UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

PRONÚNCIA VARIÁVEL DE (NDO) NA FALA PAULISTANA

DANY THOMAZ GONÇALVES

Versão Revisada

SÃO PAULO
2018

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

PRONÚNCIA VARIÁVEL DE (NDO) NA FALA PAULISTANA

Dany Thomaz Gonçalves

Dissertação apresentada ao Programa de Pós-Graduação em Semiótica e Linguística Geral do Departamento de Linguística da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo para a obtenção do título de Mestre em Linguística.

Orientador: Prof. Dr. Ronald Beline Mendes

Versão Revisada

SÃO PAULO
2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

G635p Gonçalves, Dany Thomaz
Pronúncia variável de (ndo) na fala paulistana /
Dany Thomaz Gonçalves ; orientador Ronald Beline
Mendes. - São Paulo, 2018.
117 f.

Dissertação (Mestrado)- Faculdade de Filosofia,
Letras e Ciências Humanas da Universidade de São
Paulo. Departamento de Linguística. Área de
concentração: Semiótica e Linguística Geral.

1. Sociolinguística. 2. Apagamento de /d/. 3. Fala
Paulistana. I. Mendes, Ronald Beline, orient. II.
Título.

BANCA EXAMINADORA

Prof. Dr. Ronald Beline Mendes (FFLCH-USP)
Presidente

Prof. Dr. Paulo Chagas de Souza (FFLCH-USP)
Membro Titular

Profª. Drª. Dayane Celestino de Almeida (UNICAMP)
Membro Titular

Profª. Drª. Livia Oushiro (UNICAMP)
Membro Titular

Prof. Dr. Sebastião Carlos Leite Gonçalves (UNESP)
Membro Suplente

Profª. Drª. Beatriz Raposo de Medeiros (FFLCH-USP)
Membro Suplente

Drª. Fernanda Canever
Membro Suplente

Para mamãe

Agradecimentos

Agradeço aos meus pais, José (*in memorian*) e Helena, por todo apoio desde sempre e por me incentivarem a sempre buscar mais.

Agradeço ao meu orientador, Ronald Beline Mendes, pelo acolhimento, pelos ensinamentos e por todas as palavras gentis que me concedeu durante o período tão corrido de curso do mestrado.

Agradeço à Livia Oushiro pela amizade, acolhida, por ter aceitado participar das bancas de qualificação e de defesa e pelas inúmeras mensagens respondidas prontamente. Sem ela, este trabalho não teria a mesma “graça”.

Agradeço à professora Dayane Celestino de Almeida por ter aceitado participar da banca de defesa e, também pela enorme simpatia e prontidão ao responder meus e-mails sem ao menos nos conhecermos pessoalmente.

Agradeço ao professor Paulo Chagas por ter aceitado participar das bancas de qualificação e defesa e pelas aulas maravilhosas de Teoria fonológica onde me ajudou a descobrir a melhor classificação para o apagamento de gerúndio.

Agradeço ao GESOL pela acolhida e pelos momentos de aprendizado e descontração. Sou eternamente grato por todos os comentários enriquecedores acerca do meu trabalho.

Agradeço especialmente à Amanda Santana, minha amiga paulistana mais que querida, por ser ela. Lembro-me ainda do dia da arguição do pré-projeto quando sentados a esperar nos conhecemos. De lá para cá nunca mais nos separamos. Obrigado pela ajuda em todos os momentos, obrigado por me incluir no contexto “uspiano” e “paulistano”, obrigado pela companhia nos congressos, que possamos ter muitos outros momentos assim.

Agradeço aos meus amigos que tiveram que aturar muitas das minhas desculpas por não poder sair ou fazer qualquer atividade pois estava estudando. Mas agradeço em especial à Emannelly, Manu, por ter sentado várias vezes com outro computador ao meu lado e ajudado a revisar os dados quando a minha cabeça já doía. Essa dissertação é para você também.

Agradeço ao João, pelo carinho, força, e compreensão dos momentos de surtos. Ainda bem, que Deus o colocou no meu caminho.

Agradeço à Érica pela ajuda sempre que eu precisei e por todo o apoio técnico.

Resumo

Essa dissertação trata da pronúncia variável de /ndo/, com /d/ apagado ou pronunciado. Baseada na teoria e metodologias da Sociolinguística Variacionista (LABOV, 2006[1966], 2008[1972]), focaliza tal variável – aqui referida como (NDO) – na variedade paulistana do português brasileiro, através das 60 entrevistas do Projeto SP2010 (MENDES & OUSHIRO, 2012), estratificadas por Sexo/Gênero, Faixa Etária e Escolaridade. A variável (NDO) tem sido estudada desde o final dos anos 70 sob diferentes perspectivas. Alguns pesquisadores se atêm à pronúncia variável do sufixo de gerúndio (por exemplo, *falano* ou *falando* em diferentes variedades do português: Rio de Janeiro (MOLLICA, 1989), Belo Horizonte (CRISTÓFARO SILVA, 1996), João Pessoa (MARTINS, 1999, 2001), Minas Gerais (MARTINS, 2006), São José do Rio Preto (FERREIRA, 2010), Dourados e Ponta Porã – MS (MARTINS e BUENO, 2011), Taboco – MS (VIEIRA, 2011), Fortaleza (NASCIMENTO ET AL., 2013), Maceió (ALMEIDA e OLIVEIRA, 2017)). Leva-se em consideração o apagamento de /d/ como uma variável cujo padrão de realização pode ter especificidades na fala paulistana – uma variedade que, ainda que mais detidamente estudada nos últimos anos, ainda não foi analisada no que toca a (NDO), nem no gerúndio, especificamente, nem em outros contextos morfológicos ou itens lexicais, tal como na palavra *quando*. O objetivo central é, então, investigar quais as variáveis linguísticas e sociais que se correlacionam à realização de (NDO). Para tanto, analisou-se qualitativa e quantitativamente a amostra, onde foram encontrados casos de apagamento de /d/ na pronúncia da Conjunção “quando” (93/1557) e em verbos no gerúndio (1001/4143). Segundo as análises estatísticas multivariadas, verificou-se que não há correlação significativa entre (NDO) e nenhuma das variáveis linguísticas, o apagamento ocorre com maior frequência na fala dos homens, sendo mais propício na fala daqueles com menor grau de instrução, assim como ocorre para aqueles falantes da Classe Social menos favorecida. Com relação à Faixa Etária, existe um favorecimento na fala dos mais jovens, principalmente naqueles da Faixa Etária intermediária, já para a terceira faixa existe um desfavorecimento, esses resultados levam a indícios de que o apagamento de /d/ em (NDO) seja uma variável estável na capital paulista.

Palavras-chave: apagamento de /d/, São Paulo, fala paulistana, sociolinguística variacionista, gerúndio.

Abstract

This master's thesis deals with /ndo/ variable pronunciation, with /d/ deletion or pronunciation. Based on Variationist Sociolinguistics theory and methodologies (LABOV, 2006[1966], 2008[1972]), we focus on this variable – here called as (NDO) – on São Paulo's speech of Brazilian Portuguese, through 60 interviews from SP2010 project (MENDES & OUSHIRO, 2012), stratified by Sex/Gender, Age and Level of Education. The variable (NDO) has been studied since the 1970's from different perspectives. Some researchers focused only on the pronunciation of the suffix of gerund (for example, *falano* or *falando* in different portuguese varieties as Rio de Janeiro (MOLLICA, 1989), Belo Horizonte (CRISTÓFARO SILVA, 1996), João Pessoa (MARTINS, 1999, 2001), Minas Gerais (MARTINS, 2006), São José do Rio Preto (FERREIRA, 2010), Dourados and Ponta Porã – MS (MARTINS e BUENO, 2011), Taboco – MS (VIEIRA, 2011), Fortaleza (NASCIMENTO ET AL., 2013), Maceió (ALMEIDA e OLIVEIRA, 2017)). Taking into account the deletion of /d/ as a variable in which the performance pattern may have specificities in Paulistano speech - a variety that, although studied in recent years, has not yet been analyzed regarding (NDO), nor in the gerund, specifically, nor in other morphological contexts or lexical items, as in the word *quando*. The main goal is to investigate the linguistic and social variables that are correlated to the performance of (NDO). For that, the sample was analyzed qualitatively and quantitatively, in which cases of deletion of /d/ were found in the pronunciation of the conjunction "quando" (93/1557) and in verbs in the gerund (1001/4143). According to the multivariate statistical analysis, it was verified that there is no significant correlation between (NDO) and the linguistic variables, deletion occurs more frequently in men's speech, being more propitious in the speech of those with lower Education Degree, as it occurs for those speakers of the less favored Social Class. Regarding the Age Range, there is a favoring of young people's speech, at all levels, the results lead to an indication that the /d/ deletion in (NDO) is a variable stability in São Paulo city.

Key Words: /d/ deletion, São Paulo, Paulistano speech, variationist sociolinguistics, gerund.

Lista de Tabelas

1.1 Distribuição Geral de apagamento de /d/ na fala carioca -----	7
1.2 Distribuição Geral de apagamento de /d/ para Fortaleza/CE -----	12
1.3 Resultados do apagamento para Maceió/AL-----	13
2.1 Apagamento por extensão do vocábulo na fala carioca-----	33
2.2 Apagamento por conteúdo seguinte na fala carioca-----	34
3.1 Distribuição geral de apagamento de /d/ em (NDO)-----	36
3.2 (NDO) por classe morfológica-----	37
3.3 Ocorrências de (NDO) por Classe Morfológica-----	38
3.4 Apagamento de /d/ nas conjunções quando e segundo-----	39
3.5 Agrupamentos sociais dos informantes não categóricos quanto à pronúncia de (NDO) em quando-----	42
3.6 Taxas de apagamento de /d/ em quando de acordo com Sexo/Gênero, Faixa Etária e Escolaridade dos 21 informantes não categóricos-----	43
3.7 Taxas de apagamento de /d/ em quando de acordo com Região da Cidade, Mobilidade Geográfica, Classe Social e Origem dos Pais dos 21 informantes não categóricos -----	44
3.8 Distribuição de d _ø em três estruturas-----	45
3.9 Distribuição de apagamento de /d/ em justaposição de acordo com os 11 informantes não categóricos-----	46
3.10 Distribuição de apagamento em justaposição por estrutura-----	48
3.11 Distribuição de dados de apagamento para as Orações Reduzidas Adverbiais por variáveis sociais-----	50
3.12 Distribuição de dados de apagamento para as Orações Reduzidas Adverbiais por variáveis linguísticas-----	57
3.13 Distribuição de dados de apagamento para as Perífrases Verbais por variáveis sociais---	60

3.14 Distribuição de dados de apagamento para as Perífrases Verbais por variáveis linguísticas- -----	66
4.1 Estimativas (em logodds) de valores de significância para Perífrases e Orações Reduzidas- -----	76
4.2 Estimativas (em logodds) de valores de significância para Sexo/Gênero e Classe Social das Orações Reduzidas-----	78
4.3 Estimativas (em logodds) de valores de significância com interação entre Faixa Etária e Escolaridade das Orações Reduzidas-----	78
4.4 Estimativas (em logodds) de valores de significância para Orações Reduzidas-----	79
4.5 Apagamento de Gerúndio com relação à Faixa Etária para duas variedades do PB-----	80
4.6 Apagamento de Gerúndio com relação à Escolaridade para outras variedades do PB-----	81
4.7 Recategorização dos fatores da variável Classe Social-----	83
4.8 Estimativas (em logodds) de valores de significância para Orações Reduzidas (II)-----	83
4.9 Estimativas (em logodds) de valores de significância para Perífrases com Sexo/Gênero e Conjugação-----	84
4.10 Estimativas (em logodds) de valores de significância para Perífrases com variáveis sociais- -----	85
4.11 Estimativas (em logodds) de valores de significância para Perífrases com Classe Social Recategorizada-----	86

Lista de Quadros

1 Variáveis Estratificadoras da amostra SP2010 e perfis sociolinguísticos dos informantes---	23
2 Variáveis Sociais-----	25
3 Grupos de Fatores Linguísticos-----	30
4 Grupos de Fatores Linguísticos para os subconjuntos de dados do gerúndio-----	34
5 Verbos Auxiliares das Orações Perifrásticas-----	67

Lista de Figuras

1 Divisão da cidade de São Paulo entre regiões mais centrais e mais periféricas-----	27
2 Origem da população economicamente ativa (30-60 anos) na região metropolitana de São Paulo-----	28
3 Fórmula para cálculo do Índice Socioeconômico-----	29
4 Frequência de d_0 e d_1 em QUANDO para os indivíduos não categóricos-----	39
5 Configurações Acústicas de QUANDO com d_0 -----	40
6 Configurações Acústicas de QUANDO com d_1 -----	41
7 Proporção de d_0 em Orações Reduzidas por Sexo/Gênero-----	51
8 Proporção de d_0 em Orações Reduzidas por Faixa Etária-----	52
9 Proporção de d_0 em Orações Reduzidas por Escolaridade-----	52
10 Proporção de d_0 em Orações Reduzidas por Classe Social-----	53
11 Proporção de d_0 em Orações Reduzidas de acordo com Renda do informante-----	54
12 Proporção de d_0 em Orações Reduzidas por Mobilidade Geográfica-----	54
13 Proporção de d_0 em Orações Reduzidas por Origem dos Pais-----	55
14 Proporção de d_0 em Orações Reduzidas por Região da Cidade-----	56
15 Proporção de d_0 em Orações Reduzidas por Conjugação-----	57
16 Proporção de d_0 em Orações Reduzidas por Tamanho do Verbo-----	58
17 Proporção de d_0 em Orações Reduzidas por Contexto Seguinte-----	59
18 Proporção de d_0 em Perífrases por Sexo/Gênero-----	61
19 Proporção de d_0 em Perífrases por Faixa Etária-----	61
20 Proporção de d_0 em Perífrases por Escolaridade-----	62
21 Proporção de d_0 em Perífrases por Classe Social-----	62

22 Proporção de $d_\emptyset$ em Perífrases por Renda dos informantes-----	63
23 Proporção de $d_\emptyset$ em Perífrases por Mobilidade Geográfica-----	64
24 Proporção de $d_\emptyset$ em Perífrases por Origem dos Pais-----	65
25 Proporção de $d_\emptyset$ em Perífrases por Região da Cidade-----	65
26 Proporção de $d_\emptyset$ em Perífrases, sem o verbo ESTAR, de acordo com Material Interveniente	67
27 Proporção de $d_\emptyset$ em Perífrases por Conjugação Verbal-----	68
28 Proporção de $d_\emptyset$ em Perífrases por Tamanho do Verbo Principal-----	69
29 Proporção de $d_\emptyset$ em Perífrases por Contexto Seguinte-----	69
30 Proporção de $d_\emptyset$ em Perífrases por Tamanho do Verbo Auxiliar ESTAR-----	70
31 Proporção de $d_\emptyset$ em Perífrases com auxiliar ESTAR de acordo com Material Interveniente	71
32 Proporção de $d_\emptyset$ no cruzamento das variáveis Tamanho do auxiliar ESTAR* Material Interveniente-----	72

Sumário

Introdução	1
1 Apagamento De /D/ Em (NDO)	4
1.1 Os Casos De Apagamento Em (NDO): Indícios Do Surgimento.....	4
1.2 Estudos Variacionistas Sobre (NDO)	6
2 Caminhos Teórico-Metodológicos	15
2.1 Teoria Da Variação E Mudança Linguística	15
2.2 Paradigma Quantitativo	19
2.3 A Amostra De Fala Paulistana.....	22
2.4 Outras Variáveis Sociais.....	26
2.5 Variáveis Linguísticas Independentes	30
2.6 Variáveis Específicas Para Subconjuntos De Dados	34
3 Conjunto De Dados	36
3.1 Conjunto Geral De Dados	36
3.2 Quando	39
3.3 Verbos No Gerúndio.....	45
3.3.1 Justaposição	46
3.3.2 Orações Reduzidas Adverbiais	49
3.3.3 Perífrases Verbais	59
3.4 Síntese.....	72
4 Análises Multivariadas: Gerúndios Em Orações Reduzidas E Perífrases Verbais	75
4.1 Tipo De Gerúndio: Orações Reduzidas X Perífrases	75
4.2 Em Busca De Interações.....	77
4.3 Gerúndio Em Orações Reduzidas.....	79
4.4 Gerúndio Em Perífrases Verbais	84
Conclusão	88
Referências	93
Anexos.....	98

INTRODUÇÃO

Desde o final dos anos 1970, têm sido realizados estudos sobre diversos fenômenos de variação do português falado no Brasil, dentre eles, o apagamento da oclusiva coronal sonora [d] na sequência –ndo, sobretudo em verbos no gerúndio. O apagamento de [d] nesse sufixo é comum no português brasileiro, podendo ser constatado em diferentes regiões: Rio de Janeiro (MOLLICA, 1989), Belo Horizonte (CRISTÓFARO SILVA, 1996), João Pessoa (MARTINS, 1999, 2001), Minas Gerais (MARTINS, 2006), São José do Rio Preto (FERREIRA, 2010), Dourados e Ponta Porã – MS (MARTINS e BUENO, 2011), Taboco – MS (VIEIRA, 2011), Fortaleza (NASCIMENTO ET AL., 2013), Maceió (ALMEIDA e OLIVEIRA, 2017).

Os autores que tratam da pronúncia variável de (NDO), com ênfase no sufixo de gerúndio, referem-se à pronúncia [-no] como um caso de redução (CRISTÓFARO SILVA, 1996), assimilação (MOLLICA e MATTOS, 1992) ou como um caso de apagamento (FERREIRA, 2010; VIEIRA, 2011; NASCIMENTO ET AL., 2013; ALMEIDA e OLIVEIRA, 2017). Antes de optar por um desses termos, empregados em trabalhos anteriores para fazer referência ao fenômeno em foco, foram levantados os tipos de processos fonético-fonológicos considerados para classificar a ausência do segmento [d] em {-ndo}. De acordo com Spencer (1996) e Cristófaró Silva (2015): (1) apagamento é um fenômeno fonológico em que um segmento consonantal ou vocálico é cancelado; o termo é sinônimo de *cancelamento* e *queda* e equivale ao fenômeno de *lenição*, ou seja, de enfraquecimento consonantal em grau máximo; (2) assimilação é um processo através do qual um determinado som “adquire” (assimila) propriedades de um som adjacente. Assim, o termo *apagamento* parece ser mais adequado para o fenômeno estudado neste trabalho, considerando-se que [d] e [n] já compartilham o mesmo ponto de articulação, de modo que o termo *assimilação* parece impreciso.

Toma-se o apagamento de /d/ como uma variável cujo padrão de realização pode ter especificidades na fala paulistana – uma variedade que, ainda que mais detidamente estudada nos últimos anos, ainda não foi analisada no que toca a (NDO), nem no gerúndio, especificamente, nem em outros contextos morfológicos ou itens lexicais, tal como na palavra

quando. O objetivo central é, então, investigar, sob a ótica da Sociolinguística Variacionista (LABOV, 2008[1972]), as variáveis linguísticas e sociais que se correlacionam à realização de (NDO), através dos dados extraídos da amostra coletada pelo Projeto SP2010¹ (MENDES e OUSHIRO, 2012). Considerando-se que as análises quantitativas não se restringem, aqui, ao sufixo de gerúndio, uma das variáveis independentes incluída no estudo foi a Classe Morfológica do item em que ocorre [-ndo], por exemplo, adjetivos (*lindo*, *vagabundo*), conjunções (*segundo* e *quando*), o numeral *segundo*, substantivos (*Fernando*, *mundo*), verbos no presente (*mando*, *recomendo*), além, é claro, de verbos no gerúndio (*amando*, *comendo*, *partindo*, *pondo*). Iniciar as análises incluindo-se todas as palavras terminadas em (ndo) no corpus tem a vantagem de sinalizar que o trabalho variacionista, em princípio e por princípio, não deveria admitir que a variação absolutamente não ocorra em certos contextos. Em outras palavras, ainda que o apagamento de /d/ possa ser muito mais frequente quando se trata do gerúndio, cabe verificar se, em outros casos, /d/ é apagado suficientes vezes para que se proponha alguma análise quantitativa, com vistas à evidenciação de padrões.

Além da Classe Morfológica da palavra terminada em /-ndo/ - uma variável que organiza o início das análises – incluem-se também a Extensão do Vocábulo e a Vogal Precedente. No caso de verbos, especificamente, interessou verificar possíveis correlações com a Estrutura Sintática (perífrases – *estava comendo devagar*, Orações Reduzidas – *gosto de crianças correndo pela sala*, justaposição – *a hora ia passando, passando, passando*), o Verbo Auxiliar em perífrases (*estar*, *andar*, *ficar*, *viver*, etc), o Tempo Verbal do Verbo Auxiliar, a existência de Material Interveniente entre o auxiliar e o gerúndio, etc.

Em termos de variáveis sociais, interessam-se sobretudo as “clássicas”: Gênero/Sexo do informante, sua Escolaridade, sua Faixa Etária (em razão da estratificação da amostra SP2010) e Classe Social (aqui, analisada de acordo com índices propostos por Oushiro, 2015). Um dos intuitos deste trabalho foi verificar se a alternância de [no] e [ndo] pressupõe estigma ou prestígio dentro da comunidade de fala paulistana para em seguida buscar uma forma de correlação com distintas variáveis, assim como Gênero/Sexo, Faixa Etária e Classe Social, tal como se propõe nos inúmeros trabalhos variacionistas, a partir de Labov (2006 [1966]). Se

¹ Disponível em <http://projetosp2010.fflch.usp.br>. Último acesso em: 25 de julho de 2018

consideramos que a pronúncia [no] é “menos cuidadosa” e menos prestigiada, espera-se que as mulheres tendam a apagar o /d/ menos frequentemente, relativamente aos homens. De certo modo, isso também tem a ver com Escolaridade: seria de esperar que os falantes menos escolarizados (na amostra SP2010, aqueles que pararam de estudar no Ensino Médio) apresentem tendência de apagar o /d/ relativamente maior do que aqueles que têm nível Superior de Escolaridade). Quanto à Faixa Etária, o intuito é verificar, através do construto do tempo aparente, se o caso de variação é estável ou se há indício de mudança em curso (LABOV, 2001). Tendo em vista que a amostra SP2010 não é estratificada quantitativamente por Classe Social e que em estudos anteriores percebeu-se que a variável não se refere apenas à renda familiar ou do informante, decidiu-se utilizar o índice de Classe Social proposto por Oushiro (2015), calculado a partir de informações tais como nível de Escolaridade, Ocupação e Renda. Essas e outras variáveis serão discutidas mais adiante.

No Capítulo 1, apresentam-se os trabalhos que tratam do apagamento de /d/ em (NDO) no português brasileiro (PB) e em seguida estudos variacionistas acerca do apagamento. No Capítulo 2, faz-se uma introdução à teoria da variação e mudança linguística e ao paradigma quantitativo, que são os pressupostos que fundamentam a pesquisa. Também se faz uma descrição da amostra da fala paulistana que aqui se utiliza, seguida das variáveis independentes que se estabeleceram a partir das hipóteses aventadas. No Capítulo 3, apresentam-se as análises descritivas e inferenciais univariadas de todas as ocorrências de (NDO) encontradas no corpus. A partir destas, foram organizados os dados para os quais se procedeu, no capítulo 4, a análises estatísticas inferenciais multivariadas, no sentido de verificar quais são as variáveis linguísticas e sociais mais relevantes para a ocorrência do apagamento de /d/. Finalmente, a Conclusão apresenta um balanço dos resultados obtidos e das discussões realizadas nas seções anteriores.

Capítulo 1

Apagamento de /d/ em (NDO)

Este capítulo apresenta uma discussão acerca de indícios do surgimento do apagamento de /d/ em (NDO), seguido de uma revisão de pesquisas anteriores que tratam deste processo fonológico através da perspectiva variacionista.

1.1 – Os casos de apagamento em (NDO): indícios do surgimento

Tal como se indica na introdução, o termo “apagamento” é utilizado para designar um processo fonológico em que um segmento consonantal ou vocálico é cancelado. Alguns autores como Bagno (1997), Martins (2004) e Hora e Aquino (2012), destacam que o caso do apagamento da consoante oclusiva coronal sonora /d/ em palavras com a terminação (NDO), como em “comendo”, “estudando”, “quando”, “segundo”, seria resultado de um processo de assimilação do fonema /d/ pelo fonema /n/ para, finalmente haver o apagamento do fonema /d/, sendo (-nd- > -nn- > -n-).

Antes de fazer uma discussão sobre os trabalhos que desenvolvem um estudo com relação ao apagamento do /d/, cabe dizer que o processo de apagamento em (NDO) não é algo recente na língua portuguesa; no trabalho de Amaral publicado em 1920, *O dialeto caipira: gramática, vocabulário*, encontram-se registros de apagamentos de /d/ em que o autor considera a pronúncia reduzida como característica daquilo que denomina como dialeto caipira do PB:

como se vê, ainda este brasileirismo, parecendo original, talvez tenha a sua genealogia transatlântica. É de notar-se, porém, que a fórmula mais comum não é a que aí fica registrada, mas a forma de gerúndio, posposta ao infinitivo “estar”, que tem o mesmo sentido, e é de uso vulgaríssimo: “Aquela torre parece que tá quereno caí” (AMARAL, 1920:159).

Conforme o autor, “as condições do meio, as transformações sociais pelas quais passava o São Paulo da época, partiriam os fios condutores da evolução” (AMARAL, 1920 apud ALMEIDA e OLIVEIRA, 2017), fato este que não se comprovou, já que pesquisas recentes

como a de Ferreira (2010) com relação ao processo de apagamento na região de São José do Rio Preto, situada no interior de São Paulo, indica que os casos de apagamento em (NDO) têm se mostrado como uma variedade persistente na fala, em que os fios condutores da evolução não foram partidos.

Outros trabalhos – como o de Marroquim (1934), para as variedades alagoana e pernambucana; Teixeira (1938), para a mineira; Melo (1946), de uma perspectiva dialetológica geral – identificaram casos de apagamento como uma influência de línguas africanas. Para Marroquim (1934:86), o que aconteceu com o grupo /nd/ foi que

os africanos teriam sentido dificuldade em pronunciar o grupo, e, por menor esforço, tê-lo-iam modificado, como o simplificou pelo mesmo motivo o resto da população, sem que uns imitassem outros. No grupo dialetal aquilano-umbro-romano dá-se o mesmo fenômeno, outrora mais generalizado na Itália, sem suspeita de influência africana.

Partilhando da mesma opinião que Marroquim, atribuindo o apagamento a uma influência africana, mas também das línguas tupi, Melo (1946:57) salienta que o fenômeno também foi verificado em outras línguas que não tinham contato com as citadas: “sirva de exemplo o latim vulgar. O osco e o úmbrio tinham *nn* quando o latim dizia *nd*”.

Amaral, Marroquim, Melo e Teixeira descreveram o fenômeno de apagamento do /d/ em morfema de gerúndio como uma característica do falar “caipira” ou “roceiro”, caracterizando o falante que faz uso da regra como inculto e rústico, sendo desta forma, considerado estigmatizado. Como uma variante geralmente tomada como não padrão no PB, a realização [no] para (NDO) tem sido estudada em diferentes regiões do Brasil, como verificada nos estudos variacionistas a seguir, em que foram identificadas diferentes visões acerca das atitudes linguísticas pelos falantes como tratando-a como forma estigmatizada por certos grupos sociais.

1.2– Estudos Variacionistas sobre (NDO)

De acordo com Ferreira (2010), cuja pesquisa inclui o caso do apagamento de /d/ em gerúndios para a fala de São José do Rio Preto (SP), os tratamentos do apagamento de /d/ se subdividem em diferentes perspectivas: alguns, como Lemle (1984) verificam, a partir de uma ótica morfossintática, os fatores que favorecem a pronúncia não padrão [no]; outros, como Mollica (1989), Cristófaró Silva (1996) e Martins (2006), que verificaram através da ótica da Sociolinguística.

Mesmo sob o prisma da sociolinguística, alguns trabalhos apresentam procedimentos metodológicos distintos. Encontram-se trabalhos que verificaram a influência de variáveis linguísticas e sociais (MARTINS, 1999; FERREIRA, 2010; VIEIRA, 2011; NASCIMENTO, ARAÚJO E CARVALHO, 2013); outros que se dedicaram exclusivamente a variáveis sociais (NASCIMENTO E MOTA, 2004; COSTA, 2009; MARTINS E BUENO, 2011) e aqueles que investigaram a frequência de uso do apagamento de /d/ (HORA E AQUINO, 2012). No que concerne à maneira como os dados foram analisados quantitativamente, algumas pesquisas exibem seus resultados sob a forma de pesos relativos (MARTINS, 1999; FERREIRA, 2010; NASCIMENTO, ARAÚJO E CARVALHO, 2013), enquanto os demais apresentam-nos sob a forma de percentuais. Esses três últimos trabalhos citados são os únicos em que se extraíram dados de grandes amostras: VALPB, IBORUNA e NORPOFOR², respectivamente. Os anteriores constituíram um *corpus* especialmente para a pesquisa sobre o apagamento de /d/.

São poucos os trabalhos que incluem em suas análises itens de outras Classes Morfológicas além de verbos no gerúndio, por ser a classe em que mais ocorre o apagamento, relativamente a outras palavras terminadas por (NDO) como em substantivo (*mundo*), adjetivos (*lindo*, *vagabundo*), numeral (*segundo*) e verbos no presente (*mando*, *vendo*), somente encontradas nas pesquisas de Mollica (1989), Martins (1999), Ferreira (2010) e Vieira (2011)

² Projeto Variação Linguística no Estado da Paraíba; IBORUNA – Amostra de fala da região de São José do Rio Preto; Projeto Norma Oral do Português Popular de Fortaleza.

que atentam-se a descrever, mesmo com poucos dados de apagamento, quais fatores caracterizam a não realização de /d/ nestes contextos.

Seguindo essas pesquisas cronologicamente, Mollica (1989) inicia seu estudo sobre o apagamento de /d/ em todos os casos de (NDO), a partir de dados extraídos de entrevistas com 64 falantes não universitários da cidade do Rio de Janeiro. A autora encontrou as seguintes ocorrências:

Tabela 1.1 - Distribuição geral de apagamento de /d/ na fala carioca

Classe Morfológica	N	%
Verbos no Gerúndio	1034/2672	38,7
Verbos no Presente	6/62	10
Substantivos	22/314	7
Outros	195/1053	19

Fonte: Adaptado de Mollica (1989)

De acordo com a tabela 1.1, quando /d/ faz parte da raiz, como em substantivos, conjunções ou verbos no presente, tende a resistir ao apagamento. No caso dos verbos no gerúndio, /d/ encontra-se numa unidade afixal. Segundo Mollica (1989), o apagamento de /d/ no Rio de Janeiro não atua de maneira uniforme no léxico, seja porque não atinge os itens potencialmente sujeitos ao processo, seja porque nem todas as categorias gramaticais são atingidas. Encontraram-se 22 casos de apagamento na palavra *mundo* (pronunciado como [‘munu]) – um número maior do que os 6 casos em verbos no presente, não exemplificados no estudo.

Segundo Cristófaró Silva (1996), em seu estudo de cunho mais teórico-descritivo, com dados do dialeto mineiro: (i) o processo de redução (aqui denominado apagamento) aplica-se apenas às formas de gerúndio; (ii) esse processo se aplica em relação à informação dada pelo componente fonológico e morfológico encontrados nos verbos do gerúndio; (iii) o processo não leva a reorganização lexical, ou seja, não há mudança na organização interna do léxico, visto

que palavras que contêm o mesmo contexto favorável ao apagamento e que não portam informação morfológica de gerúndio não seriam afetadas pelo processo.

Partindo dessas considerações de Cristófaró Silva (1996), mas levando-se em consideração os resultados obtidos por Mollica (1989), uma das questões que se considerou para encaminhar a presente pesquisa é a seguinte: se a manutenção de /d/ em outras palavras que não verbos no gerúndio não é categórica, seria o caso de que o apagamento estaria se estendendo a outros vocábulos que possuem o mesmo tipo de terminação? Se sim, interessaria verificar quais outras classes e itens lexicais favoreceriam o apagamento de /d/.

Em uma mesma linha de pesquisa, Martins (1999; 2001) analisou a fala de 24 informantes do projeto VALPB, reunindo 1988 dados de (NDO). Em sua amostra, estratificada por Sexo/Gênero, três Faixas Etárias e três níveis de Escolaridade, essas três variáveis têm correlação significativa com a variável dependente, além de Classe de Palavras e Extensão do Vocábulo. A autora reporta os seguintes pesos relativos para apagamento, que é favorecido: na fala dos homens (.59) relativamente à das mulheres (.42); informantes com até 49 anos (.53) relativamente àqueles que têm mais de 49 anos (.45); analfabetos (.69), relativamente aos informantes que detêm ensino fundamental completo (.53) e universitários (.31). Martins (1999:340) salienta que “estamos lidando com um fenômeno de variação estável na cidade de João Pessoa”. Nesse trabalho, os resultados mostram que o apagamento de /d/ é particularmente produtivo nos gerúndios. Os verbos no presente do indicativo (*pretendo* [pretenu]) e a Conjunção *quando* ([quanu]), por outro lado, desfavorecem a queda da consoante. Quanto à Extensão do Item Lexical, itens com mais de duas sílabas favorecem o apagamento da oclusiva, em oposição aos vocábulos dissílabos, que tendem a manter o segmento.

Em comparação ao estudo de Mollica (1989), Martins (1999) verifica um número maior de casos de apagamento na Conjunção *quando* do que no substantivo *mundo*. Nesse sentido, interessa verificar quão produtivo é o apagamento de /d/ também nesses casos, além do gerúndio, na fala paulistana. Tal como mostram as análises apresentadas nos capítulos 3 e 4, das seis Classes Morfológicas em que ocorre (NDO) – adjetivos (*vagabundo*, *lindo*), conjunções (*quando*, *segundo*), numeral (*segundo*), substantivos (*mundo*), verbos no presente (*pretendo*, *mando*) e verbos no gerúndio (*entendendo*, *chovendo*, *falando*), foram encontrados casos de

apagamento em verbos no presente (1), e na Conjunção *quando* (93) e, claro, nos verbos no gerúndio (1070). Diferentemente da pesquisa sobre a fala carioca, que verificou certa frequência de apagamento em substantivos, na amostra paulistana aqui em foco, depois dos verbos no gerúndio, é na Conjunção *quando* que o apagamento é mais frequente.

No estudo de Mota e Nascimento (2004), que utilizam inquéritos que compõem o ALIB³, o apagamento ocorreu em maior proporção em um contexto mais natural de fala, aquele em o falante presta um grau menor de atenção a sua fala, mais casual. Analisando 20 informantes da amostra, distribuídos por Sexo/Gênero (11 indivíduos do Sexo feminino e 9 do Sexo masculino), 2 Faixas Etárias (I- 18 a 30 anos, com 11 indivíduos e II-50 a 65 anos, com 9) e dois níveis de Escolaridade (Ensino Fundamental I, com 16 falantes e Superior, com 4), encontraram uma média de 61,5% de apagamento para os homens e 42% para as mulheres. As demais variáveis sociais seriam estudadas em um segundo momento da pesquisa. Para a situação de fala, verificou-se 33% de apagamento em contextos mais formais, quando o grau de atenção à fala é maior, e 65,5% quando a fala era mais casual. Segundo as autoras:

“fica constatado que a variação encontra-se apenas no nível fonético, não sendo assimilada como forma lexical e forma não-padrão do gerúndio, pois, em todos os informantes observados, as duas formas se revezam, conforme a tensão do discurso, mostrando que há por parte do falante um relativo grau de consciência de que a forma de gerúndio utilizada no cotidiano diverge da forma admitida na norma padrão da língua portuguesa.” (MOTA e NASCIMENTO, 2004:7)

Com relação a esta afirmação, a questão dos diferentes estilos (cf. LABOV, 2008 [1972]) com base na atenção à própria fala, corrobora-se quando os falantes estão lendo textos propostos ou quando estão numa conversa mais espontânea.

As pesquisas de Mollica (1989) para o Rio de Janeiro, Ferreira (2010) para São José do Rio Preto (SP) e Vieira (2011) para Taboco (MS) são aquelas das quais essa dissertação mais se aproxima, em razão da proposta de verificar, através de análises multivariadas de dados

³ Atlas Linguístico do Brasil. Foram analisados inquéritos de 10 cidades: Jequié - BA, Santo Amaro - SP, Salvador - BA, Belém - PA, Imituva - SC, João Pessoa - PB, Marília - SP, Niterói - RJ, Porto Alegre - RS e Recife - PE.

extraídos de amostras maiores, quais fatores linguísticos e sociais favorecem o apagamento da oclusiva coronal /d/ em (NDO). Desta forma, incluíram-se variáveis linguísticas como: Classe Morfológica, Extensão do Vocábulo, Contexto Fonológico Precedente e Seguinte. Ferreira (2010) ainda adicionou uma classificação mais detalhada para os verbos no gerúndio, também utilizada nesta pesquisa, com a seguinte classificação das estruturas que continham esses tipos de verbo:

- Perífrases – verbo auxiliar mais um verbo principal no gerúndio:
(1) Pedro *está brincando* pela rua.
- Justaposição – oração que pode ser caracterizada por uma perífrase ou oração reduzida, em que se encontram duas ou mais ocorrências de gerúndio em sequência:
(2) Eles estavam *discutindo fazendo* o maior pampeiro.
- Orações Reduzidas – o verbo no gerúndio é nuclear de uma oração; no exemplo abaixo, funciona como um adjunto adverbial:
(3) Dara passou meses *saindo* com o Cesar.

Para aplicação do seu estudo a partir das variáveis delimitados, Ferreira (2010) selecionou 76 entrevistas estratificadas por Sexo/Gênero, 5 Faixas Etárias e 4 níveis de Escolaridade. Obteve 999 ocorrências de verbos no gerúndio, com 72% de apagamento. As variáveis que têm correlação significativa com o apagamento são: Sexo/Gênero (favorecimento na fala dos homens – .59 – e desfavorecimento na das mulheres – .40); Escolaridade (1º Ensino Fundamental .72; 2º EF .57; Ensino Médio .48; Ensino Superior .32); Faixa Etária (26 a 35 anos: .64; 7 a 15 anos: .63; 16 a 25 anos: .56; 36 a 55: .44) e Estrutura Sintática (justaposição: .80; perífrases: .51; Orações Reduzidas: .45). Portanto, a partir da análise, a autora identificou que a frequência de apagamento é alta na variedade estudada, especialmente em estruturas perifrásticas, sendo recorrente na fala dos homens, dos menos escolarizados e dos mais jovens (de 7 a 35 anos).

O trabalho de Vieira (2011) analisa dados extraídos de entrevistas com 16 informantes, todos com Ensino Fundamental completo, agrupados em duas Faixas Etárias (15 a 35 anos e

acima de 50) para o falar de Taboco - MS. Verificou apagamento de /d/ em 92% dos casos de gerúndio (297/326), e 42% em conjunções (65/158). Desta forma, percebe-se que o apagamento parece ser uma característica predominantemente encontrada nos verbos no gerúndio para esta variedade de fala ainda que a porcentagem de apagamento encontrada para as conjunções seja relativamente alta. Já para Faixa Etária, seu resultado foi inesperado, pois encontrou a mesma taxa de apagamento para ambas (74%). Com relação ao Sexo/Gênero, assim como em todos os outros trabalhos mencionados, o apagamento é mais frequente entre os homens (80% - 182/229) do que entre as mulheres (70% - 180/255). Vieira afirma que “o equilíbrio da variante nas duas Faixas Etárias nos leva a perceber a generalidade do fenômeno, bem como seu caráter não estigmatizado na comunidade linguística estudada” (VIEIRA, 2011:20). A partir dos resultados de Mollica (1989), Ferreira (2010) e Vieira (2011), considerando os pressupostos de Labov (2006[1966]) no que diz respeito ao prestígio de uma determinada variável, um dos objetivos que se considerou para encaminhar esta pesquisa é o seguinte: seria o caso de o apagamento ser considerado uma variante estigmatizada na comunidade paulistana, podendo ser considerada como característica caipira como salientava Amaral (1920)? Se não, interessaria investigar como a variante se comporta a partir das análises das variáveis sociais.

Com uma amostra estratificada muito semelhante à de Vieira (2011), Martins e Bueno (2011) investigaram o apagamento de /d/ na região de Dourados e Ponta Porã, também no Mato Grosso do Sul. As autoras obtiveram 104 dados de apagamento, em entrevistas com 12 informantes agrupados por Sexo/Gênero, 2 Faixas Etárias (20 a 50 anos; 51+) e 2 níveis de Escolaridade (analfabetos e alfabetizados). Com um total de 53,8% (56/104) de casos de apagamento, verificaram que essa variante é mais frequente na fala das mulheres (82,5% - 33/41) do que na dos homens (35,9% - 23/64). Apesar da amostra ser relativamente pequena (apenas 12 falantes), as autoras entendem que “a redução da forma de gerúndio já se tornou um fenômeno bastante recorrente no português falado em Dourados e região” (MARTINS e BUENO, 2011:22). No tocante à Faixa Etária, os indivíduos do sexo masculino com 20 a 50 anos (53% - 16/30) empregam mais o apagamento do que aqueles com mais de 50 anos (20,6% - 7/34). Entre as mulheres a distribuição dos dados é equilibrada: 79,2% (19/24) para as de 20 a 50 anos e 82,3% (14/17) para as mais velhas. Observa-se algo semelhante para elas com respeito à sua escolarização: 79% (15/19) de apagamento entre as alfabetizadas e 81,8% (18/22)

para as não alfabetizadas. Já entre os homens, encontraram-se 34,2% (14/41) de apagamento para os alfabetizados e 39,1% (9/23) para os não alfabetizados. Contrariando os resultados dos outros trabalhos mencionados, as autoras encontraram indícios de que para Dourados e Ponta Porã o apagamento é verificado em maior número na fala de mulheres, inclusive as de idade mais avançada; em relação à variável Escolaridade, para as mulheres os resultados foram bem próximos, mas quanto aos homens, a partir dos resultados estatísticos os dados não se revelaram significativos.

Com dados da amostra do Projeto Norma Oral do Português Popular de Fortaleza, Nascimento, Araújo e Carvalho (2013), investigaram o apagamento na fala de 24 informantes, estratificados, assim como nas outras, por Sexo/Gênero, 3 Faixas Etárias (15 a 25, 26 a 49, 50+) e 2 níveis de Escolaridade (alfabetizados ou com Ensino médio). No total de 477 ocorrências de (NDO), há 74,2% de supressão de /d/.

Tabela 1.2 - Distribuição geral de apagamento de /d/ para Fortaleza/CE

		N	%
Sexo/Gênero	Feminino	179/221	81
	Masculino	166/244	68
Faixa Etária	15-25	105/137	77
	26-49	144/166	87
	50+	96/162	59
Escolaridade	Alfabetizados	170/191	89
	Ensino Médio	175/274	64

Os testes estatísticos mostram que os informantes com baixa Escolaridade (.75), as mulheres (.57) e os da faixa de 26 a 49 anos (.71) favorecem o apagamento em relação aos de alta Escolaridade (.31), os homens (.43) e os informantes das faixas de 15 a 25 anos (.54) e os com mais de 49 anos (.25). Considerando os resultados encontrados, assim como no trabalho de Martins e Bueno (2011), verificou-se que as mulheres fortalezenses estão mais suscetíveis a apagar o /d/ em verbos no gerúndio que os homens.

O estudo mais recente sobre apagamento de /d/ de que se tem notícia é o de Almeida e Oliveira (2017), que analisa dados de fala espontânea de 30 informantes maceioenses. São 524 suas ocorrências de gerúndio, com 42% de apagamento. A tabela 1.3 transcrita abaixo mostra as variáveis significativas em seu estudo: Sexo/Gênero, Extensão do Vocábulo e Contexto Fonético-Fonológico Seguinte.

Tabela 1.3 – Resultados do apagamento para Maceió/AL.

		Apagamento	Total	%	Signif.	Peso Relativo
Sexo/Gênero	Feminino	75	247	30,4	<0,001	.39
	Masculino	144	277	52	<0,001	.61
Extensão do Vocábulo	Polissílabos	75	146	51,4	0,004	.63
	Trissílabos	128	324	39,5	0,625	.51
	Dissílabos	16	54	29,6	0,023	.37
Contexto Fonético-Fonológico Seguinte	Haplologia	77	88	87,5	<0,001	.91
	Vogal	85	168	50,6	0,048	.57
	Consoante	43	144	29,9	0,006	.37
	Pausa	14	124	11,3	<0,001	.12

Fonte: Adaptado de Almeida e Oliveira (2017)

O que os autores trouxeram com este estudo é que em relação às variáveis linguísticas, os contextos favorecedores são proporcionalmente ligados à Extensão do Vocábulo: quanto maior ele, maior a chance de /d/ ser apagado, e o Contexto Seguinte de haplologia (onde consoantes seguintes compartilham propriedades fonéticas com a consoante /d/ dos gerúndios, por exemplo, “*falando do João*”) indica um maior favorecimento por parte desse contexto.

Essa revisão de trabalhos mostra que as variáveis sociais atuam fortemente sobre o fenômeno analisado. Com relação à variável Sexo, os estudos de Martins (1999), Nascimento e Mota (2004), Ferreira (2010) e Vieira (2011) mostram que os homens apagam o /d/ mais frequentemente que as mulheres, ao contrário do que observaram Martins e Bueno (2011) e Nascimento, Araújo e Carvalho (2013), o que traz a esta pesquisa uma dupla hipótese de que pode ser possível que o apagamento seja favorecido na fala dos homens ou das mulheres. No

tocante à Escolaridade, Martins (1999), Ferreira (2010) e Nascimento, Araújo e Carvalho (2013) mostram que, quanto menor a Escolaridade, maior a frequência de apagamento de /d/. Com respeito à Faixa Etária, os trabalhos citados aqui, sem exceção, revelam que a queda de /d/ no morfema de gerúndio ocorre mais na fala dos jovens do que na fala dos informantes acima de 50 anos. Um fator que pode ser indicativo de uma mudança em progresso.

Até o momento, não há estudos sobre a fala paulistana, no que concerne a essa variável já tão estudada em outras comunidades. Cabe verificar, portanto, quais variáveis regulam o apagamento de /d/ numa amostra de falantes da capital paulista. Tal como se ressalta nos estudos anteriormente reportados, é em verbos no gerúndio que o apagamento mais ocorre, mas interessa verificar o que o favorece nessas palavras, além de analisar quão frequente é o apagamento em outros casos. Para os verbos no gerúndio, as hipóteses que serão atestadas buscam responder se os homens apagam mais o /d/ que as mulheres, como em Mollica (1989) e Ferreira (2010), ou se seria o caso das mulheres favorecerem o apagamento como em Martins e Bueno (2011); quanto à idade, busca-se verificar se a variável é estável, como verificou Mollica (1989) para o Rio de Janeiro, ou em caminho de uma possível mudança em progresso e, se quanto maior o grau de Escolaridade a variante considerada padrão seria utilizada com mais frequência, indicando ou não certo prestígio por parte da comunidade, já que, através dos trabalhos aqui comentados, o apagamento não se configura como algo necessariamente estigmatizado e muito menos de uma fala considerada “caipira”.

Capítulo 2

Caminhos teórico-metodológicos

Neste Capítulo, descreve-se a Teoria da Variação e Mudança, assim como os preceitos do Paradigma Quantitativo, que compõem as bases das análises que se apresentam mais adiante; em seguida, descreve-se a amostra da fala paulistana utilizada com as hipóteses sociais e linguísticas sobre os dados de apagamento de /d/ em (NDO).

2.1- Teoria da Variação e Mudança Linguística

A análise da variável (NDO), cujas variantes são a pronúncia com /d/ pronunciado ou apagado é aqui desenvolvida com base no arcabouço teórico da Sociolinguística Variacionista (Labov 1966; 1972 e todos os trabalhos que com esses se iniciaram).

Segundo Tagliamonte (2006), a Sociolinguística é uma subárea da Linguística que estuda a língua em seu contexto social, em que a análise variacionista está inserida e toma como ponto de partida as regras gramaticais e, a partir delas, estuda suas relações com o contato com a sociedade.

A Sociolinguística surgiu com o propósito de reagir à ausência do componente social no Estruturalismo e da Teoria Gerativa, para dar conta da dimensão sócio-histórica do fenômeno linguístico, ou seja, dos fatos concernentes à variação e à mudança linguísticas. O precursor dessa “linguística social” é William Labov, que disseminou os estudos voltados para a relação entre língua e sociedade a partir de seu trabalho seminal – *The social motivation of a sound change*, de 1963, sobre a pronúncia dos ditongos (aw), como em *house* ‘casa’, e (aj), como em *ice* ‘gelo’, no inglês falado pelos habitantes da ilha de Martha’s Vineyard, em Massachusetts nos Estados Unidos.

No Brasil, a teoria variacionista encontrou um dos campos mais férteis para a aplicação de seus postulados e, desde a década de 1970, diversas pesquisas nessa linha têm sido desenvolvidas. Com o Projeto Censo da Variação Linguística do Estado do Rio de Janeiro, no Programa de Estudos sobre o Uso da Língua (PEUL), seu coordenador, o Professor Anthony Naro (UFRJ), foi o pioneiro na adoção da metodologia da Sociolinguística Variacionista no Brasil, com o objetivo de estudar o português falado no Rio de Janeiro.

Considera-se que a variação linguística é um fato característico de todas as línguas, fato este que não era observado anteriormente na pesquisa linguística. Já a mudança linguística, embora fizesse parte dos estudos da Linguística Histórica do século XIX, era analisada como decorrente apenas de pressões internas do sistema linguístico, mas não necessariamente sociais. O que diferencia a teoria Gerativa da teoria Sociolinguística é a primeira característica que aqui se retoma: aquela se ocupava da competência linguística, ou seja, um sistema abstrato e homogêneo; esta, nas palavras de Mendes (2015:125), “ocupa-se em desvendar como a heterogeneidade – ou seja, a variação – se organiza”.

Weinreich, Labov e Herzog (2006[1968]) afirmam que essa heterogeneidade faz parte da competência linguística, sendo assim, necessária para o funcionamento real de qualquer língua e o indivíduo tem a capacidade para codificar e decodificá-la. Labov discorda de Saussure, Chomsky e outros que insistem na homogeneidade necessária do objeto linguístico, que ignoram a heterogeneidade e que consideram a fala como caótica e “desmotivada” (cf. LABOV (2008[1972])). Para os sociolinguistas, a variação e a mudança são inerentes às línguas. A variação não é vista como um efeito do acaso, mas como um fenômeno cultural motivado tanto por fatores linguísticos quanto sociais. A variação não é, portanto, assistemática.

Por este viés, os linguistas dessa vertente buscam estudar a diversidade linguística na estrutura social, privilegiando a fala em diferentes contextos, o significado social das variáveis e o emprego das variantes por uma determinada comunidade de fala. A interação não existe sem a comunidade. Através da vida em sociedade, da interação entre os indivíduos e da herança cultural é que ocorrem as mudanças na língua. Daí a importância de se analisar a relação existente entre língua, sociedade e cultura. Nas palavras de Hora (2004:18):

“a Teoria da Variação opõe-se à ausência do componente social e à concepção de língua que até então impera na linguística estrutural e gerativa. Situa-se em relação ao conjunto língua e sociedade, considerando a variedade das formas em uso como objeto complexo, decorrente dos fatores internos, próprios do sistema linguístico, e dos fatores sociais que interagem no ato da comunicação. A variação da língua constitui, portanto, um dado relevante da teoria e da descrição sociolinguística.”

O foco de interesse da Sociolinguística baseia-se num conflito entre duas faces, a princípio, contraditórias: uma aparente estabilidade *versus* uma constante variação, tanto no nível da fala do indivíduo como na comunidade, fazendo jus ao que Labov (1982) sugere: que nesta linguística “social” seja estudada a estrutura e a evolução da língua no contexto social da comunidade. E é a partir disso que a teoria afirma que as línguas humanas estão em constante mudança, que se propaga de forma gradativa e pressupõe períodos longos de variação em diversos eixos sociais.

Seguindo os pressupostos teóricos de Labov (1982) com relação à variação, têm-se os objetos da descrição linguística, a noção de dialeto e idioleto. Aquele pressupõe a variedade gramatical e fonológica e é distinta de outras variedades da língua; este é a forma de falar característica de um determinado falante. A partir deste ponto de vista, encontra-se um “percalço”: existe uma espécie de julgamento quanto às características sociais de um indivíduo relacionado ao seu modo de falar. Isto implica dizer que uma cultura socialmente dominante consideraria inadequadas as variantes mais comumente empregadas pelas camadas sociais subordinadas. Leva-se em consideração que, dessa perspectiva, quanto às normas de comportamento apropriado de fala, a escolha de uma variedade como norma culta pode ser determinada de várias formas, entre elas: a língua falada pelas camadas sociais mais altas e a língua não falada pelos mais carentes socioeconomicamente. Assim, emergem os conceitos de estigma e prestígio na sociolinguística variacionista: a fala de um grupo estigmatizado socialmente será também estigmatizada, ao passo que a variedade falada pelo grupo mais “prestigiado” socioeconomicamente será também prestigiada.

De certo modo, esta pesquisa busca verificar em que ponto se encontra o apagamento de /d/ em (NDO), nessa dicotomia entre prestígio e estigma. Tal como vimos no capítulo anterior, Amaral (1920) considerou tal variante, quando se trata da pronúncia do gerúndio, como traço de um falar inculto, rude, com estigma. Entretanto, também vimos, com Mollica (1989), que esse não se trata de um fenômeno “socialmente marcado”. Interessa, portanto, discutir os dados paulistanos também dessa perspectiva.

A noção de estigma e prestígio também se correlaciona com os processos de mudança linguística. Estes podem ser classificados como *mudança de cima* ou *mudança de baixo*, termos esses que foram estabelecidos a partir do estudo de William Labov em Nova Iorque em 1966. Conforme o autor (LABOV, 1994:78) “*mudanças de cima* são introduzidas pela Classe Social dominante, geralmente com plena consciência por parte do falante⁴”. Essas mudanças não são incorporadas abruptamente no vernáculo da classe dominante, mas sim na fala cuidadosa envolvendo uma série de adaptações até tornar-se consistente. Os casos de *mudança de baixo* são (LABOV, 1994:78) “mudanças sistemáticas que aparecem primeiramente no vernáculo, e representam o funcionamento de fatores internos – linguísticos⁵.” Esses casos de mudança estão abaixo do nível de consciência do falante, de forma que ninguém percebe ou fala sobre eles. Podem ser introduzidos por qualquer Classe Social, mas, geralmente, as Classes Sociais mais baixas atuam como grupo inovador.

Além disso, estigma e prestígio envolve, destacadamente, a análise perante a variável Classe Social, na expectativa de um cenário em que os falantes das Classes mais altas exibem maior proporção de uso das formas de prestígio do que os falantes de Classes mais baixas. Considere-se ainda que, de acordo com o trabalho de Labov (1982) em Nova Iorque, falantes de Classe Média podem ultrapassar os da Classe mais alta, no emprego da forma que detém prestígio na comunidade. É o que ele verificou no caso da pronúncia rótica de /-r/ pós-vocálico. Além disso, é geral a verificação de que os processos de mudança tendem a ser liderados pelos

⁴ Tradução própria de: *Changes from above* are introduced by the dominant social class, often with full public awareness.

⁵ Tradução própria de: *Changes from below* are systematic changes that appear first in the vernacular, and represent the operation of internal, linguistic factors.

indivíduos mais integrados da Classe Média Baixa e/ou das seções mais elevadas da classe operária (cf. LABOV, 1982:77), constituindo um processo de mudança de baixo.

Quanto à Sexo/Gênero, em razão dos muitos estudos que isso verificaram, espera-se que as mulheres tendam a evitar formas desprestigiadas ou estigmatizadas numa comunidade, se por acaso uma certa variante de uma certa variável é mais empregada pelas mulheres, relativamente aos homens, entende-se que assim o fazem justamente porque tal forma não é, no geral, positivamente avaliada naquela comunidade. Entretanto, pode-se dizer que há certa controvérsia no que diz respeito a tendências gerais na comunidade, no que concerne ao sexo gênero: as mulheres empregam mais uma forma variante porque ela tem prestígio na comunidade, ou certa forma acaba tendo avaliação positiva na comunidade em virtude de ser mais empregada pelas mulheres? De qualquer forma, em termos de padrões gerais, muitos trabalhos observam essa correlação – em diferentes comunidades, para diferentes línguas (MOLLICA, 1989; FERREIRA, 2010).

Além de tratar da noção de prestígio/estigma, a Teoria da Variação e Mudança rompe com a dicotomia entre sincronia/diacronia proposta por Saussure (1916), aproximando-as (cf. TARALLO, 1994). As mudanças não ocorrem de forma abrupta, como preconiza a perspectiva gerativa (CHOMSKY, 1957; LIGHTFOOT, 1999), já que toda mudança implica processos de variação. O inverso, contudo, nem sempre se observa.

Por meio do estudo da variação, pode-se entender a direção, bem como fazer generalizações, acerca da mudança – que não necessariamente consiste na substituição de uma variante por outra, mas que pode consistir numa evolução do sistema de (des)favorecimento do uso, da ocorrência de uma certa variante, de uma certa variável.

2.2 - Paradigma Quantitativo

A partir dos estudos de William Labov, nas décadas de 1960 e 1970 em Nova Iorque e na Filadélfia, o paradigma quantitativo desenvolveu-se nos estudos sociolinguísticos,

rapidamente estabelecendo-se na área (cf. BAYLEY, 2002). Hoje em dia, ao falar em “sociolinguística”, sobretudo quando se usa o adjetivo “variacionista”, pressupõe-se que se procede a uma análise quantitativa das formas em foco: quão (in)frequentemente ocorrem em múltiplos contextos (definidos pelos chamados fatores, ou variáveis independentes) e, idealmente, como proporções de dados extraídos de uma amostra podem ser tomadas como indícios de tendências em uma comunidade de falantes.

De acordo com Guy (1993), para dar início a uma pesquisa deste cunho, são necessários três passos: a coleta de dados; a redução e extração dos dados; sua interpretação.

A coleta nada mais é do que a forma como se obtém os dados que serão posteriormente analisados. Alguns cuidados devem ser tomados para isso: deve ser feito um planejamento prévio para verificar se os dados coletados refletem o fenômeno que se quer estudar, se os procedimentos utilizados para a coleta são confiáveis e replicáveis, o que pode ser feito a fim de minimizar o enviesamento de dados, e se a amostra de dados idealizada é representativa da população. Tal como já se comentou anteriormente, o presente estudo utiliza uma amostra da fala paulistana previamente construída e disponibilizada para fins de pesquisas linguísticas (MENDES e OUSHIRO, 2012). Comumente chamada de SP2010, essa amostra é brevemente descrita no item a seguir.

O segundo passo envolve a extração dos dados que serão utilizados para compor as análises e verificar se é necessária sua possível redução. Nessa etapa, costumam-se utilizar métodos para lidar com um grande número de dados a fim de tornar possível a tarefa de análise e compreensão, e permitir a identificação de tendências e padrões gerais. Por exemplo, Oushiro (2015) trabalhou com um corpus composto com 118 entrevistas que continha mais de 25 mil ocorrências de /e/ nasal e mais de 60 mil de /-r/, analisar todas seria uma tarefa desnecessária, visto que padrões tendem a se revelar a partir de um número pequeno de dados e a se manter constantes a partir de certo ponto. A autora, por exemplo, verificou que com 50 ocorrências de cada variante (/e/ nasal, /-r/, por exemplo) para cada um dos 118 informantes - um total de 5900 dados - foi possível identificar os padrões esperados e sua manutenção constante.

Em seguida, a fase final engloba a descrição dos resultados alcançados, em que se procura formular generalizações a partir do princípio da modelagem quantitativa e o princípio de múltiplas causas.

O princípio da modelagem quantitativa (cf. BAYLEY, 2013[2002]) mostra que se podem examinar de perto as formas variantes (de uma mesma variável) e observar as características do contexto em que elas ocorrem – desde que haja dados suficientes para sustentar as hipóteses propostas. Neste caso, “contexto” diz respeito ao ambiente linguístico em que se situa o fenômeno estudado.

Já o princípio de múltiplas causas (cf. BAYLEY, 2013[2002]) sustenta a improbabilidade de que um fator contextual aleatório possa explicar a variabilidade observada em um dado específico da língua. Por exemplo, na variação encontrada em (NDO), com ou sem apagamento de /d/, pode haver maior menor probabilidade de ocorrência do apagamento de acordo com o Sexo/Gênero do falante, Classe Social, Escolaridade, etc.

Outra questão importante é verificar se falantes, individualmente, diferem em sua taxa de uso das variantes de certa variável (Guy, 1993). Geralmente, espera-se que os falantes de uma mesma comunidade de fala se assemelhem na aplicação de determinada regra variável. Assim, oferece uma maneira de entender como grupos de falantes que usam uma variante com taxas muito diferentes podem ser considerados membros da mesma comunidade de fala. Segundo Labov (1972:120-121), convém salientar que

[uma] comunidade de fala não é definida por qualquer marca no uso de elementos da língua, mas pela participação em um conjunto de normas compartilhadas. Tais normas podem ser observadas em momentos explícitos de avaliação do comportamento e pela uniformidade de padrões de variação que são invariantes em relação a níveis particulares do uso.

Essa afirmação nos lembra da necessidade de que a amostra de fala seja representativa da população, ou seja, da comunidade de fala.

Para uma análise quantitativa, é imprescindível a inclusão de um procedimento estatístico que permita alguma generalização a partir dos dados. Na presente pesquisa, na qual se defende a hipótese de que a variação na realização de (NDO) é possivelmente influenciada por algumas variáveis independentes (estabelecidas com base em estudos anteriores e numa análise qualitativa dos dados), vem ao caso uma análise multivariada realizada através da plataforma R (R Core Team, 2018). Oushiro (2014:134) explica que “o R é uma linguagem de programação voltada à análise de dados, que pode ser utilizada para realizar computações estatísticas e gráficas, compilar e anotar corpora, produzir listas de frequências, entre diversas outras tarefas”.

2.3 - A amostra da fala paulistana

A amostra da fala paulistana utilizada para este trabalho, tal como já se mencionou anteriormente, é chamada de SP2010 e foi coletada e organizada pelo Grupo de Estudos e Pesquisa em Sociolinguística da Universidade de São Paulo (GESOL-USP). Ela contém 60 entrevistas com falantes paulistanos (cujos pais podem ser paulistanos também ou não), todas elas disponíveis no site do projeto⁶.

O roteiro das entrevistas foi elaborado no sentido de obter, além de uma amostra recente da fala paulistana, informações pessoais dos informantes que eventualmente possam ser utilizadas em diferentes projetos de pesquisas. Esse roteiro é dividido em duas partes: a primeira é de cunho mais pessoal, em que o falante é conduzido a falar sobre o bairro onde mora/morou, sua infância, sua família, seu trabalho e seus momentos de lazer. A segunda trata de assuntos mais específicos acerca da cidade de São Paulo. De acordo com Mendes e Oushiro (2012: 976),

“busca uma fala cujo estilo – no sentido de Labov (2001b) – seja potencialmente mais distanciado e dissertativo/argumentativo. Essa parte também contém a leitura de uma lista de palavras, de uma notícia de jornal e de um texto com características da oralidade, na qual se

⁶ www.projetosp2010.fflch.usp.br

espera maior grau de monitoramento da fala, em diferentes graus (LABOV, 2001b, 2006); ela se encerra com questões sobre as avaliações dos falantes a respeito de três variantes sociolinguísticas específicas: a pronúncia ditongada de /e/ nasal (em palavras como “entendendo”), a ausência de concordância nominal (p.ex. “dois pastel”) e a pronúncia retroflexa de /r/ em coda silábica (em palavras como “porta”).

Para composição da amostra, foram gravadas as falas de 60 informantes estratificados de acordo com: Sexo/Gênero, Faixa Etária e Escolaridade, formando os 12 perfis definidos pelas variáveis no quadro abaixo:

**Quadro 1 – Variáveis estratificadoras da amostra SP2010
e perfis sociolinguísticos dos informantes**

Sexo/ Gênero	Faixa Etária	Escolaridade	Perfil
Feminino	20-34 anos	Até Ens. Médio	1. F1C
		Ens. Superior	2. F1S
	35-59 anos	Até Ens. Médio	3. F2C
		Ens. Superior	4. F2S
	60+ anos	Até Ens. Médio	5. F3C
		Ens. Superior	6. F3S
Masculino	20-34 anos	Até Ens. Médio	7. M1C
		Ens. Superior	8. M1S
	35-59 anos	Até Ens. Médio	9. M2C
		Ens. Superior	10. M2S
	60+ anos	Até Ens. Médio	11. M3C
		Ens. Superior	12. M3S
12 perfis x 5 informantes = 60 entrevistas			

Fonte: Mendes e Oushiro (2012:978)

Foram gravados 5 informantes para cada perfil, um de cada Zona da Cidade (Leste, Oeste, Norte, Sul, Centro). Além disso há paulistanos de uma região mais central e de uma região mais periférica entre esses cinco informantes de cada perfil. Essa característica auxilia a desvendar a hipótese de que numa grande metrópole como São Paulo, os habitantes de diferentes partes da cidade possam apresentar diferentes comportamentos sociolinguísticos.

Para as variáveis que estratificam a amostra, pode-se dizer que é enorme a quantidade de trabalhos sociolinguísticos que incluem a variável Sexo/Gênero – de acordo com Mendes e Oushiro (2012:979), “desde aquele que se considera o primeiro manual desenvolvido para a área (CHAMBERS, 1995), até manuais mais recentes (CHAMBERS; TRUDGILL; SCHILLING-ESTES, 2004; HOLMES; MEYERHOFF, 2005)”. É desejável que uma amostra inclua Sexo/Gênero na sua estratificação porque há expectativas de resultados relacionados às diferenças entre os papéis sociais desempenhados por homens e mulheres numa comunidade de fala. Sobre a comunidade em foco:

(...) papéis sociais estão em constante processo de construção e reconstrução e, no Brasil, São Paulo talvez seja uma das cidades em que estilos de vida, hábitos urbanos e práticas sociais em geral mudem mais rapidamente. (MENDES e OUSHIRO, 2012:979)

Todos os trabalhos que envolvem o papel do Sexo/Gênero atentam para três pontos:

- (i) a frequência de variantes prestigiadas é maior na fala das mulheres, relativamente à dos homens nas sociedades ocidentais e urbanas;
- (ii) as mulheres tendem a se revelar mais conservadoras em mudanças linguísticas “de baixo”, aquelas que Labov (2008[1972]:152) entende por “abaixo do nível da percepção consciente” do falante;
- (iii) as mulheres tendem a ser mais inovadoras em mudanças “de cima” vistas por Labov (2008[1972]:152) como “processo explícito de correção social aplicado a formas linguísticas individuais”.

Quanto à Faixa Etária, os paulistanos foram agrupados em três: 20 a 34 anos (1^a.); 35 a 59 anos (2^a.); 60 anos ou mais (3^a.). Não se incluíram crianças e adolescentes para não aumentar sobremaneira o número de perfis sociolinguísticos da amostra. De acordo com Mendes e Oushiro (2012:980)

a divisão em pelo menos três faixas faz-se necessária a fim de verificar variantes em retração ou difusão (em que normalmente se verifica um padrão de curva ascendente ou descendente) e variantes de variáveis estáveis (em que normalmente se observa um padrão de curva em “U”).

A importância desse grupo de fatores é semelhante à de Sexo/Gênero. Ambos, conjuntamente ao grau de Escolaridade, estão entre os mais analisados nos estudos variacionistas. Esse grupo de fatores ajuda-nos a compreender o conceito de mudança em tempo aparente, através do qual “é possível detectar mudanças linguísticas em progresso através da comparação da fala de pessoas de diferentes idades pertencentes a mesma comunidade”. (Labov, 2001a, apud MENDES e OUSHIRO, 2012:980).

Finalmente, a variável Escolaridade foi dividida em dois níveis: aqueles que possuem até o Ensino Médio e aqueles que têm Ensino Superior, completo ou não. A divisão em apenas dois níveis de Escolaridade se deu por conta das mudanças nas políticas públicas de educação nas últimas décadas, como a Progressão Continuada (no estado de São Paulo) e a ampliação do acesso aos cursos universitários pelo ProUni (em âmbito nacional), que indicam que está cada vez mais raro encontrar paulistanos que não tenham pelo menos ingressado no Ensino Médio (cf. MENDES e OUSHIRO, 2012).

A hipótese concernente a esta variável é a de que falantes com graus mais elevados de instrução empreguem com maior frequência as formas consideradas padrão ou de prestígio na comunidade.

2.4 – Outras variáveis sociais

Além das variáveis acima, que estratificam a amostra, outras foram incluídas neste estudo de (NDO), de acordo com o quadro 2:

Quadro 2 – Variáveis sociais

Sexo / Gênero	Feminino
	Masculino

Faixa Etária	25 a 34 anos
	35 a 59 anos
	60 anos ou mais
Escolaridade	até Ensino Médio
	Ensino Superior
Região	Centro Velho
	Centro Expandido
	Periferia
Origem dos Pais	Estrangeiros
	Interior SP/MG/PR
	Norte/Nordeste
	São Paulo – capital
	Mista (pais de origem diferente)
Mobilidade	Sempre morou no mesmo bairro
	Sempre morou na mesma zona
	Morou em diferentes zonas
Classe Social	Alta
	Média alta
	Média media
	Média baixa
	Baixa alta
Índice Socioeconômico	1 – 5

A variável Região da Cidade é composta por três fatores: Centro velho (Bela Vista, Consolação, Bom Retiro), Centro expandido (Santana, Moema, Tatuapé, Lapa, Pinheiros, Vila Mariana), que é a área da cidade localizada ao redor do centro histórico (ou velho) e Periferia (Sapopemba, Pirituba, Jabaquara, Parque do Carmo, Itaquera, Perus). Apesar de não estratificar a amostra, assim como as outras variáveis que se seguem, esta variável serve para dar efetividade à questão relacionada ao fato de se bairros mais periféricos, por serem mais recentes historicamente e possuírem uma infraestrutura urbana menos desenvolvida, poderiam ser reveladores de maior ocorrência de apagamento na fala dos indivíduos residentes nesta região. Segundo Oushiro (2015: 45)

bairros mais centrais têm, em geral, um custo de vida bastante superior àquele de bairros mais periféricos, algo que se verifica em preços médios de habitação e de serviços; desse modo, assim como o Nível de Escolaridade, a Região de Residência também é um índice indireto de status socioeconômico do falante.

Tendo em vista esta consideração, o apagamento como variante não padrão poderia ser considerada predominante na fala daqueles provenientes de Classe Social mais baixa, neste caso supostos moradores da Região Periférica.

Os bairros das Regiões Periférica e Central podem ser identificados na figura 1:

Fonte: Oushiro (2015)

Figura 1 Divisão da cidade de São Paulo entre regiões mais centrais e mais periféricas

Seguindo o pensamento de Oushiro (2015) e sabendo que São Paulo é considerada um lugar de “mistura” étnica, e que aproximadamente metade da população não é nativa da cidade, a variável origem dos pais investiga se falantes cujos pais não são paulistanos apresentam

tendências para o apagamento em maior taxa que os filhos de paulistanos. A figura 2, a seguir, revela a diversidade em relação à origem da população economicamente ativa na cidade.

Figura 2 Origem da população economicamente ativa (30-60 anos) na Região Metropolitana de São Paulo. (Fonte: IPEA 2011)

Quanto a Mobilidade Geográfica, a hipótese por trás desta variável é a de que “falantes cuja mobilidade é menor tendem a se orientar por normas locais, ao passo que aqueles com maior mobilidade, menos enraizados, adotam valores externos e das classes mais altas” (MILROY, 1987[1980] apud OUSHIRO, 2015: 47). Dessa forma, propõe-se que o maior número de ocorrências de apagamento estaria indicado nas falas de entrevistados com menor mobilidade geográfica.

Assim como as duas variáveis anteriores, Origem dos Pais e Mobilidade Geográfica, a variável Classe Social é incomumente analisada no contexto brasileiro pela dificuldade de se definir os critérios utilizados para seus fatores. No lugar desta variável, diversos estudos utilizam-se do nível de Escolaridade dos falantes como Índice de Classe Econômica. Segundo Oushiro (2015:48) em

estudos que examinaram Classe Social sociolinguisticamente, é consenso que a variável não se refere apenas à renda familiar ou à renda do informante; ela costuma ser operacionalizada como um índice composto que normalmente abarca nível de Escolaridade, ocupação e

renda (...)e, além desses, valor do imóvel e sua manutenção em geral, mobilidade em relação aos pais e bairro de residência.

Seguindo esses critérios salientados pela autora, optou-se por incrementar nas análises desta pesquisa a variável Índice Socioeconômico, que é composta com base em quatro parâmetros (cf. figura 3):

- (i) nível de Escolaridade do entrevistado;
- (ii) ocupação do entrevistado;
- (iii) renda média familiar do bairro de residência;
- (iv) média dos índices de Escolaridade e ocupação dos pais do informante.

$$\frac{Escolaridade + Ocupação + Renda + \frac{EscPai+OcupPai+EscMãe+OcupMãe}{4}}{4}$$

Figura 3 Fórmula para cálculo do Índice Socioeconômico (Fonte: Oushiro, 2015:51)

O intuito dessas duas variáveis é verificar os indivíduos de classe mais baixa favoreçam a regra para o apagamento, assim como verificado em outros trabalhos como o de Labov (2006[1966]).

Outra característica que pode ser observada através da análise perante a Classe Social é se existe hipercorreção na propagação da mudança linguística (LABOV, 1966; 1972). A hipercorreção se dá quando “falantes da classe média baixa vão além do grupo de status mais elevado em sua tendência a usar as formas consideradas corretas e apropriadas para estilos formais” (LABOV, 1972:155), a partir da distribuição dos dados de apagamento quanto às Classes Sociais se constatará a existência ou não de hipercorreção por parte dos falantes para a fala paulistana.

2.5. Variáveis linguísticas independentes

Além das variáveis sociais listadas acima, foram incluídas diferentes variáveis linguísticas que caracterizam os dados de NDO encontrados na amostra que podem ser verificadas no quadro abaixo:

Quadro 3 – Grupos de fatores linguísticos

Classe Morfológica	Adjetivos Conjunções Numerais Substantivos Verbos no presente Verbos no gerúndio
Tipo de Verbo no Gerúndio	Orações Reduzidas Adverbiais Orações Perifrásticas Justaposição
Verbo Auxiliar da Perífrase	Acabar Estar Ficar Ir (...)
Tamanho do Verbo Auxiliar ESTAR	Pleno Reduzido Sem
Material Interveniente	Com Sem
Conjugação	1ª - ar 2ª - er 3ª - ir
Tamanho do Verbo Principal	2 sílabas 3 sílabas

	4 ou + sílabas
Contexto Seguinte	Silêncio
	Não Silêncio

Partindo das Classes Morfológicas em que se encontram palavras terminadas por (NDO), a considerar que a maioria dos trabalhos acerca do apagamento não verificam outros contextos morfológicos que não os verbos no gerúndio, decidiu-se verificá-las, assim como Mollica e Mattos (1992) e Ferreira (2010) fizeram.

Logo em seguida, os casos de verbos no gerúndio permitiram que fossem analisadas diferentes Estruturas Sintáticas (Orações Reduzidas Adverbiais, Perífrases Verbais ou Justaposição), nos exemplos (a), (b) e (c) abaixo, que poderiam ser mais propícias ao apagamento de /d/.

a) Orações Reduzidas Adverbiais

(8)S1:eu saio de casa todo dia *amaldiçoando* os prefeito que autorizaram fazer esses prédio

(...)

S1:então eu passei uma semana *zoneando* a classe e a professora chamou minha mãe e falou que tinha que me tirar da escola (Luciana N., F2S)

b) Perífrases

(9)S1: eu acredito porque a ce/ a cidade de São Paulo *está ficando* estressante mesmo né principalmente... eh na área de transporte né que está terrível

(...)

S1: vai ser muito difícil conforme *vai passando* o tempo... o governo não *vai fazendo* nada entendeu tem que se começar alguma coisa e nada se começa né (Celina B., F2C)

c) Justaposição⁷

⁷ este tipo de oração de gerúndio é o que Cunha e Cintra (2010:506) classificam como construção afetiva no qual o aspecto inacabado do verbo no gerúndio exprime a ideia de progressão indefinida.

(10)S1: e eu não ia ter que me preocupar com dinheiro então... eu podia ficar dois anos *pesquisando escrevendo*...
(Ariane G., F2S)

(...)

S1: eu sempre tive... uma certa dificuldade principalmente porque a minha irmã... era a mais velha que eu ela só dava uma lida e lembrava tudo eu precisava...

S1: ficar *repetindo repetindo repetindo*... mas como eu já não tinha vontade então... piorou né? (Darcy M., F3C)

As Perífrases Verbais, das três Estruturas Sintáticas que envolvem os verbos no gerúndio, são o contexto mais produtivo quanto a sua estrutura, e inclusive Ferreira (2010) verifica que é também o contexto mais propício ao apagamento. A partir de sua produtividade estrutural foram selecionadas certas características a fim de serem analisadas como variáveis linguísticas, como o caso do Verbo Auxiliar da Perífrase. Com esta variável, pretende-se verificar se os verbos considerados auxiliares propriamente ditos (ser, estar; cf. LOBATO, 1975) favorecem mais a aplicação do apagamento em gerúndios que os verbos de movimento (ir, vir, andar) e que os verbos semi-auxiliares (ficar, começar, acabar, etc) que também funcionam como auxiliares nas Orações Perifrásticas.

Tendo em vista que o verbo auxiliar estar é considerado o verbo auxiliar mais frequentemente utilizado em Orações Perifrásticas de gerúndio, optou-se por averiguar se a redução fonológica de estar, “tá”, seria um fator favorecedor para o apagamento no verbo subsequente ao auxiliar. Trabalhos envolvendo a gramaticalização de “estar + gerúndio”, como o de Mendes (1999), verificaram que a forma reduzida ocorre com mais frequência quando seguida de verbos no gerúndio, desta forma, traz-se para esta pesquisa a hipótese de que o verbo auxiliar estar em sua forma reduzida seja favorecedor ao apagamento em verbos no gerúndio.

Ainda nas Orações Perifrásticas pode ocorrer Material Interveniente entre o verbo auxiliar e o verbo no gerúndio. A inserção desta variável justifica-se pela hipótese sintática que sugere que quando há Material Interveniente existe um enfraquecimento na unidade semântica da perífrase. Assim, quanto mais distante o verbo auxiliar estiver do verbo na forma de gerúndio, menor seria a aplicação da regra de apagamento.

Partindo para características do Verbo Principal da Oração Perifrástica, com relação à Conjugação dos mesmos, segundo Ferreira (2010), a expectativa quanto a esta variável era de que não houvesse diferença relevante entre as diferentes conjugações para a fala rio-pretense. Para esta pesquisa, decidiu-se por uma hipótese diferente: considerando que a primeira Conjugação latina possuía aproximadamente 3.620 verbos e, conforme Coutinho (1976), é essa a que ainda se conserva no português, mostrando-se mais produtiva até mesmo na formação de verbos estrangeiros que são incorporados ao léxico da língua portuguesa (ex. delete > deletar), seria a primeira Conjugação verbal um fator favorecedor ao apagamento?

Os verbos conjugados no gerúndio apresentam tamanhos diferentes quanto ao número de sílabas, a partir dessa característica direciona-se a hipótese verificada em Mollica e Mattos (1992: 56), e corrobora-se com os dados da tabela 2.1:

quanto maior é o item, maior a chance de se operar a assimilação⁸. Isso pode ser explicado se levarmos em consideração fatores de processamento que atuam no nível da unidade da palavra: os segmentos tendem à não realização, quando as cadeias vocabulares são grandes.

Tabela 2.1 – Apagamento por Extensão do Vocábulo na fala carioca

	Ocorrências	%
Dissílabo	308/1619	19
+ 2 sílabas	949/2482	37

Assim, considera-se que o número de apagamento será maior para os verbos com 3, 4 ou mais sílabas como também foi verificado por Vieira (2011).

Outra variável linguística analisada é o Contexto Seguinte aos verbos no gerúndio. É possível verificar a utilização desta variável em diferentes pesquisas sociolinguísticas (MOLLICA e MATTOS, 1992; LUCENA e VASCONCELOS, 2007). Para esta pesquisa, segue-se o mesmo critério de Mollica e Mattos (1992), onde tem-se dois fatores: silêncio (ou

⁸ Assimilação é o termo utilizado por estas autoras para referir-se ao que chamamos aqui de Apagamento.

pausa) e não silêncio (qualquer outro ambiente fonológico que não seja a pausa como apresentado na tabela abaixo).

Tabela 2.2 – Apagamento por Contexto Seguinte na fala carioca

	Ocorrências	%
Silêncio	139/412	34
Não silêncio	857/2184	39

Fonte: Adaptado de Mollica e Mattos (1992)

Foi observado que /d/ em (NDO) tende a ser mais preservado quando seguido de silêncio, “pois neste ambiente não há qualquer fonema próximo que favoreça a assimilação” (MOLLICA e MATTOS, 1992: 57). Esta é a hipótese que esta pesquisa se propõe a responder.

2.6 Variáveis específicas para subconjuntos de dados

Nem todas as variáveis propostas e descritas acima fazem parte de todos os subconjuntos de dados. Para as 3 estruturas sintáticas de gerúndio: *Perífrases*, *Orações Reduzidas* e *Justaposição* propostas neste trabalho, algumas variáveis foram estabelecidas conforme suas características, sistematizadas no quadro 4 abaixo:

Quadro 4 – Grupo de Fatores Linguísticos para os Subconjuntos de Dados do Gerúndio

Subconjunto de Dados	Grupo de Fatores
Orações Reduzidas Adverbiais	Conjugação Tamanho do Verbo

	Contexto Seguinte
Orações de Gerúndio em Justaposição	Material Interveniente Tempo do Verbo Auxiliar Verbo Auxiliar da Perífrase Tamanho do Verbo Auxiliar Conjugação Tamanho do Verbo Principal Contexto Seguinte
Orações Perifrásticas	Material Interveniente Tempo do Verbo Auxiliar Verbo Auxiliar da Perífrase Tamanho do Verbo Auxiliar Conjugação Tamanho do Verbo Principal Contexto Seguinte

Capítulo 3

Conjunto de Dados

O capítulo anterior descreveu brevemente a amostra da fala paulistana utilizada para compor esta pesquisa, a partir de uma revisão dos apontamentos teóricos essenciais da sociolinguística variacionista e de seus procedimentos teórico-metodológicos. Em seguida foram descritas as variáveis sociais e linguísticas utilizadas para compor esta pesquisa. A seguir, o presente capítulo concentra-se nos dados extraídos e na discussão descritivas das ocorrências de palavras com (NDO).

3.1- Conjunto Geral de Dados

As ocorrências de palavras terminadas em (NDO) foram manualmente extraídas das 60 entrevistas do ProjetoSP2010. Do total de 6.586 dados, aproximadamente 18% são representados por casos de apagamento de /d/ - que, daqui em diante, serão chamados de *dø* e, para os casos em que /d/ se mantém serão chamados de *d₁*:

Tabela 3.1 - Distribuição geral de apagamento de /d/ em (-ndo)

Dø	%	/d/ pronunciado	%	Total
1.164	17,9	5.422	82,1	6.586

Esses dados encontrados se organizam em 6 subconjuntos, de acordo com as classes morfológicas das palavras terminadas em (NDO): (1) adjetivos, (2) conjunções, (3) numerais, (4) substantivos, (5) verbos no presente e (6) verbos no gerúndio⁹:

- ☉ (1) D1: legal agora queria que você lesse... essa notícia pra mim em voz alta também por favor

⁹ Clipes de áudio estão disponíveis para os exemplos (1) a (6) em <https://drive.google.com/open?id=1GnZFCRLZZRQUSEF8LLwL8aEooePhMyQm>

S1: esse meu nariz entupido está *lindo* (d₁) ler (Janaína B., F1C)

- ⊙ (2) S1: aqui assim não era nem festa assim o que tinha/ teve aqui uma época *quando* (dø) eu era pequeno era (Alberto M., M2C)

(...)

- ⊙ D1: ah está errado falar assim?

S1: *segundo* (d₁) eles né (Dalson O., M1C)

- ⊙ (3) S1: e aí meu avô morreu

S1: o meu pai... que é o *segundo* (d₁) filho (Lucas S., M1C)

- ⊙ (4) S1: cumprimento todo *mundo*¹⁰ (d₁) não sei o quê mas raríssimas vezes eu lembro o nome de alguém
S1: e então dentro do prédio realmente amizade... praticamente não tenho (Darcy M., F3C)

- ⊙ (5) S1: me dou bem adoro fazer isso que eu faço

S1: quanto a isso não não me *arrependo* (dø) não

D1: e você

D1: e você gostaria de fazer uma outra coisa? (Alberto M., M2C)

- ⊙ (6) S1: a minha mãe tem contato com todo mundo então minha mãe está sempre me *ligando* (dø) *falando* (dø) como que está lá (Ana S., F1C., F3S)

A Tabela 3.2 dispõe as ocorrências de dø e d₁:

Tabela 3.2 - (NDO) por classe morfológica

	d ø	%	d ₁	%	Total
Adjetivos	0	0	67	100	67
Numeral	0	0	60	100	60
Substantivos	0	0	661	100	661
Verbos no presente	1	1	97	99	98
Conjunções	93	6	1.464	94	1.557
Verbos no gerúndio	1.070	26,3	3.073	73,7	4.071
Total	1.164	17,9	5.422	82,1	6.586

¹⁰ Diversas ocorrências da palavra *mundo* foram observadas em contexto de pronome indefinido (*todo mundo*). Como não houve apagamento nesses casos, todos foram classificados como “substantivo”, ou seja, sem identificar especificamente os que constituem tal pronome.

Não há, portanto, ocorrências de *dø* em adjetivos, numerais (*segundo*) e substantivos terminados em (NDO). Encontrou-se apenas 1 caso de *dø* em verbos no presente (exemplificado em 5 acima). A variação na pronúncia de (NDO) é visivelmente mais robusta quando se trata de conjunções (6% de *dø* em 1557 dados) e verbos no gerúndio (26% de *dø* em 4.071 casos).

A tabela 3.3 identifica os casos de (NDO) que não são quantitativamente analisados aqui, em razão de que, neles, a variável é quase categoricamente realizada com *d_l* (excetuando-se a ocorrência de “*arrepeno*” (os números entre parênteses indicam quantas vezes cada palavra foi encontrada):

Tabela 3.3 – Ocorrências de (NDO) por classe morfológica

Adjetivo	Numeral	Substantivo	V. Presente
Fundo (8)	Segundo (60)	Bando (4)	Ando (21)
Lindo (47)		Comando (1)	Aprendo (1)
Oriundo (1)		Contrabando (2)	Arrependo (2)
Profundo (2)		Fernando (3)	Atendo (11)
Redondo (5)		Fundo (20)	Confundo (1)
Tremendo (1)		Gumercindo (1)	Defendo (1)
Vagabundo (3)		Mundo (622)	Dependo (1)
		Segundos (2)	Entendo (19)
		Submundo (1)	Escondo (3)
		Vagabundo (5)	Mando (5)
			Ofendo (1)
			Pretendo (18)
			Recomendo (1)
			Vendo (1)

De acordo com o que se apresenta na tabela 3.2, as análises a seguir concentram-se nos casos de Conjunção (quando e segundo) e verbos no gerúndio, que são os mais numerosos.

3.2 - Quando

Entre as 1557 ocorrências de conjunções, encontram-se apenas 6 casos de *segundo* (todos de d1). Dos 1151 casos de *quando*, ocorre dØ em apenas 6%.

Tabela 3.4 Apagamento de /d/ nas conjunções *quando* e *segundo*

	d Ø	%	Total
segundo	0	0	6
Quando	93	6	1.551

Esses dados são aqui analisados de maneira descritiva – mesmo os de *quando*, já que 93 casos de apagamento representam muito pouco, quando o total é mais de dez vezes maior.

Entre os 60 falantes da amostra, apenas na fala de 21 deles ocorre apagamento de /d/ em *quando*. Esses falantes estão dispostos na figura 4 (com as barras correspondentes a apagamento em negrito). Por conta desse número que representa apenas um terço do quantitativo de falantes que compõem a amostra de fala, decidiu-se por não analisar os dados de forma inferencial.

Figura 4 – Frequências de dØ e d1 em *quando* para os indivíduos não categóricos da amostra SP2010

Os falantes que mais apagam o /d/ em *quando* são SandroS (16 casos), MadalenaM (14 casos), AlbertoM (12 casos), CarolinaA (9 casos) e RuthC (9 casos) – em ordem decrescente por número de casos. A figura abaixo mostra as configurações acústicas de ocorrência de *quando* com /d/ apagado e /d/ pronunciado na fala de MadalenaM:

Figura 5 Configurações acústicas de quando com **d**

Na figura 5, verificam-se as configurações de uma realização da palavra quando no momento em que /d/ não é pronunciado. Para fins de comparação, a parte em preto e branco do lado esquerdo da figura, temos um padrão de realização de /n/ segundo o Manual de Fonética Acústica e Experimental (BARBOSA; MADUREIRA, 2015), à direita, temos a identificação

de /n/ dentro do retângulo vermelho seguido da vogal /u/ demonstrada com as setas vermelhas e nenhuma ocorrência de /d/, que pode ser verificada na figura acima.

Figura 6 - Configurações acústicas de quando com **d₁**

Para o caso de *quando* com /d/ pronunciado, verifica-se no lado esquerdo da figura 6 uma configuração padrão da realização acústica de /d/ e do lado direito tem-se no retângulo vermelho a representação acústica do momento da pronúncia de /n/ e de /d/.

Como são poucos os casos de apagamento na Conjunção *quando* e também poucos os indivíduos não categóricos, a fim de tentar verificar as diferenças entre esses informantes, a tabela 3.5 mostra como eles se agrupam, de acordo com as variáveis sociais da amostra.

Tabela 3.5 – Agrupamentos sociais dos informantes não categóricos quanto à pronúncia de (NDO) em *quando*

	N Informantes
Sexo/Gênero	
Feminino	11
Masculino	10
Escolaridade	
Até Ensino Médio	11
Ensino Superior	10
Faixa Etária	
20 a 34	10
35 a 59	11
60+	-
Mobilidade	
Mesmo Bairro	6
Mesma Zona	10
Zonas Diferentes	5
Região da Cidade	
Centro Expandido	8
Centro Velho	11
Periferia	2
21 Informantes	

A tabela 3.5 indica que esses 21 informantes em cuja fala ocorre *dø* em *quando* se distribuem equilibradamente de acordo com seu Sexo/Gênero e Escolaridade. Não há falantes mais velhos entre esses 21 informantes. Para Mobilidade e Região, os agrupamentos desses informantes não são balanceados. A tabela 3.6, a seguir, mostra a distribuição de dados de apagamento encontrados na Conjunção *quando* referente às variáveis Sexo/Gênero, Faixa Etária e Escolaridade, em que os dados mostraram distribuição equilibrada quanto ao número de informantes:

Tabela 3.6 - Distribuição de apagamento de /d/ em *quando* de acordo com Sexo/Gênero, Faixa Etária e Escolaridade dos 21 informantes não categóricos (N Total = 539)

	d	ø	%	Total
Sexo/Gênero				
Feminino	50	16		307
Masculino	43	19		232
$X^2= 1.3928e-31(1), p= 0.08937$				
Faixa Etária				
20 a 34	24	9		260
35 a 59	69	25		279
$X^2=20.6(1), p <0.001$				
Escolaridade				
Até Ensino Médio	64	21		300
Ensino Superior	29	12		239
$X^2= 9.7416(1), p <0.001$				

Verificando a organização dos dados, percebe-se que para ambos os sexos a percentagem de apagamento é bem próxima, diferente do que se encontrou para as duas Faixas Etárias, em que os indivíduos da 2ª faixa apagaram quase três vezes mais o /d/ em *quando* do que os da 1ª faixa. Para Escolaridade, os dados de apagamento predominaram na fala daqueles com menos grau de instrução.

Seguindo com a distribuição dos dados para as outras variáveis sociais foram encontradas células vazias para Renda do Informante e Índice Socioeconômico quando foram feitos os gráficos destas variáveis, desta forma, como a variável Classe Social relaciona-se a estas duas outras variáveis, preferiu-se a utilização desta para a análise. A seguir, na tabela 3.7, têm-se os dados referentes às variáveis voltadas para a origem dos informantes: Região da Cidade, Mobilidade Geográfica e Origem dos Pais, seguidos da variável Classe Social.

Tabela 3.7 - Distribuição de apagamento de /d/ em *quando* de acordo com Região da Cidade, Mobilidade Geográfica, Classe Social e Origem dos Pais dos 21 informantes não categóricos (N Total = 539)

	d ø	%	Total
Região da Cidade			
Centro Expandido	16	10,2	157
Centro Velho	11	12,6	87
Periférica	66	22,4	295
$X^2= 33.084(2), p <0.001$			
Mobilidade Geográfica			
Mesmo Bairro	21	14,2	148
Mesma Zona	39	13,9	281
Zonas Diferentes	33	30	110
$X^2= 17.983(2), p <0.001$			
Origem dos Pais			
Pais não Paulistanos	63	19,1	329
Mãe ou Pai Paulistano	3	11,1	27
Pais Paulistanos	16	36,4	44
Um(a) Avô(ó) Paulistano(a)	5	8,3	60
2 ou + Avós Paulistanos	6	7,6	79
$X^2= 54.372(4), p <0.001$			
Classe Social			
A	5	8,1	62
B1	3	3,5	85
B2	18	31,6	57
C1	30	20,8	144
C2	37	19,4	191
$X^2= 26.413(4), p <0.001$			

De acordo com a tabela 3.7, verificou-se que para as falas dos 21 informantes analisados, os dados de apagamento distribuídos conforme a Região da Cidade são verificados em maior número na fala daqueles cuja procedência são de bairros mais periféricos. Os indivíduos moradores dos bairros do Centro Velho e do Centro Expandido apresentam certo equilíbrio em relação à percentagem de apagamento. No que diz respeito à Mobilidade Geográfica, aqueles que moraram em diferentes Zonas da Cidade durante suas vidas parecem mostrar um número maior de produção, quanto à variante não padrão, que os indivíduos que permaneceram

morando no mesmo Bairro ou pela mesma Zona. Quanto à Origem dos Pais dos informantes pode se dizer que o apagamento em *quando* aparece em maior proporção quando ambos os pais são paulistanos. A última variável diz respeito à Classe Social dos entrevistados e, de certa forma, os dados mostram que as classes mais baixas indicam um maior número de não pronunciamento de /d/.

De forma geral, a análise descritiva dos casos de apagamento referentes a Conjunção *quando*, para as Variáveis Sociais, mostra que pode haver certas especificidades que indiquem favorecimento para a pronúncia não padrão. Estas especificidades, como maior produção de apagamento na fala de homens, daqueles da Faixa Etária intermediária, com menor grau de instrução e pertencentes às Classes Sociais mais baixas, que foram encontradas para os casos da Conjunção, serão verificadas para os casos de apagamento dos verbos no gerúndio a fim de verificar possíveis padrões de semelhança.

3.3 – Verbos no Gerúndio

Tal como vimos anteriormente, é em verbos no gerúndio que *dø* é mais frequente. Entretanto, gerúndios ocorrem em três estruturas diferentes: em “Justaposição” (de acordo com FERREIRA, 2010), Orações Adverbiais Reduzidas e, claro, em Perífrases. A tabela 3.8 mostra a distribuição de *dø* nesses contextos.

Tabela 3.8 - Distribuição de *dø* em verbos no gerúndio em três estruturas

	d ø	%	Total
Justaposição	22	40,7	54
Reduzidas Adverbiais	231	21,5	1.076
Perífrases	817	27,8	2.941
Total	1.070	26,3	4.071
$X^2 = 22.099(2), p < 0.001$			
Justaposição e Perífrases: $X^2 = 3.798(1), p = 0.05132$			
Perífrases e Reduzidas: $X^2 = 15.947(1), p < 0.001$			

O total de casos Justaposição é muito menor que nas outras duas estruturas, mas ocorre apagamento em proporção que parece alta. A seguir, faz-se uma descrição mais detalhada desses dados.

3.3.1 – Justaposição

Das 54 ocorrências de gerúndio em Justaposição, *dø* representa 40,7%. Essa estrutura ocorre nas entrevistas com 26 dos 60 informantes da amostra e, entre esses, há apagamento na fala de apenas 11¹¹. Esses dados são aqui analisados de maneira descritiva – assim como os de *quando*, já que as 54 ocorrências da estrutura verbal em Justaposição representam muito pouco em relação as outras estruturas. A tabela 3.9, abaixo, mostra a distribuição de apagamento para os 11 informantes não categóricos:

Tabela 3.9 Taxas de apagamento de /d/ em *Justaposição* de acordo com os 11 informantes não categóricos (N Total = 28)

Informante	d ø	Total
AlbertoM	3	3
AnaS	1	1
ArianeG	1	1
CarlosJ	2	2
CarolinaA	3	7
CelinaB	3	3
DouglasA	4	4
JaimeN	1	2
LeonS	1	2
LucasS	1	1
NelsonF	2	2
Total	22	28

¹¹ As ocorrências de Verbos no Gerúndio em Justaposição encontram-se nos Anexos desta pesquisa.

A distribuição de dados por falante revela um número muito baixo tanto de ocorrência da estrutura como de casos de apagamento¹². O que se verifica é que alguns dos falantes produziram mais verbos com estrutura de Justaposição (AlbertoM, CarlosJ, CelinaB, DouglasA e NelsonF) e também foram os que mais apagaram nesse contexto.

Estratificando esses falantes, num primeiro momento, de acordo com as variáveis sociais: Sexo/Gênero, Faixa Etária, Escolaridade, Classe Social e Região da Cidade, verificou-se que a distribuição de apagamento para os verbos em Justaposição se parece com a distribuição observada para a Conjunção *quando* com as mesmas variáveis, assim, encontrou-se uma percentagem maior de ocorrências na fala de entrevistados do Sexo Masculino, indivíduos da Faixa Etária Intermediária, com menor grau de instrução e aqueles provenientes da Classe Social mais baixa e de bairros mais periféricos.

Num segundo momento, estratificou-se os mesmos 11 falantes pelas variáveis linguísticas: Verbo Auxiliar, Tamanho do Auxiliar, Material Interveniente, Conjugação, Tamanho do Verbo Principal e Contexto Seguinte. Assim, os Verbos Auxiliares *acabar* e *ficar* indicaram 100% de apagamento para suas ocorrências juntamente às orações que não apresentaram estrutura composta com Verbos Auxiliares. Quanto ao Tamanho do Verbo Auxiliar ESTAR, foram encontradas 10 ocorrências com esse Auxiliar, em que as orações eram compostas pela forma fonológica plena ou pela forma reduzida (“tá”, “tô”, “tava”). Perante esses dados, verificou-se que as duas ocorrências para a forma plena apresentaram apagamento e, para os casos da forma reduzida, encontrou-se 2 apagamentos em 8 ocorrências totais. Outra característica das orações de Justaposição são os Materiais Intervenientes entre o verbo auxiliar e o verbo no gerúndio, em que foram averiguados um número maior da variante não padrão para as estruturas em que não existem materiais entre os dois verbos que compõem a Perífrase das orações de Justaposição. Para a Conjugação do Verbo Principal, não houve apagamento para a terceira Conjugação, e a predominância desses casos deu-se para a primeira Conjugação,

¹² As tabelas de distribuição de dados para as ocorrências de apagamento em Gerúndios de Justaposição encontram-se nos Anexos.

inclusive a que tem maior número de ocorrência da estrutura, seguida da segunda Conjugação. Quanto ao Tamanho por sílabas do Verbo Principal, os dados indicaram que quanto maior a extensão do verbo, maior o número de apagamentos encontrados. E, por último, a variável Contexto Seguinte aos Verbos no gerúndio mostrou que em estruturas compostas por um *Complemento* em seguida o número de apagamento é considerado menor que nas estruturas onde se verifica a *Pausa*.

Algumas outras características das orações em Justaposição foram exploradas de acordo com as estruturas verificadas entre as ocorrências, pois elas não são consideradas padrão, por exemplo, o que caracteriza uma oração em Justaposição seriam dois ou mais itens lexicais, preferência para os verbos no gerúndio, com terminação -ndo produzidos um após o outro, de forma seguida, a princípio hipotetizou-se que seriam repetições do primeiro item, mas não foi somente este fator encontrado. Foram encontradas 35 ocorrências de orações com 2 verbos em Justaposição, 14 ocorrências com 3 verbos, e 1 ocorrência com 4 verbos. Além disso verificou-se a existência de 3 outros tipos de estruturas, uma composta por um verbo e a Conjunção *quando* com 2 ocorrências, uma composta por um verbo e o substantivo *mundo* com 1 ocorrência e outra com um verbo e a Conjunção *segundo* com outra 1 ocorrência, totalizando as 54 ocorrências encontradas. Por essa classificação, foram encontrados os números de apagamento abaixo:

Tabela 3.10 - Distribuição de apagamento em Justaposição por estrutura (N= 54)

		dø	%	Total
Verbo + Quando		0	0	2
Verbo + Mundo		0	0	1
Verbo + Segundo		0	0	1
Verbo + Verbo	Iguais	2	25	8
	Diferentes	13	48	27
Verbo + Verbo + Verbo	Iguais	4	50	8
	Diferentes	1	17	6
Verbo + Verbo + Verbo+ Verbo	Diferentes	1	100	1

Assim, de acordo com a tabela 3.10, verifica-se que em construções com dois itens lexicais em que o segundo item não é um verbo no gerúndio não foram encontrados casos de apagamento; em relação às construções com dois verbos, encontrou-se uma porcentagem maior de apagamento para as ocorrências com verbos diferentes do que para as ocorrências com dois verbos iguais, diferentemente do que aconteceu com as construções com três verbos, onde o fator relacionado a três verbos iguais favoreceu o apagamento em virtude do fator com três verbos diferentes. Comparando as estruturas com dois verbos e a estrutura com três verbos, constatou-se que entre esses dois contextos o favorecedor ao apagamento proporcionalmente seria o com dois verbos. A última linha da tabela, mostra que a única ocorrência com quatro verbos, todos diferentes, teve as quatro realizações de /d/ apagadas.

Por consequência do número pequeno de dados encontrados para a estrutura de Justaposição, os dados não foram analisados de forma inferencial. Contudo, como nos casos da Conjunção *quando*, os padrões encontrados serão utilizados para comparação das análises inferenciais feitas com os dados de Orações Reduzidas Adverbiais e as Perífrases Verbais.

3.3.2 – Orações Reduzidas Adverbiais

Em relação às 1076 ocorrências de gerúndio em Reduzidas Adverbiais (ver tabela 3.8), 21,5 % representam $\emptyset$ (231 casos). Esses dados podem ser verificados na tabela 3.11 abaixo em sua distribuição pelas variáveis sociais: Sexo/Gênero, Faixa Etária, Escolaridade, Classe Social, Renda dos Informantes, Mobilidade Geográfica, Origem dos Pais e Região da Cidade.

Tabela 3.11 – Distribuição de dados de apagamento para as Orações Reduzidas Adverbiais por Variáveis Sociais (N= 1076)

		d	ø	%	Total
Sexo/Gênero	Feminino	107	19	554	
	Masculino	124	24	522	
Faixa Etária	20 a 34	108	27	395	
	35 a 59	107	31	345	
	60+	16	5	336	
Escolaridade	Ensino Médio	144	27	533	
	Ensino Superior	87	16	543	
Classe Social	A	39	20	195	
	B1	13	10	126	
	B2	35	12	303	
	C1	66	29	227	
	C2	78	35	225	
Renda dos Informantes	1 a 2 SMs	23	37	63	
	2 a 4 SMs	52	23	222	
	4 a 10 SMs	107	23	457	
	10 a 20 SMs	25	10	246	
	+20 SMs	24	27	88	
Mobilidade Geográfica	Mesmo Bairro	63	22	283	
	Mesma Zona	107	17	629	
	Zonas Diferentes	61	37	164	
Origem dos Pais	Pais não Paulistanos	154	26	596	
	Mãe ou Pai Paulistano	36	14	251	
	Pais Paulistanos	4	10	41	
	Um(a) Avô(ó) Paulistano(a)	14	15	95	
	2 ou + Avós Paulistanos	23	25	93	
Região da Cidade	Centro Expandido	53	15	349	
	Periférica	36	15	237	
	Centro Velho	142	29	490	

Para os dados da tabela acima foram criados gráficos para visualização das proporções em conjunto aos resultados dos testes de qui quadrado com intuito de verificar se há diferenças entre as proporções de apagamento de cada variável social.

Para a variável Sexo/Gênero, a figura 7 mostra que a taxa de $d\emptyset$ é pouco maior entre os homens; a diferença entre eles e as mulheres não é significativa, pelo teste qui-quadrado ($p > 0.05$), diferentemente do que se esperaria, em face dos resultados obtidos por trabalhos anteriores, revistos no capítulo 1, já que os resultados aqui verificados se assemelham aos estudos de Oushiro (2011), sobre a variação no emprego das interrogativas-Q em São Paulo, e o de Santos (2014), sobre o emprego das formas de subjuntivo e do indicativo nas orações adverbiais na fala de São Paulo e São Luís.

$$X^2=2.8857(1), p= 0.08937$$

Figura 7 – Proporção de $d\emptyset$ em Orações Reduzidas de acordo com Sexo/Gênero

Em relação à Faixa Etária dos informantes, a diferença entre cada um dos fatores mostrou-se estatisticamente relevante ($X^2= 82.354(2)$, $p<0.001$), na figura 8 verifica-se que o índice de apagamento é muito maior entre os mais jovens, com uma taxa de apagamento bem baixa entre os mais velhos.

$$X^2 = 82.354(2), p < 0.001$$

Figura 8 – Proporção de $d_\emptyset$ em Orações Reduzidas por Faixa Etária

Quanto ao fator Escolaridade, esperava-se que o maior número de apagamento fosse encontrado na fala dos indivíduos com menos Escolaridade – o que se observa na figura 9., de modo que a diferença observada é significativa estatisticamente.

$$X^2 = 18.639(1), p < 0.001$$

Figura 9 – Proporção de $d_\emptyset$ em Orações Reduzidas por Escolaridade

Para as 5 Classes Sociais classificadas conforme a amostra de fala, assim como em outros trabalhos sociolinguísticos (OUSHIRO, 2015), esperava-se que o maior número de casos de apagamento estaria concentrado nas classes mais baixas. Entretanto, para Orações Reduzidas, $d_\emptyset$ é mais frequente na classe mais alta (A); as classes mais baixas (C1 e C2) estão

mais próximas de A do que as classes que podem ser consideradas médias (B1 e B2) nessa amostra. O número de informantes, contudo, não é equilibrado entre as classes, de modo que isso possa estar enviesando tal distribuição.

$$X^2=58.256(4), p<0.001$$

Figura 10 – Proporção de $d_{\emptyset}$ em Orações Reduzidas por Classe Social

Já o gráfico em que se considera a Renda dos Informantes (figura 10) mostra que os falantes do grupo que recebe de 1 a 2 salários mínimos produzem um maior número de apagamento do que os falantes dos outros quatro grupos, para estes há uma determinada oscilação entre as proporções de apagamento observadas sendo elas mais altas para aqueles que recebem de 4 a 10 salários mínimos e aqueles que recebem mais de 20. Para esta variável, a diferença entre cada um de seus fatores se mostrou estatisticamente relevante ($X^2= 30.39(4)$, $p<0.001$). Separadamente do grupo que recebe de 1 a 2 salários mínimos, foi calculado o qui quadrado para os falantes dos outros quatro grupos, de forma que obteve-se $X^2= 19.908(2)$, $p<0.001$), reafirmando que a diferença entre os fatores é estatisticamente relevante.

$$X^2 = 30.39(4), p < 0.001$$

Figura 11 – Proporção de $d_\emptyset$ em Orações Reduzidas de acordo com Renda dos informantes

Quanto à Mobilidade Geográfica, uma variável proposta por Oushiro (2015), esperava-se que, quanto maior a mobilidade, maior seria o número de ocorrências da variante padrão, pois, segundo (MILROY, 1987 [1980]), falantes cuja mobilidade é menor tendem a se orientar por normas locais, ao passo que aqueles com maior mobilidade, menos enraizados, adotam valores externos e das classes mais altas. Para as Orações Reduzidas, os dados indicam uma maior incidência de apagamento para com Mobilidade alta e não para os com Mobilidade baixa, como esperado.

$$X^2 = 31.577(2), p < 0.001$$

Figura 12 – Proporção de $d_\emptyset$ em Orações Reduzidas por Mobilidade Geográfica

Quanto à Origem dos Pais, a investigação que se faz é se falantes cujos pais não são paulistanos apresentam tendências para o apagamento em maior taxa que os filhos de paulistanos. A distribuição dos dados sugere que, quando o paulistano tem dois ou mais avós paulistanos, parece maior taxa de apagamento, seguido por aqueles que tem pais não paulistanos.

$$X^2 = 20.789(4), p < 0.001$$

Figura 13 – Proporção de $d_{\emptyset}$ em Orações Reduzidas por Origem dos pais

Finalmente, outra variável social que se considera importante – tendo em vista a discussão de forma estigmatizadas e prestigiadas em uma comunidade de fala – é a das regiões da cidade de São Paulo. Em seu trabalho, Oushiro (2015:13) verificou que

sobre (-r) (...) esta é uma variável aparentemente estável em São Paulo quando se considera a comunidade como um todo, mas que os jovens de regiões periféricas e de classes mais baixas, por um lado, e aqueles de regiões mais centrais e de classes mais altas, por outro, exibem tendências em direções contrárias: enquanto os primeiros favorecem fortemente a variante retroflexa, os segundos a evitam radicalmente. Defende-se que a forte presença de migrantes do Norte/Nordeste na comunidade conduziu a uma reinterpretação do valor social do retroflexo, por parte dos falantes de classes mais baixas, como uma variante local e de relativo prestígio em relação às variantes fricativas, o que tem motivado o movimento de certos grupos de falantes em sua direção.

No caso do apagamento de /d/ em (NDO), esperava-se que haveria mais casos na periferia, retomando o que foi dito no capítulo 2, que indivíduos residentes em bairros mais centrais, têm, em geral, um custo de vida superior àqueles de bairros mais periféricos, remetendo a um nível indireto de status socioeconômico do falante. Desta forma, os indivíduos de bairros mais centrais poderiam ser incluídos, indiretamente, no grupo daqueles provenientes das Classe mais altas, que, geralmente, são os que mais fazem uso da variante padrão. De acordo com os dados encontrados, a figura 13 mostra que o fator mais propício ao apagamento é o dos bairros da Periferia, como já se esperava. E para tanto, a diferença entre as distribuições é significativa ($p < 0,001$).

$$X^2=30.108(2), p<0.001$$

Figura 14 – Proporção de $d_{\emptyset}$ em Orações Reduzidas por Região da Cidade

Seguindo com as variáveis linguísticas propostas para compor esta pesquisa, nem todas preenchem os critérios referentes às Orações Reduzidas de gerúndio, sendo somente analisadas as variáveis: Conjugação Verbal, Tamanho do Verbo Principal e Contexto Seguinte mostradas na tabela 3.12

Tabela 3.12 – Distribuição de dados de apagamento para as Orações Reduzidas Adverbiais por Variáveis Linguísticas (N= 1076)

		d ø	%	Total
Conjugação	1 ^a	164	23	703
	2 ^a	54	20	269
	3 ^a	13	13	104
Tamanho do Verbo Principal	2 sílabas	18	15	124
	3 sílabas	131	23	583
	4 ou + sílabas	82	22	369
Contexto Seguinte	Complemento	203	22	924
	Pausa	28	18	152

Dividido em 3 conjugações, esperava-se que verbos da primeira Conjugação, por ser considerada a Conjugação mais produtiva em relação à sua terminação, viessem a ter o /d/ mais apagado do que nos outros dois contextos. Conforme o valor de $p= 0.0382$, a variável foi considerada significativa estatisticamente. E os resultados podem ser confirmados no gráfico a seguir:

$$X^2= 6.5301(2), p= 0.0382$$

Figura 15 – Proporção de d ø em Orações Reduzidas por Conjugação

Verificou-se um número de apagamento bem maior nos dados encontrados para a primeira Conjugação que para a segunda e a terceira Conjugações. Como os dados de apagamento encontrados para a segunda e a terceira Conjugações são bem parecidos, analisou-

se a partir de um teste de qui quadrado com a finalidade de comparar a diferença significativa entre esses dois fatores e, obteve-se o resultado $X^2 = 2.3197(1)$, $p = 0.1277$. Com este resultado, chegou-se à conclusão que somente é significativa a diferença entre a primeira Conjugação e às demais, mas não entre a segunda e a terceira.

Assim como no trabalho de Ferreira (2010), aventou-se a hipótese de que quanto maior o verbo em relação ao seu número de sílabas, maior a chance de apagamento. De acordo com o gráfico a seguir, percebe-se que verbos com 3 ou mais sílabas possuem um maior número de apagamento, como verificado na tabela 3.12. Em relação aos resultados obtidos pelo teste de qui quadrado, o valor de $p = 0.128$ não indicou que a variável fosse considerada significativa perante a distribuição dos dados.

$$X^2 = 4.1109(2), p = 0.128$$

Figura 16 – Proporção de $d_\emptyset$ em Orações Reduzidas por Tamanho do Verbo

Assim como proposto nos estudos de Mollica e Mattos (1992), para a variável referente ao Contexto Seguinte aos verbos no gerúndio, esperava-se encontrar um número de ocorrências maior de apagamento onde houvesse Contexto Fonológico Seguinte, aqui denominado “Complemento” e a ausência de contexto como “Pausa”. Os dados encontrados para as Orações Reduzidas, mostram que os casos em que havia Complemento são de pouco mais de 20% de não realizações de /d/, já nos casos de Pausa se encontrou um número um pouco abaixo de 20%

de apagamento. Quanto ao teste univariado, a variável não se mostrou significativa com valor $p=0.3784$.

$$X^2 = 0.77583(1), p = 0.3784$$

Figura 17 – Proporção de $d_\emptyset$ em Orações Reduzidas por Contexto Seguinte

De modo geral, a partir das análises descritivas e inferenciais univariadas acerca das Orações Reduzidas, socialmente, a distribuição de dados possibilita verificar uma maior frequência de dados de apagamento para os informantes mais jovens, pertencentes a Classe Social mais baixa, aqueles com menor grau de instrução e que possuem Mobilidade Baixa. Linguisticamente, a ênfase dos dados de apagamento parece indicar que verbos de primeira Conjugação estão mais propícios a esta variante, junto daqueles com maior extensão de sílabas e sem complemento quanto ao Contexto Seguinte, mas nenhuma dessas variáveis linguísticas mostrou resultado significativo estatisticamente. A intenção a partir desses resultados é averiguar possíveis padrões de semelhança com o contexto mais produtivo de estruturas dos verbos no gerúndio: as Perífrases.

3.3.3 – Perífrases Verbais

Tal como vimos anteriormente, na tabela 3.7, o contexto mais produtivo de verbos no gerúndio – e, portanto, o conjunto de dados mais robusto sobre a variável (NDO), é das

construções perifrásticas: são 2941 ocorrências, com 27,8% de d $\emptyset$ (817 casos). Estes casos estão distribuídos conforme as variáveis sociais na tabela 3.13 a seguir:

Tabela 3.13 – Distribuição de dados de apagamento para as Perífrases Verbais por Variáveis Sociais (N= 2941)

		d $\emptyset$	%	Total
Sexo/Gênero	Feminino	296	22	1334
	Masculino	521	32	1607
Faixa Etária	20 a 34	332	34	973
	35 a 59	419	37	1149
	60+	66	8	819
Escolaridade	Ensino Médio	520	33	1568
	Ensino Superior	297	22	1373
Classe Social	A	114	24	471
	B1	36	10	347
	B2	139	18	756
	C1	180	31	576
	C2	348	44	791
Renda dos Informantes	1 a 2 SMs	137	42	326
	2 a 4 SMs	198	35	566
	4 a 10 SMs	327	28	1175
	10 a 20 SMs	86	13	656
	+20 SMs	69	32	218
Mobilidade Geográfica	Mesmo Bairro	280	33	841
	Mesma Zona	341	22	1558
	Zonas Diferentes	196	36	542
Origem dos Pais	Pais não Paulistanos	564	34	1680
	Mãe ou Pai Paulistano	125	19	673
	Pais Paulistanos	30	25	118
	Um(a) Avô(ó) Paulistano(a)	71	27	262
	2 ou + Avós Paulistanos	27	13	208
Região da Cidade	Centro Expandido	185	18	1014
	Periférica	146	22	655
	Centro Velho	486	38	1272

Assim como para as Orações Reduzidas, prevalece o caso de apagamento sobre o sexo masculino. O que difere nas análises para as Perífrases é que o número de apagamento entre os homens e as mulheres mostraram-se significativos neste grupo com valor de $p < 0.001$.

$$X^2 = 38.547(1), p < 0.001$$

Figura 18 – Proporção de $d_\emptyset$ em Perífrases por Sexo/Gênero

A primeira e a segunda Faixas Etárias indicam maior número de apagamento em relação à terceira Faixa Etária, a princípio, olhando para os resultados e comparando-os com os encontrados para as Orações Reduzidas, percebemos um padrão similar de distribuição. E, verificando o teste univariado, percebe-se um resultado significativo com o valor $p < 0.001$.

$$X^2 = 220.86(2), p < 0.001$$

Figura 19 – Proporção de $d_\emptyset$ em Perífrases por Faixa Etária

Nas Perífrases, os dados referentes à Escolaridade, mostrou-se uma incidência do apagamento na fala dos informantes com menor grau de instrução, e aqui, com uma diferença um pouco mais acentuada que nas Orações Reduzidas, assim como verificou-se como uma variável significativa.

$$X^2=47.122(1), p<0.001$$

Figura 20 – Proporção de $d_{\emptyset}$ em Perífrases por Escolaridade

No que diz respeito à Classe Social/ Índice Socioeconômico, a hipótese relativa a quanto menor a classe social, maior o uso da forma não padrão, parece que será confirmada. Algo que será melhor explicado após os testes estatísticos multivariados será o porquê dos informantes da classe A, sendo mais alta que a B1, tendem a apagar mais que os da classe subsequente.

$$X^2= 194.5(4), p<0.001$$

Figura 21 – Proporção de $d_{\emptyset}$ em Perífrases por Classe Social

No tocante à Renda dos Informantes, verifica-se o mesmo ocorrido para a Classe Social, as menores rendas que se correlacionam as menores classes possuem índices de apagamento maiores que as maiores rendas. Inclusive, a relação entre o grupo de informantes que possuem uma renda maior que 20 salários mínimos em relação aos que recebem entre 10 e 20 salários mínimos é maior assim como a classe social A em relação a B1.

$$X^2 = 119.61(4), p < 0.001$$

Figura 22 – Proporção de $d_{\emptyset}$ em Perífrases por Renda dos informantes

A variável Mobilidade geográfica para as Perífrases mostra dados parecidos com os encontrados para as Orações Reduzidas, onde indivíduos que moraram em zonas diferentes ou no mesmo bairro tendem a apagar mais o /d/ do que indivíduos que durante suas vidas moraram na mesma zona. Esta variável mostrou-se significativa com valor $p < 0.001$.

$$X^2 = 59.522(2), p < 0.001$$

Figura 23 – Proporção de $d_{\emptyset}$ em Perífrases de acordo com Mobilidade Geográfica

Desta forma, parece que a hipótese utilizada por Oushiro (2015) faz algum sentido em relação aos dados analisados por aqui: falantes com Mobilidade mais alta (aqueles que moraram em Zonas Diferentes) apresentam um número menor de produção da variável não padrão, o apagamento de /d/, que aqueles com Mobilidade média (que moraram em bairros diferentes da mesma Zona) e os de Mobilidade baixa (que sempre moraram no mesmo bairro).

Diferente das Orações Reduzidas, onde o fator 2 ou mais avós paulistanos mostrou um número maior de apagamento, para as Perífrases, a distribuição dos dados sugere que, quando o paulistano tem os pais não paulitanos, parece maior taxa de apagamento, seguido por aqueles que tem pais paulitanos e pelo menos um(a) avô(ó) paulistano(a).

$$X^2 = 79.612(4), p < 0.001$$

Figura 24 – Proporção de $d_\emptyset$ em Perífrases por Origem dos Pais

Seguindo com a variável Região da Cidade, e comparando-se os dados distribuídos das Perífrases com os das Orações Reduzidas, pode-se constatar que na Periferia a variante não padrão tem um número maior de ocorrências em relação ao Centro Velho e ao Centro Expandido. A diferença, na distribuição dos dados, se destaca em relação ao número encontrado nos Centros, pois nas Perífrases o número de apagamento é maior no Centro Expandido e nas Orações Reduzidas no Centro Velho.

$$X^2 = 124.73(2), p < 0.001$$

Figura 25 – Proporção de $d_\emptyset$ em Perífrases por Região da Cidade

As variáveis linguísticas utilizadas para analisar os dados de Perífrases são todas aquelas propostas por Ferreira (2010), na tabela 3.14 podemos verificar a distribuição de dados para as variáveis: Tamanho do Auxiliar Estar, Material Interveniente entre o Verbo Estar e o Verbo Principal, Material Interveniente entre outros Verbos Auxiliar e o Verbo Principal, Conjugação Verbal do Verbo Principal, Tamanho do Verbo Principal e Contexto Seguinte ao verbo no gerúndio:

Tabela 3.14 – Distribuição de dados de apagamento para as Perífrases Verbais por Variáveis Linguísticas

		d ø	%	Total
Material Interveniente entre outros V.AUX e V.PRIN.	Com	53	28	190
	Sem	299	30	982
Conjugação	1 ^a	474	29	1622
	2 ^a	287	28	1013
	3 ^a	56	18	306
Tamanho do Verbo Principal	2 sílabas	92	27	339
	3 sílabas	457	28	1633
	4 ou + sílabas	268	28	969
Contexto Seguinte	Complemento	667	28	2389
	Pausa	150	27	551
Tamanho do Auxiliar ESTAR	Pleno	45	30	152
	Reduzido	415	26	1617
Material Interveniente entre V.ESTAR e V.PRIN.	Com	44	28	157
	Sem	416	26	1612

A primeira variável linguística verificada refere-se ao Material Interveniente entre o Verbo Auxiliar e o Verbo Principal, ou verbo no gerúndio. Neste primeiro momento, verifica-se somente os casos de verbos auxiliares sem o verbo ESTAR, que constitui uma variável em separada mais à frente. Quanto ao Material Interveniente, a hipótese aventada por Ferreira (2010) e que se retoma aqui é a que Orações Perifrásticas que possuem Material Interveniente entre o Verbo Auxiliar e o Verbo Principal conjugado no gerúndio tendem a apagar mais o /d/. Além de se encontrar um número equilibrado de apagamento para os casos com e sem Material

Interveniente, o resultado do valor de $p=1$ mostrou que esta variável não é de fato significativa para os demais verbos auxiliares, sem levar em consideração o auxiliar ESTAR.

$X^2=7.5735e-30(1), p= 1$

Figura 26 – Proporção de $d_{\emptyset}$ em Perífrases, sem o verbo ESTAR, de acordo com Material Interveniente

Os Verbos Auxiliares das Orações Perifrásticas também constituem uma variável linguística que será utilizada como efeito aleatório de acordo com a frequência do uso dos mesmos em relação ao apagamento. Foram encontrados os seguintes verbos auxiliares nas Perífrases verificadas na amostra de fala paulistana:

Quadro 5 - Verbos Auxiliares das Orações Perifrásticas

Acabar	Participar
Andar	Partir-se
Chegar	Passar
Começar	Pegar
Continuar	Perceber
Crescer	Saber
Deixar	Sair
Dizer	Sentir-se
Entrar	Ser
Estar	Ter
Ficar	Terminar
Imaginar	Ver
Ir	Vir
Oferecer	Viver
	Voltar

Em seguida ao Material Interveniente, encontram-se a distribuição de dados para as conjugações. Diferente do ocorrido com as Orações Reduzidas, com as Perífrases, os dados de apagamento encontrados na segunda Conjugação equiparam-se aos de primeira Conjugação, assim, para estas a hipótese quanto a um maior caso de apagamento na primeira Conjugação parece não se confirmar. Porém, a variável foi selecionada como estatisticamente significativa, pois há diferença significativa entre a terceira Conjugação e as demais.

$$X^2=15.016(2), p<0.001$$

Figura 27 – Proporção de $d_\emptyset$ em Perífrases por Conjugação Verbal

A variável linguística que se refere ao Tamanho do Verbo Principal da Perífrase não demonstrou diferença significativa entre seus fatores assim como os resultados obtidos com as Orações Reduzidas. Para as Perífrases, os números de dados de apagamento encontrados para verbos com 3 sílabas e 4 ou mais sílabas são os mesmos e, a diferença entre eles e os verbos com 2 sílabas são mínimas.

$$X^2=0.056625(2), p=0.9721$$

Figura 28 – Proporção de $d_{\emptyset}$ em Perífrases por Tamanho do Verbo Principal

Como proposto anteriormente, esperava-se encontrar para a variável referente ao Contexto Seguinte aos verbos no gerúndio, um número de ocorrências maior de apagamento onde houvesse Complemento, como verificou Mollica (1989). Assim como para as Orações Reduzidas, o apagamento se mostrou mais incidente nesses contextos, mas quanto ao teste univariado, a variável não se mostrou significativa com valor $p=0.3787$.

$$X^2=0.77504(1), p=0.3787$$

Figura 29 – Proporção de $d_{\emptyset}$ em Perífrases por Contexto Seguinte

Foi sugerido nas variáveis linguísticas, a análise do Tamanho do Verbo Auxiliar. Portanto, separou-se somente as ocorrências de Perífrases com o verbo ESTAR, em que foram analisadas em um modelo estatístico em separado das outras Perífrases Verbais compostas por outros auxiliares (ir, andar, vir, ficar...) para verificar se nesses contextos o apagamento se comporta de maneira diferente. Para esta análise, sua pronúncia foi classificada em Pleno ou Reduzido. Pleno quando o verbo é pronunciado em sua forma fonética considerada padrão e, Reduzido quando é pronunciado em sua forma fonética reduzida como: “tá”, “tô”, “tava”. Analisando as entrevistas, foram encontradas 1770 ocorrências do verbo ESTAR, em que foram verificados 1618 casos de redução e 152 de forma plena. O esperado era que na forma reduzida houvesse um maior número de apagamento do verbo no gerúndio, hipótese esta não confirmada, consequentemente os resultados do qui quadrado não selecionaram esta variável como significativa.

$$X^2 = 0.90126(1), p = 0.3424$$

Figura 30 – Proporção de $d_\emptyset$ em Perífrases por Tamanho do Verbo Auxiliar ESTAR

Assim como verificado com os outros verbos auxiliares, a variável Material Interveniante para as Perífrases com o verbo ESTAR não demonstrou certa significância com o valor de $p > 0.05$. Uma característica interessante é que com o verbo ESTAR, a proporção de apagamento inverteu, mesmo que de forma mínima, mostrando que a hipótese de Ferreira (2010), sobre o contexto favorecedor para o apagamento com Material Interveniante seria ratificada.

$$X^2=0.24701(1), p= 0.6192$$

Figura 31 – Proporção de $d_\emptyset$ em Perífrases com auxiliar ESTAR de acordo com Material Interveniente

Concernente ao Material Interveniente entre o Verbo Auxiliar ESTAR e o verbo no gerúndio, poder-se-ia identificar indícios de gramaticalização da forma plena do verbo ESTAR para sua forma reduzida, como estudou Mendes (1999). Partindo da noção de gramaticalização estudada por Lehmann (1982, apud MENDES, 1999:101), “um dos processos que se observa, na passagem de uma gramaticalização fraca para uma gramaticalização forte, é o da coalescência”, assim segundo Mendes (1999:101) “a ocorrência de elementos entre o auxiliar e a forma nominal (gerúndio) seria um empecilho à morfologização do primeiro”. Desta forma, o esperado pelo autor era um número menor de casos com Material Interveniente nas construções Perifrásticas compostas por ESTAR reduzido + gerúndio. Para verificar essa hipótese, a Variável Tamanho do Auxiliar ESTAR foi cruzada com a variável Material Interveniente, onde foram encontrados os seguintes dados de apagamento para a fala paulistana:

Figura 32 – Proporção de $d_{\emptyset}$ no cruzamento das variáveis Tamanho do auxiliar ESTAR x Material Interveniente

A partir desses dados verifica-se que

o uso da forma *tá* se relaciona com a ausência de elementos interferentes entre o auxiliar e o gerúndio. A inserção de tais elementos é, portanto, uma medida de coalescência entre *tá* e *-ndo*: se não houvesse nenhum caso de interveniência entre *tá* e a forma nominal em questão, o grau de dependência sintagmática entre ambos seria máximo. (MENDES, 1999:102)

Os resultados a que se chegou com o cruzamento das variáveis Tamanho do Auxiliar Estar e Material Interveniente, comparam-se ao que se confirmou com os dados de Mendes (1999), em que parece ser uma tendência nos dados para o apagamento na fala paulistana, em razão da ocorrência de mais casos de apagamento entre a **forma reduzida + gerúndio** do que a **forma plena + gerúndio**.

3.4 – Síntese

Em seu conjunto, as análises descritivas apresentadas partiram da organização dos conjuntos de dados para as análises multivariadas. Em razão do pequeno número de

ocorrências, os casos de *quando* e de gerúndios em *Justaposição* são analisados no presente capítulo de maneira mais descritiva. Para *quando*, a distribuição, mostra que a maior produção de apagamento pode ser verificada na fala de homens, daqueles da Faixa Etária intermediária, com menor grau de instrução e pertencentes às Classes Sociais mais baixas. Para os verbos em contexto de Justaposição, por conta do número bem reduzido de apagamento, é difícil verificar minuciosamente se algum fator das variáveis sociais ou linguísticas seria mais favorável ao apagamento, porém quando analisou-se os dados para as estruturas de Justaposição encontradas, verificou-se características mais interessantes. Por exemplo, em construções com dois itens lexicais em que o segundo item não é um verbo no gerúndio não foram encontrados casos de apagamento; em relação às construções com dois verbos, encontrou-se uma porcentagem maior de apagamento para as ocorrências com verbos diferentes do que para as ocorrências com dois verbos iguais, diferentemente do que aconteceu com as construções com três verbos, onde o fator relacionado a três verbos iguais favoreceu o apagamento em virtude do fator com três verbos diferentes. E a única ocorrência com quatro verbos, todos diferentes, teve as quatro realizações de /d/ apagadas.

Para os casos de apagamento encontrados nas Orações Reduzidas Adverbiais e nas Perífrases foram verificados de forma descritiva e inferencial através de testes univariados, em que se verificou que para as Orações Reduzidas, socialmente, a distribuição de dados possibilita verificar uma maior frequência de dados de apagamento para os informantes mais jovens, pertencentes a Classe Social mais baixa, aqueles com menor grau de instrução, que possuem dois ou mais avós paulistanos e que possuem Mobilidade Alta. Linguisticamente, a ênfase dos dados de apagamento parece indicar que verbos de primeira Conjugação, com $X^2 = 6.5301(2)$, $p = 0.0382$, estão mais propícios a esta variante, mas nenhuma das outras variáveis linguísticas mostrou resultado significativo estatisticamente (Tamanho do Verbo Principal ($X^2 = 4.1109(2)$, $p = 0.128$) e Contexto Seguinte ($X^2 = 0.77583(1)$, $p = 0.3784$)). Para as Perífrases, aconteceu o mesmo, somente a variável Conjugação, com $X^2 = 15.016(2)$, $p < 0.001$, mostrou-se estatisticamente significativa (Tamanho do Verbo principal ($X^2 = 0.056625(2)$, $p = 0.9721$), Material Interveniente ($X^2 = 0.37523(1)$, $p = 0.5402$) e Contexto Seguinte ($X^2 = 0.77504(1)$, $p = 0.3787$)), já as variáveis sociais parecem indicar que os casos de apagamento tem mais força na fala de homens, aqueles pertencentes à 1ª e a 2ª Faixas Etárias, com menor grau de instrução,

pertencentes às duas Classes Sociais mais baixas (C1 e C2), provenientes da região do Centro, filhos de pais não paulistanos e com Mobilidade mais alta. O capítulo 4, a seguir, apresenta análises multivariadas para Orações Reduzidas Adverbiais e as Perífrases, através de modelos de regressão logística.

Capítulo 4

Análises Multivariadas

Gerúndios em Orações Reduzidas e Perífrases Verbais

O capítulo anterior apresentou uma análise descritiva dos dados de apagamento encontrados em *quando* e em estruturas verbais compostas por Justaposição. Esses dados não serão analisados inferencialmente através de testes estatísticos por conta do número reduzido de dados encontrados. Já para as Orações Reduzidas e Perífrases realizaram-se análises descritivas e inferenciais univariadas. A partir destas, verificou-se que existe uma possibilidade de direção do apagamento para as falas de indivíduos do Sexo masculino, aqueles pertencentes às 1ª e 2ª Faixas Etárias, e às duas Classes Sociais mais baixas (C1 e C2). No presente capítulo, apresentam-se os resultados de análises quantitativas multivariadas dos conjuntos de ocorrências em Orações Reduzidas e Perífrases, desenvolvidas com o programa R (R Core Team, 2018).

4.1- Tipo de Gerúndio: Orações Reduzidas x Perífrases

As análises multivariadas aqui reportadas objetivam averiguar quais variáveis linguísticas e sociais têm correlação com (NDO) nesses dois conjuntos de dados que são mais numerosos – tanto da perspectiva dos totais, quanto da perspectiva da variante dØ. Nesse tipo de análise, testam-se os efeitos de mais de uma variável ao mesmo tempo e suas interações. Essas interações mostram que as variáveis chamadas independentes podem não ser, de fato, independentes entre si; quando isso ocorre, significa que elas não influenciam isoladamente a variação, mas sim quando cruzadas entre si. Isso equivale a dizer, por exemplo, que a variável

Sexo/Gênero, que não mostrou correlação com (NDO) nas Orações Reduzidas, pode interagir com Faixa Etária. Em outras palavras, pode ser o caso de que, a depender da Faixa Etária, os dados se distribuam de maneira significativamente diferente nas falas de homens, de um lado, e mulheres, de outro. Além disso, as análises reportadas a seguir foram feitas através de modelos de efeitos mistos (BAAYEN, 2008), que incluem dois tipos de variáveis independentes: as fixas e as aleatórias. As fixas dizem respeito às características linguísticas e sociais que são representativas dos indivíduos que compõem a amostra utilizada (como o Sexo/Gênero, Faixa Etária, Tamanho do Verbo principal, Contexto Seguinte, etc.) e que podem ser replicados em estudos futuros. Já as variáveis aleatórias são aquelas que dizem respeito à amostra coletada, ou seja, os Participantes e os Itens Lexicais. A inclusão desses dois tipos de variáveis nos modelos multivariados permite determinar se as correlações encontradas de fato se referem às variáveis sociais ou linguísticas ou à contribuição de algum(s) participante(s) ou ocorrência(s) em específico, que se comportaria(m) de maneira muito diferente dos demais, enviesando os resultados (cf. OUSHIRO, 2017).

Antes de investigar as correlações com (NDO) nos dois subconjuntos de dados - Orações Reduzidas Adverbiais e Perífrases – separadamente, fez-se em um modelo com todos esses dados, a fim de verificar se a variável Tipo de Gerúndio (Reduzidas vs. Perífrases) é significativa na amostra.

Tabela 4.1: Estimativas (em logodds) de valores de significância para Perífrases e Orações Reduzidas (N = 4016)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.64279	0.08307	-7.738	<0.001	***
sexogenM	0.52864	0.07791	6.785	<0.001	***
fai.eta2a	0.06815	0.08143	0.837	0.40262	
fai.eta3a	-1.94873	0.13077	-14.902	<0.001	***
Escolaridades	-0.63513	0.07767	-8.178	<0.001	***
tipo.gerR	-0.26157	0.08979	-2.913	<0.005	**

glm(ndo~sexogen+fai.eta+Escolaridade+tipo.ger, data=gpr, family = binomial)

Nesse modelo, não foram incluídas as variáveis aleatórias. Na primeira coluna da tabela, aparecem as variantes das variáveis fixas (isto é, os níveis de Sexo/Gênero, Faixa Etária, Escolaridade e Tipo de Gerúndio), sendo que *Intercept* é o termo que se refere ao valor de referência estabelecido nessa análise, que é o valor de apagamento estimado para um falante do Sexo feminino, da primeira Faixa Etária, com Escolaridade até o Ensino Médio, para as Perífrases. O que mais interessa nesse modelo é justamente a variável Tipo de Gerúndio: é significativa a diferença entre apagar o /d/ nos gerúndios em perífrases e em Orações Reduzidas; nessas últimas, relativamente às primeiras, desfavorece-se o apagamento dado o valor da estimativa negativo para Tipo de Gerúndio: Orações Reduzidas ($-0.64279 + (-0.26157) = -0.38122$).

No conjunto total de dados, de acordo com esse modelo, os resultados verificados para Sexo/Gênero se mantêm em relação ao já verificado no Capítulo 3 nas análises univariadas, quando o Sexo Feminino se mostrou menos favorável ao uso do apagamento que os homens. O mesmo foi verificado para Faixa Etária e Escolaridade, em que informantes da terceira faixa desfavorecem o uso do apagamento como aqueles com maior grau de instrução.

4.2 - Em busca de interações

Para as Orações Reduzidas, verificaram-se dois casos de interações significativas possíveis, nos testes sem efeitos aleatório, entre Sexo/Gênero e Classe Social, bem como entre Faixa Etária e Escolaridade. Para a primeira interação, a tabela 4.2 mostra o valor de significância das variáveis e de possíveis interações entre elas. As interações são marcadas por “____:____”, como em “sexogenM:classesocialB1”.

Tabela 4.2: Estimativas (em logodds) de valores de significância para Sexo/Gênero e Classe Social das Orações Reduzidas (N = 1076)

	Estimativa	Erro padrão	Valor-z	Significância (p)
(Intercept)	-3.7329	1.1981	-3.116	0.00183 **
sexogenM	2.4829	1.3736	1.808	0.07066 .
classesocialB1	1.4415	1.7770	0.811	0.41727
classesocialB2	0.4883	1.3813	0.354	0.72372
classesocialC1	1.8711	1.3323	1.404	0.16019
classesocialC2	2.5409	1.4104	1.802	0.07161 .
sexogenM:classesocialB1	-3.4764	2.1440	-1.621	0.10493
sexogenM:classesocialB2	-1.9515	1.7188	-1.135	0.25622
sexogenM:classesocialC1	-1.8304	1.8204	-1.006	0.31466
sexogenM:classesocialC2	-1.7339	1.7191	-1.009	0.31317

glmer(ndo ~ sexogen * classesocial +(1 | falante) + (1 | ocor), data = gr, family = binomial)

Os resultados obtidos pelo teste estatístico revelaram que não existe correlação entre as variáveis Sexo/Gênero e Classe Social como esperado através dos resultados dos testes sem os efeitos, mas sim a interação entre Faixa Etária e Escolaridade. Todos os resultados estatisticamente significativos foram destacados em negrito na tabela 4.3:

Tabela 4.3: Estimativas (em logodds) de valores de significância com interação entre Faixa Etária e Escolaridade das Orações Reduzidas (N = 1076)

	Estimativa	Erro padrão	Valor-z	Significância (p)
(Intercept)	-0.1847	0.4781	-0.386	0.6993
fai.eta2a	-0.5728	0.6878	-0.833	0.4050
fai.eta3a	-5.0082	1.0492	-4.773	< 0,001 ***
EscolaridadeS	-1.7377	0.6976	-2.491	0.0127 *
fai.eta2a:EscolaridadeS	0.7374	0.9794	0.753	0.4515
fai.eta3a:EscolaridadeS	3.2517	1.3022	2.497	0.0125 *

glmer(ndo~fai.eta*Escolaridade+(1|falante)+(1|ocor), data = gr, family = binomial)

4.3 – Gerúndio em Orações Reduzidas

Para o conjunto de dados das Orações Reduzidas, os testes univariados mostraram que não há correlação entre as variáveis linguísticas e a variável dependente (NDO). A partir disso, foram criados diferentes modelos multivariados sem a inclusão delas, seguindo o Princípio da Navalha de Occam, “também conhecido como Princípio da Simplicidade ou Princípio da Parcimônia, de acordo com o qual são preferíveis as teorias mais simples às mais complexas” (cf. GUY e ZILLES, 2007). Como a adição de múltiplas variáveis com células vazias resultava na não convergência dos modelos com efeitos mistos, a partir dos testes multivariados GLM, foram sendo retiradas as variáveis que não mostraram significância num primeiro momento. As variáveis sociais, por sua vez, apresentam-se como as mais relevantes para o emprego variável de (NDO). Desta forma, as três variáveis que não fizeram com que o modelo convergisse foram Sexo/Gênero, Faixa Etária e Escolaridade, como na tabela 4.4.

Tabela 4.4 Estimativas (em logodds) de valores de significância para Orações Reduzidas (N = 1076)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.9461	0.4756	-1.989	0.0467	*
sexogenM	0.7248	0.4620	1.569	0.1167	
fai.eta2a	-0.2817	0.5121	-0.550	0.5822	
fai.eta3a	-3.3421	0.6758	-4.945	< 0,001	***
EscolaridadeS	-0.8467	0.4603	-1.840	0.0658	.

glmer(ndo~ sexogen+fai.eta+Escolaridade+(1|falante)+(1|ocor), data = gr, family = binomial)

Os resultados obtidos com a inclusão da variável Sexo/Gênero foi com permite analisar a existência ou não de prestígio na fala dos paulistanos, já que uma das hipóteses em relação a esta pesquisa envolvia esta comparação dos resultados para a variável em outras variedades do PB. Num primeiro momento, em uma análise de qui-quadrado, a variável não se revelou significativa, $p=0.08937$.

A tabela mostra que entre a primeira Faixa Etária e a segunda não há diferença significativa. Mostra-se relevante, a diferença verificada entre as duas primeiras Faixas Etárias em relação à terceira Faixa. Assim, encontrou-se uma predominância de apagamento para as duas primeiras Faixas, comparem-se estes com os resultados obtidos por Mollica (1989), para a variedade carioca e por Ferreira (2010), para a variedade rio-pretense na tabela 4.5:

Tabela 4.5 Apagamento de Gerúndio com relação à Faixa Etária para duas variedades do PB

Faixa Etária	SJRP – SP		Rio de Janeiro - RJ		São Paulo – SP	
	Apl.Total	%	Apl.Total	%	Apl.Total	%
1^a	168/201	83	219/1117	20	108/395	27
2^a	423/579	73	775/2035	38	107/345	31
3^a	51/143	36	263/949	28	16/336	5

Fonte: adaptado de Ferreira (2010)

Já se mencionou anteriormente que a metodologia adotada em cada pesquisa não foi a mesma, pois a divisão de Faixas Etárias foi diferente da utilizada para esta pesquisa que segue os critérios de estratificação da amostra de fala paulistana, em que as faixas se subdividem em: primeira (20 a 34 anos), segunda (35 a 59 anos) e terceira (acima de 60 anos). Entretanto, comparou-se as tendências percentuais entre as variedades identificadas na tabela acima para descrever a direção do apagamento entre as Faixas Etárias para essas regiões. A variedade paulistana se aproxima dos resultados obtidos para variedade riopretense, no que toca ao maior número de casos de apagamento para a primeira e segunda Faixas Etárias e um número bem menor para os informantes da terceira faixa que tendem a preservar a variável plena. Diferentemente das variedades paulistas, os dados de Mollica (1989) para o Rio de Janeiro, mostram que os falantes da primeira e da última Faixa Etária são os que mais tendem a preservar o /d/, enquanto para São Paulo e Rio Preto, somente os falantes mais velhos são os que tendem a preservar a forma plena.

Constata-se, então, que quanto as Orações Reduzidas, os dados de apagamento confirmam a hipótese, proposta por Labov (1972), que prevê que falantes mais velhos preservam a variante padrão – nesse caso com a pronúncia de /d/ em (NDO).

Para a variável Escolaridade, encontrou-se o que Guy e Zilles (2007) chamam de significância marginal. A significância marginal ocorre quando o valor de p resulta em um valor próximo de 0.05. Onde para EscolaridadeS, obteve-se $p=0.0658$. Isso pode significar que

o grupo no todo, pode representar um efeito significativo sobre o processo, mas inclui muitos fatores entre os quais há vários que não têm efeito significativamente distinto dos outros. O resultado é que o grupo, inicialmente, não chega ao nível de $p < 0.05$, mas se redefinido teria sim esse nível de significância. (GUY e ZILLES, 2007:184)

A partir desse contexto, olhou-se para a variável e verificou-se que os fatores são, de fato, bem distribuídos, não aleatoriamente, portanto decidiu-se mantê-las nas análises descrevendo suas características.

O valor negativo para a estimativa da Escolaridade relacionada ao Ensino Superior revela que o apagamento vem sendo utilizado em maior número na fala de pessoas com até o Ensino Médio. Verificando os resultados de São Paulo Capital com outras variedades, para as quais há estudos sobre essa variável, obteve-se:

Tabela 4.6 Apagamento de Gerúndio com relação à Escolaridade para outras variedades do PB

Escolaridade	São Paulo -SP		SJRP-SP		Custódia – PE		João Pessoa – PB	
	Apl.Total	%	Apl.Total	%	Apl.Total	%	Apl.Total	%
1º EF			111/129	86	148/220	67	539/799	67
2º EF	144/533	27	177/232	76			596/877	68
EM			241/332	73	178/291	61	313/544	58
ES	87/543	16	113/230	49			288/806	36

Os números apresentados na tabela 4.7 mostram que, nas quatro regiões, de modo geral, o comportamento da variável Escolaridade é o mesmo, isto é, diminui a aplicação do apagamento à medida que aumenta o nível de Escolaridade. Por meio dessas características entre os fatores da Escolaridade, devido ao grau do apagamento apresentar-se inversamente proporcional ao aumento dos segmentos de escolarização, mostrando o pouco prestígio da variante. Desta forma, seguindo a afirmação de Ferreira (2010:108)

concluimos que essa possa ser a tendência para o fenômeno em questão e uma possível explicação para esses resultados seria o fato de os informantes mais escolarizados terem maior conhecimento da gramática normativa e dos seus valores sociais, uma vez que as formas privilegiadas são aquelas provindas da gramática e dos grupos sociais com maior poder econômico, que consequentemente, na maioria das vezes, são os indivíduos que possuem maior grau de Escolaridade.

Além da variável Escolaridade, um dos objetivos desta pesquisa foi utilizar o Índice Socioeconômico e a variável Classe Social como variáveis que poderiam influenciar significativamente o apagamento de /d/ em (NDO), já que a variável Classe Social não se refere apenas à renda familiar ou à renda do informante, no geral, ela costuma ser operacionalizada como um índice composto que normalmente abarca nível de Escolaridade, Ocupação e Renda. Quando se incluiu a variável Classe Social, o modelo não convergiu. Averiguou-se que a quantidade de indivíduos por célula não estava bem equilibrada em relação às duas primeiras classes mais altas e as demais. Seguindo-se um critério mencionado por Guy e Zilles (2007:153) “não se deve ter a grande maioria dos dados (por exemplo, acima de 90%) concentrada num só fator, nem se deve ter um fator ou mesmo vários fatores com muito pouco dado”, assim para solucionar o problema da não convergência, decidiu-se por recategorizar o número de indivíduos por célula a fim de equilibrar os dados. Na tabela 4.7, tem-se o antes e o depois desta recategorização.

Tabela 4.7 Recategorização dos fatores da variável Classe Social

<i>Antes da Recategorização</i>	
Classe Social	Informantes
A	10
B1	7
B2	16
C1	14
C2	13
<i>Depois da Recategorização</i>	
Classe Social	Informantes
A + B1	17
B2	16
C1	14
C2	13

A partir da recategorização da variável, realizou-se outro modelo, com as variáveis sociais: Sexo/Gênero, Escolaridade e Classe Social. Num modelo anterior a este, inclui-se também a variável Faixa Etária, que fez com que o modelo não convergisse – possivelmente porque o número de ocorrências de apagamento na terceira Faixa Etária ser muito pequeno em relação às outras duas faixas etárias. Para tanto, realizou-se um modelo sem a Faixa Etária, que resultou nos dados visualizados na tabela 4.8.

Tabela 4.8 Estimativas (em logodds) de valores de significância para Orações Reduzidas (N = 1076)

	Estimativa	Erro padrão	Valor-z	Significância (p)
(Intercept)	-2.8665	0.8470	-3.384	0.000713 ***
sexogenM	0.7989	0.5384	1.484	0.137848
Escolaridades	0.1278	0.6476	0.197	0.843510
classesocialB2	-0.4998	0.7063	-0.708	0.479176
classesocialC1	0.9614	0.8354	1.151	0.249757
classesocialC2	1.6906	0.8516	1.985	0.047131 *

glmer(ndo~sexogen+Escolaridade+classesocial+(1|falante), data = gr, family = binomial

A tabela 4.8 mostra que a Classe Social C2 mostra um valor de $p < 0.05$ significativo em relação ao *Intercept*. Desta forma, verificou-se que a Classe Social mais baixa é favorecedora do apagamento de /d/ em Verbos no Gerúndio das Orações Reduzidas.

4.4 – Gerúndio em Perífrases Verbais

A análise univariada de (NDO) nas perífrases revelou que, das sete variáveis independentes de natureza linguística, somente três parecem ser preditoras: Tempo Verbal do Auxiliar, Tamanho do Auxiliar e Conjugação do Verbo no Gerúndio. Partindo para uma análise multivariada, foram criados modelos com a inclusão de variáveis sociais e Tempo verbal e Conjugação para verificar o comportamento do apagamento. Para as análises multivariadas das Perífrases Verbais foi incluído mais uma variável como efeito misto: Verbo Auxiliar da Perífrase. O primeiro modelo contendo todas as variáveis aqui citadas não convergiu, assim, foram sendo criados modelos para verificar se alguma variável poderia ser descartada caso não demonstrasse significância estatística. Desta forma, o modelo mais simples que se conseguiu obter para verificar a distribuição estatística foi esse que se mostra na tabela 4.9.

Tabela 4.9 Estimativas (em logodds) de valores de significância para Perífrases com Sexo/Gênero e Conjugação (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-1.9951	0.3733	-5.344	< 0,001	***
sexogenM	0.7090	0.5151	1.376	0.16872	
conjugacao2a	-0.1663	0.1562	-1.065	0.28702	
conjugacao3a	-0.9104	0.2401	-3.792	< 0,001	***

glmer(ndo ~ sexogen + conjugacao + (1 | falante) + (1 | ocor) + (1 | verboaux), data = gp, family = binomial)

É significativa a diferença entre apagar /d/ em gerúndios de verbos da terceira Conjugação e da primeira. O valor da estimativa ($-1.9951 + (-0.9104) = -1.0847$) indica que em relação a primeira Conjugação, o apagamento para a terceira é desfavorecido. Levando-se em consideração a produtividade exercida sobre a primeira Conjugação verbal e a sua frequência de uso que pode indicar que verbos mais frequentes tendem a apagar mais.

Para o conjunto de variáveis sociais, todas as variáveis revelaram valor $p < 0.001$ nos testes de qui-quadrado. Para realização dos modelos multivariados foram acrescentadas uma a uma para verificar até que ponto o modelo convergiria. As variáveis sociais que não estratificaram a amostra: Mobilidade, Origem dos Pais, Região e Classe Social provocaram a não convergência do modelo. Como primeira análise, portanto, para as variáveis sociais das Perífrases, utilizou-se o modelo proposto na tabela 4.10.

Tabela 4.10: Estimativas (em logodds) de valores de significância para Perífrases com variáveis sociais (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.9107	0.4695	-1.940	0.0524	.
sexogenM	0.7579	0.4368	1.735	0.0827	.
fai.eta2a	-0.5179	0.5164	-1.003	0.3158	
fai.eta3a	-2.6071	0.5533	-4.711	< 0,001	***
EscolaridadeS	-0.5453	0.4354	-1.253	0.2103	

glmer(ndo~sexogen+fai.eta+Escolaridade+(1|falante)+(1|ocor), data=gp, family = binomial)

Somente a variável Faixa Etária mostrou valor p significativo. E a variável Sexo/Gênero resultou em um valor marginal de significância, $p = 0.0827$, diferente do encontrado nas análises univariadas, quando o valor $p = < 0.001$. Sexo/Gênero não se mostra uma variável preditora do apagamento.

A partir do resultado de Faixa Etária, verificou-se que os valores proporcionais entre faixas perante o apagamento são similares aos valores observados para as Reduzidas: a terceira Faixa Etária desfavorece o apagamento em relação à primeira. Desta forma, os dois conjuntos

de dados, os de Perífrases e os de Orações Reduzidas, trazem a mesma característica acerca do apagamento em relação às faixas etárias, que pode ser reafirmada com uma comparação dos dados da tabela 4.5, em que a terceira Faixa Etária desfavorece o apagamento de /d/ nos verbos de gerúndio em comparação às outras duas faixas etárias. Quanto às duas faixas mais jovens, percebe-se uma propensão em utilizar mais o apagamento na faixa intermediária, o que caracterizaria indícios de uma variável estável.

Considerando-se os mesmos critérios utilizados para as Orações Reduzidas quanto a variável Classe Social, numa primeira análise a variável interferiu na convergência do modelo. Em vez de descartá-la, optou-se por recategorizá-la da mesma forma como mostrada na tabela 4.8, já que com essa recategorização de seus fatores, verificou-se resultados diferentes dos originados na primeira análise em que o modelo não convergiu para as Orações Reduzidas, acreditou-se que poderia acontecer o mesmo para as Perífrases. A partir dessas observações, realizou-se um modelo com as variáveis: Sexo/Gênero, Escolaridade e Classe Social, mostrado na tabela 4.11.

Tabela 4.11 Estimativas (em logodds) de valores de significância para Perífrases com Classe Social recategorizada (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-3.0625	0.7574	-4.044	<0.001	***
sexogenM	0.7988	0.4983	1.603	0.10892	
Escolaridades	0.3412	0.5837	0.584	0.55891	
classesocialB2	0.4185	0.6407	0.653	0.51358	
classesocialC1	1.0255	0.7698	1.332	0.18279	
classesocialC2	2.1446	0.7816	2.744	0.00607	**

glmer(ndo~sexogen+Escolaridade+classesocial+(1|falante), data = gp, family = binomial

O que pôde se averiguar com os resultados obtidos com esse modelo foi que, assim como nas Orações Reduzidas, o apagamento da Classe Social C2 obteve um valor p significativo com $p = 0.00607$ em relação aos dados de apagamento da Classe Social A. Comparando estes resultados com os obtidos em relação às Orações Reduzidas, há uma diferença entre os níveis da variável, já que para as Reduzidas encontrou-se um número baixo

de apagamento para a Classe B2. Antes da recategorização dos informantes das células, o número de apagamento era menor para os informantes da classe B1 (considerada Média alta), o que de qualquer forma não se chegou a resultados parecidos para os dois tipos de estruturas sintáticas de gerúndio. Assim, chega-se à conclusão de que para os resultados da variável Classe Social, o apagamento de /d/ em verbos no gerúndio se comportou da mesma forma, na Classe mais baixa.

Em síntese, os resultados dos testes estatísticos multivariados indicam que as variáveis preditoras do apagamento para as Perífrases centram-se na Conjugação Verbal, Faixa Etária e Classe Social. Quanto à Conjugação verbal, verificou-se que os verbos de primeira Conjugação (-ar) favorecem às realizações de apagamento na fala dos paulistanos. Para a Faixa Etária, obteve-se um resultado parecido com os reportados para as Orações Reduzidas, onde os informantes das duas faixas mais jovens tendem a apagar mais que os mais velhos, e, por último, a variável Classe Social nos mostrou ser significativa estatisticamente após a recategorização de seus fatores. Com os resultados obtidos, verificou-se que os falantes da Classe mais alta (a combinação das Classes A + B1) tendem a pronunciar mais o /d/ nos verbos de gerúndio que os falantes das outras Classes mais baixas (B2, C1 e C2), esses, por fim, favorecem o uso do apagamento.

Para as Perífrases, as variáveis linguísticas Tempo Verbal do Auxiliar, Tamanho do Auxiliar, Tamanho do Verbo Principal, Material Interveniente entre Verbo Auxiliar e Verbo no Gerúndio e Contexto Seguinte não se revelaram significativas. Da mesma forma que as variáveis sociais Sexo/Gênero, Escolaridade, Mobilidade, Origem dos Pais e Região de Nascimento também não foram selecionadas.

Conclusão

Esta pesquisa de mestrado verificou de que forma se comporta o apagamento de /d/ em (NDO) na fala paulistana de acordo com fatores sociais e linguísticos. Num primeiro momento, os dados de (NDO) extraídos das 60 entrevistas que compõem o corpus do PROJETOSP2010, foram categorizados em Classes de Palavras, onde encontraram-se: Adjetivos, Conjunções, Substantivos, Verbos no Presente e Verbos no Gerúndio. Os casos de apagamento só foram encontrados em uma ocorrência para 98 Verbos no Presente, 93 ocorrências de 1557 da Conjunção *quando* e 1070 ocorrências de 4071 Verbos no Gerúndio (Justaposição, Orações Reduzidas e Perífrases).

As ocorrências de *quando* foram analisadas descritivamente, já que, seguindo-se os critérios do Paradigma Quantitativo, foram excluídas das análises multivariadas por apresentarem um número extremamente pequeno de dados. Entre os 60 falantes da amostra, apenas na fala de 21 deles ocorre apagamento de /d/ em *quando*. De forma geral, a análise descritiva dos casos de apagamento referentes a Conjunção *quando*, para as Variáveis Sociais, mostra que pode haver certas especificidades que indiquem favorecimento para a pronúncia sem o /d/. Estas especificidades, deu-se com os resultados de maior produção de apagamento na fala de homens, daqueles da Faixa Etária intermediária, com menor grau de instrução e pertencentes às Classes Sociais mais baixas.

Para os verbos no gerúndio verificou-se um número mais frequente de $\emptyset$ que nas outras Classes Morfológicas. Entretanto, gerúndios ocorrem em três estruturas diferentes: em “Justaposição” (de acordo com FERREIRA, 2010), Orações Adverbiais Reduzidas e, claro, em Perífrases. Começando pelos gerúndios de Justaposição, que em um total de 54 ocorrências foram encontrados 22 casos de apagamento representando 40,7%, verificou-se que essa estrutura ocorre nas entrevistas com 26 dos 60 informantes da amostra e, entre esses, há apagamento na fala de apenas 11. Esses dados foram aqui analisados de maneira descritiva – assim como os de *quando*, já que as 54 ocorrências da estrutura verbal em Justaposição representam muito pouco em relação as outras estruturas. Estratificando esses falantes, num

primeiro momento, de acordo com as variáveis sociais: Sexo/Gênero, Faixa Etária, Escolaridade, Classe Social e Região da Cidade, foi possível verificar que a distribuição de apagamento para os verbos em Justaposição se parece com a distribuição observada para a Conjunção *quando* com as mesmas variáveis, assim, encontrou-se uma percentagem maior de ocorrências na fala de entrevistados do Sexo Masculino, indivíduos da Faixa Etária Intermediária, com menos grau de instrução e aqueles provenientes da Classe Social mais baixa e de bairros mais periféricos. Olhando para as variáveis linguísticas, os Verbos Auxiliares *acabar* e *ficar* indicaram 100% de apagamento para suas ocorrências juntamente às orações que não apresentaram estrutura composta com Verbos Auxiliares. Quanto ao Tamanho do Verbo Auxiliar ESTAR, foram encontradas 10 ocorrências com esse Auxiliar, em que as orações eram compostas pela forma fonológica plena ou pela forma reduzida (“tá”, “tô”, “tava”). Perante esses dados, verificou-se que as duas ocorrências para a forma plena apresentaram apagamento e, para os casos da forma reduzida, encontrou-se 2 apagamentos em 8 ocorrências totais. Outra característica das orações de Justaposição são os Materiais Intervenientes entre o verbo auxiliar e o verbo no gerúndio, em que foram averiguados um número maior da variante não padrão para as estruturas em que não existem materiais entre os dois verbos que compõem a Perífrase das orações de Justaposição. Para a Conjugação do Verbo Principal, não houve apagamento para a terceira Conjugação, e a predominância desses casos deu-se para a primeira Conjugação, inclusive a que tem maior número de ocorrência da estrutura, seguida da segunda Conjugação. Quanto ao Tamanho por sílabas do Verbo Principal, os dados indicaram que quanto maior a extensão do verbo, maior o número de apagamentos encontrados. E, por último, a variável Contexto Seguinte aos Verbos no gerúndio mostrou que em estruturas compostas por um *Complemento* em seguida o número de apagamento é considerado menor que nas estruturas onde se verifica a *Pausa*. Além das variáveis sociais e linguísticas, as orações de Justaposição demonstraram algumas características estruturais, como verbo + segundo, verbo + verbo, verbo + verbo + verbo, etc. Conforme essas características, verificou-se que construções com dois itens lexicais em que o segundo item não é um verbo no gerúndio não foram encontrados casos de apagamento; em relação às construções com dois verbos, encontrou-se uma percentagem maior de apagamento para as ocorrências com verbos diferentes do que para as ocorrências com dois verbos iguais, diferentemente do que aconteceu com as construções com três verbos, onde

o fator relacionado a três verbos iguais favoreceu o apagamento em virtude do fator com três verbos diferentes. Comparando as estruturas com dois verbos e a estrutura com três verbos, constatou-se que entre esses dois contextos o favorecedor ao apagamento proporcionalmente seria o com dois verbos. E quanto a única ocorrência com quatro verbos, todos diferentes, teve as quatro realizações de /d/ apagadas.

Com relação às Orações Reduzidas, as variáveis que se mostraram preditoras foram a Faixa Etária e a Classe Social do falante. Para a Faixa Etária verificou-se que os mais jovens tendem a usar mais o apagamento que os mais velhos. Na variedade paulistana os resultados obtidos se aproximam dos resultados encontrados para a variedade riopretense, no que toca ao maior número de casos de apagamento para a primeira e segunda Faixas Etárias e um número bem menor para os informantes da terceira faixa que tendem a preservar a variante plena. Diferentemente das variedades paulistas, os dados de Mollica (1989) para o Rio de Janeiro, mostram que os falantes da primeira e da última Faixa Etária são os que mais tendem a preservar o /d/, enquanto para São Paulo e Rio Preto, somente os falantes mais velhos são os que tendem a preservar a forma padrão. Para a Classe Social, os indivíduos referentes a classe mais baixa são os favorecedores para o apagamento. De acordo com os modelos de efeitos mistos analisados, verificou-se que para as Orações Reduzidas, nenhuma das variáveis linguísticas se mostrou significativa e quanto as sociais, Mobilidade e Origem dos Pais impediram os modelos de convergir.

Os resultados dos testes estatísticos multivariados indicam que as variáveis preditoras do apagamento para as Perífrases centram-se na Conjugação Verbal, Faixa Etária e Classe Social. Quanto à Conjugação verbal, verificou-se que os verbos de primeira Conjugação (-ar) favorecem às realizações de apagamento na fala dos paulistanos. Para a Faixa Etária, obteve-se um resultado parecido com os reportados para as Orações Reduzidas, onde os informantes das duas faixas mais jovens tendem a apagar mais que os mais velhos. Constata-se, então, que tanto nas Orações Reduzidas quanto nas Perífrases. Desta forma, os dois conjuntos de dados, os de Perífrases e os de Orações Reduzidas, trazem a mesma característica acerca do apagamento em relação às Faixas Etárias, em que a terceira Faixa Etária desfavorece o apagamento de /d/ nos verbos de gerúndio em comparação às outras duas Faixas Etárias. Quanto às duas faixas mais

jovens, percebe-se uma propensão em utilizar mais o apagamento na faixa intermediária, o que caracterizaria indícios de uma variável estável. E, por último, a variável Classe Social nos mostrou ser significativa estatisticamente após a recategorização de seus fatores. Com os resultados obtidos, verificou-se que os falantes da Classe mais alta (a combinação das Classes A + B1) tendem a pronunciar mais o /d/ nos verbos de gerúndio que os falantes das outras Classes mais baixas (B2, C1 e C2), esses, por fim, favorecem o uso do apagamento. O valor negativo para a estimativa da Escolaridade relacionada ao Ensino Superior revela que o apagamento vem sendo utilizado em maior número na fala de pessoas com até o Ensino Médio. Por meio dessas características entre os fatores da Escolaridade, devido ao grau do apagamento apresentar-se inversamente proporcional ao aumento dos segmentos de escolarização, comprova-se o pouco prestígio da variante.

Para as Perífrases, as variáveis linguísticas Tempo Verbal do Auxiliar, Tamanho do Auxiliar, Tamanho do Verbo Principal, Material Interveniente entre Verbo Auxiliar e Verbo no Gerúndio e Contexto Seguinte não se revelaram significativas. Da mesma forma que as variáveis sociais Sexo/Gênero, Escolaridade, Mobilidade, Origem dos Pais e Região de Nascimento também não foram selecionadas.

Os estudos sociolinguísticos acerca do apagamento de /d/, geralmente, analisam o comportamento da consoante coronal dental sonora somente em ocorrências de Verbos no Gerúndio. Cabe ressaltar que este trabalho traz à tona questões que ainda não tinham sido discutidas, como o caso de verificar se o apagamento estaria se espalhando a outras classes morfológicas que possuem palavras terminadas por (NDO). Foi possível verificar que, no contexto paulistano, além dos Verbos no Gerúndio é possível encontrar um número interessante de apagamento de /d/ nas conjunções *quando* – 93 casos de apagamento em 1553 ocorrências. Apesar de não terem sido analisadas a partir de testes estatísticos multivariados, mostraram uma distribuição um pouco parecida com os dados encontrados para as Perífrases e as Orações Reduzidas Adverbiais.

Viu-se, ainda, que, em comparação aos outros trabalhos sobre apagamento citados no Capítulo 1 e também outros trabalhos sociolinguísticos para o português brasileiro, este está um passo à frente dos demais pois verifica padrões de variação correlacionados à Classe Social.

Como ressalta Oushiro (2015:327) “não se pretende haver apresentado aqui um modelo acabado para a análise da variável Classe Social, em São Paulo ou para o português brasileiro, mas se propõem critérios que devem ser refinados em estudos futuros”.

Finalmente, vale destacar que este trabalho abre caminho para futuras pesquisas envolvendo outras Classes Morfológicas com terminação (NDO) e correlações com diferentes variáveis sociais como Mobilidade, Origem dos pais e Zona da Cidade que ainda apresentam algumas lacunas.

Referências

- ALMEIDA, A. N. S., OLIVEIRA, A. J. – **Você fala cantano? Uma análise sistemática no morfema –ndo formador de gerúndio no português**. Letronica: Revista Digital do Programa de Pós-Graduação em Letras da PUCRS Porto Alegre, v. 10, n. 1, p. 200-209, janeiro-junho, 2017.
- AMARAL, A. (1920). **O dialeto caipira: gramática, vocabulário**. 4. ed. São Paulo: HUCITEC; Brasília: INL, 1982.
- BAAYEN, R. Harald. **Analysing Linguistic Data: a Practical Introduction to Statistics**. Cambridge: Cambridge University Press, 2008.
- BAGNO, Marcos. **A língua de Eulália: novela sociolinguística**, São Paulo: Contexto, 1997.
- BARBOSA, P; MADUREIRA, S. **Manual de fonética acústica experimental: aplicação a dados do português**. São Paulo: Cortez, 2015.
- BAYLEY, Robert. “**The quantitative paradigm**.” In: Chambers, J.K. & Schilling-Estes, N. (Eds.), **The handbook of language variation and Chang**. pp. 85–107. Malden, MA: Blackwell, 2 ed., 2013 [2002].
- CHAMBERS, J. K. **Sociolinguistic theory**. Oxford: Blackwell, 1995.
- CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (Ed.). **The handbook of language variation and change**. Oxford: Blackwell, 2004.
- CHOMSKY, N. **Syntactic Structures**. The Hague: Mouton, 1957.
- COUTINHO, I. de L. **Pontos de gramática histórica**. 6. ed. Rio de Janeiro: Livraria acadêmica, 1976.
- CRISTÓFARO SILVA, T. **Fonologia: por uma análise integrada à morfologia e à sintaxe**. In: Cadernos do Departamento de Letras Vernáculas. Anais da 2ª. Semana de Estudos Portugueses. v. 2, p. 56-65. Belo Horizonte: FALE/UFMG, 1996. Disponível em: <http://www.projetoaspa.org/cristofaro/publicacao/pdf/originais/artigos/integrada.pdf>. Acesso em: 03 nov. 2017.
- _____. **Dicionário de Fonética e Fonologia**. 1. ed. São Paulo: Editora Contexto, 2015.
- CUNHA, C.; CINTRA, L. F. L. **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira, 2010.
- FERREIRA, J. S. **O apagamento do /d/ em morfema de gerúndio no dialeto de São José do Rio Preto**. 2010. 142f. Dissertação (Mestrado em Estudos Linguísticos). Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2010.

- GNERRE, Maurízzio. **Linguagem, escrita e poder**. São Paulo: Martins Fontes, 1987
- GUY, Gregory R. “**The quantitative analysis of linguistic variation**.” In: Preston, D. (Ed.), *American Dialect Research*, pp. 223–249. Amsterdam: Benjamins, 1993.
- GUY, Gregory; ZILLES, Ana. **Sociolinguística quantitativa: instrumental da análise**. São Paulo: Parábola, 2007.
- HOLMES, J.; MEYERHOFF, M. (Orgs.). **The handbook of language and gender**. Oxford: Blackwell Publishers, 2006.
- HORA, Dermeval da; AQUINO, Maria de Fátima S. **Da fala para a leitura: análise variacionista**. Alfa: Revista de Linguística, v. 56, n. 3, p. 1099-1115, 2012.
- IPEA. “**Comunicados do IPEA no. 115 – Perfil dos migrantes em São Paulo**.”, 2011. Último acesso em 16 nov./2011. URL http://www.ipea.gov.br/portal/images/stories/PDFs/comunicado/111006_comunicadoi pea115.pdf
- LABOV, William. **The social motivation of a sound change**. Word, 19, 1963.
- . **The social stratification of English in New York City**. Washington, D. C.: Center of Applied of Linguistics, 2006 [1966].
- . **Building on empirical foundations**. In: LEHMANN, Winfred P.; MALKIEL, Yakov (Eds.) **Perspectives on historical linguistics**. Amsterdam/ Philadelphia: John Benjamins, 1982
- . **Principles of linguistic change: internal factors**. Vol. 1. Oxford: Blackwell Publishers, 1994.
- . “**The anatomy of style-shifting**.” In: Eckert, P. & Rickford, J. R. (Eds.), **Style and Sociolinguistic Variation**, pp. 85–108. Cambridge: Cambridge University Press, 2001a.
- . **Principles of linguistic change: social factors**. Oxford & Cambridge: Blackwell, 2001b.
- . **Sociolinguistics Patterns**. Philadelphia: University of Pennsylvania, 1972 [ed. Br.: 2008. **Padrões Sociolinguísticos**. Trad. de Marcos Bagno, Maria Marta Pereira Scherre e Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008.
- LEHMANN, C. 1982a. **Thoughts on Grammaticalization**. A programmatic sketch. Köln, *Arbeiten des Kölner Universalien-Projekts*, vol. I.
- LEMLE, M. **Análise sintática: teoria geral e descrição do português**. São Paulo: Ática, 1984. (Ensaio; 106).
- LEVSHINA, N. **How to do Linguistics with R: data exploration and statistical analysis**. Amsterdam/Philadelphia: John Benjamin Publishing Company, 2015.

- LIGHTFOOT, David. **The development of language acquisition, change and evolution.** Malden (MA), Oxford: Blackwell, 1999.
- LOBATO, L. M. P.. **Os verbos auxiliares em Português Contemporâneo. Critérios de auxiliaridade.** In____. **Análises linguísticas.** Tradução: Maria Angela Botelho Pereira. Petrópolis: Vozes, 1975.
- LUCENA, R.M. de, VASCONCELOS, D.C de. **Apagamento da oclusiva dental no dialeto do brejo paraibano: uma regra variável.** Revista: A Cor das Letras – UEFS, n. 8, 2007
- MARROQUIM, M. **A língua do Nordeste (Alagoas e Pernambuco).** São Paulo: Nacional, 1934.
- MARTINS, E. F.. **Atlas linguístico do Estado de Minas Gerais: o princípio da uniformidade da mudança lingüística nas características fonéticas do português mineiro.** Revista Virtual de Estudos da Linguagem – ReVEL. Ano 4, n. 7, ago. 2006. Disponível em: <<http://paginas.terra.com.br/educacao/revel/>>. Acesso em: 10 out. 2006.
- MARTINS, I. da S.; BUENO, E. S. da S. **Estudo do gerúndio – a transformação de [no] em [n] no português falado na região de fronteira.** Sociodialeto, v.1, n.4, jul. 2011. Disponível em <http://www.sociodialeto.com.br/edicoes/9/28092011064716.pdf>. Acesso em: 05 set. 2017.
- MARTINS, I. F. de M. **Apagamento da oclusiva dental /d/ no grupo “ndo” na fala de João Pessoa.** Dissertação (Mestrado em Língua Portuguesa) – Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 2001.
- MARTINS, I. F. de M. **Variação do apagamento da oclusiva dental /d/ no grupo “ndo” na fala de João Pessoa.** In: SOARES, M. E.; ARAGÃO, M. do S. S. de (orgs.). XVI JORNADA DE ESTUDOS LINGÜÍSTICOS. Fortaleza: UFC, p. 337-342. 1999.
- _____. **Apagamento da oclusiva dental /d/: perspectivas variacionista e fonológica.** In: HORA, D. da. Estudos sociolingüísticos: perfil de uma comunidade. João Pessoa: ILAPEC, 2004. p. 55-82.
- MATTOS, P. B. & M. C. de M. MOLLICA (1989) **Dois processos de assimilação fonológica no português falado semi-espontâneo do Rio de Janeiro.** Relatório final do projeto Mecanismos funcionais do uso da língua. Rio de Janeiro: UFRJ, inédito.
- _____. (1992) **Pela Conjugação das abordagens variacionista e difusionista.** Revista de estudos da linguagem, 1. Belo Horizonte: UFMG: 53-64.
- MELO, G. C. de. (1946). **A língua do Brasil.** 2. ed. Rio de Janeiro: Fundação Getúlio Vargas, 1971.
- MENDES, R.B. e OUSHIRO, Livia. **O paulistano no mapa sociolinguístico brasileiro.** Alfa, São Paulo, v.56, n.3, p.963-991, 2012

- MENDES, Ronald Beline. **A gramaticalização de estar + gerúndio no português falado**. 1999. 110f. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP.
- _____. Ronald Beline. **Língua e variação**. In: FIORIN, José Luiz (Org.) et al. **Linguística? Que é isso?** São Paulo: Contexto, 2015.
- MILROY, Leslie. **Language and social networks**. Oxford: Basil Blackwell, 2 ed., 1987 [1980].
- MOLLICA, M. C. **Um padrão etário recorrente em fenômenos de variação fonológica**. Estudos Linguísticos XVII Anais de Seminários do GEL. São Paulo, 1989.
- NASCIMENTO, K. R. S. do; ARAÚJO, A. A. de; CARVALHO, W. J. de A. **A redução do gerúndio no falar popular de Fortaleza: um olhar variacionista**. Veredas. Juiz de Fora, v.2. p. 398-413. 2013. Disponível em: <http://www.ufjf.br/revistaveredas/files/2014/04/21%C2%BA-ARTIGO.pdf>. Acesso em: 05 set. 2017.
- NASCIMENTO, L.; MOTA, J. **A ausência do ‘d’ no gerúndio: com base em inquéritos experimentais do projeto ALIB**. Hyperion Letras. Salvador, s/v, n.7, s/p. 2004. Disponível em: <<http://www.hyperion.ufba.br/revista_7_04.htm>>. Acesso em: 10 fev. 2017.
- OUSHIRO, L. **Introdução à Estatística para Linguistas**. Zenodo, 2017. Disponível em <<http://doi.org/10.5281/zenodo.822070>>
- _____. **Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São de Paulo**. Tese de Doutorado. São Paulo: FFLCH-USP, 2015a. 390f.
- _____. **“Tratamento de dados com o R para análises sociolinguísticas.”** In: FREITAG, Raquel MeisterKo (Ed.), **Metodologia de coleta e manipulação de dados em Sociolinguística**. São Paulo: Blucher, 2014. URL: <http://dx.doi.org/10.5151/BlucherOA-MCMDS-10cap>
- _____. **Uma análise variacionista para as Interrogativas-Q**. Dissertação de Mestrado. São Paulo: FFLCH-USP, 2011. 174f.
- R CORE TEAM. **“R: A language and environment for statistical computing.”** R Foundation for Statistical Computing, Vienna, Austria, 2018. URL: <http://www.R-project.org/>
- SANTOS, W.S. **A morfologia do indicativo na expressão do modo subjuntivo em São Paulo e São Luís**. Dissertação de Mestrado. São Paulo: FFLCH-USP, 2014. 141f.
- SAUSSURE, Ferdinand de. **Curso de Linguística Geral**. Trad. Antônio Chelini, José Paulo e Izidoro Beinkstein. São Paulo; Cultrix, 1975[1916].
- SPENCER, Andrew. **Phonology**. Oxford: Blackwell Publishers, 1996

- TAGLIAMONTE, Sali. **Analysing sociolinguistic variation**. São Paulo: Cambridge University Press, 2006.
- TARALLO, F. **A pesquisa sociolinguística**. São Paulo: Martins Fontes, 1994.
- TEIXEIRA, J. A. **O falar mineiro**. Revista do Arquivo Municipal de São Paulo, n. 45, ano 4, p. 5-100, mar. 1938.
- VIEIRA, M. S. **Apagamento do /d/: abordagem sociolinguística sob a perspectiva do gênero sexual**. Sociodialeto. Campo Grande, v.1, n. 4, p. 1-27, jul. 2011. Disponível em: <http://www.sociodialeto.com.br/edicoes/9/28092011063729.pdf>. Acesso em: 05 set. 2017.
- WEINREICH, U; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da mudança linguística**. Tradução de Marcos Bagno. São Paulo: Parábola, 2006[1968].

Anexos

Anexo A: Roteiro de entrevista

PARTE I

Bairro (aprox. 10 min.)

Objetivos: descobrir o grau de enraizamento do informante no bairro onde vive/outros bairros; descobrir o grau de mobilidade da pessoa

(1) Há quanto tempo você mora na (Mooca, Pinheiros, Bexiga...)?

(2) Você gosta de morar aqui?

(3) Por que você escolheu morar neste bairro? (manter em mente que o informante pode não ter escolhido morar ali: pode ter sido porque a família já morava lá etc.)

(a) Se o informante mora há bastante tempo: Como era o bairro antes/quando você se mudou pra cá? Mudou muito?

(b) Se o informante não mora lá há muito tempo, ou se nem sempre viveu ali: Em que outros lugares você já morou? Como era lá comparado com esse bairro aqui? Onde você preferia morar?

(4) Você tem bastante contato com as pessoas aqui do bairro?

(5) Hoje, tem algum outro bairro em que você gostaria de morar?

Infância (aprox. 5 min.)

Objetivos: relaxar o informante (em geral, as pessoas gostam de falar sobre sua infância); obter informações sobre mudanças no bairro/cidade de São Paulo; grau de mobilidade do informante; obter informações sobre Escolaridade

(6) E como foi a sua infância (no bairro X)? Você pode contar um pouco de como foi, o que você fazia...?

(a) brincava na rua/dentro de casa? Do que vocês brincavam?

(b) como eram os seus pais? eram rígidos...?

(7) Você foi pra escola no mesmo bairro? Como era a escola? Você sempre estudou na mesma escola? Você gostava de ir pra escola? Tem algum professor que te marcou? Até que série você estudou?

(8) Enquanto ainda era criança/adolescente, você ia pra outros lugares dentro da cidade de São Paulo?

(pra onde, pra fazer o quê...)

Família (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante, grau de enraizamento no bairro/cidade

(9) Você tem irmãos? (é possível que esta informação já tenha aparecido na parte sobre infância; neste caso, falar algo como: “Você disse que tem uma irmã... Você tem outros irmãos?”) Quantos anos eles têm?

(10) Seus pais são daqui de São Paulo mesmo? (Se sim, perguntar também sobre avós, bisavós... até encontrar a primeira geração da família que veio pra cá). Quantos anos eles (pais) têm? Quando seus (pais/avós/bisavós) vieram pra São Paulo? Você sabe por que eles vieram?

(11) Você tem filhos? Quantos anos eles têm?

(12) E o resto da família, tios, primos, também vivem aqui em São Paulo? (Se sim, em que bairros? Moram próximo? Se não, onde? Tem contato sempre?)

Trabalho/Ocupação (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante; características socioeconômicas

(13) Você trabalha aqui por perto? (se não souber onde a pessoa trabalha)

(14) O que você faz?

(15) Você gosta do seu serviço? (Se não, o que a pessoa preferiria fazer?)

Lazer (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante; mobilidade na cidade; características socioeconômicas

(16) E nas horas de lazer, o que você e sua família gostam de fazer? (Se saem, vão pra que lugares?) Você acha que a cidade de São Paulo tem boas opções de lazer? Quais?

(17) A maioria dos seus amigos mora aqui nesse bairro mesmo? (Se não, onde?)

(18) Você tem alguma atividade de recreação em grupo, algum clube...?

(19) Quais são seus amigos mais antigos? Você mantém contato com os amigos de colégio?

(20) Você costuma viajar? Pra que lugares já viajou? Que lugares gostaria de conhecer?

PARTE II

A Cidade de São Paulo (aprox. 20 min.)

(21) Uma pesquisa recente mostrou que 57% das pessoas em São Paulo deixariam a cidade caso pudessem. O que você acha disso? (se o informante perguntar, a pesquisa é do Ibope e foi publicada em janeiro/2010¹³)

(22) Você gosta de morar em São Paulo? (Se não, em que lugar preferiria morar?) Por quê?

(23) O que você acha que caracteriza a cidade (tanto as coisas boas quanto ruins)?

(24) O que você acha que caracteriza o paulistano (tanto as coisas boas quanto ruins)?

(25) Olhando pra mim, você diria que eu sou paulistano(a)? Por quê?

(26) O que você mais gosta em São Paulo?

(27) O que você não gosta em São Paulo? (a depender do tópico mencionado pelo informante – violência, trânsito, poluição etc. – procurar explorar mais o assunto. P.ex.: você já foi assaltado? O que aconteceu? (para obter narrativa pessoal) O que o governo poderia fazer pra solucionar esse problema? (para obter uma fala mais distanciada)) (explorar o subtópico por aproximadamente 10 min.)

(28) Pras pessoas que não vivem em São Paulo, como você acha que elas imaginam que seja a cidade? Qual é a imagem que as pessoas de fora de São Paulo têm da cidade?

(29) Você falou que já foi pra (X, Y, Z). Quando você foi pra esses lugares, as pessoas percebiam que você era paulistano? (Se sim) como elas percebiam?

(30) Quando você conhece alguém, você percebe se a pessoa não é daqui de São Paulo?

(a) (Se sim) como você percebe? (Se o informante mencionar o modo de falar, seguir nessa linha e tentar conseguir informações mais precisas. É normal as pessoas não saberem definir o porquê, mas devemos tentar tirar mais informações).

(b) (Se não) Quando você ouve uma pessoa falando, por exemplo, você percebe que ela é de fora pelo sotaque?

Avaliações Linguísticas (aprox. 10 min.)

(31) Qual o sotaque do Brasil que você mais gosta? E tem algum que te irrita? Como é que (o gaúcho/o carioca/o mineiro/o caipira etc. – a depender dos sotaques mencionados) fala(m)?

(32) Aqui em São Paulo tem muito (i)migrante, né? De onde vêm a maioria das pessoas? Tem algum bairro específico em que eles (italianos, coreanos, nordestinos etc.) se concentram?

¹³ Disponível em http://www.nossasaopaulo.org.br/portal/arquivos/Pesquisa_IRBEM_Ibope_2010_completa.pdf

(33)E como é que as pessoas falam na cidade de São Paulo? (evitar usar a palavra “sotaque”)

(34)E aqui dentro da cidade de São Paulo, você vê diferenças entre as diferentes regiões, diferentes bairros? (Se sim) Você poderia dar alguns exemplos? (É normal as pessoas responderem que não, mas se responderem que sim, tente obter informações mais precisas)

(Imprimir a lista de palavras, a notícia e o trecho, todos abaixo, para mostrar ao informante)

LISTA DE PALAVRAS: Apresentar ao informante a lista de palavras e dizer: “Agora eu queria te pedir pra ler algumas coisas. Eu tenho aqui uma lista de palavras, e eu queria que você lesse cada uma delas. Você pode dar uma olhada primeiro, antes de começar a ler.” Esperar que o informante leia a lista de palavras. Depois perguntar:

(35)A gente estava falando de sotaques agora há pouco... Como você acha que uma pessoa do interior falaria algumas dessas palavras? Tem diferenças?

(36)E como um carioca falaria algumas dessas palavras?

(37)Tem mais algum sotaque no Brasil que você conhece?

LEITURA DE NOTÍCIA DE JORNAL: Dizer ao informante: “Agora eu queria que você lesse esse texto. Você pode ler em voz baixa antes de falar em voz alta.” Esperar que o informante leia a notícia.

DEPOIMENTO: Dizer ao informante: “Agora eu queria que você lesse esse texto. Você pode ler em voz baixa antes de falar em voz alta.” Esperar que o informante leia a notícia.

(38)O que você acha desse modo de falar: “Você tá entendendo o que eu tô dizendo?” (com “en” ditongado e exagerado). Se a pessoa manifestar uma atitude negativa, perguntar: Como você acha que deveria ser? O que tem de errado aí?

(39)Quem você acha que fala assim? (se falar “paulistanos”, perguntar: “você acha que todos os paulistanos falam assim ou é uma coisa de uma região mais específica na cidade?”)

(40)Você fala desse modo?

(41)E o que você acha de “Me vê dois pastel e um chopps?” (Fazer as mesmas perguntas 38–40 acima)

(42)E o que você acha de “A porta tá aberta.” (com tepe exagerado). Fazer as mesmas perguntas 38–40 acima.

(43)E tem mais algum modo de falar que você acha assim bem paulistano?

Palavras

alma
amargo
animação
argola
atitude
barqueiro
biscoito
cacique
carteiro
cerca
chácara
circo
cisne
curto
defender
discoteca
elefante
enchente
entender
entretenimento
erguer
fazenda
felicidade
firme
fornalha
furgão
fusquinha
geleira
gérmen de trigo
gordo

gula
Hércules
hilário
irmã
justiça
lento
mortadela Marba
mosca
necessidade
noite
orca
Ordem e progresso
Orgânico
órgão
ostracismo
penteadado
pertencimento
perto
porto
presente
rapidez
riqueza
sabor de menta
soberba
trabalho
turco
urgente
utilidade
vulto
zebra

Notícia

08/02/2010

Moradores cobram solução para área alagada em SP e aguardam encontro com prefeito

Moradores da região do Jardim Pantanal protestaram nesta segunda-feira em frente à Prefeitura de São Paulo contra a inundação em bairros da zona leste, que completa dois meses hoje. Uma comissão de manifestantes foi recebida pela Secretaria de Relações Institucionais e apresentou reivindicações, mas o prefeito Gilberto Kassab (DEM) não participou do encontro. Representantes dos manifestantes esperam que o governador José Serra (PSDB) também participe da reunião de sexta. De acordo com os integrantes da comissão, não havia integrantes do governo estadual no encontro realizado hoje na prefeitura. Além de cobrar a limpeza imediata das águas nos bairros, os moradores cobram uma solução para a falta de moradia. A maioria dos desalojados continua em escolas municipais e, com o início do ano letivo, temem não ter para onde ir.

Depoimento

Tá chovendo muito! Choveu tanto, tanto na semana passada que ficou uma piscina na minha casa. Ó, pra você ver: Molharam todos os armários, a cama, os colchões, tudo... Foi um sacrifício... O que a gente fez? Nós tivemos que erguer os móveis pra limpar tudo: a geladeira, o forno... minha irmã até veio me ajudar, sabe? E meus filhos compraram umas cadeiras novas, mas é aquela coisa, assim... quando chover de novo, vai molhar tudo outra vez. Você fica sem ter o que fazer. E tem um rio lá perto que sempre alaga... quer dizer, é água dentro e fora de casa! Daí, o que acontece? Fica aquele trânsito, os carros todos parados, a gente demora um tempão pra chegar em casa... Não aguento mais enchente nessa cidade... Agora que eu vou fazer? Os políticos falam, falam, mas eles tinham que fazer alguma coisa urgente. Você tá entendendo o que eu tô dizendo?

Anexo B: Informantes do corpus de entrevistas sociolinguísticas

O Quadro apresenta o significado dos códigos utilizados na Tabela 1 a seguir.

S – Sexo / Gênero	F – Feminino M – Masculino
I – Idade	Em anos
E - Escolaridade	C – até Ensino Médio S - Ensino Superior
R – Região	V – Centro Velho E – Centro Expandido P – Periferia
Z – Zona	C – Central N – Norte S – Sul L – Leste O – Oeste
P – Origem dos Pais	E – Estrangeiros I – Interior SP/MG/PR N – Norte/Nordeste P – São Paulo – capital X – Mista (pais de origem diferente)
M – Mobilidade	B – sempre morou no mesmo bairro Z – sempre morou na mesma zona M – morou em diferentes zonas
C – Classe Social	A – alta B1 – média alta B2 – média media C1 – média baixa C2 – baixa alta
IS – Índice Socioeconômico	Índice 1 – 5

Tabela Características sociais dos informantes e respectivas taxas de emprego de apagamento de /d/ em QUANDO e nos Verbos no Gerúndio

Informante	S	I	E	R	Z	P	M	C	IS	QUANDO %(TOTAL)	GERÚNDIO %(TOTAL)
AlbertoM	M	41	C	P	O	0	B	C1	2,7	63(12)	72(89)
AliceM	F	30	C	P	L	0	B	C1	2,6	0	20(9)
AmadeuC	M	89	S	E	O	0	M	B2	3,4	0	9(3)
AnaS	F	32	C	P	O	1	M	C2	2,2	0	70(40)
AnselmoN	M	63	S	P	N	0	M	B2	3,1	0	24(10)
ArianeG	F	36	S	P	N	0	B	B1	3,7	0(1)	4(2)
BenjaminT	M	84	C	E	L	0	M	B1	3,6	0	4(2)
BiancaV	F	33	S	V	C	0	M	A	4,3	0	7(4)
CarlosJ	M	31	S	P	S	1	B	A	4,3	0	2(2)
CarolinaA	F	24	C	V	C	0	M	C2	2,5	12(9)	35(42)
CelinaB	F	50	C	E	O	0	Z	C2	2,2	23(6)	60(43)
DalsonO	M	24	C	P	L	0	B	C1	2,7	4(1)	22(16)
DarcyM	M	74	C	P	O	0	M	C2	2,1	0	7(2)
DouglasA	M	24	C	P	N	0	Z	C2	2,6	0	100(57)
EdnaC	F	81	C	E	L	0	Z	C2	2,3	0	7(3)
ElaineS	F	56	S	P	L	0	B	B2	3,2	0	43(29)
EthelM	F	60	S	P	N	1	B	B2	3,4	0	16(8)
FabianaB	F	26	S	E	L	3	M	B2	3,5	0	4(2)
FelixL	M	81	C	V	C	1	M	B2	3,3	0	0
FredericoV	M	37	S	E	O	0	Z	B1	3,9	0	7(7)
GilvanS	M	60	S	P	L	0	M	B2	3,4	0	2(1)
HugoF	M	83	C	E	O	0	M	C2	2,5	0	5(4)
IaraM	F	65	S	P	S	0	Z	C1	2,8	0	0
JaimeN	M	33	S	V	C	1	B	B2	3,4	0	24(16)
JanainaB	F	28	C	P	S	4	M	C1	3	7(3)	21(15)
JorgeV	M	72	S	V	C	1	M	A	4,5	0	47(39)
JoseN	M	46	S	P	N	0	M	C2	2,5	8(3)	24(30)
LeonS	M	47	S	E	L	0	M	A	4,2	5(1)	44(30)
LucasS	M	25	C	V	C	0	B	C2	2,4	0	43(63)
LucianaM	F	58	S	E	S	0	M	B2	3,3	0	7(3)
LucianoT	M	39	C	E	O	0	M	B1	4	0	3(2)
MadalenaM	F	51	S	P	O	2	Z	B2	3,5	52(14)	38(23)
MarciaM	F	86	C	V	C	1	Z	C1	2,7	0	4(2)
MarietaS	F	54	C	P	N	0	Z	B2	3,3	0	2(1)
MauricioB	M	36	C	E	L	4	B	C1	3	10(2)	11(6)
MeireC	F	48	C	E	S	1	M	C1	3	0	16(14)
MilenaF	F	19	C	E	N	3	M	C1	3	14(1)	68(27)
MuriloM	M	50	S	V	C	1	M	A	4,2	0	24(12)
NelsonF	M	37	C	P	S	0	Z	C2	2,4	11(3)	79(75)
NiltonR	M	36	S	P	S	0	B	B2	3,2	0	0
PedroS	M	66	S	E	S	2	M	A	4,4	0	44(43)
PolianaM	F	45	S	V	C	4	B	A	4,4	0	0
RaissaO	F	22	S	P	S	2	M	B2	3,1	12(2)	16(11)
RenataC	F	35	C	V	C	1	B	B2	3,2	15(2)	16(9)
RicardoT	M	29	S	E	O	1	Z	B1	3,8	7(1)	31(12)
RobertoS	M	22	C	E	O	4	B	A	4,1	0	29(19)
RobsonF	M	31	C	P	S	0	M	B2	3,3	0	20(18)
RodolfoR	M	63	C	P	N	0	M	C2	2,5	0	2(1)
RodrigoR	M	30	S	E	N	3	B	A	4,2	12(3)	34(32)

RomuloS	M	60	C	E	S	3	M	B1	3,6	0	0
RuthC	F	41	C	P	L	0	Z	C1	2,6	56(9)	57(51)
SandroS	M	52	C	P	N	0	M	C2	2,4	57(16)	82(75)
SergioA	M	27	S	E	L	4	M	A	4,3	7(1)	23(10)
SilviaB	F	25	S	P	N	0	M	C1	2,6	17(2)	62(29)
SoraiaS	F	72	C	E	N	0	M	C1	2,6	0	0
SueliO	F	66	S	E	L	1	M	B2	3,2	0	1(1)
VaniaD	F	61	S	V	C	1	M	C1	3	0	0
VeraD	F	77	C	P	S	0	M	C1	3	0	12(6)
VivianeC	F	26	S	E	O	3	M	B1	3,6	4(1)	31(25)
VivianT	F	76	S	E	O	0	M	C1	2,9	0	2(1)

Anexo C: Verbos no Gerúndio em Justaposição

1- Ocorrências das Estruturas de Gerúndio em Justaposição por Falante (N=54)

(As frases em Negrito indicam que as ocorrências tiveram o /d/ apagado)

AlbertoM

- **a gente ficava andando por aí brincando andando de bicicleta**
- **tem que ficar estudando estudando**
- **acabar se interessando querendo ver**

AnaS

- **está sempre me ligando falando como**

ArianeG

- **ficar dois anos pesquisando escrevendo**

BiancaV

- ficar vendo filme em casa estudando lendo
- pode estar ligando falando ah você

CarlosJ

- **uma bola de neve vai aumentando aumentando aumentando...mas acho**
- **mais assim viajando viajando com a minha família**

CarolinaA

- foi conhecendo conhecendo aí foi
- esse outro estava em outro canto... esperando olhando pra ver
- **ficasse em casa... brincando quebrando fazendo besteira**
- **e acabava voltando entrando**
- **e foi indo indo indo**
- foi passando passando passando
- **foi passando passando passando**

CelinaB

- **estar trabalhando fazendo**
- **entendendo entendendo o que**
- **né entendendo comendo**

DarcyM

- a gente ia a pé numa boa cantando fazendo fe/sei lá
- ficar repetindo repetindo repetindo

DouglasA

- ...fazendo arrumando trabalhei
- **muita gente fica escolhendo escolhendo emprego**
- **eu fui me enturmando conhecendo mais gente**
- eu estava passando subindo a escolinha

EdnaC

- fiquei cuidando cuidando do meu pai

FelixL

- e acabaram se relacionando conversando e batendo papo
- despachava os papéis em pé assinando vendo colocando
- a vida e lutando brigando

IaraM

- está querendo querendo me mandar

JaimeN

- **churrasquinho gritando conversando dançando cantando**
- vai crescendo ouvindo outros

JoseN

- fica passando passando
- você vai indo segundo uma semana

LeonS

- estou lendo estudando ou viajando então não tenho
- **está lá a secretária escrevendo fazendo o relatório**

LucasS

- **vai enchendo enchendo enchendo**

MarietaS

- as pessoas ficam na rua conversando fazendo batendo papo
- está acontecendo mundo afora

MeireC

- mandei e-mail reclamando falando
- acabei desistindo quando eu vim
- estão sempre cantando quando estão

NelsonF

- **aí fomos ficando ficando ela foi**
- **nós acabamos acostumando chamando de tia**

RenataC

- fiquei pensando pensando pensando por conta

RicardoT

- ficava todo mundo na rua comemorando soltando rojão

RobertoS

- e acabei desistindo saindo da

- das pessoas falando descrevendo determinadas

Robsonf

- descemos a ribanceira rolando capotando

RuthC

- ficar me maquiando arrumando cabelo

SueliO

- você vai ficar conversando brigando discutindo e sem

VivianT

- ficava lá conversando contando historinhas

2 – Tabelas de Distribuição

Distribuição de apagamento em Justaposição por Variáveis Sociais para os 11 informantes não categóricos (N = 28)

		d o	%	Total
Sexo/Gênero	Feminino	10	71	14
	Masculino	12	86	14
Faixa Etária	20 a 34	6	60	10
	35 a 59	16	89	18
Escolaridade	Ensino Médio	17	81	21
	Ensino Superior	5	71	7
Classe Social	A	3	75	4
	B1	1	100	1
	B2	1	50	2
	C1	3	100	3
	C2	14	78	18
Região da Cidade	Centro Expandido	4	80	5
	Centro Velho	5	50	10
	Periférica	13	100	13

Distribuição de apagamento em Justaposição por Variáveis Linguísticas para os 11 informantes não categóricos (N= 28)

		d ø	%	Total
Verbo Auxiliar	Acabar	3	100	3
	Estar	4	67	6
	Ficar	4	100	4
	Ir	5	56	9
	Sem	6	100	6
Tamanho do Auxiliar ESTAR	Pleno	2	100	2
	Reduzido	2	25	8
Material Interveniente	Com	2	67	3
	Sem	20	80	25
Conjugação	1 ^a	16	94	17
	2 ^a	6	60	10
	3 ^a	0	0	1
Tamanho do Verbo Principal	2 sílabas	0	0	3
	3 sílabas	10	91	11
	4 ou + sílabas	12	86	14
Contexto Seguinte	Complemento	12	75	16
	Pausa	10	83	12

Anexo D: Gráficos da variável Índice Socioeconômico para Perífrases e Orações Reduzidas

**Proporção de apagamento de /d/ em R.Adverbiais
de acordo com o índice socioeconômico dos informantes**

Anexo E: Modelos Multivariados que não Convergiram

Verbo ESTAR

(N = 1770)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.540887	0.525601	-1.029	0.30344	
sexogenM	0.695090	0.434333	1.600	0.10952	
fai.eta2a	-0.530160	0.509282	-1.041	0.29788	
fai.eta3a	-2.648625	0.557516	-4.751	<0.001	***
EscolaridadeS	-0.587420	0.432966	-1.357	0.17487	
conjugacao2a	0.003754	0.193954	0.019	0.98456	
conjugacao3a	-0.986899	0.315006	-3.133	0.00173	**
mat.intS	-1.598328	0.859908	-1.859	0.06307	.
tamanhoauxR	-0.214936	0.261634	-0.822	0.41135	
mat.intS:tamanhoauxR	1.875801	0.898305	2.088	0.03678	*

glmer(ndo~sexogen+fai.eta+Escolaridade+conjugacao+mat.int*tamanhoaux+(1|falante)+(1|ocor), data=gestar, family = binomial)

Perífrases e Reduzidas

Estimativas (em logodds) de valores de significância para Perífrases e Orações Reduzidas (N = 4017)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.7479678	0.0007220	-1036	<0.001	***
sexogenM	0.7534859	0.0007218	1044	<0.001	***
fai.eta2a	-0.5559118	0.0007217	-770	<0.001	***
fai.eta3a	-2.8947691	0.0007219	-4010	<0.001	***
EscolaridadeS	-0.6355478	0.0007219	-880	<0.001	***
tipo.gerR	-0.2633953	0.0007218	-365	<0.001	***

glmer(ndo~ sexogen+fai.eta+Escolaridade+tipo.ger +(1|falante)+(1|ocor), data = gpr, family = binomial)

Estimativas (em logodds) de valores de significância para Perífrases com variáveis sociais (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.9107	0.4695	-1.940	0.0524	.
sexogenM	0.7579	0.4368	1.735	0.0827	.
fai.eta2a	-0.5179	0.5164	-1.003	0.3158	
fai.eta3a	-2.6071	0.5533	-4.711	< 0,001	***
EscolaridadeS	-0.5453	0.4354	-1.253	0.2103	

glmer(ndo~sexogen+fai.eta+Escolaridade+(1|falante)+(1|ocor), data=gp, family = binomial)

Estimativas (em logodds) de valores de significância para Perífrases com variáveis sociais (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
--	-------------------	--------------------	----------------	--------------------------	--

(Intercept)	-0.9107	0.4695	-1.940	0.0524	.
sexogenM	0.7579	0.4368	1.735	0.0827	.
fai.eta2a	-0.5179	0.5164	-1.003	0.3158	.
fai.eta3a	-2.6071	0.5533	-4.711	< 0,001	***
EscolaridadeS	-0.5453	0.4354	-1.253	0.2103	.

glmer(ndo~sexogen+fai.eta+Escolaridade+(1|falante)+(1|ocor), data=gp, family = binomial)

Perífrases

Estimativas (em logodds) de valores de significância para Perífrases com todas as variáveis e Índice Socioeconômico (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	0.71454	1.71125	0.418	0.67627	
sexogenM	0.49591	0.31735	1.563	0.11814	
fai.eta2a	-0.88263	0.37393	-2.360	0.01826	*
fai.eta3a	-3.17389	0.44839	-7.078	<0.001	***
EscolaridadeS	-0.01892	0.43601	-0.043	0.96540	
indice2,2	-0.16539	1.20380	-0.137	0.89072	
indice2,3	-0.97337	1.51068	-0.644	0.51936	
indice2,4	0.24856	1.19167	0.209	0.83478	
indice2,5	-1.13143	1.13174	-1.000	0.31744	
indice2,6	-0.76595	1.14936	-0.666	0.50514	
indice2,7	0.04746	1.16905	0.041	0.96762	
indice2,8	-16.62093	1051.10624	-0.016	0.98738	
indice2,9	-1.20044	1.67290	-0.718	0.47302	
indice3	-0.50926	1.11935	-0.455	0.64914	
indice3,1	-0.69156	1.22028	-0.567	0.57091	
indice3,2	0.03494	1.19036	0.029	0.97658	
indice3,3	-3.58512	1.19860	-2.991	<0.005	**
indice3,4	-0.03460	1.23672	-0.028	0.97768	
indice3,5	-1.41052	1.29330	-1.091	0.27543	
indice3,6	-0.89338	1.21385	-0.736	0.46174	
indice3,7	-16.57758	863.71953	-0.019	0.98469	
indice3,8	-2.68657	1.60195	-1.677	0.09353	.
indice3,9	-4.07732	1.56990	-2.597	0.00940	**
indice4	-3.66119	1.51615	-2.415	0.01574	*
indice4,1	-0.25011	1.40482	-0.178	0.85869	
indice4,2	-0.21382	1.28967	-0.166	0.86832	
indice4,3	-1.88228	1.24163	-1.516	0.12953	
indice4,4	-15.27975	1046.65449	-0.015	0.98835	
indice4,5	3.14984	1.43562	2.194	0.02823	*
mobilidadeM	0.63845	0.36292	1.759	0.07855	.
mobilidadeZ	1.51961	0.51598	2.945	<0.005	**
origem.pais	-0.33945	0.15101	-2.248	0.02459	*
regiaoP	0.37601	0.33990	1.106	0.26862	
regiaoV	-0.84395	0.43883	-1.923	0.05446	.
tempoauxN	-0.38078	1.25654	-0.303	0.76186	

tempoauxPA	-0.38621	1.24788	-0.309	0.75695	
tempoauxPR	-0.58552	1.24561	-0.470	0.63830	
tamanhoauxR	-0.05845	0.12136	-0.482	0.63009	
conjugacao2a	-0.10434	0.11018	-0.947	0.34366	
conjugacao3a	-0.87399	0.18930	-4.617	<0.001	***

glmer(ndo~sexogen+fai.eta+Escolaridade+indice+mobilidade+origem.pais+regiao+tempoaux+tamanhoaux+conjugacao+(1|falante)+(1|ocor), data=gp, family = binomial)

Estimativas (em logodds) de valores de significância para Perífrases com todas as variáveis e Classe Social (N = 2940)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-0.87843	1.57436	-0.558	0.5769	
sexogenM	0.94850	0.39039	2.430	0.0151	*
fai.eta2a	-0.82431	0.45634	-1.806	0.0709	.
fai.eta3a	-2.67943	0.50823	-5.272	<0.001	***
EscolaridadeS	0.28052	0.46992	0.597	0.5505	
classesocialB1	-1.38049	0.78226	-1.765	0.0776	.
classesocialB2	-0.08774	0.62526	-0.140	0.8884	
classesocialC1	0.43149	0.75120	0.574	0.5657	
classesocialC2	1.15271	0.79371	1.452	0.1464	
mobilidadeM	0.06891	0.46755	0.147	0.8828	
mobilidadeZ	0.65426	0.59658	1.097	0.2728	
origem.pais	-0.16881	0.17455	-0.967	0.3335	
regiaoP	0.51970	0.47917	1.085	0.2781	
regiaoV	-0.30675	0.56264	-0.545	0.5856	
tempoauxN	-0.33408	1.25451	-0.266	0.7900	
tempoauxPA	-0.32726	1.24588	-0.263	0.7928	
tempoauxPR	-0.51212	1.24345	-0.412	0.6804	
tamanhoauxR	-0.08557	0.12077	-0.709	0.4786	
conjugacao2a	-0.09028	0.10976	-0.823	0.4108	
conjugacao3a	-0.88767	0.18896	-4.698	<0.001	***

glmer(ndo~sexogen+fai.eta+Escolaridade+classesocial+mobilidade+origem.pais+regiao+tempoaux+tamanhoaux+conjugacao+(1|falante)+(1|ocor), data=gp, family = binomial)

Orações Reduzidas

Estimativas (em logodds) de valores de significância para Orações Reduzidas com todas as variáveis e Índice Socioeconômico (N = 1077)

	Estimativa	Erro padrão	Valor-z	Significância (p)
(Intercept)	0.06891	0.46755	0.147	0.8828
sexogenM	0.65426	0.59658	1.097	0.2728

fai.eta2a	-0.16881	0.17455	-0.967	0.3335
fai.eta3a	0.51970	0.47917	1.085	0.2781
EscolaridadeS	-0.30675	0.56264	-0.545	0.5856
indice2,2	-0.33408	1.25451	-0.266	0.7900
indice2,3	-0.32726	1.24588	-0.263	0.7928
indice2,4	-0.51212	1.24345	-0.412	0.6804
indice2,5	-0.08557	0.12077	-0.709	0.4786
indice2,6	-0.09028	0.10976	-0.823	0.4108
indice2,7	-0.88767	0.18896	-4.698	<0.001
indice2,8	0.06891	0.46755	0.147	0.8828
indice2,9	0.65426	0.59658	1.097	0.2728
indice3	-0.16881	0.17455	-0.967	0.3335
indice3,1	0.51970	0.47917	1.085	0.2781
indice3,2	0.185040	0.528321	0.350	0.7262
indice3,3	-1.438854	0.823010	-1.748	0.0804
indice3,4	-0.818386	0.678628	-1.206	0.2278
indice3,5	0.377538	0.800071	0.472	0.6370
indice3,6	0.915374	0.828238	1.105	0.2691
indice3,7	0.234719	0.482747	0.486	0.6268
indice3,8	0.185040	0.528321	0.350	0.7262
indice3,9	-1.438854	0.823010	-1.748	0.0804
indice4	-0.818386	0.678628	-1.206	0.2278
indice4,1	0.377538	0.800071	0.472	0.6370
indice4,2	0.915374	0.828238	1.105	0.2691
indice4,3	-0.16539	1.20380	-0.137	0.89072
indice4,4	-0.97337	1.51068	-0.644	0.51936
indice4,5	0.24856	1.19167	0.209	0.83478
mobilidadeM	-1.13143	1.13174	-1.000	0.31744
mobilidadeZ	-0.76595	1.14936	-0.666	0.50514
origem.pais	0.04746	1.16905	0.041	0.96762
regiaoP	-16.62093	1051.10624	-0.016	0.98738
regiaoV	-1.20044	1.67290	-0.718	0.47302
conjugacao2a	-0.50926	1.11935	-0.455	0.64914
conjugacao3a	-0.69156	1.22028	-0.567	0.57091

glmer(ndo~sexogen+fai.eta+Escolaridade+indice+mobilidade+origem.pais+regiao+conjugacao+(1|falante)+(1|ocor), data=gr, family = binomial)

Estimativas (em logodds) de valores de significância para Orações Reduzidas com todas as variáveis e Classe Social (N = 1077)

	Estimativa	Erro padrão	Valor-z	Significância (p)	
(Intercept)	-2.056352	1.011390	-2.033	0.0420	*
sexogenM	1.030030	0.417821	2.465	0.0137	*
fai.eta2a	-0.429514	0.460657	-0.932	0.3511	
fai.eta3a	-3.142463	0.617697	-5.087	<0.001	***
EscolaridadeS	0.185040	0.528321	0.350	0.7262	
classesocialB1	-1.438854	0.823010	-1.748	0.0804	.
classesocialB2	-0.818386	0.678628	-1.206	0.2278	

classesocialC1	0.377538	0.800071	0.472	0.6370	
classesocialC2	0.915374	0.828238	1.105	0.2691	
mobilidadeM	0.234719	0.482747	0.486	0.6268	
mobilidadeZ	1.399589	0.632212	2.214	0.0268	*
origem.pais	0.053425	0.179196	0.298	0.7656	
regiaoP	0.901351	0.547874	1.645	0.0999	.
regiaoV	-0.003514	0.629721	-0.006	0.9955	
conjugacao2a	-0.628793	0.262901	-2.392	0.0168	*
conjugacao3a	-0.991605	0.408009	-2.430	0.0151	*

glmer(ndo~sexogen+fai.eta+Escolaridade+classesocial+mobilidade+origem.pais+regiao+conjugacao+(1|falante)+(1|ocor), data=gr, family = binomial)