbal como categórica, sendo prescrita como regra válida apenas a concordância padrão do pronome de primeira pessoa do plural *nós*, com verbos acompanhados da desinência *-mos*, como em “*nós* comemos”. (Brustolin, 2009). No entanto, trabalhos variacionistas apontam o fenômeno da concordância verbal como regra variável, uma vez que existe mais de uma forma para expressar a concordância de primeira pessoa do plural, e essas formas competem pelo espaço na comunidade de fala, como demonstram Fernandes (1999), Rubio (2012), Foeger (2014), Benfica (2016), Vieira (2019), para citar alguns.

Na fala culta fortalezense, com dados provenientes do Projeto Português Oral Culto de Fortaleza nas décadas de 1990 e de 2020, para efeito de análise e descrição linguística do fenômeno, identificamos quatro padrões de concordância verbal com os pronomes *nós* e *a gente*:

***Nós* com *-mos*** (concordância padrão):

(159) na sociedade capitalista... ***nós vimos*** recentemente... todo dia ***nós vemos*** esse desperdício... o BraSIL é o país do desperdício... mas nos países comunistas... eles começaram... a CONtrolar... o abastecimento... (Inq. 17 – F1 – EF, homem, 33 anos)

- **Nós sem -mos** (concordância não padrão):

(160) Inf.: e assim era feito no ano que tivesse essas festas **nós tava** lá né SEMpre tinha um amigo de um conhecido do conhecido do papagai do conhecido que arranjava um convite pra gente entrar (Inq. 13 – F1 – DID, mulher, 41 anos)

- **A gente sem -mos** (padrão emergente):

(161) Inf.: a grande parte da população ela valoriza muito o rádio... **a gente pensa** que não mais o... esse tipo de comentário que o locutor FAZ esse animador desses programa faz é:: repercute (Inq. 15 – F1 – DID, homem, 58 anos)

- **A gente com -mos** (concordância não padrão):

(162) Inf.: aí tinha um local lá que **a gente vamos conhecer** aqui... é um local parecia assim tipo a costa mendes daqui... cheio de DOce doce... cada coisa/ pra gente degu/ comer né... comprar... era assim pra conhecer as coisas... muito bem lá... eu gostei tambémm..... (Inq. 62 – F2 – DID, mulher, 63 anos)

Nosso objetivo, nesta análise, é observar os contextos sociais e linguísticos mais favoráveis ao emprego da concordância padrão e não padrão. Para isso, observamos as ocorrências das variantes em função das variáveis preditoras de ordem extralinguística (*sexo/gênero, faixa etária e tipo de inquérito*), e das de natureza linguística (*tempo verbal e tipo de paradigma verbal, tonicidade do verbo, conjugação verbal e estrutura do verbo*).

Muitos trabalhos realizados acerca dos padrões de concordância verbal são desenvolvidos no Brasil à luz da Teoria da Variação e Mudança Linguística, tendo como foco análises multivariadas. O levantamento da literatura apontou tendências distintas de análise da concordância variável com os pronomes *nós* e *a gente* nos trabalhos consultados: a) análises binárias que observam a presença ou a ausência da desinência *-mos* com o pronome padrão *nós* (Fernandes, 1999; Zilles, Maya, Silva, 2000; Zilles, 2005; Rodrigues, 2007; Silva, 2009; Foeger, 2014; Oushiro, 2015); b) o uso do pronome *a gente* com presença *versus* ausência da desinência verbal *-mos* (Coelho, 2006; Rubio, 2012; Mattos, 2013); e c) análise ternária envolvendo as variantes *nós* com *-mos*, *nós* sem *-mos*, *a gente* sem *-mos* (Scherre e Naro, 2016; Scherre, Yacovenco e Naro, 2018), e d) análise binária da estrutura de concordância padrão *nós* com *-mos* *versus* a estrutura emergente *a gente* sem *-mos* (Carvalho, Freitas e Favacho, 2020; Carvalho e Santos, 2023). Em ambos os trabalhos, o enfoque é dado na marcação ou a ausência da desinência plural *-mos* junto aos pronomes *nós* e *a gente*.

Nesta análise, observamos as ocorrências de cada uma das quatro variantes que compõem o fenômeno variável da concordância verbal, inspirados nos trabalhos de Scherre, Yacovenco e Naro (2018), com as devidas adaptações, que conduziram estudo sobre a variação na concordância verbal na Baixada Cuiabana e em Vitória, e de Carvalho, Freitas e Favacho (2020), que investigaram os padrões de concordância na fala de Fortaleza. Em um segundo momento, prevendo a distribuição desbalanceada das ocorrências, uma vez que trabalhos variacionistas constataam poucas ocorrências da concordância não padrão (*nós* sem *-mos* e *a gente* com *-mos*), procedemos análise multivariada através de um modelo de regressão logística que analisa, em um recorte binário tradicional, em termos de favorecimento, a atuação da concordância padrão (*nós* com *-mos*) vs. padrão emergente na língua portuguesa (*a gente* sem *-mos*).

Nosso objetivo, nesta análise, é testar as tendências inicialmente observadas por Naro, Göski e Fernandes (1999), posteriormente desenvolvidas por Naro, Scherre, Foeger e Benfica (2014); Naro, Scherre, Foeger e Benfica (2017); e Scherre, Yacovenco e Naro (2018). Para os autores, a inserção do pronome *a gente* atuaria na questão da ambiguidade potencial existente em alguns verbos, que apresentam a mesma forma para o pretérito perfeito (*nós* cantamos) e para o presente do indicativo (*nós* cantamos), sendo o pronome *a gente* sem a desinência *-mos* usado essencialmente para verbos no presente com ambiguidade temporal, e o padrão *nós* empregado com uso preferencialmente junto a desinência *-mos* “para pretérito e presente de formas diferentes (*nós* tivemos/*nós* temos; *nós* fomos/*nós* vamos ou *nós* somos), sem possibilidade de ambiguidade, majoritariamente de oposição singular/plural mais saliente (*nós* foi/*nós* fomos; *nós* tem/*nós* temos; *nós* vai/*nós* vamos; *nós* é/*nós* somos)” (Scherre, Yacovenco e Naro, 2018, p.14). Para tanto, replicamos o grupo de fatores *tempo e paradigma verbal*, com as devidas adaptações, já que a replicabilidade de pesquisas constitui um dos princípios centrais da Sociolinguística de cunho variacionista.

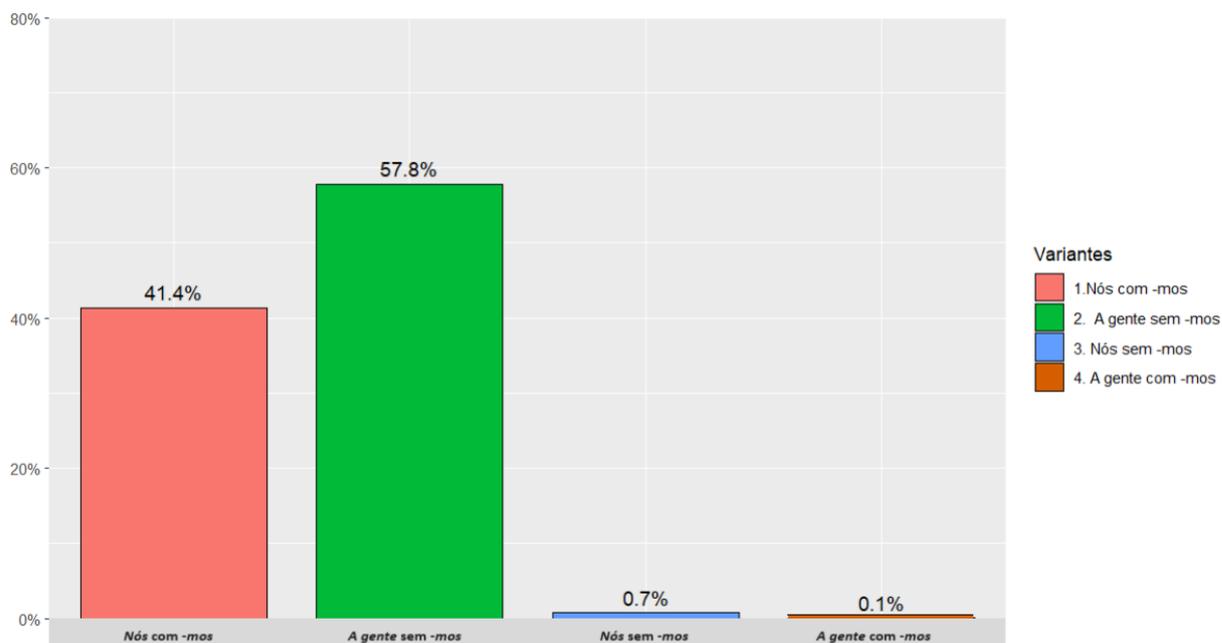
Primeiramente, apresentamos a descrição dos dados de concordância verbal referentes à década de 1990, em termos de proporção e frequência de uso, bem como a análise do modelo de regressão logística. Em um segundo momento, descrevemos os dados referentes à década de 2020, observando as mesmas variáveis predictoras. Por fim, procedemos à comparação entre as duas décadas, com o objetivo de analisar as tendências no uso dos padrões de concordância em duas décadas.

## 6.1 A concordância verbal de primeira pessoa do plural no falar culto de Fortaleza: análise descritiva dos dados de fala da década de 1990

Os dados de concordância verbal de primeira pessoa do plural na fala culta fortalezense referentes a amostra advinda do projeto PORCUFORT Fase I, que comporta a fala representativa da década de 1990, foram coletados por meio da ferramenta computacional *RStudio*, através do pacote *dmsocio* (Oushiro, 2018). Como explanado na subseção referente a descrição das variáveis (ver subseção 4.4.1), o fenômeno da concordância verbal com os pronomes *nós* e *a gente* comporta quatro variantes, que foram analisados à luz da teoria variacionista:

- a. *Nós* com *-mos*: concordância padrão, construção sem estigma (Bechara, 1999).
- b. *Nós* sem *-mos*: concordância não padrão, construção com estigma (Peluco, 2022).
- c. *A gente* com *-mos*: concordância não padrão, com estigma pela tradição gramatical (Peluco, 2022).
- d. *A gente* sem *-mos*: padrão de concordância emergente, sem estigma (Bechara, 1999).

Todas as ocorrências dessas variantes foram analisadas em nossa Amostra 1990. Após levantamento quantitativo realizado por meio da linguagem de programação *R*, através do ambiente *RStudio*, identificamos um total de 1938 observações de concordância com os pronomes de primeira pessoa do plural. Do total de ocorrências observadas, 802 dados correspondem à realização de concordância padrão *nós* com *-mos* (*nós cantamos*), 14 dados se referem a ocorrências de *nós* sem *-mos* (*nós canta*), 1121 observações são da estrutura *a gente* sem *-mos* (*a gente canta*), a maioria dos dados em nossa amostra, e apenas 1 ocorrência de *a gente* com *-mos* (*a gente cantamos*). O gráfico abaixo ilustra a proporção de cada uma das variantes da concordância verbal em nossa amostra representativa da década de 1990:

**Gráfico 29:** Proporção de uso da concordância verbal na Amostra 1990 (N=1957)

Fonte: elaboração própria

O resultado geral do levantamento dos padrões de concordância verbal em amostra do PORCUFORT da década de 1990 revela expressiva preferência dos informantes pelo padrão emergente da língua *a gente sem -mos*, totalizando 57,8% das realizações, sendo a variante mais destacada entre os falantes cultos de Fortaleza. Em seguida, o uso do pronome canônico associado a verbos com desinência *-mos*, dados de concordância padrão, representa a segunda variante mais empregada entre os informantes de nossa amostra, sendo responsável por 41,4% das ocorrências observadas em nossa base de dados representativa dos anos 1990. Os dados de concordância verbal não padrão são pouco recorrentes na amostra em análise, sendo a variante *nós sem -mos* observada em apenas 0,7% dos casos, e a estrutura *a gente com -mos* responsável por apenas 0,1% do total de ocorrências, fato que confirma nossa hipótese inicial, pois esperávamos que as observações dos padrões de concordância verbal com os pronomes *nós* e *a gente* se concentrariam nas realizações de concordância padrão e do padrão emergente.

Esse resultado demonstra que, na fala culta de Fortaleza, a disputa variável pelo emprego da concordância de primeira pessoa do plural está concentrada entre as formas *a gente sem -mos* e *nós com -mos*, sendo os dados de concordância não padrão pouco frequentes diante do total de dados (15 ocorrências de um total de 1938). Essa tendência relacionada a maior proporção de uso da concordância padrão também foi observada em estudos sociolinguísticos conduzidos em diversas regiões do país, tanto na fala quanto na escrita. O Quadro 12 apresenta

o comparativo de nossos resultados com os de pesquisas sociolinguísticas realizadas em diversas comunidades de fala no Brasil:

**Quadro 12:** Comparação do resultado de nosso estudo com pesquisas sociolinguísticas

ESTUDO	<i>NÓS COM -MOS</i>	<i>NÓS SEM -MOS</i>	<i>A GENTE SEM -MOS</i>	<i>A GENTE COM -MOS</i>
<b>Nossa pesquisa Fala - Fortaleza</b>	<b>41,4%</b>	<b>0,7%</b>	<b>57,8%</b>	<b>0,1%</b>
Brustolin (2009) Escrita - Florianópolis/SC	84%	1%	11%	2%
Brustolin (2009) Fala - Florianópolis/SC	33,96%	0,78%	58,24%	7,05%
Mendonça (2010) Fala - Vitória/ES	28,6%	1%	69,5%	0,9%
Rubio (2012) São José do Rio Preto/SP	22,5%	3,8%	69,2%	4,5%
Mattos (2013) Fala goiana	17,8%	25,3%	76,54%	0,27%
Carvalho, Freitas e Favacho (2020) Fala - Fortaleza/CE	37,8%	0,1%	62%	0,1%
Vitorio (2015) Escrita - Maceió/AL	77%	8%	4%	9%

Fonte: elaboração própria

Como é possível observar nos dados apresentados no Quadro 12, os resultados obtidos em nossa pesquisa sobre os padrões de concordância verbal na fala culta fortalezense se assemelham aos de muitas pesquisas desenvolvidas com *corpus* de língua falada, tendo em vista que, em todos os trabalhos analisados com dados de fala, há preferência significativa pelo uso do padrão emergente formado pelo pronome *a gente* associado a verbos na terceira pessoa do singular, sem a marcação da desinência *-mos* (*a gente* sem *-mos*). Em seguida, disputando a preferência dos falantes, temos o uso do pronome *nós* com concordância de primeira pessoa do plural, ou seja, concordância verbal padrão (*nós* com *-mos*). Em todas as pesquisas que analisam dados de língua falada, assim como a nossa, se observam poucas ocorrências de concordância verbal não padrão, com proporções de uso próximas a zero.

A título de comparação, Brustolin (2009) conduziu investigação sobre o uso dos pronomes de primeira pessoa do plural na fala de alunos do ensino fundamental de Florianópolis e verificou resultados próximos aos obtidos neste estudo, a saber, alto percentual de uso de *a gente* sem *-mos* (58,24%), seguido do padrão *nós* sem *-mos*, com 33,96% do total. Os dados observados sem concordância, assim como em nosso estudo, se mostraram pouco recorrentes

na fala, sendo a variante *a gente* com *-mos* empregada em 7,05% dos dados, e *nós* sem *-mos* usado em somente 0,78% das ocorrências, conforme indica as informações do quadro 12.

Essa mesma tendência foi observada na fala de Vitória/ ES, quando Mendonça (2010) analisou os pronomes *nós* e *a gente* e as respectivas concordâncias padrão e não padrão em dados de fala advindos do Projeto O Português Falado na Cidade de Vitória (PortVix). Após levantamento dos dados, o autor constatou maior preferência pelo uso de *a gente* sem *-mos* (69,5%), seguido por *nós* com *-mos* (28,6%), *nós* sem *-mos* (1%) e, por fim, *a gente* com *-mos* (0,9%). De maneira parecida, Rubio (2012), em investigação sobre a alternância e concordância verbal com os pronomes *nós* e *a gente* a partir de amostra extraída do Banco de Dados Iboruna, composto pelo Projeto ALIP (Amostra Linguística do Interior Paulista), identificou uso mais recorrente de *a gente* com verbo na terceira pessoa do singular (69,2%) e de *nós* com concordância padrão *-mos* (22,5%). Assim como demonstra a tendência geral no comportamento do fenômeno, dados de concordância não padrão são escassos, tendo a forma *a gente* com *-mos* ocorrido em apenas 4,5% dos dados, enquanto o pronome *nós* sem marca de plural *-mos* se realizou em apenas 3,8% dos casos.

Em relação à fala goiana, a pesquisa realizada por Mattos (2013) revelou alta proporção de uso do padrão emergente de concordância *a gente* sem *-mos*, com 76,54% dos dados. Diferentemente das outras pesquisas, o segundo padrão de concordância mais utilizado na amostra analisada pela autora trata-se do pronome *nós* sem desinência de plural (*nós* sem *-mos*), que obteve um total de 25,3% das realizações do fenômeno em análise. De acordo com a autora, a alta proporção de uso do pronome *nós* sem *-mos* (*nós* canta) reflete uma tendência atípica, uma vez que dados de concordância não padrão são pouco frequentes em diversas regiões do país, e pode estar associada à origem caipira da fala goiana:

A grande diferença da fala goiana relativamente à IPP é o uso do singular verbal com *nós*, que remete às raízes rurais da cultura e que os goianos praticam sem estigmatização. Essa identidade cultural e linguística de base rural estaria sendo utilizada na fala dos mais jovens, na contramão do crescente efeito da escolarização para o aumento do nível de concordância verbal (Mattos, 2013, p. 123)

Em relação ao uso do pronome *nós* com desinência plural *-mos* (concordância padrão), a linguista identificou uso menos recorrente que *nós* com *-mos*, com 17,8% das ocorrências na amostra. O uso do pronome inovador *a gente* junto a verbos com marcação da desinência *-mos* (*a gente* cantamos) se mostrou pouco recorrente, tendo apenas 5 ocorrências, o que corresponde a 0,27% do total de realizações.

Ainda sobre a língua falada, Carvalho, Freitas e Favacho (2020), assim como nós, investigaram o uso dos pronomes *nós* e *a gente* e as suas respectivas concordâncias padrão e não padrão na fala culta de Fortaleza. Os resultados obtidos com a análise dos dados confirmam os achados desta pesquisa, sendo o pronome *a gente* com concordância de terceira pessoa do singular preferido entre os informantes (62%), estando em competição com o pronome *nós* com concordância de primeira pessoa do plural (37,8%). Os dados referentes a concordância não padrão (*a gente* com *-mos* e *nós* sem *-mos*) são escassos, tendo cada uma dessas variantes apenas 1 ocorrência, o que corresponde a 0,1% do total de dados. Esse comparativo evidencia uma tendência na língua falada no que concerne ao uso dos padrões de concordância: o uso pronome *a gente* sem a desinência *-mos* tem se destacado em praticamente todas as comunidades de fala, confirmando os resultados desta pesquisa, enquanto o pronome *nós* com desinência *-mos* figura em segundo lugar na disputa pelo emprego preferencial nas localidades pesquisadas, com exceção da fala goiana, que tem a concordância não padrão impulsionada por elementos culturais (Mattos, 2013).

Na escrita, por outro lado, há maior favorecimento do pronome *nós* com concordância padrão, sendo o pronome *a gente* e suas respectivas concordâncias pouco recorrente nesta modalidade da língua. O uso preferencial do pronome padrão com concordância de primeira pessoa do plural na modalidade escrita pode refletir a pressão normativa exercida pela escola que prescreve o pronome *nós* como única referência à primeira pessoa do plural reconhecida formalmente (Vitorio, 2015). Além disso, questões de percepção e avaliação podem restringir o uso do pronome *a gente* em ambientes mais formais, como o texto escrito, uma vez que o pronome inovador *a gente*, mesmo que recorrente na língua falada, é avaliado como informal, enquanto o pronome padrão *nós* goza de prestígio. Sobre essa questão, Freitas, Favacho e Carvalho (2022, p. 48) explicam:

Destacamos a percepção do uso do *a gente* com concordância avaliado como informal e associado a um uso da periferia urbana. Ainda que essa seja uma forma encaixada na estrutura do Português Brasileiro em todos os estratos sociais, com mais frequência de uso do que o *nós* em estudos de produção por participantes com ensino superior, chegando a tendência de uso de cerca de 80% em dados do Projeto NURC, com falantes escolarizados de cinco capitais brasileiras na década de 1970 (Zilles, 2007) e com 62% das ocorrências em corpus do Projeto PorcuFort coletado na década de 1990, com dados de fala de fortalezenses graduados (Carvalho; Freitas; Favacho, 2020), ainda assim, esse uso segue sem prestígio oficial talvez por fugir do controle da escola, não ser pauta do ensino regular (Freitas, Favacho e Carvalho, 2022, p. 48).

A análise da avaliação e da percepção das formas *nós* e *a gente* e dos padrões de concordância podem explicar os resultados das pesquisas sobre o emprego da concordância

verbal em textos escritos. Brustolin (2009), ao investigar redações escolares de alunos do ensino fundamental de Florianópolis, observou maior recorrência do pronome canônico *nós* com concordância de 1PP (84%), diferentemente do que ocorre na língua falada pelos mesmos informantes, que preferem o padrão emergente (*a gente* sem *-mos*). Essa mesma tendência é observada em textos escolares de alunos do ensino médio de Maceió (Vitorio, 2015), sendo a variante *nós* com *-mos* preferida entre os informantes (77%). O uso de *a gente* com ou sem concordância nessa modalidade é pouco frequente, conforme é possível observar no quadro 12.

Como visto, apesar dos diferentes números que imprimem o comportamento do fenômeno da concordância verbal em cada uma das regiões e em cada modalidade da língua analisada, se nota um padrão na variação da concordância em todas as pesquisas revisadas. As tendências gerais demonstram maior favorecimento dos padrões *a gente* sem *-mos* e *nós* com *-mos*. Neste ponto, é necessário observar o efeito de cada variável isolada, em termos de proporção de uso, sobre a realização da concordância verbal de primeira pessoa do plural e a distribuição das variantes em função de cada preditor observado nesta análise, para verificar quais contextos sociais e linguísticos são mais favoráveis à realização de cada variante de concordância verbal. Utilizamos a linguagem de programação R para verificar a atuação de cada preditor, bem como plotar gráficos de barras e de linhas, além de tabelas com a apresentação dos resultados. Neste sentido, o *RStudio* nos forneceu frequências e proporções, assim como resultados de modelo de regressão logística referentes à atuação de todos os fatores que elencamos como possíveis condicionadores da variação na concordância verbal de primeira pessoa do plural no falar culto de Fortaleza, possibilitando uma análise estatística dos dados coletados em nossa amostra.

### ***6.1.1 Atuação das variáveis predictoras em termos de frequência e proporção de uso: a concordância verbal na Amostra 1990***

Com o objetivo de analisar a atuação de cada variável de ordem linguística e social sobre a realização dos padrões de concordância verbal de primeira pessoa do plural, observamos, em termos de proporção de uso, a distribuição das variantes em função de cada variável predictor elencada neste estudo. Além disso, procedemos à realização de testes de qui-quadrado, que visam examinar a correlação entre as variantes de concordância verbal e cada preditor observado nesta pesquisa. Como discutido anteriormente, o teste de qui-quadrado de

Pearson é importante para avaliar a associação significativa entre as variáveis, identificando o grau de interação entre elas (Oushiro, 2021).

Através da linguagem de programação *R*, por meio do *software R Studio*, submetemos os dados representativos da fala culta de Fortaleza da década de 1990 à análise em termos de estatística descritiva básica. Esta análise nos forneceu os valores das proporções e frequências de uso relacionados à influência de cada variável previsora que atua no fenômeno da concordância verbal com os pronomes *nós* e *a gente*, que apresentamos aqui em formas de gráficos e tabelas. Apresentamos também, sempre que possível, exemplos de ocorrências de dados de fala, com foco nos dados de concordância não padrão, tendo em vista que a concordância padrão (*nós* com *-mos*) e o padrão emergente na língua (*a gente* sem *-mos*) são abundantes em nossa amostra, sendo recorrentemente exemplificada. Entre essas variáveis, a primeira que investigamos, que tem natureza social, trata-se do *sexo/gênero*, conforme descrito na tabela abaixo.

**Tabela 35:** Distribuição das proporções da concordância uso em função do *sexo/gênero*

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Sexo/Gênero</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Homem</b>	483	44,1%	608	55,5%	5	0,4%	0	0%
<b>Mulher</b>	319	37,9%	513	60,9%	9	1,1%	1	0,1%
$\chi^2 (3) = 9.3501, p = 0.02498$								

Fonte: elaboração própria.

Conforme podemos observar nos dados dispostos na tabela 35, em termos de proporção de uso, homens e mulheres apresentam comportamentos discretamente distintos. Em termos de tendência de uso, tanto homens quanto mulheres se inclinam ao uso dos mesmos padrões de concordância na fala culta fortalezense da década de 1990. O padrão emergente de concordância (*a gente* sem *-mos*) é a variante usada em maior proporção na amostra, tanto entre os homens (55,5%), quanto entre as mulheres (60,9%), que tem seu uso ainda mais intensificado. Em relação ao padrão antigo na língua (*nós* com *-mos*), este é usado com expressividade, o segundo padrão de concordância mais observado na comunidade de fala, sendo mais intenso entre os homens (44,1%), e menos intenso entre as mulheres (37,9%), seguindo a tendência dos trabalhos desenvolvidos sobre o tema (Brustolin, 2009; Mendonça, 2010; Rubio, 2012, para citar alguns).

Os dados de concordância não padrão são menos frequentes em nossa amostra, conforme apontam os dados da tabela 35. A forma *nós* sem *-mos* ocorreu em apenas 14 casos,

sendo 5 dados de fala entre os homens (0,4%) e 8 dados entre as mulheres (1,1%). Por fim, a variante *a gente* com *-mos* só obteve 1 dado em nossa amostra, sendo de uma informante mulher (0,1%), representando uma ocorrência pouco usual em nossa base de dados. As ocorrências (160) e (161) ilustram as observações do padrão de concordância *nós* sem *-mos*, e no exemplo (162) se observa a realização da estrutura de *a gente* com *-mos*, único dado referente a esse padrão de concordância em nossa amostra da década de 1990.

- (160) Inf.: eh:: a oraÇÃO... ou para a nossa composiÇÃO... ***nós num precisava*** escrever mais nada porque o dicionário traz todas as palavras... então num seria... o que falta... é a nossa... criativiDAde... em estruRAR (Inq. 114 – F1 – EF, homem, 62 anos)
- (161) Inf.: era um PALmo de daquele BEM fininho como é que a gente andava em cima daquilo eu num sei... sabe? ***nós era*** verdadeiro equilibrista... sabe? pra andar em cima daquilo (Inq. 07 – F1 – DID, mulher, 41 anos)
- (162) Inf.: e que a gente procurou prestar todo esse serviços... que ele realmente necessitava lá:: ***a gente tínhamos*** que ser polivalente... com a farmácia... do meu esposo (Inq. 23 – F1 – DID, mulher, 61 anos)

O dado (162) é um pouco controverso, tendo em vista que se trata de uma ocorrência de concordância não padrão, passível de estigma, que foi produzido por um informante de nível superior, do sexo feminino, com mais de 50 anos. De maneira geral, estudos sociolinguísticos têm apontado “forte correlação entre padrões de estratificação social e gênero, com um grande número de estudos, agora clássicos, em que as mulheres, independentemente de outras categorias sociais, como idade, classe etc., tendem a usar mais formas padrão do que os homens” (Freitag, 2015. p. 29).

Por outro lado, essa ocorrência vai ao encontro dos achados de Rodrigues (2004), ao investigar a concordância verbal de primeira pessoa do plural na fala de 40 informantes adultos moradores de São Paulo, com amostra representativa do falar da década 1980. Diferentemente da tendência atestada em trabalhos sociolinguísticos, o autor identificou na análise dos padrões de concordância verbal que, no lugar de demonstrar “preferência pelas formas de prestígio, as mulheres das comunidades estudadas tendem a usar as formas verbais não-padrão, sem marcas formais de concordância do verbo com seu sujeito” (Rodrigues. 2004, p. 126).

Além disso, para a variável previsora *sexo/gênero*, o teste de qui-quadrado de Pearson resultou em um valor de  $\chi^2 = 9.3501$ . Quanto mais distante de zero for o valor de qui-quadrado, mais significativa será a interação entre a variável previsora e a variável resposta e,

consequentemente, mais relevante será o grupo de fatores para a análise. A análise apontou um (1) grau de liberdade, o que indica que a variável possui dois fatores (homem e mulher).

Em relação ao resultado do teste do estatístico de qui-quadrado, este indicou uma correlação estatisticamente significativa entre a variável *sexo/gênero* e a realização dos padrões de concordância verbal com os pronomes *nós* e *agente* na Amostra 1990, tendo em vista que o valor de p atribuído a esta variável é de  $p = 0.02498$ , estando dentro do intervalo de significância adotado ( $p < 0,05$ ). Diante dos dados, existem evidências estatísticas para rejeitar a hipótese nula de que não há associação entre a variável *sexo/gênero* e os padrões de concordância de primeira pessoa do plural, e acatar a hipótese alternativa.

Dentre as variáveis de natureza extralinguística controladas nesta pesquisa, destaca-se por sua importância em estudos sociolinguísticos a *faixa etária*, variável essencial em uma pesquisa à luz da teoria da variação e mudança linguística. A observação da faixa etária permite prever o comportamento do fenômeno linguístico e ter indícios de mudança linguística em curso ou de variação estável, tendo em vista que as diferenças entre os falantes de diferentes gerações podem ser reflexos das diferenças nas regras adotadas pela comunidade de fala ao longo do tempo (Labov, 1981). Em amostra representativa da década de 1990, a tabela 36 apresenta a distribuição das variantes de concordância verbal em função das três faixas etárias observadas neste estudo.

**Tabela 36:** Distribuição das proporções de uso da concordância verbal por *faixa etária*

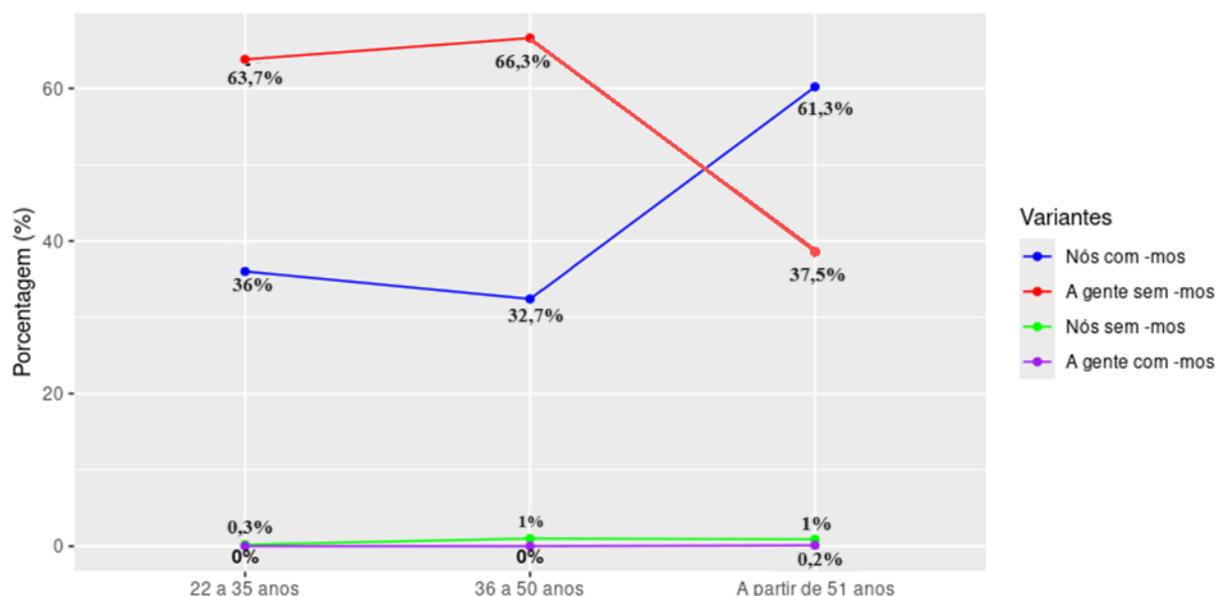
PORCUFORT I - Anos 1990 (N = 1938)								
<i>Faixa etária</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
22 a 35 anos	266	36%	471	63,7%	2	0,3%	0	0%
36 a 50 anos	227	32,7%	461	66,3%	7	1%	0	0%
51 anos ou +	309	61,3%	189	37,5%	5	1%	1	0,2%
<b><math>\chi^2 (6) = 117.46, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

As informações da tabela 36 demonstram que os dados de concordância verbal de primeira pessoa do plural se concentram no padrão antigo na língua (*nós com -mos*) e no padrão emergente (*a gente sem -mos*), tanto entre os informantes de menor faixa etária quanto entre os informantes de maior faixa etária. De acordo com Rubio (2012), os mais jovens são mais tendenciosos a utilizar as variantes inovadoras, que, no caso da concordância verbal, podemos associar ao padrão emergente *a gente sem -mos*, que não é passível de estigma. Por outro lado, os informantes com mais de 50 anos de idade possuem maior tendência a aderir às variantes

conservadoras, mais antigas na língua, que se referem à concordância verbal padrão *nós com -mos*. Esse fato é observado nos dados analisados, conforme podemos observar no gráfico a seguir:

**Gráfico 30:** Proporção de uso da concordância verbal por *faixa etária* (Amostra 1990)



Fonte: Elaboração própria

Entre os informantes mais jovens da década de 1990, de 22 a 35 anos de idade, há maior inclinação ao uso de *a gente sem -mos* (63,7%), sendo o padrão de concordância mais utilizado nessa faixa etária. Disputando a preferência destes informantes na comunidade, está a concordância padrão *nós com -mos* que, embora menos usada que o padrão emergente, ainda apresenta uso expressivo entre os jovens, com 266 ocorrências, o que corresponde a 36% dos dados. Como dito anteriormente, dados de concordância verbal não padrão com marcas de estigma são escassos em nossos dados, principalmente pelo fato de estarmos lidando com informantes classificados como cultos. Por isso, a variante *nós sem -mos* só obteve duas realizações em entre os informantes mais jovens (0,3%), e não encontramos realização da variante *a gente com -mos* entre os informantes de 22 a 35 anos. A título de ilustração, os excertos (163) e (164) são exemplos das duas ocorrências de concordância não padrão *nós sem -mos* encontradas entre os jovens da Amostra 1990:

- (163) Inf.: a vontade popular fosse... a expressão maior... então nós tivemos... com isso... aquilo que *nós chama* de SOvietes... DESde o sistema militar... então... um grupo de miliTAres... SE... REBELOU... contra oficiais... qual era... o interesse dessa rebe-lião?... (Inq. 17 – F1 – EF, mulher, 33 anos)

- (164) Inf.: vamos aprender a reger coRAL *nós é* sempre ensinado:: a gente que pra que a gente consiga alcançar os aGUDos sopra::nos tem que:: é FEchar as vogais todas as vogais são fechadas né e as vezes (Inq. 26 – F1 – D2, mulher, 33 anos)

Em relação aos informantes de segunda faixa etária, de 36 a 50 anos de idade, estes apresentam comportamento linguístico parecido com o comportamento dos mais jovens, sendo o padrão emergente *a gente* sem *-mos* usado com proporção ainda maior, com um total de 461 observações, o que corresponde a 66,3% do total. Seguindo a tendência geral da amostra, a segunda variante mais usada entre os informantes cultos fortalezenses da década de 1990 foi a concordância padrão *nós* com *-mos*, tendo se realizado em 227 ocorrências (32,7%). Por fim, os dados de ocorrência não padrão representam poucas ocorrências em nossa amostra, sendo a estrutura *nós* sem *-mos* com apenas 7 dados, o que representa 1% do total de ocorrências. Não observamos ocorrências de *a gente* com *-mos* entre informantes da faixa etária II (36 a 50 anos) em nossa amostra.

Os informantes a partir de 51 anos de idade, em relação aos mais jovens, apresentam inversão no comportamento linguístico. Enquanto os mais jovens apresentam preferência pelo uso do pronome *a gente* com apagamento da desinência *-mos*, os mais velhos se inclinam ao uso do padrão de concordância mais antigo na língua (*nós* com *-mos*), sendo expressivamente mais utilizado entre os informantes desta faixa de idade (61,3%). Em seguida, o padrão emergente (*a gente* sem *-mos*) continua concorrendo pela preferência dos informantes de faixa etária III, tendo ocorrido em 189 dados, o que equivale a 37,5% dos dados, em conformidade com Rubio (2012). Os casos de concordância não padrão são inexpressivos estatisticamente em nossa amostra. O padrão *nós* sem *-mos* obteve apenas 5 dados (1%), enquanto a estrutura *a gente* com *-mos* se realizou apenas 1 vez em toda nossa amostra representativa da década de 1990, conforme exemplifica a ocorrência (162), sendo um dado de um informante com mais de 51 anos com essas características.

Esse dado de concordância não padrão, pouco recorrente na amostra e passível de estigma como a estrutura *a gente* com *-mos* (Carvalho, Freitas, Favacho, 2020) se torna ainda mais interessante de se analisar quando consideramos que foi produzido por uma informante da faixa etária III. Trabalhos sociolinguísticos, de modo geral, reconhecem os mais jovens como mais receptivos a formas não padrão. Rubio (2012, p. 358), por exemplo, em pesquisa sobre os padrões de concordância verbal no português europeu e brasileiro, trata sobre essa questão ao observar que “em relação à atuação do fator social faixa etária, na amostra brasileira,

observamos tendência de aumento do emprego da forma inovadora nas faixas etárias mais jovens, o que denota implementação de *a gente* sobre *nós*, por processo de sucessão geracional”. Nos dados analisados nesta pesquisa, no entanto, é possível perceber que nem sempre os mais jovens iniciam ou impulsionam o uso de formas não padrão, tendo em vista que estruturas do tipo “a gente tínhamos” é uma forma de concordância não padrão e foi produzida por uma informante de nível superior, do sexo feminino, que geralmente tendem a aderir à formas padrões da língua, e de maior faixa etária, que tradicionalmente tendem a conservar as formas linguísticas gramaticalmente tradicionais (Rubio, 2012).

A próxima variável de natureza extralinguística testada neste estudo foi o *tipo de inquérito*. O controle desta variável previsorora está diretamente ligado ao grau de monitoramento que o informante emprega no momento da elocução, ou seja, à capacidade inerente do falante de observar e avaliar a própria fala dependendo do contexto em que se expressa. Assim, em ambientes de maior formalidade, o nível de monitoramento da fala é amplificado, levando o informante a uma fala mais cuidadosa, evitando assim variantes não padrão (Labov, 1972). Já em ambientes menos formais, há menos monitoramento, uma vez que o informante tende a ajustar sua fala às expectativas sociais e normativas do ambiente de interlocução. A distribuição das variantes de concordância verbais em função do tipo de inquérito pode ser conferida na tabela abaixo:

**Tabela 37:** Distribuição das proporções da concordância verbal por *tipo de inquérito*

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Inquérito</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>D2</b>	138	26,5%	376	72,3%	6	1,2%	0	0%
<b>DID</b>	312	38,9%	487	60,6%	3	0,4%	1	0,1%
<b>EF</b>	352	57,2%	258	42%	5	0,8%	0	0%
<b><math>\chi^2 (6) = 118.52, p &lt; 0,001</math></b>								

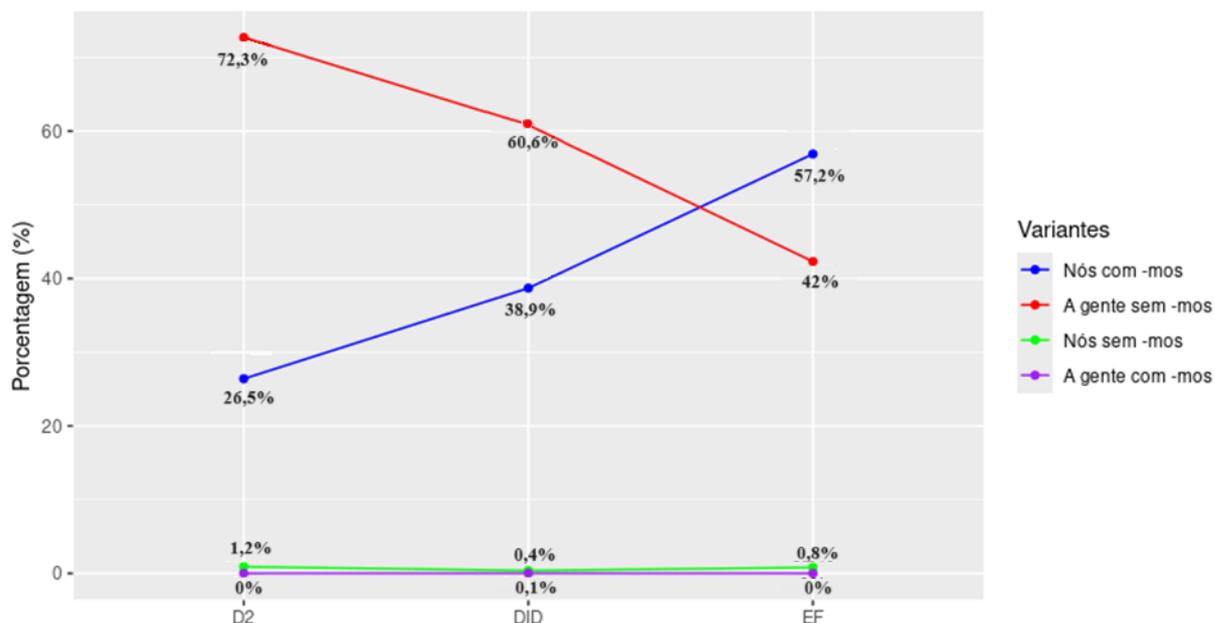
Fonte: elaboração própria.

Como vimos na análise da alternância pronominal entre *nós* e *a gente* (ver capítulo 5), adotamos o entendimento de Coelho et al. (2015) de que o grau de formalidade de determinado registro configura um *continuum*, que vai do menos formal ao mais formal. Tendo isso em vista, os inquéritos do tipo D2 (diálogo entre dois informantes) são os registros que apresenta o menor grau de formalidade entre os registros, pois são gravações que registram conversas entre duas pessoas que possuem certo grau de intimidade, como amigos ou parentes.

Em seguida, inquéritos do tipo DID (diálogo entre dois informantes) possuem grau intermediário de formalidade, e, por fim, os inquéritos do tipo EF (elocução formal), que representam os registros mais formais e monitorados entre os tipos disponíveis, pois são gravações de aulas, palestras, sermões em igrejas e em outros ambientes que exigem certa formalidade (Araújo, 2011).

Dito isto, os dados da tabela 37 apontam os inquéritos do tipo D2 como os mais favoráveis ao uso do padrão de concordância emergente (*a gente* sem *-mos*), tendo sido realizado em 376 ocorrências de concordância verbal com a primeira pessoa do plural, um total expressivo de 72,3%. Esse resultado era esperado, tendo em vista que a análise dos dados de alternância pronominal revelou forte associação entre o pronome *a gente* e inquéritos do tipo D2 (ver seção 5.1).

Há uma queda, porém, no padrão emergente de concordância em inquéritos do tipo DID que, embora não pareça significativa, representa uma diferença de mais de 10 pontos percentuais (60,6%). Isso ocorre ao observarmos os dados em um outro nível que compõe este *continuum* que representa esta variável, uma vez que os registros do tipo DID possuem grau intermediário de formalidade. Essa tendência é ainda mais perceptível quando se observa os dados referentes aos inquéritos do tipo EF, que possuem alto grau de monitoramento da fala. Nestes inquéritos, há uma queda significativa do padrão de concordância emergente *a gente* sem *-mos* (42%), deixando de ser a concordância preferida entre os falantes cultos fortalezenses de 1990. Lima (2017, p. 4), por exemplo, afirma que os falantes preferem o uso “da variante *a gente* em discursos menos formal e com fala sem monitoramento”. Tendo isso em mente, é natural que o uso de estruturas de concordância verbal associadas ao pronome *a gente* diminuam conforme aumentem o grau de formalidade do registro, uma vez que a formalidade e o monitoramento parecem representar resistência à entrada de estruturas com o pronome *a gente* na comunidade em questão. Para uma representação visual da expressividade das variantes de concordância em nossos dados, observemos o gráfico 30:

**Gráfico 30:** Proporção de uso da concordância verbal por faixa etária (Amostra 1990)

Fonte: elaboração própria

Em relação à concordância verbal padrão *nós com -mos*, o gráfico 30 aponta a tendência inversa: quanto mais formal o tipo de registo, maior é o uso da concordância verbal padrão. Em inquéritos do tipo D2, há a menor proporção para a variante mais antiga na língua, com 138 ocorrências (26,5%). Tagliamonte e Denis (2008), ao analisarem mensagens de textos de adolescentes canadenses, concluíram que os contextos mais informais são ambientes que enfrentam menos pressão a aderir às formas linguísticas tradicionais, ambientes então menos propensos ao uso de variantes padrões. Isso poderia explicar a baixa ocorrência da concordância padrão em registros do tipo D2. Em inquéritos do tipo DID, há aumento na taxa de uso da concordância verbal padrão, passando a ser observado em 312 ocorrências, o que representa um total 38,9% dos dados.

Já em Elocuções Formais (EF), a preferência linguística dos informantes cultos se inverte. Ou seja, em ambientes de maior monitoramento da fala, como em aulas, palestras, sermões na igreja, etc, o padrão emergente deixa de ser preferido entre os informantes e a concordância verbal padrão *nós com -mos* assume a preferência entre esses falantes cultos fortalezenes. Em inquéritos do tipo EF a concordância *nós com -mos* obteve proporção de uso de 57,2%. Esse resultado vai ao encontro da pesquisa de Araujo (2019, p. 54) sobre os falares de diversas cidades brasileiras, quando afirma que “as formas verbais clássicas são empregadas quando há maior monitoramento do discurso”.

A concordância não padrão *nós* sem a marcação da desinência *-mos* obteve, em nossa amostra, apenas 6 ocorrências em inquéritos do tipo D2 (1,2%), 3 dados encontrados em registros do tipo DID (0,4%), e apenas 5 observação nos dados de inquéritos do tipo EF (0,8%). Ocorrências da estrutura de concordância não padrão com o pronome *a gente* (*a gente* com *-mos*) só foram observadas em 1 dado de nossa amostra, que pode ser conferido no exemplo (160). Esse dado ocorreu em um inquérito do tipo DID, que possui grau médio de formalidade, por uma mulher por mais de 50 anos.

Além disso, o teste estatístico de qui-quadrado de Pearson, que verifica a atuação da variável *tipo de inquérito* sobre a realização dos padrões de concordância verbal de primeira pessoa do plural apontou o valor de  $\chi^2 = 118.52$ , e um valor de  $p < 0,001$ . Estando dentro do intervalo de confiança adotado neste estudo, o valor  $p$  indica correlação estatisticamente significativa entre as variáveis, o que é indício de que a variável *tipo de inquérito* deve ser incluída no modelo de regressão logística.

A variação na concordância verbal de primeira pessoa do plural também pode ser motivada por preditores de natureza linguística. Dentre as variáveis destacadas neste estudo está o *tempo e o tipo de paradigma verbal*. Esta variável tem suas raízes nos estudos de Naro (1981) e de Naro, Görski e Fernandes (1999), que analisam a escala de saliência fônica e o tempo verbal como variáveis distintas. Posteriormente, Scherre e Naro (2010) conduziram estudo sobre a concordância verbal em que realizaram o cruzamento da saliência fônica e do tempo/modo verbal em conjunto. Essas duas variáveis são complexas, tendo em vista que envolvem vários níveis de análise linguística, sendo motivadas por forças de natureza cognitiva, gramatical e funcional (Yacovenco, Scherre, Naro, 2018). Além disso, podem estar diretamente relacionados a outro fenômeno variável complexo, a queda ou redução da proparoxítone, que tem se mostrado expressivo no português brasileiro (Silva, 2021).

Naro & Scherre (2016), propõem uma remodelagem da variável *tempo e tipo de paradigma verbal*, em que levam em consideração o tempo verbal, a ambiguidade temporal e a saliência fônica, passando a denominar a variável de escala de proeminência. A escala de proeminência proposta pelos autores reorganiza os níveis em cinco categorias, sendo três categorias de presente, uma categoria de pretérito imperfeito e uma categoria que engloba todas as ocorrências de pretérito perfeito, distribuídas em uma variável com cinco níveis, conforme detalham Scherre, Yacovenco e Naro (2018, p. 437):

- a) *imperfeito* – sem ambiguidade temporal com presente e com pretérito perfeito, todos de oposição menos saliente [-Ambíguo, -Saliente];

- b) *presente igual ao pretérito* - com ambiguidade temporal [+Ambíguo], com predominância de formas menos saliente [-Saliente] e alguns casos de oposição mais saliente [+Saliente];
- c) *presente diferente do pretérito* - sem possibilidade de ambiguidade temporal e com oposição menos saliente [-Ambíguo, -Saliente];
- d) *presente diferente do pretérito* - sem possibilidade de ambiguidade temporal e com oposição mais saliente [-Ambíguo, +Saliente];
- e) *pretérito perfeito* - com ambiguidade temporal e oposição mais saliente [+Ambíguo, +Saliente), sem ambiguidade temporal com oposição mais saliente [-Ambíguo, +Saliente] e com ambiguidade temporal com oposição menos saliente [+Ambíguo, -Saliente].

Nosso objetivo é analisar a atuação da estrutura emergente *a gente* sem *-mos* colocando como foco os verbos com ambiguidade temporal, por esse motivo, não consideramos a escala de proeminência desenvolvida pelos autores. Neste primeiro momento, apresentamos todos os tempos e tipos de paradigmas observados em nossa amostra. Em um segundo momento, na análise multivariada através do modelo de regressão logística, trataremos apenas dos dados com tempos verbais no modo indicativo, assim como fizeram Scherre, Yacovenco e Naro (2018). Com isso, a tabela 38 apresenta a distribuição das variantes de concordância verbal em função do *tempo e tipo de paradigma verbal* em nossa amostra dos anos de 1990:

**Tabela 38:** Distribuição das variantes de concordância verbal por tempo e paradigma verbal

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Paradigma</i>	<i>Nós com -mos</i>	<i>%</i>	<i>A gente sem -mos</i>	<i>%</i>	<i>Nós sem -mos</i>	<i>%</i>	<i>A gente com -mos</i>	<i>%</i>
<b>Presente igual</b>	95	27,9%	243	71,5%	2	0,6%	0	0%
<b>Presente Diferente</b>	331	52,6%	295	46,9%	3	0,5%	0	0%
<b>Pret. Perf. Igual</b>	85	49,1%	88	50,9%	0	0%	0	0%
<b>Pret. Perf. Diferente</b>	95	66,4%	48	33,6%	0	0%	0	0%
<b>Pret. imperfeito</b>	80	25,6%	226	72,2%	6	1,9%	1	0,3%
<b>Futuro</b>	75	51,4%	71	48,6%	0	0%	0	0%
<b>Presente do subjuntivo</b>	4	22,2%	14	77,8%	0	0%	0	0%
<b>Imperfeito do subjuntivo</b>	8	32%	16	64%	1	4%	0	0%
<b>Futuro do subjuntivo</b>	3	12,5%	21	87,5%	0	0%	0	0%
<b>Futuro do pretérito</b>	11	36,7%	19	63,3%	0	0%	0	0%
<b>Infinitivo</b>	15	15,5%	80	82,5%	2	2%	0	0%
<b><math>\chi^2 (33) = 201.28, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

Como exposto na tabela 38, em verbos no presente que apresentam a mesma morfologia para os tempos passado e presente (*nós* falamos/corremos/saímos), se destaca a estrutura de concordância emergente *a gente* sem *-mos*, tendo 243 ocorrências (71,5%). Já a estrutura de concordância padrão *nós* com *-mos* se realizou em 95 ocorrências, sendo responsável por 27,9% das ocorrências neste contexto. Ocorrências de concordância verbal não padrão *nós* sem *-mos* são escassas, tendo apenas 2 dados (0,3%), enquanto a estrutura *a gente* com *-mos* não foi observada nesse contexto em nossa amostra. Para exemplificar cada uma das variantes no contexto de tempo presente com a mesma forma na relação presente/passado, as ilustrações (165), (166) e (167) são exemplos de ocorrências de *nós* com *-mos*, *a gente* sem *-mos* e *nós* sem *-mos*, respectivamente:

- (165) Inf.: a minha preferência... sempre foi alfabetização... e depois então o:: primeiro grau:: maior como ***nós chamamos*** aqui né? o primeiro grau maior que é quinta à s:/ à:: oitava SÉrie... e EU susbtituÍA os professores (Inq. 39 – F1 – D2, mulher, 66 anos)
- (166) Inf.: eu FAço... e PAgo minhas promessa direitinho:: eu sei que atualmente a igreja condena a promessa... chama negócio com Deus... chamam negócio... a igre/ é pelo menos na na Te::ologia... ***a gente chama*** negócio com Deus (Inq. 33 – F1 – D2, mulher, 69 anos)
- (167) Inf.: com isso... aquilo que ***nós chama*** de SOvietes... DESde o sistema militar... então... um grupo de miliTares... SE... REBELOU... contra oficiais... qual era... o interesse dessa rebelião? (Inq. 17 – F1 – EF, homem, 33 anos)

Em relação ao tempo presente com morfologias diferentes para os tempos presente e passado (*nós* fazemos/ *nós* fizemos), esse contexto é preferido para realização da concordância padrão *nós* com *-mos*, com 331 dados de fala culta fortalezense, o que representa 52,6% do total. Em seguida, o padrão de concordância emergente disputa a preferência na comunidade de fala, tendo se realizado em 295 das ocorrências (46,9%), enquanto a concordância não padrão *nós* sem *-mos* obteve apenas 3 dados em nossa amostra, totalizando somente 0,5% do total. A estrutura *a gente* com *-mos* não foi observada em contextos de presente com formas diferentes na relação presente/passado. Abaixo, os excertos (168), (169) e (170) ilustram respectivamente as realizações de cada uma das variantes de concordância verbal neste contexto verbal:

- (168) Inf.: aliás num é nem uma conotação... eh:: o sentido REAL... né?... porque:: eh::... o sentido... conotativo é um:: sentido figurado né?... e ***nós podemos*** até lembrar isso de português né? (Inq. 39 – F1 – D2, mulher, 62 anos)

- (169) Inf.: os defensor do Realismo CRIticam... os românticos e ficam essa... confusão por mais de Ano... chega a PON::to... de:: o Castilho... né?... é que **a gente diz** na linguagem... usual chamar pros paus... o Quental... chama o o:: Castilho chama pa/ duelar com... o Quental... e eles vão duelar né? com esGRIma né? (Inq. 36 – F1 – EF, homem, 24 anos)
- (170) Inf.: TOdo o êxodo rural... né?... e denuncia exatamente... a::... todo o desTIIno... desse homem... quer dizer ele quesTIOna exatamente tudo isso... na poeSIa... **nós tem** aqui (Inq. 56 – F1 – EF, mulher, 54 anos)

Contextos verbais que representam o pretérito perfeito com mesma morfologia na relação pretérito perfeito/presente do indicativo (*nós* beijamos/cantamos) são ambientes de disputa acirrada entre o padrão de concordância mais antigo na língua e o padrão emergente. A estrutura padrão *nós* com *-mos* foi observada em 85 ocorrências (49,1%), enquanto a variante de concordância verbal *a gente* sem *-mos* se realizou em 88 dados em nossa amostra (50,9%), demonstrando ser o ambiente linguístico mais disputado pelas formas de concordância verbal entre os informantes cultos de Fortaleza. A estrutura não padrão *nós* sem *-mos*, assim como a forma *a gente* com *-mos* não se realizaram nesse contexto na amostra analisada. Abaixo, apresentamos ocorrências que ilustram cada realização neste contexto nos dados de fala culta, respectivamente.

- (171) Inf.: a:: as palavras... quando **nós estuDAmos**... no nosso tempo quando a gente estudou... eh::... a colocaÇÃO... a única coisa que vem ... cabeça é a colocação dos pronomes... AÍ pronto... quer dizer... pra nós praticamente AINDA HOje... quando se fala em colocaÇÃO... a gente só pensa na colocação pronominal.. (Inq. 114 – F1 – EF, homem, 62 anos)
- (172) Inf.: essa TAxa... de transferência de calor por radiação... **a gente já FALOU**... no:: início... do curso quando /tava apresentando as fórmula/ de resistência de calor... quando a gente não tem GRANde variação de temperatura ela vai ser desprezada (Inq. 54 – F1 – EF, homem, 28 anos)

Em relação a verbos no pretérito perfeito com formas diferentes para o pretérito perfeito e para o presente do indicativo (*nós* sabemos/ *nós* soubemos), só observamos em nossa amostra de fala culta fortalezense ocorrências de concordância verbal com padrão canônico e com padrão emergente. Na Amostra 1990, encontramos 95 dados do padrão *nós* com *-mos* associados a verbos no pretérito com formas diferentes para o passado e presente (66,4%), enquanto o padrão emergente *a gente* sem *-mos* é menos recorrente, sendo observado em 48 ocorrências, o que equivale a 33,6% do total. Os dados (173) e (174), extraídos de nossa base de dados em análise, ilustram ocorrências de nossas variantes neste contexto linguístico:

- (173) Inf.: de onde saiu o Pelé?... né?... nosso BRASilzinho que de repente o homem se tornou o REI... do futebol... né?... **nós tivemos** aí oh... CRAques e mais craques aí GaRRINcha... né?... **nós tivemos** aí::... o próprio ZICO que... que infelizmente na última copa dele... DEU aquela zebra toda... perdeu um pênalti... bestamente... né? (Inq. 20 – F1 – DID, mulher, 32 anos)
- (174) Inf.: não houve dificuldade nenhuma **a gente veio** aqui tanto eu quanto o meu esposo ele ensi/... eu ensinava... no José Barcelos e eu ensinei também... no Paulo Benevides SÓ que eu não achei melhor porque eu ensinei em Tauá eu ensinei principalmente psicologia né?... e então eu achava muito bom no quarto normal... (Inq. 23 – F1 – DID, mulher, 61 anos)

Verbos no pretérito imperfeito do indicativo (cantava/cantávamos), como discutido na seção destinada a análise da alternância pronominal (ver capítulo 5), tendem a ser ambientes mais produtivos para realização da forma *a gente*, devido à natureza dos inquéritos que, no geral, recorrem a técnica da narrativa pessoal, com o intuito de evitar a atenção dos informantes à sua fala e, com isso, evitar o monitoramento. Em relação aos padrões de concordância verbal, o mesmo resultado se repete, tendo em vista a alta proporção da estrutura emergente *a gente* sem *-mos* com verbos no pretérito imperfeito (72,2%). Em seguida, a concordância verbal padrão figura com expressiva proporção de uso, tenho 80 ocorrências, o que representa 25,6% do total. As estruturas de concordância não padrão, como demonstra a tendência geral da Amostra 1990, são pouco recorrentes. A variante *nós* sem *-mos* apresentou apenas 6 ocorrências em nossos dados, conforme exemplificada no excerto (175), sendo responsável por apenas 1,9% do total. Já a estrutura *a gente* com *-mos*, vale destacar, foi observado apenas uma única vez em nossos dados, conforme ilustra o exemplo (176), sendo este dado de pretérito imperfeito (0,3%), tão forte a associação deste tempo verbal com o pronome *a gente*.

- (175) Inf.: principalmente cinema de ARte... quase todo sábado **nós ia**... mas eu me lembro de UM que é Jhonny vai à GUerra... deve ter saído de circulação... eh::... o próprio filme Love story (Inq. 08 – F1 – DID, homem, 40 anos)
- (176) Inf.: mas um povo BOM... e que a gente procurou prestar todo esse serviços... que ele realmente necessitava:: lá **a gente tínhamos** que ser polivalente... com a farmácia do meu esposo (Inq. 23 – F1 – DID, mulher, 61 anos)

No futuro do indicativo as ocorrências de concordância verbal parecem estar bem distribuídas em torno da concordância verbal padrão e do padrão emergente na língua. A forma antiga na língua *nós* com *-mos* obteve 75 realizações (51,4%) e a estrutura *a gente* sem *-mos* se realizou em 71 dados (48,6%). Não se observaram, no entanto, ocorrências de concordância verbal não padrão nesses contextos. De maneira parecida, não houve ocorrências de *nós* sem -

*mos* e *a gente* com *-mos* em verbos no presente do subjuntivo, sendo observados apenas 14 realizações de *a gente* sem *-mos* (77,8%) e 4 ocorrências de *nós* com *-mos* (22,2%).

O futuro do subjuntivo apresentou apenas 24 ocorrências, sendo 21 para a estrutura de concordância emergente *a gente* sem *-mos* (87,5%) e 3 para a estrutura padrão *nós* com *-mos* (12,5%), não sendo observadas realizações de *nós* sem *-mos* ou de *a gente* com *-mos*. Por outro lado, no imperfeito do subjuntivo, a estrutura emergente se realizou em 16 ocorrências (64%), enquanto a estrutura padrão obteve 8 dados (32%) e a concordância não padrão *nós* sem *-mos* foi observada em um único dado de futuro do subjuntivo. Não houve registros, no entanto, da estrutura *a gente* com *-mos* neste tempo verbal.

Já o futuro do pretérito foi registrado em 30 ocorrências de padrões de concordância verbal com os pronomes de primeira pessoa do plural. Neste contexto, o padrão emergente na língua é preferido, tendo obtido 19 ocorrências, o que corresponde a 63,3% do total. Em seguida, o padrão antigo na língua *nós* com *-mos* foi registrado em 11 ocorrências, sendo responsável por 36,7% dos dados. Não houve registro, no entanto, de realizações de estruturas de concordância não padrão *nós* sem *-mos* e *a gente* com *-mos* associados a verbos no futuro do pretérito.

Por fim, os dados da forma nominal infinitivo se mostraram pouco produtivas em relação aos padrões de concordância verbal na comunidade de fala de Fortaleza da década de 1990. A estrutura *a gente* sem *-mos* se mostrou expressivamente favorita (82,5%), enquanto o padrão antigo na língua ocupa o segundo lugar na disputa pela preferência entre os falantes (15,5%). A estrutura não padrão *a nós* sem *-mos* se realizou em 2 ocorrências (2%). Já a forma *a gente* com *-mos* não foi registrada em verbos no infinitivo.

Além disso, submetemos os dados da variável *tempo e tipo de paradigma verbal* à análise do teste estatístico de qui-quadrado de Pearson, que obteve um valor de  $\chi^2 = 201,28$ , valor significativamente diferente de zero. O teste estatístico apresentou *p-value*  $< 0,001$ , indicando que a variável previsorora atua significativamente sobre a realização dos padrões de concordância, uma vez que está dentro do intervalo de confiança arbitrado nesta pesquisa. Isto é indicativo de que este preditor é importante para explicar a variação na concordância verbal e que, por isso, deve ser incluída no modelo de regressão logística.

Além do *tempo e tipo de paradigma verbal*, dentre os fatores linguísticos observados nesta pesquisa está a *estrutura do verbo* que, segundo Zilles, Maya e Silva (2000), pode influenciar no padrão de concordância verbal. Nesta análise, verbo simples se refere a realização isolada da forma verbal, desacompanhada de qualquer outro elemento verbal. (*nós cantamos*). Locuções verbais correspondem a estrutura verbal formada pelo verbo auxiliar

acompanhado pelo verbo principal na sua forma nominal (nós *devemos analisar* todos os documentos). Já os verbos compostos correspondem a realizações do verbo ter ou haver acompanhado do particípio do verbo principal (nós *temos analisado* muitos processos). Nossa hipótese inicial é a de que os verbos simples favoreciam o padrão emergente *a gente sem -mos*, devido à alta proporção desta estrutura em nossa amostra. Os resultados desta variável em função dos padrões de concordância podem ser conferidos na tabela abaixo:

**Tabela 39:** Distribuição das proporções da concordância verbal por *estrutura do verbo*

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Estrutura</i>	<i>Nós com - mos</i>	%	<i>A gente sem - mos</i>	%	<i>Nós sem - mos</i>	%	<i>A gente com - mos</i>	%
<b>Simple</b>	632	41,8%	867	57,4%	12	0,8%	0	0%
<b>Locução</b>	162	39,4%	246	59,9	2	0,5%	1	0,2%
<b>Composto</b>	8	50%	8	50%	0	0%	0	0%
<b><math>\chi^2 (6) = 5.1781, p = 0.5212</math></b>								

Fonte: elaboração própria.

Conforme é possível observar na tabela 39, em relação ao padrão de concordância verbal emergente *a gente sem -mos*, este padrão é usado em maior proporção em praticamente todos os níveis desta variável. Quando se trata de verbos simples, a variante emergente se realiza em 867 ocorrências em nossa amostra de fala culta fortalezense da década de 1990, equivalente a 57,4% das realizações de concordância verbal. Em seguida, em locuções verbais, encontramos 246 dados de *a gente sem -mos*, o que corresponde a 59,9% do total. Já em contextos de verbos compostos, há apenas 8 dados em nossa amostra, que totaliza 50% das ocorrências. Os excertos (177), (178) e (179) ilustram as ocorrências de *a gente sem -mos* associados a verbos simples, a locuções verbais e a verbos compostos, respectivamente:

- (177) Inf.: *a gente FEZ* o curso de de computação um mês... eu conheci a Inês lá... sim *a gente fez* o curso de computação foram vinte dias vinte dias MESmo... cinco cinco {cinco e cinco... tá entendendo? enTÃO... ELES passaram quando *a gente terminou* o CURso (Inq. 02 – F1 – D2, mulher, 23 anos)
- (178) Inf.: fazendo uma analogia com:: a espessura crítica também da esFERa né?... a mesma coisa *a gente pode fazer* uma vez que A ÁREA... vai vaRIAR... e quando a gente varia... eh:: essa espessura de isolamento... TÁ? (Inq. 54 – F1 – EF, homem, 28 anos)
- (179) Inf.: na aula anterior... *a gente já havia discutido*... sobre a diferença entre esgoto secundário e esgoto primário né?... é bom lembrar que o esgoto secundÁRIO... é aquele que não recebe os gases... provenientes desses consumidores (Inq. 152 – F1 – EF, homem, 41 anos)

Seguindo a tendência geral de nossa amostra, a concordância verbal padrão com o pronome *nós* é a segunda variante, em termos de proporção de uso, mais expressiva em nossa amostra. De acordo com as informações expostas na tabela 39, a estrutura de concordância padrão *nós* com *-mos* é mais frequente em verbos simples, tendo um total 632 dados, o que representa 41,8% dos casos de concordância registrados. Em relação às locuções verbais, as realizações da concordância verbal padrão *nós* com *-mos* ocorrem de forma mais discreta, sendo observadas em 162 dados de fala culta fortalezense, totalizando 39.4% das ocorrências. Já o padrão *nós* com *-mos* associado a verbos compostos se realizou em apenas 8 ocorrências, o que totaliza 50% dos dados neste contexto linguístico. Para exemplificar, as ocorrências (180), (181) e (182) ilustram a realização da concordância verbal *nós* com *-mos* em verbo simples, em locução verbal e em verbo composto, respectivamente.

- (180) Inf.: Será que DEUS num é uNIVERSAL mesmo não por que... de repente só o brasileiro tem que ganhar? por que *nós* temos que ser os melhores? *nós jogamos* o melhor futeBOL?... *nós* já *joGAmos* o melhor futebol aTÉ... *nós joGAmos* né?... mas cada UM tem seu esPAço tem seu dia... sabe? (Inq. 20 – F1 – DID, mulher, 32 anos)
- (181) Inf.:vai SER:: NUM::... pré-vestibular exatamente dividido em:: *nós estamos vendo* a possibilidade de seis salas... né? por enquanto só tem cinco porque são cinco televisõ::es e cinco video-cassete né? (Inq. 07 – F1 – D2, mulher, 26 anos)
- (182) Inf.: na Escola de Engenharia que trabalhava/ na IBM... *nós* sabíamos que *nós tínhamos sido* aproVAdo... todos os todos os cinco... inclusive a a::... é acho... o GRUpo que tinha tinha ti/ obtido o melhor resultado (Inq. 46 – F1 – DID, homem, 48 anos)

Os padrões de concordância que são passíveis de estigma obtiveram poucos dados em nossa amostra. A estrutura não padrão *nós* sem *-mos* foi observada em apenas 12 dados em amostra dos anos de 1990, o que corresponde a 0,8% do total em contexto de verbos simples. Já as observações do pronome *nós* sem a marcação da desinência de plural *-mos* (*nós* sem *-mos*) associadas a locuções verbais apareceram em apenas duas ocorrências, o que equivale a somente 0,5% dos dados. Realizações com essa estrutura em verbos compostos não foram observadas em nossa amostra. Com isso, para exemplificar as ocorrências das variantes de concordância verbal nessas variáveis, os excertos (183) e (184) ilustram dados da concordância não padrão *nós* sem *-mos* com verbos simples e com locuções verbais, respectivamente:

- (183) Inf.: é uma poeSia... que... expressa muito bem... Todo o êxodo rural... né?... e denuncia exatamente... a:... todo o desTIIno... desse homem... quer dizer ele quesTIOna exatamente tudo isso... na poeSia... **nós tem** aqui (Inq. 56 – F1 – EF, mulher, 54 anos)
- (184) Inf.: eh:: a oraÇÃO... ou para a nossa composiÇÃO... **nós** num **precisava escrever** mais nada porque o dicionário traz todas as palavras... então num seria... o que falta... é a nossa... criativiDAde... em estrutuRAR... (Inq. 114 – F1 – EF, homem, 62 anos)

Em relação à concordância não padrão com o pronome *a gente* (*a gente* com *-mos*), conforme já havíamos comentado, encontramos apenas um dado em nossa amostra representativa do falar culto de Fortaleza da década de 1990. Trata-se de um dado de locução verbal e pode ser conferido no excerto (176), não havendo ocorrências dessas variantes associadas a verbos simples ou a verbos compostos.

Por fim, o teste estatístico em que a variável preditora foi submetida, o teste de qui-quadrado de Pearson, chegou a um valor de  $\chi^2 = 5.1781$  com seis graus de liberdade, sendo significativamente diferente de zero. Além disso, o valor de *p* atribuído pelo teste estatístico está acima do valor de alfa adotado nesta pesquisa, tendo o *p-value* = 0.5212. O baixo valor de qui-quadrado associado ao alto valor de *p* nos dá indícios suficientes para não rejeitar a hipótese nula e determinar que não existe correlação estatisticamente significativa entre os padrões de concordância e a estrutura do verbo. Motivo pelo qual, possivelmente, esta variável preditora poderá ser excluída no teste de seleção automática através da função *step*.

Outra variável de natureza linguística que testamos neste estudo trata-se da *conjugação verbal*. Essa variável preditora se mostrou relevante em estudo conduzido por Zilles, Maya e Silva (2000) sobre os padrões de concordância verbal de primeira pessoa do plural nas cidades de Panambi e Porto Alegre, no Rio Grande do Sul. Para os autores, os verbos de terceira conjugação apresentam tendência ao apagamento da desinência plural *-mos*, enquanto os verbos de primeira e de terceira conjugação favorecem a marcação da desinência *-mos*. Com o objetivo de observar se esses resultados se repetem na amostra analisada, apresentamos a distribuição dos padrões de concordância em função da variável *conjugação verbal* em nossos dados na tabela 40:

**Tabela 40:** Distribuição das proporções da concordância verbal por *conjugação verbal*

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Conjugação</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>1ª conjugação</b>	192	29,8%	447	69,4%	5	0,8%	0	0%
<b>2ª conjugação</b>	485	48,9%	498	50,3%	7	0,7%	1	0,1%
<b>3ª conjugação</b>	125	41,3%	176	58,1%	2	0,6%	0	0%
<b><math>\chi^2 (6) = 61.28, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

Em termos de proporção de uso, os dados da tabela 40 revelam que, em todos os contextos, o uso do padrão emergente *a gente sem -mos* é mais favorável. Em usos com verbos na primeira conjugação, o padrão emergente se mostrou fortemente favorecido, com 447 ocorrências e uma proporção de uso de 69,4%, sendo o contexto mais favorável a este padrão de concordância. Já na segunda conjugação, o uso de *a gente sem -mos* representa um ambiente de disputa linguística, tendo ocorrido em 50,3% dos dados. Por fim, os verbos de terceira conjugação representam a menor quantidade de nossa amostra. Nesses verbos, o padrão emergente também é favorecido, apresentando 176 ocorrências, o que corresponde a uma proporção de uso de 58,1%. Esses dados revelam que, em todos os contextos dessa variável, o padrão emergente apresenta proporções de uso maiores que 50%.

Em relação à concordância verbal padrão *nós com -mos*, este é o segundo padrão de concordância mais utilizado entre os informantes cultos fortalezenses na Amostra 1990. Em verbos de primeira conjugação, há 192 ocorrências de concordância padrão, o que corresponde a 29,8%. Já com verbos na segunda conjugação, observamos 485 ocorrências desse padrão, ou seja, uma proporção de uso de 48,9%. Por fim, verbos que possuem a terceira conjugação associados ao padrão de concordância *nós com -mos* somam 125 ocorrências, o que representa 41,3% dos dados.

Dados de concordância verbal não padrão apresentam a mesma tendência observada nas outras variáveis, ou seja, poucas ocorrências. O padrão de concordância *nós com -mos* foi flagrado em apenas 5 dados de primeira conjugação em nossa Amostra 1990 (0,8%), enquanto o mesmo padrão, junto a verbos da segunda conjugação, foi observado em 7 ocorrências (0,7%). Realizações da concordância não padrão *nós sem -mos* associadas a verbos de terceira conjugação se realizaram apenas 2 vezes em nossos dados, o que representa 0,6% dos dados neste contexto. Para ilustrar essas ocorrências de concordância não padrão, os

exemplos (185), (186) e (187) demonstram ocorrências de *nós* sem *-mos* com verbos na primeira, segunda e terceira conjugações, respectivamente:

- (185) Inf.: e... FOmos pra Universidade COMO?... gratuitamente eu sempre analiso isso ***nós estuda***... EU por exemplo estudei fiz a minha Faculdade... gratuitamente... mas quem pagou a Universidade? quem foi? (Inq. 39 – F1 – D2, mulher, 66 anos)
- (186) Inf.: a qualidade desse curso que é muito BOM muito escla-reDOR... aí depois você POde juntar aqui coMIGO ou com Eri-NALdo a Tânia uma pessoa mais esclarecida... pra ***nós fazer*** aquela reunião com os professores (Inq. 16 – F1 – D2, mulher, 30 anos)
- (187) Inf.: a tantos filmes que:: principalmente cinema de ARte... quase todo sábado ***nós ia***... mas eu me lembro de UM que é Jhonny vai à GUerra... deve ter saído de circulação... eh::... o próprio filme Love story (Inq. 08 – F1 – DID, homem, 40 anos)

Em nossa amostra representativa do falar culto fortalezense da década de 1990, encontramos apenas um dado de concordância não padrão com o pronome *a gente* (*a gente* com *-mos*), que já expomos no exemplo (176), e pertence a uma ocorrência associada a um verbo da segunda conjugação, o verbo *ter*: “a gente tínhamos”. Além do mais, o teste estatístico de qui quadrado identificou valor de  $\chi^2 = 61.28$  e  $p\text{-value} < 0,001$ , estando dentro do intervalo de confiabilidade. Isso demonstra forte associação entre a conjugação do verbo e os padrões de concordância verbal, motivo pelo qual esse preditor deve ser incluído no modelo de regressão logística para ser analisada em um contexto multivariado.

A última variável preditora de natureza linguística que pretendemos incluir no modelo de regressão logística para a análise multivariada é a *tonicidade do verbo*. Este preditor se mostrou significativo em diversas pesquisas sociolinguísticas que investigam os padrões de concordância verbal com a primeira pessoa do plural, como Zilles, Maya e Silva (2000), Rubio (2012), Agostinho (2013), entre outros. Zilles, Maya e Silva (2000) observaram que as proparoxítonas estão relacionadas à marcação da desinência *-mos*. Na mesma direção, Agostinho (2013) identificou maior associação da marca de terceira pessoa do singular a verbos oxítonos. Com o intuito de observar se essa mesma tendência se repete em nossos dados, a tabela 41 detalha a distribuição das variantes de concordância verbal em função da variável preditora *tonicidade*:

**Tabela 41:** Distribuição das proporções da concordância verbal por *tonicidade do verbo*

PORCUFORT I - Anos 1990 (N = 1938)								
<i>Tonicidade</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Monossílabo</b>	0	0%	385	99,5%	2	0,5%	0	0%
<b>Oxítono</b>	0	0%	173	98,9%	2	1,1%	0	0%
<b>Paroxítono</b>	723	55,8%	563	43,4%	10	0,8%	0	0%
<b>Proparoxítono</b>	79	98,8%	0	0%	0	0%	1	1,2%
$\chi^2 (9) = 603.55, p < 0,001$								

Fonte: elaboração própria.

Os dados da tabela 41 demonstram comportamentos diferentes das variantes de concordância verbal a depender da tonicidade do verbo. As informações da tabela apontam as palavras monossílabas como ambientes linguísticos categóricos ao uso de padrão emergente *a gente sem -mos*, obtendo 385 dados, um total de 99,5% das observações, enquanto a concordância não padrão *nós sem -mos* foi observado em 2 dados, totalizando 0,5%. Dados de concordância padrão *nós com -mos* e de concordância não padrão *a gente sem -mos* não ocorreram com verbos monossílabos em nossa amostra, o que já esperávamos, tendo em visto que o acréscimo da desinência *-mos* alteraria o número de sílaba da palavra, fazendo com que determinada palavra deixe de ser monossílaba. Os excertos (188) e (189) são exemplos de concordância não padrão com o pronome nós associados a verbos monossílabos:

(188) Inf.: vamos aprender a reger coRAL ***nós é*** sempre ensinado:: a gente que pra que a gente consiga alcançar os aGUDos sopra::nos tem que:: é FEchar as vogais todas as vogais são fechadas né e as vezes (Inq. 26 – F1 – D2, mulher, 33 anos)

(189) Inf.: todo o desTIno... desse homem... quer dizer ele quesTIOna exatamente tudo isso... na poeSIa... ***nós tem*** aqui (Inq. 56 – F1 – EF, mulher, 54 anos)

Ainda em relação à *tonicidade do verbo*, as oxítonas têm um comportamento parecido ao dos monossílabos. Em verbos oxítonos, há expressiva proporção de uso do padrão de concordância emergente (*a gente sem -mos*), com uma frequência de 173 dados observados em nossa Amostra 1990 e proporção de uso praticamente categórica de 98,9%. Por outro lado, a variante *nós sem -mos* apresenta baixa frequência, sendo realizada em apenas 2 dados com verbos oxítonos (0,5%), conforme pode ser observado no excerto (190). Esses dados se referem às realizações de verbos no infinitivo sem a marcação da desinência *-mos*. Já as estruturas de concordância padrão *nós com -mos* e de concordância verbal não padrão *a gente com -mos*

associadas a verbos oxítonos não foram encontrados em nossa amostra, tendo em vista a impossibilidade de realização de verbos oxítonos com desinência *-mos* no português brasileiro (Silvano, 2014).

- (190) Inf.: uma pessoa mais esclarecida... pra ***nós fazer*** aquela reunião com os professores digo "O:lhe Simões... se for uma reunião... pra ficar com aquilo ali MESmo... que nem ao menos vão ter o trabalho de ler o texto... é preferível nem começar... (Inq. 56 – F1 – EF, mulher, 54 anos)

Ocorrências com verbos paroxítonos representam a maior quantidade de dados em nossa amostra, totalizando 1296 realizações dos padrões de concordância verbal de primeira pessoa do plural. Esse contexto linguístico representa um ambiente de disputa linguística acirrada entre a concordância verbal padrão *nós* com *-mos*, que obteve 723 realizações (55,8%), e o padrão emergente de concordância *a gente* sem *-mos*, que foi observado em 563 ocorrências (43,4%) em nossa amostra do PORCUFORT da década de 1990. Não foram observadas, no entanto, realizações da estrutura de concordância verbal *a gente* com *-mos* associados a verbos com acento paroxítono, enquanto dados de concordância não padrão *nós* sem *-mos* são pouco recorrentes em nossa amostra, tendo ocorrido em apenas 10 ocorrências (0,8%) neste contexto, conforme ilustra exemplo (191):

- (191) Inf.: existia muito bambolê... inclusive era coisa de Rico... logo que começou logo que surgiu sabe?... aí depo/ e ***NÓS sabe*** que que nós pobre fazíamos/ pegava um... uma uma borracha que tinha um ferro dentro num sei pra quê que servia aquilo (Inq. 12 – F1 – DID, mulher, 41 anos)

Por fim, em palavras proparoxítonas, confirmando os estudos de Zilles, Maya e Silva (2000) e Agostinho (2013), há uma tendência praticamente categórica à marcação da desinência de plural *-mos*. Confirme é possível conferir na tabela 41, o padrão de concordância mais antigo na língua portuguesa *nós* com *-mos* está fortemente associado a verbos proparoxítonos, tendo 79 ocorrências (98,8%). Por outro lado, o padrão de concordância emergente na língua, formado pelo pronome inovador *a gente* sem a desinência *-mos*, não foi registrado em nossa amostra de fala culta junto a verbos proparoxítonos, assim como dados de concordância verbal não padrão *nós* sem *-mos*. A estrutura *a gente* com *-mos* foi registrada em apenas uma ocorrência com verbos proparoxítonos, como mostra o exemplo (176). Os excertos (192) e (193), a título de ilustração, exemplificam ocorrências das variantes *nós* com *-mos*, e *a gente* com *-mos*, respectivamente, com verbos de tonicidade proparoxítona:

- (192) Inf.: não podíamos sentar... sem que ele dissesse... "boa tar::de gente" ele falava assim... e depois enTÃO... **nós sentávamos**... mas ANtes disso ah:: ele saía ele fazia o seguinte ele saía da classe... (Inq. 39 – F1 – D2, mulher, 58 anos)
- (193) ele realmente necessitava lá:: **a gente tínhamos** que ser polivalente... com a farmácia... do meu esposo (Inq. 23 – F1 – DID, mulher, 61 anos)

### 6.1.2 Análise multivariada em um modelo de regressão logística: a concordância verbal em um recorte binário - Amostra 1990

Neste recorte, para a análise multivariada, trabalharemos apenas com dados de concordância verbal junto a verbos com tempos verbais no modo indicativo, tendo em vista que pretendemos testar o efeito da ambiguidade temporal causada por verbos que possuem a mesma morfologia para os tempos presente do indicativo e passado. É nesse modo verbal que se concentram verbos com essa característica (Scherre, Yacovenco e Naro, 2018). Nossa hipótese é a de que a inserção do padrão emergente *a gente* sem *-mos* atuaria na questão da ambiguidade potencial semântico temporal desses verbos. Portanto, para efeito de análise e descrição dos dados da concordância verbal, removemos de nossa amostra as ocorrências associadas a verbos no futuro (146 dados), presente do subjuntivo (18), imperfeito do subjuntivo (25), futuro do subjuntivo (24), futuro do pretérito (30) e infinitivo (97), restando um total de 1598 dados de concordância verbal de primeira pessoa do plural, distribuídos da seguinte forma:

**Tabela 42:** Distribuição de concordância verbal com dados de modo indicativo

Variantes	<i>Nós com -mos</i>	<i>Nós sem -mos</i>	<i>A gente sem -mos</i>	<i>A gente com -mos</i>
Frequência	686	11	900	1
Proporção	42,9%	0,7%	56,3%	0,1 %

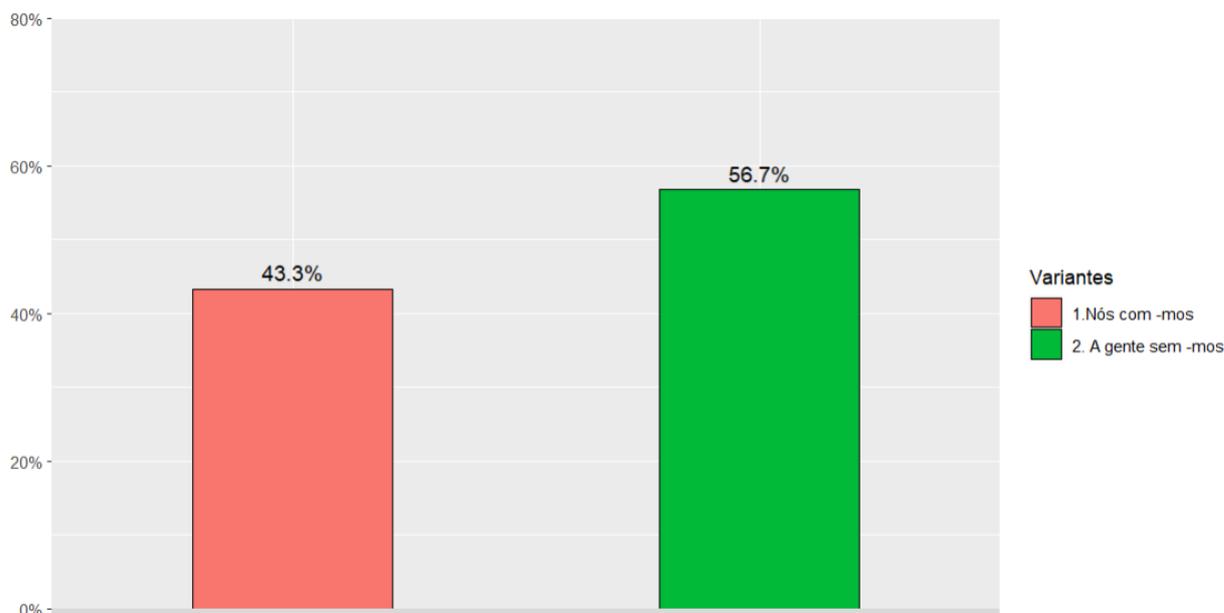
Fonte: elaboração própria

Além disso, nossa ideia inicial para esta investigação seria a realização de uma análise multivariada quaternária, envolvendo todas as quatro variantes de concordância verbal de primeira pessoa do plural, através de um modelo de regressão multinomial, quando envolve uma variável resposta com mais de dois níveis (Levshina, 2015). No entanto, os poucos dados de concordância não padrão causaram um expressivo desequilíbrio na distribuição das variantes

na amostra. Como exemplo desse desequilíbrio, citamos a variável *tonicidade verbal* (ver tabela 40), que possui várias células com valores vazios, ou seja, sem ocorrências de concordância verbal. Para Gries (2013), a falta de balanceamento dos dados ou valores vazios podem afetar a convergência do modelo de regressão multinomial.

Na amostra representativa do falar culto da década de 1990, o desbalanceamento na distribuição de dados afeta diretamente na verossimilhança de um possível modelo multinomial. Variantes com poucas ocorrências influenciam diretamente na qualidade das previsões feitas pelo modelo de regressão e na estabilidade dos valores estimados. Em nossos dados, especialmente a variante *a gente sem -mos* pode ser problemática, tendo em vista que possui apenas 1 dado a ser contrastado em um ambiente multivariado com, por exemplo, 900 dados da variante *a gente sem -mos*. Da mesma forma, a variante *nós sem -mos* apresentou 11 observações em toda amostra, número irrisório se comparado com as ocorrências de *nós com -mos* (686) e *a gente sem -mos* (900). Com tamanho desbalanceamento, não foi possível proceder uma análise multinomial envolvendo as quatro variantes, ou sequer foi possível realizar uma análise ternária como fizeram Scherre, Yacovenco e Naro (2018) com os padrões *nós com -mos*, *a gente sem mos* e *nós sem -mos*.

Em casos como esse, uma das técnicas mais sensatas a serem empregadas é a exclusão dos níveis da variável resposta com poucas ocorrências para realizar uma regressão logística binária. Isso resolveria o problema da falta de balanceamento dos dados, que é um dos princípios básicos do modelo de regressão logística. Além disso, ofereceria verossimilhança aos coeficientes estimados e evitaria o problema da estabilidade dos valores (Gries, 2013). Diante do cenário da distribuição de nossos dados, optamos por excluir as observações de *nós sem -mos* (11 dados) e de *a gente com -mos* (1 dado), passando então a trabalhar, a título de análise, com 1586 dados da fala culta fortalezense dos anos 1990 em um recorte binário, conforme gráfico abaixo:

**Gráfico 31:** Proporção de *nós* com *-mos* e *a gente* sem *-mos* (N=1586)

Fonte: elaboração própria

Conforme é possível observar no gráfico 31, a exclusão dos dados de concordância não padrão provocou um discreto aumento na proporção de uso da concordância padrão *nós* com *-mos* (de 41,4% para 43,3%), ao passo que a proporção de uso do padrão de concordância emergente *a gente* sem *-mos* diminuiu (de 57,8% para 56,7%). Destacamos que essa diferença representa 1,4 pontos percentuais, o que não parece ser tão expressiva a ponto de comprometer significativamente a análise dos dados. De qualquer forma, para a construção e modelagem do modelo de regressão logística binária, consideramos apenas esses dados apresentados no gráfico 31, tendo em vista as escassas ocorrências das formas de concordância não padrão em nossa amostra.

Para a criação do modelo de regressão logística, consideramos as variáveis predictoras e os testes estatísticos realizados. Apesar de o teste de qui-quadrado de Pearson sugerir não haver interação significativa entre a variável *estrutura verbal* e a realização do fenômeno variável de concordância verbal, decidimos mantê-la no modelo de regressão para o observar sua atuação em um contexto multivariado. Por meio do *software R Studio*, através da função *glmer* (Generalized Linear Model), criamos um modelo logístico com as variáveis predictoras *sexo/gênero*, *faixa etária*, *tipo de inquérito*, *tempo* e *tipo de paradigma verbal*, *estrutura verbal*, *tonicidade verbal*, *conjugação verbal* e a variável aleatória *informante*. O modelo apresenta a seguinte sintaxe inserido em script no ambiente *R Studio*:

```
glmer(VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO + PARADIGMA +
ESTRUTURA.VERBO + TONICIDADE + CONJUGACAO + (1|INFORMANTE), data =
Concordancia_1990, family = binomial)
```

Após proceder a criação do modelo logístico, para posteriormente realizar os testes de multicolinearidade e de seleção automática das variáveis, o modelo não convergiu corretamente, uma vez que um erro foi constatado no ambiente *R Studio*. A seguinte mensagem é mostrada:

**Figura 8:** Mensagem de erro ao convergir o modelo de regressão logística

```
Warning messages:
1: In sqrt(vv) : NaNs produced
```

Fonte: elaboração própria

O erro reportado pode indicar que problemas no balanceamento dos dados continuam, o que sugere que houve problemas ao ajustar o modelo. A expressão "NaNs produced" pode indicar diversos problemas com os dados, como valores não numéricos, valores inexistentes/ausentes (células vazias), dados inadequados ou até mesmo presença de multicolinearidade (Maltefort, 2012). Após análise exploratória de nossos dados, identificamos muitas categorias vazias na variável *tonicidade* (ver tabela 34), o que pode estar causando problemas com a convergência do modelo, tendo em vista que esta variável apresenta muitas células vazias. Optamos então por criar um novo modelo sem a inclusão da variável *tonicidade verbal*, o que fez o modelo convergir normalmente.

Após o programa convergir com sucesso, submetemos o modelo logístico a análise através da função *vif* (Variance Inflation Factor), que integra o pacote CAR. O teste *Vif* permite verificar se o modelo de regressão criado possui multicolinearidade entre as variáveis previsoras, ou seja, testa se existe algum grau de redundância entre as variáveis testadas, o que pode comprometer a análise. O resultado do teste de multicolinearidade resultou no seguinte *output*:

**Tabela 43:** Resultado do teste de colinearidade *vif* da concordância verbal – Amostra 1990

Resultado do teste de inflação da variância			
	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
<b>Sexo/gênero</b>	1.103519	1	1.050485
<b>Faixa etária</b>	1.160577	2	1.037931
<b>Tipo de inquérito</b>	1.192145	2	1.044918
<b>Paradigma verbal</b>	2.192255	4	1.103091
<b>Estrutura Verbal</b>	1.130652	2	1.031175
<b>Conjugação</b>	1.918169	2	1.176852
<b>car::vif(modelo)</b>			

Fonte: Elaboração própria.

De acordo com Levshina (2015), pesquisas mais rigorosas adotam 5 como valor máximo de GVif ajustado (GVIF<sup>1/(2\*Df)</sup>) para determinar quando há ou não multicolinearidade no modelo. Em nosso caso, nenhum dos valores ultrapassou o número 2, indicando que não possuímos qualquer problema com variáveis redundantes (mais de uma variável que testa a mesma informação), o que ocasionaria problemas nas estimativas dos coeficientes. O maior coeficiente apresentado é referente a variável *tempo e tipo de paradigma verbal do modo indicativo*, que obteve valor igual a 2.192255, mas que, depois de ajustado, o valor caiu para 1.103091, comprovando ausência de multicolinearidade no modelo, que é um dos princípios básicos da regressão logística.

Após comprovada a ausência de multicolinearidade em nossos dados, submetemos o modelo de regressão logística à seleção automática de variáveis predictoras através da função *step*. Como tratado no capítulo 5, essa função faz a seleção de forma automatizada das variáveis que são estatisticamente significativas para explicar o fenômeno da variação e, com isso, ajuda na criação de um modelo logístico final para a análise das variáveis em um contexto multivariado. Como vimos, a seleção pode ser feita por meio do *step forward*, *step backward* ou *step both*, e em todos os testes a mesma informação é fornecida, ou seja, a sugestão de quais variáveis devem permanecer no modelo e quais devem ser removidas. Para determinar que variáveis devem ser incluídas no modelo final, consideramos aquelas que apresentam os menores índices de AIC (Levshina, 2015).

Procedemos, então, à análise da seleção automatizada das variáveis linguísticas e extralinguísticas de nosso modelo através dos três parâmetros: *forward*, *backward* e *both*. Em todas as análises, foi recomendado a permanência de todas as variáveis predictoras incluídas no modelo inicial. Com isso, apesar do teste de qui-quadrado não ter apontado interação relevante entre a variável *estrutura verbal* e os padrões de concordância, no ambiente multivariado, esta

se mostrou significativa. A tabela 44 apresenta o resultado do teste de seleção automática das variáveis através da função *step*:

**Tabela 44:** Resultado da seleção das variáveis de concordância por meio da função *step*

<b>Resultado Step forward (AIC = 2175.09)</b>			
	<b>Df</b>	<b>Df Deviance</b>	<b>AIC</b>
<b>Paradigma verbal</b>	4	2044.0	2054.0
<b>Faixa etária</b>	2	2058.0	2064.0
<b>Tipo de inquérito</b>	2	2063.1	2069.1
<b>Conjugação</b>	2	2111.3	2117.3
<b>Sexo/gênero</b>	1	2163.6	2173.1
<b>Estrutura Verbal</b>	2	2167.1	2173.1
<b>&lt;none&gt;</b>		2173.1	2175.1
<b>Modelo sugerido: VR ~ PARADIGMA + FAIXA.ETARIA + TIPO.INQUERITO + CONJUGACAO + SEXO.GENERO + ESTRUTURA.VERBAL</b>			

Fonte: elaboração própria

A tabela 44 apresenta os resultados de AIC de vários modelos distintos. O último dos modelos, chamado de *<none>*, ou de modelo nulo, representa um modelo vazio, que não contém qualquer variável previsora. Em seguida, um modelo que inclui a variável estrutura verbal é testado e apresenta um AIC menor que o do modelo nulo, indicando que a inclusão desta variável melhora o poder explicativo do modelo de regressão. Isso acontece à medida em que se inclui uma nova variável, de que modo que não há piora do modelo com a inclusão de qualquer uma das variáveis testadas. Isso é indicativo de que todas as variáveis incluídas no modelo inicial devem permanecer no modelo final. Desta forma, a função *step* apontou como modelo ideal para explicar os nossos dados de fala culta da década de 1990 aquele que contenha as variáveis *tempo e tipo de paradigma do modo indicativo, faixa etária, tipo de inquérito, conjugação verbal, sexo/gênero e estrutura verbal*. Permanecemos, então, com o mesmo modelo inicial.

Com o objetivo de comprovar a qualidade do modelo indicado na função *step*, aplicamos a modelagem à função *lrm* (Logistic Regression Model), do pacote *rms* (Regression Modeling Strategies) para obter o grau de verossimilhança que as medidas estatísticas do modelo podem indicar. Dentre os valores tomados como referência, como explicado (ver seção 5.1) está o índice de concordância C, medida que adotamos para determinar o poder explicativo do modelo de regressão logística. A tabela a seguir apresenta o resultado do modelo submetido a função *lrm*:

**Tabela 45:** Resultado do teste multivariado de *nós* e *a gente* através da função *lrm*

	Model Likelihood Ratio Test			Discrimination Indexes		Rank Discrim. Indexes	
<b>Obs</b>	1586	<b>LR chi2</b>	682.32	<b>R2</b>	0.287	<b>C</b>	0.769
<b>A gente sem -mos</b>	900	<b>d.f.</b>	13	<b>R2(13,1589)</b>	0.207	<b>Dxy</b>	0.769
<b>Nós com -mos</b>	686	<b>Pr(&gt;chi2)</b>	>0.0001	<b>R2(13,1169.5)</b>	0.271	<b>Gamma</b>	0.541
<b>Max reriv</b>	2e-13			<b>Brier</b>	0.192	<b>Tau-a</b>	0.541

Fonte: Elaboração própria.

Os dados da tabela 45 indicam uma série de informações sobre o modelo de regressão criado para o tratamento de nossos dados. Como vimos, a informação mais relevante para nós é a estimativa referente ao índice de concordância C. Para Levshina (2015), valores de concordância C ente 0,7 e 0,8 possuem poder de discriminação dos dados aceitável. Em nosso caso, o índice C é igual a 0,769, o que atesta o poder explicativo que o modelo de regressão criado para tratar os dados de concordância possui, sendo eficaz na análise desses dados.

Tenho comprovada a qualidade do nosso modelo logístico na explicação dos dados, procedemos a análise multivariada da atuação das variáveis predictoras sobre a variável resposta, através da função *glmer*. O modelo submetido a função *glmer* inclui todas as variáveis testadas neste estudo sobre concordância verbal, com exceção da variável predictor *tonicidade*, que apresentou problemas de valores vazias em seus dados, o que fez com que o modelo não pudesse convergir adequadamente. Além disso, o valor da razão de Verossimilhança (LR chi2) atribuído ao modelo foi um valor igual a 682.32, o que representa um índice confiável de verossimilhança no tratamento dos dados.

Com isso, a análise do modelo de regressão logística gerou um valor de *intercept*, que é composto pelo segundo nível da variável resposta em relação ao primeiro nível de cada variável predictor. Na análise dos padrões de concordância verbal de primeira pessoa do plural com dados da Amostra 1990, nossa variável de referência é a estrutura *nós* com *-mos*. Já observamos o pronome *a gente* como variável de referência ao analisar a alternância pronominal. Buscamos então, ver o comportamento da variante canônica, a concordância padrão *nós* com *-mos*, em contraste com a o padrão emergente *a gente* sem *-mos*, observando esse contraste em um modelo multivariado. Desta forma, neste modelo de regressão, o valor de *intercept* se refere ao uso da estrutura de concordância padrão *nós* com *-mos* nos seguintes contextos: (1) homem, (2) 22 a 35 anos, (3) inquérito do tipo DID, (4) presente com morfologia

diferente, (5) primeira conjugação e (6) verbos com estrutura simples. A tabela abaixo apresenta, em termos de favorecimento, a análise multivariada dos dados de concordância verbal na fala culta da década de 1990:

**Tabela 46:** Estimativas de *nós* com *-mos* no modelo de regressão logística: 1990 (N = 1586)

<i>Coefficients</i>	Estimate	Std. Error	z value	Pr(> z )	
<i>(Intercept)</i>	-0.6989	0.2099	-3.330	0.000867	***
<i>Sexo/gênero</i>					
Mulher	-0.2494	0.1231	-2.027	0.042677	*
<i>Faixa etária</i>					
36 a 50 anos	0.1799	0.1414	1.272	0.203372	
A partir de 50 anos	1.5180	0.1528	9.932	< 2e-16	***
<i>Tipo de inquérito</i>					
D2	-0.5535	0.1454	-3.807	0.000140	***
EF	0.8828	0.1467	6.016	1.78e-09	***
<i>Tempo e Paradigma verbal</i>					
Presente igual	-0.8986	0.1947	-4.616	3.91e-06	***
Pret. Perf. Igual	-0.3583	0.2190	1.636	0.101852	
Pret. Perf. Diferente	0.8397	0.2228	3.769	0.000164	***
Pret. imperfeito	-0.9625	0.1755	-5.483	4.18e-08	***
<i>Conjugação verbal</i>					
2ª conjugação	0.4600	0.1632	2.819	0.004814	**
3ª conjugação	-0.2326	0.2433	-0.956	0.339052	
<i>Estrutura verbal</i>					
Locução verbal	-0.6319	0.1706	-3.705	0.000212	***
Verbo composto	0.2833	0.5518	0.513	0.607701	
Modelo: glm(VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO + ESTRUTURA.VERBAL + CONJUGACAO + PARADIGMA.VERBAL)					
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					

Fonte: elaboração própria.

O resultado da regressão logística apresentada na tabela 46 demonstra que todas as variáveis previsoras incluídas no modelo apresentam interação estatisticamente significativa com a realização da estrutura de concordância padrão, indicada pelo valor de *p* (*p value*). As estimativas de uso, apresentadas em *logodds*, apontam a atuação de cada fator sobre a variável resposta, de modo que estimativas com valores positivos indicam favorecimento, enquanto valores negativos apontam tendência ao desfavorecimento em relação ao *intercept*. A seguir, apresentamos os resultados, em termos de favorecimentos, da atuação de cada variável sobre a realização dos padrões de concordância verbal em um recorte binário.

a) *Atuação do sexo/gênero sobre os padrões de concordância verbal*

A variável *sexo/gênero* é de expressiva relevância em estudos sociolinguísticos e frequentemente testadas em análises sobre a concordância verbal com os pronomes *nós* e *a gente*, conforme demonstram Zilles, Maya e Silva (2000), Vianna e Lopes (2003), Zilles e Batista (2006), Rubio (2012), Agostinho (2013), entre outros. Esta variável, embora aparente ser simples, é bastante complexa, tendo em vista que revela informações importantes sobre a caracterização da comunidade de fala. Para além da diferenciação entre a fala de homens e mulheres e os papéis sociais que homens e mulheres podem exercer na comunidade, a observação desta variável pode revelar indícios, por exemplo, de estabilidade ou de prováveis mudanças linguísticas, se considerarmos os pressupostos de Labov (1990).

Para o sociolinguista, existem dois princípios básicos associados aos efeitos da variável *sexo/gênero* sobre a realização dos fenômenos de variação linguística: a) em fenômenos linguísticos em estado de variação estável, os homens tendem a usar a variante não padrão mais frequentemente que as mulheres; e b) em fenômenos que apresentam indícios de mudança linguística, as mulheres tendem a aderir a formas inovadoras de forma mais frequente que os homens.

Partindo dessa visão, a análise dos nossos dados indica que, entre os homens (valor de referência), nos contextos de realização do *intercept*, há desfavorecimento da concordância padrão *nós* com *-mos*, com estimativa de -0.6989, e valor de p igual a 0.000867, o que demonstra significância estatística. No entanto, em relação aos homens, as mulheres tendem a inibir ainda mais o uso da concordância verbal padrão, demonstrando inclinação ao uso da forma emergente, considerada inovadora na língua (*a gente* sem *-mos*), com estimativa de uso de -0,9484, valor menor do que o *intercept*, o que demonstra maior desfavorecimento.

Comparando os dois cenários apresentados por Labov (1990) com o resultado da análise dos nossos dados, poderíamos dizer que a situação b) representaria melhor a realidade de nossa amostra, uma vez que, em nossa análise, as mulheres apresentaram maior tendência ao uso da forma inovadora que os homens. No entanto, é superficial afirmar que existe mudança linguística, ou até mesmo indícios de mudança observando apenas a atuação de uma única variável. Há de se analisar também variáveis relevantes em contexto multivariado.

Carvalho e Santos (2023), em estudo sobre a concordância verbal com os pronomes *nós* e *a gente* na mesma comunidade de fala que nós, em dados de fala da década de 1980, extraídos a partir do projeto Dialeto Sociais Cearenses (DSC), composto por dezoito falantes em entrevistas sociolinguísticas (Santos; Viana; Araújo, 2021), observou tendência inversa à

encontrada em nossa análise. No modelo de regressão logística desenvolvido pelos autores, os homens apresentaram maior tendência a desfavorecer a forma padrão *nós* com *-mos*, estando mais propensos ao uso da forma *a gente* sem *-mos* que as mulheres. Obviamente, há de se considerar que a metodologia dos *corpora* é distinta, os traços sociais dos informantes são diferentes, mas o fato de se tratar da mesma comunidade também é uma informação relevante, mesmo que os dados tenham sido obtidos através de processos metodológicos distintos.

Por outro lado, Silvano (2016), ao estudar os padrões de concordância verbal em textos alunos de Florianópolis, observou comportamentos distintos para homens e mulheres, o que pode ser um demonstrativo de que o fenômeno se manifesta de formas diferentes na fala e na escrita. Segundo o autor, as mulheres tendem a ser mais conservadoras, com maior tendência a aderir a forma de concordância padrão (P.R. 0,72), diferente dos informantes do sexo masculinos, que são mais propícios ao uso da forma não padrão (P.R. 0,24). De forma parecida, Agostinho (2013), em investigação sobre a escrita escolar, identificou maior tendência entre as alunas ao uso de *nós* com *-mos* (P.R. 0,58), enquanto os alunos demonstraram desfavorecimento da forma (P.R. 0,40).

A variável *sexo/gênero* é comprovadamente importante em estudos sociolinguísticos. No entanto, não é possível considerá-la isoladamente, ainda mais tendo por análise um modelo de regressão logística que investiga a atuação das variáveis em um contexto multivariado. Assim, a próxima variável que apresentamos, igualmente importante para o entendimento de um fenômeno sociolinguístico, é a *faixa etária*.

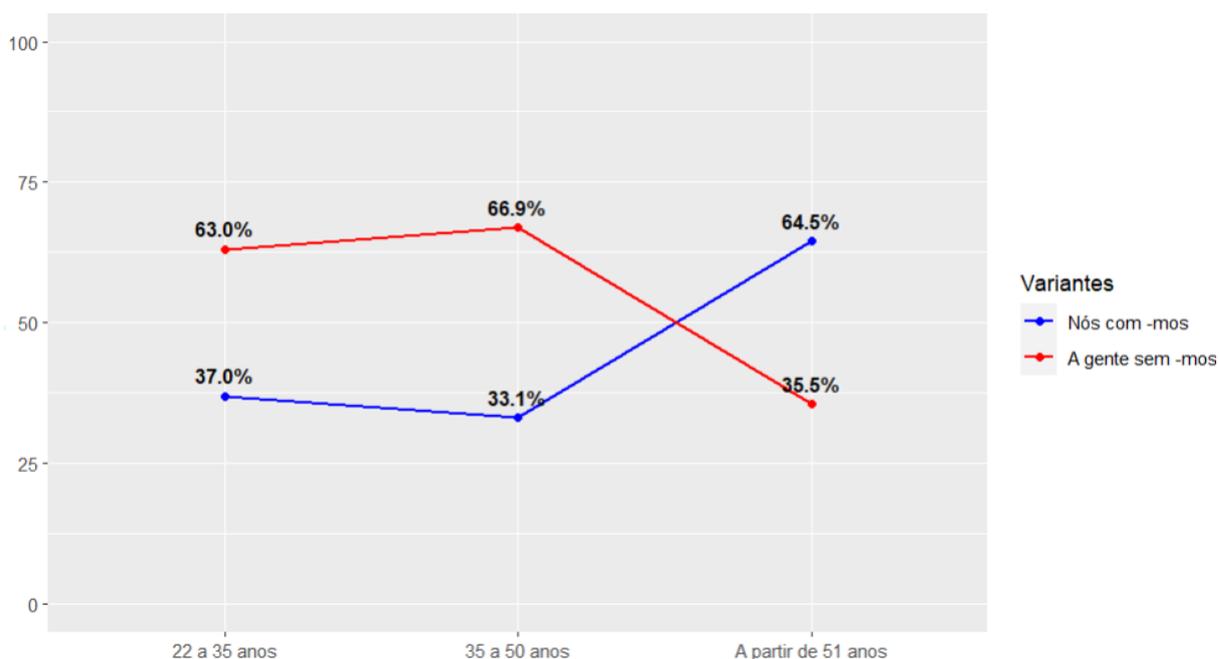
#### *b) Atuação da faixa etária sobre os padrões de concordância verbal*

A observação da variável previsora *faixa etária* constitui uma etapa essencial em um estudo sociolinguístico nos moldes labovianos, especialmente quando o estudo trata da variação na concordância verbal de primeira pessoa do plural, tendo em vista que a distribuição das ocorrências em função das faixas etárias pode auxiliar na elaboração de projeções acerca do fenômeno variável, fornecendo informações que são tomadas como norte para estudos pautados a luz da teoria da variação e da mudança linguística, além de fornecer dados que podem indicar indícios de estabilidade ou de mudança linguística (Araujo, 2014). Com isso, a análise da *faixa etária* pode possibilitar a formulação de hipóteses sobre a formação cultural e histórica da comunidade de fala em estudo, levando a uma compreensão mais ampla da realidade linguística na comunidade. Mattos (2013), por exemplo, observou expressivas

ocorrências da concordância verbal não padrão *nós* sem *nós* (“*nós é*”, “*nós fala*”), distribuídas em todas as faixas etárias na fala goiana. Essa característica, segundo a autora, tem aspecto cultural, uma vez que é fruto das raízes caipiras da região e é compartilhado sem estigma entre os falantes, como marca da identidade.

Ao observar esta variável em tempo aparente, a análise sociolinguística pode indicar o estágio da variação e mudança linguística do fenômeno. A análise em tempo real, por outro lado, pode apontar os contextos mais favoráveis à expansão das variantes ou mesmo a restrição de determinada forma, considerando uma série de fatores, como a avaliação que o falante faz da própria língua. É importante destacar que, de modo geral, tanto em estudos que investigam a fala popular quanto nos que investigam a fala culta, os jovens têm mostrado maior tendência ao uso de formas inovadoras (Foeger, 2014; Mattos, 2013; Rubio, 2012);

Nos dados representativos da fala culta fortalezense de 1990, essa tendência atestada em outros estudos pode ser observada. Conforme é possível observar nos dados a tabela 46, os informantes mais jovens, aqueles que possuem de 22 a 35 anos, constituem o *intercept* do modelo de regressão logística. Nos mesmos contextos de realização do *intercept*, o valor da estimativa para os mais jovens é de -0.6989, conforme exposto na tabela. O valor negativo aponta que os informantes mais jovens de nossa amostra tendem a desfavorecer o uso da concordância verbal padrão *nós* com *-mos* em nosso modelo, sendo estes mais favoráveis a forma *a gente* sem o *-mos*. Esse resultado já era esperado, pois a análise da alternância pronominal (ver capítulo 5), apontou forte interação entre os informantes mais jovens e o uso do pronome *a gente*. Além disso, o valor atribuído ao *p-value* = 0.000867 demonstra uma interação estatisticamente significativa, indicando que, em relação à *faixa etária*, as análises da alternância e da concordância verbal se confirmam. O gráfico 32 apresenta a tendência no uso das variantes de acordo com a *faixa etária*.

**Gráfico 32:** Proporção de *nós com -mos* e *a gente sem -mos* por faixa etária (Amostra 1990)

Fonte: elaboração própria

De modo geral, as tendências apresentadas pelos resultados do modelo de regressão confirmam a distribuição dos padrões de concordância *nós com -mos* e *a gente sem -mos* em função das faixas etárias. O gráfico 32 aponta menor adesão das faixas mais jovens ao padrão canônico de concordância, sendo observada em apenas 37% das ocorrências nas falas dos mais jovens, e isso foi comprovado no modelo de regressão, que indicou os mais jovens como desfavoráveis ao uso de *nós com -mos*. Esse resultado se assemelha ao de Foeger (2014), que, ao investigar a concordância verbal com os pronomes de primeira pessoa do plural em Santa Leopoldina, chegou a resultados próximos aos encontrados na amostra de Fortaleza dos anos de 1990, ou seja, o padrão mais antigo de concordância *nós com -mos* é menos frequente entre os mais jovens, tendo se realizado em apenas 44% dos dados, sendo estes desfavoráveis ao uso da concordância padrão (P.R.0,39). Para a autora, isso pode sugerir uma possível mudança em curso no sentido de o padrão de concordância ser substituído pelo padrão emergente *a gente sem -mos*.

Os informantes da faixa etária II, de 26 a 35 anos, embora tenham se destacado no sentido de apresentar a menor proporção de uso da forma de concordância padrão *nós com -mos* entre as faixas pesquisadas (33,1%), não demonstraram relevância estatística em nosso modelo multivariado, tendo o valor de p maior do que o adotado como nível de confiança desta pesquisa. Isso faz com que desconsideremos os seus resultados, uma vez que sua verossimilhança não é estatisticamente relevante.

Em relação aos informantes com mais de 51 anos, o modelo de regressão logística apontou forte interação entre a realização das variantes de concordância verbal e esses informantes. De acordo com os dados da tabela 46, informantes com mais de 50 anos de idade tendem a favorecer expressamente o uso da estrutura padrão de concordância *nós* com *-mos*, apresentando uma estimativa de uso de 0,8191, indicando que, em relação aos mais jovens, os mais velhos da amostra estão mais inclinados ao uso de *nós* com *-mos*. Esse resultado está alinhado com os achados na análise da alternância pronominal da década de 1990 (ver seção 5.1), tendo em vista que a análise identificou forte associação entre o pronome padrão *nós* e os informantes da faixa etária III.

Nossos resultados também confirmam os achados por Mattos (2013) na fala goiana, em que encontrou maior percentual de uso do padrão canônico *nós* com *-mos* em informantes com idade entre 41 e 86 anos, com percentual de uso que chega a 94% das ocorrências. Lima (2023), na mesma direção, observou que os informantes entre 50 e 65 anos, os mais velhos, eram mais favoráveis ao uso de *nós* com desinência de primeira pessoa *-mos*, tendo observado a estrutura de concordância padrão em 74,7% das ocorrências em amostra do Atlas Linguístico do Maranhão, enquanto os mais jovens eram responsáveis por 56,9% desses dados. No entanto, a autora pondera que, apesar de “nossos dados mostrarem um uso bastante acentuado da não-concordância na primeira faixa etária, não podemos afirmar a existência de uma mudança em progresso em direção à perda da concordância”

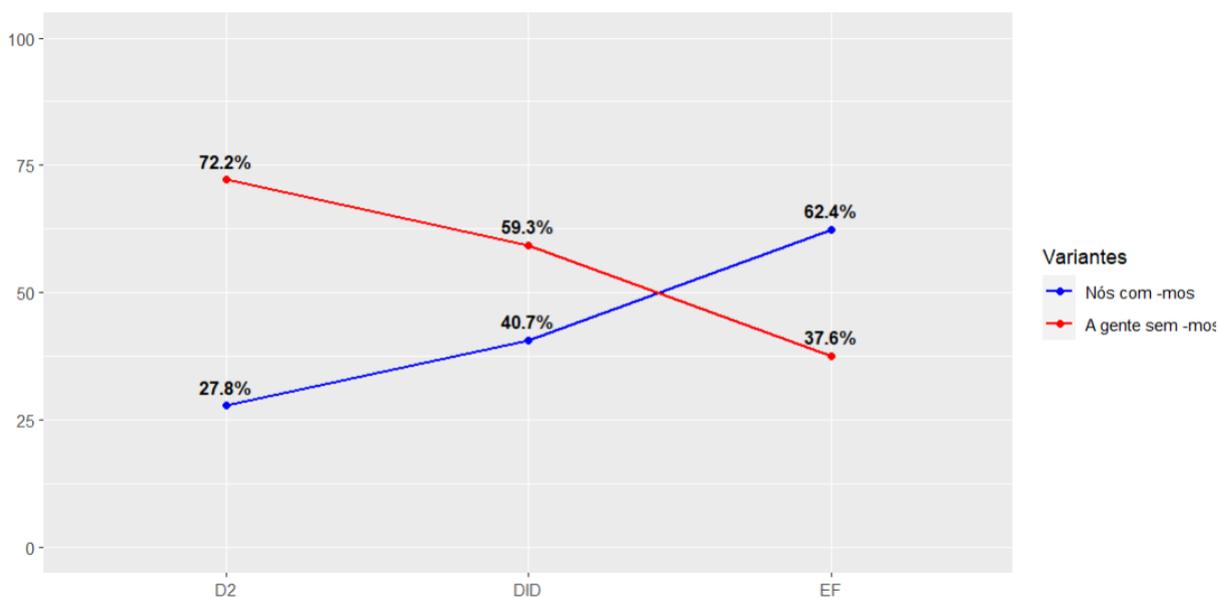
### *c) Atuação do tipo de inquérito sobre os padrões de concordância verbal*

Como discutido anteriormente, esta variável representa um *continuum* de formalidade (Coelho et al., 2015), que engloba os tipos de registro, desde aqueles que apresentam menor grau de formalidade e/ou monitoramento da própria fala (D2), perpassando por registro com grau intermediário de monitoramento (DID), até aqueles registros que exigem alto grau de formalidade (EF). Apesar de *nós* e *a gente* terem o mesmo referente e ocuparem um lugar de variação na língua, a situação de comunicação, o tipo de discurso, ou grau de monitoramento dado a fala pode condicionar o uso de uma ou de outra variante com suas respectivas concordâncias. Omena (1996) explique que há situações em que há maior condicionamento ao uso de uma forma em detrimento da outra:

em alguns contextos de fala, não se usa a gente por nós; em outros, há uma distribuição condicionada por fatores linguísticos e/ou sociais ou, ainda, uma aparente situação em que uma forma pode alternar-se indiferentemente com a outra; em certos pontos, porém a substituição parece ter se completado ou estar em vias de fazê-lo. (Omena, 1996, p. 211).

Neste modelo de regressão, o valor de referência se refere a probabilidade do uso da variante padrão *nós* com *-mos* para inquéritos do tipo DID (diálogo entre informante e documentador). Ou seja, considerando os fatores em que o *intercept* se realiza, há desfavorecimento no uso de *nós* com *-mos* em inquéritos do tipo DID, uma vez que sua estimativa de uso é um número negativo (-0.6989). Isso quer dizer que em registros em que consideramos médio o grau de formalidade, há menor propensão ao uso da concordância padrão e maior inclinação ao uso do padrão emergente de concordância (*a gente* sem *-mos*). Nesse sentido, a tendência é confirmada pelas proporções de uso. O gráfico abaixo demonstra as proporções de cada variante de concordância verbal em cada tipo de inquérito:

**Gráfico 33:** Proporções de uso de *nós* com *-mos* e *a gente* sem *-mos* por tipo de inquérito



Elaboração própria

Conforme é possível observar na tabela 46, há desfavorecimento ao padrão de concordância canônico indicado pelo modelo de regressão em informantes de inquéritos do tipo DID, o que é confirmado em termos de proporção de uso. O gráfico 33 indica proporção de uso de 40,7% de concordância padrão em inquéritos do tipo DID, demonstrando que nestes tipos de registro, o uso do padrão emergente se destaca, tendo 59,3% das ocorrências. Ao considerar esses inquéritos como um *continuum* de formalidade, em que os inquéritos de tipo DID

representam o ponto médio, é importante observar os polos dessa escala e contrapor os resultados, para ver se a tendência, em termos de favorecimento, se observa também em termos de proporção de uso.

Os inquéritos menos formais dentre os três tipos analisados nesta pesquisa, os inquéritos do tipo D2 (diálogo entre dois informantes), demonstraram expressiva adesão ao padrão emergente *a gente* sem *-mos*, conforme demonstra o gráfico 33, sendo responsável por 72,2% das ocorrências. Em termos de favorecimento, conforme demonstra o resultado do modelo de regressão exposto na tabela 46, inquéritos do tipo D2 apresentam intenso desfavorecimento do padrão canônico de concordância *nós* com *-mos*, com estimativa de -1,2521, número bem mais baixo que o do *intercept*, o que demonstra que este tipo de inquérito está ainda mais desfavorável ao uso do padrão canônico que os inquéritos do tipo DID. Este fato aponta os inquéritos menos formais como mais inclinados à adesão da forma de concordância emergente *a gente* sem *-mos*, padrão geral desta amostra. Esse resultado pode ser considerado estatisticamente confiável, tendo em vista que o valor de  $p = 0.000140$ , respeitou o limite de confiabilidade adotado neste estudo.

Já os inquéritos que registam situações de elocuições formais (EF), os mais formais de todos, uma vez que constituem gravações de aulas, palestras, cerimônias religiosas e outras situações que exigem grande atenção à fala, estes apresentaram características inversas aos outros tipos de inquéritos, o que é de se esperar, uma vez que ambientes que exigem maior formalidade tendem a favorecer variantes padrões. Sobre o uso de *nós* e *a gente* e suas respectivas concordâncias, Brustolin (2009), faz considerações importantes:

Nossas variáveis *nós* e *a gente* são duas formas alternativas para designar o pronome pessoal de primeira pessoa do plural e são usadas em certas situações de interação entre os indivíduos – dependendo do nível de formalidade usa-se uma em detrimento da outra. Além disso, há também a questão do uso de *a gente* entre as diferentes gerações, em que os mais velhos preferem o uso de *nós* e os mais jovens o uso de *a gente*. Ressaltamos, ainda, que o uso desses pronomes não é estigmatizado, mas está correlacionado a variáveis extralinguísticas, por exemplo, nível de formalidade, escrita versus fala e faixa etária dos falantes. Contudo, verificamos que o estigma no uso desses pronomes, em geral, está relacionado à concordância verbal: realização – *mos* (*a gente cantamos*) ou zero (*nó(i)s canta*). (Brustolin, 2009, p. 90-91).

Confirmando as palavras de Brustolin (2009), em nossos dados, identificamos que, em inquéritos menos formais, há maior tendência ao uso do padrão emergente *a gente* sem *-mos*, enquanto em situações mais formais, como em inquéritos do tipo EF, há maior proporção de uso da estrutura de concordância verbal canônica, sendo 62,4% de *nós* com *-mos*, enquanto a estrutura emergente é menos recorrente nestes contextos de maior formalidade, com 37,6%

de realizações de *a gente* sem *-mos*. Em termos de estatística inferencial, a análise multivariada através do modelo de regressão logística indicou intenso favorecimento da concordância padrão *nós* com *-mos* em inquéritos do tipo EF, com estimativa 0,1839. O número positivo indica que as elocuições formais são os únicos contextos em que o padrão *nós* com *-mos* se destaca, confirmando um *continuum* na escala de formalidade que, em nossos dados, obedecem a seguinte ordem em termos de favorecimento da estrutura canônica D2 => DID => EF.

*d) Atuação do tempo e tipo de paradigma verbal sobre os padrões de concordância verbal*

Como explicado anteriormente, a observação deste preditor se deu inspirado nos trabalhos de Scherre, Yacovenco e Naro (2018), com devidas adaptações, e de Carvalho, Freitas e Favacho (2020), pressupondo o efeito significativo que os usos nos tempos do modo indicativo exercem sobre a realização dos padrões de concordância verbal associados aos pronomes *nós* e *a gente*. Por conta disso, vale destacar, neste recorte binário, analisamos apenas dados de *nós* com *-mos* e *a gente* sem *-mos* no modo indicativo. Scherre, Yacovenco e Naro (2018), baseados nos estudos de Naro, Fernandes e Görski (1999), Naro & Scherre (2016), desenvolveram estudo com a variável *tempo e tipo de paradigma verbal do modo indicativo* para testar as tendências atestadas nestes estudos de que a) o padrão de concordância canônico *nós* com *-mos* parece estar associado do tempo pretérito perfeito b) a marcação da desinência *-mos* (*nós dormimos*) vs a ausência da desinência (*nós dorme*) resolveria o problema da ambiguidade temporal entre os verbos que possui a mesma morfologia para o pretérito perfeito e para o presente do indicativo (Scherre, Yacovenco e Naro, 2018, p. 429).

Nesse caso, segundo a tendência sugerida por Naro, Görski e Fernandes (1999), a queda da desinência *-mos*, especialmente em verbos no presente com mesma morfologia, atuaria na questão da ambiguidade semântico temporal. Nesse sentido, a inserção do pronome *a gente* no quadro pronominal brasileiro levou a criação de um padrão emergente de concordância verbal, que naturalmente está associado à ausência da desinência *-mos*. Dessa forma, o uso da estrutura de concordância *a gente* sem *-mos* em verbos regulares no modo indicativo teria um importante aspecto funcional, o de agir sobre a questão da ambiguidade temporal natural desses verbos. Daremos este enfoque funcional no tratamento dos dados à luz desta variável, não considerando a questão da escala da proeminência proposta por Naro e Scherre (2016).

Dito isto, verbos no presente do indicativo com morfologia diferente para o pretérito perfeito e para o presente do indicativo representam o valor de referência deste modelo de regressão (*intercept*). Nesse sentido, considerando todos os outros contextos que constituem o *intercept*, verbos no presente com formas diferentes tendem a desfavorecer a concordância padrão *nós* com *-mos*, com estimativa de  $-0.6989$ . Este resultado vai na contramão dos encontrados por Nato et al., (2014), que observaram o tempo presente com formas diferentes para o pretérito perfeito e presente como favoráveis ao uso da desinência *-mos* (P.R. 734), e por Benfica (2016), que constatou comportamento próximo ao neutro para este fator (P.R. 0.508). No entanto, um dos nossos objetivos é observar como os padrões de concordância agem em verbos com formas iguais no sentido de desfazer a ambiguidade temporal.

Em relação aos dados no presente do indicativo com mesma morfologia para o pretérito perfeito e para o presente do indicativo, se observa forte inibição de *nós* com *-mos*, indicando que este é um contexto expressivamente favorável ao uso do padrão emergente *a gente* sem o *-mos*, com estimativa de uso de  $-1,5975$ . Esse valor indica que esse contexto verbal favorece significativamente ao uso do padrão emergente *a gente* sem *-mos*. O resultado corrobora os achados obtidos na investigação de Scherre, Yacovenco e Naro (2018), em investigação sobre os padrões de concordância de primeira pessoa do plural com foco na variável *tempo e tipo de paradigma verbal do modo indicativo*. Para os autores, a inserção de *a gente* sem *-mos* como padrão emergente de concordância atuaria em uma questão funcional da língua, a ambiguidade temporal provocada por alguns verbos regulares quando conjugados na primeira pessoa do plural. De acordo com os linguistas, a porta de entrada da estrutura emergente seria:

em especial pelo presente que pode ter a mesma forma do pretérito. Na possibilidade real de produção de *nós* sem *-mos*, seja por força natural da prosódia do português brasileiro, que evita construções proparoxítonas, seja para desfazer ambiguidade potencial e assegurar uma leitura de tempo presente, o que gera estruturas com estigma, sujeitas ao preconceito e à intolerância linguística pela comunidade de fala brasileira letrada (Scherre / Naro 2014), a ampliação e a expansão dos usos de *a gente* sem *-mos*, com especial força em áreas urbanas mais amplas, se apresentam como uma estratégia intuitiva, que permite resolução de conflitos sociolinguísticos, por meio de mais usos de estruturas com concordância (Scherre, Yacovenco e Naro, 2018, p.23).

Assim, com a inserção do pronome *a gente* no quadro pronominal do português, o verbo passa a ser conjugado no singular, se ligando a terceira pessoa. Com o verbo na terceira pessoa do singular, a ausência da desinência *-mos* auxiliaria na questão de dupla interpretação temporal “*nós cantamos*”, por exemplo, tendo em vista que com, o uso do pronome *a gente*, a

frase passaria a ser expressa “a gente cantou”, sem qualquer problema de ambiguidade. Com isso, verbos no presente com mesma morfologia para pretérito perfeito e presente do indicativo constituem ambientes favoráveis ao uso do padrão emergente, uma vez que resolve uma questão de uso da língua. Resultados próximos aos nossos foram encontrados por Naro et al, (2014) e Benfica (2016), conforme exposto a tabela 47:

**Tabela 47:** Efeito do *tempo e tipo de paradigma verbal* sobre concordância verbal *nós* com *-mos* em amostras do português brasileiro

	<b>Nossa Pesquisa (Fortaleza)</b>	<b>Naro et al, 2014 (Baixada Cuiabana)</b>	<b>Benfica (2016) (Vitória)</b>
Presente diferente	52,9% (331/627) <b>-0.6989</b>	80,3% (106/132) <b>0,789</b>	95,9% (163/170) <b>0,508</b>
Presente igual	28,1% (95/340) <b>-1.5975</b>	23,8% <b>(20/84)</b> <b>0,199</b>	93,9% (31/33) <b>0,387</b>
Pret. Perf. Igual	49,1% (85/173) <b>-0.3506</b>	97,9% (47/48) <b>0,980</b>	98,3% (171/174) <b>0,690</b>
Pret. Perf. Diferente	66,4% (94/143) <b>0.1408</b>	96,3% (26/27) <b>0,963</b>	98,9% (89/90) <b>0,797</b>
Pret. imperfeito	26,1% (80/306) <b>-1.6614</b>	13,2% (17/129) <b>0,071</b>	31,5% (17/54) <b>0,009</b>
<b>TOTAL</b>	43,2% 686/1589	51,4% (216/420)	90,4% (471/521)

Fonte: Elaboração própria

Conforme apontam os resultados da tabela 47, a atuação do tempo presente com morfologia diferente para o pretérito perfeito e o presente do indicativo, em nossa amostra, diverge dos resultados encontrados nas pesquisas de Naro et. al (2014) e de Benfica (2016), tendo cada uma delas apresentado uma característica distinta para este fator. O resultado da análise dos dados de fala culta fortalezense apontou desfavorecimento da forma padrão *nós* com *-mos* neste contexto. Naro et. al (2015), com dados de fala da Baixada Cuiabana, por outro lado, observou esse contexto como expressivamente favorável ao uso da concordância verbal padrão *nós* com *-mos*, a marcação da desinência *-mos*, com peso relativo de 0,789. Por outro lado, Benfica (2016), com dados de fala de Vitória - ES, identificou que nesse contexto

comportamento praticamente neutro no condicionamento da forma de concordância padrão *nós* com *-mos*, com peso relativo igual a 0,508.

Em relação ao presente do indicativo com formas iguais para o pretérito perfeito e para o presente do indicativo, nosso resultado confirma os encontrados pelos autores. Ou seja, tanto na comunidade de fala de Fortaleza, quanto na Baixada Cuiabana e em Vitória, se observou expressivo desfavorecimento da estrutura padrão de concordância verbal *nós* com *-mos*, indicando ser esse um contexto expressamente favorável ao uso do padrão de concordância emergente *a gente* sem *-mos*. Os resultados desses estudos podem confirmar a hipótese de Scherre, Yacovenco e Naro (2018) sobre o tempo/modo verbal presente do indicativo com a mesma morfologia na relação presente/pretérito perfeito. De acordo com os autores, a ausência da marcação *-mos* nos verbos (característica do padrão emergente) atuaria nesse contexto como mecanismo para desfazer a ambiguidade temporal de verbos com mesma forma para pretérito perfeito e presente do indicativo, o que poderia favorecer o uso da estrutura *a gente* sem *-mos*. Isso se confirma nas três amostras (ver tabela 47).

Em relação ao pretérito perfeito com formas iguais para o pretérito e o presente do indicativo, o modelo de regressão logística apontou desfavorecimento da forma padrão de concordância *nós* com *-mos*, tendo apresentado estimativa de uso negativa (-0.3506), diferentemente dos estudos realizados em amostras de fala da Baixada Cuiabana e de Vitória, que apontaram forte favorecimentos da concordância verbal padrão com o pronome *nós*. No entanto, em nossos dados, o modelo de regressão logística revelou pouca interação entre este fator e a realização da variável resposta. O valor de *p* atribuído a esse fator é igual a 0.101852, estando fora do intervalo de confiança adotado nesta pesquisa e, por isso, esse resultado não possui confiabilidade estatisticamente significativa.

Quando se trata do pretérito perfeito com formas diferentes para o presente e para o passado, o modelo de regressão apontou forte favorecimento do padrão de concordância canônico *nós* com *-mos*, com estimativa de uso de 0.1408. Esse resultado também é confirmado por Naro et al (2014) que identificou esse contexto como fortemente associado à forma padrão de concordância (P.R. 0,963), e a Benfica (2016), que observou na fala de Vitória forte influência de verbos no pretérito com formas diferentes sobre a realização de *nós* com a marcação da desinência *-mos* (P.R. 0,797). Para Scherre, Yacovenco e Naro (2018), uma das questões que se destaca ao analisar o *tempo e o tipo de paradigma verbal do modo indicativo* é que a desinência *-mos* é preferencialmente usada como marca de pretérito perfeito, e isso é confirmado nas pesquisas sobre concordância verbal. De acordo com os autores, “Há um

processo de iconicidade: o tempo mais marcado, o pretérito perfeito, recebe preferencialmente a marca explícita, o -mos” (Scherre, Yacovenco e Naro, 2018, p. 23).

Por fim, em nosso modelo de regressão logística, que analisa a atuação das variáveis previsoras em um contexto multivariado, os verbos no pretérito imperfeito do modo indicativo são fortemente desfavoráveis a marcação da desinência plural *-mos*, sendo este contexto linguístico fortemente associado à utilização do padrão emergente de concordância verbal *a gente* sem *-mos*, que naturalmente apaga a desinência do verbal. Esse intenso desfavorecimento também foi observado em análise da Baixada Cuiabana, em que o padrão *nós* com *-mos* é intensamente desfavorecido (P.R. 0,071), e na cidade de Vitória, que apresenta o mesmo comportamento (P.R. 0,009). Resultados próximos aos nossos foram observados por Benfica (2016), que discorre:

A respeito do uso do pretérito imperfeito, é importante colocar que sua forma na 1PP é pouco frequente no português brasileiro em decorrência da forma alternante à 1PP *a gente* ser mais utilizada nesse tempo verbal, o qual apresenta a desinência de 3PP. Ou seja, é mais comum ouvirmos “*a gente brincava*” do que “*nós brincávamos*” (Benfica, 2016, p. 80)

Essa observação da autora também foi vista em nossa análise sobre a alternância pronominal (ver seção 5.1), em que encontramos forte favorecimento do pronome *a gente* com verbos no pretérito imperfeito. Uma das explicações para a forte associação entre esses verbos e a ausência da desinência *-mos* pode estar na tendência natural do português em reduzir as proparoxítonas. Scherre, Yacovenco e Naro (2018) também observaram essa tendência em seu trabalho. De acordo com os linguistas “a baixa saliência (*falava/falávamos*) e a esquiva da proparoxítona com o morfema de plural *-mos* (*falávamos*), motivada pelo padrão fonológico preferencialmente paroxítono do português brasileiro, provocam a ausência de *-mos*” (Scherre, Yacovenco e Naro, 2018, p. 21). Essa razão poderia explicar o forte desfavorecimento do padrão de concordância *nós* com *-mos* em nossos dados.

#### *e) Atuação da conjugação verbal sobre os padrões de concordância verbal*

A variável conjugação verbal é considerada em estudos sociolinguísticos sobre a concordância verbal de primeira pessoa do plural, como Zilles e Batista (2006), Zilles, Maya e Silva (2000), Agostinho (2013) e Silvano (2016), para citar alguns. Zilles e Batista (2006), por exemplo, em análise sobre a concordância verbal na fala culta de Porto Alegre, observaram

maior tendência ao apagamento da desinência número pessoal *-mos* em verbos da terceira conjugação, enquanto verbos de segunda conjugação se mostraram mais favoráveis ao uso da desinência *-mos*. Já em Zilles, Maya e Silva (2000), se encontra maior tendência à marcação da desinência *-mos* em verbos de primeira e segunda conjugação. Com o objetivo de testar essas tendências em nossos dados, incluímos essa variável em nosso modelo de regressão logística.

A análise multivariada (ver tabela 46) revela que, em contextos de primeira conjugação (*intercept*), há desfavorecimento da forma de concordância padrão *nós* com *-mos*. Vale destacar que a formação do *intercept* leva em conta outros contextos linguísticos e sociais que interferem diretamente na atuação individual dos verbos da primeira conjugação, porém o resultado se mostra estatisticamente confiável ( $p\text{-value} = 0.000867$ ). Esse resultado difere do encontrado por Zilles e Batista (2006) e de Zilles, Maya e Silva (2000), que constataram os verbos na primeira conjugação como aliados da manutenção da desinência *-mos* (P.R. 0,44).

Por outro lado, nosso modelo de regressão apontou os verbos de segunda conjugação como favoráveis à estrutura padrão de concordância verbal *nós* com *-mos*, resultado que está em conformidade com o encontrado em Zilles, Maya e Silva (2000). Em nossa análise, os verbos de segunda conjugação, em relação ao *intercept*, parecem favorecer expressamente o uso do padrão canônico de concordância, com estimativa de uso igual a 0,1611. O valor positivo indica favorecimento da variante de referência testada no modelo (*nós* com *-mos*). Uma das hipóteses que explicaria esse resultado é a alta frequência de ocorrências do verbo do tipo *ter* associado ao pronome *nós*. Sendo estes verbos pertencentes à segunda conjugação, estariam atuando estatisticamente no favorecimento do padrão *nós* com *-mos*. Observe a distribuição do *tipo de verbo* sobre as variantes:

**Tabela 48:** Distribuição de *nós* com *-mos* e *a gente* sem *-mos* por tipo de verbo

<i>Tipo de verbo</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%
<b>Ação</b>	248	29,9%	581	70,1%
<b>Estado</b>	85	85,2%	61	41,8%
<b>Dicendi</b>	39	30%	91	70%
<b>Epistêmico</b>	32	33,3%	64	66,7%
<b>Ter</b>	282	72,7%	106	27,3%

Fonte: Elaboração própria

Conforme é possível ver na tabela 48, os verbos do tipo *ter*, isoladamente, representam uma grande proporção do total de verbos em nossa amostra. A maioria desses verbos está associada a estrutura de concordância canônica *nós* com *-mos*, 282 ocorrências, o

que equivale a 72,7% do total de dados de verbo ter. Essa alta proporção de uso, sendo os verbos do tipo ter pertencentes à segunda conjugação, pode estar exercendo pressão estatística para o favorecimento da estrutura padrão em nossa amostra. Além disso, o modelo de regressão forneceu um valor de  $p$  igual a 0.004814, estando dentro do intervalo de confiança, apontando que o resultado possui confiabilidade estatística.

Fenômeno parecido foi observado por Zilles, Maya e Silva (2000), ao identificarem forte associação entre os verbos da terceira conjugação e o apagamento da desinência número pessoal. De acordo com os autores “O peso relativo alto, de 0,60, para a terceira conjugação (280 ocorrências) possivelmente está sendo influenciado pela grande incidência do verbo ir (são 134/280 ocorrências deste verbo, das quais 114 são casos de vamos)” (Zilles, Maya e Silva, 2000, p. 213). No entanto, não podemos testar essa hipótese sobre os verbos de terceira conjugação em nossos dados, tendo em vistas que o modelo apontou esse fator como não estatisticamente confiável ( $p$ -value = 0.339052).

*f) Atuação da estrutura do verbo sobre os padrões de concordância verbal*

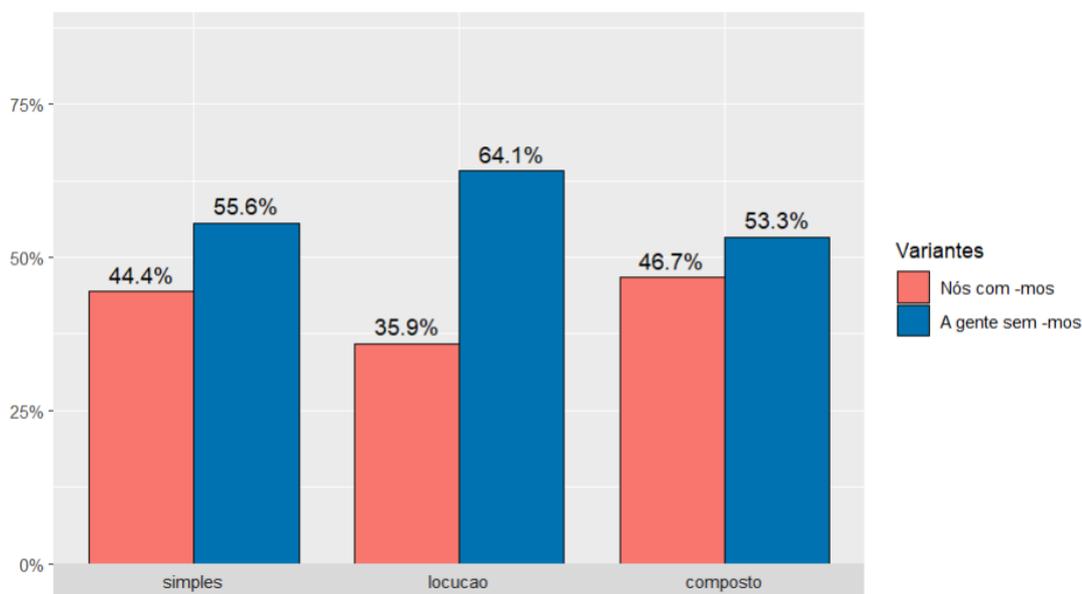
Consideramos observar esta variável previsora por identificar, em nossa amostra, grande número de ocorrências de locuções verbais associadas ao padrão emergente de concordância *a gente* sem *-mos*, quase o dobro quando comparada ao número de ocorrências do padrão canônico. Essa característica poderia apontar um ambiente mais favorável à expansão da estrutura emergente na língua, que seria um achado importante em nossa pesquisa. Ocorrências como a ilustrada em (194) é um exemplo frequente em nossa amostra:

(194) Inf.: parece que você já coNHEce... TUdo você já se sente mais tranqüila... lá no Rio quando *a gente vai SAIR*... é parece porque a gente num TEM aquele... conhecimento... com as rua diREItto nem nada né? (Inq. 32 – F1 – DID, mulher, 51 anos)

A fim de confirmar a hipótese alternativa de que existe uma interação estatisticamente significativa entre a estrutura do verbo e a realização dos padrões de concordância verbal, incluímos este preditor no modelo de regressão. Em relação aos verbos simples (*intercept*), esses agem de modo a inibir o uso da concordância verbal padrão, sendo, nos mesmos contextos que compõem o *intercept*, aliados do padrão emergente *a gente* sem *-mos*. Esse resultado também era esperado, uma vez que os verbos simples são maioria em nossa

amostra e, como maioria, ocorrem em maior proporção junto ao pronome *a gente*, conforme demonstra o gráfico 34:

**Gráfico 34:** Distribuição de *nós* com *-mos* e *a gente* sem *-mos* por estrutura do verbo



Fonte: Elaboração própria

Conforme observado no gráfico 34, a estrutura emergente de concordância tem maior proporção de uso em todas as estruturas verbais, o que pode estar colaborando para o seu favorecimento no modelo multivariado. Em relação às locuções verbais, tanto em termos de proporção de uso, quanto em termos de favorecimento, observamos tendência a inibir intensamente o uso do padrão de concordância canônico *nós* com *-mos* em nossa amostra, com estimativa de uso de -1,3308. Essa estimativa de uso bem acima do valor de *intercept* pode indicar que verbos com esta estrutura são ambientes expressivamente favoráveis ao uso do padrão emergente de concordância *a gente* sem *-mos*, o que é confirmado pela proporção de uso, conforme expõe o gráfico 34. Além disso, é importante destacar que o modelo atribuiu a esse resultado um valor de *p* igual a 0.000212, indicando ser um resultado estatisticamente confiável para a análise.

Por fim, em relação aos verbos compostos, também conhecidos como tempo composto, ou seja, aqueles formados pelo verbo *ter* ou *haver* mais o particípio do verbo principal (*a gente* tinha saído), estes não demonstraram relevância estatística no modelo de regressão desenvolvido para tratar os dados de nossa amostra, estando o valor de *p* atribuído a sua estimativa fora do intervalo de confiança (*p-value* = 0.607701), e por isso não consideramos o resultado desse fator estatisticamente significativo.

### 6.1.3 Síntese dos resultados – A concordância verbal na Amostra 1990

A amostra representativa da fala culta fortalezense da década de 1990 revelou 1938 observações de concordância verbal com a primeira pessoa do plural, com as ocorrências concentradas na estrutura emergente *a gente* sem *-mos* (57,8%), seguido de *nós* com *-mos* (41,4%). Realizações de concordância não padrão são raras em nossa amostra, distribuídas da seguinte forma: *nós* sem *-mos* (0,7%) e *a gente* com *-mos* (0,1%). Devido ao baixo número de ocorrências, foi necessário a retirada dos dados de concordância não padrão para a análise de um modelo multivariado em um recorte binário.

Com o recorte binário realizado, nos restaram um total de 1586 dados de concordância, sendo 56,7% ocorrências de *a gente* sem *-mos* e 43,3% de *nós* com *-mos*. Esses dados foram submetidos a análise através de um modelo de regressão que, inicialmente, não convergiu corretamente devido às células vazias presentes na variável *tonicidade do verbo*. Após a remoção dessa variável, o modelo logístico convergiu e apontou relevância estatística de todas as variáveis predictoras observadas.

Em relação ao *sexo/gênero*, as mulheres apresentaram tendência a desfavorecer mais intensamente o padrão de concordância *nós* com *-mos* em contexto multivariado. Na mesma direção, os mais jovens apresentaram forte tendência a inibir padrão de concordância verbal mais antigo na língua, assim como os falantes de 35 a 50 anos, que tendem a desfavorecer a concordância de *nós* com *-mos*. Diferente dos informantes com mais de 50 anos da amostra representativa da década de 1990, que se mostraram favoráveis ao uso de *nós* com *-mos*. O *tipo de inquérito*, da mesma maneira, se mostrou relevante, no sentido de os inquéritos do tipo D2, os menos formais, se apresentarem como favoráveis ao padrão emergente *a gente* com *-mos*, enquanto os inquéritos mais formais (EF) são expressamente aliados da forma canônica de concordância *nós* com *-mos*.

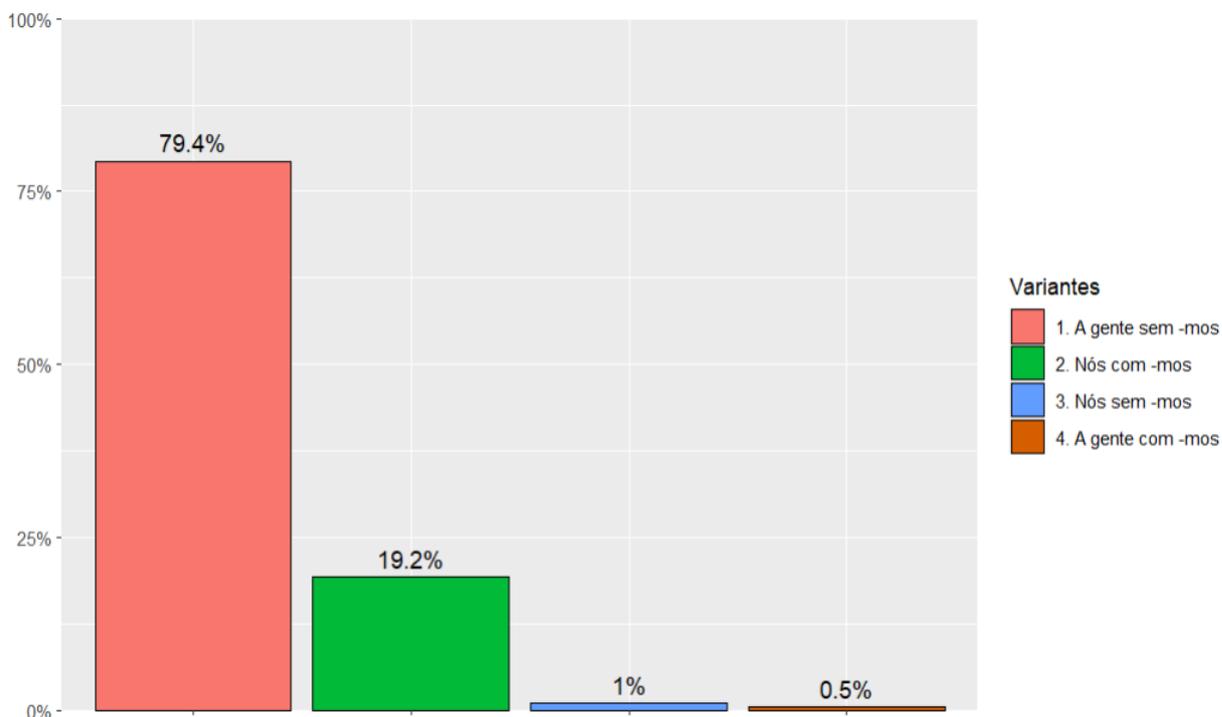
Em relação ao *tempo e tipo de paradigma verbal do modo indicativo*, percebe-se forte favorecimento de *a gente* sem *-mos* em usos com verbos que apresentam a mesma forma para o pretérito perfeito e para o presente do indicativo, especialmente no presente. Já os usos da concordância junto a verbos no pretérito perfeito com formas diferentes para o pretérito e o presente do indicativo tendem a ser favorável ao uso de *nós* com *-mos*. Por outro lado, verbos no pretérito imperfeito são forte inibidores do padrão de concordância canônico *nós* com *-mos*, se revelando como o contexto verbal mais receptível, em termos de favorecimento, ao uso da estrutura *a gente* sem *-mos*.

Ao analisar a *conjugação verbal*, percebe-se desfavorecimento da estrutura padrão de concordância com verbos da primeira conjugação e leve inclinação ao padrão canônico em verbos da segunda conjugação. Verbos da terceira conjugação não apresentaram efeito significativo. Por fim, em relação a *estrutura verbal*, verbos simples e locuções verbais demonstraram forte tendência ao favorecimento da estrutura emergente *a gente* sem *-mos*, enquanto os verbos compostos não apresentaram interação significativa no modelo de regressão logística.

## **6.2 A concordância verbal de primeira pessoa do plural no falar culto de fortaleza: análise descritiva dos dados de fala da década de 2020**

Nesta seção, apresentamos a análise descritiva dos dados de concordância verbal referentes a amostra de fala culta fortalezense da década de 2020. Com objetivo de ser mais dinâmico na apresentação dos resultados, enfocaremos na descrição dos dados e nos resultados dos testes estatísticos realizados.

Dito isto, após submeter os dados de concordância verbal à análise estatística por meio do ambiente *R Studio*, constatamos um total 2596 observações de concordância verbal com os pronomes de primeira pessoa do plural. Desse total de dados, 2061 se referem ao uso do padrão emergente de concordância *a gente* sem *-mos*, sendo o mais recorrente padrão de concordância empregado na Amostra 2020. Em seguida, observamos 498 ocorrências de padrão canônico de concordância *nós* com *-mos*, sendo a segunda variante mais utilizada na nossa amostra. Dados de concordância não padrão, tanto relacionados ao pronome *nós* quanto ao *a gente* são pouco recorrentes em nossa amostra e difíceis de serem flagrados de modo geral (Scherre e Naro, 2016). Em dados representativos da década de 2020, a estrutura não padrão *nós* sem *-mos* (*nós canta*) se realizou em apenas 25 ocorrências, enquanto a variante *a gente* com *-mos* (*a gente cantamos*) obteve um total de 12 dados. O gráfico 35 ilustra a distribuição de cada padrão de concordância em nossa Amostra 2020:

**Gráfico 35:** Distribuição dos padrões de concordância verbal: Amostra 2020

Fonte: elaboração própria

Conforme observado no gráfico 35, o padrão emergente de concordância é expressamente preferido na comunidade de fala, sendo responsável por 79,4% das realizações de concordância na amostra. Essa tendência está associada ao aumento na proporção de uso do pronome *a gente*, que se implementou e se expandiu em todos os contextos de uso, conforme demonstrou análise da alternância pronominal (ver seção 5.2). Em seguida, ainda disputando lugar como padrão de concordância na comunidade de fala, temos a estrutura canônica, formada pelo pronome *nós* acrescido da desinência de plural *-mos*, com 19,2% dos dados. Mesmo tendo proporção de uso expressiva menor que o padrão emergente, o fenômeno ainda se caracteriza como uma regra variável nos termos de Labov (1977). Os dados de concordância não padrão com o pronome *nós* somam apenas 1% das ocorrências, enquanto a estrutura *a gente* com *-mos* representa 0,5%. A mesma tendência na distribuição das variáveis foi observada em outros estudos, conforme apresenta a tabela 49:

**Tabela 49:** Comparação do resultado de nosso estudo com pesquisas sociolinguísticas

ESTUDO	<i>A GENTE SEM -MOS</i>	<i>NÓS COM -MOS</i>	<i>NÓS SEM -MOS</i>	<i>A GENTE COM -MOS</i>
<b>Nossa pesquisa Fala - Fortaleza</b>	<b>79,4%</b>	<b>19,2%</b>	<b>1%</b>	<b>0,5%</b>
Itabaiana/SE (Fretaj, 2016)	82%	17%	0,5%	0,5%
Aracaju/SE (Freitaj et al, 2016)	83%	15%	1,5%	0,5%
Fortaleza/CE (Carvalho e Santos, 2023)	68,4%	22,7%	8,3%	0,5%

Fonte: elaboração própria

O que se percebe ao analisar a tabela 49 é um padrão na distribuição dos padrões de concordância verbal com a primeira pessoa do plural nas diversas comunidades de fala. Tanto os resultados de Sergipe, com dados da década de 2010, quanto os de Fortaleza, com dados da década de 1980 (Carvalho e Santos, 20223) e da década de 2020 (nosso estudo), demonstram maior preferência pelo padrão de concordância emergente na língua *a gente sem -mos*, seguido do padrão canônico, sendo este bem menos expressivos. Além do mais, os dados de ambas as amostras apresentados na tabela 49 demonstram a concordância não padrão ser pouco recorrente na língua, principalmente a estrutura *a gente com -mos* (nós cantamos).

Após observar a distribuição geral das variantes em nossa Amostra 2020 e discutir seus resultados, a próxima seção será destinada à apresentação dos resultados referentes à atuação de cada variável previsoras sobre a realização da variável resposta. Sendo assim, apresentaremos as frequências e proporções de uso de cada variáveis, além do resultado dos testes estatísticos de qui-quadrado. Destacamos que, para a análise dos dados da Amostra 2020, observamos as mesmas variáveis que foram testadas nos dados da Amostra 1990.

### ***6.2.1 Atuação das variáveis previsoras em termos de frequência e proporção de uso: a concordância verbal de primeira pessoa do plural na Amostra 2020***

Com o objetivo de analisar a atuação de cada uma das variáveis previsoras elencadas para este estudo sobre a concordância verbal com os pronomes *nós* e *a gente*, observamos a distribuição das variantes de concordância em função de cada variável previsoras de natureza extralinguística e linguística, além de conduzir teste estatístico que visa identificar

o grau de interação que cada um dos preditores testados possuem sobre a realização dos padrões de concordâncias analisados neste estudo.

Os dados de fala culta fortalezense representativos da década de 2020 foram submetidos a análise através da linguagem de programação *R*, que, por meio do *software R Studio*, forneceu proporções de uso e resultados dos testes realizados sobre as variáveis. Dentre as variáveis de natureza social testada nesta pesquisa, destaca-se a variável *sexo/gênero*, como pode ser conferido na tabela 50:

**Tabela 50:** Distribuição das proporções de uso da concordância por *sexo/gênero*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Sexo/Gênero</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Homem</b>	195	18,1%	855	79,5%	18	1,7%	7	0,7%
<b>Mulher</b>	303	19,9%	1202	79,3%	7	0,5%	5	0,3%
$\chi^2 (3) = 12.071, p = 0.007144$								

Fonte: elaboração própria.

Nos dados representativos da década de 2020, percebe-se comportamento semelhante no uso dos padrões de concordância verbal entre homens e mulheres. Em ambos os sexos, conforme os dados apresentados na tabela 50, há maior preferência pelo uso do padrão emergente *a gente sem -mos*, sendo realizado em 855 ocorrências entre os homens (79,5%), e em 1202 ocorrências entre as mulheres (79,3%), representando a maioria das realizações em termos de frequências de uso. Esse resultado pode ser dado em função da ampla expansão do pronome *a gente* na década de 2020, como comprovou a análise da alternância pronominal (ver seção 5.2).

Ocupando o segundo lugar na preferência dos informantes, tanto entre os homens quanto entre as mulheres, está o padrão canônico *nós com -com*. Neste caso, a proporção de uso desta estrutura é levemente mais intensa nas informantes mulheres, com 303 dados, o que corresponde a 19,9% das realizações de concordância verbal. Já para os homens, observamos 195 dados de *nós com -mos*, sendo responsável por 18,1% dos padrões de concordância verbal em nossa amostra. Destacamos que a diferença no emprego das formas de concordância verbal entre homens e mulheres é pequena, representado apenas 1,7 pontos percentuais.

Casos de concordância não padrão, como já explicado, são menos frequentes na Amostra 2020. No caso da estrutura *nós sem -mos* (*nós fala*), foram encontrados 18 dados entre os homens (1,7%), enquanto entre as mulheres foram flagrados 7 ocorrências com essa estrutura (0,5%). Por outro lado, a estrutura não padrão *a gente com -mos* (*a gente falamos*) é

ainda menos recorrente, tendo 7 ocorrências entre os homens (0,7%), e apenas 5 realizações entre as mulheres (0,3%), demonstrando que, em termos de proporção de uso, os comportamentos linguísticos de homens e mulheres se assemelham.

Além da variável *sexo/gênero*, outra variável social que tem se destacado em estudos sociolinguísticos é a variável *faixa etária*. Essa variável foi relevante na análise sobre a alternância pronominal entre *nós* e *a gente*, e acreditamos que, para a análise dos padrões de concordância de primeira pessoa do plural, esse preditor se mostrará relevante. A tabela abaixo apresenta a distribuição das variantes do fenômeno sociolinguístico da concordância verbal em função das faixas etárias:

**Tabela 5:** Distribuição das proporções de uso da concordância por *faixa etária*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Faixa etária</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>22 a 35 anos</b>	75	11,2%	589	88,3%	1	0,2%	2	0,3%
<b>36 a 50 anos</b>	110	13%	732	86,2%	6	0,7%	1	0,1%
<b>51 anos ou +</b>	313	29%	740	68,5%	18	16,7%	9	0,8%
$\chi^2 (6) = 134, p < 0,001$								

Fonte: elaboração própria.

A análise da distribuição das ocorrências em função das faixas etárias controladas neste estudo revela predominância do padrão emergente de concordância *a gente sem -mos* em todas as faixas etárias analisadas, especialmente entre os informantes mais jovens da amostra, de 22 a 35 anos, onde atinge proporção de uso de 88,3% das realizações, com 589 ocorrências de concordância. Entre os informantes da faixa etária de 36 a 50 anos, o comportamento do fenômeno se assemelha ao observado entre os mais jovens, ou seja, nota-se alta proporção de uso da forma estrutura de concordância emergente, com 732 ocorrências, o que equivale a 86,2% dos dados analisados. Já entre os mais velhos da amostra, embora os números os indiquem como favoráveis a concordância emergente, atingiu o menor índice entre as três faixas pesquisadas (68,5%).

Em contrapartida, a estrutura de concordância padrão *nós com -mos* é usada com menos frequência à medida que a idade do informante diminui. Isso pode ser um indicativo de mudança linguística nos termos de Labov (1974), tendo em vista que os mais jovens tendem a impulsionar as formas inovadoras e promover com mais intensidade as mudanças linguísticas. Entre os informantes de 22 a 35 anos, encontramos 76 ocorrências de *nós com -mos*, o que equivale a 11,2% das ocorrências de concordância verbal na Amostra 2020. Já para os

informantes de 36 a 50 anos, há uma elevação tanto na frequência quanto na proporção de uso do padrão de concordância canônico, sendo observado em 110, ou seja, em 13% do total de dados. Por fim, informantes com mais de 50 anos de idade atingiram o maior índice de utilização da concordância padrão, tendo realizado a estrutura padrão *nós* com *-mos* em 313 ocorrências, sendo responsáveis por 29% das realizações. Embora a alta proporção de uso do padrão emergente, esses dados estão em conformidade com a teoria laboviana, uma vez que as gerações mais velhas apresentam tendência mais acentuada a seguir as normas mais tradicionais da língua, enquanto os jovens tendem a ser mais inovadores.

As variantes de concordância não padrão são pouco recorrentes na amostra. Conforme é possível observar nos dados apresentados na tabela 50, a forma de concordância *nós* sem *-mos*, por exemplo, obteve 0,2% dos dados entre os informantes de 22 a 35 anos; 0,7% entre os de 36 a 50 anos e 1,7% entre os informantes com mais de 50 anos. De forma parecida, o padrão *a gente* com *-mos* é pouco frequente, sendo realizado apenas em 0,3% dos dados entre os mais jovens, 0,1% entre informantes de 36 a 50 anos e em 0,8% das ocorrências entre os informantes com mais de 50 anos de idade.

Além disso, o teste de qui quadrado de Pearson realizado para verificar o grau de interação da variável *faixa etária* sobre a realização dos padrões de concordância verbal apontou forte associação entre as variáveis. O resultado do teste apontou o valor de  $\chi^2 = 134$ , com seis graus de liberdade e valor de  $p < 0,001$ . Esse resultado é suficiente para rejeitar a hipótese nula de que não há diferença no uso dos padrões de concordância a depender da faixa etária do informante, e acatar a hipótese alternativa de que a faixa etária condiciona significativamente a escolha entre o padrão de concordância usado.

A próxima variável de ordem extralinguística observada nesta pesquisa trata-se do *tipo de inquérito*. Como já dito anteriormente, estava variável previsora não caracteriza qualquer trato social do informante, mas apenas qualifica a modalidade de gravação das entrevistas quanto ao grau de formalidade e ao nível de monitoramento empregado na fala, acreditando serem fatores que condicionam o uso das variantes de concordância verbal de primeira pessoa do plural, tendo em vista que é um fenômeno que envolve estigma. Os resultados referentes a distribuição das ocorrências de concordância verbal em função do *tipo de inquérito*, bem como o resultado do teste de interação de qui-quadrado, podem ser conferidos na tabela 51:

**Tabela 51:** Distribuição das proporções da concordância verbal por *tipo de inquérito*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Inquérito</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>D2</b>	60	7%	777	90,8%	14	1,6%	5	0,6%
<b>DID</b>	217	20,3%	842	78,6%	7	0,7%	5	0,5%
<b>EF</b>	221	33%	442	66,1%	4	0,6%	2	0,3%
<b><math>\chi^2 (6) = 169.51, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

Como explicado, os inquéritos do tipo D2 (diálogo entre dois informantes), apresentam o menor grau de formalidade entre os inquéritos analisados. Já os inquéritos do tipo DID (diálogo entre informante e documentador) possuem grau intermediário de formalidade. Por fim, os inquéritos do tipo EF (elocução formal) são os mais formais entre os inquéritos analisados (Araujo, 2011).

Nos dados representativos da década de 2020, o padrão emergente *a gente sem -mos*, considerado sem estigma (Bechara, 1999), se destacou em todos os tipos de inquéritos, mais intensamente entre os inquéritos do tipo D2, com 777 ocorrências e um total quase categórico de 90,8% dos dados. Em inquéritos do tipo DID, com grau intermediário de formalidade, também há predominância da forma *a gente sem -mos*, tendo obtido 842 ocorrências, o que representa um total de 78,6% dos dados. Por fim, de forma menos intensa em inquéritos do tipo EF, porém ainda bastante expressiva, observamos 442 realizações do padrão de concordância emergente, o que equivale a 66,1% dos dados.

De acordo com Labov (1966), o grau de formalidade está diretamente relacionado ao contexto comunicativo no momento da elocução. Ou seja, em situações mais formais, os informantes naturalmente apresentam tendência ao uso de variantes mais conservadoras na língua. Em situações como um discurso público, uma aula ou palestra, que são ambientes de gravação dos inquéritos do tipo EF, os informantes costumam adotar formas linguísticas que são socialmente valorizadas. Nos dados da década de 2020, o pronome *a gente* se expandiu com tanta intensidade, que até mesmo em ambientes mais formais passou a ser adotado como forma preferencial de primeira pessoa do plural, levando a criar um padrão de concordância emergente, sem estigma, igualmente adotado (*a gente sem -mos*). No entanto, comparando a atuação desse padrão emergente, percebe-se, ainda, que os inquéritos mais formais (EF) possuem uma certa resistência ao uso da forma não canônica, pois, mesmo apresentando alta proporção de uso desse padrão emergente, os inquéritos do tipo EF são os que apresentam menor proporção de uso entre os três investigados.

Em relação ao uso da concordância padrão *nós* com *-mos*, esse é pouco recorrente em inquéritos do tipo D2, tendo sido observado em apenas 60 ocorrências, o que corresponde a 7% dos dados (ver tabela 51). Labov (1966), introduz o conceito de monitoramento da fala para caracterizar a atenção que o falante dá à fala no momento da elocução. Para o autor, em contextos menos formais e descontraídos há baixo monitoramento da fala, levando os informantes a ficarem mais relaxados, levando-os a uma tendência maior a aderir variantes não padrão. Geralmente, inquéritos do tipo D2 são constituídos por gravações de pessoas que já se conhecem e possuem algum tipo de intimidade, como amigos, familiares, colegas. Sendo assim, é de se esperar que a variante canônica ocorra em menor proporção neste tipo de inquérito.

Já em inquéritos do tipo DID, em que há grau intermediário de formalidade, há um leve aumento na proporção de uso do padrão de concordância canônico, tendo sido realizado em 217 dados, o que equivale a um total de 20,3% das realizações de concordância verbal em nossa amostra. Por fim, inquéritos do tipo EF apontaram maior recorrência da concordância padrão *nós* com *-mos* entre os três tipos analisados, tendo 221 ocorrências, o que corresponde a um total de 33% dos dados. Esse resultado também está de acordo com a teoria laboviana, uma vez que inquéritos do tipo EF apresentam alto grau de monitoramento da fala, o que condiciona o uso de variantes tidas como padrão (Labov, 1966).

Em relação aos dados de concordância não padrão, a estrutura *nós* sem *-mos* obteve 14 realizações em inquéritos do tipo DID (1,6%), 7 ocorrências em registros do tipo DID (0,7%), e apenas 4 dados em inquéritos do tipo EF (0,6%), se mostrando pouco presente na amostra. Isso é característica dos estudos sobre concordância verbal no Brasil, tendo em vista que dados de concordância não padrão são difíceis de serem flagrados na língua falada (Oushiro, 2015). Em relação a estrutura não padrão *a gente* com *-mos*, observamos 5 ocorrências em registros do tipo D2 (0,6%), também 5 realizações em inquéritos do tipo DID (0,5%), e somente 2 observações em registros do tipo EF (0,3%).

Além das variáveis de natureza extralinguística, o fenômeno variável da concordância verbal pode ser condicionado por variáveis previsoras de ordem linguística. Dentre as variáveis que se mostraram relevantes para explicar os padrões de concordância verbal de primeira pessoa do plural para Benfica (2016), Naro et al (2014), Scherre, Yacovenco, Naro (2018), Carvalho e Santos (2023), está o *tempo e tipo de paradigma verbal*. A hipótese geral é a de que o uso do padrão emergente atuaria na questão da ambiguidade temporal de verbos com as mesmas formas para o presente e para o passado. Aqui, apresentamos todos os tempos verbais observados, para ter uma visão geral da distribuição das variantes. Mais à

frente, no modelo de regressão logística, analisamos apenas dados do modo indicativo. A tabela 52 apresenta os dados de concordância verbal em função desta variável previsora *tempo e tipo de paradigma verbal*:

**Tabela 52:** Proporções da concordância verbal por *tempo e tipo de paradigma verbal*

**PORCUFORT II - Anos 2020 (N = 2596)**

<i>Paradigma</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Presente igual</b>	38	8,6%	406	91,4%	0	0%	0	0%
<b>Presente Diferente</b>	191	29,1%	459	69,9%	3	0,5%	4	0,6%
<b>Pret. Perf. Igual</b>	91	30,6%	203	68,4%	0	0%	3	1%
<b>Pret. Perf. Diferente</b>	74	34,1%	140	64,5%	1	0,5%	2	0,9%
<b>Pret. imperfeito</b>	37	5,7%	585	91,3%	17	2,6%	2	0,3%
<b>Futuro</b>	59	36,4%	102	63%	0	0%	1	0,6%
<b>Presente do subjuntivo</b>	2	8,3%	22	91,7%	0	0%	0	0%
<b>Imperfeito do subjuntivo</b>	3	20%	12	80%	0	0%	0	0%
<b>Futuro do subjuntivo</b>	0	0%	17	100%	0	0%	0	0%
<b>Futuro do pretérito</b>	1	8,3%	11	91,7%	0	0%	0	0%
<b>Infinitivo</b>	2	1,8%	104	94,5%	4	3,6%	0	0%
<b><math>\chi^2 (33) = 312.96, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

Os resultados apresentados na tabela 52 seguem a tendência dos resultados observados na análise geral dos dados, sendo o padrão emergente expressivamente mais recorrente na década de 2020. Nesse sentido, em todos os contextos verbais que constituem essa variável previsora, o padrão emergente de concordância verbal *a gente sem -mos* é utilizado em maior proporção, sendo categórico em alguns casos (futuro do subjuntivo). Neste ponto, já é possível identificar que os resultados vão caminhar para expressivo favorecimento deste padrão de concordância em detrimento dos outros.

Em relação aos verbos no presente com mesma forma para o pretérito perfeito e para o presente do indicativo, o padrão de concordância *a gente sem -mos* é expressivamente mais utilizado, sendo observado em 406 ocorrências, o que representa 91,4% do total de dados. Diferente do padrão de concordância canônico, que ocorreu apenas em 38 dados, o que representa 8,65% do total. Esse resultado está em conformidade com o hipotetizado por Scherre, Yacovenco e Naro (2018), tendo em vista que a estrutura *nós com -mos* apresenta a mesma forma para o presente e para o passado, o padrão emergente atuaria no sentido de desfazer essa

ambiguidade. O excerto (195), por exemplo, extraído de nossa fonte de dados, não fica claro formalmente que se trata de uma ocorrência no presente do indicativo, diferente de excerto (196), que com o uso do padrão emergente, se faz emergir com mais clareza o tempo verbal.

- (195) Inf.: Não... **nós compramos** por lote... esse lote tem uma quantidade TAL... aí a gente a/adequa a produção... só que aí quando o dólar SObe... essa mercadoria que a gente vai comprar FOra... encaREce... (Inq. 03 – F2 – DID, homem, 26 anos)
- (196) Inf.: .... e aí você escolhe o seu fornecedor... professora e como é **a gente compra** de um e de outro (ali)? não a gente orienta o compraDOR... aí ele fecha a compra com UM fornecedor... como? pedindo pro outro fornecedor... (Inq. 53 – F2 – EF, mulher, 61 anos)

Conforme evidenciam os exemplos (195) e (196), o uso da estrutura emergente de concordância, criada a partir da inserção do pronome *a gente* na língua portuguesa, é responsável por destacar o tempo expresso pelo verbo na própria estrutura do verbal, diferente do padrão canônico de concordância, que exige que se recorra a elementos do contexto para inferir o tempo verbal. Com isso, o uso da estrutura inovadora auxiliaria no sentido de atuar sobre a questão da ambiguidade temporal causada por esses verbos. Não se observou ocorrência do padrão de concordância *a gente* com *-mos* neste contexto.

Em contextos de tempo presente com morfologia diferente para o pretérito perfeito e o presente do indicativo, a estrutura emergente continua sendo predominante, mesmo que sua inserção nesses ambientes não tenha a função de evitar a ambiguidade temporal. Esse resultado se deve a forte expansão da forma *a gente* sem *-mos* em diversos contextos sociais e linguísticos na década de 2020. De acordo com a tabela 52, a estrutura emergente foi observada neste contexto em 459 ocorrências (69,9%). Em contrapartida, a estrutura padrão de concordância, formada pelo pronome *nós* e a desinência *-mos*, se realizou em 191 dados, o que corresponde a 29,1% do total. As ocorrências (197) e (198) ilustram ocorrências de *nós* com *-mos* e *a gente* sem *-mos* neste contexto:

- (197) Inf.: quando vocês ouvirem falar em cri-ti-cis-mo ... é ... a filosofia do Kant racionalismo **nós temos** vários:: Decart Espinosa empirismo **nós temos** vários né? o Lock o Bacon o Riumer com parêntesis né? (Inq. 79 – F2 – EF, mulher, 36 anos)
- (198) Inf.: tem um momento certo de me pagar:: o meu salá::rio meu décimo tercei::ro as minhas férias... e por aí vai... por isso que **a gente diz** que é um

contrato de TRAtO sucessivo... porque há uma/uma REciprocidade de obrigações (Inq. 66 – F2 – EF, homem, 49 anos)

Os dados de concordância não padrão, como já dito sucessivas vezes, são muito escassos na amostra representativa da década de 2020. Após levantamento, observamos apenas 3 ocorrências de *nós* sem *-mos*, que aparece ilustrada no excerto (199), o que corresponde a apenas 0,5% dos dados de tempo presente com formas diferentes para o passado e para o presente. Por outro lado, a estrutura *a gente* com *-mos*, ilustrada no excerto (200), também foi pouco recorrente em nossos dados neste contexto, obtendo apenas 4 realizações, o que equivale apenas a 0,6%:

(199) Inf.: eu também queria MUIto ser ser uma:: uma empreSÁria mas no momento ainda não... *nós* não *tem* condições... é pode ser um sonho futuro (Inq. 04 – F2 – DID, mulher, 31 anos)

(200) Inf.: no início quando nós éramos menores nós brigávamos muito... coisa de crian::ça mas hoje em dia no caso nós viramos melhores amigas... no caso *a gente somos* muito unidas até porque... esto::u na minha casa... (Inq. 07 – F2 – DID, mulher, 24 anos)

Os verbos com pretérito perfeito com morfologia igual para o presente e o passado também são verbos que possuem ambiguidade temporal, razão pela qual acreditamos que este contexto beneficiaria o uso da estrutura *a gente* sem *-mos*, pois atuaria no sentido resolver a ambiguidade potencial desses verbos. Como esperado, este contexto verbal é expressivamente favorável ao padrão emergente na língua, sendo observado em 203 dados de concordância verbal (68,4%). A estrutura de concordância padrão *nós* com *-mos*, embora seja empregada em menor proporção, ainda disputa a preferência pelo padrão de concordância adotado pelos informantes cultos de Fortaleza na década de 2020, com 91 observações (30,6%). Ambientes verbais com a mesma morfologia para ambos os tempos seriam mais propensos a entrada do pronome *a gente*, uma vez que o padrão de concordância canônico não marca o tempo verbal, enquanto o padrão emergente, criado pela inserção do pronome *a gente* no quadro pronominal do português, esclarece a questão do tempo verbal expresso pelo verbo. Os excertos (201) e (202) são exemplos, respectivamente, de ocorrências do padrão de concordância canônico e do padrão emergente, e como o uso desses padrões alteram a ambiguidade potencial que alguns verbos possuem:

- (198) Inf.: a cidade não é violenta eu tô falando que há um exaGero... principalmente pela aquela parte política que **nós falamos**... falamos no início de tenTAR... éh:... de qualquer forma... pra fazer com que a gente pense que a Única solução pra violência seja o armamento (Inq. 37 – F2 – DID, homem, 26 anos)
- (199) Inf.: AÍ eu vou só:: abrir a parte da logística de suprimentos... pra gente concluir aquele assunto porque **a gente falou** de estoque MÍ::nimo lembra? né... estoque mínimo estoque máximo fizemos exercícios... (Inq. 53 – F2 – EF, mulher, 61 anos)

Conforme é possível observar no exemplo (198), ao nos depararmos com a estrutura “nós falamos”, de início, não temos a certeza de que se trata de um tempo verbal pretérito ou presente. Somente após averiguar o contexto, especificamente os termos “falamos no início” é que se confirma que se trata de um verbo no pretérito perfeito. Diferentemente do exemplo (199), em que a estrutura emergente de concordância verbal “a gente falou”, por si só, já carrega o tempo verbal associado ao próprio verbo, causado pela ausência da desinência de plural *-mos*. Um terceiro padrão de concordância foi observado neste contexto, trata-se da concordância não padrão *a gente* com *-mos*, que foi observada em apenas 2 ocorrências (0,9%). O excerto (200), retirado de nossa base de dados, ilustra esse dado:

- (200) Inf.: eu dei um chute que embarquei a bola... eles foram pegar a bola e **a gente corremos** né e fomos embora... mas foi dezenove a nove né... o C. fortim fez dez gol eu fiz nove (Inq. 62 – F2 – D2, homem, 69 anos)

Em relação ao pretérito perfeito com morfologia diferente para ambos os verbos, este foi o contexto verbal que se mostrou favorável ao uso do padrão emergente *a gente* sem *-mos* com menor intensidade dentre os contextos verbais pesquisados na amostra, tendo ocorrido em 140 dados (64,5%). Em contrapartida, apesar de apresentar menor proporção que o padrão emergente, este contexto foi o mais favorável à estrutura padrão *nós* com *-mos* em termos de proporção de uso dentre os verbos do modo indicativo, com 74 dados de concordância (34,1%). Esse resultado pode ser explicado pela ausência de ambiguidade temporal em verbos com essa característica, o que poderia de alguma forma diminuir a atuação da concordância emergente, tendo em vista que a estrutura *a gente* sem *-mos* pode estar sendo usada como recurso para desambiguar o verbo. A estrutura de concordância não padrão *nós* sem *-mos* foi flagrada em apenas 1 ocorrência (0,5%), que pode ser conferida no exemplo (201), enquanto o padrão de concordância verbal *a gente* com *-mos* foi flagrado em 2 dados (0,9%), e está representado no excerto (202).

- (201) Inf.: não/ não... porque quando nosso tio morreu **nós** não **foi** mais a praia... não gostava de praia não não não... ninguém/ alias o pessoal não ia muito a praia não (Inq. 62 – F2 – DID, homem, 69 anos)
- (202) Inf.: ... era a gente sentava na calça::da tudo isso a gente fazia tudo tudo tudo tudo isso **a gente fizemos** aí quando chegava o terço...tá na hora do terço... minha filha era briga ((risos)) era confuSÃO (Inq. 29 – F2 – DID, homem, 62 anos)

Verbos no pretérito imperfeito do indicativo apresentam alta associação ao padrão emergente de concordância verbal *a gente* sem *-mos*, tendo obtido 585 ocorrências, o que corresponde a um total de 91,3% dos dados. Essa característica também foi observada na análise da alternância pronominal, em que os verbos no pretérito imperfeito estavam fortemente correlacionados ao pronome inovador *a gente*. O padrão canônico de concordância, por sua vez, foi pouco recorrente neste contexto, sendo realizado em apenas 37 dados, o que equivale a 5,7% dos dados. Esse resultado sugere que o pretérito imperfeito do indicativo pode ser uma porta de entrada para a expansão das formas inovadoras de primeira pessoa do plural.

Além disso, cabe destacar que, dentre os contextos verbais investigados, o pretérito imperfeito foi o que apresentou maior número de ocorrências de concordância não padrão, sendo 19 realizações. Deste total de observações de concordância não padrão, observamos 17 dados referentes à estrutura *nós* sem *-mos* (2,6%), enquanto a estrutura de concordância *a gente* com *-mos* obteve apenas 2 dados (0,3%). Os exemplos (203) e (204), retirados da amostra de fala culta fortalezense dos anos 2020, ilustram realizações das variantes *nós* com *-mos* e de *a gente* sem *-mos*, respectivamente, no pretérito imperfeito do modo indicativo:

- (203) Inf.: só entrava com de paletó e graVA::ta né... e assim era feito no ano que tinha essas festas **nós tava** lá né SEMpre tinha um amigo de um conhecido do conhecido que arranjava um convite pra gente (Inq. 13 – F2 – DID, homem, 59 anos)
- (204) Inf. a gente chegou lá uma vez a gente tava **a gente estávamos** em quatro na época dois casais aí chegamos numa cidadezinha que não me lembro bem o nome AÍ eu acostumado com as pizzas daqui achava que a pizza grande seria POUco pra nós quatro (Inq. 50 – F2 – DID, homem, 56 anos)

Nossa análise para esta variável tem como foco os tempos verbais no modo indicativo. Por conta disso, a partir deste ponto, descreveremos os resultados deste preditor, mas não apresentaremos exemplos de ocorrências, uma vez que os outros fatores não são foco desta análise. Dito isto, em relação ao futuro do presente, encontramos 102 ocorrência de *a*

*gente* sem *-mos* (63%), 59 realizações de concordância padrão (36,4%), e apenas 1 dado de concordância verbal não padrão *a gente* com *-mos* (0,6%). Já o futuro do pretérito apresentou 11 dados de *a gente* sem *-mos* (91,7%), e apenas 1 dado de *nós* com *-mos* (8,3%), sem registro de concordância não padrão. Esse resultado confirma a tendência geral de nossa amostra: altas proporções de uso para o padrão emergente, proporções de uso discretas para o padrão canônico, e escassas ocorrências de concordância não padrão.

Essa mesma tendência pode ser observada ao analisar os tempos subjuntivos. No caso do futuro do subjuntivo, o ambiente é tão favorável para o padrão de concordância emergente que o resultado é categórico para *a gente* com *-mos* (100%). Já o presente do subjuntivo apresentou 22 ocorrências de *a gente* sem *-mos*, sendo responsável por 91,7% dos dados, e apenas 2 realizações de concordância padrão *nós* com *-mos* (8,3%). De forma parecida, o padrão emergente é favorito no imperfeito do subjuntivo, com 12 dados de *a gente* sem *-mos* (80%), e 3 ocorrências de *nós* com *-mos* (20%). Destacamos que realizações de concordância verbal não padrão não foram observadas no modo subjuntivo de maneira geral.

Por fim, em relação à forma nominal infinitivo, resultados apontam para a mesma direção na atuação dos padrões de concordância. Seguindo a tendência geral da amostra, verbos no infinitivo são mais receptivos ao uso do padrão emergente de concordância (94,5%), sendo o padrão canônico pouco recorrente neste contexto (1,8%), assim como a estrutura de concordância não padrão *nós* sem *-mos* (3,6%). Não foi registrado ocorrência de *a gente* com *-mos* em dados de infinitivo.

Além do *tempo e tipo de paradigma verbal*, outra variável previsora associada ao verbo que testamos neste estudo foi a *estrutura do verbo*. A observação desta variável visa identificar se existe algum tipo de correlação significativa entre a estrutura verbal e a escolha do padrão de concordância utilizado. A tabela abaixo apresenta a distribuição das ocorrências das variantes de concordância verbal em função da variável previsora estrutura do verbo.

**Tabela 53:** Distribuição das proporções da concordância verbal por *estrutura do verbo*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Estrutura</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Simple</b>	389	19%	1620	79,3%	24	1,2%	10	0,5%
<b>Locução</b>	107	19,9%	428	79,6%	1	0,2%	2	0,4%
<b>Composto</b>	2	13,3%	13	86,7%	0	0%	0	0%
$\chi^2 (6) = 5.1564, p = 0.5239$								

Fonte: elaboração própria.

Conforme se observa na tabela 53, assim como a tendência geral da amostra, há ampla preferência pelo padrão emergente de concordância em todos os contextos de estrutura verbal. Em verbos com estrutura simples, o uso de *a gente* sem *-mos* é responsável pela maioria das ocorrências (79,3%), enquanto a estrutura padrão *nós* com *-mos* é menos recorrente (18,9%). Ocorrências de concordância não padrão, embora poucas, são um mais frequentes em verbos com estrutura simples, sendo a variante *nós* sem *-mos* responsável por 24 ocorrências (1,2%), e a estrutura *a gente* sem *-mos* com 10 realizações (0,5%). Devido ao grande número de exemplos apresentados de concordância padrão e do padrão emergente, ilustramos ocorrências referentes apenas a concordância verbal não padrão. Dessa forma, os exemplos (205) e (206) demonstram, respectivamente, realizações de *nós* sem *-mos* e *a gente* com *-mos* associados a verbos com estrutura simples:

- (205) Inf.: só eu e tu que não levamos palmatória... só eu e o H. que ***nós fazia*** tudo direitinho... era da época da palmatória mas nós NUNca... nem eu nem tu nem o H. levamos palmatória (Inq. 62 – F2 – D2, homem, 60 anos)
- (206) Inf.: vamos dizer a gente tinha um... vamos dizer ***a gente saímos*** pra:... vamos dizer para as festas em Parangaba tranquilamente íamos e voltávamos a qualquer hora da noite... (Inq. 56 – F2 – D2, mulher, 50 anos)

As locuções verbais também se mostraram ambientes férteis ao uso de *nós* sem *-mos*, sendo observadas 428 ocorrências (79,6%), enquanto o padrão *nós* com *-mos*, seguindo a tendência da amostra, foi registrado em 107 observações (19,9%). Dados de concordância não padrão em locuções verbais foram observados em apenas 3 casos, sendo 1 ocorrência de *nós* sem *-mos* (0,2%), e 2 realizações de *a gente* com *-mos* (0,4%). Os excertos (207) e (208), retirados de nossa base de dados, ilustram dados de concordância não padrão das variantes *nós* sem *-mos* e *a gente* com *-mos* em locuções verbais:

- (207) Inf.: então oito horas apagava as luzes ***nós tinha que ir*** tudo dormir... e faltava muita energia... muita luz... faltava... faltava muita energia e a gente ficava nas calçada contando história de trancoso... de lobisomem... (Inq. 62 – F2 – D2, homem, 60 anos)
- (208) Inf.: vocês tão percebendo que as nossas aulas tá relacionada a teoria com a prática... essa PRÁxis... ***a gente estamos tentando***... eh/... eh... com ajuda do laboratório... Né? trazer... eh... melhor esclarecimento da teoria (Inq. 87 – F2 – EF, homem, 63 anos)

Por fim, ainda em relação à *estrutura verbal*, em verbos compostos foram observadas apenas ocorrências de concordância verbal padrão e do padrão emergente, não havendo registros de concordância não padrão. A variante emergente *a gente* sem *-mos* obteve 13 ocorrências, a maioria dos dados de concordância nesse contexto, o que corresponde a 86,7% do total. Já o padrão de concordância canônico se realizou em apenas 2 dados, sendo responsável por apenas 13,3% das ocorrências.

Além da *estrutura verbal*, outra variável preditora relacionada ao verbo que parece exercer alguma influência sobre a realização dos padrões de concordância é a *conjugação verbal*. Macêdo (2022), em estudo sobre os padrões de concordância verbal em Sergipe, identificou que a estrutura padrão de concordância verbal *nós* com *-mos* pode ser motivada por verbos na primeira e na segunda conjugação. Com o objetivo de analisar se essa tendência se repete em nossos dados, observamos a distribuição das variantes em função dessa variável, conforme demonstra tabela 54:

**Tabela 54:** Distribuição das proporções da concordância verbal por *conjugação verbal*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Conjugação</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>1ª conjugação</b>	135	13,9%	819	84,6%	9	0,9%	5	0,6%
<b>2ª conjugação</b>	233	22,8%	777	76%	11	1,1%	2	0,2%
<b>3ª conjugação</b>	130	21,5%	465	79,9%	5	0,8%	5	0,8%
<b><math>\chi^2 (6) = 31.787, p &lt; 0,001</math></b>								

Fonte: elaboração própria.

Como esperado, a distribuição das ocorrências de concordância verbal demonstra forte preferência pelo uso do padrão emergente de concordância verbal em todas as conjugações verbais, característica que é tendência de nossa amostra. A variante *a gente* sem *-mos* foi observada em 819 ocorrências em verbos de 1ª conjugação, totalizando 84,6% dos dados, enquanto a estrutura padrão de concordância foi registrada em 135 dados, o que corresponde a 13,9% do total de realizações. Estruturas de concordância não padrão foram registradas em 14 ocorrências, sendo 9 dados de *nós* sem *-mos* (0,9%) e 5 dados de *a gente* com *-mos* (0,6%). Os exemplos (209) e (210) ilustram realizações de concordância não padrão com verbos na 1ª conjugação.

(209) Inf.: O colégio Farias Brito que não era só Farias Brito que tinha era o mais famoso *nós ficava* resolvendo os TDS... o vestibular no ar e nos passava o

final de semana o sábado e domingo no/numa no terceiro ano que era o vestibular (Inq. 13 – F2 – DID, homem, 59 anos)

- (210) Inf.: as memórias vai falhando né? vai falhando... **a gente** já **falamos** dos carro... ah... também uma coisa que:: que:: num existe hoje... dona T... T. C. (Inq. 62 – F2 – D2, homem, 69 anos)

Os verbos que pertencem à segunda conjugação, assim como se observou com os verbos de primeira conjugação, são contextos mais produtivos a realização do padrão emergente de concordância verbal, com 777 dados (76%). Em seguida, o uso do padrão canônico de concordância é expressivo em nossa amostra, embora seja menos utilizado que a estrutura *a gente* sem *-mos*. Dados da variante *nós* com *-mos* apresentam 233 ocorrências, o que equivale a 22,8% dos dados. Realizações de concordância não padrão associados a verbos de segunda conjugação são escassos, tendo apenas 11 ocorrências de *nós* sem *-mos* (1,1%), e apenas 2 ocorrências de *a gente* com *-mos* (0,2%), como demonstram os exemplos (211) e (212):

- (211) Inf.: acho que durante a esCOla... faculdade... sempre/... / sempre **nós é** lecioNAdo que essa questão do de um conflito entre empregadOR e empregado... né? (Inq. 76 – F2 – D2, homem, 39 anos)
- (212) Inf.: aí com cinco anos de namoro a gente se casou e hoje **a gente** já **temos**... no caso dia seis vamos completar três anos de casados... temos uma relação muito bo::a mu::ito tran::quila e bem amorosa... (Inq. 07 – F2 – DID, mulher, 24 anos)

Com relação aos verbos que pertencem à terceira conjugação, esses seguem a mesma tendência observada nos outros verbos: alta proporção de uso de padrão emergente *a gente* sem *-mos*, com 465 realizações (79,9%). Disputando a preferência dos falantes como padrão de concordância de primeira pessoa do plural, a estrutura canônica *nós* com *-mos* foi observada em 130 casos, sendo responsável por 21,5% dos dados. Já as estruturas de concordância não padrão foram registradas em 5 ocorrências para cada variante de concordância não padrão. Os excertos (213) e (214), extraídos do projeto PORCUFORT da década de 2020, são exemplos de ocorrências das variantes *nós* sem *-mos* e *a gente* com *-mos*:

- (213) Inf.: que quando também sur/ via a televisão ninguém tinha televisão... **nós ia** pra casa do N. assistir televisão né... era na casa do N. e a primeira televisão que surgiu aqui foi da dona I. (Inq. 62 – F2 – D2, mulher, 60 anos)
- (214) Inf.: **a gente saíamos** pra::... vamos dizer para as festas em Parangaba tranquilamente íamos e voltávamos a qualquer hora da noite... nós não

tínhamos o DESprazer... de ter essa INsegurança que nós temos hoje... (Inq. 56 – F2 – D2, mulher, 50 anos)

Outra variável previsora que pode estar correlacionada a realização dos padrões de concordância verbal com a primeira pessoa do plural é a *tonicidade do verbo*. Para Rubio (2012), existe uma tendência natural dos falantes a evitarem a proparoxítone e, para isso, recorrem às formas verbais na terceira pessoa no singular, favorecendo o apagamento da desinência *-mos*, o que beneficiaria o uso do padrão emergente de concordância. A fim de testar esta hipótese, a tabela 55 apresenta a distribuição das realizações de nossos dados pela variável tonicidade do verbo:

**Tabela 55:** Distribuição das proporções da concordância verbal por *tonicidade do verbo*

PORCUFORT II - Anos 2020 (N = 2596)								
<i>Tonicidade</i>	<i>Nós com -mos</i>	%	<i>A gente sem -mos</i>	%	<i>Nós sem -mos</i>	%	<i>A gente com -mos</i>	%
<b>Monossílabo</b>	0	0%	636	99,4%	4	0,6%	0	0%
<b>Oxítono</b>	0	0%	279	98,9%	3	1,1%	0	0%
<b>Paroxítono</b>	462	28,2%	1146	70%	18	1,1%	11	0,7%
<b>Proparoxítono</b>	36	97,3%	0	0%	0	0%	1	2,7%
$\chi^2 (9) = 466.48, p < 0,001$								

Fonte: elaboração própria.

Conforme é possível observar na tabela 55, verbos monossílabos são ambientes praticamente categóricos para o uso do padrão emergente de concordância verbal, registrando 636 ocorrências de *a gente sem -mos*, o que corresponde a 99,4% dos dados. Outros 4 dados foram registrados para a variante de concordância não padrão *nós sem -mos*, o que representa apenas 0,6% das ocorrências de concordância verbal. Registros dos pronomes com marcação *-mos*, ou seja, *nós com -mos* e *a gente com -mos* não foram flagrados com verbos monossílabos nos dados representativos de fala culta fortalezense da década de 2020. Obviamente, qualquer verbo de apenas uma sílaba deixaria de ser um monossílabo com a adição da DNP *-mos*, o que torna esse resultado previsível. O excerto (215) ilustra exemplo de ocorrência de *nós sem -mos* neste contexto:

(215) Inf.:e o restante ali seria o lucro variável de mês a mês e TAL... e:: até MENos da meTAdede um pouco... e AÍ você via que... eh/ eh:: **nós é** ensinado que AH... NÃO ou você diminui o lu::cro e aumenta o salá::rio OU abaixa o salá::rio (Inq. 76 – F2 – D2, homem, 38 anos)

De forma parecida aos monossílabos, o comportamento dos verbos oxítonos também demonstram ser contextos verbais bastante favoráveis ao uso de *a gente* sem *-mos*, sendo praticamente categóricos ao uso do padrão emergente. Em nossa amostra, se observam 279 ocorrências de *a gente* sem *-mos*, o que totaliza 98,9%. Os dados referentes às realizações de *nós* sem *-mos* correspondem aos verbos no infinitivo sem a desinência *-mos* (1,1%), conforme ilustra o exemplo (216):

- (216) Inf.: aí o A diz pro B e para o C... a proposta é pra ***nós matar*** esse sujeito agora... aí o B e o C tudo cidadão de bem né... pulou e disse NÃO rapaz... não nasci pra isso não... (Inq. 69 – F2 – EF, homem, 58 anos)

Em relação aos verbos paroxítonos, estes são os mais produtivos da amostra e, assim como os outros, são mais favoráveis ao uso da estrutura emergente *a gente* sem *-mos*, tendo sido observado em 1146 dados, uma proporção de uso de 70%. Por outro lado, o padrão de concordância canônico, formado pelo pronome *nós* e a desinência *-mos*, é o segundo padrão de concordância preferido entre os informantes cultos de Fortaleza da década de 2020, tendo obtido 462 dados, ou seja, uma proporção de uso de 28,2%. Dentre os contextos verbais analisados, cabe destacar, os verbos paroxítonos foram os mais produtivos às formas não padrão de concordância, sendo a estrutura *nós* sem *-mos* flagrada em 18 ocorrências (1,1%) e a variante *a gente* com *-mos* sendo observada em 11 dados (0,7%), conforme demonstram os exemplos (217) e (218):

- (217) Inf.: ***nós morava*** num terreno grande... na frente ficar na casa da minha avó e atrás ficava a casa dos meus pais e aí eu passei minha inFÂNcia... em torno disso... em torno entre minha vó né entre minha mãe (Inq. 29 – F2 – DID, mulher, 62 anos)
- (218) Inf.: nossa mente é uma folha em branca ***a gente somos*** um nem bom nem mau vai depender das nossas experiências e das nossas escolHAS Russou vai dizer que a nossa natureza é boa foi a sociedade que nos corrompeu... (Inq. 79 – F2 – EF, mulher, 36 anos)

Por fim, em relação a verbos proparoxítonos, esses foram os menos produtivos na realização dos padrões de concordância, devido a tendência natural do falante de evitar as proparoxítonas (Tamanine, 2010; Mendonça, 2010; Rubio, 2012). Mattos (2013), em dados da fala goiana, identificou correlação significativa entre a marcação da desinência *-mos* e palavras proparoxítonas. Em nossos dados, entre os contextos verbais pesquisados, este é o único, em termos de proporção, mais favorável ao uso da concordância padrão com o pronome *nós*

(97,3%). Com esta tonicidade verbal, não há realizações de *a gente* sem *-mos*, padrão emergente tão abundante em nossa amostra. Tampouco há ocorrências de *nós* sem *-mos*, restando apenas 1 dado de concordância verbal não padrão, que se refere a realização de *a gente* com *-mos*, que pode ser conferido no exemplo (219):

(219) Inf.: a gente chegou lá uma vez a gente tava ***a gente estávamos*** em quatro na época dois casais aí chegamos numa cidadezinha que não me lembro bem o nome Aí eu acostumado com as pizzas daqui achava que a pizza grande seria POUco pra nós quatro (Inq. 50 – F2 – DID, homem, 56 anos)

Apresentados os dados em termos de frequência e proporção de uso, bem como alguns exemplos extraídos da base de dados em análise, a próxima seção será dedicada à construção e análise de um modelo de regressão logística, a partir do qual analisamos os dados em termos de estatística inferencial, através de um modelo de regressão logística, com foco em dados nos tempos verbais do modo indicativo.

### ***6.2.2 Análise multivariada em um modelo de regressão logística: a concordância verbal em um recorte binário – Amostra 2020***

Para a composição do modelo de regressão logística, pretendemos analisar a variável *tempo e tipo de paradigma verbal do modo indicativo*, inspirados nos trabalhos de Scherre, Yacovenco e Naro (2018), Carvalho, Freitas e Favacho (2020) e Carvalho e Santos (2023), que identificaram usos dos verbos no modo indicativo como atuantes sobre a realização dos padrões de concordância. Como citado anteriormente, a hipótese seria a de que o padrão emergente *a gente* sem *-mos* seria favorecido em contextos de presente com morfologia igual para pretérito perfeito e presente do indicativo, pois resolveria a questão da ambiguidade temporal desses verbos. Para testar essa hipótese em um contexto multivariado, através de um modelo de regressão, optamos por permanecer apenas com verbos no modo indicativo, assim como os autores. Com isso, excluímos os dados de futuro (162 dados), presente do subjuntivo (24), imperfeito do subjuntivo (15), futuro do subjuntivo (17), futuro do pretérito (12) e infinitivo (110), nos restando a 2256 dados de concordância verbal, distribuídos conforme demonstra a tabela 56:

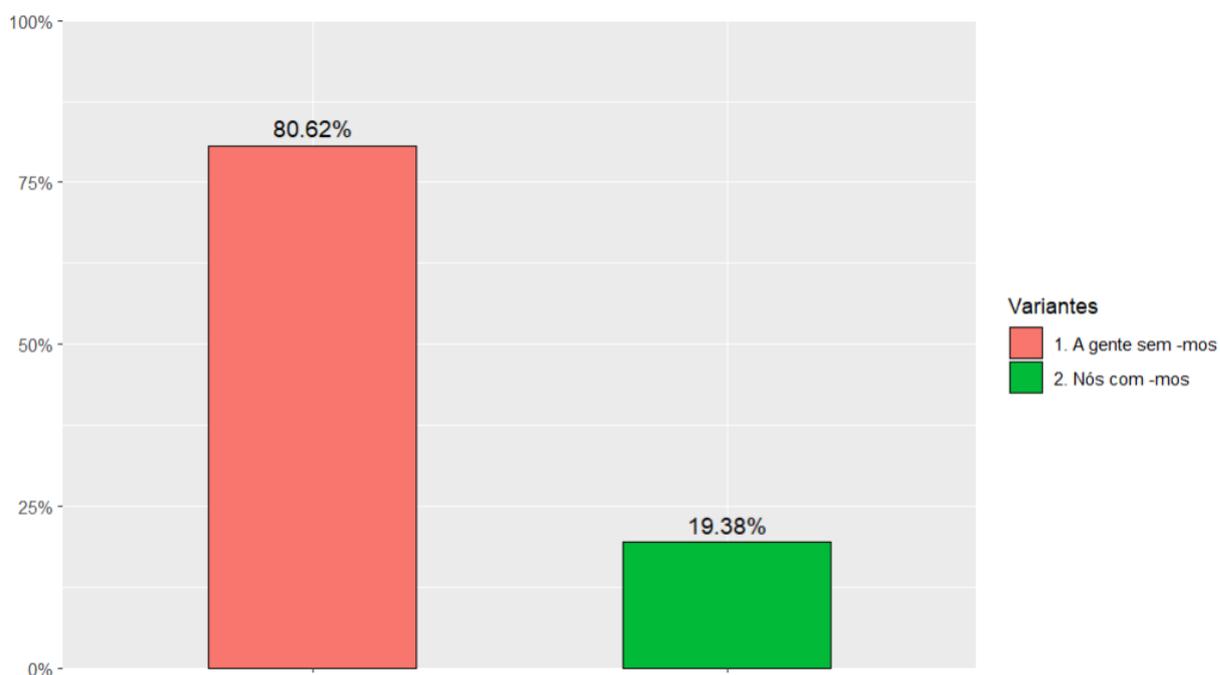
**Tabela 56:** Distribuição das proporções da concordância com dados no modo indicativo

Variante	<i>Nós com -mos</i>	<i>Nós sem -mos</i>	<i>A gente sem -mos</i>	<i>A gente com -mos</i>
Frequência	431	21	1793	11
Proporção	19,1%	0,9%	79,5%	0,5%

Fonte: Elaboração própria

Conforme observado na tabela 56, a distribuição das variantes de concordância verbal de primeira pessoa do plural, em nossa amostra, é desproporcional. O padrão emergente, formado pelo pronome *a gente* e o verbo na terceira pessoa do singular (*a gente com -mos*), é expressivamente mais utilizado entre os informantes, com 79,5% do total de ocorrências. Em seguida, o padrão de concordância mais antigo na língua (*nós com -mos*) é usado de forma significativa entre os informantes da amostra, sendo realizado em 19,1% dos dados. Os dados de concordância não padrão, conforme observado na tabela 56, apresentam poucas ocorrências. A estrutura *nós sem -mos* (*nós canta*) foi observada em 21 ocorrências, o que corresponde a 0,9% do total, enquanto o padrão *a gente com -mos* (*a gente cantamos*) foi registrado em 11 realizações, sendo responsável por apenas 0,5% dos dados.

Assim como aconteceu com análise dos dados de concordância verbal da Amostra 1990, a falta de balanceamento entre as variantes que compõem o fenômeno da concordância verbal na Amostra 2020 impede de que seja realizado um modelo de regressão multinomial, que evidencia e compara as quatro variantes em um contexto multivariado. Esse resultado já era esperado, tendo em vista que ocorrências de concordância não verbal em dados de fala são pouco recorrentes (Scherre, Yacovenco e Naro, 2018). Por conta disso, procedemos um recorte binário, considerando os padrões de concordância mais recorrentes em nosso *corpus* para comparar no modelo de regressão, assim como fizeram Carvalho, Freitas e Favacho (2020) e Carvalho e Santos (2023). Com isso, passamos a ter uma variável resposta binária, ou seja, que comporta duas variantes: *nós com -mos* e *a gente sem -mos*. Com a remoção dos dados de concordância não padrão, nos restaram para a análise do modelo de regressão 2224 dados de concordância, distribuídos conforme o gráfico abaixo:

**Gráfico 36:** Proporções da concordância verbal em recorte binário – Amostra 2020

Fonte: elaboração própria

Conforme podemos observar no gráfico 36, após a remoção das ocorrências de *nós* sem *-mos* e *a gente* com *-mos*, nos restaram 2224 dados de concordância verbal de primeira pessoa do plural, distribuídos da seguinte forma: 1793 dados de *a gente* sem *-mos* (80,62%) e 431 dados de *nós* com *-mos* (19,38%). Com esse recorte binário, submetemos os dados a função *glmer* (Generalized Linear Model), através do *software R Studio*, para a elaboração de um modelo de regressão logística incluindo as variáveis predictoras testadas nesta pesquisa, a saber, *sexo/gênero*, *faixa etária*, *tipo de inquérito*, *tempo* e *tipo de paradigma verbal do modo indicativo*, *tonicidade do verbal*, *conjugação verbal* e *estrutura verbal*, de modo que o modelo criado apresentou a seguinte sintaxe:

```
glmer(VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO + PARADIGMA +
ESTRUTURA.VERBAL + TONICIDADE + CONJUGACAO + (1|INFORMANTE), data =
Concordancia_2020, family = binomial)
```

No entanto, após a execução do modelo logístico na ferramenta *R Studio*, houve problemas ao convergir a modelagem proposta. Basicamente, o mesmo erro observado na análise dos dados da Amostra 1990 (ver 6.2.1). Com a exclusão dos dados de concordância não verbal, a variável predictor *tonicidade do verbo* apresentou muitas células vazias, o que

compromete a verossimilhança do modelo estatístico, tendo em vista que o balanceamento na distribuição dos dados é um dos princípios básicos para a criação de um modelo de regressão. Diante de tantas células vazias, nossa opção foi remover a variável *tonicidade* do modelo de regressão logística. Após a retirada deste preditor, o modelo convergiu corretamente.

Iniciamos, então, os testes para comprovar a eficiência do modelo de regressão, os mesmos aos quais foram submetidos os dados da década de 1990. O primeiro deles, o teste *vif* (Variance Inflation Factor), do pacote *CAR*, é utilizado para observar se existe multicolinearidade nos dados a serem analisados, ou seja, se existem variáveis que são redundantes. A ausência de multicolinearidade é essencial para a análise de regressão logística, tendo em vista que é um dos pilares de um modelo de regressão. Após submeter nossos dados de concordância função *Vif*, o programa *R Studio* gerou o seguinte *output*:

**Tabela 57:** Resultado do teste de colinearidade *vif* das variáveis preditoras - Amostra 2020

<b>Resultado do teste de inflação da variância</b>			
	<b>GVIF</b>	<b>Df</b>	<b>GVIF<sup>1/(2*Df)</sup></b>
<b>Sexo/gênero</b>	1.036446	1	1.018060
<b>Faixa etária</b>	1.087271	2	1.021138
<b>Tipo de inquérito</b>	1.255235	2	1.058477
<b>Paradigma verbal</b>	2.373692	4	1.114110
<b>Estrutura Verbal</b>	1.149711	2	1.035493
<b>Conjugação</b>	1.992943	2	1.188157
<b>car::vif(Concordancia_2020)</b>			

Fonte: Elaboração própria.

Como visto, o valor arbitrado em pesquisas mais rigorosas para determinar quando há multicolinearidade nos dados é GVif ajustado igual a 5. Em nossos dados, todos os valores estão próximos a 1, o que indica que, no modelo criado, não há colinearidade, e que podemos prosseguir com a análise dos dados. Submetemos, então, nosso modelo à seleção automatizada das variáveis previsoras que se mostram relevantes para a análise multivariada em nosso modelo através da função *step*. Como vimos, essa função faz parte da instalação base do *R* e utilizamos como parâmetro para composição do modelo ideal para explicar os nossos dados os menores valores descritos no Critério de Informação Akaike (AIC). Após submeter o modelo as funções *step forward*, *step backward* e *step both*, obtivemos o seguinte resultado:

**Tabela 58:** Resultado do teste de seleção das variáveis através da função *step* - Amostra 2020

<b>Resultado Step forward (AIC = 1747.44)</b>			
	<b>Df</b>	<b>Df Deviance</b>	<b>AIC</b>
<b>&lt;none&gt;</b>		1725.4	1747.4
<b>-Sexo/gênero</b>	1	1724.9	1748.9
<b>-Conjugação</b>	2	1723.2	1749.2
<b>Modelo sugerido: VR ~ PARADIGMA.VERBAL + TIPO.INQUERITO + FAIXA.ETARIA + ESTRUTURA.VERBAL</b>			

Fonte: elaboração própria

O resultado dos testes das funções *step forward*, *step backward* e *step both* indicaram o mesmo valor limite de AIC (1747.44), e nos três resultados foi recomendada a exclusão das variáveis *sexo/gênero* e *conjugação verbal*. Conforme podemos observar na tabela 58, por considerarmos o menor valor de AIC para seleção de variáveis, o teste de *step* apontou que até mesmo um modelo vazio, ou seja, um modelo que não contém qualquer variável *<none>*, tem maior poder explicativo que um modelo que inclui as variáveis *sexo/gênero* e *conjugação verbal*. Por esse motivo, acatamos a indicação do teste e removemos essas variáveis do nosso modelo de regressão final. Mantemos então este modelo para efeito de análise dos dados da Amostra 2020.

Tendo composto o modelo ideal para o tratamento de nossos dados, submetemos o modelo à análise através da função *lrm* (Logistic Regression Model), que nos forneceu informações referentes à eficiência do modelo na explicação dos dados. Dentre as informações fornecidas, focaremos no índice de discriminação C, que comprova o poder explicativo do modelo logístico no processo de variação dos dados. Após testar a função *lrm*, o seguinte resultado foi gerado:

**Tabela 59:** Resultado do teste multivariado de *nós e a gente* através da função *lrm*

	<b>Model Likelihood Ratio Test</b>		<b>Discrimination Indexes</b>		<b>Rank Discrim. Indexes</b>		
<b>Obs</b>	2224	<b>LR chi2</b>	661.98	<b>R2</b>	0.300	<b>C</b>	0.808
<b>A gente sem -mos</b>	1793	<b>d.f.</b>	10	<b>R2(13,1589)</b>	0.184	<b>Dxy</b>	0.616
<b>Nós com -mos</b>	431	<b>Pr(&gt;chi2)</b>	>0.0001	<b>R2(13,1169.5)</b>	0.352	<b>Gamma</b>	0.626
<b>Max reriv</b>	6e-08			<b>Brier</b>	0.121	<b>Tau-a</b>	0.192

Fonte: Elaboração própria.

Os dados da tabela 59 apresentam informações que atestam a qualidade do modelo de regressão logística criado para o tratamento de nossos dados de concordância verbal da década de 2020. A primeira informação que devemos analisar é o LR chi2, que se refere ao teste de Razão De Verossimilhança. Esse teste compara o efeito do modelo multivariado que inclui todas as variáveis previstas com um modelo nulo, sem qualquer variável, para determinar a eficácia na explicação dos dados. O valor de LR Chi2 apontou um valor igual a 661.98 que,

tendo como ponto médio da escala o valor 500, podemos assegurar que a verossimilhança do modelo é aceitável do ponto de vista estatístico. Além disso, o Índice de Concordância C é importante para determinar o poder explicativo do modelo criado. No caso do nosso modelo, o valor do índice C é igual a 0.808 que, segundo Levshina (2015), apresenta um bom poder discriminatório dos dados da Amostra 2020.

Tendo atestado a eficácia estatística no tratamento dos dados, passamos então a analisar o uso da concordância verbal em um recorte binário através de um modelo multivariado. Através da função *glmer*, por meio do ambiente *R Studio*, foram geradas estimativas de uso e valores estatísticos que destacam a atuação de cada fator sobre a realização das variantes de concordância. Destacamos que, ao convergir, o modelo de regressão logística gerou um valor de *intercept*, que será tomado como referência em termos de estatística inferencial. No caso deste modelo logístico, o valor de *intercept* corresponde ao uso da variante de concordância *nós* com *-mos* nos seguintes contextos sociais e linguísticos: (1) 22 a 35 anos de idade, (2) inquéritos do tipo DID, (3) presente com paradigma diferente e (4) verbos com estrutura simples. Os resultados das estimativas de uso no ambiente multivariado podem ser conferidos na tabela 60:

**Tabela 60:** Estimativas de *nós* com *-mos* no modelo de regressão logística: 2020 (N = 2224)

<i>Coefficients</i>	Estimate	Std. Error	z value	Pr(> z )	
<b>(Intercept)</b>	-1.62110	0.17916	-9.048	< 2e-16	***
<b>Faixa etária</b>					
36 a 50 anos	0.56341	0.18581	3.032	0.00243	**
A partir de 50 anos	1.53756	0.16775	9.166	< 2e-16	***
<b>Tipo de inquérito</b>					
D2	-1.10347	0.16751	-6.588	4.47e-11	***
EF	0.78908	0.14567	5.417	6.06e-08	***
<b>Tempo e Paradigma verbal</b>					
Presente igual	-1.76226	0.20288	-8.686	< 2e-16	***
Pret. Perf. Igual	-0.01274	0.17414	-0.073	0.94167	
Pret. Perf. Diferente	0.29296	0.18909	1.549	0.12131	
Pret. imperfeito	-1.62600	0.20985	-7.748	9.32e-15	***
<b>Estrutura verbal</b>					
Locução verbal	-0.86771	0.18475	-4.697	2.64e-06	***
Verbo composto	0.31084	0.81684	0.381	0.70355	
Modelo: glm(VR ~ SEXO.GENERO + FAIXA.ETARIA + TIPO.INQUERITO + ESTRUTURA.VERBAL + CONJUGACAO + PARADIGMA.VERBAL)					
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>					

Fonte: elaboração própria.

De forma geral, a análise inicial do modelo de regressão logística pontuou a interação estatisticamente relevante de todas as variáveis predictoras incluídas na versão final do modelo que comporta os dados representativos da década de 2020. Essa tendência confirma a análise em termos de frequência e proporção de uso. Ou seja, assim como observamos na distribuição das variantes de concordância verbal em função dos preditores, em que o padrão emergente apresentou maior proporção de uso em relação ao padrão canônico de concordância em praticamente todos os contextos observados, na análise multivariada também observamos que os contextos estudados, de uma forma geral, tendem a inibir o uso do padrão canônico de concordância *nós* com *-mos*, e a favorecer o uso do padrão emergente *a gente* sem *-mos*.

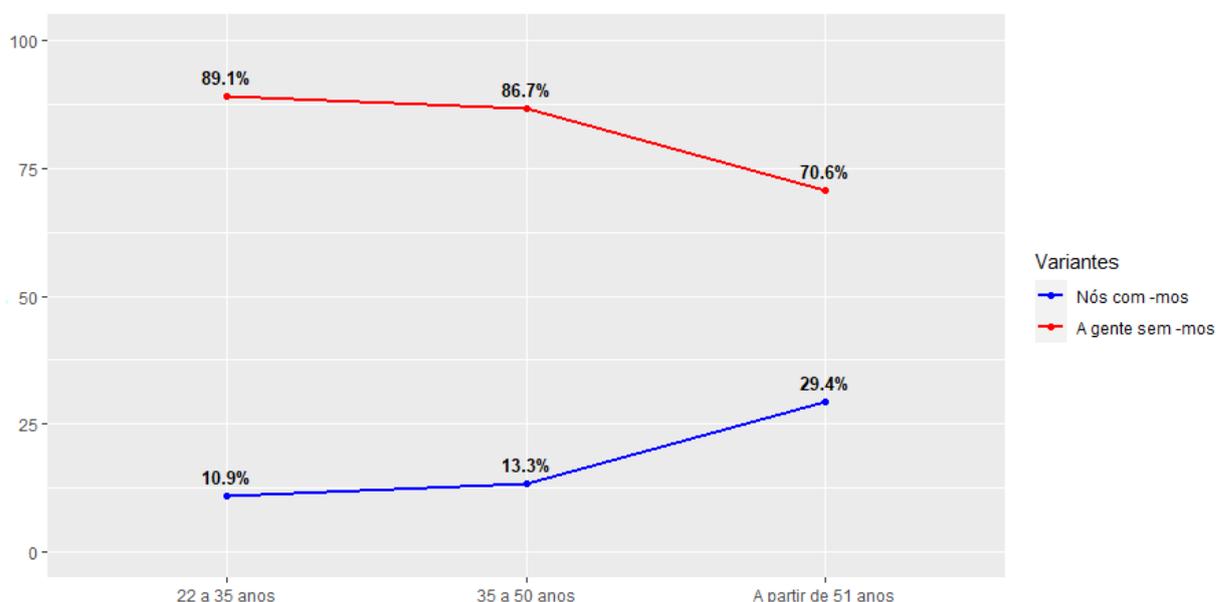
A seguir, apresentamos a atuação de cada variável predictorora em um contexto de interação. Os resultados são dispostos na escala de *logodds* (*log-odds-ratio*), em que valores positivos indicam tendência ao favorecimento e, ao contrário, valores negativos representam tendência ao desfavorecimento em relação ao *intercept*. Nossa variante de referência é a estrutura padrão de concordância *nós* com *-mos*, ou seja, todos os resultados apresentados se dão em função desta variante. Cabe destaque que, para se atribuir o valor de estimativa a um fator de uma variável, é necessário somar seu valor ao valor do *intercept*. A primeira variável predictorora descrita na análise da concordância verbal de primeira pessoa com dados da Amostra 2020 é a *faixa etária*.

#### *a) Atuação da faixa etária sobre os padrões de concordância verbal*

A tabela 60 apresenta o resultado da atuação de todas as variáveis predictororas atuando em conjunto em um contexto multivariado a partir de um modelo de regressão logística, que busca evidenciar os contextos mais favoráveis a realização da concordância verbal padrão *nós* com *-mos*. A primeira variável exposta pelo modelo é de ordem extralinguística e trata-se da *faixa etária*. Conforme podemos observar, os informantes mais jovens (22 a 35 anos) não aparecem explicitamente no resultado do modelo, uma vez que estes compõem o *intercept*, que é o valor de referência de nossa análise. O valor de estimativa atribuído ao *intercept* é igual a -1.62110, tendo atuação estatisticamente relevante ( $p\text{-value} < 0,001$ ). Isso quer dizer que, nos contextos de realização do *intercept*, os informantes mais jovens tendem a desfavorecer expressivamente o uso da concordância verbal padrão *nós* com *-mos*, sendo estes informantes mais propensos ao uso do padrão emergente *a gente* com *-mos*.

Em relação aos informantes de faixa etária de 36 a 50 anos, estes também atuam de forma significativa na realização dos padrões de concordância verbal. Em relação ao *intercept* (mais jovens), os informantes com essa faixa etária tendem a desfavorecer em menor grau o uso da concordância verbal padrão, mas ainda agem de forma a inibir a estrutura *nós* com *-mos*, com estimativa de uso igual a  $-1,05769$ . O valor de estimativa negativa indica que, embora façam com menor intensidade que os mais jovens, os falantes com faixa etária entre 36 a 50 anos também estão inclinados ao uso do padrão emergente *a gente* sem *-mos* na Amostra 2020. Esses resultados são confirmados ao observar as proporções de uso de cada faixa etária em um recorte binário em nossa amostra, conforme demonstram o gráfico 36:

**Gráfico 36:** Proporção de uso de *nós* com *-mos* e *a gente* sem *-mos* por faixa etária



Fonte: elaboração própria

Conforme pode ser observado no gráfico 36, percebe-se alto percentual de uso do padrão emergente de concordância *a gente* sem *-mos* nas duas primeiras faixas etárias (89,1% e 86,7%, respectivamente), com uma significativa redução entre os informais com mais de 50 anos (70,6%). Em termos de favorecimento, os falantes com idade a partir de 51 anos apresentam estimativa de uso igual a  $-0,08354$ , valor que indica que, em relação ao *intercept*, há leve desfavorecimento de *nós* com *-mos*, tendo em vista que o ponto de neutralidade é 0. Esse resultado demonstra que, em um contexto multivariado, todas as faixas etárias apresentam tendência a inibir a forma padrão de concordância, o que nos leva a inferir que há favorecimento do padrão emergente *a gente* sem *-mos* em todas as idades pesquisadas.

Para Labov (1994) uma mudança em progressão pode ser observada se contrastarmos a fala de informantes de diferentes faixas etárias e observar o comportamento linguístico destes no que diz respeito ao uso das variantes inovadoras e conservadoras na língua. No caso do fenômeno em estudo, a inserção do pronome *a gente* no quadro pronominal do português criou um padrão emergente de concordância verbal na língua que pode ser considerado inovador, e passou a disputar espaço com o padrão de concordância mais antigo, o *nós* com *-mos*. Assim, a distribuição das variantes em função das faixas etárias pesquisadas pode trazer indícios de mudanças, conforme discorre Oliveira (2022):

À medida que uma variante inovadora se mostra gradativamente mais produtiva em grupos mais jovens e menos produtiva em grupos mais velhos e, em contrapartida, a variante conservadora realiza o caminho contrário, ou seja, é mais frequente na fala de indivíduos mais velhos e menos frequente na fala de indivíduos mais novos, tem-se um indicativo de mudança em curso (Oliveira, 2022, p. 38)

Relacionando essas palavras com os nossos resultados, no entanto, é preciso ter cuidado ao afirmar que se trata de uma mudança. De fato, os mais jovens usam a variante inovadora com mais frequência, são quase 90% das ocorrências de concordância verbal. No entanto, os informantes da faixa etária de 51 + da amostra também utilizam a variante inovadora com alta proporção de uso (70,6%), além de favorecer, mesmo que de forma discreta, o uso do padrão emergente em termos de estimativa (-0,08354), no modelo de regressão logística. Por outro lado, há de se considerar que os falantes com mais de 50 anos apresentam uma proporção de uso da forma canônica *nós* com *-mos* de 29,4%, que pode parecer pouco, mas é uma proporção bastante expressiva, mesmo levando em conta que, no fenômeno analisado, a variante emergente não sofre estigma linguístico, conforme afirma Bagno (1999).

#### *b) Atuação do tipo de inquérito sobre os padrões de concordância verbal*

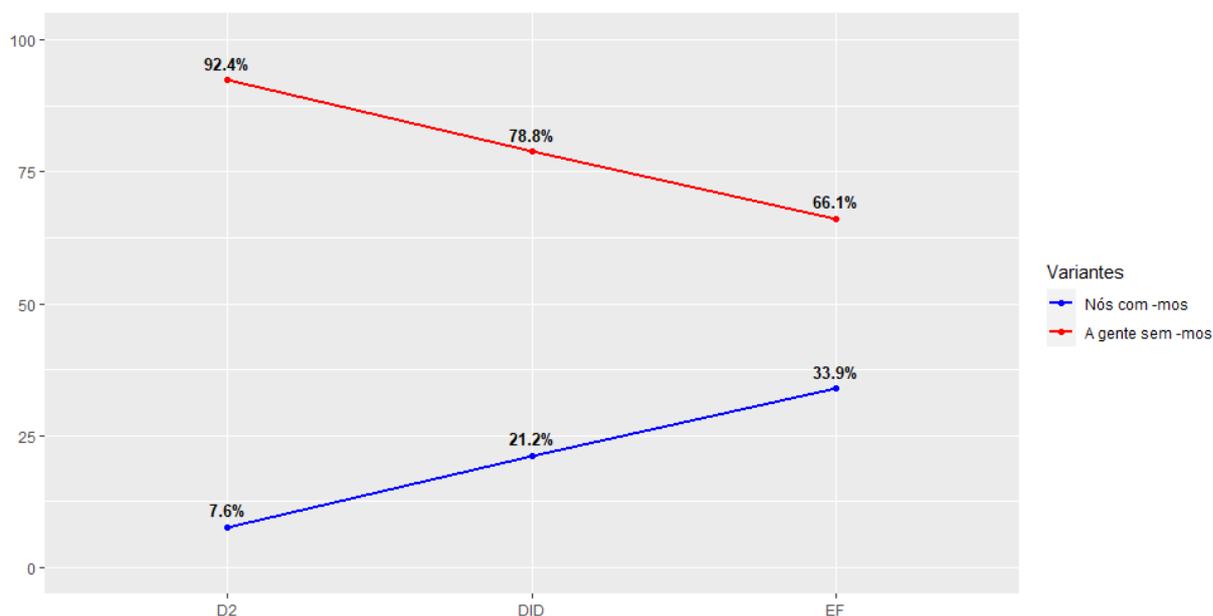
A variação na concordância verbal de primeira pessoa do plural é um fenômeno variável fortemente influenciado pela escolaridade dos informantes (Scherre & Naro, 2014). Isso pode explicar as poucas ocorrências de concordância não padrão na amostra, tendo em vista que se trata de um fenômeno fortemente estigmatizado socialmente. Principalmente se tratando de uma amostra composta por informantes graduados. Estudos sociolinguísticos têm apontado tendência dos mais escolarizados ao uso de variantes não estigmatizadas (Freitag, 2016). Outro fator que também leva o falante a evitar variantes socialmente estigmatizadas é o

grau de formalidade empregado na fala, tendo em vista que ambientes que exigem mais formalidade condicionam o uso de variantes linguísticas mais conservadoras (Labov, 1972). A variável *tipo de inquérito* objetiva analisar a atuação do grau de formalidade sobre o uso de *nós* com *-mos*.

A tabela 60 demonstra a atuação do preditor *tipo de inquérito* sobre a realização da variante de referência desta análise, a estrutura padrão de concordância *nós* com *-mos*. Conforme podemos observar, o valor de *intercept* refere-se a inquéritos do tipo DID (diálogo entre informante e documentador), que nesta pesquisa consideramos ter grau intermediário de monitoramento a fala, uma vez que a presença do entrevistador e do gravador podem inibir o verdadeiro vernáculo do informante. Esse tipo de inquérito, nos mesmos contextos de realização do *intercept*, tende a desfavorecer fortemente o uso da estrutura padrão de concordância padrão com o pronome *nós*, com estimativa de uso de  $-1.62110$ . Esse resultado aponta para a forte expansão no uso de *a gente* com *-mos* em todos os contextos, inclusive os que exigem certo nível de formalidade.

Em relação aos inquéritos do tipo D2 (diálogo entre dois informantes), estes são os registros classificados como os de menor grau de monitoramento à própria fala, já que se tratam de conversas entre pessoas que possuem certa intimidade, como amigos e parentes. Estes registros são ainda mais desfavoráveis ao uso da forma de concordância padrão *nós* com *-mos*, com valor de estimativa de  $-2,72457$ , valor bastante inferior ao *intercept* deste modelo. Esse resultado pode sugerir que o padrão emergente *a gente* sem *-mos* pode estar associado mais intensamente a contextos de menor formalidade, embora seja uma estrutura que não é avaliada negativamente (Peluco, 2022). Esse resultado, vale destacar, confirma as proporções de uso de cada variante distribuídas em função do *tipo de inquérito*, ou seja, a estrutura *a gente* sem *-mos* é mais intensamente favorecido em inquéritos do tipo D2 e menos intensamente em registros do tipo DID:

**Gráfico 37:** Proporção de *nós* com *-mos* e *a gente* sem *-mos* por tipo de inquérito - Amostra 2020



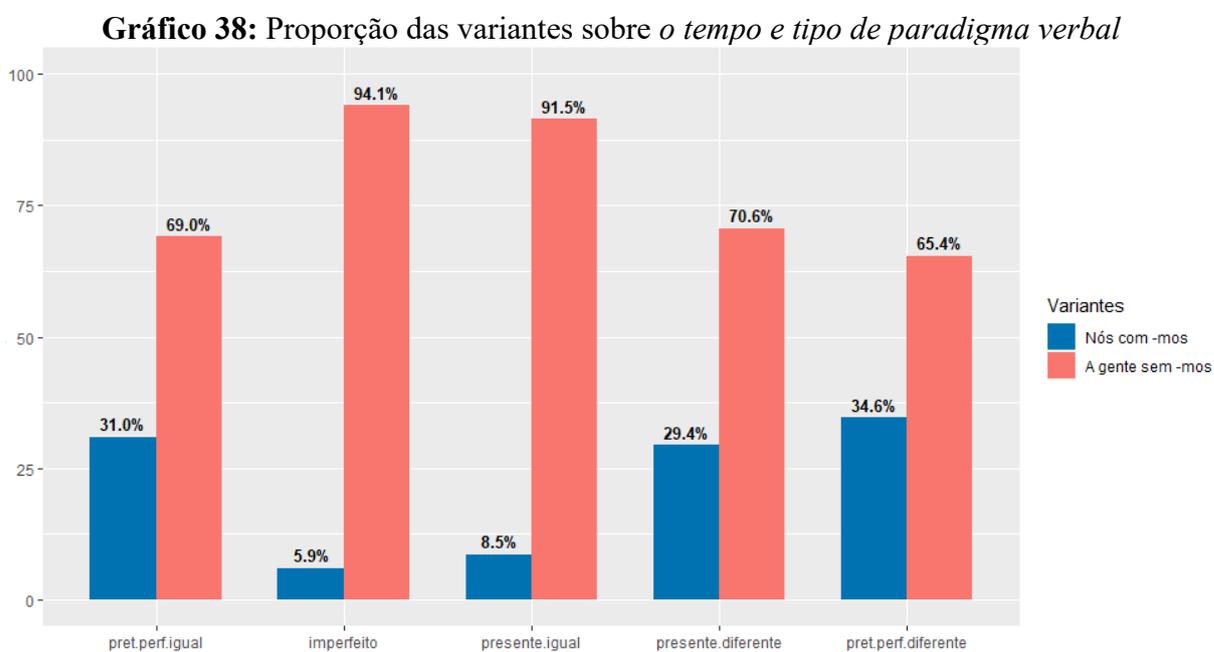
Fonte: elaboração própria

Por fim, as elocuições formais (EF) são os inquéritos mais formais entre os analisados nesta pesquisa. Neste tipo de registro, se comparado ao *intercept*, há um aumento na estimativa de uso da concordância padrão *nós* com *-mos*, porém esta continua sendo desfavorecida também neste contexto, com valor estimado de atuação igual a  $-0,83202$ , inibindo seu efeito quase pela metade (em relação ao *intercept*). Esse resultado está em conformidade com o gráfico 37, uma vez que o padrão de concordância *nós* com *-mos* é menos frequentemente usado em inquéritos do tipo D2, e mais recorrente em registros do tipo EF. Por outro lado, também, fica evidente o forte favorecimento do padrão emergente *a gente* sem *-mos*, tendo em vista que “essa forma não é avaliada negativamente, podendo transitar muito bem em diferentes estilos de monitoração e eventos de oralidade e letramento, pois quem a utiliza não sofrerá estigma” (Peluco, 2022, p. 133-134).

#### *d) Atuação do tempo e tipo de paradigma verbal do modo indicativo sobre os padrões de concordância verbal*

A variável previsora *tempo e tipo de paradigma verbal do modo indicativo* é uma variável complexa que envolve diferentes níveis linguísticos. Neste estudo, assumimos seu

caráter funcional, no sentido de a estrutura emergente *a gente* sem *-mos* estar atuando em verbos com mesma morfologia na relação presente/passado para inibir a potencial ambiguidade semântico temporal (Scherre, Yacovenco e Naro, 2018). Neste modelo de regressão logística, apenas dois fatores, além do *intercept*, se mostraram significativos para a realização dos padrões de concordância verbal, a saber, presente com paradigma igual e pretérito imperfeito. Antes de analisar o favorecimento de cada variante, decidimos relacionar as estimativas de uso às proporções de ocorrências em cada contexto, tendo em vista a alta recorrência de *a gente* sem *-mos* na amostra representativa da fala culta da década de 2020. O gráfico 38 demonstra a proporção das variantes de concordância em cada contexto verbal:



Fonte: elaboração própria

O gráfico 38 demonstra que as proporções de uso são muito elevadas para o uso do padrão emergente nos contextos que se mostraram estatisticamente significativos para o fenômeno variável. Nesse sentido, é necessário considerar que a alta frequência de uso de *a gente* sem *-mos* pode estar se sobressaindo sobre o efeito da variável no contexto multivariado, o que, de fato, não anula a hipótese das autoras, que adotamos nesta análise, mas a confirma. A alta frequência de ocorrências de uma variante, *a gente* sem *-mos*, em determinado ambiente é indicativo de que esse ambiente é favorável ao uso desta variante.

Nesse sentido, o *intercept* deste modelo corresponde a atuação de verbos no presente com formas diferentes para presente e passado. No contexto do *intercept*, os verbos no presente com morfologias diferentes apresentam forte tendência a inibir o uso da concordância

padrão *nós* com *-mos* (-1.62110). Na mesma direção, os verbos no presente com mesma forma para o presente e o passado tendem a inibir fortemente o uso da concordância verbal padrão, apresentando estimativa de uso de -3,38336, o dobro da força inibidora apresentada pelo *intercept* do modelo, sendo, dentre os fatores investigados, o contexto que mais desfavorece a variante de referência *nós* com *-mos*. Esse resultado pode indicar que, na década de 2020, a padrão emergente pode estar atuando como mecanismo linguístico para desfazer a ambiguidade semântica temporal desses verbos, confirmando nossa hipótese. Há se de considerar também que a alta proporção de uso de *a gente* sem *-mos* neste contexto (91,5%), o que poderia facilmente desfavorecer a variante ao qual está sendo confrontada. Como dito, seja em termos de favorecimento, seja em termos percentuais, podemos acatar nossa hipótese sobre o uso dessa variável.

O pretérito imperfeito tem comportamento parecido. No modelo de regressão logística, se observa forte desfavorecimento no uso da concordância padrão com verbos no pretérito imperfeito, sendo atribuída estimativa de uso de -3,2471, expressamente menor que o intercepto do modelo. Primeiro aspecto a se considerar é a baixa proporção de uso de *nós* com *-mos* associadas a verbos no imperfeito na Amostra 2020 (5,7%), o que favorece facilmente o padrão emergente nesse contexto (ver tabela 52). Outro aspecto a se observar é a natureza dos dados, a maioria dos inquiridos usam a técnica da narrativa pessoal, que são ambientes propícios ao uso do imperfeito, uma vez que relatam situações que aconteciam corriqueiramente. Por outro lado, a narrativa pessoal condiciona o falante a não monitorar a fala, o que pode favorecer o uso de variantes inovadoras. Isso pode estar agindo no sentido de favorecer o padrão *a gente* sem *-mos* e, por outro lado, inibindo o uso do padrão canônico de concordância *nós* com *-mos*.

Por fim, é necessário levar em consideração que os verbos no pretérito imperfeito estão relacionados com outro fenômeno linguístico: a redução das proparoxítonas em paroxítonas. Nesse sentido, haveria uma tendência natural a evitar a marcação de *-mos* para reduzir as proparoxítonas em paroxítonas, favorecendo o pronome *a gente*, que é marcado com concordância de terceira pessoa do singular, ou seja, com ausência da desinência *-mos* (*a gente* sem *-mos*). Resultados parecidos são encontrados em Peluco (2022), que explica:

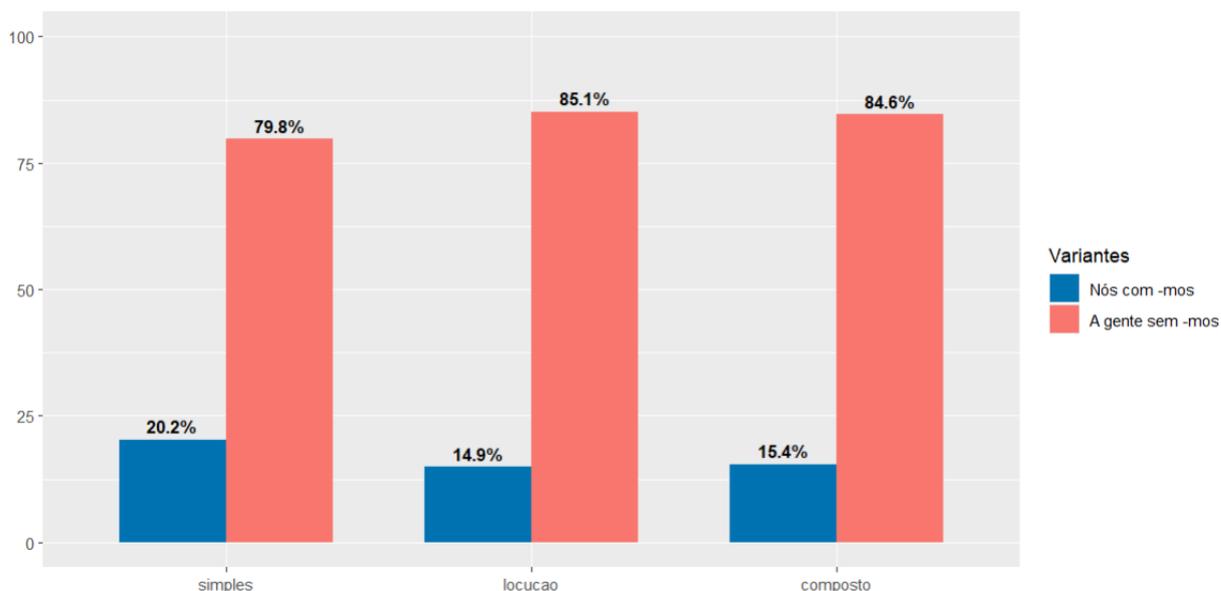
o falante evita formas proparoxítonas nas formas verbais de 1PP, ou seja, o uso de verbos no pretérito imperfeito, a fim de que ajam de acordo com uma tendência fonológica do português brasileiro falado, a de evitar acentuar a antepenúltima sílaba, eliminando, assim, a última sílaba. (Peluco, 2022, p. 93)

Essa hipótese se confirma em nossos dados, tendo em vista a alta proporção uso de *a gente* sem *-mos* (91,3%) com verbos no pretérito imperfeito. Além disso, o modelo atribuiu a este fator um valor de  $p$  igual a  $9.32e-15$ , o que o classifica como estatisticamente confiável para a explicação de nossos dados. Os demais fatores não apresentaram relevância estatística no modelo de regressão, uma vez que os valores de  $p$  atribuídos a estes fatores estavam fora do intervalo de confiança adotado nesta pesquisa, motivo pelo qual não serão descritos aqui.

A próxima variável de cunho linguístico analisada modelo multivariado foi a *estrutura do verbo*, acreditando haver relação entre o uso da concordância verbal e a estrutura verbal usada pelo informante.

#### *e) Atuação da estrutura do verbo sobre os padrões de concordância verbal*

Em relação a variável previsora *estrutura do verbo*, esta foi a última incluída neste modelo de regressão logística que analisa os dados de concordância verbal da década de 2020 em um contexto multivariado. Conforme demonstra os dados apresentados na tabela 60, o valor de referência do modelo (*intercept*) se refere a atuação de verbos com estrutura simples. A estimativa de uso atribuída a atuação de verbos simples é igual a  $-1.62110$ , o que significa dizer que, nos mesmos parâmetros do *intercept*, este tipo de estrutura verbal tende a desfavorecer expressamente a realização do padrão de concordância *nós* com *-mos*, sendo ambientes mais favoráveis a utilização do padrão emergente. Esse resultado pode estar relacionado a baixa proporção de uso de *nós* com *-mos* em nossa amostra, conforme demonstra os resultados do gráfico 39:

**Gráfico 39:** Proporção de uso das variantes em função da *estrutura verbal*

Fonte: Elaboração própria

De acordo com os percentuais de uso expostos no gráfico 39, em contextos de estrutura verbal simples, a proporção de uso do padrão emergente de concordância é intensamente mais acentuada que o padrão canônico, motivo pelo qual a estrutura *nós com -mos* pode sofrer inibição quando associado a verbos com essa estrutura. Além disso, há que se considerar que o valor de  $p$  atribuído a este fator ( $p\text{-value} < 0,001$ ) está dentro do intervalo de confiança, motivo pelo qual esse resultado pode ser considerado estatisticamente confiável.

Um padrão parecido pode ser observado no comportamento dos padrões de concordância associados a locuções verbais que, em relação ao valor de referência do modelo, apresentam estimativa de uso igual a  $-2,48881$ . Esse valor de estimativa está bem abaixo do valor de *intercept*, o que sugere que as locuções verbais demonstram uma tendência ainda maior a desfavorecer o uso da concordância padrão *nós com -mos*. Esse resultado está em conformidade com a distribuição das proporções de uso em função de cada nível deste preditor (ver gráfico 39), tendo em vista que, dentre as estruturas verbais analisadas, as locuções verbais foram as menos recorrentes para a concordância verbal padrão *nós com -mos* (14,9%). Esse resultado é estatisticamente significativo, com valor de  $p < 0,001$ .

Uma terceira estrutura verbal foi observada em nossa amostra, os verbos compostos, que são construções verbais constituídas pelo verbo *ter* ou *haver* em conjunto com o particípio do verbo principal. No entanto, nosso modelo demonstrou que esse tipo de estrutura verbal não demonstra relevância estatística em um contexto multivariado, apesar da alta proporção de uso do padrão emergente *a gente sem -mos* em nossa amostra (86,7%). Por conta

disso, não podemos considerar o resultado para este fator confiável em termos de estatística inferencial.

### 6.2.3 Síntese dos resultados – A concordância verbal na Amostra 2020

Na década de 2020, os dados apresentaram 2596 ocorrências de concordância verbal com os pronomes *nós* e *a gente*. Desse total de dados, 2061 se referem ao uso do padrão emergente de concordância *a gente* sem *-mos* (79,4%), sendo o mais recorrente padrão de concordância empregado na Amostra 2020. Em seguida, observamos 498 ocorrências de padrão canônico de concordância *nós* com *-mos* (19,2%), sendo a segunda variante mais utilizada em nossa amostra. A estrutura não padrão *nós* sem *-mos* se realizou em apenas 25 ocorrências (1%), enquanto a variante *a gente* com *-mos* obteve um total de 12 dados (0,5%).

Após proceder um recorte binário, nos restaram 2224 dados de *a gente* sem *-mos* e de *nós* com *-mos*, apenas com dados relacionados a verbos do modo indicativo, dos quais foram submetidos aos testes estatísticos. O teste de seleção automática das variáveis recomendou a remoção das variáveis *sexo/gênero* e *conjugação verbal* do modelo de regressão final, se mostrando relevante para a análise de nossos dados de concordância verbal apenas os preditores *faixa etária*, *tipo de inquérito*, *tempo* e *tipo de paradigma verbal do modo indicativo* e *estrutura do verbo*.

O modelo de regressão logística, que analisa a interação das variáveis em um contexto multivariado, apontou inibição da estrutura padrão de concordância *nós* com *-mos* em todos os contextos analisados na década de 2020, motivado pela ampla expansão do padrão de concordância emergente na amostra mais atual, que passou a atuar e se consolidou nos mais diversos contextos linguísticos e sociais. O modelo logístico apontou ainda como ambientes que favorecem mais intensamente a realização de *a gente* sem *-mos* a faixa etária mais jovens (22 a 35 anos), inquéritos do tipo D2 (diálogo entre dois informantes), verbos no presente com paradigma igual e locuções verbais.

### 6.3 A concordância verbal nas décadas de 1990 e 2020: um estudo em tempo real de curta duração de tendência

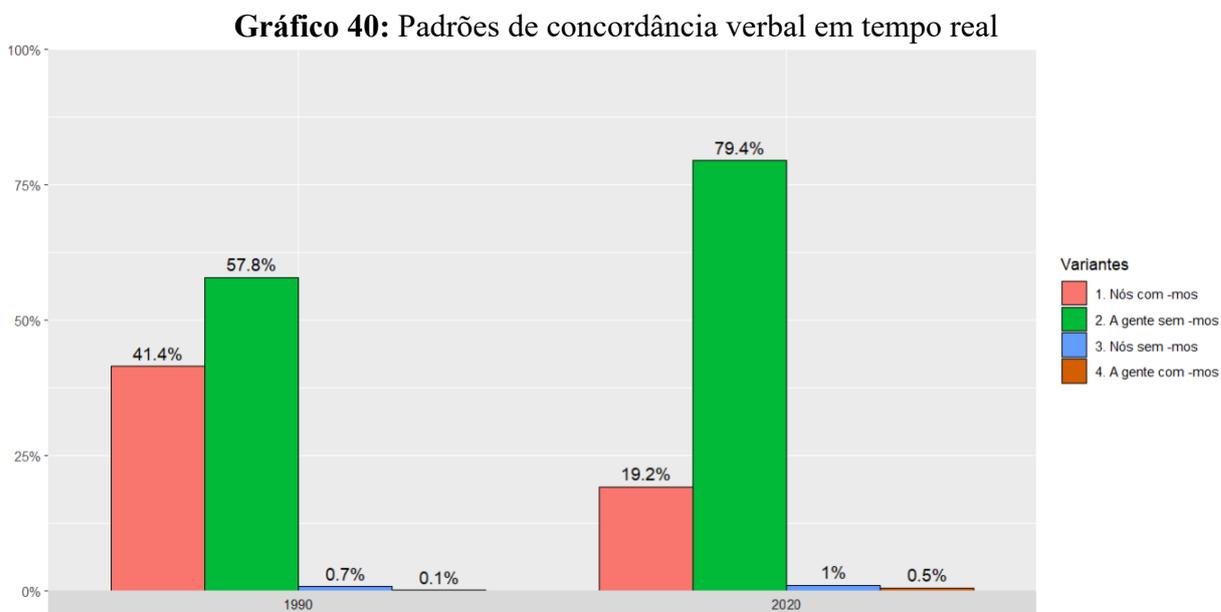
Para Labov (1994), a pesquisa sociolinguística em tempo real é uma abordagem metodológica que permite observar e estudar as dinâmicas da mudança linguística que envolve determinada comunidade de fala, permitindo ao pesquisador sociolinguística identificar de maneira mais eficaz como as práticas linguísticas evoluem e como os fatores sociais e linguísticos podem influenciar na mudança ao longo do tempo, fornecendo uma base empírica mais robusta para entender os processos de mudança linguística.

O parâmetro de análise linguística em tempo real se contrapõe à análise em tempo aparente, uma vez que, a medida em que se realizam estudos sociolinguísticos em diversas comunidades de fala, se evidenciam aspectos favoráveis e desfavoráveis ao estudo em tempo aparente, principalmente quando se trata de avaliar a questão da mudança linguística (Paiva e Duarte, 2003). Geralmente, se atribui indícios de mudança a partir da observação da distribuição das variantes pelas faixas etárias estudadas, quando os mais jovens usam variantes mais inovadoras e os mais velhos apresentam comportamento linguístico mais conservador. No entanto, nem sempre essa atribuição pode ser efetiva, tendo em vista que essa distribuição pode ser característica de uma gradação etária, o que torna difícil a tarefa de classificar uma mudança em curso.

Labov (1994) afirma que a associação das evidências encontradas pela observação de fenômenos linguísticos em tempo aparente e as evidências fornecidas em tempo real pode ser um caminho mais efetivo para observar determinada mudança linguística em progresso dentro de uma comunidade de fala. Nas palavras do autor a “combinação de observações no tempo aparente e no tempo real é o método básico para o estudo da mudança em progresso” (Labov, 1994, p.63). Neste sentido, tratamos de observar os padrões de concordância verbal de primeira pessoa do plural em duas sincronias, na mesma comunidade linguística, utilizando os mesmos procedimentos metodológicos empregados na construção das duas bases de dados, Amostra 1990 e Amostra 2020.

Em termos gerais, analisamos 4534 dados de concordância verbal com os pronomes *nós* e *a gente* em Fortaleza. Desse total, foram observados 1938 dados na década de 1990 e 2596 na década de 2020. Em termos de proporção de uso, percebemos redução significativa do padrão canônico *nós* com *-mos* de uma década para outra, ao passo que observamos expressivo aumento da proporção de uso do padrão emergência, e aumento discreto nas proporções de

concordância não padrão. Em termos de tendência, percebe-se um padrão no comportamento do fenômeno da concordância verbal na comunidade de fala, ou seja, uso mais acentuado do padrão emergente e do padrão canônico, e uso pouco recorrente de concordâncias não padrão, conforme aponta o gráfico 40:



Fonte: elaboração própria

Conforme é possível observar no gráfico 40, o uso do padrão de concordância *nós* com *-mos* (*nós cantamos*) diminuiu após quase 30 anos (de 41,4% para 19,2%). Por outro lado, confirmando nossa hipótese, o uso do padrão emergente *a gente* sem *-mos* (*a gente canta*) obteve um aumento significativo em sua proporção de uso na comunidade de fala, saltando de 57,8% para 79,4%, se destacando ainda mais entre os falantes cultos de Fortaleza. A estrutura não padrão *nós* sem *-mos* (*nós canta*) aumentou discretamente sua proporção de uso, de 0,7% para 1%, assim como a forma *a gente* sem *-mos* (*a gente cantamos*), que na década de 1990 apresentou proporção de uso igual a 0,1%, e na década de 2020 aumentou para 0,5%.

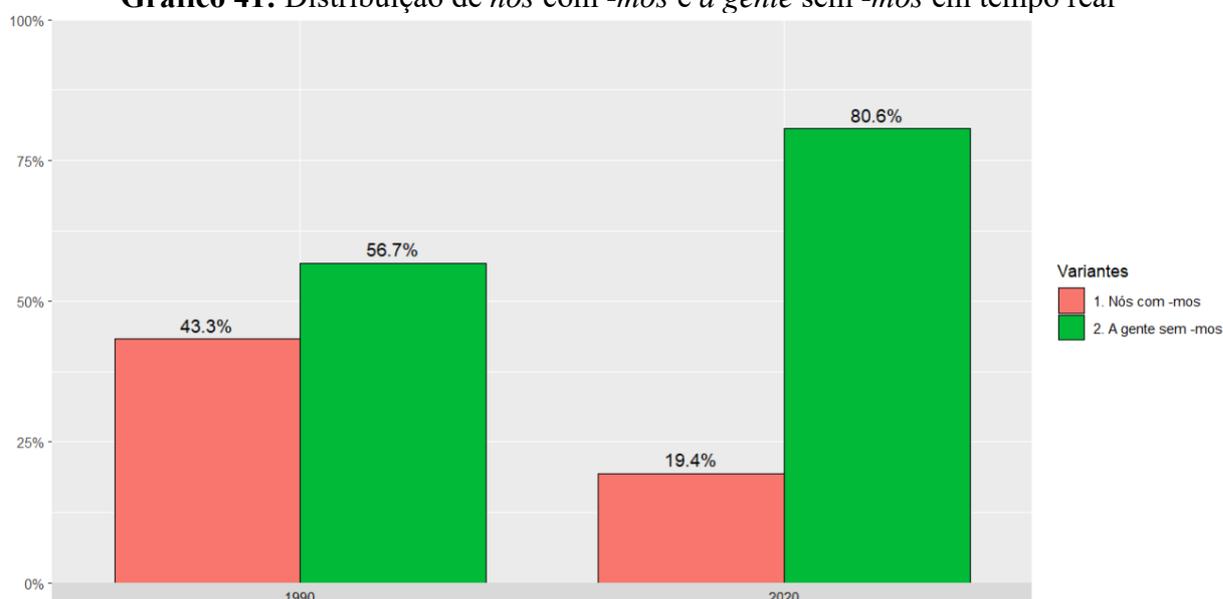
Embora se observe aumento no uso das estruturas não padrão de concordância verbal *nós* sem *-mos* e *a gente* com *-mos*, a tendência na amostra permaneceu após mais de duas décadas: ocorrências escassas de estruturas não padrão. Em nossa base de dados, o fato de considerarmos apenas informantes de nível superior completo pode ser um fator desfavorável às ocorrências de concordância verbal não padrão, tendo em vista que esse fenômeno está diretamente relacionado com a escolaridade, conforme explica Freitag (2016):

Diferentemente da variação entre *nós* e *a gente*, a variação na concordância é fortemente influenciada pela escolarização (Scherre & Naro 2014). A concordância é um domínio gramatical fortemente sensível à avaliação social no Brasil, e os resultados de estudos de covariação entre a alternância pronominal e a presença vs. ausência de marca morfêmica de concordância apontam para uma avaliação social estigmatizada para combinações como “*nós fala*” e “*a gente falamos*”, com base na configuração de padrões distribucionais sociodemográficos: ocorrências deste tipo costumam ser associadas a falantes menos escolarizados, como apontam Naro, Gorski & Fernandes (1999), que analisaram amostra de fala do Rio de Janeiro/RJ. (Freitag, 2016, p. 903)

Outros fatores também devem ser considerados, como a inserção de inquéritos do tipo EF (elocuições formais), que naturalmente tendem a inibir variantes avaliadas negativamente, e do tipo DID (diálogo entre informante e documentador), em que a própria presença do documentador e do equipamento de gravação tendem a inibir formas com marcas de estigmas social. Devido a poucas ocorrências de *nós* sem *-mos* e *a gente* com *-mos*, não foi possível proceder a análise multivariada, tendo em vista que as ocorrências de concordância verbal se concentram na variante de concordância emergente *a gente* sem *-mos* e no padrão de concordância mais antigo na língua, *nós* com *-mos*. Por conta disso, fomos levados a retirar os dados de concordância não padrão.

Para além disso, vale destacar que, os dados apresentados a partir deste ponto levam em conta apenas as ocorrências do modo indicativo, tendo em vista que apenas as ocorrências de concordância verbal associadas a este modo verbal foram analisadas no modelo de regressão logística (ver 6.1.2 e 6.2.2). Com a retirada dos dados de concordância não padrão, nos restaram 3810 dados analisados em um contexto multivariado, sendo 1586 ocorrências na década de 1990 e 2224 na década de 2020, distribuídas conforme o gráfico abaixo:

**Gráfico 41:** Distribuição de *nós* com *-mos* e *a gente* sem *-mos* em tempo real



Fonte: elaboração própria

A comparação em tempo real aponta para uma forte expansão no uso do padrão emergente de concordância *a gente* sem *-mos* na comunidade de fala, de 56,7% em 1900 para 80,6% em 2020, um aumento significativo. Por outro lado, é possível observar no gráfico 41 a redução significativa no emprego da concordância padrão *nós* com *-mos*, tendo seu uso reduzido pela metade em quase 30 anos, caindo de 43,3% para 19,4%. Embora seja notória a diferença na proporção de uso entre as duas formas de concordância verbal, não podemos simplesmente afirmar que existe mudança em curso, uma vez que a estrutura padrão permanece sendo utilizada e com certa expressividade, quase 20%. Nesse caso, o fenômeno permanece como uma regra variável, nos termos de Labov (1969), e a tendência se repete na comunidade de fala, ou seja, há uma disputa entre *nós* com *-mos* e *a gente* sem *-mos* pela preferência do padrão de concordância adotado na comunidade, com aumento significativo para o uso de *a gente* sem *-mos*.

Para Labov (1994), é necessário comparar os fatores mais relevantes em cada uma das sincronias analisadas e observar as tendências e o comportamento do fenômeno em função desses fatores. Submetemos todos os dados a seleção automatizada das variáveis, em um modelo unificando em uma só amostra os dados da década de 1990 e de 2020, e criando a variável *década de gravação*. Nosso objetivo é observar se a tendência geral da amostra manifesta se repete em cada década, assim como Lopes (2003), e investigar se a tendência observada na década de 1990 se repete na década de 2020. Os resultados podem ser conferidos no quadro 13:

**Quadro 13:** Confronto entre os resultados das análises geral e específicas de concordância

Análise geral	Década 1990 (Fase I)	Década 2010/2020 (Fase II)
1. Paradigma verbal	1. Paradigma verbal	1. Paradigma verbal
2. Década de gravação	2. Faixa etária	2. Tipo de inquérito
3. Faixa etária	3. Tipo de inquérito	3. Faixa etária
4. Tipo de inquérito	4. Conjugação verbal	4. Estrutura verbal
5. Estrutura verbal	5. Sexo/gênero	
6. Conjugação	6. Estrutura verbal	

Fonte: elaboração própria.

A hipótese da taxa constante (Kroch, 1989) teoriza que a influência global da amostra se repete nas amostras específicas. Como pode ser visto no quadro 13, em todas as amostras se destacou *tempo e tipo de paradigma verbal do modo indicativo*, apontando forte influência dessa variável previsora sobre os padrões de concordância verbal na comunidade de fala. Além disso, a seleção da variável *década de gravação* na amostra geral com todos os dados indica que o uso da concordância verbal é significativamente diferente em cada uma das décadas, o que pode indicar mudanças em curso. Além do mais, a variável previsora *faixa etária* pode indicar sinais de mudança, sendo selecionada nas três rodadas, apontando que o fenômeno apresenta comportamento diferente a depender da idade do informante.

O teste realizado com os dados da Amostra 2020 selecionou quatro preditores como significativos para a realização do fenômeno. Isso também pode ser indicativo para uma mudança em curso, pois, segundo Labov (2006), no desenvolvimento da mudança pode haver fatores que deixam de ter correlação com o fenômeno. No caso de nossa amostra, percebemos que dois fatores perderem efeito sobre a realização dos padrões de concordância: o *sexo/gênero* e a *conjugação verbal*. Isso indica que o padrão emergente adentrou em diversos contextos linguísticos e sociais, a ponto de não haver diferenciação entre homens e mulheres no uso dos padrões de concordância, nem tampouco da conjugação verbal.

Abaixo, apresentamos os resultados em termos de proporção de uso, de favorecimento e de tendência relacionadas ao uso da variante padrão de concordância verbal *nós* com *-mos*, *intercept* dos modelos, sempre contrapondo os resultados ao uso de *a gente* com *-mos*, padrão emergente. As variáveis analisadas em tempo real de tendências são *faixa etária*, *tipo de inquérito*, *tempo e tipo de paradigma verbal do modo indicativo* e *estrutura verbal*. A primeira apresentada é variável *faixa etária*.

#### a. Faixa etária

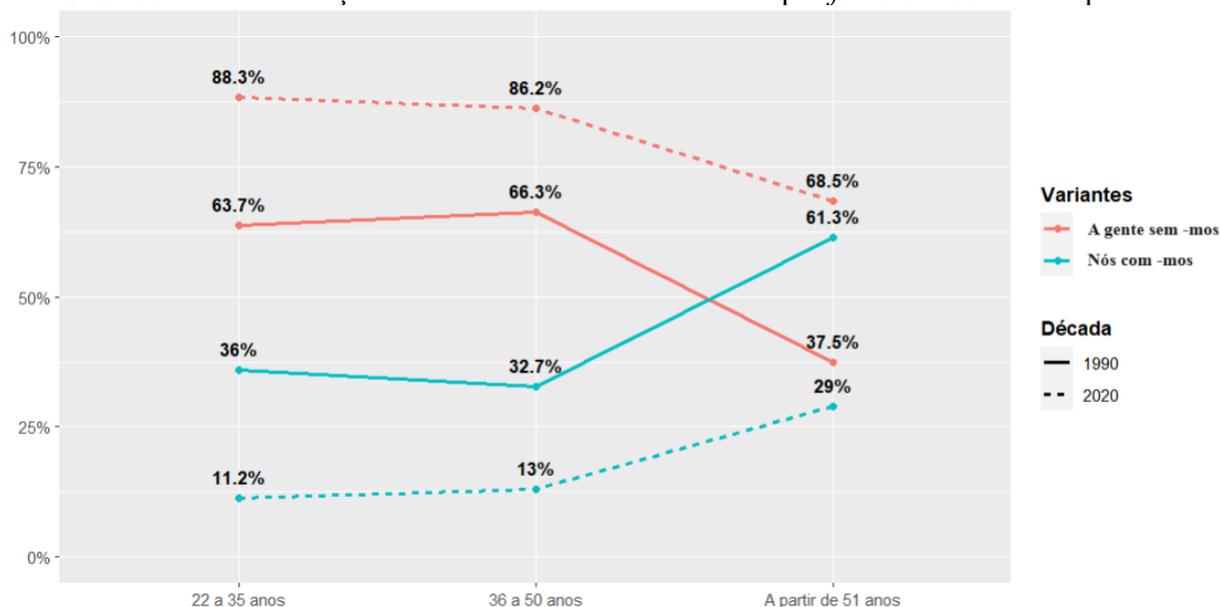
Um dos fatores de natureza social mais importantes quando se trata de mudança linguística é a *faixa etária*, tendo em vista que trabalhos sociolinguísticos constantemente associam mudança à idade dos informantes. Nosso objetivo ao confrontar o comportamento do fenômeno de concordância verbal em função da faixa etária em duas décadas é observar possíveis indícios de mudança linguística em tempo real ou de gradação etária. Abaixo, a tabela 61 apresenta o comparativo, em termos de proporção e de estimativa de uso, da variante de referência *nós* com *-mos* em tempo real.

**Tabela 61:** Atuação da *faixa etária* sobre o uso de *nós* com *-mos* em tempo real.

Faixa etária	Anos 1990			Anos 2020		
	Proporção	Estimativa	(*)	Proporção	Estimativa	(*)
22 a 35 anos	36%	-0.6989	***	11,2%	-1.62110	***
36 a 50 anos	32,7%	0.1799		13%	0.56341	**
A partir de 51 anos	61,3%	1.5180	****	29%	1.53756	***
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

Conforme é possível observar na tabela 61, o comparativo entre as duas décadas analisadas (1990 e 2020) aponta para uma expressiva retração no uso da variante padrão de concordância *nós* com *-mos* em todas as faixas etárias, o que indica que, em todas as idades, houve expansão no uso do padrão emergente *a gente* com *-mos*. Entre os mais jovens, por exemplo, há uma queda no uso da concordância padrão de 36% em 1990 para 11,2% em 2020. De forma parecida, os informantes de 36 a 50 anos de idade reduziram o uso de *nós* com *-mos* mais do que a metade em quase três décadas, caindo de 32,7% para 13% na amostra mais recente. Os informantes mais velhos, na década de 1990 eram os únicos a usarem preferencialmente o uso da concordância padrão, com 61,3%. Em 2020, houve mudança nesse comportamento, e os informantes com mais de 50 anos passaram a preferir o uso variante emergente, sendo a concordância padrão menos utilizada (29%). O gráfico 42 demonstra a proporção de uso de cada variante em função da faixa etária nas duas sincronias investigadas:

**Gráfico 42:** Distribuição das variantes de concordância por *faixa etária* em tempo real

Fonte: elaboração própria.

Conforme é possível observar, apesar de haver diminuição na proporção de uso de *nós* com *-mos* e consequente aumento de *a gente* com *-mos* em todas as faixas etárias, em termos de tendência, podemos apontar uma certa estabilidade na variação. Conforme demonstra gráfico 42, há altas proporções de uso do padrão emergente nas duas primeiras faixas etárias nos anos de 1990, e se observa uma queda significativa em seu uso na terceira faixa etária. O mesmo é observado na década de 2020. Mesmo que os percentuais tenham aumentado, as duas primeiras faixas etárias continuam com percentuais muito altos para o uso de *a gente* sem *-mos* e, na terceira faixa etária há uma queda significativa. Ou seja, as linhas estatísticas que demonstram a evolução no uso das variantes de concordância verbal em tempo real de tendência se assemelham, ambas apresentando uma curva estatística significativa no mesmo ponto.

A mesma tendência pode ser observada ao analisar as estimativas de uso. Os mais jovens, tendem a desfavorecer expressivamente *nós* com *-mos*, tanto na década de 1990 (-0.6989), quanto na década de 2020 (-1.62110), de forma mais acentuada na amostra mais atual. Os informantes de 35 a 50 anos, em relação aos mais jovens, desfavorecem com menor intensidade o padrão canônico de concordância, com estimativa de uso de 0.1799 na Amostra 1990, e de 0.56341 na Amostra 2020. Por fim, os mais velhos, em relação aos mais jovens, tendem a favorecer a estrutura padrão de concordância, com estimativa igual a 1.5180 em 1990, e 1.53756 em 2020.

Nesse sentido, as proporções de uso podem indicar para uma mudança a vista, uma vez que se percebe proporções próximas ao categórico para o padrão emergente entre os mais jovens, sugerindo a expansão gradativa de *a gente* sem *-mos*. Por outro lado, embora se observe significativa diminuição no uso de *nós* com *-mos* entre os informantes com mais de 50 anos, a proporção de uso continua relativamente alta na década de 2020, e a alta significância ( $p < 0,001$ ) pode sugerir que os mais velhos estejam resistindo à mudança linguística, o que está em conformidade com Labov (2006), que sugere que os mais velhos tendem a conservar formas mais tradicionais na língua.

#### *b. Tipo de inquérito*

O *tipo de inquérito* visa observar como o grau de formalidade pode interferir na escolha entre o uso de determinado padrão de concordância verbal, haja vista “a percepção do uso do *a gente* com concordância avaliado como informal e associado a um uso da periferia

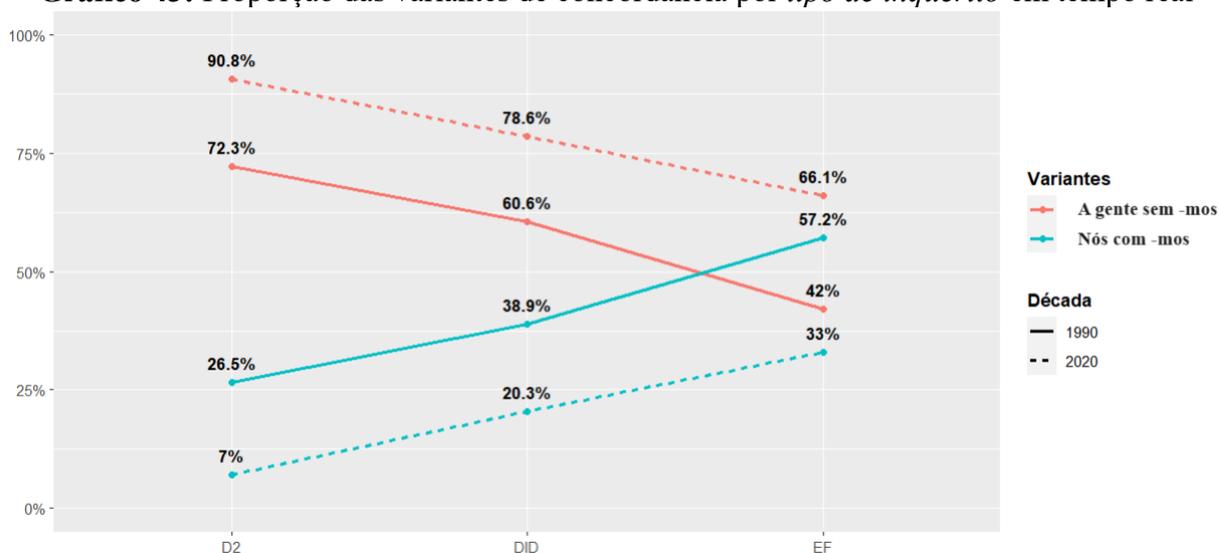
urbana” (Freitas, Favacho e Carvalho, 2022, p.48), o que tornaria ambientes menos formais propensos ao uso do padrão emergente, enquanto “o *nós* é a forma de prestígio, considerada escolarizada, contudo a ausência de concordância verbal (*nós* sem *-mos*) é avaliada negativamente” (Freitas, Favacho e Carvalho, 2022, p. 52). Nesse ponto, buscamos observar se as formas variantes adentraram e se expandiram independentemente da formalidade empregada no momento da elocução. A distribuição das proporções de *nós* com *-mos* em duas décadas pode ser conferida na tabela 62:

**Tabela 62:** Atuação do *tipo de inquérito* sobre o uso de *nós* com *-mos* em tempo real

Tipo de inquérito	Anos 1990			Anos 2020		
	Proporção	Estimativa	(*)	Proporção	Estimativa	(*)
DID	38,9%	-0.6989	***	20,3%	-1.62110	***
D2	26,5%	-0.5535	***	7%	-1.10347	***
EF	57,2%	0.8828	***	33%	0.78908	***
<b>Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1</b>						

Fonte: elaboração própria.

De acordo com os dados presentes na tabela 62, o uso do padrão de concordância *nós* com *-mos* diminuiu de uma década a outra em todos os tipos de inquéritos. Em inquéritos do tipo DID (diálogo entre informante e documentador), que corresponde ao *intercept* dos nossos modelos de regressão, o uso da concordância padrão caiu quase pela metade, saindo de 38,9% na década de 1990 para 20,3% em 2020. De forma ainda mais acentuada, o uso da concordância padrão *nós* com *-mos* reduziu em inquéritos do tipo D2 (diálogo entre dois informantes), que são os menos formais entre os registros investigados nesta pesquisa, tendo seu uso reduzido mais do que a metade (de 26,5% para 7%). Os inquéritos mais formais da amostra, os registros EF (elocução formal) apresentaram mudança no seu comportamento ao longo das décadas, ou seja, se observa preferência pelo uso da concordância padrão na década de 1990 (57,2%), reduzindo seu uso de forma expressiva em 2020 (33%), passando o padrão emergente a ser preferido nestes registros, conforme demonstra gráfico 23:

**Gráfico 43:** Proporção das variantes de concordância por *tipo de inquérito* em tempo real

Fonte: elaboração própria

Em termos de tendência, o gráfico demonstra que a atuação do fenômeno, embora tenha diminuído consideravelmente o uso de *nós com -mos* e aumentado expressivamente o uso de *a gente sem -mos*, se mantém semelhante nas duas décadas analisadas. Inquéritos menos formais (D2) são menos favoráveis ao uso do padrão de concordância *nós com -mos* tanto na Amostra 1990 (-0.5535), quando na Amostra 2020 (-1.10347), o que sugere forte interação entre a escolha da concordância e o grau de formalidade, refletindo uma tendência à adoção de formas mais inovadoras em ambientes menos formais. O percentual de uso da variante padrão vai aumentando conforme aumenta o grau de formalidade na década de 1990, e esse comportamento se repete na análise da década de 2020, conforme demonstra o gráfico 43.

Os inquéritos do tipo DID representam um grau intermediário de formalidade. Nestes registros, há desfavorecimento do padrão canônico de concordância, tanto em termos de proporção de uso, quanto em termos de favorecimento, com estimativa de uso de -0.6989 na década de 1990, e -1.62110 na década de 2020, mantendo a tendência ao desfavorecimento, porém de forma mais intensa que na década mais atual. Já os inquéritos EF, os mais formais das amostras, refletem uma tendência que se repete nas duas amostras analisadas: entre os inquéritos analisados, estes são os mais favoráveis ao uso do padrão *nós com -mos*. Embora haja expressivo aumento na proporção de uso do padrão emergente na década de 2020, esse resultado sugere resistência à mudança linguística em ambientes mais formais, onde variantes que detêm prestígio social são mais valorizadas.

c. *Tempo e tipo de paradigma verbal do modo indicativo*

A variável de natureza linguística *tempo e tipo de paradigma verbal* foi o preditor considerado de maior relevância sobre a realização dos padrões de concordância verbal, tanto na análise geral envolvendo os dados das duas décadas, quanto nas análises específicas realizadas para o tratamento dos dados em cada uma das sincronias. Nesse ponto, considerando sua importância estatística, é essencial observar quais contextos as variantes de concordância verbal de primeira pessoa ganharam ou perderam espaço, passaram a agir de modo a favorecer ou a inibir os seus usos, bem como buscar indícios de mudanças nos traços associados a cada uma das variantes. Como dito, assumimos o caráter funcional deste preditor, acreditando que a inserção do padrão emergente *a gente* sem *-mos* atuaria sobre a questão da ambiguidade temporal de verbos com mesma morfologia na relação presente/pretérito, por isso, o padrão de concordância *nós* com *-mos* seria inibido nestes contextos. Os dados referentes a análise em tempo real da ação da variável preditora *tempo e tipo de paradigma verbal do modo indicativo* podem ser conferidos na tabela 63:

**Tabela 63:** Atuação do *tempo e paradigma verbal* sobre *nós* com *-mos* em tempo real

Paradigma verbal	Amostra 1990			Amostra 2020		
	Proporção	Estimativa	<i>p-value</i>	Proporção	Estimativa	<i>p-value</i>
Presente diferente	52,6%	-0.6989	***	29,1%	-1.62110	***
Presente igual	27,9%	-0.8986	***	8,6%	-1.76226	***
Pret. perf. diferente	66,4%	0.8397	***	34,1%	0.29296	
Pret. Perf. igual	49,1%	-0.3583		30,6%	-0.01274	
Pret. imperfeito	25,6%	-0.9625	***	5,7%	-1.62600	***
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

Em todos os tempos e paradigmas observados, há diminuição na proporção de uso do padrão de concordância *nós* com *-mos* na década de 2020 se comparada à década de 1990, conforme é possível observar nos dados da tabela 63. Em contrapartida, há aumento expressivo de uso do padrão emergente *a gente* sem *-mos*, tendência observada na amostra como um todo. Em relação ao presente com formas diferentes (*intercept*), percebe-se diminuição na proporção de uso da forma canônica de 52,6% em 1990, para 29,1% em 2020. Quando o verbo está no presente e tem a mesma forma na relação presente/passado, há uma diminuição significativa na proporção, caindo de 27,9% para 8,6%, o que demonstra ser um contexto favorável ao uso do padrão emergente, de forma ainda mais acentuada na década de 2020, confirmando a nossa

hipótese de que a estrutura emergente pode estar atuando para desfazer a ambiguidade do verbo com mesma forma para ambos os verbos, especialmente no tempo presente.

Em relação ao pretérito perfeito com formas diferentes, este foi contexto em que houve retração significativa no uso do padrão de concordância *nós* com *-mos*, preferido entre os informantes na década 1990, teve a proporção de uso reduzida de 66,4% para 34,1% em 2020, apontando para expansão do padrão emergente. Na mesma direção, verbos com pretérito perfeito com formas iguais, que na década de 1990 eram ambientes de disputa acirrado pelo padrão preferido de concordância (49,1%), passam a ser menos favoráveis ao uso de *nós* com *-mos* (30,6%), destacando a ampla utilização de *a gente* sem *-mos*. Assim como os verbos no pretérito imperfeito, que na década de 1990 já eram desfavoráveis ao uso da concordância verbal canônica (25,6%), se tornaram ainda mais na década de 2020 (5,7%), se apresentando como contexto categórico para o uso do padrão emergente *a gente* sem *-mos*.

Em termos de tendências de uso, as estimativas, apresentadas em *loggods*, apontam para um padrão no uso das estruturas de concordância relacionadas ao *tempo e tipo de paradigma verbal do modo indicativo* que se assemelha na análise das duas décadas. Esse padrão foi o mesmo observado em Scherre, Yacovenco e Naro (2018): a marcação da desinência *-mos* está associada, de forma geral, ao pretérito perfeito, enquanto a ausência da desinência, ou seja, o verbo na terceira pessoa singular está associada ao presente, independente do paradigma. O presente com formas diferentes apresentou tendência a inibir a forma padrão *nós* com *-mos* na década de 1990 (-0.6989), e a intensificar ainda mais essa inibição na década de 2020 (-1.62110), o que confirma a hipótese das autoras, tendo em vista que este contexto é favorável ao uso do padrão emergente *a gente* sem *-mos*, que naturalmente está com o verbo na terceira pessoa do singular, sendo favorecido em tempo presente.

De forma parecida, os verbos no presente com formas iguais são intensamente desfavoráveis ao uso de *nós* com *-mos*, sendo o fator que mais desfavorece a concordância padrão entre todos os contextos verbais investigados neste estudo. Nesse contexto, a estrutura emergente pode estar, de fato, sendo beneficiada, pois estaria evitando a ambiguidade temporal que verbos com ambiguidade, pois a marcação da desinência *-mos*, responsável pela ambiguidade, cairia pelo uso do verbo na terceira pessoa do singular associado ao pronome *a gente*. Essa mesma tendência se assemelha à encontrada em Carvalho e Santos (2023, p.49), com dados de fala popular de Fortaleza da década de 1980, ao observar que a “estrutura padrão *nós* com *-mos* tem menor probabilidade de ocorrer com verbos no presente igual ao pretérito (cf. intercepto -2,1937) e com verbos no pretérito imperfeito (-3,3859)”. No caso da fala culta, as estimativas de uso apontam desfavorecimento na década de 1990, com estimativa de uso de

-0.8986, e inibição ainda mais acentuada na década de 2020, com estimativa igual a -1.76226, confirmando os trabalhos de Scherre, Yacovenco e Naro (2018), Carvalho, Freitas e Favacho (2020) e Carvalho e Santos (2023).

O pretérito perfeito, independente do tipo de paradigma, não teve efeito significado no modelo de regressão que analisou os dados da Amostra 2020, motivo pelo qual não será possível compará-lo em termos de favorecimento no contexto multivariado, embora já o tenhamos comprado em termos de proporção de uso as décadas de 1990 e de 2020. No entanto, o pretérito imperfeito se mostrou estatisticamente significativo como motivador dos padrões de concordância na análise das duas décadas. Na Amostra 1990, se observa forte inibição do uso da concordância padrão *nós* com *-mos* com verbos no pretérito imperfeito, com estimativa de uso de -0.9625, demonstrando que esses verbos estão fortemente associados ao padrão emergente de concordância *a gente* sem *-mos*. Na década de 2020, esse padrão se repete, ou seja, os verbos no imperfeito continuam fortemente desfavoráveis ao uso da concordância padrão, com estimativa igual -1.62600. Esses resultados podem indicar que, apesar das expressivas mudanças nas proporções de uso, a tendência no comportamento variável da concordância verbal permanece estável na comunidade de fala.

#### d. Estrutura verbal

A *estrutura verbal* foi a última variável previsora selecionada nos dois modelos de regressão constituídos para a análises dos dados referentes às décadas de 1990 e de 2020. Cabe destacar que na análise geral, a que engloba todos os dados, este previsor se destacou como relevante, o que demonstra que a tendência geral da amostra se repete a cada década. A tabela 64 apresenta os dados referente à análise em tempo real da atuação da variável *estrutura do verbo* sobre a realização da estrutura padrão de concordância *nós* com *-mos*:

**Tabela 64:** Atuação da *estrutura verbal* sobre o uso de *nós* com *-mos* em tempo real

Estrutura verbal	Anos 1990			Anos 2020		
	Proporção	Estimativa	(*)	Proporção	Estimativa	(*)
Simple	41,8%	-0.6989	***	19%	-1.62110	***
Locução verbal	39,4%	-0.6319	***	19,9%	-0.86771	***
Verbo composto	50%	0.2833		13,3%	0.31084	
<b>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</b>						

Fonte: elaboração própria.

As informações dispostas na tabela 64 evidenciam uma significativa redução na proporção de uso do padrão de concordância mais antigo na língua, em todos os contextos de estrutura verbal analisados. Esse fato indica ampla expansão no emprego do padrão emergente de concordância *a gente* sem *-mos*, que se implementou e se espalhou em diversos contextos linguísticos ao longo de quase trinta anos.

Em relação aos verbos com estrutura simples, percebe-se uma redução na proporção de uso da concordância padrão de 41,8% na década de 1990, para 19% na década de 2020, o que representa uma diminuição de mais da metade de sua proporção. De forma parecida, as locuções verbais se tornaram ambientes ainda mais desfavoráveis ao uso de *nós* com *-mos*, apresentando uma retração significativa de 39,4% para 19,9% na amostra mais atual. Os verbos compostos apresentaram uma redução ainda maior de uma década para a outra, caindo de 50% na Amostra 1990, para 13,3% na Amostra 2020, três vezes menos.

Em termos de tendência, as duas amostras se assemelham quando se trata de favorecimento. Os verbos com estrutura simples, que representam o *intercept* dos modelos analisados, demonstram tendência a inibir o uso da concordância padrão *nós* com *-mos* na década de 1990, com estimativa de uso de -0.6989, com ação inibidora ainda maior na década de 2020, em que apresenta estimativa igual a -1.62110. Essa intensificação no desfavorecimento de *nós* com *-mos* pode ser reflexo da alta proporção de uso do padrão emergente na Amostra 2020, que evoluiu consideravelmente.

Na mesma direção, as locuções verbais também apresentam comportamento parecido ante ao fenômeno da concordância verbal nas duas décadas. Ou seja, em relação ao *intercept*, na década de 1990 se observa forte tendência ao desfavorecimento da estrutura padrão, com estimativa igual a -0.6319, indicando esse contexto como favorável ao uso de *a gente* sem *-mos*. Na década de 2020 se observa a mesma tendência, porém de forma mais intensa: forte inibição do padrão *nós* com *-mos* que, em relação ao *intercept*, apresenta estimativa igual a -0.86771, e inclinação ao uso do padrão emergente. Os verbos compostos não apresentaram interação estatisticamente significativa sobre a realização dos padrões de concordância em qualquer das análises, motivo pelo qual não reportaremos seu resultado em termos de favorecimento.

### 6.3.1 Síntese dos resultados – A concordância verbal em tempo real de curta duração

A análise em tempo real contabilizou 4534 ocorrências de concordância verbal de primeira pessoa do plural em nossa amostra de fala culta fortalezense em duas sincronias. Na década de 1990, foram flagrados 1938 dados distribuídos da seguinte forma: *a gente* sem *-mos* (57,8%), *nós* com *-mos* (41,4%), *nós* sem *-mos* (0,7%) e *a gente* com *-mos* (0,1%). Na década de 2020, se notou 2596 dados de concordância verbal, com aumento significativo do padrão emergente, aumento discreto da concordância não padrão e diminuição expressiva no uso da estrutura padrão de concordância, distribuídos da seguinte forma: *a gente* sem *-mos* (79,4%), *nós* com *-mos* (19,2%), *nós* sem *-mos* (1%) e *a gente* com *-mos* (0,5%).

De modo geral, houve diminuição significativa da proporção de uso da concordância padrão *nós* com *-mos*, a variante de referência, em todos os contextos observados. Esse resultado tem relação direta com o significativo aumento na proporção de uso do pronome *a gente*, conforme constatamos ao analisar a alternância pronominal (ver seção 5.3). Em relação a *faixa etária*, se observa retração significativa entre os mais jovens, caindo de 36% para 11,2%, assim como os de faixa etária 51 +, que apresentaram redução na proporção de *nós* com *-mos* de 61,3% para 29%. O mesmo se observa em função da variável *tipo de inquérito*, em que o padrão canônico de concordância teve sua proporção de uso reduzida de forma significativa entre todos os tipos de registro, com destaque para os inquéritos menos formais (D2), que apresentou uma redução de 26,5% para 7% na década mais atual, demonstrando ser um ambiente praticamente categórico para o uso do padrão emergente *a gente* sem *-mos*. Por outro lado, em registros mais formais (EF), embora se observe redução no uso de *nós* com *-mos*, de 57,2% para 33%, este contexto parece ainda resistir a manutenção do padrão canônico de concordância.

O mesmo padrão de retração no uso da variante mais antiga na língua se observa na análise das variáveis de natureza linguística. Em relação *ao tempo e tipo de paradigma verbal do modo indicativo*, se nota significativa redução na sua proporção de uso associada a verbos no presente com morfologia diferente para pretérito perfeito e presente do indicativo (de 52,8% para 29,4%). De forma mais intensa, se percebe expressiva redução no uso de *nós* com *-mos* em verbos no presente com formas iguais para passado e presente, caindo de 27,9% na década de 1990 para 8,5% na década de 2020. Por fim, no que se refere a *estrutura do verbo*, tanto em verbos simples quanto em locuções verbais houve inibição da estrutura padrão de concordância *nós* com *-mos*, apontando para a ampla expansão do padrão emergente em todos os contextos analisados.

## 7 CONCLUSÃO

Nesta pesquisa, objetivamos analisar em tempo real de curta duração, à luz da teoria da variação e mudança linguística, os fenômenos variáveis que envolvem a primeira pessoa do plural, ou seja, os padrões de *alternância pronominal* e de *concordância verbal* com os pronomes *nós* e *a gente* em amostras de fala culta de Fortaleza, capital do estado do Ceará, a partir do banco de dados do Projeto Descrição do Português Oral Culto de Fortaleza - PORCUFORT, em duas décadas distintas, a saber, 1990 e 2020. Para descrever os fenômenos variáveis em questão, assumimos que a realização das variantes de primeira pessoa do plural (*nós* e *a gente*) e suas respectivas concordâncias padrão e não padrão não acontece de modo aleatório, mas é motivada por forças de natureza linguísticas e sociais, que condicionam a escolha entre os pronomes.

Considerando os pressupostos de Labov (1994), para conduzir um estudo de tendência considerando tempo real de curta duração, é necessário seguir alguns parâmetros metodológicos. Baseado na teoria da variação e mudança linguística, traçamos alguns objetivos que conduzem uma análise de curta duração, são eles:

- a) Analisar quais e como as variáveis de ordem linguística e extralinguística atuam de modo a favorecer o uso dos pronomes de primeira pessoa do plural e suas respectivas concordâncias em amostras de fala culta fortalezense da década de 1990;
- b) Investigar quais e como variáveis de caráter estrutural e social condicionam o comportamento variável da *alternância pronominal* e de *concordância verbal de primeira pessoa do plural* em amostra de fala culta fortalezense da década de 2020;
- c) Comparar, sempre que possível, os resultados obtidos a partir dos dados das amostras de 1990 e 2020, a fim de verificar as tendências que se mantiveram na fala culta de Fortaleza ao longo de mais de 20 anos.

O tratamento de um fenômeno variável em tempo real exigiu encaminhamentos específicos para a realização da análise. Entre esses encaminhamentos, tivemos que levantar questões de pesquisa, a partir da análise dos objetivos, bem desenhadas para dar conta de uma análise em tempo real de curta duração. A partir dos objetivos elencados acima, definidos para esta análise, traçamos três questões de pesquisa que englobam os fenômenos da *alternância pronominal* e da *concordância verbal de primeira pessoa do plural*:

- a) Quais e como as variáveis de ordem linguística e extralinguística atuam sobre o comportamento variável dos fenômenos da *alternância pronominal* e da concordância *verbal de primeira pessoa do plural* na amostra de fala culta representativa da década de 1990?
- b) Quais e como variáveis sociais e estruturais que atuam significativamente sobre a realização dos pronomes *nós* e *a gente* e suas respectivas concordâncias em amostra de fala culta fortalezense representativa da década de 2020?
- c) De que forma as tendências observadas na análise dos dados da década de 1990 se repetem ou se distanciam dos resultados obtidos com a análise dos dados da década de 2020?

A fim buscar explicações fundamentadas às questões que moveram a realização desta análise, construímos duas amostras de fala representativas da capital do Ceará. A primeira delas representa a variedade culta da língua da década de 1990 e foi composta por 50 informantes em inquéritos dos tipos DID, D2 e EF, extraídos do Projeto PORCUFORT Fase I. A segunda amostra de fala culta, por sua vez, é representativa da década de 2020 e corresponde a 54 informantes, advindos de inquéritos também dos tipos DID, D2 e EF, extraídos do Projeto PORCUFORT Fase II.

As amostras que compreendem o Projeto PORCUFORT, como tratado no desenvolvimento desta tese, foram construídas seguindo os pressupostos básicos para a composição de banco de dados de fala de caráter sociolinguístico, os mesmos procedimentos instituídos na criação do projeto Norma Urbana Culta - NURC, que é referência em bancos de dados sociolinguísticos. O primeiro banco de dados utilizado nesta pesquisa, o PORCUFORT Fase I, tem sua origem na década de 1990, enquanto o segundo foi constituído nos anos finais da década de 2010 e nos anos iniciais da década de 2020. De qualquer maneira, os dois *corpora* que compõem o projeto PORCUFORT são uma rica fonte de pesquisas, tanto em tempo aparente quanto em tempo real, os quais podemos extrair dados reais da língua em uso acerca dos mais diversos fenômenos variáveis que permeiam a variedade culta de língua de Fortaleza.

Em relação à análise de dados com o fenômeno da *alternância pronominal*, nos resultados obtidos para a amostra representativa da fala culta da década de 1990, verificamos ao longo dos 50 inquéritos do tipo DID, D2 e EF, um total de 1957 observações de primeira pessoa do plural, sendo 1140 ocorrências do pronome inovador *a gente*, e 817 dados do pronome canônico *nós*. A variante tida como inovadora (*a gente*) foi mais utilizada entre os informantes, com 58,3% das observações, ao passo que a variante padrão (*nós*) apresentou

menor proporção de uso, tendo 41,7% das ocorrências.

Em relação ao percentual de uso das variantes analisadas na variedade culta da capital cearense, a amostra representativa da década de 2020 revelou 2618 observações dos pronomes de primeira pessoa do plural. Desse total, 525 dados correspondem à realização do pronome *nós*, enquanto o pronome *a gente* foi registrado em 2093 observações. Notavelmente, a variante inovadora foi expressivamente mais utilizada que a variante padrão em nossa amostra do PORCUFORT da década de 2020. Os resultados gerais da análise, em termos percentuais, revelam significativa preferência pelo pronome inovador na língua, o pronome *a gente*, entre os fortalezenses de ensino superior completo, com 79,95% do total de dados. Em relação ao pronome mais antigo na língua, o pronome *nós*, considerado padrão, este se mostrou pouco recorrente em nossa amostra, tendo se realizado em 20,05% das ocorrências de primeira pessoa do plural.

É importante destacar que a predominância da variante inovadora confirma uma tendência já observada por outros estudiosos desse fenômeno. Diversos estudos variacionistas sobre a variação primeira pessoa do plural, realizados em diferentes localidades do Brasil, indicam que, de fato, quando se trata da linguagem falada por sujeitos com ensino superior completo, a variante inovadora *a gente* é usada em maior proporção do que a variante padrão *nós*. Uma das possíveis explicações para esse fato está relacionada à expansão da forma inovadora que, não sendo passível de estigma social, apesar de ser avaliada como informal, adentrou nos mais diversos contextos linguísticos e estratos sociais, e ganhou cada vez mais espaço na comunidade de fala (Peluco, 2022).

Além da proporção geral de uso dos pronomes *nós* e *a gente*, verificamos, por meio da ferramenta computacional *R Studio*, que a variante inovadora *a gente* é beneficiada nos seguintes contextos sociais na década de 1990:

- a) Faixa etária I - 22 a 35 anos (63,9%) e faixa etária II - 36 a 50 anos (66,6%);
- b) Tipo de inquérito - D2 (72,8%) e DID (61%).
- c) Grau de referencialidade dos pronomes - referência genérica (71,3%);
- d) Tipo de verbo - verbos de ação (68,9%), *dicendi* (69%), e epistêmico (70,6%);
- e) Paralelismo formal - primeira da série (56%), manutenção do mesmo pronome (59,7%) e referência isolada (60,6%);
- f) Tempo verbal e formas nominais do verbo - presente do indicativo (55,4%), pretérito imperfeito (72,9%), presente do subjuntivo (77,8%), futuro do pretérito (63,3%), gerúndio (100%) e infinitivo (82,5%);

- g) Saliência fônica - Nível 1 - (cantando) (100%), Nível 2 - (cantar/cantarmos) (82%), Nível 3 - (falava/ falávamos) (68,5%) e Nível 4 - (fala/ falamos) (69,9%).

Por outro lado, o pronome canônico *nós* é usado em maior proporção entre os informantes da faixa etária III - a partir de 51 anos (61%); e em inquéritos do tipo EF (57,7%).

Na análise com os dados representativos da década de 2020, em quase todos os contextos sociais e linguísticos o pronome inovador *a gente* é empregado com maior proporção de uso, sendo o pronome padrão *nós* favorecido, em termos de proporção de uso, apenas na variável linguística Paralelismo formal, quando há *quebra no paralelismo* (*nós*: 51,8%; *a gente*: 48,2%).

Em termos totais, houve um aumento significativo nas frequências e proporções da forma inovadora na amostra 2020, quando confrontada com a amostra 1990. Quando se compara o desempenho dos falantes, no curto intervalo de tempo pesquisado, a tendência de uso da forma *a gente* se mantém: há preferência pelo pronome *a gente* em ambas as amostras.

Um dado importante a destacar nos resultados obtidos nesta pesquisa é o fato de os informantes com mais de 50 anos, na amostra 1990, terem preferido o uso de *nós* em vez de *a gente*, sendo a única faixa etária pesquisada a apresentar esse comportamento linguístico. Em relação à amostra de 2020, no entanto, os informantes mais velhos (a partir de 51 anos) demonstraram mudança em seu comportamento, uma vez que deixaram de preferir o uso de *nós* (30,3%) e passaram a assumir o uso do *a gente* (69,7%).

A implementação de *a gente* se expandiu em todas as faixas etárias, inclusive entre os mais velhos da amostra. Essa expansão se mostrou também na variável *grau de referencialidade do pronome*: o uso do *a gente* com referência específica aumentou se compararmos o resultado das duas amostras, indicando que a forma inovadora se implementou e se expandiu em contextos em que o referente é específico.

Os dados da fala de informantes cultos de Fortaleza em tempo real revelam indícios de que estamos diante de um fenômeno que envolve mudança linguística em progresso. Mesmo que possamos observar certa estabilidade na variação entre *nós* e *a gente*, especialmente porque a tendência praticamente se mantém nas duas décadas, é possível observar alguns indicativos de mudança linguística em tempo real em curso. O primeiro deles é constatado logo na rodada inicial, a que envolve todos os dados duas décadas em uma única amostra “Dados Gerais” (Amostra 1990 e Amostra 2020). Ao analisar todos os dados em um mesmo conjunto, nesta rodada, criamos a variável “Década de gravação”, sendo a primeira variável destacada como significativa na análise inicial (ver quadro 11). Isso aponta que o fenômeno da alternância

entre os pronomes *nós* e *a gente* possui comportamento diferente em cada uma das décadas, sendo, de maneira geral, a forma *a gente* mais utilizada em 2020 que em 1990. Essa informação já é um indício de mudança linguística em progresso, sugerindo que a ampla expansão de *a gente* (de 58,25% para 79,95%) pode estar caminhando progressivamente para a substituição de *nós* na dala culta de Fortaleza. Cabe destacar que o mesmo fenômeno não é observado, por exemplo na escrita escolar de Fortaleza, que demonstrou uso expressivo do pronome padrão *nós* (89,4%) e menos recorrente do pronome inovador *a gente* (10,6%). Essa constatação demonstra que a mudança observada acontece no campo da língua falada, mas não na língua escrita, em que persiste o *nós* na comunidade de fala (Araujo, 2023).

Um outro indicativo tem relação com a variável *faixa etária*. Além da notória expansão do pronome *a gente* entre os estratos mais jovens na década de 2020, a mudança no comportamento da camada de maior faixa etária da comunidade (acima de 50 anos) é um forte indicativo de mudança linguística. Na década de 1990, esses informantes se mostravam mais conservadores diante do fenômeno, com baixa proporção de uso do pronome inovador *a gente* (37,7%). Já na década de 2020, os informantes com esse mesmo perfil, que resistiam ao uso de *a gente* na década passada, passaram a usá-lo preferencialmente (69,4%), demonstrando a forte adesão da forma pronominal inovadora em todos os ambientes sociais, o que é mais o indicativo de mudança linguística em andamento.

Percebe-se indícios de mudança linguística também atrelados à alguns traços intrínsecos característicos dos pronomes, como no grau de referencialidade por ele expresso. Alguns ambientes linguísticos eram prototípicos à realização de determinada variante. Na década de 1990, por exemplo, o pronome *a gente* estava associado a referencialidade genérica, sendo os contextos de referencialidade específica ambiente de restrição ao uso da forma inovadora. Na década de 2020, a expansão de *a gente* se deu em todos os contextos linguísticos, passando a predominar inclusive em contextos em que antes sofria inibição, como o de referencialidade específica. Essa mudança no traço de especificidade está associada a alta expansão de *a gente*, que atingiu todos os estratos da comunidade, apontando para uma mudança progressiva também constatado no Rio de Janeiro (Lopes, 2003) e em Salvador (Nascimento, 2013).

Embora tenhamos uma amostra limitada, em um intervalo de tempo considerado curto, em termos de tendência, este estudo demonstrou um padrão geral nas duas amostras, indicando para uma certa estabilidade da variação e mudança linguística das formas pronominais *nós/a gente* na comunidade de fala fortalezense em termos de tendência de uso, com mudança progressiva de *nós* por *a gente* em tempo real.

Em relação ao fenômeno da *concordância verbal com a primeira pessoa do plural*, a análise dos dados da Amostra 1990 revelou 1938 realizações de concordância verbal, sendo o padrão emergente *a gente* sem *-mós* preferido entre os informantes (57,8%). Em seguida, com proporção de uso expressiva, se destaca o padrão canônico *nós* com *-mos* (41,4%), seguindo dos dados de concordância não padrão *nós* sem *-mos* (0,7%) e *a gente* com *-mos* (0,1%). As baixas ocorrências podem estar relacionadas ao estigma que as formas não padrão recebem, sendo avaliadas como desprestigiadas. Associado a isso, o fato de tratarmos de uma amostra apenas com informantes cultos pode colaborar para as poucas ocorrências de concordância não padrão. Freitas, Favacho e Carvalho (2022), em estudos sobre percepção e avaliação dos padrões de alternância e de concordância verbal com falantes cultos de Fortaleza, demonstram que as formas de concordância não padrão são sensíveis à avaliação social. De acordo com as autoras:

A ausência de concordância verbal (*nós* sem *-mos* –*nós* fala/ *nós* come) recebe avaliação negativa, estando associada a um uso não escolarizado, informal e de periferia urbana, o que sugere ser esse um comportamento avaliativo explícito de escolarizados. A ausência de concordância verbal é um fenômeno acima do nível de consciência de falantes com escolarização. Freitas, Favacho e Carvalho (2022, p. 52)

Tendo em vista que nossa base de dados contempla informantes escolarizados, moradores da zona urbana da capital, é esperado que estes adotem preferencialmente formas de maior prestígio e evitem formas estigmatizadas, razão que explicaria o baixo número de dados dessa natureza. Além disso, de forma geral, realizações de concordância não padrão são variantes difíceis de serem flagradas em dados de fala, com estudos demonstrando pouca recorrência no português brasileiro (Freitag, 2016). É necessário considerar também que a concordância padrão, em algumas comunidades de fala do Brasil, é usada sem estigma, sendo marca de identidade, conforme demonstram os estudos de Mattos (2013) e Foeger (2024). No entanto, a comunidade de falantes cultos fortalezenses não parece adotar um padrão de concordância como marca de identidade social, dada a concentração dos dados nos padrões emergente e canônico.

O levantamento realizado por meio da ferramenta estatística determinou os seguintes contextos como favoráveis ao uso da estrutura padrão de concordância *nós* com *-mos* na década de 1990: Faixa etária III - a partir de 51 anos (60,2%); Tipo de inquérito - EF (56,9%). Em contrapartida, o padrão emergente *a gente* sem *-mos* é usado em maior proporção entre os informantes da faixa etária I - 22 a 35 anos (63,8%) e faixa etária II - 36 a 50 anos

(66,6%); e em inquéritos do tipo D2 (72,7%) e DID (60,8%). Já os preditores linguísticos que se mostraram favoráveis ao uso de *nós* com *-mos* nos dados da Amostra 1990 foram: *tempo e paradigma verbal do modo indicativo* - presente do indicativo com formas diferentes para o passado e o presente (52,5%) e pretérito perfeito com formas diferentes para o pretérito e o presente (66,4%); Por outro lado, o padrão emergente é fortemente favorecido em contextos de presente do indicativo com formas iguais (71,8%), em pretérito perfeito com formas iguais (50,6%), e pretérito imperfeito (72,2%). Vale destacar, esse resultado pode ser interpretado como uma confirmação a hipótese de Scherre, Yacovenco e Naro (2018), de que o pronome *a gente* poderia estar atuando como elemento para desambiguar os verbos com as mesmas formas para o presente e o passado, uma vez que o padrão emergente *a gente* sem *-mos* é preferido em contextos de verbos com potencial ambiguidade semântica-temporal. Esse resultado também reafirma o encontrado por Benfica (2014):

nos casos em que as formas do pretérito perfeito e presente são iguais, a marca *-mos* é reservada preferencialmente ao pretérito perfeito, em que a oposição singular/plural é mais saliente; nos casos em que as formas não são iguais, é bem frequente o uso do *-mos* também no presente do indicativo; nos casos de pretérito imperfeito, é onde se observa mais tendência de menos concordância para evitar a proparoxítone e se aproximar do padrão paroxítono do português brasileiro (Benfica, 2014, p.99).

Na variável *estrutura do verbo*, não se observou favorecimento do padrão *nós* com *-mos*, tendo apenas os verbos compostos representado disputa entre as formas de concordância (50%). Na mesma direção, todos os fatores da variável *conjugação verbal* apresentaram maior proporção de uso para o padrão de concordância emergente *a gente* sem *-mos*, sendo o padrão canônico menos recorrente em todos os contextos dessa variável preditora. Diferentemente da variável *tonicidade*, em que o padrão *nós* com *-mos* se mostrou beneficiado em paroxítonos (55%) e em proparoxítonos (97,3%), estando o padrão emergente associado a monossílabos (99,5%) e a oxítonos (97,7%).

Na década de 2020, houve uma redução significativa no uso do padrão de concordância *nós* com *-mos*, uma queda de 41% na Amostra 1990 para 19,1% na Amostra 2020. Essa expressiva diminuição na proporção de uso de *nós* com *-mos* está relacionado ao aumento significativo do pronome inovador *a gente* na mesma comunidade da fala, conforme demonstrou análise da variação na *alternância pronominal* (ver seção 5.1). A análise dos dados da Amostra 2020 verificou um total 2618 observações de concordância verbal com os pronomes de primeira pessoa do singular do plural. Desse total de dados, 2081 se referem ao uso do padrão emergente de concordância *a gente* sem *-mos* (79,5%), sendo o mais recorrente padrão de

concordância empregado na Amostra 2020. Em seguida, observamos 500 ocorrências de padrão canônico de concordância *nós* com *-mos* (19,1%), sendo a segunda variante mais utilizada em nossa amostra. Em nossos dados, a estrutura não padrão *nós* sem *-mos* (*nós canta*) se realizou em apenas 25 ocorrências (1%), enquanto a variante *a gente* com *-mos* (*a gente cantamos*) obteve um total de 12 dados (0,5%).

Em termos de favorecimento, a forma canônica sofreu retração no uso em praticamente todos os contextos analisados, com exceção da variável previsora *tonicidade*, em que os proparoxítonos permanecem associados categoricamente ao padrão de concordância *nós* com *-mos*, tendo 97.3% de ocorrência nesse contexto. Nos demais ambientes analisados, todos indicaram maior proporção de uso de *a gente* com *-mos*, em maior ou menor intensidade. Nesse ponto, percebe-se ambientes em que ainda se encontra algum tipo de resistência ao uso do padrão emergente *a gente* com *-mos*, seja por motivação social ou por restrição formal da própria língua.

Essa forte expansão no uso de *a gente* com *-mos* de uma década para outra em praticamente todos os contextos analisados destaca uma tendência já atestada em trabalhos da área (Omena, 2003; Lopes, 2003; Nascimento, 2013). Essa tendência aponta para uma mudança linguística em progresso, tendo em vista que estas pesquisas demonstram que o pronome inovador *a gente* se expande, em termos de proporção de uso, em cada década. Esse aumento, por um lado, pode confirmar a hipótese da taxa constante de Kroch (1989), tendo em vista que identificamos determinados padrões na taxa de uso das variantes em diversas pesquisas sobre o uso dos pronomes e suas respectivas concordâncias, o que pode levar a interpretação de que o uso do pronome está aumentando e em determinada taxa. Por outro lado, a mudança na proporção de uso na comunidade, embora razoavelmente uniforme, não pode ser considerada constante, tendo em vista que não tivemos como comparar a evolução dessa taxa ano após ano.

De qualquer forma, indícios de mudança podem ser evidentes. A concordância emergente *a gente* sem *-mos* se expandiu e parece continuar sua expansão no sentido de substituir a concordância padrão *nós* com *-mos*, *ao menos é o que sugere os percentuais de uso*. No entanto, é necessário analisar esses indícios de mudança linguística com cuidado, uma vez que o fenômeno da *concordância verbal de primeira pessoa do plural* sofre restrições estruturais do sistema linguístico. Por exemplo, em contextos de proparoxítono, não encontramos dados referentes ao padrão emergente *a gente* sem *-mos*, sendo este contexto verbal categórico ao uso de *nós* com *-mos*, dada a impossibilidade de reproduzir um proparoxítono ao usar o padrão emergente, tendo em vista que a ausência de verbos proparoxítonos conjugados na terceira pessoa do singular no português brasileiro. Por outro

lado, não existe a possibilidade de uso da estrutura padrão *nós* com *-mos* junto a verbos monossílabos, o que torna o ambiente categórico para o uso do padrão emergente na língua. Em uma possível mudança, não se sabe que rearranjo o sistema linguístico faria para driblar essas questões, ou talvez o próprio fenômeno da queda da proparoxítona resolveria essa uma das questões diante de uma mudança de *nós* para *a gente*. São questões importantes e que merecem ser investigadas com maior profundidade, uma vez que trariam respostas importantes sobre a natureza do fenômeno.

Contextos de natureza social também podem colaborar para a manutenção da forma *nós* com *-mos*, no sentido de inibir uma possível mudança linguística de *nós* com *-mos* para *a gente* sem *-mos*. Em relação à *faixa etária*, embora o padrão emergente tenha se destacado na década de 2020, ao comparar os padrões com os dados da Amostra 1990, percebe-se com clareza um padrão no uso dos pronomes distribuídos em função das faixas etárias. Na década de 1990, assim como na década de 2020 percebe-se baixa proporção de uso da concordância padrão *nós* com *-mos* entre os informantes mais jovens (22 a 35 anos), que aumenta à medida que aumenta a faixa etária, atingindo maior proporção de uso dentre os informantes com mais de 50 anos. Essa tendência foi observada nas duas amostras analisadas, sendo mais expressiva na amostra de 1990. Já nos dados de 2020, apesar de se observar menor proporção de uso do padrão *nós* com *-mos* em todas as faixas etárias, os informantes com mais de 50 anos foram os que se mostraram menos desfavoráveis ao uso do padrão de concordância canônico. Essa tendência pode apontar para a faixa etária como um favorecedor da concordância *nós* com *-mos*, tendo em visto que os informantes de maior faixa etária ainda resistem em assumir completamente a forma emergente, o que poderia restringir uma possível mudança linguística.

Outro fator que se mostrou significativo para o uso de *nós* com *-mos* e *a gente* sem *-mos* foi o *tipo de inquérito*. Divididos de acordo com o grau de formalidade que cada registro representa, os dados das duas décadas apontaram para a mesma tendência referente ao tipo de inquérito analisado. Ou seja, inquéritos menos formais (D2) são menos receptíveis a forma de concordância padrão (*nós cantamos, nós vendemos, nós partimos*), diferente dos inquéritos mais formais (EF), que apresentam tendência a aderir à variante canônica de concordância verbal. Esses resultados também podem ser relacionados ao teste de percepção e avaliação realizado com falantes cultos de Fortaleza (Freitas, Favacho e Carvalho, 2022, p.52), em que os informantes cultos “avaliam o *nós* como a forma de prestígio, formal, mais correta, refletindo a influência dos aparelhos formais de educação, que enfocam prioritariamente o quadro teórico dos pronomes pessoais”. Nesse sentido, o uso do pronome *nós* com *-mos*, por receber maior prestígio, tende a ser usado com mais recorrência em inquéritos mais formais (EF). Esse

contexto pode representar um ambiente de restrição à entrada do padrão emergente *a gente* sem *-mos* na comunidade de fala, uma vez que o pronome *a gente* ainda é avaliado como informal, enquanto o *nós* recebe status de variante de prestígio.

O fenômeno variável da primeira pessoa do plural é um fenômeno linguístico complexo e multifacetado, que apresenta muitas nuances e abordagens. É necessário considerar que o uso da *alternância* e da *concordância verbal de primeira pessoa do plural* apresentam motivações que podem ter origem linguística, social e, até mesmo, ideológica, sendo usada até mesmo como marca identitária (Mattos, 2013; Foeger, 2014). Outras abordagens sobre o fenômeno o observam sobre o viés funcionalista, assumindo que o pronome *a gente* pode ser usado como estratégia para indeterminar o sujeito, conforme sugerem Mendonça e Nascimento (2015), ou como instrumento desambiguar verbos com a mesma forma para o passado e para o presente, conforme defendem Scherre, Yacovenco e Naro (2018). Esta tese tratou de apresentar o efeito das variáveis sobre os fenômenos linguísticos, em uma abordagem descritiva dos resultados. Dada a complexidade do fenômeno de primeira pessoa do plural e as diversas lentes que sobre ele podem ser colocadas, convidamos os demais pesquisadores a contribuir com esta pesquisa, a partir de abordagens que explicam outras características do fenômeno variável da primeira pessoa do plural.

## REFERÊNCIAS

ÁLBAN, Maria del Rosário e FREITAS, Judith. Nós ou A gente? **Estudos Linguísticos e Literários**. nº 5, Salvador, UFBA. Agosto, 1991. p. 75-89.

ANTUNES, Marcello. **Estrangeiros compram mais títulos públicos**. Gazeta Mercantil, São Paulo, 19 abr. 2001

APPEL, René; MUYSKEN, Pieter. **Bilingüismo y contacto de lenguas**. Barcelona: Ariel, 1996.

ARAÚJO, L. E. S. A Variável faixa etária em estudos linguísticos. *Estudos Linguísticos* XXXVI, p.389-398, mai./ago., 2007. Disponível em: <<http://www.gel.org.br/estudoslinguisticos/edicoesanteriores/4publica-estudos-2007/sistema06/71.PDF>>. Acesso em: 03 mar. 2024.

ARAÚJO, Aluiza Alves de. In: CONGRESSO NACIONAL DE LINGÜÍSTICA E FILOSOFIA, 15, 2001a, Rio de Janeiro. **Anais... O Projeto Norma Oral Do Português Popular de Fortaleza – NORPOFOR**. Cadernos do CNLF. Rio de Janeiro: CiFEFiL, v. 5, p. 835-845, 2011. Disponível em: <[http://www.filologia.org.br/xv\\_cnlf/tomo\\_1/72.pdf](http://www.filologia.org.br/xv_cnlf/tomo_1/72.pdf)>. Acesso em: 12 jan. 2021.

ARAÚJO, Silvana Silva de Farias. *A concordância verbal no português falado em Feira de Santana-Ba: sociolinguística e sócio-história do português Brasileiro*. 341 f. il. 2014. Tese (Doutorado) – Instituto de Letras. Universidade Federal da Bahia, Salvador, 2014.

ARAÚJO, Aluiza Alves de; VIANA, Rakel de Bezerra Macedo; PEREIRA, Lidiane de Sousa Pereira. O banco de dados NORPOFOR. In: ARAÚJO, Aluiza Alves de; VIANA, Rakel de Bezerra Macedo; PEREIRA, Lidiane de Sousa Pereira (org.). **Fotografias sociolinguísticas do falar de Fortaleza – CE**. Fortaleza: EdUECE, 2018, p. 183-65.

ARAÚJO, Aluiza Alves de; VIANA, Rakel de Bezerra Macedo; PEREIRA, Lidiane de Sousa Pereira. O projeto descrição do Português Oral Culto de Fortaleza – PORCUFORT: das origens aos dias atuais. **Sociodialeto**, Moto Grosso do Sul, v. 8, n. 24, p.174-198, 2018<sup>a</sup>

ARAÚJO, Aluiza Alves de; VIANA, Rakel de Bezerra Macedo; RODRIGUES, L. S. **O falar culto de Fortaleza em foco**. São Paulo: Pimenta Cultural, 2021.

ARAÚJO, Marden Alyson Matos de. **Será que a gente usa mais o nós?** Uma fotografia sociolinguística do falar popular de Fortaleza. 2016. 148 f. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-Graduação em Linguística Aplicada, Universidade Estadual do Ceará, Fortaleza, 2016.

ARAÚJO, M. A. M. de. Nós e a gente no falar dos fortalezenses. In: ARAÚJO, A. A. de; VIANA, R. B. de M.; PEREIRA, M. L. de S. (Orgs.). **Fotografias sociolinguísticas do falar de Fortaleza – CE**. Fortaleza: EdUECE, 2018.

AZEREDO, José Carlos de. *Gramática Houaiss da Língua Portuguesa*. São Paulo: Publifolha, 2012.

BAGNO, Marcos. **A norma oculta: língua e poder na sociedade brasileira**. São Paulo: Parábola Editorial, 2003.

BECHARA, Evanildo. **Moderna gramática portuguesa**. 37 ed. Rio de Janeiro: Nova Fronteira, 2009

BORGES, Paulo Ricardo Silveira. **A gramaticalização de a gente no português brasileiro: análise histórico-social linguística da fala das comunidades gaúchas de Jaguarão e Pelotas**. 2004. 227 f. Tese (Doutorado em Letras). Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

BRASIL. Ministério da Educação. **Referencial curricular nacional para a educação infantil**. Brasília: MEC/SEF, 1998.

BRUSTOLIN, Ana Kelly Borba da Silva. **Itinerário do uso e variação de nós e a gente em textos escritos e orais de alunos do Ensino Fundamental da Rede Pública de Florianópolis**. 2009. 232 f. Dissertação (Mestrado em Linguística). Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2009.

BUTZGE, Clóvis. Aspectos sociolinguísticos em propagandas radiofônicas. In: **Especialização em língua, literatura e ensino**. v. III, p. 1-14, 2009

CALLOU, Dinah. O projeto NURC no Brasil: da década de 70 à década de 90. **Revista de Linguística** (impresa). São Paulo-SP, v. 11, 1999. p. 231-250.

CAMACHO, Roberto Gomes. A relevância social da sociolinguística: o efeito de escolaridade na marcação de número. **Caderno de Estudos Linguísticos**. nº 58.3, Campinas, pp. 461-479 - set./dez. 2016.

CARVALHO, H. M. de; FREITAS, M. L.; FAVACHO, L. de L. A variação dos pronomes sujeitos nós e a gente: a fala culta de Fortaleza em cena. **Revista (Con)Textos Linguísticos**, Vitória, v. 14, n. 27, p. 30-45, 2020. Disponível em: <https://periodicos.ufes.br/contextoslinguisticos/article/view/29213>. Acesso em: 10 fev. 2023.

CEZÁRIO, Maria Maura; VOTRE, Sebastião; COSTA, Marcos Antônio. In: MARTELOTTA, Mário Eduardo (Org.). **Manual de linguística**. São Paulo: Editora Contexto, 2008, p. 141-155.

COELHO, Izete Lehmkuhl; SOUZA, Christiane Maria N. de; GÖRSKI, Edair Maria; MAY, Guilherme Henrique. **Sociolinguística**. Florianópolis: LLV/CCE/UFSC, 2010.

COELHO, Izete Lehmkuhl; SOUZA, Christiane Maria N. de; GÖRSKI, Edair Maria; MAY, Guilherme Henrique. **Para conhecer sociolinguística**. São Paulo: Contexto, 2015.

CORBUCCI, Paulo. **O Ensino Superior Brasileiro na Década de 90**. Políticas Sociais - acompanhamento e análise. Brasília: IPEA, 2001.

CUNHA, Celso; CINTRA, Luís Felipe Lindley. **Nova gramática do português contemporâneo**. 5. ed. Rio de Janeiro: Lexikon, 2008.

CYRANKA, Lúcia Furtado de Mendonça; RONCARATI, Cláudia. Atitudes Linguísticas: uma pesquisa em escolas públicas de Juiz de Fora (MG-Brasil). In: Congresso Internacional da ABRALIN, 5., 2009, João Pessoa-PB. **Anais...** João Pessoa: IDEIA, 2009. v. 1. p. 1-20.

FERNANDES, Eliene. Fenômeno variável: nós e a gente. In: HORA, Dermeval (org.). **Estudos sociolinguísticos: perfil de uma comunidade**. Santa Maria: Pallotti, 2004. p. 149-156.

FIGUEIREDO, Candido de. **Novo dicionário da língua portuguesa**. 2.ed. Lisboa: Liv. Ed. Tavares Cardoso & Irmão, 1913.

FREITAG, Raquel Meister Ko. **Idade: uma variável sociolinguística complexa**. Línguas & Letras, v. 6, pp. 105-121, 2005.

FREITAG, Raquel Meister Ko. et al. Avaliação e variação linguística: estereótipos, marcadores e indicadores em uma comunidade escolar. In: FREITAG, R. M. K.; SEVERO, C. G.; GORSKI, E. M. (org.). **Sociolinguística e política linguística: olhares contemporâneos**. São Paulo: Edgard Blücher, 2016. p. 141-160

FREITAS, Judith. “Nós” e “a gente” em elocuições formais. **Estudos Linguísticos e Literários**. Salvador: UFBA, n. 11, p. 91-102, 1991.

FREITAS, Maylle Lima; RODRIGUES, Lorena da Silva; SANTOS, Hugo Leonardo Gomes. Nós e a gente no falar culto de Fortaleza: variação ou mudança?. In: ARAÚJO, Aluiza Alves de; VIANA, Raket de Bezerra Macedo; RODRIGUES, L. S. (org.). **O falar culto de Fortaleza em foco**. São Paulo: Pimenta Cultural, 2021.

GONÇALVES, Cintia Schivinski; BRESCANCINI, Cláudia Regina. Considerações sobre o papel da sociofonética na comparação forense de locutores. **Linguagem e Direito**, Porto, v. 1, n. 2, p. 67-87, 2014.

GUIMARÃES, Tatiane de Almeida Araújo Studart de. **Tu e você no falar de Fortaleza: variação e avaliações linguísticas**. 2019, 219f. Tese (Doutorado em Linguística Aplicada) Programa de Pós-Graduação em Linguística Aplicada, Universidade Estadual do Ceará, Fortaleza, 2019. Disponível em: <<http://bdtd.ibict.br/vufind/Record/UECE0013154779951bd62ce7c3ec96e68b1d7>>. Acesso em: 13 jun. 2021.

GUY, Gregory Riordan; ZILLES, Ana Maria Stahl. **Sociolinguística Quantitativa: instrumental de análise**. São Paulo: Parábola Editorial, 2007.

HOPPER, Paul; TRAUGOTT, Elizabeth Closs. **Grammaticalization**. Cambridge University Press, 1993.

HYMES, Dell. "Competence and performance in linguistic theory" **Acquisition of**

**languages: Models and methods.** Ed. Huxley and E. Ingram. New York: Academic Press. 1971.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 1991** - Características Gerais da população e Instrução: resultados da Amostra. n. 11. Rio de Janeiro, 1991. 133 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Contagem da População 1996.** Rio de Janeiro, 1997. 70 p. Disponível em: <<http://biblioteca.ibge.gov.br/visualizacao/livros/liv26412.pdf>> Acesso em: 31 out. 2021.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 2019.** Brasília: Inep, 2020. Disponível em: <<http://portal.inep.gov.br/basica-censo-escolar-sinopse-sinopse>>. Acesso em: 13 de out. 2021.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 1995.** Brasília: Inep, 1996. Disponível em: <[https://download.inep.gov.br/informacoes\\_estatisticas/sinopses\\_estatisticas/sinopses\\_educacao\\_superior/sinopse\\_educacao\\_superior\\_1995.xls](https://download.inep.gov.br/informacoes_estatisticas/sinopses_estatisticas/sinopses_educacao_superior/sinopse_educacao_superior_1995.xls)>. Acesso em: 16 de nov. 2021.

LABOV, Willian. **Padrões Sociolinguísticos.** Tradução de M. Bagno, Maria M. P. Scherre, Caroline R. Cardoso. São Paulo: Editora Parábola Editorial, 2008 [1972].

LABOV, Willian. The social stratification of English in New York city. 2 ed. Cambridge: University Press, 2006 [1966]. Disponível em: <<http://idiom.ucsd.edu/~bakovic/variation/Labov-2006.pdf>>. Acesso em: 23 abr. 2021.

LAMBERT; W. E.; HODGSON, R. C.; GARDNER, R. C.; FILLENBAUM, S. Evaluation reactions to spoken languages. **Journal of Abnormal Social Psychology**, n. 60, p. 44-51, 1960.

LEITE, Marli Quadros. Variação linguística: dialetos, registros e norma linguística. In: SILVA, Luiz Antônio (Org.) **A língua que falamos.** Português: história, variação e discurso. São Paulo: Globo, 2005. p. 183-210.

LOPES, Célia Regina dos Santos. **“Nós” e “a gente” no português falado culto do Brasil.** Rio de Janeiro. 1993. 82f. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1993.

LOPES, Célia Regina dos Santos. **A inserção de ‘a gente’ no quadro pronominal do português.** Frankfurt am Main/Madrid: Vervuert/Iberoamericana, 2003, v.18. 174 f. Disponível em: <<https://laborhistorico.letas.ufrj.br/producao/Lopetese.pdf>> . Acesso em: 21 maio 2021.

MACHADO, Márcia dos Santos. **Sujeitos pronominais “nós” e “a gente”:** variação em dialetos populares do norte fluminense”. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1995.

MAIA, Francisca Paula Soares. **A variação nós e a gente no quadro pronominal do português: percurso histórico**. 2003. 144 f. Dissertação (Mestrado em Letras) - Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2003.

MARENCO, Sandro Marcio Drumond Alves; SOUZA, Natália Larizza Sanches de; FONSECA, Mariana Augusta da Conceição; "Linguística Forense em Diacronia: Apontamentos Terminométricos da Variação Defloramento/Estupro no Sergipe Oitocentista". In: **Língua e Sociedade: Diferentes Perspectivas, Fim Comum**. São Paulo: Blucher, 2019. p.147-166.

MARTINS, Carlos Benedito. O ensino superior brasileiro nos anos 90. **São Paulo em Perspectiva**, São Paulo, v. 14, n. 1. p. 41-60, jan./mar. 2000.

MATTOS, Shirley Eliany Rocha. **Goiás na primeira pessoa do plural**. 2013. 137 f. Tese (Doutorado em Linguística) - Faculdade de Letras, Programa de Pós-Graduação em Linguística, Universidade de Brasília, Brasília, 2013.

MEILLET, Antoine. **Linguistique historique et linguistique générale**. Paris: Edouard Champion, 1912.

MALTEFORT, Mitchell. Unable to fit model using "lrm.fit". ETH Zurich, 2012. Disponível em: <https://www.stat.math.ethz.ch/pipermail/r-help/2012-May/314080.html>. Acesso em: 27 maio 2024.

MENDES, Rute Paranhos Silva. **O perfil da alternância do sujeito nós e a gente em Santo Antônio de Jesus: um recorte do português popular no interior da Bahia**. 2007. 140 f. Dissertação (Mestrado em Letras) - Instituto de Letras, Programa de Pós-Graduação em Letras, Universidade Federal da Bahia, 2007.

MENDONÇA, Alexandre Kronemberger de. **Nós e a gente em Vitória: análise sociolinguística da fala capixaba**. 2010. 98 f. Dissertação (Mestrado em Estudos Linguísticos) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal do Espírito Santo, Vitória, 2010.

MENON, Odete Pereira da Silva. A gente: um processo de gramaticalização. In: Estudos linguísticos. **XXV Anais dos Seminários do GEL**. Taubaté: UNITAU, p. 622-628, 1996.

MENON, Odete Pereira da Silva. A gente, eu, nós: sintomas de uma mudança em curso no português do Brasil? **Anais do ELFE**. Maceió: UFAL: 1995, p. 397-403.

MENON, Odete et al. Alternância nós/ a gente nos quadrinhos: análise em tempo real. In: RONCARATI, Cláudia; ABRAÇADO, Jussara. (org.). **Português brasileiro: contato lingüístico, heterogeneidade e história**. Rio de Janeiro: 7Letras, p. 96-105, 2003.

MORENO FERNÁNDEZ, F. **Principios de sociolingüística y sociología del lenguaje**. Barcelona: Ariel, 1998.

MUFWENE, Salikoko. **Language birth and death**. Annual Review of Anthropology 33, 2004, p. 201-222.

NARO, Anthony Julius; GÖRSKI, Edair; FERNANDES, Eulalia. Change without Change. **Language Variation and Change**. v. 11, nº 2, New York, 1999, p. 197- 211

NARO, Anthony Julius. Modelos quantitativos e tratamento estatístico. *In*: MOLLICA, Maria Cecília; BRAGA, Maria Luiza. (org.). **Introdução à sociolinguística: o tratamento da variação**. São Paulo: Editora Contexto, 2004. p.15-31.

NARO, Anthony J; SCHERRE, Maria Marta Pereira. General principles governing variation in Brazilian Portuguese. *In*: **New Ways Of Analyzing Variation** 45, nov. 2016. Vancouver: Simon Fraser University e University of Victoria

NASCIMENTO, Carina Sampaio. **Nós e A gente em Salvador: confronto entre duas décadas**. 2013. 128f. Dissertação (Mestrado em Letras) - Programa de Pós-Graduação em Língua e Cultura do Instituto de Letras, Universidade Federal da Bahia, Salvador, 2013.

NUNES, José Joaquim. (1906) **Crestomatia arcaica** – Excertos da literatura portuguesa desde o que mais antigo se conhece até ao Século XVI. Reedição: Lisboa, Ed. Clássica. 1970.

OLIVEIRA, Karoline Abrantes Oliveira. **Variação entre nós e a gente na fala da apresentadora do programa Resenhas do RN (2019-2020)**. 2022. 93 f. Dissertação (Mestrado em Estudos da Linguagem) – Departamento de Letras, Universidade Federal do Rio Grande do Norte, Natal, 2022.

OMENA, Nelize Pires de. A referência à primeira pessoa do plural. *In*: SILVA, Giselle Machline de Oliveira; SCHERRE, Maria Marta Pereira. (org.). **Padrões Sociolinguísticos**. Rio de Janeiro: Tempo Brasileiro, 1996. p. 185-215.

OMENA, Nelize. P. de A. Referência à primeira pessoa do discurso no plural. *In*: SILVA, M. O.; SCHERRE, M. M. P. (Orgs.). **Padrões sociolinguísticos**. Análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro. UFRJ, 1998a, p. 185-215.

OMENA, Nelize. P. de A. As influências sociais na variação entre nós e *a gente* na função de sujeito. *In*: SILVA, M. O.; SCHERRE, M. M. P. (Orgs.). **Padrões sociolinguísticos**. Análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro. UFRJ, 1998b, p. 311- 323.

OMENA, Nelise Pires de. (1986): “A referência variável da primeira pessoa do discurso no Plural”, *in*: NARO, Anthony Julius *et alii*: **Relatório Final de Pesquisa: Projeto Subsídios do Projeto Censo à Educação**, Rio de Janeiro, UFRJ, 2:286–319.

OMENA, Nelize. P. de A. Referência à primeira pessoa do plural: variação ou mudança? *In*: PAIVA, M. C.; DUARTE, E. L. (Orgs.). **Mudança em tempo real**. Rio de Janeiro: Capa Livraria, 2003, p. 63-80

PAIVA, Maria da Conceição de; DUARTE, Maria Eugênia Lamoglia. Mudança linguística: observações no tempo real. *In*: MOLLICA, Maria Cecília; BRAGA, Maria Luiza. (org.). **Introdução à sociolinguística: o tratamento da variação**. São Paulo: Editora Contexto, 2004. p. 179-190.

PENA, Luís Carlos Martins (1815-1848): **As Melhores Comédias de Martins Pena**, Apres. Guilhermino César, Rio de Janeiro, Mercado Aberto, 1987.

PHILIPPSEN, Neusa Inês; SANTOS, Josilene Pereira dos. **O uso de nós e a gente em cinco comunidades rurais de Terra Nova do Norte/Mato Grosso/Brasil**. Polifonia, Cuiabá-MT, v. 26, n.41, p. 01-188, janeiro-março, 2019.

REIGHARD, John. Contraintes sur le changement syntaxique. **Cabier de Linguistique**, Québec, v. 8. p. 407-436, 1978.

ROCHA, Fernanda da Cunha Faria. **A alternância nos pronomes pessoais e possessivos do português de Belo Horizonte**. 2009. 107f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística e Língua Portuguesa, Faculdade de Letras, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2009.

RODRIGUES, Luciana; MEDEIROS, Cleyber Nascimento de; LIMA, Laislânia Holanda de. **Perfil Municipal de Fortaleza: Tema XI: Perfil do Analfabetismo nos Bairros**. Informe nº 44. Fortaleza: IPECE, 2012.

ROCHA LIMA, Carlos Henrique da. **Gramática Normativa da Língua Portuguesa**. 47. ed. Rio de Janeiro: José Olympio, 2008.

SALDANHA, José Herand Diógenes; BENEVIDES, Marinina Gruska. As Políticas de Habitabilidade em Fortaleza: breves relatos dos diferenciais das gestões de Juraci Magalhães e Luizianne Lins. **Conhecer: debate entre o público e o privado**, [S. l.], v. 3, n. 08, p. 161–195, 2013.

SANKOFF, David. Sociolinguistics and syntactic variation. In: Newmeyer, Frederick J. (Ed.) **Linguistics: the Cambridge survey**. Volume IV (Language: the socio-cultural context). New York, Cambridge University Press, 1988.

SANKOFF, David; TAGLIAMONTE, Sali A.; SMITH, Eric. **Goldvarb X - A multivariate analysis application**. 2005. Toronto: Department of Linguistics; Ottawa: Department of Mathematics. 2005. Disponível em: <[http://individual.utoronto.ca/tagliamonte/Goldvarb/GV\\_index.htm#ref](http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm#ref)>. Acesso em: 02 out. 2021.

SANTANA, A. M. **Nós e A gente: um retrato do português popular de Salvador**. 2014. Dissertação (Mestrado em Letras) – Departamento de Ciências Humanas/Colegiado de Letras, Universidade Estadual da Bahia, Salvador, 2014

SANTOS, Rodrigo Medeiros dos; BRANCHES, Messias Viana. **Problemas identificados em gráficos estatísticos publicados nos meios de comunicação**. Revista Brasileira de Ensino de Ciência e Tecnologia, v. 11, n. 3, p. 1-18, 2018. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/7004331.pdf> . Acesso em: 11 out. 2021.

SAUSSURE, Ferdinand de. **Curso de lingüística geral**. Organização de Charles Bally e Albert Secheyay com a colaboração de Albert Riedlinger. Trad. de Antônio Chelini, José Paulo Paes e Izidoro Blikstein. 24ª ed. São Paulo: Pensamento-Cultrix, 2002.

SCHERRE, Maria Marta Pereira; NARO, Anthony Julius. Análise quantitativa e tópicos de interpretação do Varbrul. *In*: MOLLICA, Maria Cecília; BRAGA, Maria Luiza. (org.). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, 2004. p. 147-178.

SCHERRE, Maria Marta Pereira; NARO, Anthony J. Efeitos da saliência fônica e do tempo/modo verbal na concordância verbal. *In*: MOLLICA, M. C. de M. (org.) **Usos da linguagem e sua relação com a mente humana**. Rio de Janeiro: Tempo Brasileiro, 2010.

SCHERRE, M. M. P.; YACOVENCO, L. C.; NARO, A. **Nós e a gente no português brasileiro: concordâncias e discordâncias**. Estudos de Linguística Galega, v. especial I, ed. F. Cidrás, F. Dubert and X. L. Regueira, 2018. p. 13-27.

SEARA, Izabel Christine. A variação do sujeito nós e a gente na fala florianopolitana. **Organon**, v.14, n.28 e 29, Porto Alegre, 2000, p.179-194.

SEN, J. A robust analysis and forecasting framework for the Indian mid cap sector using times series decomposition approach. Disponível em: <https://scite.ai/reports/10.36227/techrxiv.15128901>. Acesso em: 19/10/2023

SILVA, Ivanilde da. **De quem nós/ a gente está (mos) falando agora?: uma investigação sincrônica da variação entre nós e a gente como estratégias de designação referencial**. 2004. 145 f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística. Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/87528/208785.pdf?sequence=1&isAllowed=y>. Acesso em: 22 out. 2021.

SILVA, Vitor Hugo Miro Couto. Perfil Municipal de Fortaleza: **Perfil da Juventude em Fortaleza: Aspectos Socioeconômicos a partir dos dados do Censo 2010**. Informe nº 57 – Edição Especial. Fortaleza: IPECE, 2013.

SILVA, Antonio de Moraes. **Dicionário da língua portuguesa** recopilado de todos os impressos até o presente. 3.ed. Lisboa: Typographia de M.P. de Lacerda, 1823.

SILVA, Giselle Machline de Oliveira e.; PAIVA, Maria da Conceição de. Visão de conjunto das variáveis sociais. *In*: SILVA, Giselle Machline de Oliveira e; SCHERRE, Maria Marta Pereira. (org.) **Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro**. Rio de Janeiro: Tempo Brasileiro, 1998. p. 335-378.

SILVA, C. A variação nós e a gente no português carioca. *Revista do Gelne*, v. 12, n.1, p. 67-74, 2010.

SILVA, Lia Barile Carvalho da. **Nós/A gente: mudança ou variação?**. 2011. 96 f. Dissertação (Mestrado em Comunicação, Linguagens e Cultura) – Programa de Pós-Graduação em Comunicação, Linguagens e Cultura, Universidade da Amazônia, Belém, 2011.

SILVA, Marinete Rodrigues da; CAMACHO, Roberto Gomes. Os pronomes nós e a gente no português falado em Rio Branco. **Estudos linguísticos**, São Paulo, 46 (1): p. 311-321, 2017

SOUZA, Adriana dos Santos; BOTASSINI, Jacqueline Ortelan Maia. A variação no uso dos pronomes-sujeito nós e a gente. SILEL. **Anais** [...]. Uberlândia: EDUFU, v. 1, 2009. Disponível em: [http://www.ileel.ufu.br/anaisdosilel/wp-content/uploads/2014/04/silel2009\\_gt\\_lg06\\_artigo\\_4.pdf](http://www.ileel.ufu.br/anaisdosilel/wp-content/uploads/2014/04/silel2009_gt_lg06_artigo_4.pdf). Acesso em: 14 jul. 2021.

SOUZA, Maria Helena Menezes. **A variação nós e a gente na posição de sujeito na comunidade quilombola Serra das Viúvas/Água Branca - AL**. 2020. 93 f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Programa de Pós Graduação em Linguística e Literatura, Universidade Federal de Alagoas, Maceió, 2020.

TAGLIAMONTE, S. A. Quantitative analysis. In: BAYLEY, R.; LUCAS, C. (Eds.). **Sociolinguistic variation: Theories, methods, and applications**. Cambridge: Cambridge University Press, 2007.

TAGLIAMONTE, S.; DENIS, D. Linguistic ruin? LOL! Instant messaging and teen language. **American Speech**, Durham, v. 83, n. 1, p. 3-34, 2008.

TAGLIAMONTE, S. A. **Variationist sociolinguistics: change, observation, interpretation**. Cambridge: Wiley-Blackwell, 2012

TAMANINE, Andréa Maristela Bauer. **Curitiba da gente: um estudo sobre a variação pronominal nós/a gente e a gramaticalização de a gente na cidade de Curitiba**. 2010. 222 f. Tese (Doutorado em Letras) – Setor de Ciências Humanas, Letras e Artes, Universidade Federal do Paraná, Curitiba, 2010. Disponível em: <<http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/24120/TeseAndreaTamanine.pdf?sequence=1>>. Acesso em: 11 out. 2021.

TARALLO, F. **A pesquisa sociolinguística**. São Paulo: Ática, 1999.

TAVARES, Nilceu Romi Kerecz. **A variação pronominal nós e a gente nos telejornais nacionais da Rede Globo**. 2014. 131 f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em linguística, Universidade Federal do Paraná, Curitiba, 2014. Disponível em: <<http://dspace.c3sl.ufpr.br/dspace/bitstream/handle/1884/37024/R%20-%20D%20-%20NILCEU%20ROMI%20KERECZ%20TAVARES.pdf?sequence=1&isAllowed=y>>. Acesso em: 24 maio 2021.

VIANNA, Juliana Barbosa de Segadas. **A concordância de nós e a gente em estruturas predicativas na fala e na escrita carioca**. 2006. 130f. Dissertação (Mestrado em Língua Portuguesa) - Programa de Pós-Graduação em Língua Portuguesa, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

WEINREICH, Uriel; LABOV, Willian; HERZOG, Marvin. **Fundamentos empíricos para uma teoria da mudança linguística**. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2006 [1968].

ZILLES, Ana Maria Stahl. The development of a new pronoun: the linguistic and social embedding of a gente in Brazilian Portuguese. **Language Variation and Change**, v. 17, n. 1, 2005. p. 19-53.

ZILLES, Ana Maria Stahl; MAYA, Leonardo Zechlinski; SILVA, Karine Quadros da. **A concordância verbal com a primeira pessoa do plural em Panambi e Porto Alegre, RS.** Organon, 14, n. 28-29, p. 195-219, 2000