

MILENA APARECIDA DE ALMEIDA

A SOCIOLINGUÍSTICA DAS ESTRATÉGIAS DE RELATIVIZAÇÃO EM CONTEXTOS LOCATIVOS NA FALA PAULISTA



MILENA APARECIDA DE ALMEIDA

A SOCIOLINGUÍSTICA DAS ESTRATÉGIAS DE RELATIVIZAÇÃO EM CONTEXTOS LOCATIVOS NA FALA PAULISTA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Análise Fonológica, Morfossintática, Semântica e Pragmática

Orientadora: Prof.^a Dr.^a Rosane de Andrade Berlinck

Co-orientador: Prof. PhD Stephen Levey

Bolsa: FAPESP (Processo nº 2020/00593-5)

ARARAQUARA – S.P.
2022

A447s	Almeida, Milena Aparecida de A sociolinguística das estratégias de relativização em contextos locativos na fala paulista / Milena Aparecida de Almeida. -- Araraquara, 2022 132 f. : tabs. Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Letras, Araraquara Orientadora: Rosane de Andrade Berlinck Coorientador: Stephen Levey 1. Sociolinguística. 2. Sintaxe. 3. Estratégias de relativização. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Letras, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

MILENA APARECIDA DE ALMEIDA

A SOCIOLINGUÍSTICA DAS ESTRATÉGIAS DE RELATIVIZAÇÃO EM CONTEXTOS LOCATIVOS NA FALA PAULISTA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Análise Fonológica, Morfossintática, Semântica e Pragmática

Orientadora: Prof.^a Dr.^a Rosane de Andrade Berlinck

Co-orientador: Prof. PhD Stephen Levey

Bolsa: FAPESP (Processo nº 2020/00593-5)

Data da defesa: 11/02/2022

MEMBROS COMPONENTES DA BANCA EXAMINADORA:

Presidente e Orientador: Prof.^a Dr.^a Rosane de Andrade Berlinck

Universidade Estadual Paulista – UNESP / FCLAR

Membro Titular:

Prof.^a Dr.^a Livia Oushiro

Universidade de Campinas – UNICAMP/ IEL

Membro Titular:

Prof.^a Dr.^a Edair Görski

Universidade Federal de Santa Catarina - UFSC

Local: Universidade Estadual Paulista
Faculdade de Ciências e Letras
UNESP – Campus de Araraquara

Àqueles que me ensinaram sobre afeto incondicional e disciplina,
Marlene e Vandecir

AGRADECIMENTOS

Dedicar essa dissertação a minha família parece simplório mediante todo o esforço mobilizado para que esse momento fosse possível. Dessa forma, tento expressar toda a minha gratidão nessas linhas aos meus pais, Marlene e Vandecir, minhas inspirações diárias nessa batalha, por nunca duvidarem do meu potencial e acreditarem fielmente que a filha de um casal “da roça”, de uma cidade pequena, com poucas chances de alfabetização, pudesse um dia receber um título de Mestre. Não menos importante, agradeço a minha irmã, Andresa, por me compreender e me apoiar em tantos momentos difíceis.

À minha orientadora, Rosane, a quem constantemente estou testando a paciência com inúmeros projetos e entregas em cima do prazo, eu sou eternamente grata. Nosso caminho se cruzou em 2016, quando era apenas uma caloura ansiosa demais, e desde então você me deu todas as ferramentas, incentivos e liberdade para que eu trilhasse meu próprio caminho dentro da academia. Obrigada pela confiança no meu trabalho, pelos conselhos constantes e por saber dosar minha expectativa com a realidade.

Ao meu coorientador, Stephen Levey, um dos pesquisadores mais atenciosos que encontrei em toda a minha trajetória, agradeço por todo o apoio e encorajamento nesses dois anos.

Minha jornada acadêmica, até o presente momento, não teria sido tão incrível e divertida se não fosse por essas duas figuras: Marcus e Letícia. Ao meu mentor, guia espiritual e companheiro de vida, Marcus, sou extremamente grata. Nunca conseguirei colocar em palavras como eu sou agradecida ao universo pela nossa amizade, por todo o tempo que você demanda apenas para entender minha pesquisa, me auxiliar das mais diversas formas possíveis, nem que para isso seja necessário deixar de lado as redações (e o dinheiro) para assistir vídeos de *dragqueen* comigo. Eu te amo e quero reconhecer publicamente que metade dessa dissertação não existiria sem a sua brilhante mente, tão desesperadoramente incrível em produzir um diálogo com uma perspectiva que você não domina! A sua radiação faz meus dias muito mais felizes, mesmo queimando meus bolos de consolação.

A minha parceira de dificuldades, Letícia, eu gostaria de agradecer e te entregar o mundo! Uma pessoa extremamente perturbada e ansiosa, que de alguma forma se assemelha muito a minha própria perturbação e ansiedade. Dividir essa jornada, dia após dia, com você, é o que torna esse caminho mais leve. Inclusive dentro de uma pandemia, não podendo de ver e ouvir pessoalmente desde o dia que nos intitulamos mestrandas. Temos muitas comemorações, tristezas e inseguranças para compartilhar, e sei que nosso caminho ainda é longo. Obrigada por tudo, miga!

Sou extremamente grata ao meu fã número 1, Raphael, a pessoa mais incrível e especial que encontrei no meio de um caminho muito turbulento. Muito obrigada por ser tão paciente, por querer desesperadamente entender meu cotidiano tão diferente do seu, por me ensinar a dosar minha ansiedade, expectativa e prioridades, e por me apoiar incondicionalmente em qualquer decisão profissional, mesmo que isso impacte seu próprio dia a dia. Essa dissertação tem um pedacinho seu, assim como você tem um pedacinho do meu coração. Te amo!

Agradeço aos meus parceiros roraimenses, Eliabe e Fabrício, por compartilharem tanto conhecimento, por acreditarem no meu potencial e por me animarem nos momentos difíceis

com fofocas e histórias absurdas. Apesar de não dividirmos açaí com tanta constância, meu amor e carinho por vocês não diminuiu, apenas multiplicou. Obrigada pela paciência!

À minha banca de qualificação e defesa, Lívia Oushiro, Roberto Camacho e Edair Görski, meu muito obrigada por todas as colocações que tornaram meu trabalho muito mais coeso! Em especial, toda a minha gratidão à Lívia Oushiro, que tem me assistido e acompanhado desde a graduação e está sempre disponível para as dúvidas sobre o R.

Aos meus amigos do Núcleo de Pesquisa em Sociolinguística de Araraquara, agradeço imensamente por toda a escuta carinhosa e as críticas produtivas que compartilhamos no decorrer dos últimos anos!

Por fim, agradeço à FAPESP (Processo nº 2020/00593-5) e à CAPES pelo apoio financeiro, fundamental para o desenvolvimento desta pesquisa.

“Vamos atentar para a sintaxe dos paulistas...”

(Língua – Caetano Veloso)

RESUMO

No contexto dos estudos sobre relativização em português, há poucos estudos específicos sobre o contexto locativo, que investiguem a variação entre as estratégias possíveis. Por outro lado, estudos recentes identificaram um uso não locativo de **onde**, apontando para um processo de gramaticalização do item na fala de diferentes regiões (SOUZA, 2003; BRAGA, MANFILI, 2004; ALMEIDA, 2020). Tendo em conta esse cenário, o objetivo do presente estudo é investigar o processo de variação linguística de construções relativas no domínio de origem do pronome relativo **onde**, isto é, na expressão do valor locativo na fala paulista. Essa caracterização é uma primeira etapa de investigação, que nos levará, em estudo posterior, a avaliar a possível correlação entre a variação no domínio de origem e os processos de multifuncionalidade de **onde**. Temos como envelope de variação as estratégias de relativização do Português Brasileiro (AMARAL, 1920; MOLLICA, 1977; TARALLO, 1983; CORRÊA, 1998; BISPO, 2003; 2009) – isto é, **padrão (com e sem preposição)**, **copiadora** e **cortadora**. Adotamos, assim, a abordagem teórico-metodológica da Teoria de Variação e Mudança Linguística (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1972) para análise do *corpus* construído tendo por base os bancos de dados IBORUNA, do projeto ALIP (Gonçalves, s.d) e Projeto SP2010: Amostra da fala paulistana (MENDES, 2013) contabilizando 1007 relativas. Estas estão distribuídas em: 3 dados de **padrão com preposição**, 35 dados de **copiadora**, 249 dados de **cortadora** e 267 dados de **padrão com onde** provenientes do banco de dados IBORUNA; e 10 dados de **padrão com preposição**, 15 dados de **copiadora**, 128 dados de **padrão com onde** e 300 dados de **cortadora** provenientes do banco de dados Projeto SP2010. Por conta da baixa frequência em nosso *corpus*, as estratégias **padrão com preposição** e **copiadora** receberam um tratamento qualitativo, enquanto uma análise quantitativa foi produzida para compreender o comportamento das estratégias **padrão com onde** e **cortadora**. Os fatores linguísticos que norteiam a pesquisa são: função sintática do pronome relativo; distância entre o SN e a oração relativa; classificação da relativa; identidade e tipo do locativo; definitude, especificidade e status informacional do SN antecedente. Os fatores extralinguísticos controlados são: sexo/gênero, escolaridade e faixa etária do falante (presentes nos dois bancos); tipologia textual (presente no IBORUNA); região e zona de residência do falante em São Paulo e roteiro da entrevista (presentes no SP2010). Em nossa análise, averiguamos, por meio de uma regressão logística de efeitos mistos, que há um desfavorecimento do uso da **padrão com onde** para falantes da primeira faixa etária, com níveis de escolaridade mais baixos e quando o antecedente se refere a um local, para o banco de dados IBORUNA, e um desfavorecimento do uso da **padrão com onde** para falantes da segunda e terceira faixa etária com ensino médio, moradores da região centro velho e em contextos indefinidos, para o banco de dados Projeto SP2010. Concluimos que a descrição desse contexto específico de relativização revela padrões diferentes em relação às tendências gerais (de predomínio de estratégias não padrão), já que a **padrão com onde** está ainda bem presente. Além disso, as variáveis sociais se mostraram muito significativas na explicação da variação, algo que contribui com uma discussão dentro do modelo teórico sobre o papel dessas variáveis em fenômenos morfológicos e sintáticos, somando-se a alguns estudos que mostram que pode, sim, haver influência desses aspectos em níveis mais altos. Ou seja, que é possível falar em variáveis sociolinguísticas (propriamente ditas) nesses níveis. Por fim, considerando que a relativização é um mecanismo presente em muitas línguas, para o qual se busca identificar princípios gerais, o estudo pode contribuir para essa discussão em âmbito interlinguístico.

Palavras – chave: pronome relativo onde; estratégias de relativização; português paulista; variável sintática.

ABSTRACT

In the context of studies on relativization in Portuguese, there are few specific studies on the locative context that investigate the variation between possible strategies. On the other hand, recent studies have identified a non-locative use of **onde**, pointing to a process of grammaticalization of the item in the speech of different regions (SOUZA, 2003; BRAGA, MANFILI, 2004; ALMEIDA, 2020). Taking this scenario into account, the objective of the present study is to investigate the process of linguistic variation of relative constructions in the domain of origin of the relative pronoun **onde**, that is, in the expression of the locative value in São Paulo speech. This characterization is a first stage of investigation, which will lead us, in a later study, to evaluate the possible correlation between the variation in the domain of origin and the multifunctionality processes from **onde**. Our envelope of variation includes the relativization strategies used in Brazilian Portuguese (AMARAL, 1920; MOLLICA, 1977; TARALLO, 1983; CORRÊA, 1998; BISPO, 2003; 2009) – that is, **standard (with or without preposition)**, **resumptive pronoun** and **pp-chopping**. We adopted the theoretical-methodological approach of the Theory of Language Variation and Change (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1972) to analyze the corpus built from the IBORUNA database (ALIP project (Gonçalves, n.d.)) and from Projeto SP2010: Amostra da fala paulistana (MENDES, 2013). The corpus counts 1007 tokens of relatives. These are distributed as: 3 occurrences of **standard with preposition**, 35 occurrences of **resumptive pronoun**, 249 occurrences of **pp-chopping** and 267 occurrences **standard with onde** from the IBORUNA database; and 10 occurrences of **standard with preposition**, 15 occurrences of **resumptive pronoun**, 128 occurrences of **standard with onde** and 300 data of **pp-chopping** from the SP2010 database. Due to the low frequency in our corpus, the **standard with preposition** and **resumptive pronoun** strategies received a qualitative treatment, while a quantitative analysis is being produced to understand the behavior of the **standard with onde** and **pp-chopping** strategies. The linguistic factors that guide the research are: syntactic function of the relative pronoun; distance between the NP and the relative clause; relative classification; identity and type of locative; definiteness, specificity and informational status of the antecedent NP. The controlled extralinguistic factors are: sex/gender, education and age group of the speaker (present in both database); textual type (present in IBORUNA); speaker's region and area of residence in São Paulo and interview script (present in SP2010). Our statistical analyses verified, through a mixed effects logistic regression, that the use of **standard with onde** is disfavored by speakers of the first age group, with lower levels of education and when the antecedent refers to a location, for the IBORUNA database, and the use of **standard with onde** is disfavored by speakers of the second and third age groups with high school, downtown residents and in undefined contexts, for SP2010. We conclude that the description of this specific context of relativization reveals different patterns in relation to the general trends (prevalence of non-standard strategies), since the **pattern with onde** is still very present. In addition, social variables proved to be very significant in explaining variation, something that contributes to a discussion within the theoretical model about the role of these variables in morphological and syntactic phenomena, in addition to some studies that show that there may, indeed, be influence of these aspects at higher levels. In other words, it is possible to speak of sociolinguistic variables (properly so-called) at these levels. Finally, considering that relativization is a mechanism present in many languages, for which it seeks to identify general principles, the study can contribute to this discussion in an interlinguistic scope

Keywords: relative pronoun **onde**; relativization strategies; Paulista Portuguese; syntactic variable.

LISTA DE FIGURAS

Figura 1	Fórmula da subordinação adjetiva no Português Brasileiro	29
Figura 2	A taxonomia dos valores da Familiaridade Assumida	64
Figura 3	Divisão da cidade de São Paulo entre regiões mais centrais e mais periféricas	72
Figura 4	Proporção de uso das estratégias de relativização no banco de dados IBORUNA	75
Figura 5	Proporção de uso das estratégias de relativização no banco de dados Projeto SP2010	76
Figura 6	Nuvem de palavras da identidade lexical do verbo na oração relativa no banco de dados IBORUNA e Projeto SP2010, respectivamente	78
Figura 7	Proporção de uso das estratégias de relativização segundo a faixa etária do falante no banco de dados IBORUNA	81
Figura 8	Proporção de uso das estratégias de relativização segundo a faixa etária do falante no banco de dados SP2010	82
Figura 9	Proporção e frequência das estratégias de relativização segundo a escolaridade do falante no banco de dados IBORUNA	84
Figura 10	Proporção e frequência das estratégias de relativização segundo a escolaridade do falante no banco de dados SP2010	84
Figura 11	Árvore de classificação condicional para o uso das estratégias de relativização segundo a faixa etária e escolaridade do falante no banco de dados IBORUNA	88
Figura 12	Árvore de classificação condicional para o uso das estratégias de relativização segundo a faixa etária e escolaridade do falante no banco de dados SP2010	88
Figura 13	Proporção de uso das estratégias de relativização segundo a região de residência do falante	90
Figura 14	Proporção de uso das estratégias de relativização segundo a zona de residência do falante no banco de dados SP2010	91
Figura 15	Proporção de uso das estratégias de relativização segundo a região de residência e a zona de residência do falante no banco de dados SP2010	92

Figura 16	Proporção de uso das estratégias de relativização segundo a identidade semântica do antecedente no banco de dados IBORUNA	93
Figura 17	Proporção de uso das estratégias de relativização segundo a identidade semântica do antecedente no banco de dados SP2010	94
Figura 18	Proporção de uso das estratégias de relativização segundo a definitude do antecedente no banco de dados Projeto SP2010	95
Figura 19	Proporção de uso das estratégias de relativização segundo a definitude e a especificidade do antecedente no banco de dados SP2010	96
Figura 20	Proporção de uso das estratégias de relativização segundo a definitude e o status informacional do antecedente no banco de dados SP2010	98
Figura 21	Proporção de uso das estratégias de relativização segundo a definitude do antecedente no banco de dados IBORUNA	99
Figura 22	Proporção de uso das estratégias de relativização segundo a definitude e a especificidade do antecedente no banco de dados IBORUNA	100
Figura 23	Proporção de uso das estratégias de relativização segundo o status informacional do antecedente no banco de dados IBORUNA	102
Figura 24	Proporção de uso das estratégias de relativização segundo a adjacência entre o sintagma nominal antecedente e a oração relativa no banco de dados IBORUNA	104
Figura 25	Proporção de uso das estratégias de relativização segundo a adjacência entre o sintagma nominal antecedente e a oração relativa no banco de dados SP2010	105
Figura 26	Frequência e proporção de uso das estratégias de relativização segundo a classificação semântica da relativa no banco de dados IBORUNA	106
Figura 27	Frequência e proporção de uso das estratégias de relativização segundo a classificação semântica da relativa no banco de dados SP2010	106
Figura 28	Frequência e proporção de uso das estratégias de relativização segundo classe morfológica do locativo no banco de dados IBORUNA	108
Figura 29	Frequência e proporção de uso das estratégias de relativização segundo classe morfológica do locativo no banco de dados SP2010	108

Figura 30	Proporção de uso das estratégias de relativização segundo a função sintática do pronome relativo no banco de dados IBORUNA	109
Figura 31	Proporção de uso das estratégias de relativização segundo a função sintática do pronome relativo no banco de dados SP2010	110
Figura 32	Frequência e proporção de uso das estratégias de relativização segundo o sexo do falante no banco de dados IBORUNA	112
Figura 33	Frequência e proporção de uso das estratégias de relativização segundo o sexo do falante no banco de dados SP2010	112
Figura 34	Frequência e proporção de uso das estratégias de relativização segundo o tipo textual no banco de dados IBORUNA	113
Figura 35	Frequência e proporção de uso das estratégias de relativização segundo o roteiro de entrevista no banco de dados Projeto SP2010	114
Figura 36	<i>Continuum</i> de marcação (givoniana) aplicado às estratégias de relativização	116
Figura 37	Proporção de uso das estratégias de relativização segundo a faixa etária e a escolaridade do falante no banco de dados IBORUNA	119
Figura 38	Proporção de uso das estratégias de relativização segundo a faixa etária e a escolaridade do falante no banco de dados SP2010	120

LISTA DE TABELAS

Tabela 1	Descrição do envelope de variação	20
Tabela 2	Frequência e proporção de uso de onde “não locativo” segundo escolaridade no bando de dados IBORUNA	26
Tabela 3	Frequência e proporção de uso de onde “não locativo” segundo a faixa etária do falante no bando de dados IBORUNA	27
Tabela 4	Distribuição da proporção de uso das estratégias de relativização de acordo com estratégia em 5 funções sintáticas	36
Tabela 5	Porcentagem de uso das estratégias copiadora, cortadora e <i>piedpiping</i> nas três posições sintáticas mais baixas para as três classes sociais	36
Tabela 6	Distribuição da proporção de uso das estratégias de relativização <i>piedpiping</i> , copiadora e cortadora através do tempo	37
Tabela 7	Relativas no corpus <i>D&G Rio de Janeiro</i> e <i>D&G Natal</i> por estratégia e nível de escolaridade	39
Tabela 8	Distribuição das estratégias de relativização em ambientes preposicionados em Barros (2000)	41
Tabela 9	Distribuição das estratégias de relativização segundo a função sintática do pronome relativo em Burgos (2003)	42
Tabela 10	<i>Coding protocol</i> utilizado para as variáveis linguísticas	53
Tabela 11	Padrões utilizados na análise da definitude do antecedente	62
Tabela 12	<i>Coding protocol</i> utilizado para as variáveis extralinguísticas	68
Tabela 13	Análise de regressão de efeitos mistos da estratégia de relativização padrão com onde no banco de dados IBORUNA	79
Tabela 14	Análise de regressão de efeitos mistos da estratégia de relativização padrão com onde no banco de dados Projeto SP2010	80
Tabela 15	Frequência e proporção de uso das estratégias de relativização segundo a adjacência entre o sintagma nominal antecedente e a oração relativa no banco de dados IBORUNA	116
Tabela 16	Frequência e proporção de uso das estratégias de relativização segundo o a adjacência entre o sintagma nominal antecedente e a oração relativa no banco de dados SP2010	117

Tabela 17	Frequência e proporção de uso das estratégias de relativização segundo a faixa etária do falante no banco de dados IBORUNA	118
Tabela 18	Frequência e proporção de uso das estratégias de relativização segundo a faixa etária do falante no banco de dados SP2010	118

SUMÁRIO

1 INTRODUÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 POR ONDE ANDA VOCÊ? UMA BREVE CONTEXTUALIZAÇÃO	22
2.1.1 Uma perspectiva interlinguística	22
2.1.2 Uma investigação paulista	24
2.2 O OBJETO DE ESTUDO	28
2.2.1 Tradição normativa	30
2.2.2 Tradição descritiva	32
2.2.3 As estratégias de relativização no Português	33
2.3 PRESSUPOSTOS TEÓRICO-METODOLÓGICOS	42
2.3.1 A Teoria de Variação e Mudança Linguística	42
2.3.1.1 A problemática da variável sintática	45
3 DECISÕES TEÓRICO-METODOLÓGICAS	48
3.1 VARIÁVEL DEPENDENTE	50
3.2 VARIÁVEIS LINGUÍSTICAS	51
3.2.1 Função sintática do pronome relativo	53
3.2.2 Adjacência entre o sintagma nominal antecedente e a oração relativa	55
3.2.3 Classe morfológica do locativo	57
3.2.4 Identidade semântica (ou tipo) do locativo	57
3.2.5 Classificação semântica da relativa	58
3.2.6 Definitude, especificidade e status informacional do antecedente	59
3.3 VARIÁVEIS EXTRALINGUÍSTICAS	66
3.3.1 Sexo/gênero	67
3.3.2 Escolaridade	68
3.3.3 Faixa etária	69
3.3.4 Região e zona de residência de São Paulo	70
3.3.5 Tipologia textual	72
3.3.6 Roteiro da entrevista	72

4 A SINTAXE DOS PAULISTAS: ANÁLISE DE PRODUÇÃO	73
4.1 DISTRIBUIÇÃO DE DADOS DAS VARIÁVEIS CORRELACIONADAS	75
4.1.1 Síntese	100
4.2 DISTRIBUIÇÃO DE DADOS DAS VARIÁVEIS NÃO CORRELAIONADAS	102
4.2.1 Adjacência entre o sintagma nominal antecedente e a oração relativa	102
4.2.2 Classificação semântica da relativa	104
4.2.3 Classe morfológica do locativo	106
4.2.4 Função sintática do pronome relativo	107
4.2.5 Sexo/gênero do falante	110
4.2.6 Tipologia Textual	112
4.2.7 Roteiro de entrevista	113
4.3 O CASO DA COPIADORA E PADRÃO COM PREPOSIÇÃO	114
5 CONSIDERAÇÕES FINAIS	121
REFERÊNCIAS	125

1 INTRODUÇÃO

Lehmann (2002), a partir de uma abordagem funcionalista da linguagem, retoma uma proposta do âmbito dos Estudos do Léxico, utilizada desde o século XIX, com o objetivo de estender tais noções para outros níveis linguísticos de análise. Assim, as formas linguísticas – em processo de competição ou não – podem ser investigadas a partir de duas perspectivas: a semasiológica e a onomasiológica. A primeira tem por base uma linha de estudos que parte de uma determinada forma (seja ela uma palavra, uma expressão ou uma construção) para o mapeamento de seus sentidos (ou funções); enquanto a segunda parte de um conceito (ou função) para identificar que formas o expressam na língua subjacente às construções (ou formas) investigadas, descrevendo como é expresso na língua alvo.

Essa distinção se mostra muito produtiva para o estudo de fenômenos de variação em níveis gramaticais mais altos dentro da Teoria de Variação e Mudança Linguística (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1972, 1982, 1994, 2001, 2010). Isso, pois, apesar de assumir que a variação é constitutiva da língua, podendo afetar qualquer nível linguístico, há uma problemática à teoria no que diz respeito aos fenômenos linguísticos acima do nível fonético-fonológico (LAVANDERA, 1978; LABOV, 1978; MILROY, GORDON, 2003; FREITAG, 2009). Diante dessa questão, a proposta de análise de Lehmann (2002) contribui para sanar um dos problemas encontrados¹, norteados os caminhos que o pesquisador pode percorrer para investigar uma ou mais formas linguísticas de nível gramatical mais alto.

Um exemplo de trabalho morfossintático que toma essa distinção é a investigação da expressão do *antepresente* e do *passado absoluto* no espanhol de Madri, Buenos Aires e San Miguel de Tucumán, de Araújo (2017). O autor buscou identificar como esses dois valores se materializam, de forma variável, por meio das formas verbais do *pretérito perfecto compuesto* (como em *he escrito*) e do *pretérito perfecto simple* (como em *escribí*), o que caracteriza uma perspectiva onomasiológica. Para chegar a essa análise, Araújo se vale da descrição semasiológica que foi feita anteriormente sobre os dois tempos verbais (como, por exemplo, GILI GAYA (1970); RAE (1986) e TORREGO (2002)), tendo em vista uma melhor compreensão do fenômeno em questão.

Entre várias possíveis aplicações, a conjugação das duas perspectivas indicadas por Lehmann (2002) se revela muito produtiva para o estudo da forma e função do pronome relativo

¹ A discussão completa se encontra na seção 2.3.1.1: *A problemática da variável sintática*

onde no Português Brasileiro, nosso objeto de estudo. O modo como os falantes têm-se utilizado do pronome para cumprir determinadas funções comunicativas que fogem da prescrição gramatical – isto é, o valor locativo – tem despertado o interesse de vários pesquisadores (BRAGA, MANFILI, 2004; LIMA, COROA, 2013; LIMA, 2007; SOUZA, 2003; ALMEIDA, BERLINCK, 2019; ALMEIDA, 2020). Alguns desses usos podem ser observados nos seguintes exemplos:

- (1.1) “terríveis... terrível **os últimos dois anos onde eu passei a desconfiá(r) que ele tinha...** caso com uma colega de trabalho” (AC-101; L. 58-61)
- (1.2) “o povo não tem votado direito... e... o país os municípios os estados... não têm sido bem sucedido... em algumas eleições... vide:: **a eleição do... Fernando Collor... onde ele pregô(u) tanto...** e depois foi... deu no que deu... ele o povo brasile(i)ro naquela... esperança da salvação que o povo vive até hoje...” (AC-113; RO: L. 218-222)

Em (1.1) o pronome relativo **onde** é usado na construção para uma referência ao sintagma “os últimos dois anos”, ou seja, a referência por ele efetuada é temporal, e não estritamente espacial. Em (1.2) o pronome adquire um valor diferente, já que, ao efetuar o processo de referência ao sintagma anterior, ele não produz o efeito locativo prescrito ao introduzir um evento, isto é, referente “a eleição do Fernando Collor”.

Adotando uma perspectiva semasiológica² (ALMEIDA, 2019), investigamos as funções que o pronome tem exercido na língua, considerando-as como parte de um possível processo de gramaticalização, a partir de um *corpus* composto exclusivamente da fala paulista de duas comunidades – uma do interior de São Paulo e outra da capital. Os usos não locativos de **onde** encontrados, tal qual (1.1) e (1.2), corroboram a literatura sobre o fenômeno e comprovam sua multifuncionalidade. Em especial, é necessário salientar que as ocorrências mais frequentes foram, justamente, da categoria semântico-cognitiva mais gramaticalizada, denominada por nós como “texto”.

- (1.3) às vez a mãe bebe demais o pai bebe demais e vai querê(r) í(r) falá(r) po filho – “num pode” –?... tipo uma que ele vai falá(r) – “por que que você pode e eu não posso?” – aí vai começá(r) a fazê(r) escondido... **e é onde o mundo tá do jeito que TÁ...** (AC-024; RO: L. 418-420)

Em (1.3), o falante parte de uma dada situação, descrevendo-a para que seu interlocutor compreenda a argumentação e conclui introduzindo uma oração encabeçada com **onde**. Não há

² Descrita em mais detalhes na seção 2: *Por onde anda você? Uma breve contextualização.*

um antecedente explícito, um núcleo ao qual uma oração relativa se liga; a articulação se dá de forma mais frouxa, com função semelhante ao de uma conjunção em construções adverbiais, expressando relações de causa, condição e tempo. O que observamos nesse tipo de ocorrência é um uso de **onde** que apresenta um caráter sintático e semântico diferente daquele considerado normativo, uma vez que não se configura mais como um pronome relativo, aproximando-se mais de um marcador discursivo, ou uma conjunção, com valores específicos na construção do discurso, tais quais como operador conclusivo, operador consecutivo e operador explicativo (MELO, BARBOSA, 2018).

Assim, como consequência da diversificação das categorias semântico-cognitivas que se associam ao uso do pronome, é de nosso interesse ampliar a investigação, enfatizando, neste momento, uma perspectiva onomasiológica. Após mapear os usos de **onde** na fala paulista (ALMEIDA, BERLINCK, 2019; ALMEIDA, 2019; ALMEIDA, 2020) e identificar um quadro de multifuncionalidade do pronome, o que aponta para uma possível mudança linguística associada ao item, devido ao processo de gramaticalização averiguado, temos o intento de observar a variação linguística no domínio de origem, isto é, no valor locativo. Desse modo, nosso estudo tem como objetivo investigar o processo de variação linguística das estratégias de relativização preposicionadas que expressam valor locativo na fala paulista.

Tal abordagem do fenômeno se justifica por meio do problema do encaixamento proposto por Weinreich, Labov e Herzog (2006 [1968], p. 132), haja vista que

A mudança linguística, ela mesma, raramente é um movimento de um sistema inteiro para outro. Em vez disso, descobrimos que um conjunto limitado de variáveis num sistema altera seus valores modais gradualmente de um polo para outro. As variantes das variáveis podem ser *contínuas* ou *discretas*; em qualquer dos casos, a variável mesma tem um espectro contínuo de valores, já que ele inclui a frequência de ocorrência de variantes individuais na fala estendida. O conceito da variável como um elemento estrutural torna desnecessário ver flutuações no uso como externas ao sistema, pois o controle de tal variação faz parte da competência linguística dos membros da comunidade de fala.

Dessa forma, argumentamos que a descrição da variação no domínio de origem de **onde** é um importante passo dentro da investigação maior dos usos do pronome para melhor compreensão da alteração que está em desenvolvimento dentro do sistema. Isso, pois nos permite identificar se há uma influência da multifuncionalidade, no sentido de abrir espaço para outras estratégias assumirem maior frequência no valor locativo, pensando nos valores modais que gradualmente se deslocam de um polo a outro. Nossa hipótese é de que ainda não nos encontramos nesse estágio da variação, dado que o valor locativo ainda é prototípico de **onde**. Contudo, temos duas razões para o presente estudo: (i) a próxima fase do trabalho tem por base

a diacronia, o que torna relevante analisar um corpus sincrônico para articular uma argumentação bem fundamentada a respeito de um possível processo de mudança; (ii) não há estudos que foquem na descrição do contexto sintático-semântico aqui especificado, o que nos impede de fazer uma afirmação categórica a respeito dos processos de variação e mudança e torna necessária a descrição.

Por isso, com base na bibliografia recente sobre as estratégias de relativização no Português Brasileiro (AMARAL, 1920; MOLLICA, 1977; LEMLE, 1978; TARALLO, 1983; CORRÊA, 1998; BARROS, 2000; BURGOS, 2003; BISPO, 2003; 2009), intencionamos averiguar se, há um espaço propício para a variação de formas de expressão do valor prototípico de lugar. Ademais, temos por objetivos:

- i. Identificar as estratégias de relativização que funcionam como variantes na expressão do valor locativo;
- ii. Determinar quais fatores linguísticos e extralinguísticos incidem sobre os usos dessas estratégias de relativização variantes;
- iii. Analisar em que medida a situação de variação no contexto locativo se articula com a multifuncionalidade de **onde** na expressão de outras categorias semântico-cognitivas;

Para que isso seja possível, nosso envelope de variação se constrói a partir das estratégias de relativização locativas do Português Brasileiro em contextos preposicionados, sendo elas:

Tabela 1 – Descrição do envelope de variação³

Estratégia	Definição	Exemplo
Padrão <i>pied-piping</i>	Em contextos preposicionados, a tradição gramatical prescreve o uso de preposição antes do pronome relativo.	Eu vou descrever a casa em que eu estou morando
Padrão com onde	A tradição gramatical ressalva que pronomes como onde não precisam da preposição, mesmo em contextos preposicionados.	Eu vou descrever a casa Ø onde eu estou morando
Copiadora	Caracterizada pelo pronome cópia dentro da oração relativa que retoma o antecedente.	Eu vou descrever a casa que eu estou morando lá/nela
Cortadora	Caracterizada pelo corte e/ou apagamento da preposição antes do pronome relativo.	Eu vou descrever a casa Ø que eu estou morando

Fonte: própria

O presente estudo foi detalhado a partir de 5 seções, sendo a segunda, *Fundamentação teórica*, dividida em três grandes frentes: *Por onde anda você? Uma breve contextualização*, que trata dos principais resultados encontrados na perspectiva semasiológica do mapeamento de **onde**, assim como outros estudos; *O objeto de estudo*, responsável por introduzir as estratégias de relativização com que estamos trabalhando, a partir de duas perspectivas diferentes: (i) uma descritiva, tendo como objetivo destrinchar as considerações da tradição normativa e descritiva a respeito das orações relativas, e (ii) um *overview* linguístico, explorando as análises de trabalhos sobre esse fenômeno no Português Brasileiro; por fim, *Teoria de Variação e Mudança Linguística*, com o intento de tecer considerações a respeito do nosso suporte teórico-metodológico, ressaltando as questões que concernem as variáveis sintáticas. A terceira seção, *Decisões teórico-metodológicas*, apresenta como foi construído o *corpus* utilizado no estudo, assim como considerações relevantes sobre a análise dos dados, desde sua “filtragem” até as variáveis independentes selecionadas. A seção subsequente, *A sintaxe dos paulistas: análise de produção*, se destina à descrição dos resultados de produção; enquanto a sétima seção aponta as considerações finais da investigação.

³ Para o contexto preposicionado, com valor locativo, não encontramos usos de *preposition stranding* ou *pied-piping* (padrão com preposição) encabeçadas por **a qual**, **na qual**. Para mais detalhes do envelope de variação, ver seção 3: *Decisões teórico-metodológicas*.

2 FUNDAMENTAÇÃO TEÓRICA

A presente seção é responsável pela exposição do suporte teórico mobilizado para o estudo, desde a construção do projeto de pesquisa até o refinamento final da análise dos dados. Desse modo, está dividida em três eixos principais: (i) uma breve contextualização do fenômeno impulsionador da investigação, isto é, o pronome relativo **onde**; (ii) descrição das estratégias de relativização, como objeto de estudo; e, por fim, (iii) a Teoria de Variação e Mudança Linguística, nosso suporte teórico-metodológico.

2.1 POR ONDE ANDA VOCÊ? UMA BREVE CONTEXTUALIZAÇÃO

Tal qual brevemente mencionado na seção 1, *Introdução*, o presente estudo se insere em uma investigação maior dos usos de **onde**. Na presente subseção, será descrito, em mais detalhes, o movimento de multifuncionalidade e gramaticalização de pronomes relativos locativos, em uma perspectiva interlinguística, e, posteriormente, alguns aspectos da etapa anterior, desenvolvida durante a Iniciação Científica, e que são importantes para a compreensão dos resultados da fase atual.

2.1.1 Uma perspectiva interlinguística

Há uma literatura consolidada que explora o percurso de gramaticalização de pronomes relativos locativos em diversas línguas (NICHOLAS, 1998; BROOK, 2011; ROUSSOU, 2012; BEAMAN, 2021). Isso aponta para um movimento de itens que atuam em um nível representacional da língua, perdendo o seu valor de referência a objetos, associados a uma realidade biossocial, para assumir usos em nível interpessoal. Esse novo valor que se adquire se caracteriza como funções que atuam no processo de organizar a comunicação dentro de um contexto específico de uso (MARTELOTTA, 2011). Abordar essa questão em uma perspectiva interlinguística nos permite aprofundar as discussões teóricas propostas ao Português Brasileiro e produzir alguns questionamentos sobre o que está por trás desse processo de multifuncionalidade, produtivo em diversas línguas.

Brook (2011) introduz uma construção tipicamente coloquial no inglês de uma oração relativa encabeçada por **where** sem seu valor prototípico locativo. Além da multifuncionalidade associada ao pronome, um interessante movimento que se observa nos exemplos, e na discussão

à qual a autora se propõe, é que nesses contextos não-locativos um pronome lembrete surge na oração encaixada. Autores como Comrie (1999) e Pullum (2008) descrevem as orações relativas encabeçadas por **where** como aquelas que não apresentam lacuna sintática. Contudo, em contextos não locativos, há uma lacuna sendo preenchida, em posições de sujeito e objeto direto, como nas ocorrências em (2.1) e (2.2):

(2.1) There were **some people** at my high school **where they just couldn't write legibly**.

(2.2) *There were **some people** at my high school **where couldn't write legibly**.

Em (2.1) o pronome relativo **where** está introduzindo uma oração encaixada em “some people” (algumas pessoas) que é retomada pelo pronome lembrete “they” (eles). Ao tentar produzir a oração sem o pronome lembrete, “they”, em (2.2), ela se torna agramatical.

Isso nos apresenta uma configuração sintática diferente, mesmo que com um fenômeno de gramaticalização similar àquele que ocorre no Português Brasileiro. Em primeiro lugar, é necessário pontuar que não temos conhecimento de antecedente humano para usos de **onde** não locativo. Em segundo lugar, há usos de copiadora em contextos não locativos, mas isso não é tão produtivo quanto se mostra em inglês.

Nicholas (1998) e Roussou (2012) produzem discussões detalhadas acerca do processo de gramaticalização do pronome *pu* em grego. Vale salientar que o elemento *pu* é responsável por introduzir orações relativas, pseudorelativas e orações completivas. Sua origem remonta ao advérbio relativo locativo *opou* (*opou* > *pu*) para culminar em um relativizador generalizado, isto é, “*to meros pu pigame* (o lugar onde/em que nós fomos), *o logos pu irthame* (a razão pela qual/porque viemos), etc.” (ROUSSOU, 2012, p. 1 [tradução nossa]⁴).

De acordo com os autores, as funções de *pu* se estendem em muitas perspectivas, desde relativizador a complementizador, além do pronome locativo e interrogativo. Desse modo, ele assume o mesmo papel nas orações relativas, correspondendo a posições sintáticas identificadas como argumentos oblíquos. Além disso,

se as orações completivas são relativas ocultas ou vice-versa, isso depende das suposições com as quais começamos em relação a essas duas construções. Embora, crucialmente, a visualização de complementadores como pronominais podem nos

⁴ Original: *to meros pu pigame* (the place **where/that** we went), *o logos pu irthame* (the reason **why/that** we came), etc.

fornecer uma visão mais fresca das propriedades dessas construções. (ROUSSOU, 2012, p. 8 [tradução nossa]⁵)

Observamos, a partir desses detalhes, que o elemento *pu* apresenta um processo de mudança já bem estabilizado no grego, em oposição ao **onde** no Português Brasileiro. De todo modo, é possível compreender que sua multifuncionalidade afetou a sua estrutura sintática, sendo possível seu uso para além de um pronome relativo, assim como já averiguamos em **onde**.

Por fim, Beaman (2021) busca investigar os usos do marcador relativo *wo* na Suábia Central, um alto dialeto alemânico falado no sudoeste da Alemanha. Seu uso típico é quando o antecedente é um lugar físico ou uma noção temporal e, de acordo com Beaman (2021, p. 158 [tradução nossa]⁶),

a falta de marcações de gênero, número e caso dos *wo*-relativos os torna um candidato ideal para *koinéisation*, um processo pelo qual “novas variedades de uma língua são trazidas como resultado do contato entre falantes de variedades mutuamente inteligíveis dessa língua” (Kerswill 2004:669).

Nesse sentido, a pesquisadora pontua como a prescrição tem um papel importante no fenômeno: ela obstrui uma trajetória natural de mudança da língua. Quanto maior o contato do falante com a escolaridade, menos ele escolhe fazer uso das variantes de *wo*. Esse processo pode ser assemelhado ao que observamos em **onde**: há uma prescrição forte sobre seu valor locativo que, de certo modo, impede seu desenvolvimento como uma mudança linguística.

A partir dessas considerações interlinguísticas, apresentaremos as especificidades dos usos de **onde** na fala paulista na seção abaixo.

2.1.2 Uma investigação paulista

O pronome relativo **onde** apresenta uma diversidade de classificações na tradição gramatical, tanto portuguesa quanto brasileira, ao mesmo tempo em que se observa um consenso no que diz respeito à semântica de suas referências anafóricas: o **valor locativo** ou o **lugar em que** (SAID ALI, 2001; MATEUS et al, 2003; CUNHA, CINTRA, 2008; BECHARA, 2009). Contudo, uma bibliografia recente constatou uma multifuncionalidade relacionada ao

⁵ Original: Whether complement clauses are hidden relatives or vice versa depends on the assumptions we start with regarding these two constructions. Crucially though viewing complementizers as pronominals can provide us with a fresher look of the properties of these constructions.

⁶ Original: Wo-relatives' lack of gender, number, and case markings make them an ideal candidate for *koinéisation*, a process by which “new varieties of a language are brought about as a result of contact between speakers of mutually intelligible varieties of that language” (Kerswill 2004:669).

item no Português Brasileiro (SOUZA, 2003; CAMBRAIA, ARAÚJO, 2004; ALMEIDA, BERLINCK, 2019; ALMEIDA, 2020, entre outros), apontando para um conflito entre o uso da norma e a norma de uso. O que esse contraste revela é termos informações consistentes comprovando um percurso de gramaticalização de **onde**. As ocorrências em (2.3) e (2.4) ilustram a discussão proposta:

(2.3) “no Nordeste por exemplo nós temos... ainda **o coronelismo... onde aqueles... Velhos coronéis ainda imperam a política...**” (AC-113; RO: L. 210-211)

(2.4) “**os últimos dois anos onde eu passei a desconfiá(r)** que ele tinha... caso com uma colega de trabalho” (AC-110; NE: L. 58-61)

Em (2.3) e (2.4) **onde** encabeça uma oração relativa encaixada em um antecedente que não é locativo, o que reflete um conflito com o discurso normativo. Mais especificamente, o antecedente em (2.3) é uma noção abstrata que designa um movimento ou uma prática política, “coronelismo”, e em (2.4) é uma noção temporal, “os últimos dois anos”. Assim, os usos apresentados estão em um intermédio entre uma esfera de realidade biossocial, isto é, valores relacionados ao mundo concreto, tal qual o locativo associado ao **onde**, e um uso em nível interpessoal, com função de recuperar contextos específicos no processo comunicativo do falante.

Partindo desse cenário, nosso estudo buscou mapear e caracterizar os usos que o pronome relativo **onde** tem assumido na fala paulista. Além disso, tínhamos como objetivo compreender as seguintes perguntas de pesquisa: (i) é possível identificar usos não locativos de **onde** em produções orais de paulistas?; (ii) se **onde** também remete a categorias não espaciais nos dados paulistas, que fatores (linguísticos e/ou extralinguísticos) se correlacionam a tais usos?; (iii) em que medida os chamados usos “desviantes” de **onde** se inserem em um processo de gramaticalização? De forma geral, essa abordagem do fenômeno se justifica, pois (i) apesar de existir uma bibliografia consistente sobre o fenômeno, ainda há contribuições a serem feitas para interpretar esse processo no Português Brasileiro; (ii) são poucos os estudos que tratam do uso de **onde** na fala, sendo a maioria sobre a sua caracterização na escrita; e, por fim (iii) os estudos carecem de uma visão sociolinguística que contribua para a descrição do uso.

Para isso, nos utilizamos de um *corpus* composto pelos bancos de fala IBORUNA, do projeto ALIP (Gonçalves, 2007) e Projeto SP2010: Amostra da fala paulistana (MENDES, 2013). Os dados foram coletados com o auxílio do programa AntConc (TANG, 2011) e a quantificação de dados foi efetuada na linguagem de programação R (CORE TEAM, 2018).

O levantamento de dados na amostra resultou em um *corpus* de 649 orações com o pronome **onde** do banco de dados IBORUNA, do projeto ALIP (Gonçalves, s.d.), das quais em 574 (88,5%) o **onde** remete a uma leitura locativa *strictu sensu* e em 74 (11,5%) se associa a uma categoria semântica não espacial. No banco de dados SP2010 foram encontradas 395 orações com o pronome, sendo apenas 29 (7%) com usos de **onde** não locativo. Considerando, então, a primeira questão de pesquisa, a resposta dos dados paulistas analisados é positiva: é possível identificar usos não locativos na fala paulista.

Desse modo, a investigação se desenvolveu em dois momentos: (a) descrever e compreender os fatores linguísticos e extralinguísticos associados aos usos não locativos de **onde** na fala paulista e (b) caracterizar, por meio de categorias semântico-cognitivas, os valores não locativos de **onde**, tendo por base o estudo de Braga e Manfili (2004).

Para o presente recorte da pesquisa, gostaríamos de reproduzir dois resultados referentes a variáveis extralinguísticas: a escolaridade e a faixa etária do falante no banco de dados IBORUNA, reproduzidos nas tabelas 2 e 3.

Tabela 2 – Frequência e proporção de uso de **onde** “não locativo” segundo escolaridade no bando de dados IBORUNA

Escolaridade	% de uso “não locativo”	N de “não locativo”	Total
Primeiro Fundamental	7%	6	89
Segundo Fundamental	10%	14	147
Ensino Médio	17%	31	182
Ensino Superior	10%	23	230
Total	11,5%	74	648
$X^2 = 8.5 (3), p = 0.03551$			

Fonte: Almeida (2019)

Nossa hipótese inicial sobre a escolaridade se relacionava com as produções de redação, principalmente aquelas com enfoque nos vestibulares, tendo em vista o metadiscorso de professores que pontuam a dificuldade de seus alunos de compreenderem a norma padrão em relação a **onde**. Assim, hipotetizamos que o aluno, buscando se aproximar da norma culta, poderia utilizar o **onde** não locativo como uma forma de se aproximar de um discurso mais “formal”. Essa interpretação equivocada, tendo em vista que vai na direção contrária à da prescrição gramatical, se justifica pelo pouco contato do aluno com o gênero textual⁷.

⁷ Um projeto em desenvolvimento aponta que quanto mais redações o aluno produz, faz menos uso de **onde** em sua produção. Uma primeira publicação já foi feita por Almeida e Sene (2021), mas o estudo continua em desenvolvimento, buscando mais dados.

O resultado exposto na Tabela 2 mostra uma concentração de uso não locativo justamente no Ensino Médio, aquele em que existe uma ênfase na produção de redações escolares. Porém, o banco de dados não se distribui de forma ortogonal, no sentido de que temos falantes de todas as faixas etárias com esse nível de escolaridade, o que não necessariamente relaciona essa faixa com o momento escolar. Além disso, a baixa quantidade de dados nos impede de produzir qualquer tipo de afirmação categórica sobre essa variável, apenas de aventar uma possível correlação, a ser mais profundamente investigada.

Tabela 3 – Frequência e proporção de uso de **onde** “não locativo” segundo a **faixa etária** do falante no bando de dados IBORUNA

Faixa etária	% de uso “não locativo”	N de “não locativo”	Total
1ª faixa etária (7-15 anos)	4%	3	80
2ª faixa etária (16-25 anos)	21%	20	96
3ª faixa etária (26-35 anos)	5%	10	192
4ª faixa etária (36-55 anos)	11%	17	159
5ª faixa etária (55 anos ou +)	19%	24	121
Total	11,5%	74	648

$$X^2 = 28.9 (4), p = 0.000008045$$

Fonte: Almeida (2019)

No que se refere à faixa etária, tínhamos como hipótese de que essa variável extralinguística poderia contribuir para a apreensão do fenômeno, colaborando para a avaliação da hipótese de gramaticalização. Contudo, observamos na tabela que não há um crescimento ou decréscimo do uso de **onde** “não locativo” de acordo com as faixas etárias.

Dois grupos etários apresentam maiores frequências de uso não locativo de **onde**: os falantes entre 16 e 25 anos e aqueles com mais de 55 anos. Em contraponto, as demais faixas, compostas pelos falantes mais jovens da amostra (de 7 a 15 anos), assim como falantes entre 26 e 35 anos e entre 36 e 55 anos, têm uma proporção baixa de emprego de **onde** “não locativo”. Nossa interpretação, para aquele momento da pesquisa, foi compreender possíveis motivações específicas a cada grupo etário, além das interações com outros grupos de fatores. Aqui também o pequeno número de dados limita o alcance dessa interpretação.

Com a presente análise do processo de variação, podemos correlacionar essas duas variáveis com os usos de **onde** locativo (padrão com **onde**), assim como de outras estratégias de relativização. Antes, é necessário alguns apontamentos teórico-metodológicos para contextualização da investigação.

2.2 O OBJETO DE ESTUDO

Nosso objeto de estudo parte de um tipo de oração subordinada que pode receber dois nomes: a subordinação adjetiva ou a oração relativa. Sua complexidade se revela pelos inúmeros trabalhos, de diferentes abordagens teóricas, que buscaram descrever e quantificar suas especificidades. A principal fonte parte de trabalhos tipológicos que identificaram a construção como um possível universal linguístico (KEENAN, COMRIE, 1977; COMRIE, 1989; GIVÓN, 1990). Por meio de descrições semânticas e pragmáticas, os autores buscaram identificar os denominadores sintáticos comuns a todas as línguas na expressão da relativização. Tal como resume Givón (1990, p. 646 [tradução nossa]⁸) ao propor duas definições para as orações relativas, tendo como ponto de partida suas funções dentro de uma situação comunicativa e a intencionalidade do falante:

- a. Semântica: “uma *oração relativa* codifica uma proposição em que um dos participantes é correferencial com o núcleo que por ela é modificado”
- b. Pragmática: “uma oração relativa restritiva envolve uma proposição que o falante assume ser conhecida ou acessível pelo ouvinte, ou de outra forma menos provável como uma informação nova controversa”⁹

A esse respeito, podemos apontar, ainda, como um exemplo dos resultados dos estudos tipológicos, a Hierarquia de Acessibilidade dos Sintagmas Nominais de Keenan e Comrie (1977). O princípio postulado pelos autores tem como fundamento que todas as línguas apresentam uma estratégia primária, e principal, para a relativização, que se sucederia na forma de uma escala. Assim, temos uma posição relativizável mais acessível que culmina na direção de posições gradualmente menos acessíveis, representadas no cline abaixo:

Sujeito > Objeto Direto > Objeto Indireto > Oblíquo > Genitivo > Objeto de Comparação

Desse modo, a leitura da escala deve partir de um ponto de vista implicacional, o que indica que, se uma língua permite uma estratégia de relativização para uma determinada função, isso implica que é possível a existência de estratégias para todas as funções mais altas (à sua

⁸ Original: “a. Semantic: “a relative clause codes a proposition one of whose participants is coreferential with the head noun that is modified by the clause” b. Pragmatic: “a restrictive relative clause involves a proposition that the speaker assumes is known or accessible to be hearer, or otherwise unlikely to be challenged as controversial new information” (GIVON, 1990, p. 646)

⁹ A segunda definição não é universal visto que há casos em que esse núcleo nominal pode ser indefinido e novo ao ouvinte, e o falante parte dessa proposição para construir o sentido da conversa; ou seja, o correferente não é pragmaticamente pressuposto, sendo uma informação nova ou não facilmente acessível naquele momento.

esquerda). Assim, se a língua tem uma estratégia para relativizar um Objeto Indireto, também terá para o Objeto Direto e para o Sujeito; mas não necessariamente para as funções mais baixas (à sua direita). A partir desse *cline*, fica implícita a ideia de que todas as línguas deveriam apresentar estratégias de relativização na posição de sujeito, por ser uma posição mais acessível à relativização.

Tendo em vista os postulados universais, é importante dizer que a expressão da relativização pode variar; isso quer dizer que cada língua apresenta uma estrutura sintática específica para esse fenômeno. Dessa forma, no Português Brasileiro, do ponto de vista estrutural, podemos descrever uma “fórmula” de como a subordinação adjetiva se realiza, sendo ela:

Figura 1 – Fórmula da subordinação adjetiva no Português Brasileiro



Fonte: adaptado de Perini (2016, p. 283)

Nota-se, na terceira parte da equação, uma estrutura que poderia ser considerada agramatical, dada a incompletude da oração. Entretanto, essa construção é aceita na língua, tendo em vista que “na estrutura relativa, a valência do verbo continua sendo satisfeita, mas de maneira sintaticamente diferente. O complemento faltante na estrutura oracional incompleta é justamente o nominal [...]” (PERINI, 2016, p. 284) expresso anteriormente. Assim, em (2.5)

(2.5) A bobagem *que o cara disse* me deixou irritado¹⁰

O nominal “a bobagem” é seguido pelo relativo **que** e pela estrutura oracional incompleta “o cara disse”. A estrutura não é inaceitável porque a relativização “conserta” a valência do verbo “disse”, voltando-se para o nominal. Portanto, “que o cara disse” é modificador do núcleo do sintagma nominal a qual está vinculado, “a bobagem”, especificando-o. Assim, introduzimos um cunho semântico para a descrição da relativização, a partir do valor que ela pode vir a expressar, isto é, a de um adjetivo. O que está sendo proposto é que a oração

¹⁰ Exemplo retirado de Perini (2016, p. 283-284)

denominada incompleta por Perini (2016) pode ser “substituída” por um adjetivo, articulando uma correferencialidade entre dois sintagmas nominais (CASTILHO, 2014, p. 366). Por exemplo,

(2.6) [**O aluno atento**] passa de ano

(2.7) [**O aluno estudioso**] passa de ano

(2.8) → [**O aluno atento que é estudioso**] passa de ano

Castilho (2014, p. 366) analisa que os sintagmas **aluno atento** e **aluno estudioso** se constituem como correferenciais ao produzirem afirmações a respeito de um mesmo indivíduo que podem ser relacionadas por meio da relativização. Contudo, uma ressalva precisa ser feita: não são todas as estruturas relativas que permitem a substituição da oração encaixada por um adjetivo. Em (2.5), por exemplo, há uma dificuldade para se manter o conteúdo semântico trocando **que o cara disse** por um adjetivo simples.

A partir dessas colocações, introduziremos algumas descrições e discussões a respeito da oração relativa no Português Brasileiro por meio de três abordagens: (i) a tradição normativa; (ii) a tradição descritiva; e (iii) os estudos linguísticos sobre as estratégias de relativização.

2.2.1 Tradição normativa

A partir de uma perspectiva normativa de descrição do Português, Bechara (2009) define a subordinação adjetiva como aquela em que uma oração, por meio de equivalência semântica e sintática com um adjetivo, é transformada, através de um pronome relativo, à funcionalidade de um adjunto adnominal, atuando, assim, em um nível sintático inferior. Dessa forma, o pronome assume a função sintática do antecessor que reintroduz. Ademais, ao produzir reflexões acerca do índice preposicional dentro da oração, o autor ressalta a existência de complementos verbais que pedem por uma preposição, e que devem ser englobados dentro da construção. Conforme o exemplo (2.9),

(2.9) **O livro de que gostas** está esgotado

O elemento “livro” tem a função sintática de complemento relativo do verbo “gostar” e, somado a isso, o próprio verbo é acompanhado da preposição “de”. Nesses casos, o gramático afirma que o uso da preposição é imprescindível.

Ademais, Rocha Lima (2020 [1972]) afirma que há uma diferença na função sintática do pronome relativo e do antecedente ao qual faz referência. O que o autor propõe é que o pronome é responsável por reproduzir o significado do sintagma nominal, porém, seu papel sintático está relacionado à oração encaixada. Utilizando o exemplo (2.9) para ilustrar a proposição do autor, o antecedente “o livro” cumpre uma função sintática de sujeito dentro da oração principal, “o livro está esgotado”, enquanto o pronome relativo “que” funciona como objeto indireto do verbo “gostar”.

De acordo com Rocha Lima (2020 [1972], p. 333), “o emprego de orações adjetivas permite que juntemos ao substantivos características mais complexas, para as quais, muitas vezes, não existem na língua adjetivos léxicos”. Isso aponta para uma caracterização das relativas através de um viés pragmático:

Vemos, então, que, do ponto de vista pragmático, há duas situações envolvidas no emprego da oração relativa: no primeiro caso, a relativa poderia, em princípio, ser substituída por um adjetivo equivalente, como em (5)¹¹; já no caso de (10)¹², não haveria na língua um elemento correspondente ao significado da oração adjetiva. Nota-se que, nos dois casos, a oração adjetiva, por tratar-se de uma proposição, acrescenta informações mais abrangentes que um adjetivo. (BISPO, CUNHA, 2019, p. 141)

O viés pragmático expande a funcionalidade da construção, apesar de prevalecer uma associação ao caráter adjetival da oração. Nesse sentido, gostaríamos de pontuar uma discussão interessante no âmbito do ensino da oração adjetiva: a sala de aula, assim como os materiais didáticos¹³, ainda privilegia a visão da construção por um viés normativo e limitado, isto é, sua substituição por um adjetivo.

Isso demonstra uma falha nos currículos escolares, tendo em vista que a Base Nacional Comum Curricular estabelece “o uso consciente e reflexivo de regras e normas da norma-padrão em situações de fala e escrita nas quais deve ser usada” e “propõe ao professor orientar o aluno quanto às adequações dos usos linguísticos conforme as situações sociocomunicativas” (SANTOS, MELO, 2019, p. 129). Desse modo, concordamos com Bagno (2012, p. 906) ao afirmar que

¹¹ O exemplo (5) utilizado pelos autores é: A água é um líquido/**que não tem cor**” [grifo dos autores]

¹² O exemplo (10) utilizado pelos autores é: “Dizei-me, águas mansas do rio/ Para onde levais essa flor/ **Que no vosso espelho caiu?**” [grifo dos autores]

¹³ A informação se pauta em análise preliminar deste tópico que foi produzida para a disciplina de Tópicos de Sociolinguística do Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras de Araraquara (UNESP) e culminou em um artigo que ainda não foi publicado. O que averiguamos é que os materiais didáticos não apresentam as estratégias de relativização e se fazem necessárias discussões tais quais de Mallmann, Garcia e Escobar (2019) sobre orientações para o ensino.

A pesquisa com as gramáticas normativas e os livros didáticos deve mostrar que, embora compartilhando uma doutrina comum, os bons gramáticos também apresentam em seus trabalhos algumas divergências, entre si, pois eles costumam reconhecer as falhas das conceituações tradicionais e oferecem alternativas para as explicações consagradas.

Por conseguinte, esperamos que a presente pesquisa possa somar à literatura de análise das orações adjetivas e essa oriente um ensino mais contextualizado do fenômeno linguístico.

2.2.2 Tradição descritiva

Ao selecionar três gramáticas descritivas, notamos uma diferença saliente no que se refere à descrição das construções relativas, comparativamente à tradição normativa: todas apresentam as discussões provenientes de estudos linguísticos sobre o uso padrão e não padrão das estratégias de relativização. Desse modo, Bagno (2012), ao introduzir o tópico anuncia um fenômeno comum às línguas românicas: a despronominalização. Segundo o autor, “a palavra que se despronominalizou, perdeu sua função anafórica, e se transformou em mero conector entre duas sentenças independentes” (BAGNO, 2012, p. 900).

A partir desse processo, surge a estratégia não padrão copiadora que se caracteriza por um pronome cópia ou lembrete no interior da oração encaixada, carregando em si a função anafórica antes efetuada pelo pronome. Assim, no exemplo (2.10):

(2.10) Eu vou descrever **a casa que eu estou morando nela**

O pronome lembrete “nela” é responsável por retomar o antecedente, isto é, “a casa”. Além disso, o autor também introduz uma segunda variante não padrão: a estratégia cortadora. Essa se descreve por meio do apagamento da preposição em contextos preposicionados, tal qual em (2.11):

(2.11) Eu vou descrever **a casa Ø que eu estou morando**

O verbo utilizado na oração (2.11), “morar”, requer uma preposição que deveria aparecer antes do pronome relativo, “em”. Entretanto, nessa estratégia, muito produtiva no Português Brasileiro, tal qual ressalta Bagno (2012), há um apagamento de “em”.

De forma análoga, Neves (2018) apresenta as duas estratégias ao tratar das funções sintáticas dos diversos pronomes relativos dentro da oração adjetiva. Contudo, a autora não

nomeia os processos de relativização ao caracterizá-los como: (i) “casos em que essas duas funções do PRONOME RELATIVO **que** – sujeito e objeto direto – vêm, logo adiante, preenchidas novamente por um pronome pessoal do caso reto” (NEVES, 2018, p. 657), isto é, a estratégia copiadora, que se refere a “construções desabonadas pelas regras prescritivistas”; e (ii) “contrariando normas prescritivistas, frequentemente a preposição é omitida antes do PRONOME RELATIVO (**que**) que funciona como objeto indireto na ORAÇÃO ADJETIVA”, isto é, a estratégia copiadora.

Por fim, Castilho (2014, p. 368), ao descrever a sintaxe das orações adjetivas, não deixa de ressaltar a importância dos estudos linguísticos para a compreensão do fenômeno:

Investigando a perspectiva aberta por Mary Kato, Tarallo (1983) mostrou que os pronomes relativos estão perdendo suas propriedades pronominais, restringindo progressivamente sua atuação gramatical à de um nexos, sem papel funcional, fato já apontado anteriormente. Ele examinou essa possibilidade no português de São Paulo, confirmando em parte a hipótese de Kato. Tarallo demonstra que o uso da adjetiva copiadora é favorecido pelas seguintes condições: (i) quando o antecedente da adjetiva é /humano, singular e indefinido/; (ii) quando o sintagma nominal relativizado ocupa funções sintáticas na seguinte hierarquia: genitivo > objeto indireto > oblíquo > sujeito > objeto direto; (iii) em adjetivas afastadas do sintagma nominal por um segmento encaixado; (iv) quando o falante procede de classes não escolarizadas; (v) quando fala informalmente. Quanto a adjetiva cortadora, ela corresponde às altas taxas de apagamento do pronome em função oblíqua.

Logo, observamos que as gramáticas descritivas se pautam nas constatações de estudos linguísticos que revelaram uma norma de uso referente às construções relativas no português brasileiro. Com a introdução do estudo de Tarallo, se faz necessária a descrição de outras investigações das estratégias de relativização no Português Brasileiro.

2.2.3 As estratégias de relativização no Português¹⁴

O interesse no processo de variação e mudança das estratégias de relativização no Português Brasileiro é antigo. Remontamos à publicação do *Dialeto Caipira*, de Amaral (1920, p. 31), como um primeiro índice da variabilidade das formas, dado que o autor afirma que “nas orações relativas não se emprega senão **que**. Nos casos que, em bom português, reclamam este

¹⁴ Como mencionamos no início da seção, as estratégias de relativização se articulam como um fenômeno fértil dentro da linguística, sendo analisado por diferentes correntes teórico-metodológicas. No que se refere ao escopo do nosso trabalho, e tendo em vista as discussões que serão produzidas na seção (4), de análise do *corpus*, foram selecionados trabalhos de cunho sociovariacionista e que apresentem discussões sobre variáveis extralinguísticas. Porém, não estamos invalidando outros tipos de abordagem ou desmerecendo sua contribuição para a descrição do fenômeno, tais quais os trabalhos de Kato (1981, 1993). Apenas consideramos que suas discussões não seriam devidamente produtivas dentro do recorte de nossa investigação.

pronome precedido de preposição, o caipira desloca a partícula, empregando-a no fim da frase com um pronome pessoal”. Além disso, Amaral (1920, p. 31) assinala que “frequentemente se suprimem de todo a preposição e o pronome pessoal, e diz-se: a casa que eu morei, o livro que eu falei, ficando assim a relação apenas subentendida”. Dessa forma, observamos que o autor, apesar da limitação de sua época no que se refere aos materiais de investigação sintática, já evidencia o uso das estratégias **copiadora** e **cortadora**, respectivamente, no dialeto caipira.

De forma semelhante, Lemle (1978), ao evidenciar possíveis linhas de pesquisa que dão conta da heterogeneidade linguística, introduz as estratégias de relativização não padrão do Português Brasileiro. Sua justificativa preliminar para o fenômeno se dá, no caso da copiadora, por meio da proposição de que “a inovação gramatical aportada pelo dialeto não-padrão no que tange à regra de relativização teve como motivação inicial a prevenção dessas construções ambíguas [estratégia padrão]” (LEMLE, 1978, p. 85); no caso da cortadora, se refere a uma aproximação à gramática padrão, dado que os falantes parecem tentar imitar as formas do dialeto padrão, realizando “a operação mais evidente, de cancelar a cópia pronominal (e a preposição que a precede, se for o caso), esquecendo-se, porém, da segunda (e menos evidente) operação necessária para atingir o seu alvo, ou seja, a colocação da preposição” (LEMLE, 1978, p. 85).

Uma descrição mais apurada do fenômeno surge em Mollica (1977) por meio da análise da fala carioca. O objetivo da autora é investigar as construções relativas, partindo da análise do uso variável da anáfora pronominal do sintagma nominal antecedente em posição sintática de sujeito, complemento preposicionado e complemento não preposicionado. Para isso, Mollica se utiliza de uma variável dependente binária, composta pela (i) relativa com cópia, isto é, a estratégia **copiadora** e (ii) relativa com apagamento, isto é, a estratégia **cortadora** e a estratégia **padrão**. Além disso, para compreender o padrão de variação da fala carioca, a pesquisadora parte do apagamento do pronome cópia como regra de aplicação.

Em resumo, seus resultados mostram uma gradação que segue a Hierarquia de Acessibilidade dos Sintagmas Nominais de Keenan e Comrie (1977). Isto é, a regra de aplicação se dá com maior frequência em contextos de sujeitos (758/795, 95,3%), seguidos por objetos diretos (338/416, 92,1%) e objetos indiretos (54/88, 61,4%). Ao retomar seu estudo, comparando amostras de 1980 e 2000, Mollica (2003, p. 137) conclui que “verifica-se a tendência de os sujeitos estarem sendo paulatinamente preenchidos no português, sendo a cláusula relativa um contexto favorável para que não haja sujeito nulo, o que estaria indicando uma modificação de parâmetro na língua portuguesa”.

Além disso, a autora pontua que indivíduos com nível superior de escolaridade produzem mais os contextos anafóricos, isto é, da estratégia copiadora, quando os antecedentes apresentam traço [+ humano] e se encontram distantes do relativo. Tal uso é justificado, então, para evitar uma ambiguidade estrutural. Segundo Mollica (2003, p. 132)

é de se supor que a escola monitora com sucesso somente os casos mais estigmatizados em que o falante utiliza a anáfora em posição de sintagma preposicionado, como em “o cachorro Ø que eu gosto dele”, mas não dá conta da forma *standard* “o cachorro de que eu gosto”, verificando-se atualmente o predomínio de emprego da forma cortadora “o cachorro que eu gosto”.

Essa colocação é importante para refletir sobre o rótulo de variante mais ou menos estigmatizada dentro do nosso fenômeno e para compreender a influência da monitoração proveniente do ambiente escolar. Desse modo, ambos os estudos de Mollica foram notórios e inspiraram outras investigações que pudessem contemplar toda a complexidade do fenômeno.

Tarallo (1983) foi o responsável por uma visão mais ampla do quadro de estratégias de relativização em Português Brasileiro ao produzir uma investigação sincrônica e diacrônica do fenômeno. No que se refere à primeira, o pesquisador construiu um *corpus* composto de: (i) entrevistas sociolinguísticas com 40 informantes de São Paulo estratificados em nível de escolaridade, renda, ocupação, classe social, idade e sexo; (ii) discurso da mídia, sendo utilizados transmissões esportivas, documentários, entrevistas e mesas redondas; (iii) por fim, ainda foram considerados materiais provenientes de novelas brasileiras. Com um total de 1700 dados, Tarallo (1983) encontrou a seguinte distribuição das estratégias de relativização em seu corpus sincrônico (tabela 4).

Tabela 4 - Distribuição da proporção de uso das estratégias de relativização de acordo com estratégia em 5 funções sintáticas

ESTRATÉGIA	SUJEITO	OBJETO DIRETO	OBJETO INDIRETO	OBLÍQUO	GENITIVO
Com lacuna	89,7% 890 dados	97,4% 374 dados	-	-	-
<i>Piedpiping</i>	-	-	3,9% 3 dados	7,4% 17 dados	5,9% 1 dado
Pronome lembrete (copiadora)	10,3% 102 dados	2,6% 10 dados	21,1% 16 dados	10,4% 24 dados	52,9% 9 dados
Cortadora	-	-	75% 57 dados	82,2% 190 dados	41,2% 7 dados
<i>Total</i>	992 dados	384 dados	76 dados	231 dados	17 dados

Fonte: adaptado de Tarallo (1993)

A partir da Tabela 4, observamos uma predominância da estratégia **com lacuna** em posições de sujeito e objeto direto. Tal estratégia é descrita pela prescrição gramatical, o que a torna uma estratégia padrão, e observamos que ela é a mais produtiva. Em contextos preposicionados, a estratégia padrão, ***piedpiping***, apresenta um total de 21 dados em oposição a 254 dados da estratégia **cortadora**. Dessa forma, observamos que na fala paulista da década de 80, em contextos preposicionados, há uma prevalência da forma não padrão.

No que se refere aos fatores extralinguísticos analisados, Tarallo afirma que descartou de sua análise as variáveis sexo e idade, tendo em vista que a diferença entre os níveis não foi estatisticamente significativa. Por outro lado, a classe social do falante se mostrou um fator interessante no contexto preposicionado, isto é, nas posições sintáticas mais baixas de objeto indireto, oblíquo e genitivo. Apesar de não controlar o conteúdo semântico das estratégias, no sentido de não fazer diferença entre as relativas locativas e não locativas, esse resultado é importante para o escopo de nosso estudo e é apresentado na Tabela 5.

Tabela 5 – Porcentagem de uso das estratégias **copiadora**, **cortadora** e ***piedpiping*** nas três posições sintáticas mais baixas para as três classes sociais

ESTRATÉGIA	BAIXA	MÉDIA	ALTA	TOTAL
<i>Piedpiping</i>	1,3% 2 dados	5,6% 5 dados	17,9% 14 dados	21 dados
Pronome lembrete (copiadora)	21% 33 dados	11,2% 10 dados	9% 7 dados	50 dados
Cortadora	77,7% 122 dados	83,2% 74 dados	73,1% 57 dados	253 dados
<i>Total</i>	157 dados	89 dados	78 dados	324 dados

Fonte: adaptado de Tarallo (1983)

Desse modo, verificamos que, em contextos preposicionados, as classes baixa, média e alta fazem maior uso da estratégia **cortadora**; contudo, é a classe baixa que faz um uso mais expressivo da estratégia **copiadora**, em comparação às demais classes sociais. Segundo Tarallo (1983, p. 132-133), é possível propor uma escala implicacional a partir desses valores, sendo que a estratégia **cortadora** é menos estigmatizada do que a estratégia **copiadora** para os dois grupos sociais mais altos.

Gostaríamos de destacar, ainda, que o uso da estratégia padrão (*pied-piping*) é mais frequente entre os falantes de classe alta (há uma diferença de 12 pontos percentuais em relação

à frequência da classe média, por exemplo). Isso talvez permita uma correlação com escolaridade dos falantes.

Buscando compreender a inserção das estratégias não padrão no Português Brasileiro, Tarallo segue para a segunda parte da investigação, a diacronia. Seu *corpus*, que contempla do ano 1725 até 1880, dividido em períodos de 50 anos, é composto por cartas e peças teatrais. De acordo com Tarallo (1983), foram coletadas 200 orações relativas por período, totalizando 1579 dados, distribuídos na Tabela 6.

Tabela 6 - Distribuição da proporção de uso das estratégias de relativização *piedpiping*, copiadora e cortadora através do tempo

ESTRATÉGIA	I 1725	II 1775	III 1825	IV 1880
<i>Piedpiping</i>	89,2% 99 dados	88,1% 89 dados	91,3% 73 dados	35,4% 63 dados
Pronome lembrete (copiadora)	9,9% 11 dados	7,9% 8 dados	1,3% 1 dados	5,1% 9 dados
Cortadora	0,9% 1 dado	4% 4 dados	7,5% 6 dados	59,5% 106 dados
<i>Total</i>	111 dados	101 dados	80 dados	181 dados

Fonte: adaptado de Tarallo (1993)

Os resultados expostos na Tabela 6 mostram o crescimento da estratégia **cortadora** por volta de 1880 e, de acordo com Tarallo (1983, 1993), seu processo de competição com a estratégia **copiadora** para substituir a estratégia **padrão com preposição (*piedpiping*)**. Ademais, nota-se que a estratégia **copiadora** se mantém constante, mesmo que à margem das demais. Desse modo, o autor pontua duas mudanças: a reestruturação no sistema pronominal e a emergência da cortadora. Isso, pois

A substituição da anáfora pronominal pela anáfora zero gerou, por assim dizer, um novo tipo de relativa: um que se parece exatamente com uma oração declarativa matriz a não ser pelo complementizador invariável que a introduz. Neste sentido, a análise defendida aqui sugere que a antiga competição entre dois tipos de relativas – uma claramente envolvendo movimento (relativa padrão) e a outra, um processo de apagamento (pronomes lembretes) – somente produziu um segundo paradigma, mas os dois processos em competição permanecem os mesmos: movimento (*piedpiping*) vs. apagamento (relativa cortadora). (TARALLO, 1993, p. 89)

Corrêa (1998) apresenta uma abordagem diferente do fenômeno, complementando as investigações apresentadas até aqui: apesar de partir de uma análise variacionista das estratégias

de relativização, o seu enfoque é para o ensino. Assim, a autora confrontou dados de falantes não escolarizados, estudantes dos primeiros anos de escolarização (1º e 2º grau) e falantes cultos para averiguar quais variáveis linguísticas e extralinguísticas estavam relacionadas com o uso da estratégia **padrão com preposição** (*piedpiping*). Para que isso fosse possível, Corrêa construiu um corpus composto de narrativas escolares dos alunos do 1º grau, exercícios dos alunos do 2º grau e a fala urbana culta, proveniente do Projeto NURC.

Dessa forma, podemos elencar como primeira contribuição de seu trabalho a comprovação de que “existem estratégias vernaculares de relativização para os que não frequentaram a escola e uma estratégia aprendida através da educação formal. Além disso, [...] entre os informantes escolarizados as duas estratégias estão em variação” (CORRÊA, 1998, p. 151). Outrossim, entre os falantes não escolarizados e alunos do 1º grau houve um uso categórico da estratégia cortadora em contextos preposicionados, o que a autora correlaciona com (i) a pesquisa ter se desenvolvido no ensino público e (ii) a tendência no Português de redução do uso da preposição. Na fala dos universitários, observou-se o movimento inverso: não houve usos de cortadora, mas houve 25 dados (83%) da estratégia padrão sem preposição (isto é, em contextos não preposicionados) e 5 dados (17%) da estratégia padrão com preposição. Desse modo,

a vernacular domina os dados dos escolares em sua quase totalidade. Isso merece um comentário, porque, afinal, a maioria desses alunos já está na escola há oito anos ou mais. Se uma das finalidades do ensino institucional é veicular a língua culta, podemos, com estes dados, concluir que isso não vem ocorrendo com as relativas. Tanto nos dados orais como nos escritos os informantes não-escolarizados e os alunos da 1ª à 8ª série produzem quase que exclusivamente relativas cortadoras, com exceção de 2 casos de relativa padrão na 6ª série. Os adultos cultos somente produziram relativas padrão, nos casos de sintagmas preposicionais. (CORRÊA, 1998, p. 84)

Dessa forma, há um reforço de que a estratégia padrão com preposição não é natural do vernáculo do falante, dado que “com o aprendizado da estratégia padrão, o falante amplia suas possibilidades linguísticas, passando a ter um leque maior de opções. Quando isso acontece, passa a haver uma variação entre as estruturas” (CORRÊA, 1998, p. 155). Um ponto interessante da conclusão da pesquisadora vale a pena ser mencionado:

Quando a variação passa a existir, a estrutura padrão ocorre sob certas condições de ordem social e também linguística. Apenas quando o falante é escolarizado e preferencialmente se tiver nível alto de escolaridade, se for homem (ou se não lidar diariamente com crianças), de idade madura, se o desempenho linguístico for condição para o desempenho profissional, se a ocasião exigir certa formalidade. Depois de alinhavado, esse perfil de falante revela um forte componente ideológico, que não tínhamos como meta atingir, ao iniciar o trabalho, mas que retrata uma

realidade que se encontra, segundo Ataliba T. de Castilho, na América, onde o ideal linguístico e a norma culta não coincidem. (CORRÊA, 1998, p. 155)

Nota-se, por meio da citação, a criação de um imaginário – que reproduz o perfil averiguado na pesquisa – daquele falante responsável pela frequência da estratégia padrão com preposição dentro de um *corpus*. Isso, novamente, nos leva a correlacionar as variantes com certos níveis de prestígio e estigma que ainda não foram averiguados por um teste de percepção.

Por fim, Bispo (2009), com base no banco de dados Discurso & Gramática – a língua falada e escrita na cidade do Natal e do Rio de Janeiro –, investigou o uso da estratégia cortadora, buscando verificar um possível processo de substituição às outras, além dos contextos envolvidos na sua frequência dentro do *corpus*. Os resultados para fala com relação à escolaridade do falante estão dispostos na Tabela 7

Tabela 7 – Relativas no corpus *D&G Rio de Janeiro* e *D&G Natal* por estratégia e nível de escolaridade

Nível		Ensino Fundamental	Ensino Médio	Ensino Superior
Estratégia		Rio de Janeiro		
Padrão	Sem preposição	79% 83 dados	78,3% 144 dados	81,8% 139 dados
	Com preposição	-	0,5% 1 dado	13,8% 15 dados
Não Padrão	Cortadora	18,1% 19 dados	18,5% 34 dados	3,7% 15 dados
	Copiadora	2,9% 3 dados	2,7% 5 dados	3,7% 4 dados
Total		105 dados	184 dados	170 dados
Estratégia		Natal		
Padrão	Sem preposição	84,8% 213 dados	75,7% 373 dados	84,8% 430 dados
	Com preposição	3,2% 8 dados	0,4% 2 dados	1,6% 8 dados
Não Padrão	Cortadora	10% 25 dados	22,7% 112 dados	10,6% 54 dados
	Copiadora	2% 5 dados	1,2% 6 dados	3% 15 dados
Total		251 dados	493 dados	507 dados

Fonte: adaptado de Bispo (2009, p. 84)

De forma geral, observamos na Tabela 8 uma predominância, nos dois bancos de dados, da estratégia padrão sem preposição em contextos não preposicionados. No que concerne às estratégias possíveis em contextos preposicionados, ressaltamos o fato da **copiadora** ser a menos utilizada pelos falantes dos três níveis de escolaridade. Além disso, no Rio de Janeiro, é possível apontar um aumento no uso da **padrão com preposição** conforme o aumento da escolaridade; enquanto em Natal, há uma redução da estratégia padrão no Ensino Médio e o mesmo número de dados do Ensino Fundamental se apresenta no Ensino Superior.

Ainda sob uma perspectiva da Sociolinguística variacionista, Barros (2000) toma como foco o uso das relativas **cortadoras** na fala pessoense. Partindo da variável dependente apagamento da cópia, e controlando sete variáveis independentes, sendo elas faixa etária, escolaridade, sexo, animacidade do SN antecedente, especificidade do SN antecedente, distância e função sintática do relativo, a autora analisou um *corpus* constituído a partir do projeto Variação Linguística da Paraíba (VALPB), com entrevistas sociolinguísticas da década de 1990. A distribuição das estratégias de relativização encontradas por Barros (2000) em ambiente preposicionado está apresentada na Tabela 8.

Tabela 8 – Distribuição das estratégias de relativização em ambientes preposicionados em Barros (2000)

Estratégia	Número de ocorrências	%
Padrão com preposição	28	11
Cortadora	197	80
Copiadora	22	9
<i>Total</i>	247	100

Fonte: adaptado de Bispo (2009, p. 79)

Seus resultados corroboram a mudança no sistema apontada por Tarallo (1983) em sua investigação diacrônica, constatando a maior recorrência da estrutura **cortadora** (80%, 197 dados em 247) frente às demais estratégias em competição. Vale pontuar como as frequências das estratégias **padrão com preposição** e **copiadora** apresentam números tão reduzidos em contextos preposicionados.

Com objetivos e resultados semelhantes, podemos apontar o estudo de Burgos (2003), que investigou uma comunidade rural de descendentes afro-brasileiros do sul da Bahia denominada Helvécia. O autor buscava verificar a validade, na comunidade destacada, dos resultados de estudos anteriores, com ênfase no trabalho de Tarallo (1983). Além disso, hipotetizou se era possível afirmar que a relativa cortadora seria uma evolução da copiadora, e

esta última uma evolução da estratégia padrão, ou se todas se articulavam em processos independentes.

Para que isso fosse possível, Burgos utilizou um *corpus* construído a partir do projeto Vertentes do Português rural da região da Bahia e Sergipe. Foram utilizados dezoito informantes que concederam entrevistas informais – conduzidas pelo pesquisador com auxílio de um membro da comunidade. Segundo Bispo (2009, p. 79), “a escolha dos informantes considerou, entre outros, os seguintes critérios: o fato de eles serem filhos de falantes nativos e terem nascido na comunidade ou morarem nela pelo menos 75% de suas vidas”. Ademais, o pesquisador controlou como variáveis linguísticas: tipo da relativa, função sintática do antecedente e do relativo, distância entre o antecedente e a relativa, caráter humano do antecedente, posição da relativa em relação à oração matriz, tipo de relativo empregado; e extralinguísticas: idade, sexo, escolaridade, estada na comunidade. Os resultados obtidos, por meio da função sintática do pronome relativo, estão apresentados na Tabela 9.

Tabela 9 – Distribuição das estratégias de relativização segundo a função sintática do pronome relativo em Burgos (2003)

Estratégias	Sujeito	Objeto direto	Objeto indireto	Objeto locativo	Genitivo	Adjunto Adverbial
Padrão (com e sem preposição)	98% 247 dados	97% 101 dados	-	-	-	10% 16 dados
Cortadora	-	-	100% 24 dados	91% 20 dados	88% 7 dados	86% 134 dados
Copiadora	2% 6 dados	3% 3 dados	-	9% 2 dados	13% 1 dados	4% 6 dados
Total	255 dados	104 dados	24 dados	22 dados	8 dados	156 dados

Fonte: adaptado de Bispo (2009, p. 80)

Observamos na tabela que em posição de **sujeito** e **objeto direto** há uma predileção pela estratégia **padrão**, contexto em que não é possível o uso da **cortadora**. Quando nos voltamos para os contextos preposicionados, há a predominância da estratégia **cortadora**. Porém, é interessante pontuar a baixa frequência de dados nessas posições sintáticas mais baixas, com exceção do adjunto adverbial. Assim, seu trabalho se alinha aos demais e às investigações de Tarallo (1983).

Em suma, o que gostaríamos de demonstrar com os estudos resenhados é que as estratégias de relativização apresentam uma complexidade em sua análise de dados. Tanto

fatores linguísticos quanto extralinguísticos são importantes para compreender os fenômenos de variação e mudança linguística no Português Brasileiro. Assim, ressaltamos, mais uma vez, que os trabalhos escolhidos para compor nossa Fundamentação Teórica tiveram como pano de fundo uma análise que concatenasse os fatores estruturais e sociais da língua, apresentando resultados que nortearam nossa investigação.

2.3 PRESSUPOSTOS TEÓRICO-METODOLÓGICOS

2.3.1 A Teoria de Variação e Mudança Linguística

Como já brevemente mencionado na seção 1, *Introdução*, o trabalho tem como base teórico-metodológica a Teoria da Variação e Mudança Linguística (doravante TVM), também conhecida como Sociolinguística variacionista ou Sociolinguística laboviana (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1972, 1994, 2001, 2010). A TVM surge como uma crítica ao fato de que quanto mais os linguistas se dedicam às descrições das estruturas da língua, mais se afastam de descrições no que concernem aos diferentes estágios de uma língua, ou seja, “a transição de uma língua de um estado para outro.” (WEINREICH et al, 2006 [1968], p. 35). Para reverter esse movimento, os autores afirmam que a solução é o “rompimento da identificação de estruturalidade com homogeneidade” (WEINREICH et al, 2006 [1968], p. 36); o que se deve buscar é justamente “descrever a diferenciação ordenada numa língua que serve a uma comunidade” (WEINREICH et al, 2006 [1968], p. 36), uma vez que uma língua que não apresenta a heterogeneidade estrutural só pode ser disfuncional. Desse modo,

Fica claro, portanto, que desejamos uma teoria da mudança linguística que lide nada menos do que com a maneira como a estrutura linguística de uma comunidade complexa se transforma no curso do tempo, de tal modo que, em certo sentido, tanto a língua quanto a comunidade permanecem as mesmas, mas a língua adquire uma forma diferente. (WEINREICH et al, 2006 [1968], p. 37)

Nessa lógica, Weinreich, Labov e Herzog (2006 [1968]) pontuam que é compreensível que uma teoria linguística socialmente agnóstica tenha se desenvolvido por meio de uma correlação negativa entre a estruturalidade da língua e a sua comunicabilidade, desde que compreendemos que a correlação se deu de forma hipotética. Dessa forma, os autores não negam uma dialetologia estrutural sincrônica, tendo em mente que isso seja um exercício, já que o que é questionado é a falta de explicações no que se relaciona ao desenvolvimento da língua. Assim, “estamos negando simplesmente que a estrutura sincrônica da língua nos forneça os critérios principais para a comunicabilidade diferencial” (WEINREICH et al, 2006 [1968],

p. 90). Conclui-se que para compreender e resolver o paradoxo da mudança, se faz necessário analisar o que é mobilizado nas situações de contato, no sentido de como um determinado falante pode entender e reconhecer como seus os elementos estruturados na fala dos outros.

Desse modo, o que está sendo proposto é a análise do caráter heterogêneo dos sistemas linguísticos como “produto das combinações, alternâncias ou mosaicos de subsistemas distintos, conjuntamente disponíveis” (WEINREICH et al, 2006 [1968], p. 103). Os autores concebem cada um desses subsistemas como um conjunto coerente e integral de regras do tipo categórico, sendo necessário, como aparato teórico adicional, identificar um conjunto de regras que busquem identificar quais são as condições para a alternância.

Para que esse empreendimento fosse possível, Weinreich, Labov e Herzog (2006 [1968]) definem a co-ocorrência entre os elementos linguísticos, ou seja, a alternância entre duas estruturas dentro do código. Os autores pontuam que muitos estudos apontam para essa relação de coocorrência, mas falham em aprofundar a questão. Para que isso seja possível, é necessário o conceito de variável linguística, isto é, um elemento controlado por uma única regra e que é variável dentro do sistema. Assim, a intenção dos pesquisadores é mostrar que “não basta apontar a existência ou a importância da variabilidade: é necessário lidar com os fatos de variabilidade com precisão o suficiente para nos permitir incorporá-los em nossas análises da estrutura linguística” (WEINREICH et al, 2006 [1968], p. 107), tendo em vista que

Uma variável linguística tem de ser definida sob condições estritas para que seja parte da estrutura linguística; de outro modo, se estará simplesmente escancarando a porta para regras em que “frequentemente”, “ocasionalmente” ou “às vezes” se aplicam. A evidência quantitativa para a *covariação* entre a variável em questão e algum outro elemento linguístico ou extralinguístico oferece uma condição necessária para admitir tal unidade estrutural. A *covariação* pode ser oposta à *co-ocorrência estrita*, ou a co-ocorrência pode ser concebida como o caso limite da covariação. Provas das relações de co-ocorrência estrita podem emergir, de fato, de uma investigação quantitativa do tipo que oferece provas de covariação. (WEINREICH et al, 2006 [1968], p. 107)

Sendo assim,

O sistema heterogêneo é então visto como um conjunto de subsistemas que se alternam de acordo com um conjunto de regras co-ocorrentes, enquanto dentro de cada um desses subsistemas podem encontrar variáveis individuais que covariam mas não co-ocorrem estritamente. Cada uma dessas variáveis acabará sendo definida por funções de variáveis independentes extralinguísticas ou linguísticas, mas essas funções não precisam ser independentes umas das outras. Pelo contrário, normalmente se esperaria encontrar íntima covariação entre as variáveis linguísticas. (WEINREICH et al, 2006 [1968], p. 108).

Por conseguinte, é introduzido o estudo de Nova Iorque de Labov (1966; 2008) como um exemplo de descrição que buscou identificar a estrutura sociolinguística (WEINREICH et al, 2006 [1968], p. 116). A premissa do trabalho é compreender a mudança linguística de uma variável a partir das correlações entre fatores linguísticos e sociais apresentados na comunidade de fala selecionada. Tendo como hipótese a reintrodução da pronúncia do /r/ pós-vocálico em posição medial e final na fala nova iorquina, o pesquisador buscou investigar de que modo esse uso poderia refletir uma estratificação social da cidade. Para tanto, foram selecionadas três lojas de departamento (Saks, Macy's e S. Klein) que refletissem três classes sociais distintas, isto é, alta, média e baixa. Os resultados apontaram para uma recorrência maior da realização da vibrante nas lojas Saks (classe alta), uma realização intermediária para Macy's (classe média) e uma recorrência menor da variante na loja S.Klein (classe baixa), o que confirma a hipótese inicial de Labov sobre a correlação entre estratificação social e a variável linguística, em uma manutenção da forma de prestígio do /r/.

Por fim, os autores sintetizam os princípios que definem os fundamentos empíricos para a teoria de mudança proposta pelos autores. Eles se dividem em 5 problemas principais: o problema dos fatores condicionantes, o problema da transição, o problema do encaixamento (na estrutura linguística e social), o problema da avaliação e o problema da implementação.

O primeiro deles faz referência ao objetivo de determinar um possível conjunto de mudanças e quais as condições que a propiciam, visto que a proposta dos autores é um estudo detalhado da mudança em progresso. O segundo diz respeito à distribuição contínua da mudança através de faixas etárias da população, que apresentam um estágio interveniente pelo qual uma estrutura parte de A para uma estrutura B. Desse modo, os pesquisadores consideram o dialeto como transicional, tendo em vista que a mudança se dá “(1) a medida que um falante aprende uma forma alternativa, (2) durante o tempo em que as duas formas existem em contato dentro de sua competência, e (3) quando uma das formas se torna obsoleta” (WEINREICH et al, 2006 [1968], p. 122).

O terceiro problema, do encaixamento, se divide em duas frentes. No que se refere à estrutura linguística, parte do modelo que língua que apresenta “(1) estratos discretos, coexistentes, definidos pela coocorrência estrita, que são funcionalmente diferenciados e conjuntamente disponíveis a uma comunidade de fala; e (2) variáveis intrínsecas, definidas por covariação com elementos linguísticos e extralinguísticos” (WEINREICH et al, 2006 [1968], p. 123). Dessa forma, não estamos lidando com uma mudança linguística que desloca um sistema inteiro para outro. Pelo contrário, Weinreich, Labov e Herzog buscam dialogar com

valores que gradualmente se movimentam de um polo para outro, sendo as variantes contínuas ou discretas como um espectro contínuo de valores.

No que se refere à estrutura social, temos uma estrutura linguística que se encaixa em um contexto mais amplo da comunidade de fala, de forma que as variáveis sociais e geográficas se tornam elementos intrínsecos da própria estrutura. Desse modo, ao explicar o processo de mudança linguística, se faz necessário agregar os fatores sociais e como esses exercem uma força dentro do sistema como um todo. Vale destacar que a significação social não se distribui de forma igualitária, isto é, há estruturas linguísticas que são encaixadas de forma desigual dentro da estrutura social, ou que nos estágios iniciais e finais da mudança linguística podem apresentar pouca ou muita correlação com fatores de cunho social. Assim, “a tarefa do linguista não é tanto demonstrar o grau de correlação social que existe e mostrar como ela pesa sobre o sistema linguístico abstrato” (WEINREICH et al, 2006 [1968], p. 123).

O quarto problema, da avaliação, parte da investigação dos correlatos subjetivos de uma variável, isto é, a sua avaliação, tendo em vista que o nível de consciência social é um fator importante no processo de mudança. O quinto e último problema discute a dificuldade de implementação da mudança, já que perpassa um movimento de incentivos e restrições, com a influência de inúmeros fatores, tanto da estrutura social quanto da língua. Isso não pode impedir o exame minucioso em busca de uma visão mais abrangente do processo de mudança. Assim, “uma proposta deste tipo para os modos como os fatores sociais incidem sobre os traços linguísticos num mecanismo cíclico se baseia em padrões repetidos observados nuns poucos casos bem estudados (LABOV, 1965)” (WEINREICH et al, 2006 [1968], p. 124).

Dentro de nossa investigação, estamos trabalhando com o problema do encaixamento, tal qual mencionado na seção 1, *Introdução*. Tendo em vista que o pronome relativo **onde** está passando por um processo de mudança e que esse movimento não é feito de forma abrupta dentro do sistema, é do nosso interesse averiguar se há um impacto no seu domínio de origem, isto é, o valor locativo. Concomitante a isso, nos alinhamos aos demais estudos variacionistas ao lidar com o problema dos fatores condicionantes. Além disso, apesar de não nos aprofundarmos na discussão por meio de testes de percepção, exploramos a questão da avaliação.

2.3.1.1 A problemática da variável sintática

A Teoria de Variação e Mudança Linguística foi consolidada através de uma base fonológica de investigação, o que implica dizer que são necessárias algumas considerações

críticas quando nos dedicamos à descrição de fenômenos não fonológicos. A primeira delas foi pontuada por Lavandera (1978) no que se relaciona à conceitualização de regra variável: duas formas que exprimem o mesmo significado. O que observamos nessa premissa clássica é uma equivalência semântica que condiz com a realidade de variáveis fonológicas, porém que falha ao tratar de variáveis não fonológicas. Tal qual pontuou a autora, “unidades além da fonologia, digamos um morfema, ou um item lexical, ou uma construção sintática, cada um tem, por definição, um significado”¹⁵ (LAVANDERA, 1978, p. 175 [tradução nossa]). Dessa forma, a primeira sugestão de reformulação surge com a expansão da equivalência semântica para comparabilidade funcional, sendo esta: duas formas que exprimem a mesma função comunicativa, embora não apresentem, necessariamente, o mesmo significado.

Ademais, outros autores somaram suas contribuições para os desafios de se trabalhar com variáveis não-fonológicas dentro da Teoria de Variação e Mudança Linguística. Milroy e Gordon (2003), por exemplo, ao tratar de variáveis sintáticas, apontam para a baixa frequência das construções nas amostras de fala, já que não há uma garantia de que um tipo específico de “dado” surja em uma porção de fala espontânea. Isso é consequência de uma infinidade de possibilidades sintáticas que o falante tem a sua disposição, articuladas a restrições pragmáticas e semânticas de determinados contextos.

Outrossim, os autores pontuam, ainda, que os fenômenos sintáticos são menos avaliados socialmente, em contraposição a um fenômeno fonológico, devido a sua baixa frequência¹⁶. Por isso, observa-se uma tendência nos estudos de fenômenos sintáticos variáveis para um viés mais estrutural do que social. No que se refere a essa problemática dentro dos estudos de orações relativas, gostaríamos de citar Beaman (2021, p. 135 [tradução nossa¹⁷])

¹⁵ Original: “units beyond phonology, let us say a morpheme, or a lexical item, or a syntactic construction, each have by a definition a meaning”.

¹⁶ Segundo os autores: “Cheshire (1990) observes that, because of their lower frequency, syntactic variables are less available than phonological variables as resources for social evaluation, and so less likely to be socially indexical (see also Hudson 1996:45)” (MILROY, GORDON, 2003, p. 171)

¹⁷ Original: Much early sociolinguistic work has suggested that relative pronoun usage is a ‘covert variable’ not readily available for social evaluation (Tottie and Rey 1997:245). However, researchers in the 1980s and 1990s working within the variationist framework began to find that variation in relative pronoun usage was not only constrained by linguistic conditioning and syntactic position but was also correlated with various social factors, such as genre, style, education, and socio-economic status. Romaine (1982) observed that wh-pronouns are generally restricted to written texts and to specific groups of speakers, that is, educated individuals with middle-class aspirations. Quirk (1957), and later Tottie (1995), found that zero relatives are strongly favoured with personal pronouns that are the subject of the relative clause, while wh-relatives are correlated with speakers’ educational level. Guy and Bayley (1995) established that the channel of communication (spoken or written), the animacy of the antecedent, the syntactic position of the relativiser, and the distance between the antecedent and the relativiser all have significant effects on speakers’ choice of relative pronouns. Cheshire’s (1996) investigation of the lexical item that revealed that relative pronouns can serve as a linguistic ‘sign-post’ in communication, reflecting “general social principles of cooperative activity between individuals” (Cheshire 1996:392)

Muitos trabalhos sociolinguísticos iniciais sugeriram que o uso de pronome relativo é uma “variável oculta” não prontamente disponível para avaliação social (Tottie e Rey 1997:245). No entanto, pesquisadores nas décadas de 1980 e 1990 trabalhando dentro do quadro variacionista começaram a descobrir que a variação no uso do pronome relativo não era apenas limitada pelo condicionamento linguístico e posição sintática, mas também estava correlacionada com vários fatores sociais, como gênero, estilo, educação e status socioeconômico. Romaine (1982) observou que os pronomes-*wh* são geralmente restritos a textos escritos e a grupos específicos de falantes, ou seja, indivíduos educados com aspirações de classe média. Quirk (1957), e mais tarde Tottie (1995), descobriram que as relativas zero são fortemente favorecidas com pronomes pessoais que são o sujeito da oração relativa, enquanto as relativas-*wh* estão correlacionadas com o nível educacional dos falantes. Guy e Bayley (1995) estabeleceram que o canal de comunicação (falado ou escrito), a animacidade do antecedente, a posição sintática do relativizador e a distância entre o antecedente e o relativizador têm efeitos significativos na escolha dos pronomes. A investigação de Cheshire (1996) do item lexical *that* revelou que os pronomes relativos podem servir como um “poste de sinalização” linguístico na comunicação, refletindo “princípios sociais gerais de atividade cooperativa entre indivíduos” (Cheshire 1996:392).

Desse modo, temos que ter em mente durante a nossa discussão que não são apenas os fatores estruturais que podem nos ajudar a compreender o fenômeno em variação. No que se refere às estratégias de relativização, há fatores extralinguísticos que apresentam relevância significativa para a compreensão dos processos.

Freitag (2009, p. 125) acrescenta, ainda, a necessidade de suporte teórico para a análise variacionista de variáveis sintáticas, visto que “é preciso buscar um paradigma teórico que dê pistas de como interpretar os achados quantitativos e que também possibilite uma análise qualitativa integrada”, ou seja, nesse caso, uma teoria sintática.

3 DECISÕES TEÓRICO-METODOLÓGICAS

Como já mencionado na seção 1, *Introdução*, e 2, *Fundamentação Teórica*, o trabalho tem como base teórico-metodológica a Teoria da Variação e Mudança Linguística, (WEINREICH, LABOV, HERZOG, 2006 [1968]; LABOV, 1972, 1994, 2001, 2010). Partindo do pressuposto de que a língua é um sistema de expressão e comunicação empregado por uma determinada comunidade de fala, os desvios a partir de um sistema homogêneo “não são todos eles erros aleatórios de desempenho, mas são num alto grau codificados e parte de uma descrição realista da competência de um membro de uma comunidade de fala” (WEINREICH et al., 2006 [1968], p. 60). Além disso, pontuamos, na seção 2.3.1.1, *A problemática da variável sintática*, um cenário teórico-metodológico que nos impôs alguns desafios para a compreensão do nosso fenômeno em sua totalidade e complexidade. Assim, se fizeram necessárias algumas decisões que serão explicitadas e descritas abaixo.

A primeira, já mencionada, foi o recorte do estudo em duas perspectivas: semasiológica e onomasiológica. A segunda, refletindo sobre um suporte teórico para a análise dos dados, nos pautamos em uma interface Sociofuncionalista, tendo em vista “a observação do aspecto social da variação e da mudança de modo mais preciso, incorporando em seu estudo, por exemplo, fatores interacionais ligados à negociação entre falante e ouvinte na situação comunicativa” (GÖRSKI, TAVARES, 2013, p. 88). A terceira, no que diz respeito à baixa frequência de variáveis sintáticas, optamos, para a construção do *corpus*, por utilizar amostras de fala provenientes de dois bancos de dados paulistas.

O primeiro é o banco de dados IBORUNA do projeto ALIP (GONÇALVES, 2007), que reúne um total de 152 entrevistas sociolinguísticas, estratificadas segundo a faixa etária, a escolaridade, o sexo/gênero e a renda do entrevistado. Além dessas características, o projeto ALIP tem uma particularidade que é de relevância para a análise: a entrevista, e a subsequente coleta de dados da presente pesquisa, foi estruturada em cinco modalidades textuais (Narrativa de Experiência, Narrativa Recontada, Descrição de Espaço, Relato de Procedimento, Relato de Opinião).

O segundo banco de dados é aquele do Projeto SP2010 (MENDES, 2013), estratificado segundo três variáveis sociais: sexo/gênero, faixa etária e níveis de escolaridade. Diferentemente do banco apresentado anteriormente, as entrevistas são conduzidas em duas partes: na primeira, o roteiro se define por tópicos como bairro, infância, família, trabalho e lazer; na segunda parte, o roteiro perpassa discussões sobre a cidade de São Paulo e questões

de produção e avaliação linguística, além de leitura de lista de palavras, notícia e trecho. Além destas informações, encontra-se no perfil social dos falantes as seguintes informações: geração da família, renda individual e renda familiar, e região e zona de residência. Para a coleta das ocorrências de **que** e **em que**, nos utilizamos do pacote *dmsocio* (OUSHIRO, 2018) na plataforma de programação R (CORE TEAM, 2021). Desse modo, após o tratamento do *corpus* em arquivos .txt sem tabulação, a partir do *script* *dmsocio*, buscamos pelos padrões: “\\bque\\b” e “\\bem\\sque\\b”¹⁸, delimitando que a busca não deveria ocorrer nas partes correspondentes à fala do entrevistador e outros participantes. Em um primeiro momento, foi feita a identificação dos padrões solicitados e depois uma extração. É válido destacar que possíveis erros podem surgir no processo, principalmente se o padrão não surge em uma determinada entrevista. Assim, é necessário retirá-la do conjunto para que o programa dê sequência ao processo solicitado.

As ocorrências de **onde** (em relativas com núcleo locativas) foram provenientes da pesquisa anterior (ver seção 2.1.2). Os dados foram refinados para que houvesse um conjunto de estratégias de relativização locativas e apenas em contextos preposicionados, o que levou à exclusão de orações relativas em posição sintática de sujeito e objeto direto (como a ocorrência em (3.1)). Tal decisão tem por base o valor semântico que estamos controlando na presente pesquisa e que surge em posições sintática mais baixas, como, por exemplo, adjunto adverbial do verbo dentro da oração encaixada (como a ocorrência em (3.2)).

(3.1) S1 mas... S1 tenho **uma outra amiga que mora em Guarulhos...** S1 então é isso... (SoraiaS, F3MC)

(3.2) S1 meu tio ficou pouco tempo em São Paulo e depois mudou pra Piracaia que é **uma cidadezinha do interior onde meus avós tinham uma chácara** (VivianeC, F1SC)

Em (3.1) o sintagma nominal “uma outra amiga” compõe a rede argumental do verbo “morar” na posição de sujeito, enquanto em (3.2) “uma cidadezinha do interior” é uma informação que está atuando como adjunto, visto que o verbo “ter”, no interior da oração relativa, já está completo. Dessa forma, orações como em (3.1) foram excluídas de nossa análise, enquanto orações como em (3.2) compõem nosso *corpus* de investigação.

A quantificação dos dados, assim como suas interpretações, foi feita com o auxílio da linguagem de programação R (CORE TEAM, 2021), utilizando o pacote *ggstatsplot* (PATIL, 2018), com amparo dos cursos e tutoriais de Freitag (2020, 2021) e Oushiro (2021).

¹⁸ Para efeitos de replicabilidade, informamos que quando a busca e extração for por apenas uma palavra, utilizar o padrão: \\b[palavra]\\b. Quando for duas palavras, utilizar: \\b[palavra1]\\s[palavra2]\\b.

As variáveis extralinguísticas controladas são provenientes dos bancos de dados, o que gerou os seguintes grupos de fatores: (i) sexo/gênero; (ii) escolaridade; (iii) faixa etária (presentes nos dois bancos); (iv) tipologia textual (presente no IBORUNA); (v) região de residência de São Paulo; (vi) zona de residência de São Paulo; (vii) roteiro da entrevista (presentes no SP2010). Para as variáveis linguísticas, controlamos: (i) função sintática do pronome relativo; (ii) adjacência entre o sintagma nominal antecedente e a oração relativa; (iii) classe morfológica do locativo; (iv) identidade semântica (ou tipo) do locativo; (v) classificação semântica da relativa (em explicativa ou restritiva); (vi) definitude do antecedente; (vii) especificidade do antecedente; e (viii) status informacional do antecedente.

Abaixo, descreveremos como se deu a constituição das variáveis dependente e independente, assim como quais eram nossas hipóteses por trás do controle das variáveis.

3.1 VARIÁVEL DEPENDENTE

Como mencionado anteriormente no presente trabalho, a nossa variável dependente é composta pelas estratégias de relativização locativas disponíveis no Português Brasileiro. Controlamos apenas os contextos preposicionados, o que definiu o seguinte envelope de variação:

- (3.3) **Padrão *pied-piping* ou padrão com preposição:** Doc.: (ruído) M. agora eu queria que você... me descrevesse um... um local Inf.: eu vou descrever **a casa em que eu estou morando...** (AC-050; DE: L. 193-194)
- (3.4) **Padrão com onde:** ah **o lugar onde eu trabalho...** é um salão tem... três partes... tem a parte da frente né? que cê entra assim é grandão aí tem dois degrau... (AC-024; DE: L. 243-244)
- (3.5) **Copiadora:** e um tipo **d'uma salinha pequenininha que elas guardava as coisas dela lá** e vendia né? (AC-015; DE: L. 639-640)
- (3.6) **Cortadora:** e depois dali nós fomo(s) pra festa que foi na minha... **na casa que eu morava na Pascoal More(i)ra...** e fizemos assim uma festa bastante simples (AC-092; NE: L. 38-40)

Em (3.3) observamos o uso da preposição “em” antes do pronome relativo, tal qual prescrito pela tradição gramatical. Em (3.4), adentramos nos casos pontuados como exceção pela norma padrão, isto é, pronomes relativos que não necessitam da preposição em contextos preposicionados. Assim, denominamos nossa variante como **padrão com onde**, visto que o valor semântico delimita essa estratégia. Na ocorrência em (3.5) encontramos um pronome cópia, ou lembrete, dentro da oração relativa, “lá” que retoma o sintagma antecedente, “salinha

pequeninha”, o que caracteriza a estratégia **copiadora**. Em (3.6), há um apagamento da preposição “em” que define a estratégia **cortadora**.

É necessário ressaltar que não foram encontradas estratégias **padrão com preposição** encabeçadas por outros tipos de pronome relativos preposicionados, tais quais **a/o qual, as/os quais, na/no qual, nas/nos quais, para a/o qual, para as/os quais**; assim como não houve ocorrências de **cujo** dentro de nossa amostra. Isso se deve ao contexto sintático-semântico que tomamos como alvo no presente estudo, o que justifica, também, nossa escolha por não utilizar em nosso *corpus* dados de **aonde**, tal pronome apresenta uma baixa frequência e mobiliza especificidades sintáticas e semânticas próprias que demandariam outros tipos de discussões.

Tendo em vista a literatura consolidada sobre a construção relativa, nossa hipótese era que haveria um maior uso das estratégias não padrão, **copiadora** e **cortadora**, em contraposição com a **padrão pied-piping**. Nos respaldando no trabalho anterior (seção 2), esperávamos um número grande de estratégia **padrão com onde** nos dois bancos de dados que concorreria com as estratégias não padrão.

3.2 VARIÁVEIS LINGUÍSTICAS

Tendo em vista que gostaríamos de produzir um espaço propício para debate e replicabilidade científica entre os pares, reproduzimos abaixo o *coding protocol* utilizado como guia em nossas análises. A tabela é composta por três colunas contendo: (i) os níveis das variáveis controladas, (ii) o código utilizado para preenchimento da planilha de dados e (ii) exemplos, ocorrências ou comentários a respeito da constituição do grupo de fatores.

Tabela 10 – Coding protocol utilizado para as variáveis linguísticas¹⁹

Função sintática do pronome relativo	Código	Exemplos
Complemento Circunstancial	complemento.circunstancial	você num vê... lá... lá na o(u)tra fazenda que eu morava ... cascavel coral jaracuçu de monte... aqui não
Quase argumento	quase.argumento	é as escolas que eu estudei sempre foram no bairro aqui escolas tradicionais
Adjunto Adverbial	adjunto.adverbial	tem um a a casa onde a minha mãe nasceu que é quando a minha vó éh... se caSÔ(u)
Adjunto Adnominal	adjunto.adnominal	fomo(s) no cinema que as cadeira mexe
Adjacência entre o sintagma nominal antecedente e a oração relativa	Código	Exemplos
Adjacente	encaixada	aí tem o escritório que fica meu computador
Distante	distante	uma sala da escola que foi destinada a uma psicopedagoga... onde ela trabalha com os alunos
Identidade morfológica do locativo	Código	Exemplos
Nome próprio	nome	eu vou todo ano pra Califórnia onde meu filho mora às vezes eu vou duas vezes por ano
Substantivo	substantivo	eu gosto de brincá(r)... num lugar que tem bastante Árvores...
Advérbio	advérbio	tudo minhas coisa tão ali... meu rádio... então é ali que eu fico a maior parte do tempo assim
Pronome	pronome	cima do no sobradinho em cima... tem três quartos... um que meu pai e minha mãe fica
Identidade semântica (ou tipo) do locativo	Código	Exemplos
Locativo é um objeto	objeto	tem o armário pequeno... que acomoda lá... prataria...
Locativo é um lugar	lugar	tem um barracão onde guarda os trator
Classificação da relativa	Código	Exemplos
Explicativa	explicativa	lá no rancho do meu genro onde tem o barco lá
Restritiva	restritiva	quando ele enTRÔ(u) dentro da sala que eu nasci
Definitude do SN	Código	Exemplos

¹⁹ Algumas variáveis podem se mostrar complexas para a análise estatística, em uma primeira leitura. Contudo, quando necessário, foram feitos os devidos “amalgamentos”, explicitados na seção 6: *A sintaxe dos paulistas: análise de produção*. Nossa justificativa para tantos níveis de análise se deu pelo interesse em conhecer o corpus em sua integridade, buscando possíveis padrões de variação.

Definido	definido	tem o meu quarto... que eu durmo numa cama de casal
Indefinido	indefinido	é um lugar que você vai pra jogar
Especificidade do SN	Código	Exemplos
Específico	especifico	aos po(u)cos fui construindo a casa onde eu estô(u) morando atualmente
Não específico	nao.especifico	eu num quero tê(r) uma horta onde eu pago alguém que cuide pra mim
Genérico	generico	qualquer esquina aí que cê anda cê vê... um fuman(d)o
Status informacional do antecedente	Código	Comentários
Nova em folha	nova.folha	A informação nova não consta até o momento [ver exemplo 3.48]
Não usada	nova.nao.usada	A informação nova faz parte do tópico discursivo [ver exemplo 3.49]
Inferência contida	inferivel.contida	A informação é pressuposta, isto é, o falante acredita que ela é acessível ao ouvinte, e pode ser recuperada dentro do próprio sintagma nominal [ver exemplo 3.50]
Inferência não contida	inferivel.nao.contida	A informação é pressuposta através de <i>frames</i> [ver exemplo 3.51]
Textualmente evocada	evocada.textualmente	A informação faz parte do tópico discursivo e já foi utilizada [ver exemplo 3.52]
Situacionalmente evocada	evocada.situacionalmente	A informação faz parte do tópico discursivo e pode ser recuperada através de uma relação dêitica [ver exemplo 3.53]

Fonte: própria

3.2.1 Função sintática do pronome relativo

A primeira variável linguística controlada é a **função sintática do pronome relativo**, isto é, a relação sintática estabelecida pelo antecedente na rede argumental do verbo da oração relativa, relação que é correferenciada pelo pronome relativo. Em um primeiro momento, optamos pela denominação desse grupo de fatores como **função sintática do antecedente**, porém uma imprecisão metodológica surgiu, visto que essa designação é utilizada nas análises funcionalistas anteriores quando o verbo controlado está na oração não encaixada. Assim, vale ressaltar que toda a nossa análise foi construída com enfoque no antecedente, não no pronome relativo, mas o nome da variável visa evitar uma confusão entre os pares e a literatura anterior.

Ao contrário do que prescreve a gramática normativa, o locativo não funciona como um termo acessório em todos os seus usos. Isto é, não poderíamos categorizá-lo como um adjunto

adverbial em todas as ocorrências porque observamos padrões diferentes em nossos dados. Para justificar o primeiro deles, tomamos por base Rocha Lima (1972), que diferencia o adjunto adverbial de complemento circunstancial. Esse termo da oração se caracteriza como um complemento do predicator, compondo a rede argumental do verbo, na medida em que o coloca em uma “circunstância” – nesse caso, em um lugar. Assim, ele se comporta como um argumento e não um adjunto.

O segundo padrão diz respeito a uma gradação observada entre as funções sintáticas: verbos como **estudar**, **trabalhar** e **ter** (existencial) não necessitam, em princípio, de um argumento locativo dentro de sua rede argumental; porém, de alguma forma, o lugar não é utilizado como um adjunto prototípico dentro da oração, sendo necessário para cumprir as funções comunicativas mobilizadas no ato de fala. Para esses casos, denominamos o nível como **quase argumento**, tendo como objetivo compreender melhor o seu comportamento. Abaixo, em (3.7)-(3.18) apresentamos ocorrências que ilustram essa discussão.

- (3.7) **Adjunto adverbial:** S1 ai é **um lugar que minha tia tem uma chácara** que agora me fugiu (JanainaB, F1MC)
- (3.8) **Adjunto adverbial:** S1 isso ou pro **interior de São Paulo onde o meu tio mora na chácara** eu às vezes eu vou pra lá (DouglasA, M1MP)
- (3.9) **Adjunto adverbial:** atravessa uma ponte gigantesca prime(i)ro né? **que a ponte... que eu tenho pavor de passá(r) nela** (AC-026; DE: L. 170-171)
- (3.10) **Adjunto adverbial:** S1 era **um lugar em que a gente brincava na terra** jogava bolinha de gude (JorgeV, M3SC)
- (3.11) **Quase argumento:** S1 não foi **nessa empresa que eu trabalho agora...** porque eu estou lá há quinze anos (RenataC, F2MC)
- (3.12) **Quase argumento:** eu queria da Sabará pra frente tipo S1 lá pra **civilização lá onde tem banco aonde tem shopping aonde tem tudo** (JanainaB, F1MC)
- (3.13) **Quase argumento:** S1 e era na **Normandia que que eu brincava bastante ali** naquele período com a molecada (RomuloS, M3MC)
- (3.14) **Quase argumento:** S1 iguaizinhos... em cima deste aqui... **tem um quarto em que eu tenho inclusive um monte de coisa velha do meu pai e da minha mãe** (FelixL, M3MC)
- (3.15) **Complemento circunstancial:** S1 gosto muito do bairro e... **o pedaço que eu moro** eu acho muito bom (VivianeC, F1SC)
- (3.16) **Complemento circunstancial:** S1 poderia trabalhar por exemplo pra me/ melhorar **o bairro onde eu moro** né eu estou lá S1 minha casa é lá... (SilviaB, F1SP)
- (3.17) **Complemento circunstancial:** S1 me levou lá pra pra S1 pro **bairro que ele morava lá** pra gente jogar sinuca tomar cerveja (RodolfoR, M3MP)
- (3.18) **Complemento circunstancial:** S1 Brasil digamos é um país simpático D1 uhum S1 lá fora pra **todo lugar em que a gente ia** S1 eh eles gostam muito do Brasil (NiltonR, M3SC)

Em (3.7), (3.8), (3.9) e (3.10) os pronomes relativos **que**, **onde** e **em que** correferenciam a função sintática de adjunto adverbial do antecedente, dado que “um lugar” e “interior de São Paulo”, “a ponte” estão funcionando como adjuntos de “ter”, “morar” e “brincar”. Dentro da oração relativa, esses verbos encontram sua rede argumental completa com o argumento externo “minha tia”, “meu tio”, “eu” e “a gente” e argumento interno “uma/na chácara”, “nela” e “na terra”. Em (3.11), (3.12), (3.13) e (3.14) observamos que os verbos “trabalhar”, “ter” e “brincar” não “pedem” um argumento locativo, mas os antecedentes “nessa empresa”, “civilização lá”, “Normandia” e “um quarto” completam as funções comunicativas mobilizadas na fala do entrevistado. Em (3.15), (3.16), (3.17) e (3.18) temos o verbo “morar” e “ir” que precisa de um complemento circunstancial: o antecedente correferenciado pelos pronomes, isto é, “o pedaço”, “o bairro” e “lugar”, respectivamente.

Temos como hipótese que a escolha entre as estratégias de relativização se relaciona com a função sintática do pronome relativo e o tipo de encaixamento da sentença. Em contextos de complemento observamos um vínculo forte com o verbo, já que a sua rede argumental apresenta essa semântica locativa. Assim, há um compartilhamento entre o antecedente e o verbo do sentido locativo. Essa característica permitiria um maior emprego da **cortadora**. Por outro lado, quando temos a função de adjunto adverbial há um vínculo um pouco mais fraco, há um maior grau de independência em relação ao verbo, principalmente semântica, em uma relação de hipotaxe (NEVES, BRAGA, 2016). Assim, ficaria para o **onde** dar o sentido locativo (ou reforçá-lo) na oração. No caso de quase argumento, teríamos um processo intermediário: não há um compartilhamento do valor locativo entre o verbo e o antecedente, porém ele ainda se comporta quase como um argumento do verbo. Esquematizando, temos que:

- i. quando **complemento circunstancial** a oração será [+ **encaixada**/ + **dependente**]
- ii. quando **quase argumento** a oração será [+ **encaixada**/ - **dependente**]
- iii. quando **adjunto adverbial** a oração será [- **encaixada**/ - **dependente**]

3.2.2 Adjacência entre o sintagma nominal antecedente e a oração relativa

A segunda variável linguística, a **adjacência entre o sintagma nominal antecedente e a oração relativa**, perpassa o fato de que a construção relativa é construída por meio de uma oração, encabeçada por um pronome relativo, encaixado em um antecedente. Assim, é necessário controlar os casos “atípicos”, isto é, em que há alguma informação entre o antecedente e a oração. Para ilustrar, temos as ocorrências (3.19)-(3.25):

- (3.19) **Adjacente:** “foi pra Disney pra Pousada do Rio Quente... **lugares assim**²⁰ **que eu queRIA tê(r) ido** sabe?” (AC-012; NR: L. 71-72)
- (3.20) **Adjacente:** “e meu tio foi saí(r) do **prédio onde mora a mãe dele** lá num num lugar... e meu primo também foi junto...” (AC-001; NR: L. 63-65)
- (3.21) **Adjacente:** nós tem uma made(i)ra igual **uma lo(u)sas que só pode pô(r) ali** (AC-003; DE: L. 270)
- (3.22) **Adjacente:** eu vou descrever a **casa em que eu estou morando...** (AC-050; DE: L. 194)
- (3.23) **Distante:** “**numa bica de água** que tinha separado ali... **que as mulheres lavavam ro(u)pa...**” (AC-102; NR: L. 147-148)
- (3.24) **Distante:** “foi colocado **os armários** todos com cadeados bem trancados... **onde esses violinos são guardados** e aonde... também eles fazem aulas de violino...” (AC-088; DE: L. 430-432)
- (3.25) **Distante:** **tem a Praça Pereira Coutinho** que é uma delícia **que eu ando todo o santo dia com o cachorro lá** (RomuloS, M3MC)

Em (3.19), (3.20), (3.21) e (3.22) identificamos o antecedente, “lugares assim”, “prédio”, “uma lousa” e “a casa” respectivamente, seguidos imediatamente pela oração relativa encabeçada pelo pronome relativo. Por outro lado, em (3.23), (3.24) e (3.25), os sintagmas nominais “uma bica de água”, “os armários” e “a Praça Pereira Coutinho” apresentam as informações “que tinha separado ali”, “todos com cadeados bem trancados” e “que é uma delícia” antes de as orações relativas “as mulheres lavavam roupa”, “esses violinos são guardados” e “eu ando todo o santo dia com o cachorro lá” serem introduzidas pelo pronome²¹, isto é, há uma descrição inserida antes da oração propriamente dita.

Averiguar em detalhes os contextos em que há informações inseridas entre o antecedente e a oração relativa nos direcionam para uma discussão de acessibilidade da informação (ARIEL, 1988, 1991), tendo em vista que a distância pode ser um dos fatores associados no processamento da construção pelo ouvinte. Assim, nossa hipótese é de que em contextos caracterizados como **distante** haveria um uso maior da estratégia **padrão com onde**, visto que o pronome relativo recupera o valor locativo que poderia ser difícil de recuperar pelo entranhamento de outras informações entre esse e o sintagma nominal antecedente. Assim, quando identificássemos a **adjacência**, haveria um uso maior da **cortadora**, dado que o valor locativo estaria mais facilmente recuperável.

²⁰ Elementos que compõem o sintagma nominal antecedente, sejam eles marcadores discursivos, substantivos ou adjetivos, não foram considerados como informações adicionais que culminam em casos “distantes”.

²¹ Em ambos os bancos de dados não houve ocorrência de padrão com preposição “distante” do antecedente.

3.2.3 Classe morfológica do locativo

Para a nossa terceira variável linguística, **a classe morfológica do locativo**, controlamos se o antecedente seria um substantivo, um nome próprio, um advérbio ou um pronome, tal como as ocorrências (3.26), (3.27), (3.28) e (3.29):

- (3.26) **Substantivo**: aquele posto tinha uma... **uma muretinha que os rapazes ficavam sentados ali** (VivianT, F3SC)
- (3.27) **Nome próprio**: “São Paulo ter se tornado uma megalópole monstruosa **em que ahh... trânsito é uma loucura**” (FelixL, M3MC)
- (3.28) **Advérbio**: “e quando foi na tarde do sábado... ele apareceu **lá onde eu tava morando...**” (AC-076; NE: L. 54-55)
- (3.29) **Pronome**: “Doc: cê poderia descrevê(r) pra mim assim como que é é cada uma das casa que você trabalha? Inf: posso... eu tenho **uma que eu trabalho** ela é muito boa... gostosa de limpá(r)... pequena a casa...” (AC-032; DE: L. 111-114)

Em (3.26) o sintagma nominal tem como núcleo um substantivo, “uma muretinha”, ao qual a oração relativa se liga, enquanto em (3.27) o antecedente é composto por um nome próprio, isto é, São Paulo. Os casos em (3.28) e (3.29) surgem quando o antecedente é um advérbio de lugar, “lá”, ou um pronome indefinido, “uma”.

Nossa hipótese partiu da premissa de que nomes próprios não apresentariam alta frequência da estratégia **padrão com onde**, visto que o valor locativo já estava devidamente expresso e o **onde**, com sua carga semântica de lugar, poderia ser redundante; assim, esse seria um contexto propício para o uso da **cortadora**. Quando o antecedente fosse um substantivo, teríamos uma frequência mais alta de **padrão com onde** para, justamente, reforçar o valor locativo.

3.2.4 Identidade semântica (ou tipo) do locativo

Nossa quarta variável linguística, **a identidade semântica (ou tipo) do locativo**, foi construída a partir de uma diferença aparente no *corpus*, proveniente, em especial, do banco de dados IBORUNA. Isso, pois o antecedente pode se referir a um local, assim como a um objeto, dado que a tipologia textual Descrição de Espaço, que compõe uma das partes da entrevista, propicia um momento em que o falante faz referências a espaços físicos e os elementos que o compõem, tal qual as ocorrências (3.30)-(3.37):

- (3.30) **Local:** “a gente andô(u) no barco viking... é... fomo(s) **no cinema que as cadê(i)ra mexe...** que era muito legal” (AC-37; NE: L. 32-33)
- (3.31) **Local:** “eu vô(u) falá(r) de como é o **o lugar onde eu trabalho...** lá é lá é a casa da minha patroa...” (AC-016; DE: L. 177-178)
- (3.32) **Local:** S1 tem **um bairro na zona leste... que estão todos concentrados ali...** (ArianeG, F2SP)
- (3.33) **Local:** S1 isso porque **nos países em que eu visitei** sempre uma das coisas que uma das coisas qu ah ah você chega a primeira coisa é a "aonde ficam os brasileiros" entendeu? (AnselmoN, M3SP)
- (3.34) **Objeto:** “ah tem a estante e **um negócio de vidro lá que minha mãe guarda os copo** remédio essas coisa” (AC-014; DE: L. 177-178)
- (3.35) **Objeto:** “tem uma estante pequenininha... **uma cômoda onde fica minhas coisa** em cima...” (AC-024; DE: L. 224-225)
- (3.36) **Objeto:** tem **o armário pequeno... que acomoda lá... prataria...** o de uso porque... num cabe tudo (AC-080; DE: L. 144-145)
- (3.37) **Objeto:** aquilo num é escrivaninha é **uma mesa mal arrumada... em que eu estudo...** (AC-050; DE: L. 229-230)

Em (3.30), (3.31), (3.32) e (3.33) o antecedente no qual a oração relativa está encaixada é um local físico que está sendo descrito pelo falante; enquanto em (3.34), (3.35), (3.36) e (3.37) o antecedente se trata de um lugar não prototípico, nesse caso, um objeto.

Tal qual a variável anterior, hipotetizamos que uma diferença entre os usos das estratégias de relativização poderia surgir em relação ao tipo do locativo. Assim, uma frequência maior da **padrão com onde** poderia ser encontrada nos contextos em que o falante fazia referência a objetos, em razão do valor locativo do pronome relativo.

3.2.5 Classificação semântica da relativa

Uma abordagem clássica que tem sido dada à construção é a sua classificação por meio de dois tipos: a relativa restritiva ou a relativa explicativa. A primeira designa um processo de restrição feita pela oração encaixada, isto é, quando ela atribui uma informação que delimita o antecedente dentro de um determinado conjunto; enquanto a segunda trata do contexto em que a oração relativa tem como função adicionar uma informação suplementar ao antecedente a que se liga. Nas ocorrências (3.38), (3.39), (3.40) e (3.41) podemos observar essa diferença:

- (3.38) **Restritiva:** “éh tem um lugar lá que abriu faz muito tempo mas que é um lugar cheio de computador mas num é l/lan house (assim) que nem tem aqui é **um lugar lá que cê vai... pra jogá(r) computador...** aí vai um monte de gente” (AC-001; DE: L. 191-194)
- (3.39) **Restritiva:** S1 era **um bairro que eu também eu fiquei pouco tempo lá** um ano (LeonS, M2SC)

- (3.40) **Explicativa:** “agora eu acho que eu posso confundí(r) se foi lá/ se foi numa vez que ele foi jogá(r) que a gente foi em várias...foi em várias... ou se foi lá em **Santa Catarina que a gente foi em várias**” (AC-006; DE: L. 254-256)
- (3.41) **Explicativa:** sociedade de amigos é **uma sociedade de prédios em que se tem uma entidade** que visa à segurança dos condomínios (NiltonR, M3SC)

Em (3.38), ao descrever um espaço que costuma frequentar, o falante restringe o antecedente “um lugar” para aquele em que se vai para fazer uma determinada atividade. O mesmo ocorre em (3.39) quando o falante restringe o bairro para aquele em que morou por um período curto. Em (3.40), por outro lado, ao tecer alguns comentários a respeito de viagens que costuma fazer, o entrevistado aponta que pode ter sido em “Santa Catarina” as várias localidades que visitaram, produzindo uma informação adicional a cidade, mas não a caracterizando dentro de um conjunto maior. De forma similar, em (3.41) o falante, ao descrever a associação entre os prédios, explica a sua função de fornecer segurança ao condomínio, sem restringir o antecedente dentro de uma categoria maior.

Tendo em vista estudos anteriores que controlaram a variável, é de nosso interesse compreender se há alguma relação entre os dois tipos expostos e os usos das estratégias de relativização locativas. Isso, pois as demais investigações não fazem distinção entre possíveis categorias semântico-cognitivas, o que nos impede de ter um espaço de comparação efetivo para as relativas locativas.

3.2.6 Definitude, especificidade e status informacional do antecedente

No que se refere às últimas três variáveis linguísticas controladas, consideramos adequado apresentá-las em conjunto, visto que se trata de categorias semântico-pragmáticas. Desse modo, são responsáveis por caracterizar o antecedente em uma perspectiva semântica e discursiva, tendo em vista uma necessidade de compreender se esses elementos envolvidos no ato comunicativo poderiam estar atuando na escolha de uso das estratégias de relativização.

Assim, a definitude se concretiza pelo controle do tipo de determinante que precede o antecedente, podendo ser definido ou indefinido. Segundo Neves (2018, p. 407), uma das propriedades primordiais de um artigo, definido ou indefinido, é fazer referência, atuando no processo de identificação do substantivo a ele associado. O primeiro, definido, se caracteriza pelo processo de determinar um substantivo comum, isto é, particularizando-o dentro de uma classe maior, enquanto o segundo produz o efeito oposto: é comumente utilizado em contextos

nos quais não há a intenção de particularizar uma situação, uma pessoa ou coisa, ou seja, uma classe.

Algumas particularidades surgem quando observamos que o artigo definido é utilizado pelo falante quanto este assume que a informação é, de alguma forma, compartilhada ou conhecida com o seu interlocutor, podendo, desse modo, ser identificável. Dessa forma, surge a noção de identificabilidade, em que o elemento tem como função construir uma informação que corresponde a um objeto no mundo do qual se pretende falar (BRAGA, ILARI, OLIVEIRA, BASSO, 2015). Também é descrito por autores que o artigo indefinido pode ser o responsável por fazer referências a elementos que não participam da situação evidenciada no texto, assumindo uma função de elemento não referencial, visto que serve como um “alertar” ao interlocutor de que será introduzida uma informação nova no discurso (BRAGA, ILARI, OLIVEIRA, BASSO, 2015; NEVES, 2018).

É necessário, porém, produzir algumas ressalvas: essas descrições não podem ser tomadas como categóricas, já que “a multiplicidade de arranjos construcionais que o fazer da linguagem aciona para a expressão dos múltiplos significados no texto e para a produção dos efeitos discursivos” (NEVES, 2018, p. 415) gera ocorrências em que o artigo definido pode introduzir uma nova informação no discurso e o artigo indefinido ser o responsável por trazer uma informação já mencionada anteriormente. Assim,

a análise de amostras de fala produzidas em situações ordinárias de comunicação revela, no entanto, que nem todas as peças de informação se encontram totalmente explicitadas quando ocorre o uso de artigos definidos; em outras palavras, durante a fala, as entidades sobre as quais os interlocutores discorrem podem ser objeto de inferências processadas por eles. Essas inferências podem ser desencadeadas quer por entidades presentes no texto que precede, [...], quer por entidades presentes na situação de fala [...] (BRAGA, ILARI, OLIVEIRA, BASSO, 2015, p. 116)

Desse modo, optamos por controlar o tipo de informação codificada, isto é, se nova ou velha dentro da situação comunicativa, através da variável status informacional (que será descrita em detalhes abaixo). No que se refere à definitude do antecedente, utilizamos a combinação de duas variáveis para compor nossa análise, sendo (i) tipo de determinante utilizado (dimensão textual-discursiva) e (ii) a especificidade do antecedente (dimensão semântica), criando os seguintes padrões:

Tabela 11 – Padrões utilizados na análise da definitude do antecedente

	Determinante	Especificidade
<i>Definido</i>	Artigo definido (o(s), a(s)), demonstrativo (este(s), esta(s), isto, esse(s), essa(s), isso, aquele(s), aquela(s), aquilo), possessivo (o(s), a(s), meu(s), teu(s), seu(s), nosso(s), vosso(s)), todo(s)	Específico, não específico, genérico
<i>Indefinido</i>	Artigo indefinido (um, uma, uns, umas), pronome indefinido (qualquer, outro, algum, vários), sintagma nominal “nu”	Específico, não específico, genérico

Fonte: própria

Se faz necessário detalhar o que estamos considerando como *especificidade do antecedente*, variável composta por três níveis: específico, não específico e genérico. O primeiro nível corresponde ao antecedente que particulariza um determinado elemento de uma categoria, reconhecendo-o como único dentro de uma situação ou referência. O segundo faz referência ao antecedente que se aplica a todo e qualquer membro de grupo ou tipo. Por fim, o terceiro nível parte de uma abrangência da categoria em sua extensão máxima, nomeando a classe como um todo (WILMET, 1986; BERLINCK, 1995). Em (3.42), (3.43), (3.44), (3.45), (3.46) e (3.47) observamos algumas ocorrências:

- (3.42) **Definido específico:** “ela morava numa casa... que não é **a casa que eu moro hoje...** e ni/ na época meu pai era caminhone(i)ro... então ele viajava ele ficava... várias semanas fora de casa...” (AC-055; NR: L. 91-93)
- (3.43) **Definido não específico:** “a gente tem um termo que... que o PAI ou o responsável **na origem... onde essa criança... embarcô(u)** a/assina tudo os termo tudo a documentação” (AC-051; RP: L. 409-411)
- (3.44) **Definido genérico:** S1 nos Estados Unidos eu fui tudo... **todos os lugar que ela foi lá estava eu** (DarcyM, F3MP)
- (3.45) **Indefinido específico:** “mas... é **um lugar onde minha irmã morava** já faz mais de dez ano faz doi anos que eu moro aqui...” (AC-078; DE: L. 178-179)
- (3.46) **Indefinido não específico:** “eu vô(u) descrevê(r) um um um:: uma coisa que eu gostaria muito de tê(r) na minha vida ainda e VÔ(u) tê(r)... éh **uma chácara... onde eu pudesse... talvez até morá(r)** nessa chácara dependen(d)o a localização dela...” (AC-099; DE: L. 197-199)
- (3.47) **Indefinido genérico:** S1 Brasil digamos é um país simpático D1 uhum S1 lá fora pra **todo lugar em que a gente ia** S1 eh eles gostam muito do Brasil (NiltonR, M3SC)

Em (3.42), (3.43) e (3.44) observamos dois tipos de determinantes definidos: “a” e “os”. O que diferencia as orações é que em (3.42) o sintagma nominal trata de uma casa em particular, enquanto em (3.43) o antecedente faz referência a qualquer cidade que configure o ponto de origem e (3.44) designa todo um grupo ou categoria, isto é, faz uma menção genérica. Em

(3.45), (3.46) e (3.47) observamos outros dois tipos de determinantes, sendo eles indefinidos: “um(a)” e “todo”. A diferença surge, pois (3.45) descreve um lugar específico em que sua irmã residia e (3.46), por outro lado, descreve uma chácara hipotética, isto é, ela pode vir a ser sua propriedade, mas, no momento, se configura como um membro qualquer dentro do grupo; por fim, (3.47) trata de forma genérica de locais que o falante percorreu em sua viagem.

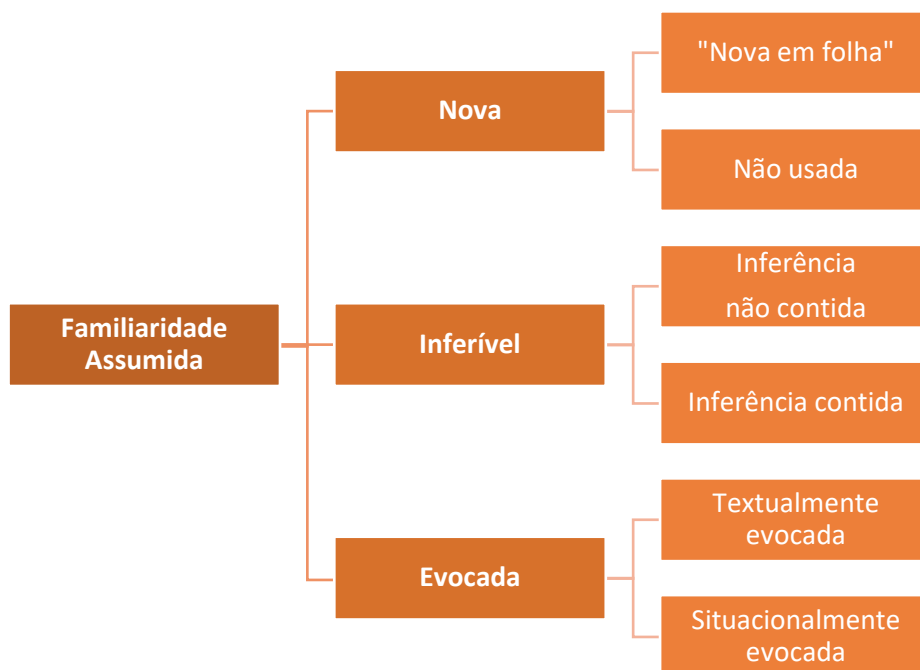
Retomando a problemática exposta sobre o status informacional do antecedente, estamos assumindo que “o empacotamento de informações em línguas naturais reflete as hipóteses do emissor sobre as suposições, crenças e estratégias do receptor” (PRINCE, 1981, p. 224 [tradução nossa]²²). Para investigar tal proposição, segundo Prince (1981), existem três linhas principais que foram exploradas por inúmeros trabalhos linguísticos, sendo elas: (i) a “novidade” da informação a partir de sua recuperabilidade e previsibilidade; (ii) a “novidade” a partir da saliência da informação e (iii) a “novidade” a partir do conhecimento compartilhado.

O presente trabalho adota a proposta teórico-metodológica de Prince (1981) que busca ampliar a noção da terceira linha, associando as outras duas e culminando em um aporte abarcando os três níveis. Para isso, a autora atualiza o termo “conhecimento compartilhado” para “Familiaridade Assumida”, visto que “não tem as conotações inúteis de simetria e fato e não soa como qualquer outra coisa” (PRINCE, 1981, p. 233 [tradução nossa]²³). Na Figura 2 apresentamos, assim, a taxonomia dos valores da Familiaridade Assumida.

²² Original: “[...] information-packaging in natural language reflects the sender's hypotheses about the receiver's assumptions and beliefs and strategies.”

²³ Original: “it lacks the unhelpful connotations of symmetry and fact and does not sound like anything else.”

Figura 2 – A taxonomia dos valores da Familiaridade Assumida



Fonte: adaptação de Prince (1981, p. 237)

De acordo com Prince (1981), se a informação é introduzida pela primeira vez no discurso, devemos classificá-la como nova e alocá-la dentro de dois tipos; isso, pois a informação nova pode ser a criação de uma entidade que não constava até então – isto é, “nova em folha” – ou uma entidade que corresponde ao modelo, referindo-se ao tópico discursivo em questão, e é colocada ou copiada dentro do discurso – isto é, “não usada”. Para exemplificar, observemos as ocorrências em (3.48) e (3.49):

- (3.48) **Nova em folha:** “Inf.: ah:: teve uma vez que eu e meus colega a gente devia tá mais ou menos em QUAtro... a gente combinô(u) de í(r) comê(r) Plizza... numa pizzaria que fica lá na avenida Alberto Andaló... chama Portugália... aí tudo bem::... aconteceu tudo normal:: a gente comeu a pizza ficô(u) cheio... a gente resolveu í(r)... pra represa a pé... aí no/ a gente foi no percurso a gente foi zuan(d)o numa boa a gente decidiu... parar... numa loja/ **num estabelecimento que vendia bebida alcoólica...** Doc.: hum Inf.: ... a gente pegô(u) e comprô(u) uma::... uma vodca né?” (AC-055; NE: L. 05-11)
- (3.49) **Nova não usada:** “Doc.: tem algum outro lugar que você possa me descrevê(r)? Inf.: é tem um/ Doc.: pode sê(r) seu quarto um lugar que você acha bonito Inf.: tem esse/ Doc.: um lugar que você já foi S1 **um lugar que eu fui muito bonito também?**... éh é no A.C.T. que é no Centro de Treinamento de Futebol de Mirassol né?” (AC-035; DE: L. 307-313)

Em (3.48) o sintagma nominal “um estabelecimento” é introduzido pela primeira vez no discurso, configurando, assim, uma informação “nova em folha”; enquanto em (3.49), o

antecedente “um lugar”, apesar de estar sendo mencionado como uma informação nova, trata-se de um elemento que faz parte do tópico discursivo, o que o configura como informação “não usada”. Desse modo, as ocorrências que compõem os dois níveis são aquelas (i) introduzidas pelo falante sem nenhuma menção anterior – tanto dele quanto do entrevistador (“nova em folha”) e (ii) colocadas no ato comunicativo a partir de um tópico especificado por um dos participantes e nomes próprios mencionados pela primeira vez (“não usada”).

Dando sequência, temos uma informação inferível quando o falante pressupõe que o ouvinte pode inferi-la através de raciocínio lógico, ou de um conhecimento plausível, a partir de uma entidade que já foi evocada dentro do discurso. Podemos dividir essa categoria em inferência “contida”, quando a informação pode ser recuperada dentro do próprio sintagma nominal, isto é, quando faz parte de um conjunto, e inferências “não contidas” – aquelas que não podem ser recuperadas dentro do próprio sintagma nominal. Incluímos nesse nível antecedentes que eram compostos por “todo lugar” e/ou “qualquer lugar”, assim como situações em que o falante fazia referência a elementos de *frame*; isto é, toda vez que descrevia espaços escolares ou residenciais e elencava itens em comum dentro desses espaços, classificamos as informações como inferências não contidas. As ocorrências em (3.50) e (3.51) ilustram os dois casos:

- (3.50) **Inferível contida**: “em cima do no sobradinho em cima... tem três quartos... **um que meu pai e minha mãe fica** ((barulho de criança na casa da vizinha)) um do meu irmão e um meu...” (AC-008; NE: L.: 106-108)
- (3.51) **Inferível não contida**: “nesse refeitório... foi colocado uma televisão... onde as crianças podem comer e assistí(r) a televisão enquanto... não bate o sinal pra elas irem pra fila... **tem o pátio... onde as crianças ficam no recreio...** o pátio tem mais ou menos a largura da escola toda... né?... é um pátio muito grande” (AC-088; DE: L. 368-371)

Em (3.50) o antecedente “um” pode ser recuperado da oração precedente “tem três quartos”, configurando um contexto de que a informação é inferível a partir de um conjunto. Por outro lado, em (3.51) “o pátio” não se apresenta como parte de um grupo mencionado de forma próxima ou dentro do próprio sintagma nominal. A informação é inferível, pois o falante está tecendo descrições a respeito da escola, o que evoca um *frame* com objetos e espaços contidos dentro do ambiente escolar.

Por fim, consideramos uma informação como evocada quando ela faz parte do modelo de discurso, podendo surgir, em um primeiro momento, a partir de bases textuais, seguindo instruções do falante, isto é, através de uma informação nova ou inferível inserida anteriormente

no discurso, ou seja, “evocada textualmente”; ou, em um outro momento, o interlocutor pode evocá-la de forma extratextual ou por meio de dêiticos dentro da situação que se descreve, ou seja, “evocada situacionalmente”. Em (3.52) e (3.53) podemos identificar os dois contextos:

- (3.52) **Evocada textualmente:** “eu gosto de ficá(r) no meu quarto quando eu tô triste... eu vô(u) pro meu quarto e choro... choro assim deitada na cama num quero falá(r) com ninguém então meu quarto é meu refugio... tô triste vô(u) pro meu quarto... (inint.) que um num largo... é o canto/ que eu fico é a última parte pra mim me escondê(r)... é o qua/ é o lugar onde que eu fico triste eu... eu tô triste d’uma coisa... é lá que eu fico Doc.: e o quê que tem nele assim que te chama atenção? Inf.: eu num sei num sei se é **o lugar onde eu tenho meu esPAço** onde as pessoas não me invadem que eu guardo minhas COIsas... que ninguém MExe...” (AC-056; DE: L. 147-154)
- (3.53) **Evocada situacionalmente:** Inf.: bom... a minha casa... assim... é duas salas né? Doc.: uhum (concordando) Inf.: como cê tá vendo aqui...duas salinhas (fala rindo)... **nessa sala que nós tamo(s)** tem... uma cortina bem grande... não muito grande” (AC-066; DE: L. 287-289)

Em (3.52), ao descrever o seu quarto, o informante explica a relação de conforto que se estabelece com o ambiente e utiliza alguns elementos para designar essa relação. Assim, observamos que o antecedente “o lugar” já foi mencionado outras vezes e está sendo evocado textualmente na oração negritada. Desse modo, a noção que é mobilizada para esse nível da variável é uma informação que foi dada, sendo o mesmo referente apresentado anteriormente. O processo em (3.53) se dá de forma diferente: o sintagma nominal “nessa sala” carrega a informação dêitica de que o informante e o entrevistado compartilham o ambiente extratextual, fazendo com que a informação seja evocada situacionalmente.

O controle das três variáveis descritas acima perpassa uma necessidade de compreender a relação que existe entre a semântica e o discurso para a acessibilidade do referente (ARIEL, 1988, 1991). Isso, pois

As menções subsequentes de referentes no discurso são feitas para representações mentais que são relativamente mais acessíveis. Uma vez evocadas, essas entidades de memória devem permanecer ativadas, pelo menos por um tempo. As referências a “novas” entidades, por outro lado, recuperam representações mentais em um armazenamento mais profundo. (ARIEL, 1991, p. 445 [tradução nossa]²⁴)

O que a autora propõe é que as informações “novas” apresentam uma demanda cognitiva maior do que um referente que já foi devidamente evocado no discurso. Assim, esse antecedente “velho” se torna mais acessível ao ouvinte. Desse modo, controlar as três variáveis expostas

²⁴ Original: “Subsequent mentions of referents in discourse are made to mental representations which are relatively more accessible. Once evoked, such memory entities must remain activates, at least for a while. References to ‘new’ entities, on the other hand, retrieve mental representations in deeper storage.” (ARIEL, 1991, p. 445)

nos ajudaria a propor uma discussão sobre como um sintagma composto de [+ definido], [+ específico] e [+ acessível] poderia exigir menos cognitivamente do ouvinte e moldaria uma frequência maior de uma estratégia de relativização em detrimento de outra.

Nossas hipóteses eram que haveria uma frequência maior de uso de **padrão com onde** em contextos indefinidos e não específicos, visto que a carga semântica locativa de **onde** poderia trazer um reforço para a descrição ou narração do falante. Quanto ao status informacional, hipotetizamos que haverá um maior número de informações evocadas (textualmente e situacionalmente), seguidas de novas não usadas, inferidas não contidas, inferidas contidas e novas em folha – seguindo a hierarquia proposta por Price (1981, p. 245). Também consideramos que, em contextos de informações novas teremos uma frequência maior da **padrão com onde**, visto que o novo é menos acessível e demanda maior esforço cognitivo. Assim, o valor locativo do pronome deveria contribuir para reforçar o valor semântico da oração e contribuir para o seu processamento.

3.3 VARIÁVEIS EXTRALINGUÍSTICAS

Para a apresentação das variáveis extralinguísticas, gostaríamos de reproduzir o *coding protocol* utilizado como guia em nossas análises. A tabela é composta por três colunas contendo: (i) as variáveis controladas, (ii) o código utilizado para preenchimento da planilha de dados e (iii) os níveis que constituem os grupos de fatores.

Tabela 12 – Coding protocol utilizado para as variáveis extralinguísticas

IBORUNA		
Variável independente	Código	Exemplos/Comentários
Sexo/gênero	masculino feminino	Masculino Feminino
Escolaridade	ensino.fundamental ensino.medio ensino.superior	1º ao 9º ano Fundamental 1º ao 3º ano Ensino Médio Ensino Superior
Faixa Etária	primeira.faixa segunda.faixa terceira.faixa quarta.faixa	7 a 25 anos 26 a 35 anos 36 a 55 anos +55 anos
Tipo textual	NE NR DE RP RO	Narrativa de Experiência Narrativa Recontada Descrição de Espaço Relato de Procedimento Relato de Opinião
SP2010		
Variável independente	Código	Exemplos/Comentários
Sexo/gênero	masculino feminino	Masculino Feminino
Escolaridade	ensino.medio ensino.superior	Até Ensino Médio completo Ensino Superior
Faixa Etária	primeira.faixa segunda.faixa terceira.faixa	19 a 34 anos 35 a 59 anos +60 anos
Região de São Paulo	centro.velho centro.expandido periferia	Centro Velho Centro Expandido Periferia
Zona de residência	central leste norte oeste sul	Zona Central Zona Leste Zona Norte Zona Oeste Zona Sul
Roteiro da entrevista²⁵	conversa sao.paulo percepcao depoimento avaliacao	Conversa Conversa sobre SP Percepção Leitura de trecho Avaliação

Fonte: adaptação de Gonçalves (s.d.) e Mendes (2013)

3.3.1 Sexo/gênero

A variável sexo apresenta uma certa controvérsia dentro dos estudos sociolinguísticos, pois as primeiras investigações de fenômenos em variação apontaram para uma maior ocorrência do uso de formas de prestígio na fala de mulheres, enquanto a forma correlata, e não prestigiada, se acentuava na fala dos homens. Essa correlação de sexo e prestígio foi

²⁵ Não foram consideradas as partes referentes a leitura de palavras e trecho de notícia.

interpretada por meio da posição que a mulher assumia dentro da comunidade: como cuidadora. Assim, a mudança, tendo em vista esse valor positivo, era liderada pelas figuras maternas.

Contudo, tal como aponta Freitag (2015), tais colocações precisam ser repensadas dentro da teoria, já que a sociedade atual se configura de forma diferente daquele período – década de 60 –, com movimentos sociais que questionam a hipótese clássica – junto de estudos que a refutam (TRUDGILL, 1972). Ademais, a problemática se acentua tendo em vista que os bancos de dados controlam apenas o sexo biológico do entrevistado, partindo da binaridade masculino x feminino, não fazendo o devido controle dos papéis de sexo/gênero mobilizados pelo falante. Como pontuado por Severo (2006, p. 8-9),

Dado que identidade (gênero é um aspecto da identidade) e linguagem implicam-se mutuamente, a sociolinguística não pode se limitar a um modelo essencialista que considera o gênero como uma categoria universal e previamente estabelecida, sem levar em conta que essa categoria é uma construção histórica, política e social, através da qual os indivíduos constituem suas identidades. Assim, argumenta-se em favor de que os trabalhos que tratam de variação/mudança devem contemplar as práticas sociais nas quais os indivíduos se engajam para constituir suas identidades (o gênero está implícito nessa constituição), pois é nessas comunidades de práticas que as variáveis assumem significado social e, a partir daí, se espalham (ou não) para o contexto social mais amplo.

Para fenômenos morfossintáticos, a questão se acentua, dado que, de um modo geral, os fatores extralinguísticos, em geral, não apresentam correlação na explicação do fenômeno (sendo esta variação ou mudança) tanto quanto os fatores linguísticos (MILROY, GORDON, 2003). Vale mencionar que existem exceções – como observaremos na seção 6 –, porém permanece a discussão sobre o significado social da variação morfossintática.

3.3.2 Escolaridade

A segunda variável extralinguística, escolaridade do falante, nos ajuda a compreender a relação da variação (e mudança) com a escolarização e, consequentemente, a norma. Nossa hipótese perpassa a correlação entre o uso de estratégias mais próximas do padrão por falantes mais escolarizados. Assim, mesmo que não tenha acesso a uma norma explícita do fenômeno, dado que as estratégias de relativização não são apresentadas nos anos escolares²⁶, o contato com a norma culta e, assim, com uma “norma implícita” guiaria seu uso.

²⁶ A informação se pauta em análise preliminar deste tópico que foi produzida para a disciplina de Tópicos de Sociolinguística do Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras de Araraquara (UNESP) e culminou em um artigo que ainda não foi publicado. O que averiguamos é que os materiais didáticos não apresentam as estratégias de relativização e se faz necessário discussões tais quais de Mallmann, Garcia e Escobar (2019) sobre orientações para o ensino.

Ademais, o uso mais frequente das estratégias não padrão é hipotetizado para falantes menos escolarizados. Isso, pois há uma correlação, que se apresenta nos estudos de relativização, entre a estratégia cortadora, em especial, com frequência maior na fala (ou escrita) de falantes com menos escolaridade (MOLLICA, 1977, 2003; CÔRREA, 1998).

Uma decisão metodológica precisou ser tomada no que se refere à estratificação apresentada pelo banco de dados IBORUNA. Os falantes são agrupados da seguinte forma: com primeiro fundamental (1º a 5º ano), com segundo fundamental (6º ao 9º ano), com ensino médio e com ensino superior. Manter o primeiro e segundo fundamental separados não foi considerado produtivo para a nossa análise, pois são duas faixas para representar um mesmo período escolar – o ensino fundamental – e culminaram em um padrão semelhante de comportamento, isto é, houve um número semelhante de usos das estratégias de relativização por esses falantes. Ademais, ao mantê-las separadas, corríamos o risco de tornar o modelo de análise multivariada escolhido (a regressão logística de efeitos mistos) muito complexo, o que deve ser evitado. Por essa razão, decidimos amalgamar os dois níveis de escolaridade, mantendo: ensino fundamental, ensino médio e ensino superior.

Para os dados provenientes do banco de dados SP2010, nenhuma alteração na estratificação precisou ser efetuada.

3.3.3 Faixa etária

Tendo em vista que estamos produzindo um estudo sincrônico, a faixa etária é um índice interessante para ser observado, dado que pode nos ajudar a compreender a disposição da variação (e mudança) dentro do grupo. Isso, pois estamos lidando com uma análise em tempo aparente (LABOV, 1972), o que pode configurar três cenários distintos: (i) a forma não padrão pode se concentrar nos mais novos, se configurando como uma inovação; ou (ii) pode ser observada nos mais velhos, o que mostra um certo conservadorismo da forma; ou, ainda, (iii) é possível não haver uma distribuição definida por faixas ou que indique alguma tendência.

Assim, em (i) temos os indícios necessários para produzir considerações sobre uma possível mudança em andamento e em (ii) uma tentativa de conservação de uma determinada forma. Contudo, é necessário ressaltar que apenas um trabalho que englobe uma perspectiva diacrônica, isto é, uma análise em tempo real, pode concluir o movimento da forma linguística observada – variação ou mudança. Para o presente trabalho, vale ressaltar a falta de um respaldo teórico-metodológico a esse respeito, visto que não há estudos sobre as estratégias de relativização no interior paulista que permitam uma comparação dos resultados. E, ainda,

mesmo que sejam possíveis as devidas referências ao trabalho de Tarallo (1983) sobre a região paulistana, não temos as informações necessárias para as especificidades do contexto locativo.

Desse modo, nossa hipótese é a de que falantes mais novos produzem mais construções consideradas não padrão, tomando por base os estudos disponíveis sobre as orações relativas no Português Brasileiro (MOLLICA, 1977; TARALLO, 1983; CORRÊA, 1998; BARROS, 2000; BURGOS, 2003; BISPO, 2003; 2009).

Por fim, é necessário salientar uma segunda decisão metodológica que se fez necessária e parte da estratificação apresentada pelo banco de dados IBORUNA. Os falantes são agrupados da seguinte forma: primeira faixa etária (7 a 15 anos); segunda faixa etária (16 a 25 anos); terceira faixa etária (26 a 35 anos); quarta faixa etária (36 a 55 anos) e quinta faixa etária (56 anos ou mais). Manter a primeira e a segunda faixa separadas não foi considerado produtivo para a nossa análise, visto que os falantes de 7 a 15 anos compõem uma faixa de falantes muito específica, isto é, que não completa as células, por exemplo, de escolaridade quando decidimos cruzar os dados em análises multivariadas.

Dessa forma, o ato de amalgamar as duas faixas parte da constatação de que esses falantes não preenchem as células de ensino médio e ensino superior completo. Apesar do grande intervalo, essa opção não desbalanceou nossa amostra, visto que o nível da variável não apresenta um número desproporcional de dados comparando aos demais. Essa questão será apresentada e explorada em mais detalhes na seção 6.

Para os dados provenientes do banco de dados SP2010, nenhuma alteração na estratificação precisou ser efetuada.

3.3.4 Região e zona de residência de São Paulo

As duas variáveis linguísticas que fazem referência ao local de residência do falante paulistano, região e zona de residência de São Paulo, são estratificadas no banco de dados Projeto SP2010 da seguinte maneira:

os critérios para a classificação de bairros como “mais centrais” ou “mais periféricos” se baseiam na história de ocupação dos bairros e no grau de desenvolvimento urbano em termos de verticalização, serviços e IDH. De modo geral, bairros mais periféricos são historicamente mais recentes, bem como seu processo de verticalização (nos últimos 20 anos), e possuem uma infraestrutura urbana relativamente menos desenvolvida (acesso a transporte público, hospitais, escolas, centros culturais etc.). Por outro lado, bairros mais centrais são mais antigos, são fortemente verticalizados (o que implica alta densidade demográfica) e possuem melhor infraestrutura urbana. Como consequência dessas características, 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. (OUSHIRO, 2015, p. 45)

Desse modo, observamos que o controle desses grupos de fatores contribui para a discussão do significado social da variação, visto que aponta para uma relação intrínseca da região da cidade com o mercado mobilizado pelo morador (BOURDIEU, 1991). Abaixo, reproduzimos um mapa que busca apontar como a divisão das regiões foi norteadada:

Figura 3 – Divisão da cidade de São Paulo entre regiões mais centrais e mais periféricas



Fonte: Oushiro (2015, p. 45)

Observamos no mapa que a região central, correspondente a coloração mais escura e delimitada no centro da imagem, concentra em si um menor número de bairros do que as regiões mais periféricas. Isso, pois elas referem-se a uma porção maior da cidade de São Paulo, abarcando uma variedade grande das zonas de residência de São Paulo, o que proporciona uma heterogeneidade tanto cultural quanto social, que deve ser levada em consideração em nossa análise.

Assim, nossa hipótese alinha-se com o que foi colocado até então, compreendendo que o perfil de falantes que produzem maior frequência de estratégias padrão se concentra em regiões mais centrais. Contudo, vale ressaltar que nossa hipótese não perpassa uma visão dicotômica entre centro x periferia em relação a padrão x não padrão; isto é, não assumiremos que o uso maior de estratégias não padrão se concentrará em regiões mais periféricas tendo em vista a sua variedade apresentada no mapa (1).

3.3.5 Tipologia textual

A penúltima variável controlada²⁷ diz respeito às tipologias textuais em que a entrevista sociolinguística do banco de dados IBORUNA está subdividida (sendo assim, não compreendendo o banco de dados Projeto SP2010). Como mencionado brevemente anteriormente, temos: Narrativa de Experiência, Narrativa Recontada, Descrição de Espaço, Relato de Procedimento, Relato de Opinião. Cada tipo concentra perguntas norteadoras, por parte do entrevistador, para que o falante cumpra determinadas funções discursivas e textuais.

Tal grupo de fatores foi importante para a compreensão dos usos do pronome relativo **onde** na pesquisa anterior (ver seção 2). Desse modo, consideramos adequado avaliar em qual tipologia textual se concretiza a ocorrência das estratégias de relativização, buscando um possível padrão. Nossa hipótese principal é de que encontraremos uma maior frequência de dados na *Descrição de Espaço*, dado o valor locativo que estamos investigando. Ademais, tendo em vista nossa hipótese anterior de que haveria um maior uso de **padrão com onde** para antecedentes classificados como objeto, hipotetizamos que em *Descrição de Espaço* se concentrará a maior frequência dessa estratégia.

3.3.6 Roteiro da entrevista

Por fim, a última variável controlada refere-se aos momentos da entrevista, guiados pelo entrevistador, para contemplar tópicos discursivos específicos, tais como: conversa – sobre lazer, infância, família, trabalho e bairro –, conversa sobre a cidade de São Paulo, percepção, leitura – de uma notícia, lista de palavras e um trecho – e avaliação. Desse modo, as perguntas do entrevistador, para cada um desses momentos, apontam um direcionamento semântico que é muito importante para a nossa análise.

Assim, controlar essa variável, tal qual a tipologia textual, nos permite averiguar um possível padrão de uso das estratégias dentro dos tópicos discursivos mobilizados. Nossa hipótese é de que encontraremos uma maior frequência de dados no momento da conversa, em especial, quando o falante está contando uma história referente ao tópico discursivo ou dando sua opinião acerca da cidade de São Paulo e bairros, isto é, contextos em que o locativo se faz mais proeminente.

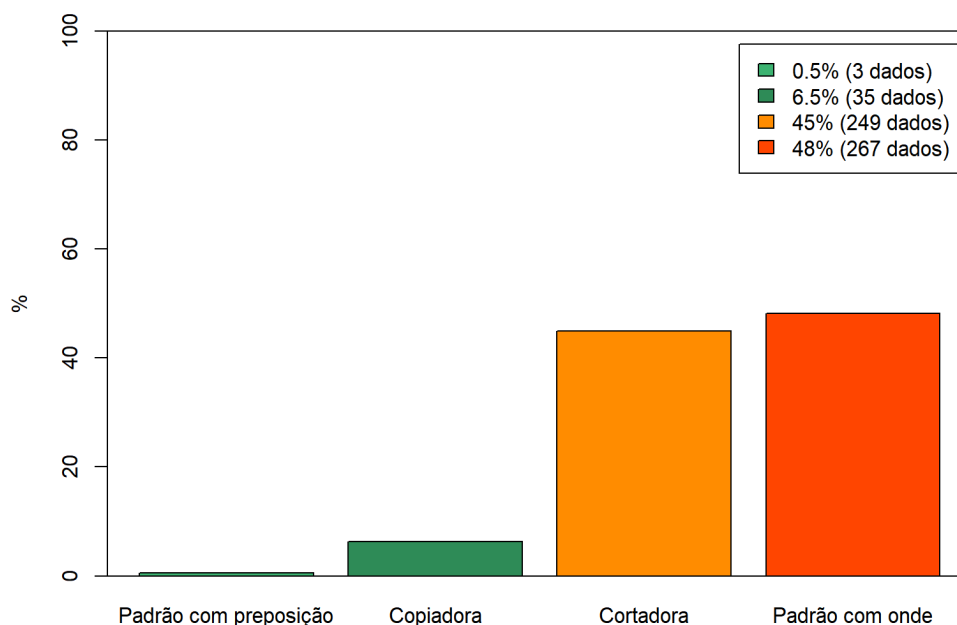
²⁷ Gostaríamos de pontuar que não estamos considerando a variável tipologia textual como extralinguística, mas consideramos adequado apresentá-la nessa seção tendo em vista que se refere à estruturação do banco de dados e não uma variável por nós criada.

4 A SINTAXE DOS PAULISTAS: ANÁLISE DE PRODUÇÃO

“nas análises variacionistas somos limitados no que descobrimos pelo que nos propusemos a procurar”
(CHESHIRE, 1998, p. 65 [tradução nossa²⁸])

A presente seção tem por objetivo expor os resultados encontrados na análise de produção da fala paulista, isto é, dos bancos de dados IBORUNA (GONÇALVES, 2007) e Projeto SP2010 (MENDES, 2013). Tendo em vista uma explicação consistente do fenômeno em questão, consideramos mais adequado apresentar os resultados de ambos os bancos de forma comparativa. Nossa intenção não é propor um trabalho contrastivo, visto que a estratificação e as variáveis independentes selecionadas entre os bancos não são idênticas; nossa proposta é investigar em minúcias as diferenças apresentadas para compreender os usos das estratégias de relativização locativas na fala paulista. Assim, uma explanação separada de cada *corpus* resultaria em algumas redundâncias que poderiam atrapalhar o poder explanatório do nosso estudo, em uma “quebra” das interpretações propostas. Assim, nas Figuras 4 e 5 encontramos a distribuição das estratégias de relativização nos dois bancos de dados:

Figura 4 – Proporção de uso das estratégias de relativização no banco de dados **IBORUNA**



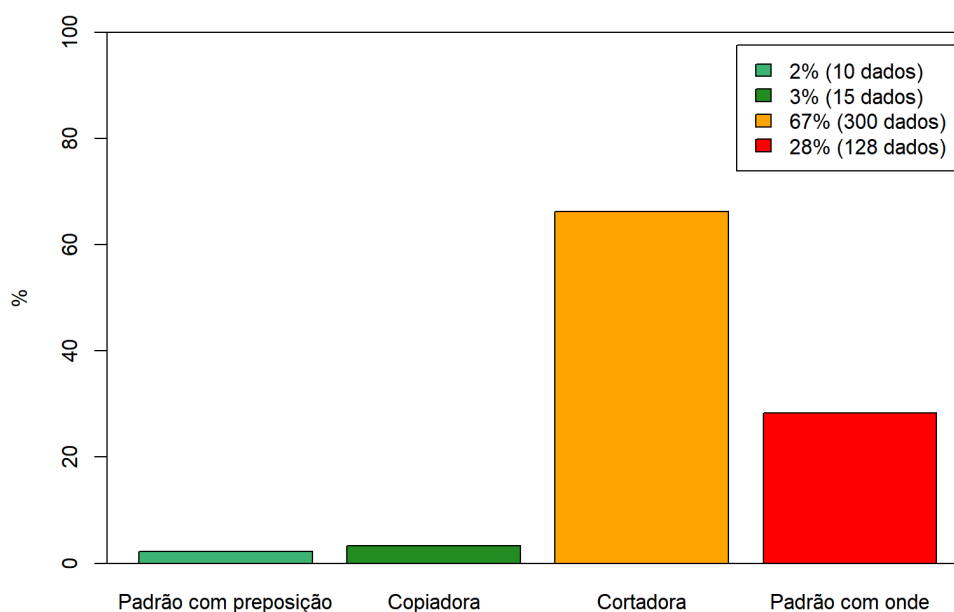
Fonte: própria

²⁸ Original: in variationist analyses we are limited in what we discover by what we set out to look for.

A partir do gráfico, observamos um uso maior de **padrão com onde** comparativamente com as outras estratégias de relativização locativas disponíveis no Português Brasileiro. Notamos, também, que os números de **padrão com preposição** e **copiadora** são baixos, o que justifica nossa decisão metodológica de retirá-las de nossa análise quantitativa, produzindo uma análise qualitativa para compreender quais fatores incidem sobre esse (baixo) uso²⁹. Assim, nossa hipótese se confirma no sentido de que o uso da **padrão com preposição** (*pied-piping*) apresenta uma frequência menor do que as estratégias não padrão.

Porém, se faz necessário averiguar o que estaria condicionando uso da **padrão com onde**, visto que ela se sobrepõe à estratégia **cortadora**. Vale pontuar que não estamos atribuindo um determinado valor à estratégia ou assumindo que a comunidade compartilhe esse sentido; é nosso interesse entender como a estratégia que, de acordo com a tradição gramatical, é considerada padrão se sobrepõe à variante não padrão.

Figura 5 – Proporção de uso das estratégias de relativização no banco de dados Projeto SP2010



Fonte: própria

Comparando os dois gráficos, observamos no banco de dados SP2010 um uso maior da estratégia de relativização **cortadora**, um resultado que se alinha com a literatura sobre a capital

²⁹ Vale mencionar que também foram retirados do *corpus*, para a análise quantitativa, 4 dados classificados como adjunto adnominal na função sintática do pronome relativo, visto que foram as únicas ocorrências nos dois bancos de dados.

paulistana (TARALLO, 1983, 1985). Verificamos, novamente, que os números de **padrão com preposição e copiadora** são baixos, o que justifica nossa decisão metodológica de retirá-las de nossa análise quantitativa, produzindo uma análise qualitativa sobre esse (baixo) uso. Assim, através dos dados paulistanos, nossa hipótese se confirma no uso menos frequente da estratégia padrão e o uso mais frequente da estratégia não padrão.

É interessante pontuar que o número total de estratégias de relativização locativas em contexto preposicionado na fala paulista é de 1007 dados. Retomando a problemática da variável sintática dentro da Teoria de Variação e Mudança Linguística, percebemos que não estamos lidando com um fenômeno de baixa frequência. Por outro lado, temos um *corpus* robusto para tecer considerações a respeito do processo de variação.

Para compreender em mais detalhes os fatores condicionantes da distribuição acima, primeiro apresentaremos a regressão logística de efeitos mistos, com a subsequente apresentação da análise univariada das variáveis que não compuseram o modelo. Essa disposição dos resultados tenciona colocar em primeiro plano a discussão da análise multivariada que visa fornecer uma compreensão maior das variáveis previsoras, isto é, ao testá-las com vistas a verificar como cada uma atua na explicação do fenômeno em variação.

4.1 DISTRIBUIÇÃO DE DADOS DAS VARIÁVEIS CORRELACIONADAS

Optamos por reportar o modelo de regressão logística com efeitos mistos para o nosso fenômeno, visto que as observações no conjunto de dados não são independentes umas das outras. Devemos nos atentar ao fato de que os nossos resultados apresentam uma variabilidade grande que é proveniente do próprio falante e isso não pode ser facilmente replicado como uma variável fixa em outros estudos, constituindo-se, assim, em uma variável aleatória.

Outrossim, a identidade lexical do verbo na oração relativa também foi controlada como uma variável aleatória dentro do nosso *corpus*. A variabilidade já se apresenta dentro da nossa pesquisa ao identificarmos que os verbos e suas frequências foram diferentes. A figura 6 apresenta uma nuvem de palavras com vistas a expor essa realidade.

Figura 6 – Nuvem de palavras da identidade lexical do verbo na oração relativa no banco de dados IBORUNA e Projeto SP2010, respectivamente



Fonte: própria

Por meio das figuras, notamos que alguns verbos são semelhantes, e isso se deve ao valor locativo com o qual estamos trabalhando, mas a disposição se dá de forma diferente. A primeira diferença se dá pelo número de verbos diferentes que se apresenta em cada um dos bancos: 139 verbos no corpus proveniente do IBORUNA e 95 verbos no corpus de dados proveniente do Projeto SP2010. A segunda é justamente na disposição desses itens pelas estratégias de relativizações, sendo que os cinco mais frequentes para o banco de dados IBORUNA foram: ficar (60), morar (52), ter (33), trabalhar (32) e estar (31); enquanto para o banco de dados Projeto SP2010 foram: ter (97), morar (89), estar (24), estudar (14) e trabalhar (14).

A partir disso, nosso modelo para o banco de dados IBORUNA é composto pela variável resposta – isto é, a variável dependente, estratégia de relativização – e as variáveis previsoras: faixa etária, escolaridade, identidade semântica do locativo – isto é, as variáveis independentes³⁰. Para o banco de dados Projeto SP2010, temos como variáveis previsoras: faixa etária, região do falante, escolaridade e definitude do antecedente.

Para o modelo de regressão logística, o segundo nível da variável resposta é o selecionado, o que deve nortear nossa leitura para o aumento ou diminuição da probabilidade

³⁰³⁰ Para chegar a esse modelo que melhor explica a variação foram aplicadas as funções step, drop1 e validate que identificaram quais as variáveis que deveriam permanecer no modelo, através de comparação, inclusão e averiguação se não houve um sobre ajuste (overfitting) das previsoras (OUSHIRO, 2021). Desse modo, foram excluídas as variáveis: função sintática do pronome relativo, adjacência entre o sintagma nominal antecedente e a oração relativa, identidade morfológica do locativo, classificação tradicional da relativa, definitude, especificidade e status informacional do antecedente, tipo textual e sexo.

de ocorrência da estratégia de relativização **padrão com onde**. A escolha dessa variante se deu pelo interesse do estudo em averiguar o comportamento do pronome relativo **onde** em contextos locativos na fala paulista, tendo por pano de fundo as demais estratégias de relativização em um contexto de concorrência e variação. Dessa forma, podemos interpretar os fatores e contextos que mais favorecem o emprego dessa variante.

O *intercept* é composto por falantes da segunda faixa etária, com ensino superior, produzindo orações relativas com antecedente objeto, para o banco de dados IBORUNA, e por falantes da primeira faixa etária, moradores da periferia, com ensino superior, produzindo orações relativas com antecedente definido, para o banco de dados Projeto SP2010. Os níveis da faixa etária, escolaridade, identidade do locativo, região do falante e definitude do antecedente foram alterados, dado que o programa os seleciona em ordem alfabética. Essa mudança teve duas motivações: a primeira, para que o *intercept* fosse composto pelos níveis que favoreceram o uso da estratégia **padrão com onde**; a segunda, pois todas as vezes em que rodamos o modelo sem alterar os níveis, apresentava-se um erro. Nas tabelas 13 e 14 são apresentados os modelos de regressão de efeitos mistos para os dois bancos de dados.

Tabela 13 – Análise de regressão de efeitos mistos da estratégia de relativização **padrão com onde** no banco de dados **IBORUNA**
Intercept = 1.68554

Variável	Estimativa	Erro padrão	Valor-z	p	Apl./N
Faixa etária					
Segunda faixa		(v. de referência)			81/123 (66%)
Primeira faixa	-0.83635	0.50846	-1.645	0.09 .	70/186 (38%)
Terceira faixa	0.04592	0.53115	0.086	0.93	69/127 (54%)
Quarta faixa	0.40514	0.57948	0.699	0.48	46/77 (60%)
Escolaridade					
Ensino Superior		(v. de referência)			115/174 (66%)
Ensino fundamental	-1.18225	0.45881	-2.577	0.009 **	88/215 (41%)
Ensino médio	-0.71401	0.50048	-1.427	0.15	63/124 (51%)
Tipo de locativo					
Objeto		(v. de referência)			53/73 (73%)
Local	-1.02283	0.39796	-2.570	0.01 *	213/440 (48%)
Modelo: ESTRATEGIA ~ FAIXA + ESCOLARIDADE + TIPO.LOCATIVO + + (1 INFORMANTE) + (1 IDENTIDADE.LEXICAL)					

Fonte: própria

Tabela 14 – Análise de regressão de efeitos mistos da estratégia de relativização **padrão com onde** no banco de dados **Projeto SP2010**

Intercept = 0.01369

Variável	Estimativa	Erro padrão	Valor-z	p	Apl./N
Faixa etária					
Primeira faixa		(v. de referência)			66/180 (37%)
Segunda faixa	-0.43279	0.39015	-1.109	0.26	34/128 (27%)
Terceira faixa	-0.71827	0.39288	-1.828	0.06 .	28/120 (23%)
Região de São Paulo					
Periferia		(v. de referência)			67/185 (36%)
Centro velho	-1.04348	0.47008	-2.220	0.02 *	16/85 (19%)
Centro expandido	-0.40865	0.36063	-1.133	0.25	45/158 (28%)
Escolaridade					
Ensino Superior		(v. de referência)			86/236 (36%)
Ensino médio	-0.82073	0.33290	-2.465	0.01 *	42/192 (22%)
Definitude do antecedente					
Definido		(v. de referência)			94/259 (36%)
Indefinido	-0.99915	0.30723	-3.252	0.001 **	34/169 (20%)
Modelo: ESTRATEGIA ~ FAIXA.ETARIA + REGIAO + ESCOLARIDADE + DEFINITUDE + (1 INFORMANTE) + (1 IDENTIDADE.LEXICAL)					

Fonte: própria

A faixa etária é um importante índice para compreendermos a distribuição da variação dentro da comunidade, o que é reforçado pelo modelo de regressão, em que essa variável foi selecionada pelos dois bancos de dados. Da mesma forma, a escolaridade tem um viés explanatório muito importante dentro do quadro de variação das estratégias de relativização.

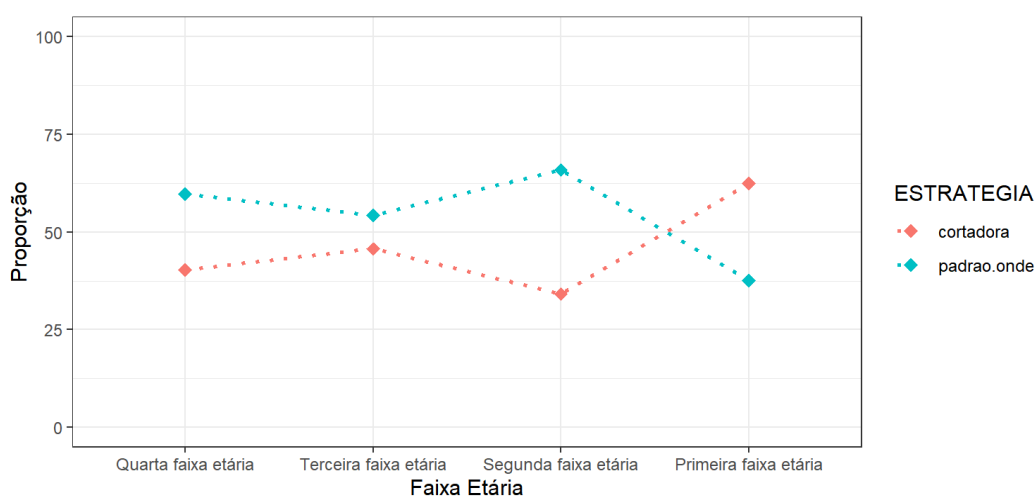
Assim, pelas estimativas, para o banco de dados IBORUNA, observamos que há um desfavorecimento do uso da **padrão com onde** para falantes da primeira faixa etária (7-25 anos) em comparação com o *intercept*. Tal diferença entre as estimativas não é significativa, estatisticamente ($p > 0.05$), porém tendo em vista que é o único nível que desfavorece o uso da estratégia, é importante reportar essa informação. Para a escolaridade, averiguamos um desfavorecimento do uso da variante padrão pelos níveis de escolaridade mais baixos, em especial o ensino fundamental. Isso aponta para uma relação do uso da **padrão com onde** com o contato com a norma culta durante os anos escolares.

Para o banco de dados Projeto SP2010, observamos que há um desfavorecimento do uso da **padrão com onde** para falantes da segunda (19-34 anos) e terceira faixa etária (35-59 anos) em comparação com o *intercept*. Tal diferença entre as estimativas não é significativa, estatisticamente ($p > 0.05$). De forma análoga ao banco de dados do interior paulista, há um

desfavorecimento do uso da variante padrão pelo nível de escolaridade mais baixo, isto é, o ensino médio.

Em nossa análise, é de suma importância compreender as correlações entre essas variáveis para caracterizar os usos das estratégias de relativização locativas pelos paulistas. Assim, apresentaremos a disposição geral dos resultados e depois apresentaremos nossa proposta de interpretação.

Figura 7 - Proporção de uso das estratégias de relativização segundo a **faixa etária** do falante no banco de dados **IBORUNA**



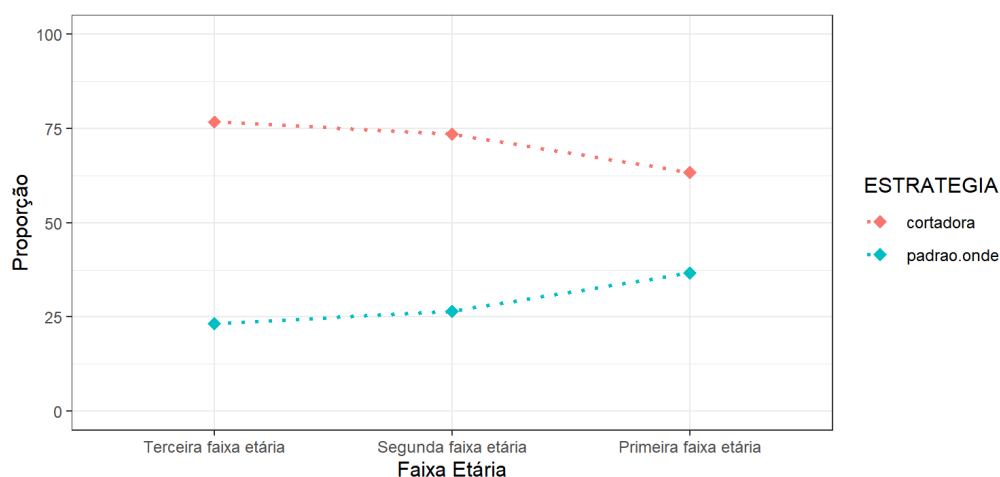
Própria: própria

Com base na figura 7, observamos dois movimentos centrais nos falantes do interior paulista: (i) um pico em rosa (62% – 116/186 ocorrências³¹), na primeira faixa etária (7-25 anos), que aponta para um maior uso da estratégia **cortadora** e (ii) um pico em azul (66% - 81/123 ocorrências), na segunda faixa etária (26-35 anos), que corresponde a um maior uso da estratégia **padrão com onde**. Podemos, ainda, elencar que a terceira faixa etária (36-55 anos) apresenta uma distribuição equilibrada de uso das estratégias; enquanto a quarta faixa etária (+55 anos) apresenta uma leve inclinação para maior uso da **padrão com onde** (59% - 46/77 ocorrências).

Ao contrapor esses resultados com aqueles encontrados no banco de dados Projeto SP2010, um padrão se repete na figura 8:

³¹ Antes de amalgamarmos os dois primeiros níveis da faixa etária, a distribuição se dava da seguinte forma: primeira faixa etária (7-15 anos) com 41% (36/87 ocorrências) de **padrão com onde** e 59% (51/87 ocorrências) de **cortadora** e segunda faixa etária (16-25 anos) com 34% (34/99 ocorrências) de **padrão com onde** e 66% (65/99 ocorrências) de **cortadora**.

Figura 8 - Proporção de uso das estratégias de relativização segundo a **faixa etária** do falante no banco de dados **SP2010**



Fonte: própria

Antes de apresentar os resultados em detalhe, gostaríamos de salientar que os falantes, de todas as faixas etárias, apresentam um uso maior de **cortadora**, fato evidenciado desde a distribuição geral dos resultados. No gráfico, podemos apontar, novamente, para dois movimentos centrais: (i) um pico em azul (38% - 69/183 ocorrências) na primeira faixa etária (19-34 anos), indicando o maior uso da estratégia **padrão com onde** e (ii) um pico em rosa (76% - 94/123 ocorrências) na terceira faixa etária (60 anos ou mais), indicando o maior uso da estratégia **cortadora**.

Desse modo, verificamos que os falantes com idades similares nos dois bancos de dados produzem mais a estratégia **padrão com onde** em contextos preposicionados e locativos. Não obstante, uma ressalva precisa ser feita: como já mencionado, a frequência maior da estratégia **cortadora** já era esperada nos paulistanos, dada a literatura consolidada na área (TARALLO, 1983,1985). Destrinchando em mais detalhes o que o estudo clássico de Tarallo (1983) nos fornece, o seu controle para a idade dos falantes se deu apenas em dois níveis não considerados estatisticamente significantes: - 35 anos e + 35 anos. Isso nos impede de produzir uma meta-análise buscando identificar, em mais detalhes, o crescimento ou não do uso da estratégia cortadora. Contudo, os valores apontados pelo autor para a classe social já nos dão um respaldo para interpretar os resultados no que concerne à nossa terceira faixa etária, uma vez que foram controlados os contextos preposicionados, isto é, as posições sintáticas mais baixas (objeto indireto, oblíquo e genitivo). Desse modo, reproduzimos abaixo, na Tabela 5, os valores encontramos pelo autor.

Tabela 5 – Porcentagem de uso das estratégias **copiadora**, **cortadora** e **piedpiping** nas três posições sintáticas mais baixas para as três classes sociais

ESTRATÉGIA	BAIXA	MÉDIA	ALTA	TOTAL
<i>Piedpiping</i>	1,3% 2 dados	5,6% 5 dados	17,9% 14 dados	21 dados
Pronome lembrete (copiadora)	21% 33 dados	11,2% 10 dados	9% 7 dados	50 dados
Cortadora	77,7% 122 dados	83,2% 74 dados	73,1% 57 dados	253 dados
<i>Total</i>	157 dados	89 dados	78 dados	324 dados

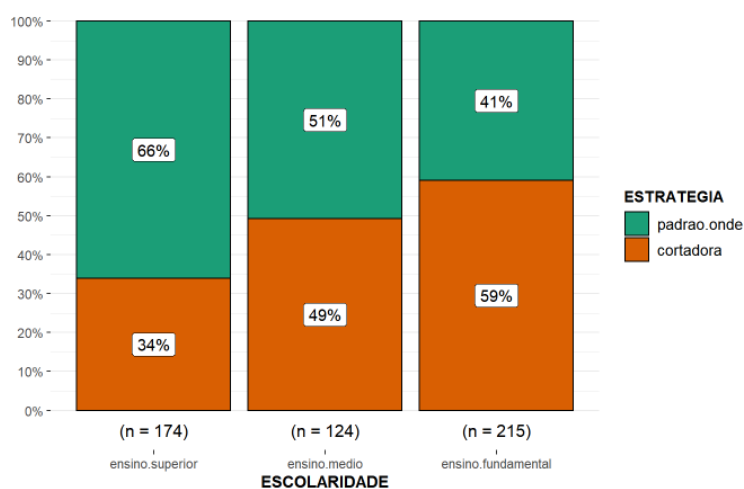
Fonte: adaptado de Tarallo (1983)

O que estamos propondo ao expor tais taxas de proporção é que os falantes da década de 80 produziam alta frequência da estratégia não padrão; e podemos aproximá-los, de alguma forma, dos falantes que compuseram a terceira faixa etária do banco de dados Projeto SP2010, visto que há um intervalo de, aproximadamente, 30 anos entre as duas amostras. Assim, os falantes com menos de 35 anos entrevistados por Tarallo (1983) correspondem, no sentido de partilharem o mesmo contexto, aos falantes com 60 anos ou mais do nosso *corpus*, o que sugere uma manutenção dessa variante em sua fala³².

Dando seguimento para a apresentação da distribuição de uso das estratégias de relativização de acordo com as variáveis extralinguísticas selecionadas, a figura 9 apresenta os resultados da escolaridade do falante para o banco de dados IBORUNA.

³² Uma questão que é preciso considerar que, ainda que estando incluídos nesse grupo de funções sintáticas mais baixas, os contextos locativos têm a particularidade de uma estratégia a mais, que é, justamente, aquela com **onde**. Mas isso não invalida nossa comparação.

Figura 9 – Proporção e frequência das estratégias de relativização segundo a **escolaridade** do falante no banco de dados **IBORUNA**

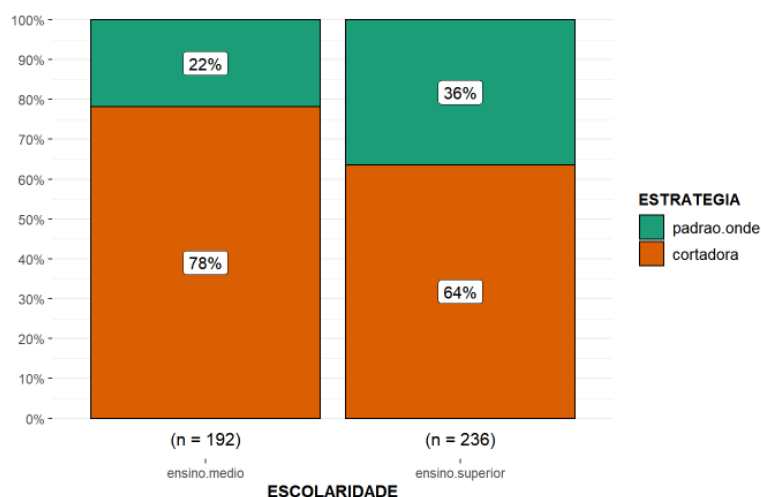


Fonte: própria

O gráfico nos apresenta uma gradação: conforme o nível de escolaridade do falante aumenta, também aumenta o uso da estratégia **padrão com onde** e diminui o uso da estratégia **cortadora**. Isso vai ao encontro da hipótese de que o uso mais frequente da variante padrão – e menos frequente da variante não padrão – situa-se na fala dos mais escolarizados e está relacionado com o tempo de contato com a norma culta.

Para os dados provenientes do banco de dados Projeto SP2010, esse padrão se mantém e é apresentado na figura 10.

Figura 10 – Proporção e frequência das estratégias de relativização segundo a **escolaridade** do falante no banco de dados **SP2010**



Fonte: própria

O gráfico apresenta uma disposição muito similar ao anterior: falantes mais escolarizados fazem maior uso da estratégia **padrão com onde**. É necessário ter em vista que esse valor se configura como relevante dentro da nossa proposta de interpretação mesmo com a frequência absoluta de **cortadora** em nosso *corpus*.

Para interpretarmos as duas variáveis extralinguísticas, vamos nos apoiar em Bourdieu (1991) e a concepção de mercado linguístico. Para o sociólogo, a relação comunicativa poderia ser entendida de duas formas: a primeira, a partir de um código que precisa ser compreendido e decifrado pelo emissor e receptor, e a segunda como uma troca econômica dentro de uma relação simbólica de poder entre aquele que produz – dotado de capital linguístico – e o consumidor ou mercado – dotado de perfil material ou simbólico. Nesse sentido, o valor que o autor propõe para a troca linguística é de que os signos não são apenas formas a serem codificadas, eles apresentam “sinais de riqueza, destinados a serem avaliados e apreciados, e sinais de autoridade, destinados a serem acreditados e obedecidos” (BOURDIEU, 1991, p. 66 [tradução nossa]³³).

A partir dessa concepção do ato comunicativo, Bourdieu compreende que o enunciado, ou discurso, recebe um valor e sentido por meio da relação com o mercado; e esse é caracterizado como uma lei particular de formação de preços (BOURDIEU, 1991, p. 67 [tradução nossa]³⁴). Sua proposição, em outras palavras, é que o enunciado recebe um valor a depender da relação de poder estabelecida através das competências linguísticas do falante, sua capacidade de produção e apropriação e apreciação. As unidades linguísticas, nesse sentido, recebem classificações sociais que, a partir da apropriação e valorização, podem definir mercados que também são classificados socialmente, contribuindo para determinar a lei de formação de preços.

Assim, o mercado linguístico se constrói de modo plural, não restrito a apenas um contexto linguístico e social. Ele pode ser utilizado para designar todo ato de comunicação humana e, de acordo com Bourdieu (1991), temos que ter em mente que, quanto mais formal é a situação investigada, mais se torna provável que a competência linguística dominante vai se construir de forma a ser o capital linguístico dentro de um determinado mercado. Assim, ela se torna capaz de impor a lei de formação de preços mais favorável a seu produto e obtém o lucro simbólico que corresponde a isso. Desse modo, a formalidade está associada com a capacidade de impor o reconhecimento da legitimidade do modo de expressão que domina nesse contexto,

³³ Original: “[...] signs of wealth, intended to be evaluated and appreciated, and signs of authority, intended to be believed and obeyed”

³⁴ Original: “[...] characterized by a particular law of price formation”

tornando-se uma regra imperativa de rigor e tendo “os destinatários desses produtos linguísticos mais inclinados a conhecer e reconhecer a legitimidade desse modo de expressão, mesmo fora dos limites da situação formal” (BOURDIEU, 1991, p. 70 [tradução nossa]³⁵).

Desse modo, compreender os valores associados a um determinado mercado é importante para investigar os significados sociais associados às formas, visto que o mercado “coloca” um preço em um produto linguístico. Ademais,

A cognição prática e o reconhecimento das leis imanentes de um mercado e as sanções pelas quais elas se manifestam determinam as modificações estratégicas do discurso, sejam elas concernentes ao esforço de 'corrigir' uma pronúncia desvalorizada na presença de representantes da pronúncia legítima e, de forma mais geral, todas as correções que tendem a valorizar o produto linguístico por uma mobilização mais intensa dos recursos disponíveis, ou, inversamente, a tendência a recorrer a uma sintaxe menos complexa para as frases curtas que os psicólogos sociais têm observado em adultos quando se dirigem crianças. Os discursos são sempre, em certa medida, *eufemismos* inspirados pela preocupação de 'falar bem', de 'falar adequadamente', de produzir os produtos que atendem às exigências de um determinado mercado; são *formações de compromisso* resultantes de uma transação entre o interesse expressivo (o que se quer dizer) e a censura inerente a relações particulares de produção linguística (seja a estrutura da interação linguística ou a estrutura de um campo especializado), uma censura que é imposta a um locutor ou escritor dotado de certa competência social, ou seja, de um poder simbólico mais ou menos significativo sobre essas relações de poder simbólico. (BOURDIEU, 1991, p. 78-79 [tradução nossa]³⁶)

É relevante para a Teoria de Variação e Mudança Linguística compreender essas relações sociais pontuadas pelo autor para a investigação dos fenômenos linguísticos. Averiguar quais são as “formações de compromisso” que estão em jogo nas situações comunicativas pode nos ajudar a interpretar os padrões linguísticos observados nos nossos resultados, além de nortear, com maior precisão, as questões de testes de percepção e processamento. Em especial, no presente estudo, contribui para a interpretação dos dois picos apontados na variável faixa etária.

³⁵ Original: “[...] making the recipients of these linguistic products more inclined to know and recognize the legitimacy of this mode of expression, even outside the constraints of the formal situation.”

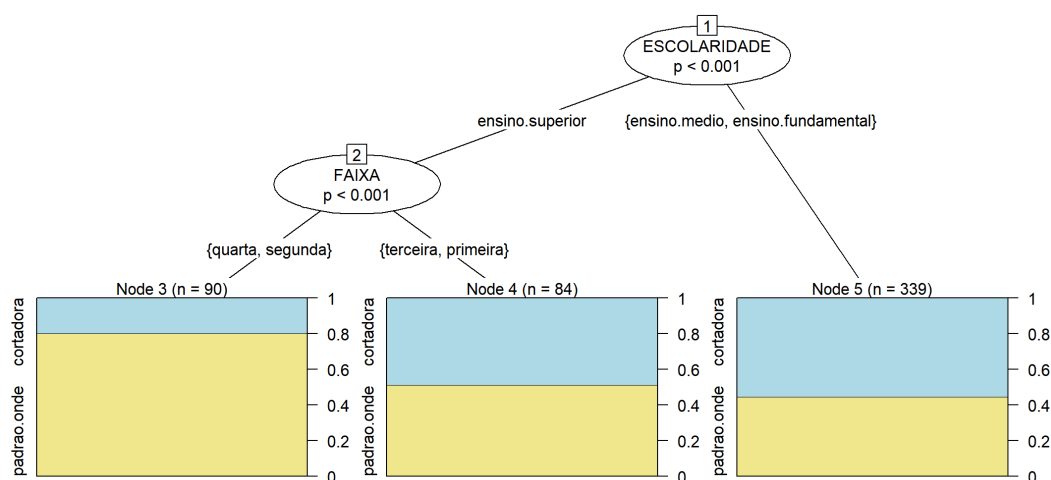
³⁶ Original: “The practical cognition and recognition of the immanent laws of a market and the sanctions through which they are manifested determine the strategic modifications of discourse, whether they concern the effort to 'correct' a devalued pronunciation in the presence of representatives of the legitimate pronunciation and, more generally, all the corrections which tend to valorize the linguistic product by a more intense mobilization of the available resources, or, conversely, the tendency to resort to a less complex syntax to the short phrases which social psychologists have observed in adults when they address children. Discourses are always to some extent *euphemisms* inspired by the concern to 'speak well', to 'speak properly', to produce the products that respond to the demands of a certain market; they are *compromise formations* resulting from a transaction between the expressive interest (what is to be said) and the censorship inherent in particular relations of linguistic production (whether it is the structure of linguistic interaction or the structure of a specialized field), a censorship which is imposed on a speaker or writer endowed with a certain social competence, that is, a more or less significant symbolic power over these relations of symbolic power.”

Isso, pois a primeira e a segunda faixas etárias, no caso do banco de dados IBORUNA, e a primeira e a terceira faixas etárias, no caso do banco de dados Projeto SP2010, apresentam momentos distintos da vida de um falante e, principalmente, referem-se a contatos com mercados com capitais linguísticos diferentes. Retomando a nossa hipótese inicial, é na fala dos mais jovens que se concentra o uso da estratégia **não padrão**. Ao correlacionar essa questão com o período escolar, que marca esse conjunto com ensino fundamental, ensino médio e ensino superior, podemos apontar para um período de conflito entre o uso da norma e a norma de uso: a situação em que, por um lado, o falante está em contato com o seu interesse expressivo e, por outro, reconhece uma censura a determinadas relações e contextos, em processos de adequação linguística à norma culta.

Desse modo, podemos interpretar a proporção de uso maior da variante **não padrão** pelo processo inerente de construção e identificação das relações de poder simbólico. Gostaríamos também de salientar que essa faixa etária – a primeira do banco de dados IBORUNA – foi responsável pela frequência mais alta de **onde** não locativo no estudo anterior (ALMEIDA, BERLINCK, 2019), o que contribui para nossa discussão de formação de norma do falante.

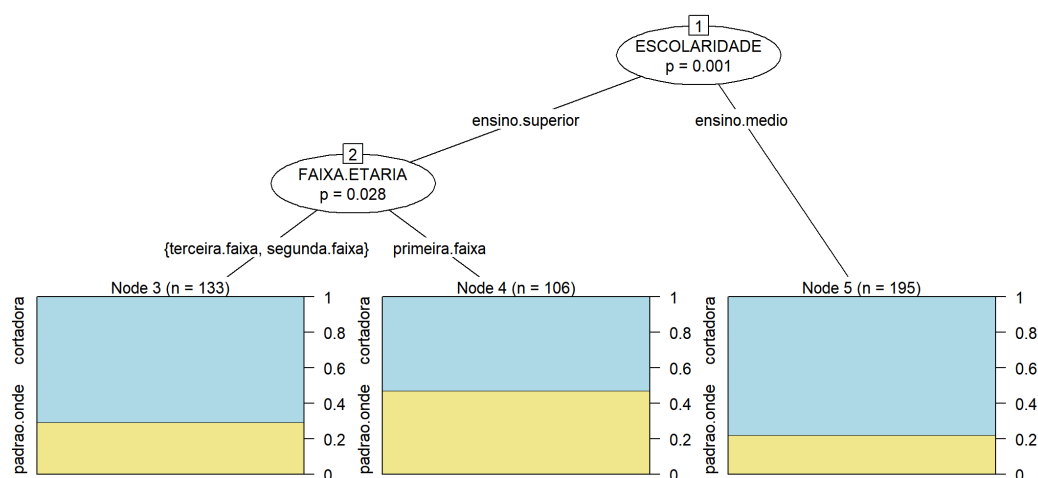
No que se refere aos falantes de 26 a 35 anos do interior paulista e 19 a 34 da capital paulistana, que apresentaram índices mais altos de uso da estratégia **padrão com onde**, podemos mobilizar dois mercados: (a) o trabalho e (b) a escolaridade. Para explicitar a relação que propomos, é necessário observar a árvore de classificação condicional nas Figuras 11 e 12.

Figura 11 - Árvore de classificação condicional³⁷ para o uso das estratégias de relativização segundo a **faixa etária** e **escolaridade** do falante no banco de dados **IBORUNA**



Fonte: própria

Figura 12 - Árvore de classificação condicional³⁸ para o uso das estratégias de relativização segundo a **faixa etária** e **escolaridade** do falante no banco de dados **SP2010**



Fonte: própria

Observamos na figura 11 que a primeira variável selecionada, escolaridade, se divide, apresentando um contexto específico no que se refere ao ensino superior. Assim, a segunda variável, faixa etária, apresenta a primeira e a terceira faixa responsáveis por um uso equilibrado das duas estratégias, enquanto a segunda e quarta faixa etária são relacionadas ao maior uso da **padrão com onde**. Assim, são os falantes com ensino superior e com 26 a 35 e + 56 anos que produzem a estratégia considerada como padrão segundo a prescrição gramatical. Isso nos ajuda

³⁷ Foram incluídos no modelo as seguintes variáveis: faixa etária e escolaridade.

³⁸ Foram incluídos no modelo as seguintes variáveis: faixa etária e escolaridade.

a compreender por que os entrevistados com +55 anos apresentavam maior frequência dessa variante, dado que há uma correlação com a sua escolaridade. Além disso, no que se refere à disposição observada para os níveis de ensino fundamental e médio, compreendemos que a distribuição equilibrada entre as duas estratégias, independentemente da faixa etária, sugere que são outros fatores que atuam nesses segmentos para a definição da variação.

Na figura 12, o movimento é semelhante: a variável escolaridade expõe o comportamento diferente do ensino superior que se relaciona à variável faixa etária. Desse modo, é possível inferir que são os falantes da primeira faixa etária (19 a 34 anos) e com um nível de escolaridade mais alto que são responsáveis pela maior frequência de uso da **padrão com onde**. Os demais níveis, de ambas as variáveis, indicam esse massivo uso da **cortadora** que, como mencionamos, remete para um cenário já estudado na literatura (TARALLO, 1983, 1985).

O que estamos propondo, no que se refere aos dois mercados mobilizados, (a) o trabalho e (b) a escolaridade, é que em (a) os falantes que compõem esse nível da variável extralinguística estão em contato com situações de autoridade, desempenhando papéis de submissão ou de autoridade dentro de um sistema trabalhista (e capitalista). Essas relações de poder norteiam os usos linguísticos para uma adequação com a norma, bem-vista e reconhecida dentro da formalidade pedida ao ambiente. Sendo assim, temos o pico do uso da estratégia **padrão com onde**.

A proposta em (b) se apresenta, pois é de nosso interesse questionar até que ponto as duas variantes aqui analisadas assumem, a partir da percepção do falante, os valores de prestígio relacionadas com formas **padrão** ou **não padrão**. Em primeiro lugar, não temos estudos que revelem se o uso de **cortadora** ou **padrão com onde**, em contextos locativos, apresenta um valor de prestígio, neutralidade ou desprestígio por parte do falante que poderia moldar uma “escolha” das estratégias para uma aproximação da norma. Em segundo lugar, estamos lidando com uma variável linguística que não é formalmente ensinada em sala de aula; isto é, os materiais didáticos não contemplam a diferença entre as estratégias de relativização em seu conteúdo³⁹. Desse modo, não seria produtivo partirmos de uma interpretação que toma o uso de uma ou outra estratégia como uma escolha consciente do falante para se adequar a uma norma explícita e ter acesso a determinados valores sociais. É nossa intenção, assim, produzir

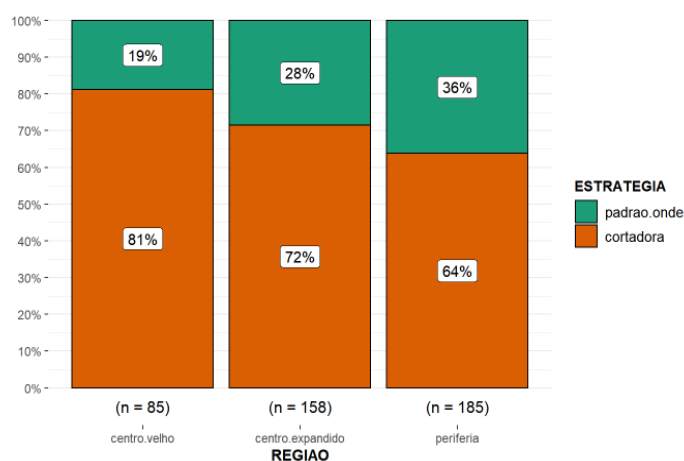
³⁹ Ressaltamos, novamente, que uma análise preliminar deste tópico foi produzida para a disciplina Seminários de Pesquisa em Sociolinguística do Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras de Araraquara (UNESP) e culminou em um artigo que ainda não foi publicado.

testes de percepção com falantes paulistas para compreender em mais detalhes as questões que aqui se abrem⁴⁰.

Portanto, a interpretação que consideramos mais adequada se relaciona com o contato que o falante tem com a norma culta, seja ele através de leitura ou (re)escrita de textos, dado que a segunda e quarta faixa etária, do banco de dados IBORUNA, e a primeira, do banco de dados Projeto SP2010, apresentam uma relação com escolaridade. Nesse sentido, quanto maior a escolaridade, maior o acesso ao letramento; isto é, estamos assumindo que os entrevistados tenham maior contato com a norma culta através de leitura e escrita.

Outro fator extralinguístico selecionado no modelo de regressão logística para o banco de dados Projeto SP2010 foi a região de residência do falante. Observamos na Tabela 14 que há um desfavorecimento da estratégia **padrão com onde** para moradores da região centro velho em comparação com o *intercept*, isto é, moradores na periferia, sendo que a diferença entre as estimativas é significativa, estatisticamente ($p < 0.05$). Desse modo, buscando compreender esse resultado, apresentamos a disposição dos dados para a região e zona de residência de São Paulo à qual o entrevistado pertence nas figuras 13 e 14.

Figura 13 - Proporção de uso das estratégias de relativização segundo a **região de residência** do falante



Fonte: própria

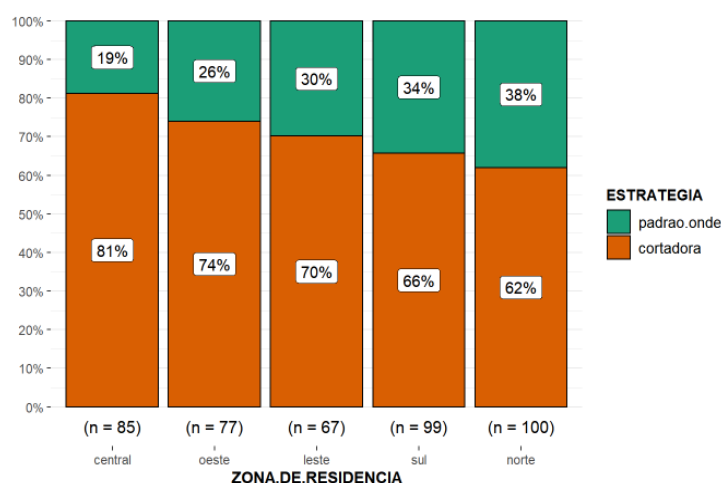
O gráfico exhibe dois movimentos: a região central produzindo mais **cortadora**, enquanto a região periférica tem um uso maior de **padrão com onde**. Nossa hipótese inicial era de que a região central poderia ser responsável pela frequência maior da estratégia padrão, dado que os bairros apresentam uma melhor infraestrutura urbana, em comparação às regiões mais

⁴⁰ Essa questão ultrapassa o escopo da dissertação, mas está em desenvolvimento como um estudo paralelo.

periféricas, o que poderia culminar no perfil mais escolarizado e com mais acesso à norma culta, isto é, responsável pelo maior uso da estratégia **padrão com onde**

A partir dessas discussões, apresentamos na figura 14 a proporção dos usos das estratégias de relativização locativas de acordo com a zona de residência dos falantes do banco de dados Projeto SP2010.

Figura 14 – Proporção de uso das estratégias de relativização segundo a **zona de residência** do falante no banco de dados **SP2010**

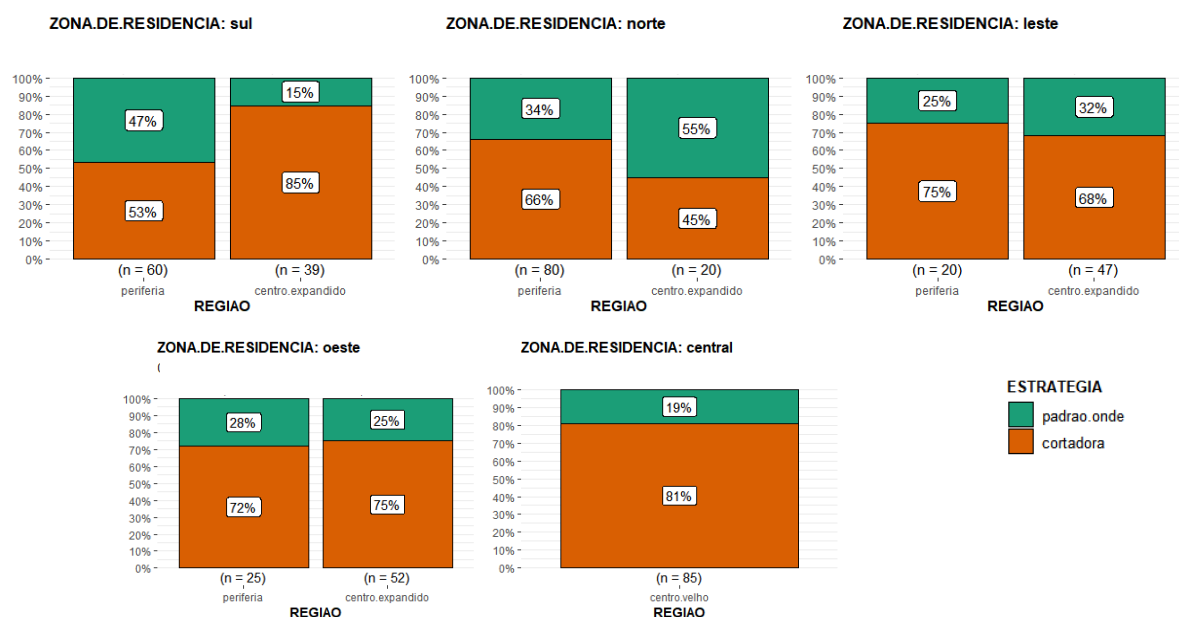


Fonte: própria

O gráfico nos apresenta, em um primeiro momento, a contraposição entre a zona central e as demais zonas da cidade de São Paulo: é ali que se concentra o maior uso da estratégia **cortadora** (81% - 69/85 ocorrências). Por outro lado, a frequência de uso da estratégia **padrão com onde** se concentra na Zona Norte (38% - 38/100 ocorrências).

Para compreender a relação entre a região e a zona de residência do falante e propor uma interpretação a respeito do uso das estratégias pelos falantes paulistanos, apresentamos na figura 15 um cruzamento entre as duas variáveis.

Figura 15 – Proporção de uso das estratégias de relativização segundo a **região de residência** e a **zona de residência** do falante no banco de dados SP2010



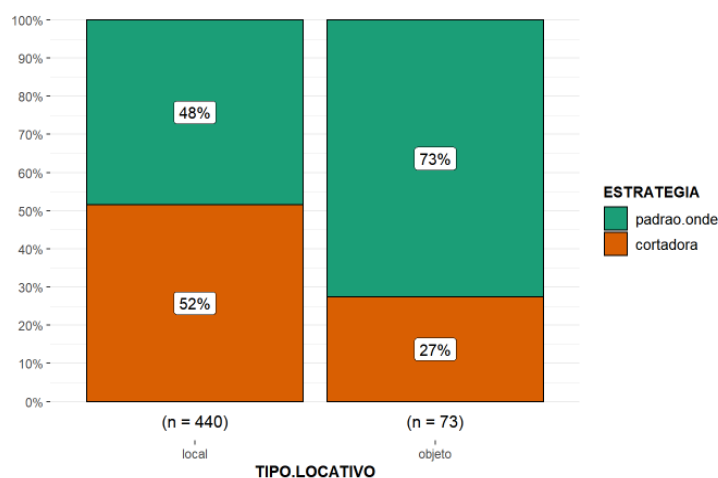
Fonte: própria

O que observamos nos gráficos é uma grande variabilidade: temos zonas em que o uso da **padrão com onde** se destaca na região periférica, tal como a Zona Sul, enquanto em outras essa estratégia se sobressai na região central, tal como a Zona Norte. O que se mantém é o uso mais frequente da **cortadora** para a região central, apresentada como um bairro mais antigo e com melhor infraestrutura urbana, alinhando-se aos resultados encontrados em Tarallo (1983). Tal distribuição dos dados nos impede de fazer afirmações contundentes a respeito da interpretação dos usos das estratégias de relativização, em especial da **padrão com onde**.

Com relação às variáveis linguísticas, há uma diferença entre os dois bancos de dados. No que se refere ao banco de dados IBORUNA, foi selecionada a identidade semântica do antecedente, isto é, o tipo de locativo que compõe o sintagma nominal, em que observamos o desfavorecimento da estratégia **padrão com onde**, em relação ao *intercept*, quando o antecedente é um local.

Como mencionamos na seção 3, *Decisões teórico-metodológicas*, o controle dessa característica se fez necessário pela estruturação da entrevista dentro do banco de dados do interior paulista, isto é, pela tipologia textual Descrição de Espaço que incentiva o entrevistado a produzir um discurso moldado tanto para a descrição de lugares quanto os elementos que o permeiam. Abaixo, nas figuras 16 e 17 exploraremos o resultado dessa variável para os dois bancos de dados.

Figura 16 – Proporção de uso das estratégias de relativização segundo a **identidade semântica do antecedente** no banco de dados **IBORUNA**



Fonte: própria

A partir do gráfico, averiguamos que quando o antecedente é um **local**, tal qual em (4.1), as duas estratégias apresentam uma disposição similar, sendo 52% (227/440 ocorrências) para **cortadora** e 48% (213/440 ocorrências) para **padrão com onde**. Por outro lado, quando o sintagma nominal se refere a um **objeto**, tal qual em (4.2), temos uma predileção pelo uso da **padrão com onde**, sendo 73% (53/73 ocorrências) e 27% (20/73 ocorrências) para **cortadora**.

- (4.1) **Local**: “pra frente você vê uma **uma área de de jantar** *que a gente janta* a gente se alimenta... tem uma pia...” (AC-008; DE: L. 99-100)
- (4.2) **Objeto**: “tem muitas árvores... tem **as mesinhas** *onde os alunos lancham...* e a... e tem a piscina... e tem muitas outras coisas” (AC-013; DE: L. 71-72)

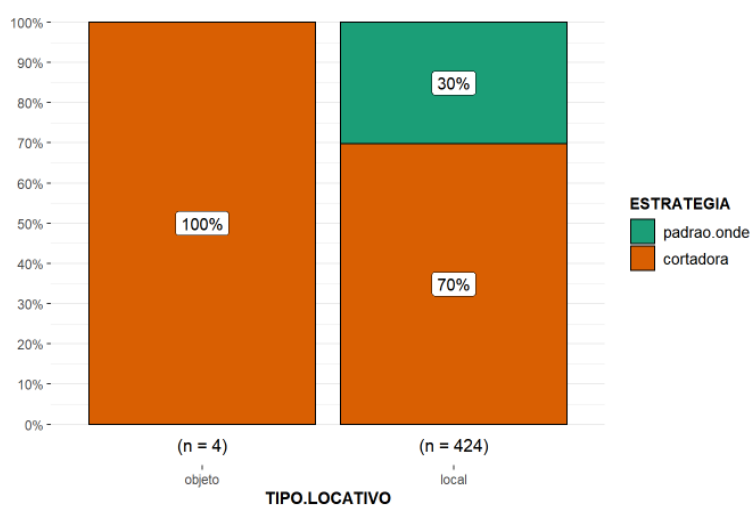
Em (4.1) o antecedente “uma área de jantar” é um espaço físico que está sendo descrito pelo falante; enquanto em (4.2) o sintagma nominal “as mesinhas” se refere a um objeto que compõe o ambiente escolar, não sendo, dessa forma, um lugar prototípico.

Tendo em vista os valores de frequência apresentados, as 73 ocorrências totais de sintagma nominal **objeto** são extremamente baixas em contraposição às 440 ocorrências totais do antecedente **local**, o que pode causar um estranhamento para a sua significância para explicação da variação. Contudo, é necessário retomar como a disposição para o SN **local** se dá de forma quase igualitária (quase 50% para as duas estratégias), enquanto para **objeto** temos valores dispares (73% e 27%). Assim, estatisticamente, a diferença entre os valores esperados e os valores observados se concentra no antecedente **objeto**.

Em vista disso, nossa interpretação desse resultado é de que o pronome relativo **onde** reforça o valor semântico locativo em um contexto em que o antecedente não é um lugar físico prototípico, como na ocorrência (4.2). Esse resultado aponta para uma conclusão que se consolidará mais à frente de que o uso da estratégia **padrão com onde** em contextos menos prototípicos se apresenta devido a uma questão de particularidade de determinadas situações discursivas. Isto é, quando temos uma construção com valores que precisam ser reforçados ou uma necessidade de resolver um possível ruído comunicativo, isto é, uma situação que pode ser interpretada como ambígua, constatamos uma prevalência da estratégia **padrão**, relacionada ao seu pronome relativo encabeçador: o **onde**.

Por outro lado, temos os resultados provenientes do Projeto SP2010 que não apresentam o mesmo padrão, o que podemos verificar na figura 17.

Figura 17 – Proporção de uso das estratégias de relativização segundo a **identidade semântica do antecedente** no banco de dados **SP2010**



Fonte: própria

Antes de nos atentarmos para os valores de proporção apresentados no gráfico, é importante pontuar os valores de frequência, sendo eles: 4 ocorrências (100%) de **objeto** para **cortadora**; 298 ocorrências (69%) de **local** para **cortadora** e 132 ocorrências (31%) para **padrão com onde**. Como há uma concentração quase categórica dos dados no nível 'local', os índices vão corresponder praticamente aos índices gerais observados nesse banco de dados. Assim, a variável não pode ser avaliada para o SP2010.

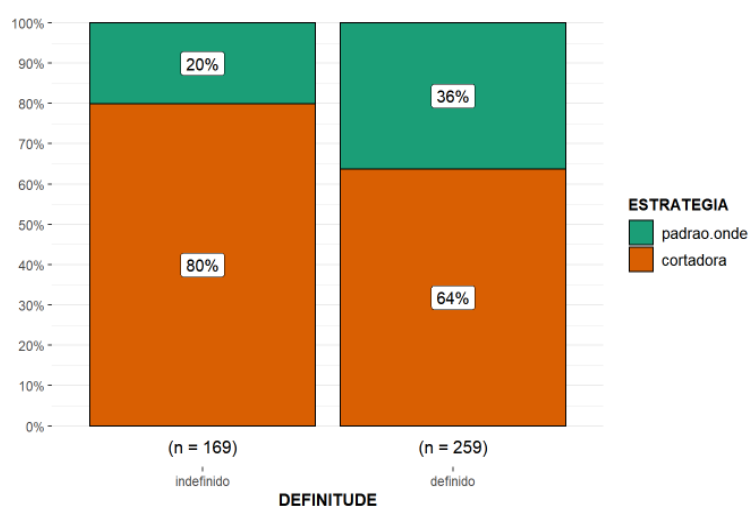
Outrossim, o tipo textual Descrição de Espaço não é compartilhado pelos dois bancos de dados. Portanto, o Projeto SP2010 não apresenta um momento propício para que o falante

possa fazer a enunciação e caracterização de objetos, mas, devido à dinâmica proposta, favorece momentos de descrição de locais da cidade de São Paulo. Desse modo, ao aplicarmos o teste de qui-quadrado, temos como resultado: $X^2 = 0.61221$, (1), $p = 0.434$. Isto é, o valor de $p > 0.05$ demonstra que a associação dos usos das estratégias de relativização e a identidade semântica do antecedente não é significativa estatisticamente para as ocorrências dos paulistanos.

Por fim, a variável linguística selecionada para compor o modelo de regressão de efeitos mistos do banco de dados Projeto SP2010 foi a **definitude** do antecedente. Observamos na Tabela 14 que há um desfavorecimento da estratégia padrão com onde quando o antecedente é indefinido. Ao contrapor com o valor do *intercept*, isto é, com antecedente definido, notamos que a diferença é a diferença entre as estimativas é significativa, estatisticamente ($p < 0.05$).

Desse modo, apresentaremos os valores de definitude do antecedente juntamente com as outras duas variáveis que compõem a categoria semântico-pragmática de nossa análise, ou seja, a **especificidade** e o **status informacional** do antecedente.

Figura 18 – Proporção de uso das estratégias de relativização segundo a **definitude do antecedente** no banco de dados **Projeto SP2010**



Fonte: própria

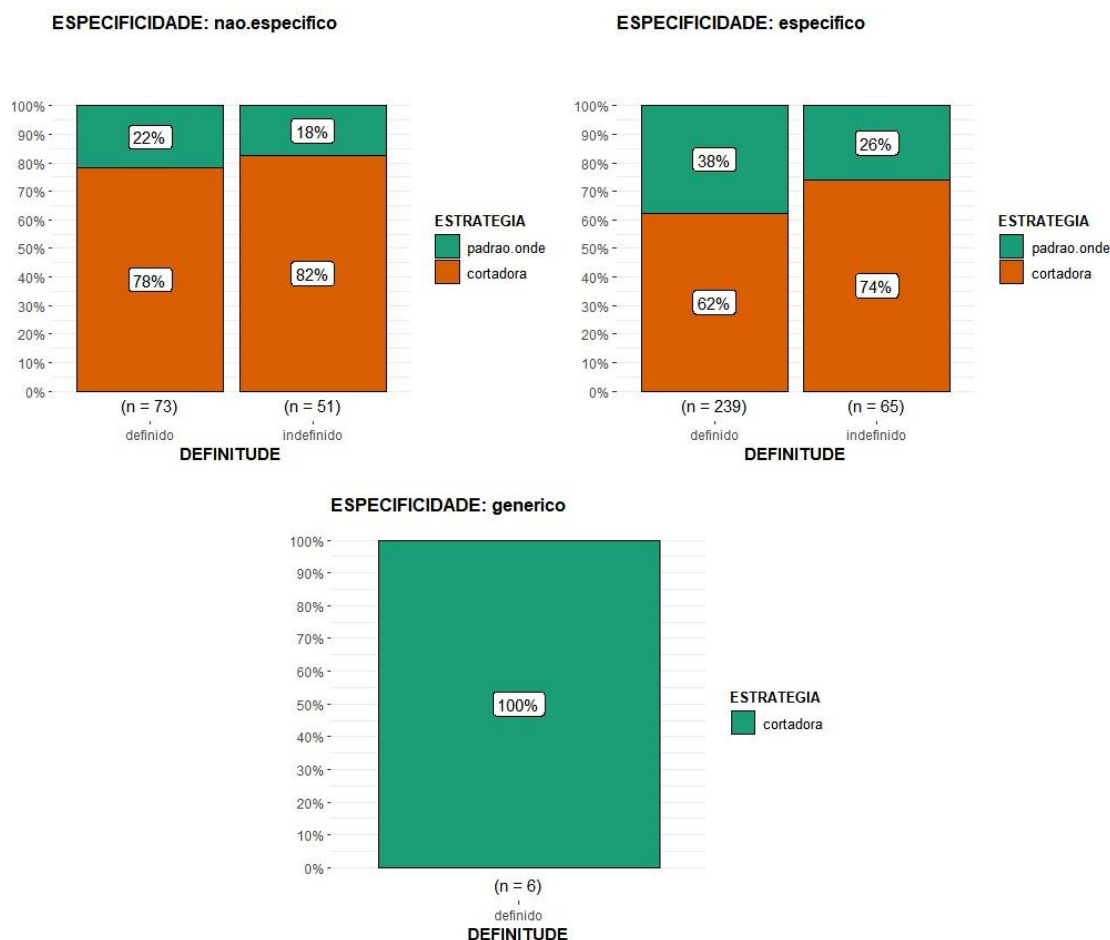
Nossa hipótese inicial era de que encontraríamos um maior uso da estratégia **padrão com onde** em contextos indefinidos, dado que carga semântica locativa de **onde** poderia ser responsável por uma ênfase na informação. Como observamos na Figura 18, o movimento na fala do interior paulista foi o contrário: em contextos definidos houve maior frequência da estratégia padrão. No caso da capital paulista, esse número não é o suficiente para se sobressair à estratégia **cortadora**, visto que essa ocorre em 165 dados (64%) em oposição a 94 dados

(36%) de **padrão com onde**, mas isso se deve ao grande número da estratégia não padrão no *corpus*. Nas ocorrências (4.3), (4.4), (4.5) e (4.6) encontramos dados que representam as possibilidades.

- (4.3) **Definido:** então gosto muito do bairro e... **o pedaço que eu moro** eu acho muito bom (VivianeC, F1SC)
- (4.4) **Definido:** S1 ah várias vezes **na região onde eu moro** é o que mais tem... por isso mesmo que eu falei pra você (PedroS, M2SP)
- (4.5) **Indefinido:** S1 ai é **um lugar que minha tia tem uma chácara** que agora me fugiu (JanainaB, F1MC)
- (4.6) **Indefinido:** S1 então eu não consigo imaginar estar **numa cidade onde todos os filmes não estejam em cartaz** peça que não tenha os museus que não tenha tudo (PolianaM, F2SC)

Em (4.3) e (4.4) observamos que os antecedentes “pedaço” e “região” são precedidos por artigos definidos, “o” e “a”, respectivamente, enquanto em (4.5) e (4.6) os antecedentes “chácara” e “cidade” são precedidos por “uma”. A partir das ocorrências, nossa interpretação do resultado foi que o uso da estratégia **padrão com onde** em contextos definidos se deve por um reforço no valor locativo. Desse modo, acreditamos que a informação colocada em pauta é conhecida pelo falante, o que o motiva a agregar um novo valor locativo sobre ela, reforçando a proximidade sobre o assunto que fala. Contudo, não podemos fazer uma afirmação categórica sem apresentar um cruzamento dessa variável com a **especificidade** e o **status informacional** do antecedente, apresentadas na Figura 19 e 20.

Figura 19 – Proporção de uso das estratégias de relativização segundo a **definitude** e a **especificidade do antecedente** no banco de dados SP2010



Fonte: própria

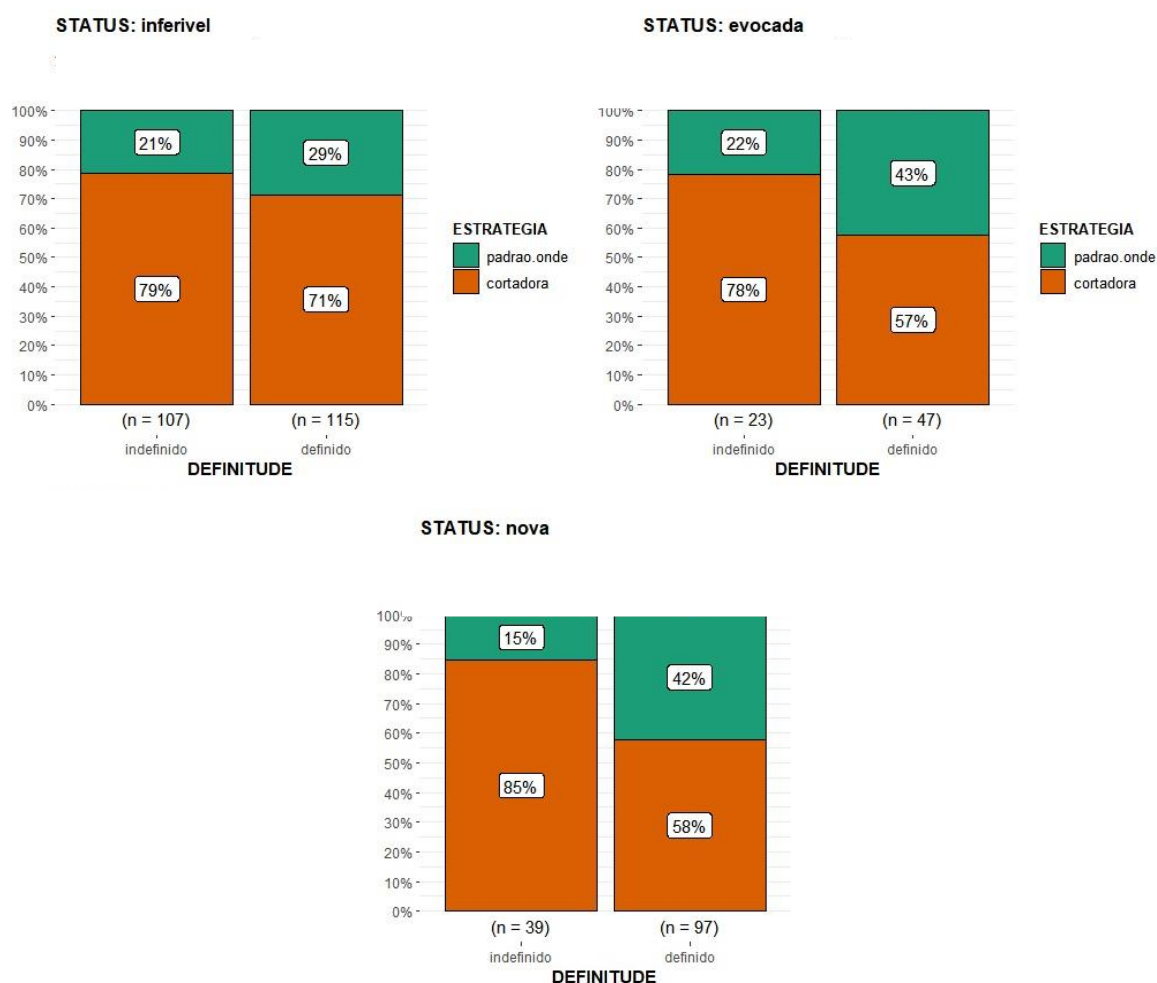
No cruzamento, observamos que contextos **genéricos** e **não específicos** apresentam um número menos frequente de estratégias de relativização, no total 130 dados, comparado ao contexto específico, 304 dados. Além disso, são predominantemente compostos pela estratégia **cortadora**. Enquanto no contexto **específico**, temos uma concentração de dados **definidos**, que apresentam um uso maior da estratégia padrão com onde, isto é, 88 dados (38%).

Recuperando nossa discussão a respeito dessa variável, estamos lidando com uma informação que particulariza um determinado elemento de uma categoria. Assim, ao utilizar um antecedente definido específico, o falante o está reconhecendo como único dentro de uma situação ou referência. Nas ocorrências (4.7) e (4.8) observamos essa situação:

- (4.7) S1 acho que sim **a faculdade onde eu estudei** a primeira faculdade que eu estudei que foi Administração na São Judas (SergioA, M1SC)
- (4.8) S1 é **as escolas que eu estudei** sempre foram no bairro aqui escolas tradicionais (SergioA, M1SC)

Em (4.7) e (4.8) temos um antecedente que é **específico** dentro de uma categoria, pois o falante não está falando sobre qualquer “faculdade” ou “escola”, ele está tecendo comentário sobre instituições específicas que frequentou. Há uma diferença sutil entre as duas ocorrências que provêm, justamente, do reforço do locativo em (4.7): a ideia de que não é uma faculdade qualquer, mas um lugar que lhe é conhecido, assim, definido e específico. No caso de (4.8), apesar de um antecedente conhecido e específico, ele soa menos específico do que a faculdade. Podemos, ainda, averiguar essa discussão com relação ao status informacional do antecedente na Figura 20.

Figura 20 – Proporção de uso das estratégias de relativização segundo a **definitude** e o **status informacional do antecedente** no banco de dados SP2010



Fonte: própria

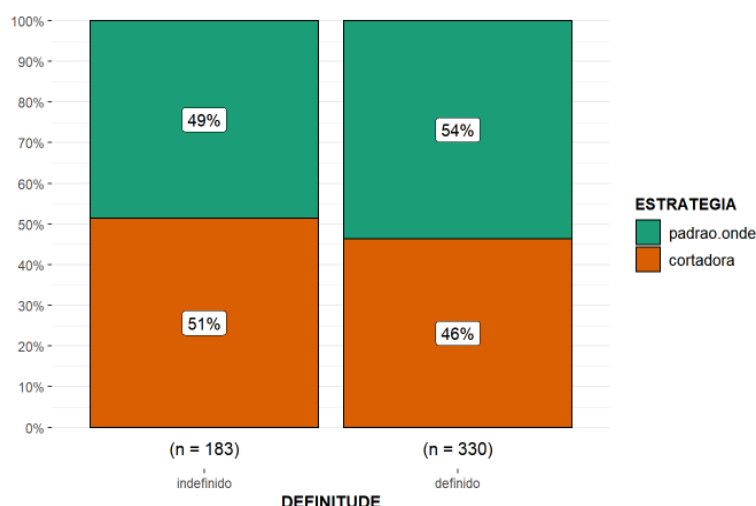
A partir do cruzamento, observamos que a maior frequência da estratégia **padrão com onde** se concentra nos contextos **definidos** e quando a informação é **evocada** e **nova**. Destrinchando um pouco mais o **status informacional** do antecedente, temos que o total de 44 dados de **padrão com onde** na variante nova se caracteriza como **nova em folha**, e em **evocada** dos 20 dados, 9 dados são **evocada situacionalmente** e 11 dados são **evocada textualmente**.

Esses números reforçam, novamente, nossa interpretação de que a estratégia **padrão com onde** reforça uma informação que é conhecida pelo falante, particular para ele no determinado tópico discursivo mobilizado. Isso, pois são contextos em que ele introduz uma nova informação, que se relaciona ao assunto em pauta (nova em folha⁴¹), que já foi mencionada anteriormente (evocada textualmente) ou que se relaciona ao espaço em que a conversa acontece (evocada situacionalmente).

Vale a pena ressaltar que essa interpretação não é categórica, tendo em vista que estamos lidando com uma quantidade pequena de dados. Estamos tecendo considerações, buscando investigar o resultado da análise multivariada, mas não podemos afirmar que esse é o processo mobilizado pelo falante.

No que se refere ao banco de dados IBORUNA, apresentamos na Figura 21 a distribuição da variável definitude do antecedente.

Figura 21 – Proporção de uso das estratégias de relativização segundo a **definitude do antecedente** no banco de dados IBORUNA

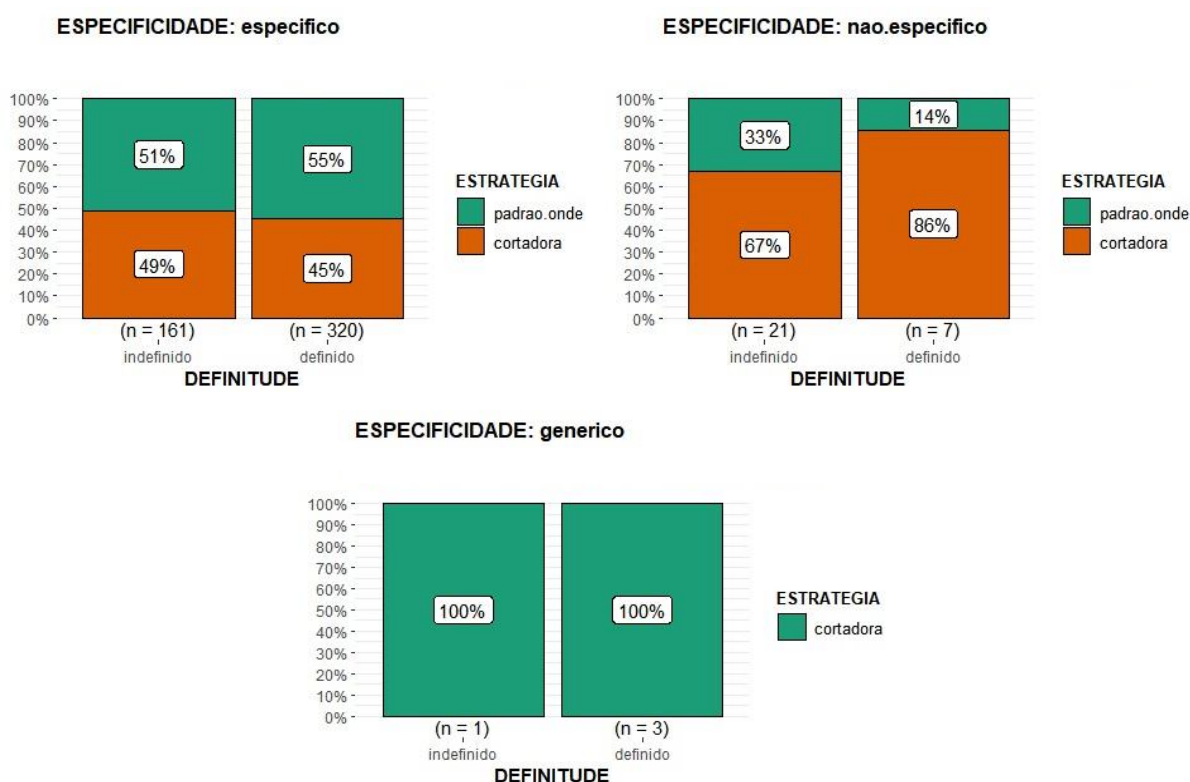


Fonte: própria

⁴¹ Consideramos a informação como nova dentro do contexto da conversa e para o ouvinte. Mas a informação é conhecida pelo falante.

O gráfico nos apresenta uma distribuição equilibrada do uso das estratégias de relativização de acordo com a **definitude** do antecedente, sendo 51% (94/183 ocorrências) para **cortadora** e 49% (89/183 ocorrências) para **padrão com onde** em contextos **indefinidos** e 46% (153/330 ocorrências) para **cortadora** e 54% (177/330 ocorrências) para **padrão com onde** em contextos **definidos**. A partir do teste de qui-quadrado, que apresenta o valor $X^2 = 0.98811$, (1), $p = 0.3202$, constatamos que a associação dos usos das estratégias de relativização e a **definitude** do antecedente não é significativa estatisticamente. Tendo em vista que a variável **definitude** foi construída em conjunto com a variável **especificidade**, apresentamos na figura 22 o seu cruzamento.

Figura 22 - Proporção de uso das estratégias de relativização segundo a **definitude e a **especificidade do antecedente** no banco de dados IBORUNA**



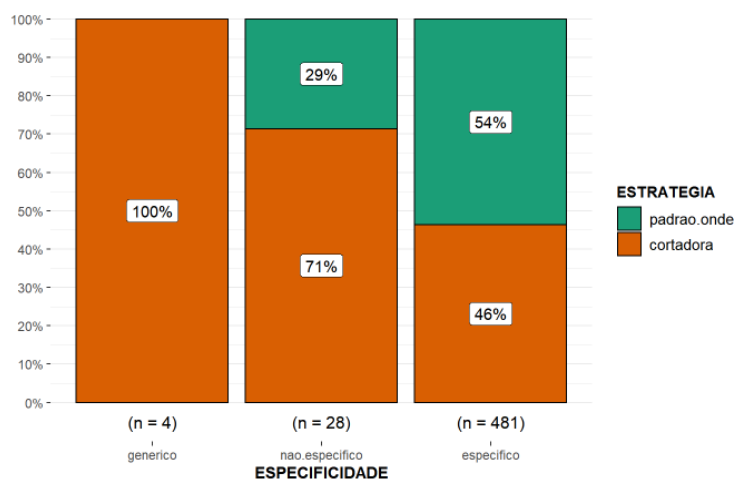
Fonte: própria

O primeiro gráfico apresenta a proporção de uso de antecedentes **indefinidos** e **definidos** para contextos **específicos** e notamos que a distribuição se dá de forma quase igualitária. Para o segundo gráfico, com valores **não específicos**, observamos um uso maior da estratégia **cortadora**. Contudo, ao observar o valor de “n”, isto é, a frequência, são

pouquíssimos dados, que impedem afirmações categóricas. O mesmo ocorre para o terceiro gráfico, com valores **genéricos**.

Tendo em vista o resultado encontrado no banco de dados Projeto SP2010, é necessário apresentar os resultados encontrados no que se refere à **especificidade** do antecedente em nosso *corpus*.

Tabela 24 – Frequência e proporção de uso das estratégias de relativização segundo a **especificidade do antecedente** no banco de dados **IBORUNA**



Fonte: própria

A primeira questão que se apresenta a partir da tabela é a célula vazia para **padrão com onde** no nível **genérico**, com apenas 4 ocorrências com a estratégia **cortadora**. Por si só, esse resultado já indicaria um enviesamento de nossa amostra. Contudo, averiguamos que o nível **não específico** não soma 30 ocorrências para as duas estratégias. Desse modo, a associação dos usos das estratégias de relativização e a especificidade do antecedente não pode ser considerada como significativa estatisticamente. Contudo, podemos apontar para um uso mais frequente da **padrão com onde** em contextos que o antecedente se caracteriza como **específico**, tal qual em (4.9), o que não era esperado⁴².

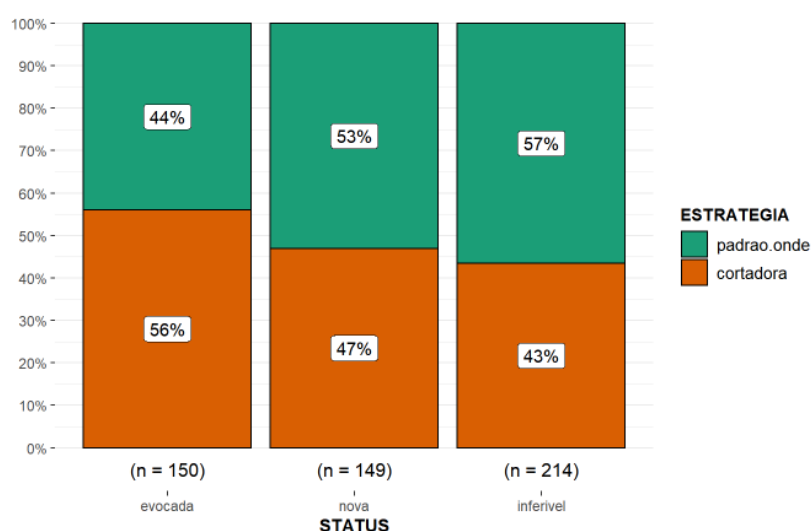
(4.9) “**é e o laboratório... onde eu fico a maior parte do tempo... é ele é composto por uma bancada... é de alvenaria...**” (AC-083; DE: L. 219-220)

⁴² A diferença (46% vs 54%) é muito pequena para apontar favorecimento. Está muito próximo do equilíbrio, o que significa que não é esse fator (nível) que explica a variação. O fato de concentrar a maior parte dos dados faz com o fator acabe reproduzindo os índices gerais, porém consideramos adequado apresentar a ocorrência e fazer o cruzamento para compreender em mais detalhes esses resultados.

Em (4.9) o antecedente “o laboratório” faz referência a um local específico e particular em que o falante trabalha. Assim, nossa interpretação sobre essa variável se torna mais coesa, ao observarmos que os dados do interior paulista também aparentam articular um reforço do valor locativo quando o antecedente se trata de um elemento particular dentro de uma categoria.

Por fim, apresentamos na Figura 23, a proporção de uso das estratégias de relativização de acordo com a variável **status informacional** do antecedente.

Figura 23 – Proporção de uso das estratégias de relativização segundo o **status informacional do antecedente** no banco de dados IBORUNA



Fonte: própria

No gráfico, observamos maior uso da estratégia **cortadora** quando o antecedente já foi **evocado** no discurso – 56% (84/150 ocorrências) – enquanto a **padrão com onde** é utilizada com maior frequência em situações de informação **nova** – 53% (79/149 ocorrências) – e **inferível** – 57% (121/214 ocorrências). Uma possível interpretação para esse resultado é de que ele vai no sentido de atribuir ao uso de **onde** um papel de indicação/reforço do valor locativo, facilitando ao interlocutor ter acesso a essa informação.

4.1.1 Síntese

Em resumo, a análise do modelo de regressão logística com efeitos mistos apontou para o favorecimento da estratégia **padrão com onde** em falantes com **ensino superior** e com idades em um intervalo similar – **26 a 35 anos** para o interior paulista e **19 a 34 anos** para a capital paulista – e nos dois bancos de dados. Com relação às variáveis linguísticas, observamos um

movimento diferente: para o banco de dados IBORUNA, a **identidade semântica do locativo** apontou para uma atuação do contexto na explicação do fenômeno, enquanto para o banco de dados Projeto SP2010 foi a **definitude do antecedente**.

Discutir essa diferença é primordial para compreendermos o processo de variação das estratégias de relativização locativas em contextos preposicionados na fala paulista. Isso, pois estamos trabalhando com uma variável sintática e, como debatemos na fundamentação teórica, há um movimento de identificar as variáveis estruturais como mais apropriadas nas discussões, tendo em vista que elas se mostram mais significativas.

Ressaltamos que não estamos desmerecendo outros trabalhos que encontraram (ou não) correlações com variáveis extralinguísticas, pois a **faixa etária** e a **escolaridade** acabam por ser fatores importantes dentro das investigações sintáticas para identificar o caminho da mudança e o impacto da norma. O que gostaríamos de discutir é como essas variáveis foram coesas em dois bancos de dados tão distintos, enquanto as variáveis linguísticas não apresentaram esse caráter e, assim, permaneceram em segundo plano.

Devemos, sim, ter em mente que a estrutura dos bancos de dados “enviesou” os resultados, no sentido de produzir situações comunicativas distintas. A entrevista, da forma como foi conduzida, no IBORUNA proporcionou momentos de Descrição de Espaço, enquanto no Projeto SP2010 tivemos um enfoque na narração do falante de sua história com a cidade de São Paulo. Isso se apresenta na própria identidade lexical do verbo da oração relativa, tão distinta nos bancos e que mobilizou estruturas sintáticas, semânticas e pragmáticas diferentes.

Desse modo, para a presente investigação, não temos uma coesão dentro das variáveis estruturais por trabalharmos com bancos de dados com propósitos diferentes. Assim, se sobressaem os fatores extralinguísticos para a compreensão do fenômeno de variação das estratégias de relativização locativas no estado de São Paulo. É na idade e escolaridade do falante, dentro de um mercado linguístico, que compreendemos o significado social que as variantes assumem na fala dos paulistas.

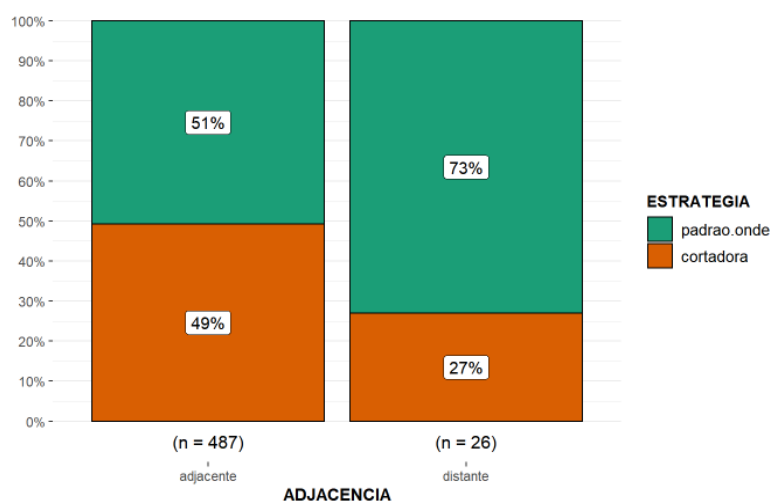
4.2 DISTRIBUIÇÃO DE DADOS DAS VARIÁVEIS NÃO CORRELACIONADAS

A presente subseção é dedicada a descrição das variáveis, linguísticas e extralinguísticas, que não compuseram a análise multivariada. Assim, iremos discutir, de forma breve, como nossas hipóteses se concretizaram ou não em nossos dados.

4.2.1 Adjacência entre o sintagma nominal antecedente e a oração relativa

O controle da variável adjacência entre o sintagma nominal antecedente e a oração relativa se construiu por meio da literatura sobre o fenômeno (MOLLICA, 1977, 2003; PINHEIRO, 1998; BARROS, 2000; BISPO, 2016) que buscou controlar a distância entre o sintagma nominal e a oração relativa. Nas figuras 24 e 25 apresentaremos a distribuição dos dados para essa variável.

Figura 24 – Proporção de uso das estratégias de relativização segundo a **adjacência entre o sintagma nominal antecedente e a oração relativa** no banco de dados IBORUNA



Fonte: própria

A partir do gráfico, identificamos que em contextos de **distância** entre o antecedente e a oração relativa, há um uso mais frequente da estratégia **padrão com onde**. Contudo, se faz necessário uma ressalva e a exposição dos números de frequência das ocorrências, tendo em vista que, quando há adjacência entre o sintagma e a oração encaixada, temos 49% (240/487 ocorrências) de uso da **cortadora** e 50,7% (247/487 ocorrências) de uso da **padrão com onde**;

e quando o sintagma e a oração encaixada estão distantes temos 27% (7/26 ocorrências) para **cortadora** e 73% (19/26 ocorrências) para **padrão com onde**.

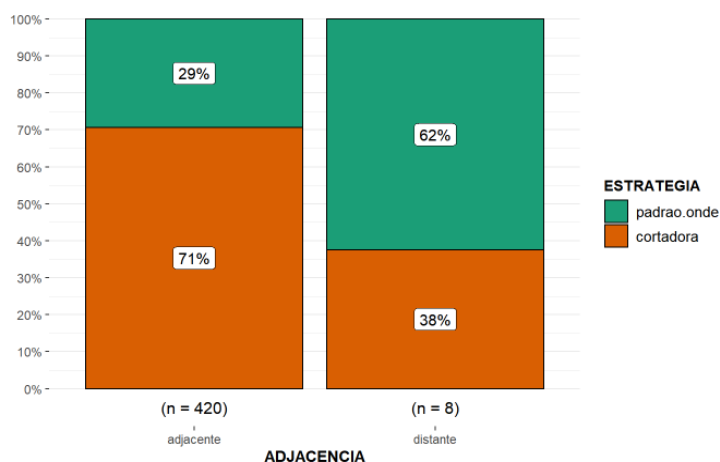
Portanto, verificamos um padrão em nossa análise: em contextos que apresentam uma particularidade dentro da situação discursiva, de menor frequência, há uma prevalência do uso da **padrão com onde**. Com relação a essa variável, em casos como em (4.10), o pronome relativo **onde** reforça o valor locativo da oração, em uma tentativa de recuperar o sintagma que está interposto por outras informações. Em (4.11), por outro lado, observamos que o uso das estratégias se torna intercambiável, dado que o antecedente está expresso e próximo à oração. Ademais, quando contraposto ao **onde**, o pronome relativo **que** não apresenta uma particularidade; pelo contrário, estamos lidando com um conectivo universal. Desse modo, hipotetizamos que a variação das duas estratégias se baseia na contraposição dos valores associados aos pronomes que as encabeçam.

(4.10) **Distante**: “**na outra parte** onde (a gente) esquento o almoço **onde tá as coisas** é tipo uma copa... é onde fica é onde fica as ca(i)xa de papelão que a gente não vai usan(d)o...” (AC-024; DE: L. 247-248)

(4.11) **Adjacente**: “Inf.: aí a gente começô(u) vim na escola... Doc.: ham Inf.: em Mirassol... **na cidade que a gente morava**... no sítio dá... cinco quilômetros” (AC-108; NE: L. 34- 35)

Em (4.10) o sintagma nominal “na outra parte” é seguido de uma informação, “onde a gente esquento o almoço” antes da oração relativa “onde tá as coisas” ser introduzida pelo pronome relativo **onde**; e em (4.11) o antecedente “na cidade” está encaixado na oração relativa “a gente morava”, encabeçada pelo pronome relativo **que**, sem informações adicionais. Baseando-nos na noção de acessibilidade de informação, tal como discutida por Ariel (1988, 1991), podemos interpretar que o valor prototípico, e prescrito pela tradição gramatical, de **onde** contribui com o processamento da construção pelo ouvinte, evitando um contexto de equívoco (ou ambiguidade) sobre o referente.

Figura 25 – Proporção de uso das estratégias de relativização segundo a **adjacência entre o sintagma nominal antecedente e a oração relativa** no banco de dados SP2010



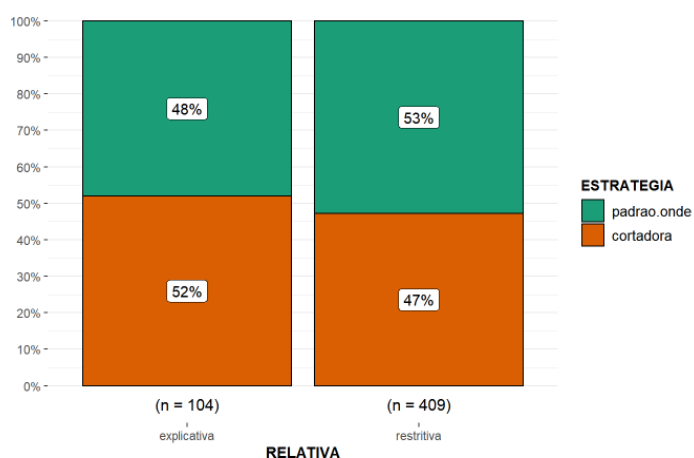
Fonte: própria

Tal qual o resultado do banco de dados IBORUNA, precisamos pontuar os valores de frequência, tendo em vista que são 299 ocorrências (70%) de contextos adjacentes para **cortadora** e 127 ocorrências (30%) para **padrão com onde**; enquanto para contextos distantes temos 3 ocorrências (38%) para **cortadora** e 5 ocorrências (62%) para **padrão com onde**.

4.2.2 Classificação semântica da relativa

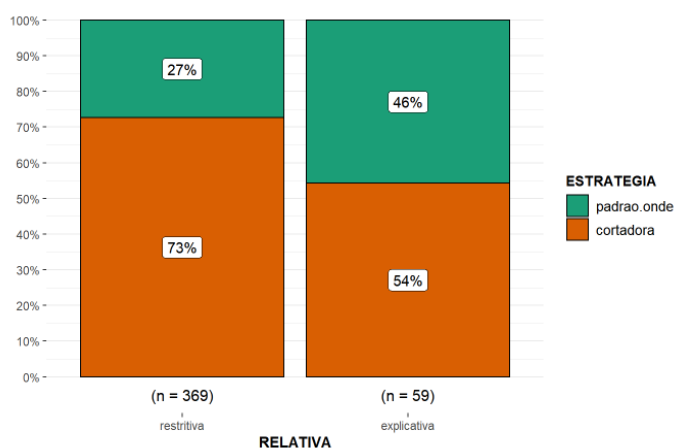
No que se refere ao estudo das orações relativas, há uma clássica distinção da construção em dois tipos: a explicativa e a restritiva. Tomando por base a literatura, foi nosso interesse controlar como essa relação se apresentaria para as estratégias de relativização em contextos locativos. As figuras 26 e 27 apresentam a distribuição nos dois bancos de dados utilizados no presente estudo.

Figura 26 – Frequência e proporção de uso das estratégias de relativização segundo a **classificação semântica da relativa** no banco de dados **IBORUNA**



Fonte: própria

Figura 27 – Frequência e proporção de uso das estratégias de relativização segundo a **classificação semântica da relativa** no banco de dados **SP2010**



Fonte: própria

Identificamos, em ambas as figuras, um uso mais frequente da relativa **restritiva** e um uso menos frequente da relativa **explicativa**. Isso pode ser interpretado como: no contexto locativo, para ambos os bancos de dados, os falantes produzem mais ocorrências como (4.12), caracterizando e restringindo lugares, do que ocorrências como (4.13), somando informações a respeito delas.

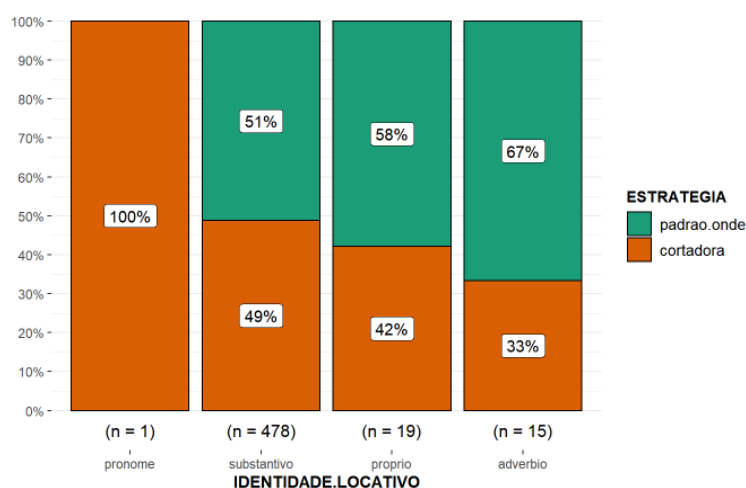
- (4.12) **Restritiva**: “e tem uma... lo(u)sa enorme e um... **muralzinho onde colocam os recados**” (AC-013; DE: L. 61-62)
- (4.13) **Explicativa**: “olha éh Macaubal onde onde eu nasci é uma cidade pequena...” (AC-149; NE: L. 60)

Em (4.12) o antecedente “um muralzinho” é restrito para aquele em que são colocados os recados, enquanto em (4.13) a oração relativa traz uma informação, que age como uma explicação, sobre o antecedente, a cidade “Macaubal”. Contudo, não podemos fazer afirmações categóricas a esse respeito, tendo em vista a baixa frequência dos dados e o fato da variável não ter apresentado diferença estatisticamente significativa entre seus níveis.

4.2.3 Classe morfológica do locativo

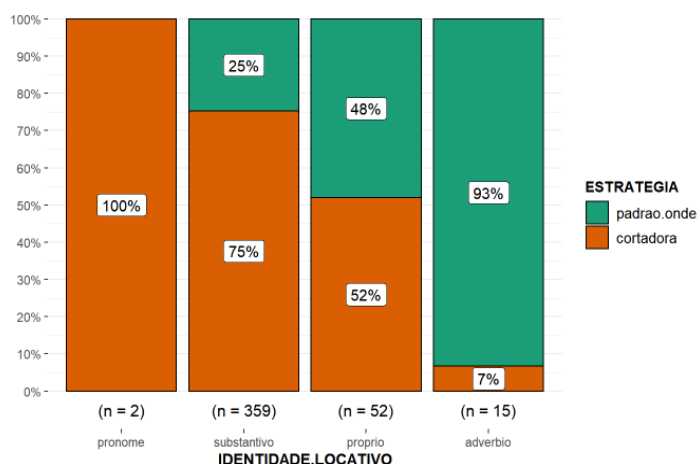
Para a variável classe morfológica do locativo, tínhamos como hipótese de que uma contraposição entre **substantivo** e **nome próprio** se apresentaria em nossos dados, sendo cada um norteado por uma estratégia de relativização. As figuras 28 e 19 apresentam a distribuição dos tipos morfológicos nos dois bancos de dados.

Figura 28 – Frequência e proporção de uso das estratégias de relativização segundo **classe morfológica do locativo** no banco de dados **IBORUNA**



Fonte: própria

Figura 29 – Frequência e proporção de uso das estratégias de relativização segundo **classe morfológica do locativo** no banco de dados **SP2010**



Fonte: própria

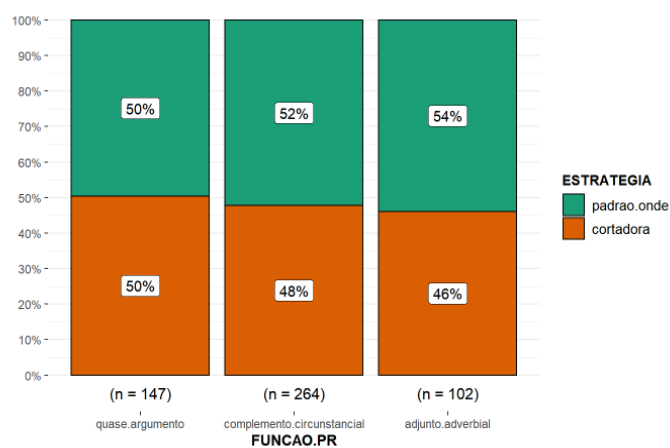
Identificamos, em ambos os gráficos, um uso majoritário do locativo como **substantivo**, enquanto a distribuição nos demais níveis se apresenta com uma frequência muito baixa. Temos células vazias e com ocorrências que não somam 30 dados que estão enviesando a amostra, principalmente no banco de dados Projeto SP2010.

Assim, mesmo que produzíssemos um amálgama dos níveis, os dados ainda se concentram de forma expressiva no locativo como **substantivo**. Portanto, constatamos que a associação dos usos das estratégias de relativização e a identidade morfológica do locativo não é significativa estatisticamente.

4.2.4 Função sintática do pronome relativo

Como pontuado na seção 3, estamos assumindo a função sintática desempenhada pelo antecedente na rede argumental do verbo da oração relativa, sendo correferenciada pelo pronome relativo. Nossa hipótese era de que o uso da estratégia **cortadora** ou da estratégia **padrão com onde** se relacionava com a função sintática do pronome relativo e o grau de encaixamento da sentença. Nas figuras 30 e 31 observamos a distribuição dos resultados para os dois bancos de dados.

Figura 30 – Proporção de uso das estratégias de relativização segundo a **função sintática do pronome relativo** no banco de dados **IBORUNA**



Fonte: própria

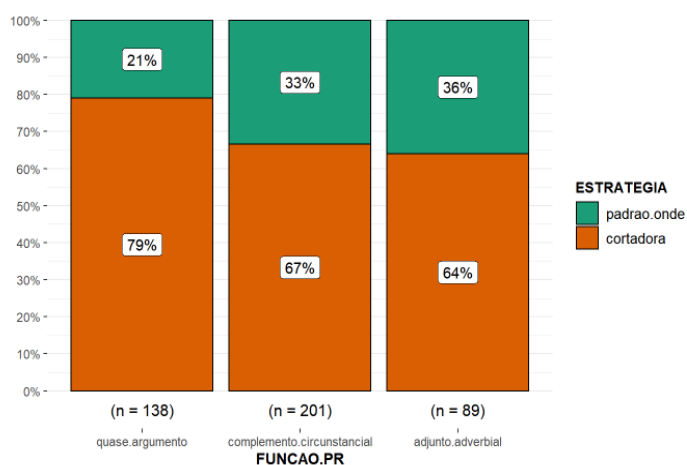
A partir do gráfico, observamos que a distribuição se apresenta de forma equilibrada entre as estratégias de relativização para as funções sintáticas, sendo 50,3% para **cortadora** (74/147 ocorrências) e 49,7% (72/147 ocorrências) para **padrão com onde** na função de **quase argumento**, 48% (126/264 ocorrências) para **cortadora** e 52% (138/264 ocorrências) para **padrão com onde** na função de **complemento circunstancial** e 46% (47/102 ocorrências) para **cortadora** e 54% (55/102 ocorrências) para **padrão com onde** na função de **adjunto adverbial**.

Desse modo, nossa hipótese não se confirma, tendo em vista que o valor de $p > 0.05^{43}$, averiguado por meio do teste de qui-quadrado, constata que a associação dos usos das estratégias de relativização e a função sintática do pronome relativo não é significativa estatisticamente.

Em contrapartida, os resultados encontrados no banco de dados Projeto SP2010 nos apresentam um padrão diferente para o uso das estratégias de relativização locativas a partir da função sintática do pronome relativo. A distribuição se apresenta, abaixo, na figura 33.

⁴³ $X^2 = 0.47666$, (2), $p = 0.7879$

Figura 31 – Proporção de uso das estratégias de relativização segundo a **função sintática do pronome relativo** no banco de dados **SP2010**



Fonte: própria

O gráfico nos apresenta o padrão que foi observado anteriormente para as demais variáveis linguísticas: quando estamos diante de um contexto singular, que apresenta particularidade dentro das situações discursivas, isto é, em que se faz necessário um reforço do valor locativo, há uma frequência maior de **padrão com onde**. Isso, pois, para a função sintática adjunto adverbial, temos 64 % (58/90 ocorrências) de **cortadora** e 36 % (32/90 ocorrências) de **padrão com onde**. No entanto, é preciso considerar que o nível **complemento circunstancial** apresenta um índice de **padrão com onde** comparável ao contexto de **adjunto adverbial**, o que parece contrariar nossa hipótese.

Esperávamos uma gradação no uso de **padrão com onde**: seria menos frequente em contexto de complemento e de quase-argumento e apresentaria uma frequência mais elevada em contexto de **adjunto**. Apenas o nível **quase-argumento** segue essa tendência.

Nossa hipótese se baseia na natureza do vínculo que o elemento representado pelo pronome relativo mantém com o verbo da oração relativa. Assim, quando a função sintática do pronome é de **complemento circunstancial**, há um vínculo forte entre esse elemento e o verbo, e um compartilhamento do traço locativo, já que o complemento faz parte da rede argumental do verbo. Desse modo o sentido locativo não precisaria ser reforçado por um pronome específico, fazendo com que um conectivo universal pudesse ser utilizado sem trazer problemas de ambiguidade à construção. Teríamos um contexto favorável ao uso da estratégia **cortadora**.

Por outro lado, quando a função sintática do pronome é de **adjunto adverbial**, há um vínculo um pouco mais frouxo entre o elemento e o verbo. A dependência é mais semântica que sintática, configurando uma relação de hipotaxe (NEVES, BRAGA, 2016). O valor locativo

não está previsto na estrutura argumental do verbo. Nesse contexto, o uso da **padrão com onde** asseguraria melhor a interpretação do vínculo semântico e do valor locativo.

Por fim, no caso do **quase argumento**, teríamos um processo intermediário: não há um compartilhamento do valor locativo entre o verbo e o antecedente, porém ele ainda se comporta como um argumento do verbo. Em (4.14), (4.15) e (4.16), as ocorrências ilustram a discussão proposta:

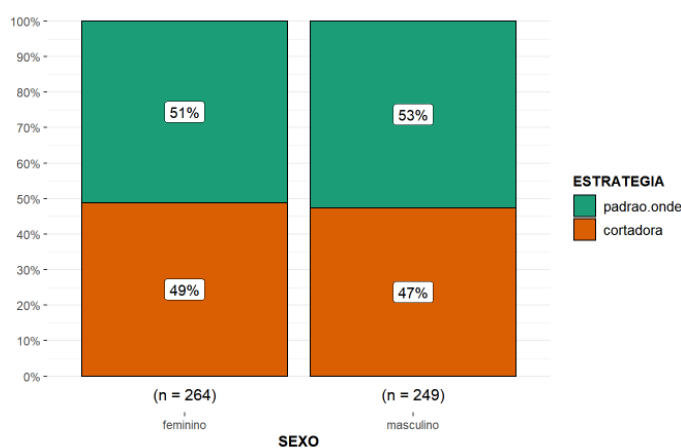
- (4.14) **Complemento circunstancial**: “eu estudei no Cambuci que foi **o primeiro bairro que eu morei** S1 depois eu fui pra Vila Mariana” (MeireC, F2MC)
- (4.15) **Adjunto adverbial**: “S1 então eu não consigo imaginar estar **numa cidade onde todos os filmes não estejam em cartaz** peça que não tenha os museus que não tenha tudo” (PolianaM, F2SC)
- (4.16) **Quase argumento**: “morei na Zona Leste S1 morei... **a Zona Oeste que eu nasci** e Zona Sul D1 nossa” (VeraD, F3MP)

Em (4.14) o pronome relativo **que** correferencia a função sintática de **complemento circunstancial** do antecedente, visto que “o primeiro bairro” faz parte da rede argumental do verbo “morar”. Em (4.15) a oração relativa apresenta uma estrutura completa, “todos os filmes não estejam em cartaz”, e o antecedente se apresenta como um adjunto ao predicado “estar em cartaz”. Por fim, em (4.16) o verbo “nascer” não apresenta um argumento locativo, mas o antecedente “Zona Oeste” complementa a ação mobilizada pelo falante em sua narração.

4.2.5 Sexo/gênero do falante

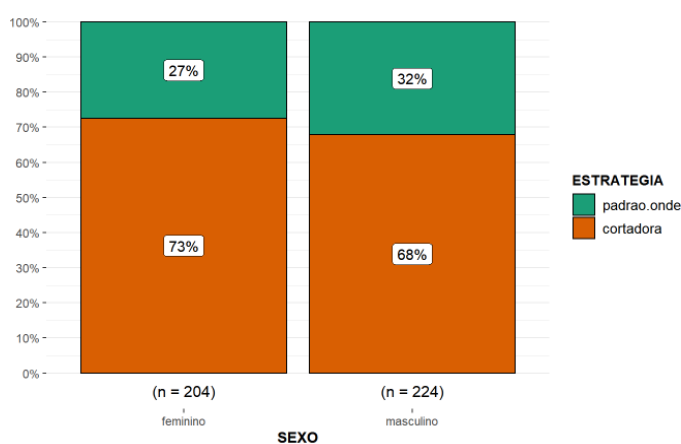
Apresentamos nas figuras 32 e 33 a distribuição dos usos das estratégias de relativização para a última variável extralinguística: o sexo/gênero do falante.

Figura 32 – Frequência e proporção de uso das estratégias de relativização segundo o **sexo** do falante no banco de dados **IBORUNA**



Fonte: própria

Figura 33 – Frequência e proporção de uso das estratégias de relativização segundo o **sexo** do falante no banco de dados **SP2010**



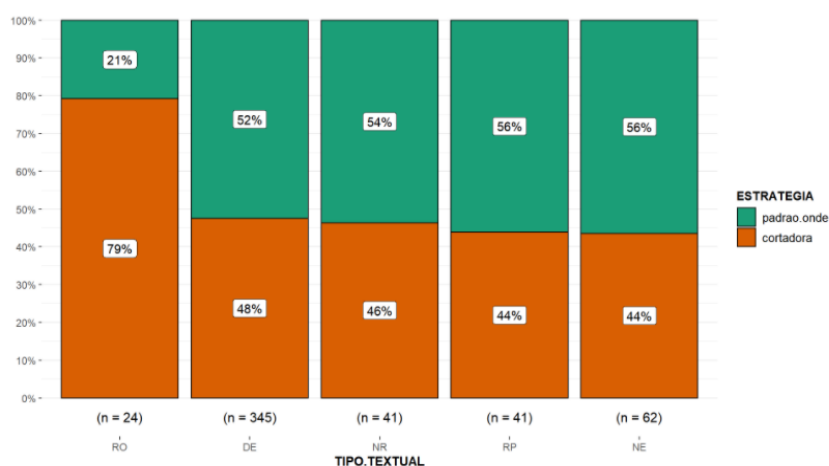
Fonte: própria

Na Figura 34, identificamos que os usos das estratégias tanto para o sexo feminino quanto para o sexo masculino se mantêm em torno da proporção de 50% para ambos, o que apresenta uma diferença que não é estatisticamente significativa. Na Figura 35, os valores são mais díspares, no sentido de que há um índice de frequência maior da **cortadora**, porém as diferenças nos usos desta estratégia entre os sexos também não se mostram significativas.

4.2.6 Tipologia Textual

Para variáveis sintáticas, é interessante observar a sua correlação com o gênero de produção textual, em contextos escritos, dado que haveria uma correlação maior de explicação dos fenômenos a partir de elementos linguísticos do que extralinguísticos (MILROY, GORDON, 2005; BRANDÃO, 2018; ALMEIDA, 2020). Para o banco de dados IBORUNA, temos uma relação de proximidade a isso quando controlamos o tipo textual em que a construção foi produzida. Essa variável se mostrou muito produtiva para a interpretação dos resultados da pesquisa anterior (ver seção 2) e podemos observar a sua distribuição para o presente momento da investigação na figura 34.

Figura 34 – Frequência e proporção de uso das estratégias de relativização segundo o **tipo textual** no banco de dados **IBORUNA**



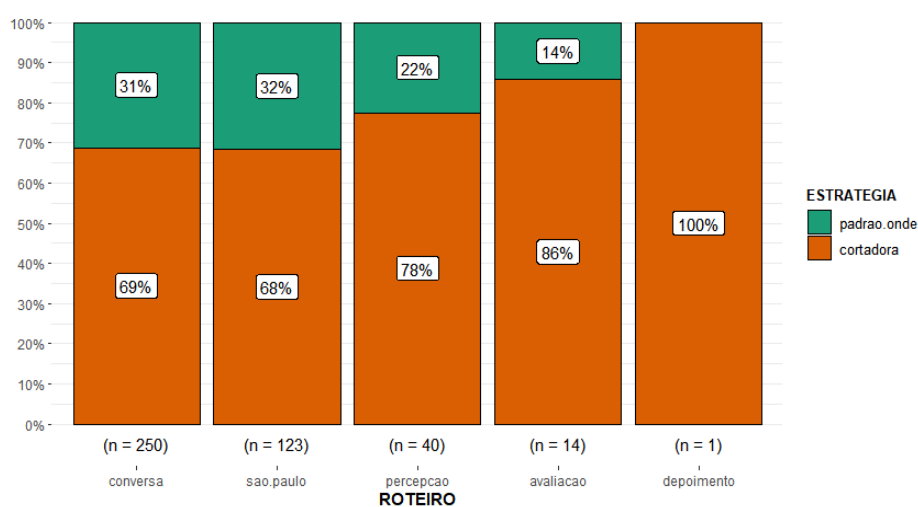
Fonte: própria

A partir do gráfico, observamos que a maior frequência de ocorrências das estratégias de relativização locativa se concentra no tipo textual **Descrição de Espaço (DE)**, assim como hipotetizamos, tendo em vista o valor semântico associado a esse tipo enunciativo. Nesse contexto, há um equilíbrio de uso das duas estratégias que se repete para os demais níveis da variável, mesmo que estes concentrem poucos dados, o que impede grandes afirmações sobre isso. O único tipo que apresenta uma relação mais díspar é o **Relato de Opinião (RO)**, com uso mais frequente da **cortadora**. Porém, nota-se que os valores dessas células não chegam a 30 ocorrências, o que impede afirmações categóricas.

4.2.7 Roteiro de entrevista

Tendo em vista a divisão dos momentos da entrevista na constituição do banco de dados SP2010, consideramos adequado controlar o roteiro como uma variável, buscando identificar em quais momentos haveria um momento mais propício para o processo de variação das estratégias.

Figura 35 – Frequência e proporção de uso das estratégias de relativização segundo o **roteiro de entrevista** no banco de dados **Projeto SP2010**



Fonte: própria

Observamos no gráfico que ocorreu um movimento similar àquele da tipologia textual do banco de dados IBORUNA: houve uma concentração de dados em uma parte específica da entrevista. No caso do Projeto SP2010, foi no momento da conversa, isto é, ao falar sobre lazer, infância, família, trabalho e bairro, os paulistas utilizaram mais estratégias de relativização. Devemos pontuar, também, o momento em que a conversa se focou na cidade de São Paulo, agrupando um número considerável de dados do fenômeno. Assim, nossa hipótese inicial se concretizou, porém sem grandes contribuições para compreensão do processo de variação.

4.3 O CASO DA COPIADORA E PADRÃO COM PREPOSIÇÃO

As duas estratégias com menor frequência dentro de nosso *corpus*, a **copiadora** e a **padrão com preposição**, apontam para algumas tendências interessantes, apesar de consistirem em um total de 13 e 50 dados, respectivamente. Esses baixos números, por si só, já apontam uma informação: em contextos preposicionados, tanto a variante não padrão quanto a variante padrão com preposição não são produtivas. É nosso interesse entender nessa subseção porque isso poderia estar acontecendo na fala paulista. Ressaltamos, contudo, que todas as propostas de interpretação não podem ser tomadas como categóricas, dada a quantidade limitada dos dados.

Na literatura do fenômeno, observamos como é recorrente a falta de um número robusto de dados da estratégia **padrão com preposição**. Desde Tarallo (1983), compreendemos a variante inovadora, isto é, a **cortadora**, em um processo de ascensão no Português Brasileiro. As hipóteses para esse processo podem ser divididas em duas esferas: a falta de naturalidade da estratégia ou um processo cognitivo de economia.

A primeira hipótese, muito discutida por Corrêa (1998) e Mollica (2003), ressalta que a estratégia **padrão com preposição** deveria ser aprendida no contexto escolar. Porém, as autoras apontam uma falha dentro desse processo, que culmina no barramento da estratégia **copiadora** e o não aprendizado da **padrão**. Nesse contexto, a estratégia **cortadora** se apresenta como “neutra”, isto é, uma alternativa que não é padrão, mas que não carrega um significado social tão negativo.

Partindo dessa discussão, a segunda hipótese pode ser sinteticamente explicada por Bispo e Cunha (2019), à luz da perspectiva givoniana. Assim, os autores mobilizam os conceitos de marcação e iconicidade para compreenderem a disposição das estratégias nos estudos linguísticos. Eles pontuam que a estratégia **padrão com preposição**

É marcada em oposição ao caráter não marcado desta última [cortadora]. Do ponto de vista da expressividade, porém, a situação se inverte: por se tratar de uma estrutura menos longa e, principalmente, por reduzir o esforço de codificação/decodificação, em nome da eficácia, a cortadora é mais expressiva que a sua correlata padrão, ou melhor, é marcada expressivamente. (BISPO, CUNHA, 2019, p. 154)

Desse modo, a discussão dos autores perpassa a relação da extensão da estratégia com seu esforço cognitivo de processamento. Seguindo uma escala, teríamos a **copiadora** como a estratégia mais longa, seguida da **padrão com preposição** e, por fim, a **cortadora**. Contudo, como reforçam os autores, a estratégia padrão é a mais complexa cognitivamente das três por

envolver um conhecimento da regência verbal. A partir disso, teríamos o *continuum* exposto na Figura 36:

Figura 36 – *Continuum* de marcação (givoniana) aplicado às estratégias de relativização



Fonte: adaptado de Bispo e Cunha (2019)

A partir das discussões, compreendendo como essas hipóteses justificam os baixos índices em nosso *corpus*, gostaríamos de propor algumas discussões a respeito de determinados fatores linguísticos e extralinguísticos. O primeiro deles é a adjacência entre o sintagma nominal e a oração relativa, exposto por meio das Tabelas 15 e 16.

Tabela 15 – Frequência e proporção de uso das estratégias de relativização segundo a adjacência entre o sintagma nominal antecedente e a oração relativa no banco de dados IBORUNA

ADJACENCIA	ESTRATEGIA		Total
	padrao.pp	copiadora	
adjacente	3 8.1 %	34 91.9 %	37 100 %
distante	0 0 %	1 100 %	1 100 %
Total	3 7.9 %	35 92.1 %	38 100 %

Fonte: própria

Tabela 16 – Frequência e proporção de uso das estratégias de relativização segundo o **a adjacência entre o sintagma nominal antecedente e a oração relativa** no banco de dados **SP2010**

ADJACENCIA	ESTRATEGIA		Total
	padrao.pp	copiadora	
adjacente	10 41.7 %	14 58.3 %	24 100 %
distante	0 0 %	1 100 %	1 100 %
Total	10 40 %	15 60 %	25 100 %

Fonte: própria

Ressaltamos, mais uma vez, a baixa frequência dos dados, motivo pelo qual não faremos nenhuma afirmação categórica a respeito das estratégias. Contudo, é interessante observar como a **copiadora** e a **padrão com preposição** foram produzidas, quase categoricamente, em contextos adjacentes nos dois bancos de dados, tal como nas ocorrências (4.17), (4.18) e (4.19).

- (4.17) **Adjacente:** nós tem uma made(i)ra igual **uma lo(u)sa que só pode pô(r) ali** (AC-003; DE: L. 270)
- (4.18) **Adjacente:** eu vou descrever a **casa em que eu estou morando...** (AC-050; DE: L. 194)
- (4.19) **Distante:** **tem a Praça Pereira Coutinho** que é uma delícia **que eu ando todo o santo dia com o cachorro lá** (RomuloS, M3MC)

Em (4.17) e (4.18) identificamos os antecedentes, “uma lousa” e “a casa” respectivamente, seguidos imediatamente pela oração relativa encabeçada pelo pronome relativo. Por outro lado, em (4.19), o sintagma nominal “a Praça Pereira Coutinho” apresenta a informação “que é uma delícia” antes de a oração relativa “eu ando todo o santo dia com o cachorro lá” ser introduzidas pelo pronome, isto é, há uma descrição inserida antes da oração propriamente dita.

Tendo em vista a discussão sobre a demanda cognitiva que as duas estratégias produzem, é justificado seu uso categórico em contextos adjacentes. Isso, pois em contextos distantes seria mobilizado ainda mais esforço pelo falante para recuperar o antecedente ao qual a oração relativa se liga.

No que se refere aos fatores extralinguísticos, apresentaremos a faixa etária do falante nas Tabelas 17 e 18.

Tabela 17 – Frequência e proporção de uso das estratégias de relativização segundo a **faixa etária** do falante no banco de dados IBORUNA

FAIXA	ESTRATEGIA		Total
	padrao.pp	copiadora	
primeira	2 13.3 %	13 86.7 %	15 100 %
segunda	0 0 %	11 100 %	11 100 %
terceira	0 0 %	8 100 %	8 100 %
quarta	1 25 %	3 75 %	4 100 %
Total	3 7.9 %	35 92.1 %	38 100 %

Fonte: própria

Tabela 18 – Frequência e proporção de uso das estratégias de relativização segundo a **faixa etária** do falante no banco de dados SP2010

FAIXA.ETARIA	ESTRATEGIA		Total
	padrao.pp	copiadora	
primeira.faixa	0 0 %	3 100 %	3 100 %
segunda.faixa	0 0 %	5 100 %	5 100 %
terceira.faixa	10 58.8 %	7 41.2 %	17 100 %
Total	10 40 %	15 60 %	25 100 %

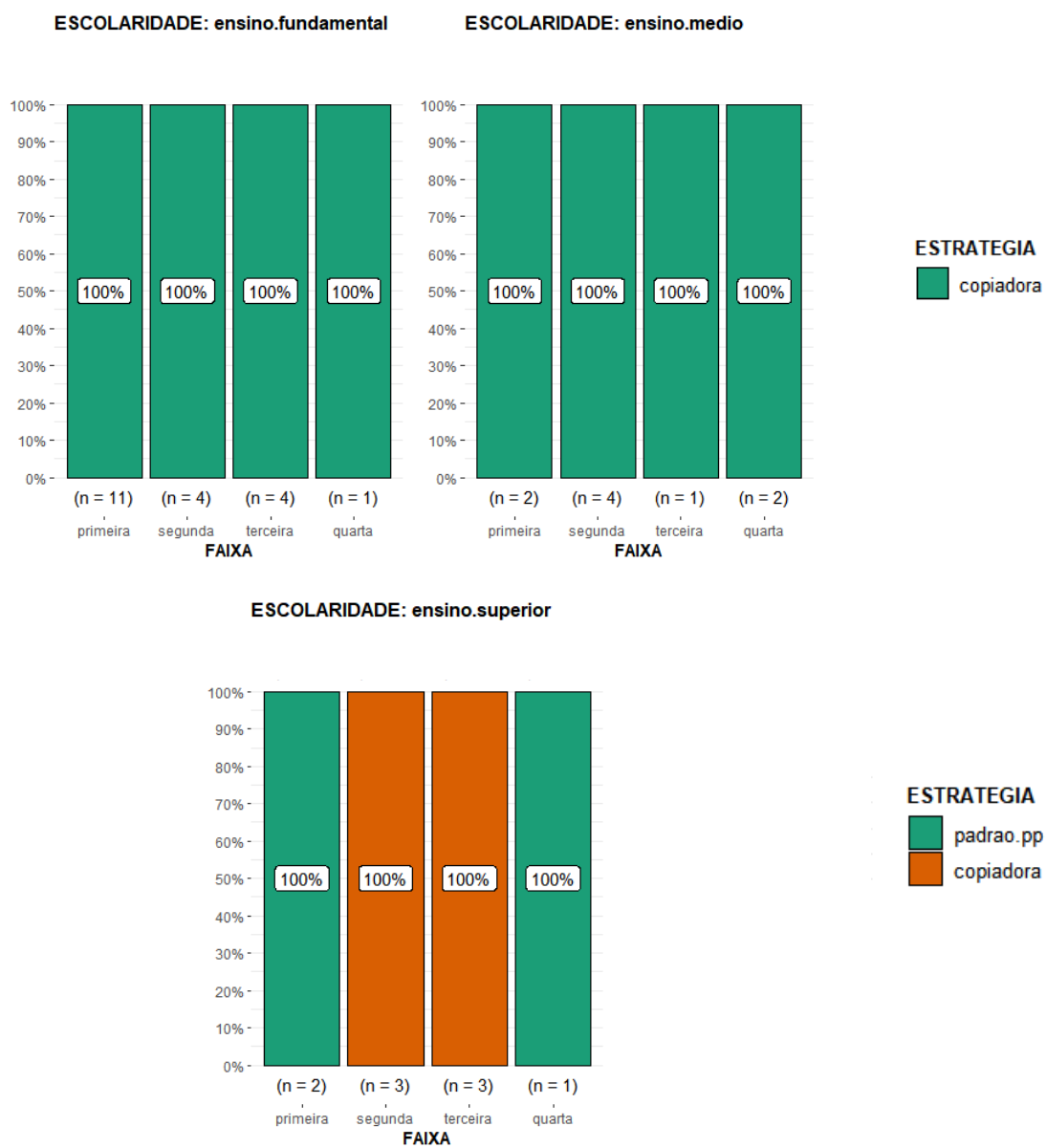
Fonte: própria

Observamos, através da tabela, movimentos distintos nos bancos de dados. No caso do IBORUNA, há uma diminuição do uso da **copiadora** conforme a faixa etária, enquanto no

Projeto SP2010 há um aumento. Com relação ao primeiro, esse percurso se justifica pelo valor negativo associado à estratégia, o que parece barrar o seu uso; além disso podemos articular uma relação com a escolaridade, visto que o contato com a norma é extremamente produtivo para essa diminuição no uso da estratégia não padrão. Com relação ao segundo caso, também interpretamos os resultados em função do contato do falante com o ensino. Abaixo reproduziremos, nos gráficos 39 e 40, o cruzamento das duas variáveis – **faixa etária e escolaridade**.

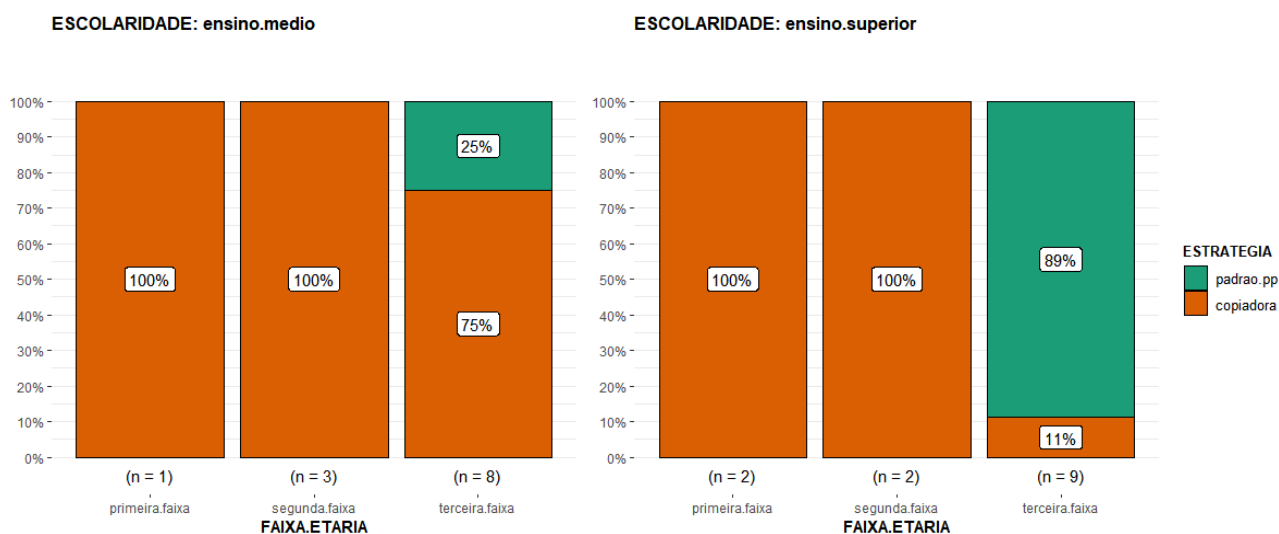
Ademais, é interessante observar que a terceira faixa etária é responsável por todos os usos de **padrão com preposição** na capital paulistana. Ainda mais curiosos os resultados se tornam, ao lembrarmos que falantes com essa faixa etária foram os responsáveis pelos maiores usos da estratégia **cortadora**.

Figura 37 – Proporção de uso das estratégias de relativização segundo a **faixa etária** e a **escolaridade** do falante no banco de dados IBORUNA



Fonte: própria

Figura 38 – Proporção de uso das estratégias de relativização segundo a **faixa etária** e a **escolaridade** do falante no banco de dados SP2010



Fonte: própria

Por meio dos gráficos, confirmamos nossa interpretação anterior. São falantes do ensino superior, de forma quase categórica, os responsáveis pelos usos da estratégia **padrão com preposição**, enquanto falantes menos escolarizados são responsáveis pela frequência da estratégia **copiadora**.

Em suma, a baixa incidência das estratégias **copiadora** e **padrão com preposição** locativas na fala paulista se dá por dois motivos: o valor negativo da estratégia não padrão e a falta de naturalidade da estratégia padrão. Observamos, ainda, que devido ao esforço cognitivo mobilizado, elas se apresentaram em nosso *corpus* em contextos adjacentes. Por fim, por meio da análise dos fatores extralinguísticos, identificamos que são falantes com ensino superior que produzem a **padrão com preposição**, enquanto os falantes menos escolarizados usam a **copiadora**.

5 CONSIDERAÇÕES FINAIS

O presente estudo tinha como objetivo principal averiguar a variação de formas na expressão do valor prototípico de lugar. Para que isso fosse possível, mobilizamos um envelope de variação com as estratégias de relativização locativas em contextos preposicionados na fala paulista, obtendo, assim, as variantes: **padrão com preposição** (*piedpiping*), **padrão com onde**, **copiadora** e **cortadora**.

Tendo em vista que esse momento da pesquisa se insere em uma investigação maior com o objetivo de compreender os usos de **onde**, foi necessário, na seção 2, apresentar uma breve contextualização a respeito de passos anteriores da pesquisa. Além disso, consideramos adequado trazer algumas provocações provenientes de uma perspectiva interlinguística, articulando como o processo que está ocorrendo com nosso pronome relativo locativo foi observado em outros idiomas.

Ainda nessa seção, apresentamos uma breve fundamentação teórica a respeito das estratégias de relativização, ressaltando como nosso fenômeno se justifica pela falta de referências a respeito do contexto preposicionado locativo no Português Brasileiro. Nosso enfoque perpassou os trabalhos que discutiam o impacto das variáveis extralinguísticas no *corpus* tendo em vista a natureza de nossa análise. Por último, apresentamos nosso apoio teórico-metodológico, a Teoria de Variação e Mudança Linguística, assim como alguns problemas condizentes ao trabalho com variáveis sintáticas.

Sobre esse último, elencamos três principais problemáticas: (i) a baixa frequência de dados; (ii) a falta de correlação com fatores sociais; e (iii) a necessidade de suporte teórico para a análise. Nesse sentido, nosso trabalho apresenta uma quebra, visto que temos um número expressivo de dados (1007 no total) e as variáveis extralinguísticas se mostraram mais fortes na explicação do fenômeno em variação.

Dando continuidade ao trabalho desenvolvido, na seção 3 apresentamos nossas decisões teórico-metodológicas. Assim, descrevemos os bancos de dados com os quais trabalhamos, sendo eles, o banco de dados IBORUNA do projeto ALIP (Gonçalves, 2007), representando a fala do interior paulista, e o banco de dados Projeto SP2010 (MENDES, 2013), representando a fala da capital paulistana. Evidenciamos as diferenças presentes na constituição e objetivo dos bancos de dados que foram extremamente importantes para nossa análise.

Ademais, descrevemos as variáveis linguísticas que controlamos em nosso *corpus*, sendo elas: (i) função sintática do pronome relativo; (ii) adjacência entre o sintagma nominal

antecedente e a oração relativa; (iii) classe morfológica do locativo; (iv) identidade semântica (ou tipo) do locativo; (v) classificação semântica da relativa (em explicativa ou restritiva); (vi) definitude do antecedente; (vii) especificidade do antecedente; e (viii) status informacional do antecedente. Reforçamos, também, as variáveis extralinguísticas provenientes dos bancos de dados: (i) sexo/gênero; (ii) escolaridade; (iii) faixa etária (presentes nos dois bancos); (iv) tipologia textual (presente no IBORUNA); (v) região de residência de São Paulo; (vi) zona de residência de São Paulo; (vii) roteiro da entrevista (presentes no Projeto SP2010).

Na seção 4, subdividimos a nossa análise em três momentos: (i) a discussão da análise de regressão logística de efeitos mistos; (ii) a descrição da análise univariada; e, por fim, (iii) a exposição de algumas considerações a respeito das estratégias copiadora e padrão com preposição.

Isso se deu pela amostra com que estávamos trabalhando, já que, para o banco de dados IBORUNA, obtivemos 3 dados **de padrão com preposição**, 35 dados de **copiadora**, 249 dados de **cortadora** e 267 dados de **padrão com onde**; para o banco de dados Projeto SP2010, obtivemos 10 dados de **padrão com preposição**, 15 dados de **copiadora**, 128 dados de **padrão com onde** e 300 dados de **cortadora**. Assim, produzimos uma análise estatística consistente das estratégias **padrão com onde** e **cortadora**, para os dois bancos de dados, que privilegiou a análise multivariada. Nossa intenção era compreender como cada uma das variáveis atuavam na explicação do fenômeno em variação. Contudo, achamos coerente também apresentar os resultados que não haviam obtido significância estatística, projetando, principalmente, a replicabilidade da pesquisa.

Desse modo, averiguamos, logo de início, que as diferenças na constituição dos bancos de dados mobilizaram contextos sintáticos, semânticos e pragmáticos diferentes. Ao explorar a identidade lexical do verbo da oração relativa, observamos frequências e itens distintos, o que impactou nos demais fatores linguísticos controlados. Assim, para o modelo de regressão, nos dois *corpora*, tínhamos como variável aleatória o falante e a identidade lexical e como valor de aplicação da variável resposta as estratégias de relativização **padrão com onde**. Como variáveis previsoras, tínhamos, para o banco de dados IBORUNA: faixa etária, escolaridade e identidade semântica do locativo; para o banco de dados Projeto SP2010: faixa etária, região do falante, escolaridade e definitude do antecedente.

Resumidamente, concluímos que, para o banco de dados IBORUNA, há um desfavorecimento do uso da **padrão com onde** para falantes da primeira faixa etária em comparação com o *intercept*. Tal diferença entre as estimativas não é significativa, mas essa faixa é responsável pelo maior uso da **cortadora**, enquanto a segunda é responsável pelo maior

uso da **padrão com onde**, o que confirma os dois movimentos de “picos” relacionando as faixas etárias com usos de estratégias diferentes. Outra diferença interessante, e que reforçou nossa análise, foi o desfavorecimento do uso da variante padrão pelos níveis de escolaridade mais baixos, em especial o ensino fundamental. Isso aponta para uma relação do uso da **padrão com onde** com o contato com a norma culta durante os anos escolares. Por fim, observamos o desfavorecimento da estratégia quando o antecedente se refere a um local.

Para o banco de dados do projeto SP2010 há um desfavorecimento do uso da **padrão com onde** para falantes da segunda e terceira faixa etária em comparação com o *intercept*. Tal diferença entre as estimativas não é significativa, estatisticamente ($p > 0.05$), porém essas faixas são responsáveis pelo maior uso da **cortadora**. A primeira faixa etária é responsável pela frequência mais alta da estratégia **padrão com onde** no *corpus*, o que confirma os dois movimentos de “picos” relacionando às faixas etárias com usos de estratégias diferentes. Também observamos um desfavorecimento do uso da variante padrão pelo nível de escolaridade mais baixo, ou seja, o ensino médio. Isso aponta, tal qual para o banco de dados IBORUNA, para uma relação do uso da **padrão com onde** com a norma. Por fim, observamos o desfavorecimento da estratégia para moradores da região centro velho e em contextos indefinidos.

Dessa forma, discutimos que os fatores extralinguísticos, por serem semelhantes nos dois bancos de dados, apresentam um poder explanatório maior no processo de variação das estratégias de relativização locativas em contextos preposicionados na fala paulista. Assim, a idade e escolaridade do falante são perspectivas interessantes para estudos futuros compreenderem o significado social que as variantes assumem na fala dos paulistas.

A análise univariada nos revelou que a adjacência entre o sintagma nominal antecedente e a oração relativa, a classificação semântica da relativa, a classe morfológica do locativo, a função sintática do pronome relativo, o sexo/gênero do falante, a tipologia textual e o roteiro de entrevista não apresentam diferenças estatisticamente significantes entre os níveis de análise. Além disso, em alguns casos, a não significância decorre de um problema de ortogonalidade, de um desbalanceamento dos dados. No caso da adjacência, por exemplo, vemos uma tendência comum aos dois bancos que poderia se confirmar se houvesse um conjunto mais robusto de dados de tipo ‘distante’. Assim, esses fatores não contribuíram, no contexto de nosso estudo, para a compreensão do fenômeno.

Por fim, com relação à **copiadora** e à **padrão com preposição**, consideramos adequado discutir os fatores envolvidos em sua baixa incidência no *corpus*. A partir da literatura, concluímos que o valor negativo associado a essa estratégia não padrão e a falta de naturalidade

da estratégia padrão são os responsáveis pela pouca frequência. Outrossim, identificamos adjacência entre o sintagma nominal antecedente como um contribuinte para o uso dessas variantes, já que contextos antecedentes exigem menos esforço cognitivo do falante. A escolaridade também influencia na escolha de uma ou outra, tendo em vista que falantes mais escolarizados são os responsáveis por produzir a **padrão com preposição**, enquanto os menos escolarizados utilizam a **copiadora**.

Após as explanações teórico-metodológicas e a investigação estatística, concluímos que existe um processo de variação linguística na expressão do valor locativo em contextos preposicionados, por meio das variantes: **padrão com preposição**, **padrão com onde**, **cortadora** e **copiadora**. A partir dos fatores linguísticos e extralinguísticos que condicionam o uso das estratégias, constatamos que afala do interior paulista se utiliza, em maior frequência, do uso prototípico do pronome **onde**; enquanto a fala da capital paulistana apresenta um uso maior da cortadora, que já havia sido identificado em momentos anteriores (TARALLO, 1983).

De todo modo, essa etapa da investigação era necessária considerando o problema do encaixamento da Teoria de Variação e Mudança Linguística. Assim, tendo em vista que o processo de mudança linguística não é um movimento de deslocamento, mas sim um processo em que “um conjunto limitado de variáveis num sistema altera seus valores modais gradualmente de um polo para outro” (WEINREICH, LABOV, HERZOG, 2006 [1968], p. 132), identificamos como o domínio de origem do pronome relativo **onde** se encontra apesar do processo de gramaticalização. Dessa forma, verificamos um processo de variação entre as estratégias relativas.

Como encaminhamentos futuros, nos dedicaremos a uma análise diacrônica para avaliarmos, com base nos estágios da língua, o desenvolvimento da multifuncionalidade de **onde** e o impacto desse desenvolvimento nas estratégias de relativização locativas em contexto preposicionado no Português Brasileiro. Isso nos dará os recursos necessários para afirmações mais categóricas a respeito dos processos que estamos investigando.

REFERÊNCIAS

- ALMEIDA, M. A. de; BERLINCK, R. de A. 2019. “Por onde anda você?” – Sobre a norma e o uso de onde na fala paulista. **Revista Diálogos**, v. 7, n. 1, p. 70-88, 2019.
- ALMEIDA, M. A. de. **Onde anda você?** um estudo sobre os usos de “onde” na fala paulista. Araraquara: Trabalho de conclusão de curso (bacharelado - Letras) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Ciências e Letras (Campus de Araraquara), 2019.
- ALMEIDA, M. A. de. “Onde está você agora?” – Uma investigação sobre os contextos não locativos de onde na fala paulista. **Revista Falange Miúda**, v. 5, n. 2, p. 157-175, 2020.
- AMARAL, A. **O dialeto caipira**. São Paulo: Hucitec-SCET-CEC, 1976 [1920].
- ARAÚJO, L. S. de. **A expressão dos valores 'antepresente' e 'passado absoluto' no espanhol**: um olhar atento a variedades diatópicas da Argentina e da Espanha. Orientador: Rosane de Andrade Berlinck. 2017. F. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, São Paulo, 2017. Disponível em: < <https://repositorio.unesp.br/handle/11449/150649?show=full> > Acesso em: 01 março 2020.
- ARIEL, M. Referring and accessibility. **J. Linguistics**, 24, p. 65-87, 1988.
- ARIEL, M. The function of accessibility in a theory of grammar. **Journal of Pragmatics**, 16, p. 443-463, 1991.
- BARROS, A. L. de. **O uso da relativa cortadora na fala pessoense**. Dissertação de Mestrado orientada pelo professor Dermeval da Hora, UFPB, João Pessoa, 2000.
- BEAMAN, Karen V. Swabian relatives: variation in the use of the wo-relativiser. In: **Advancing Socio-grammatical Variation and Change**. Routledge, 2021. p. 134-164.
- BECHARA, E. **Moderna gramática portuguesa**. 37 ed. Rio de Janeiro: Nova Fronteira, 2009.
- BERLINCK, R.A. **La position du sujet en portugais**. Étude diachronique des variétés brésilienne et européenne. Thèse de Doctorat. Katholieke Universiteit Leuven, 1995.
- BISPO, E. B. **Relativa copiadora**: uso, regularização e ensino. Dissertação de Mestrado orientada por Maria Angélica Furtado da Cunha. Programa de Pós-Graduação em Estudos da Linguagem, UFRN, Natal, 2003.
- BISPO, E. B. **Estratégias de relativização no português brasileiro e implicações para o ensino**: o caso das cortadoras. 164f. Tese (Doutorado em Estudos da Linguagem), UFRN, Natal-RN, 2009.
- BOURDIEU, P. L'économie des échanges linguistiques. **Langue Française**, Paris, n.34, 1977.
- BOURDIEU, P. Price Formation and the Anticipation of Profits. In: BOURDIEU, P. **Language and Symbolic Power**. Cambridge: Polity Press, 1991.

BRAGA, M.; MANFILI, K. Essa é a preocupação onde eu quero chegar: “onde” em referências anafóricas no português do Brasil. **Veredas** 8 (1-2), 2004. p.233-243.

BRAGA, M. L.; ILARI, R.; OLIVEIRA, R. P. de; BASSO, R. M. Os especificadores – parte 1: artigo definido. In: ILARI, R. (org.). **Palavras de classe fechada** Gramática do português culto falado no Brasil, v. 4). São Paulo: Contexto, 2015, p. 107-128.

BRANDÃO, S.M. Mercado linguístico: uma interpretação da imbricada relação estrutura linguística e estrutura social. **Linguagem: Estudos e Pesquisas**. Catalão-GO. V. 21, n.1, 2017. p.225-255.

BRANDÃO, S. M. **Alternância verbal em construções condicionais - um fenômeno variável?**. Araraquara: Dissertação (mestrado) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara, 2018.

BRANDÃO, S. M. **Template - análises linguísticas multivariadas** (variáveis binomiais) no R. [S.l.]: Zenodo. , 30 Sep. 2018. DOI: <https://doi.org/10.5281/zenodo.1439457>

BROOK, Marisa. One of those situations where a relative pronoun becomes a complementizer: A case of grammaticalization in progress... again. In: **Proceedings of the 2011 Canadian Linguistics Association annual conference**. 2011.

BURGOS, L.E.S.de. **Estratégias de uso das relativas em uma comunidade de fala afro-brasileira**. Dissertação de Mestrado, Universidade Federal da Bahia, 2003.

CALDEIRA, T. P. do R. **Cidade de muros: crime, segregação e cidadania em São Paulo**. 3 ed. São Paulo: Editora 34, 2011.

CAMBRAIA, C.N; ARAÚJO, L.E.S. Variação em locativos no português de Belo Horizonte: estudo sociolinguístico. **Paralelo** 20, Belo Horizonte, n. 2, dez. 2004. p.123-132.

CASTILHO, A.T. de. **Nova gramática do português brasileiro**. São Paulo: Contexto, 2014.

CORRÊA, W.R. **Oração relativa: o que se fala e o que se aprende no português do Brasil**. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, 1998.

CUNHA, C.; CINTRA, L. **Nova gramática do português contemporâneo**. 5. ed. Rio de Janeiro: Lexikon, 2008.

FREITAG, Raquel Meister Ko. Problemas teórico-metodológicos para o estudo da variação linguística nos níveis gramaticais mais altos. **Matraga - Revista do Programa de Pós-Graduação em Letras da UERJ**, [S.l.], v. 16, n. 24, dez. 2009. ISSN 2446-6905. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/matraca/article/view/27799/19920>>. Acesso em: 29 jul. 2021.

FREITAG, R. M. K. (Re)discutindo sexo/gênero na sociolinguística. In: FREITAG, R. M. K.; SEVERO, C. G. (org). **Mulheres, linguagem e poder: estudos de gênero na sociolinguística brasileira**. São Paulo: Blucher, 2015, p. 17-73.

FREITAG, R. M. K. **Teste de hipóteses e valores de significância**. Disponível em: <https://rkofreitag.github.io/Pvalor.html/>. Acesso em: 2020-10-07.

FREITAG, R. M. K. **Variáveis categóricas**. Disponível em: <https://rkofreitag.github.io/Categorica.html/>. Atualizado em: 2021-04-11.

FREITAG, R. M. K. **Redução de variáveis e de dados**. Disponível em: <https://rkofreitag.github.io/reducao.html/>. Atualizado em: 2021-04-11.

HEINE, B.; CLAUDI, U.; HÜNNEMEYER, F. **Grammaticalization**: a Conceptual Framework. Chicago: The University of Chicago Press, 1991.

HOPPER, P. J.; TRAUGOTT, E. C. **Grammaticalization**. New York, Cambridge University Press, 1993.

HYNDMAN et al (2021). **forecast**: Forecasting functions for time series and linear models. Disponível em: < <https://pkg.robjhyndman.com/forecast/> >

GILI GAYA, Samuel. **Curso superior de sintaxis española**. 9 ed. Barcelona: Biblograf, 1970.

GONÇALVES, Sebastião Carlos Leite. **Banco de dados Iboruna**: amostras eletrônicas do português falado no interior paulista. Disponível em: <http://www.alip.ibilce.unesp.br>.

KATO, M. A. Orações relativas: variação universal e variação individual no português. **Estudos lingüísticos**, V, 1981, p. 1-16.

KATO, M. Recontando a história das relativas em uma perspectiva paramétrica. In: ROBERTS, I.; KATO, M. **Português brasileiro**: uma viagem diacrônica. Campinas: Editora da Unicamp, 1996.

LABOV, W. **Padrões Sociolinguísticos**. São Paulo: Parábola Editorial, 2008 [1972].

LABOV, W. Where does the linguistic variable stop? A response to Beatriz Lavandera. **Working Papers in Sociolinguistics**, 44. Austin: Southwest Educational Development Laboratory, 1978.

LABOV, W. **Principles of Linguistic Change**. Vol. 1: Internal Factors. Cambridge, MA/Oxford: Blackwell Publishers. 1994

LABOV, W. **Principles of Linguistic Change**. Vol. 2: Social Factors. Cambridge, MA/Oxford: Blackwell Publishers. 2001.

LABOV, W. Some Sociolinguistic Principles. In: PAULSTON, C.B.; TUCKER, G.R. (eds) **Socilinguistics**: The Essential Readings. Oxford: Blackwell Publishing, 2003. p.234-250.

LABOV, W. **Principles of Linguistic Change**. Vol. 3: Cognitive Factors. Cambridge, MA/Oxford: Blackwell Publishers. 2010.

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

LAKOFF, G.; JOHNSON, M. *Metaphor we live by*. Chicago: The University of Chicago Press, 1980.

LAVANDERA, B. Where does the sociolinguistic variable stop?. **Language in Society**, v.7, p. 171-182, 1978.

LEHMANN, C. **Thoughts on grammaticalization**. 2 ed. Erfurt: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfut, 2002.

LEMLE, M. “Heterogeneidade dialectal: uma apelo à pesquisa” In *Linguística e Ensino do Vernáculo*, organizado por Lobato, **Tempo Brasileiro**, 53/53: 60-94, 1978.

LEVSHINA, N. **How to do Linguistics with R**. Data exploration and statistical analysis. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.

LIMA, S.C. **O impacto do vernáculo sobre o uso do onde na escrita monitorada**. 2007. 142 f. Dissertação (Mestrado em Linguística) – Instituto de Letras, Universidade de Brasília, Distrito Federal. 2007.

LIMA, S.; COROA, M.L.M.S. Recategorização, semantização e discursivização na trajetória de gramaticalização do onde. **Via Litterae**, Anápolis, v. 5, n. 2, p. 307-333, jul./dez, 2013. Disponível em: <http://www2.unucseh.ueg.br/vialitterae>

MALLMANN, A. C. L. G.; GARCIA, J. C. V.; ESCOBAR, R. de C. P. S. Variação em estratégias de relativização no Português Brasileiro. VIEIRA, S. R.; LIMA, M. D. A. de O. (orgs.). **Variação, gêneros textuais e ensino de Português: da norma culta à norma-padrão**. Rio de Janeiro: Letras UFRJ, 2019, p. 63-88.

MARINHO, J. H. C. **O uso do onde no texto acadêmico**. Revista Estudos Linguísticos, Belo Horizonte, v.8, n.1, p.159-170, jan./jun. 1999.

MARTELOTTA, M. E. **Mudança linguística: uma abordagem baseada no uso**. São Paulo: Cortez, 2011.

MATEUS, M.H.M et al. **Gramática da Língua Portuguesa**. 6 ed. Lisboa: Editorial Caminho, 2003.

MELO, R. M.; FRANCA, J. C. O item onde e suas rotas de mudança: uma abordagem à luz da teoria da gramaticalização. In: II Congresso Brasileiro sobre Letramentos e Dificuldades de Aprendizagem (CONBRALE), 2018, CAMPINA GRANDE. **Anais II CONBRALE. CAMPINA GRANDE**: Realize, 2018. v. 1. p. 1-10.

MENDES, R.B. **Projeto SP2010: Amostra da fala paulistana**. 2003. Disponível em <<http://projetosp2010.fflch.usp.br>>. Acesso em 13/04/2019.

MILROY, L.; GORDON, M. Beyond Phonology: Analyzing and Interpreting Higher Level Variation. In: MILROY, L.; GORDON, M. **Sociolinguistics: Method and Interpretation**. Oxford: Blackwell Publishing, 2003.

MOLLICA, M. C. **Estudo da cópia nas construções relativas em português**. Rio de Janeiro, 1977. Dissertação (Mestrado) Pontifícia Universidade Católica.

MOLLICA, M. C. **Influência de fatores de processamento na variação em português**. In: TARALLO, F. (org). **Fotografias sociolinguísticas**. Campinas: Pontes, 1989.

MOLLICA, M. Relativas em tempo real no português brasileiro contemporâneo. In: PAIVA, M. C.; DUARTE, M. E. **Mudanças linguísticas em tempo real**. Rio de Janeiro: Contra Capa, 2003, p.129-138.

NEVES, M.H.M.; BRAGA, M.L. As construções hipotáticas/adverbiais. In: NEVES, M.H.M. (org.) **A construção das orações complexas** (Gramática do português culto falado no Brasil, v. 5). São Paulo: Contexto, 2016. p.124-166.

NEVES, M. H. de M. **A gramática do português revelada em textos**. São Paulo: Editora Unesp, 2018.

NICHOLAS, N. **The story of pu**. Melbourne: The University of Melbourne dissertation. 1998

OLIVEIRA e SILVA, G.M.; PAIVA, M.C. Visão de conjunto das variáveis sociais. In: OLIVEIRA e SILVA, G.M.; SCHEREE, M.M.P. **Padrões Sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro**. Rio de Janeiro: Tempo Brasileiro, 1996. p.335-378.

OLIVEIRA, M. R. de. Categorias cognitivas em debate: a trajetória dos pronomes locativos em destaque. In: LIMA-HERNANDEZ, M.C. **Gramaticalização em perspectiva: cognição, textualidade e ensino**. São Paulo: Paulistana, 2010.

OUSHIRO, L. **Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São Paulo**. 2015. Tese (Doutorado em Semiótica e Linguística Geral) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

OUSHIRO, L. (2018) **dmsocio**. v0.2.0. Available at <oushiro.shinyapps.io/dmsocio>

OUSHIRO, L. (2021). Introdução à Estatística para Linguistas (1.0.9). Zenodo. <https://doi.org/10.5281/zenodo.4584104>

PATIL, I. (2018). **ggstatsplot: "ggplot2" Based Plots with Statistical Details**. CRAN. Retrieved from <<https://cran.r-project.org/web/packages/ggstatsplot/index.html>>

PINHEIRO, C. L. **A relativização no português oral culto de Fortaleza**. Dissertação de Mestrado – Universidade Federal do Ceará, 1998.

PRINCE. E. F. Toward a Taxonomy of Given-New Information. In: COLE, P. (ed.). **Radical pragmatics**. New York: Academic Press, 1981, p. 223-233.

RAE. **Esbozo de una nueva gramática de la lengua española**. Madrid: Espasa, 1986.

R Core Team (2021). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>.

ROCHA LIMA, C. H. **Gramática normativa da língua portuguesa**. Rio de Janeiro: José Olympio, 2020 [1972].

ROUSSOU, A. Complements, relatives, and nominal properties. **Handout from talk presented at the University of Ghent**, 2012.

SAID ALI, M. **Gramática histórica da língua portuguesa**. 8. ed. Brasília: Companhia Melhoramentos, 2001.

SANKOFF, G.; KEMP, W., CEDERGREN, H. The syntax of *ce que qu'est-ce que* variation and its social correlates. In: SHUY, R. E; FIRSCHMIG, R.W. (eds) **Dimensions of variability and competence**. Washington: Georgetown University Press, 1980.

SANTOS, Aymmé Silveira; MELO, Raniere Marques de. O ensino da variação linguística na Base Nacional Comum Curricular. **Entrepalavras**, Fortaleza, v. 9, n. 3, p. 115-132, setdez/2019.

SEVERO, C. G. O papel do gênero/sexo nos estudos sociolinguísticos de variação/mudança. **Revista de Letras**, v. 8, p. 1-08, 2006.

SOUZA, E. H. P. M. de. **A multifuncionalidade do onde na fala de Salvador**. 2003. Tese (Doutorado em Letras e Linguística) - Instituto de Letras, Universidade Federal da Bahia, Bahia.

SOUZA, I. G. de; CARVALHO, C. dos S. Usos de Onde na Escrita: Multifuncionalidade, Abstratização Semântica e Gramaticalização. In: LOPES, N. da S.; SANTOS, E. S. dos S.; CARVALHO, C. S. de. **Língua e Sociedade: Diferentes Perspectivas**, Fim Comum. São Paulo: Blucher, 2019, p. 197-214.

TANG, W. **A Simple Guide to Using Antconc**. 2011. Disponível em: <http://www.laurenceanthony.net/software/antconc/resources/help_AntConc321_english.pdf>

TARALLO, F.L. **Relativization strategies in Brazilian Portuguese**. Philadelphia: University of Pennsylvania, 1983. PhD Dissertation.

TARALLO, F.L. The filling of the gap PRO-DROP rules in Brazilian Portuguese. In: KING, L.D.; MALEY, C.A. (eds) **Selected papers from the XIIIth Linguistic Symposium on Romance Languages**. (Current Issues in Linguistic Theory, 36), Amsterdam/Philadelphia: John Benjamins, 1985.

TARALLO, F. L. **A pesquisa sociolinguística**. São Paulo: Editora Ática, 1986.

TORREGO, L. G. **Gramática didáctica del español**. 8 ed. Madrid: SM, 2002.

TRUDGILL, Peter. Sex, covert prestige and linguistic change in the urban British English of Norwich. **Language in society**, v. 1, n. 02, p. 179-195, 1972.

WEINREICH, U.; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da mudança linguística**. Trad. Marcos Bagno. São Paulo: Parábola, 2006 [1968]

WILMET, M. **La détermination nominale**. Quantification et caractérisation. Paris: Presses Universitaires de France, 1986.

ZILLES, A. M. S.; KERSCH, D. F. Onde: prescrição, proscrição, descrição e ensino. In: ZILLES, A. M. S.; FARACO, C. A. (Org.). **Pedagogia da Variação Linguística**: língua, diversidade e ensino. São Paulo: Parábola Editorial, 2015, p. 145-187.