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

**PERCEPÇÕES SOCIOFONÉTICAS DO (-R)
EM SÃO PAULO**

LARISSA GRASIELA MENDES SORIANO

Orientador: Prof. Dr. Ronald Beline Mendes

Versão Revisada

SÃO PAULO

2016

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

**PERCEPÇÕES SOCIOFONÉTICAS DO (-R)
EM SÃO PAULO**

LARISSA GRASIELA MENDES SORIANO

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

Orientador: Prof. Dr. Ronald Beline Mendes

Versão Revisada

SÃO PAULO

2016

Aos meus pais,
que me deram o que podiam,
e acabaram me dando o mundo.

Agradecimentos

Esta dissertação é o produto final de tempos intensos e de muito aprendizado, que compartilhei com pessoas queridas.

Ao meu orientador, Ronald Beline Mendes, quero agradecer por ter me acolhido no GESOL (Grupos de Estudos Sociolinguísticos da USP) a anos atrás e me fazer pensar como uma sociolinguista em formação desde a primeira discussão de texto. Obrigada por me dar inúmeras oportunidades de crescimento desde então, pelas correções de texto sangrentas e por desdenhar de certas inseguranças também.

À Lívia Oushiro, por ser uma inspiração. Obrigada por me ensinar sobre estatística, linguística, a vida acadêmica e, principalmente, sobre generosidade e determinação.

Ao professor Paulo Chagas, por ter me acompanhado desde o começo desta pesquisa, pelos apontamentos durante o exame de qualificação e pelo semestre em que o auxiliei nas aulas de Sociolinguística, durante o estágio PAE.

A todos os meus companheiros do GESOL, que fazem das reuniões de trabalho, encontros. Em especial, Mariane, Marília e Wendel, com quem compartilhei também tardes de risadas e um pouquinho de chororô (porque é preciso).

Aos meus colegas do Departamento de Linguística Júlia Fernandez, Rodrigo Madrid, Juliana Ángel, Aline Benevides, Thiago Chaves, Wellington Silva e o “Pará Doidera” pela alegria instantânea a cada encontro.

Ao Enio, por, a uns 10 anos atrás – enquanto subíamos a Avenida Angélica a pé –, me falar sobre variação linguística, sobre como “certo” e “errado” são noções muito esquisitas, e colocar uma pulguinha atrás da minha orelha que eu nunca mais deixei de coçar. Obrigada por sua amizade e suas provocações deliciosas.

Ao Rodney, pela amizade desde os tempos em que nós escrevíamos sobre outras coisas, pela cumplicidade e pelas tardes de conforto na USP.

Ao Vinícius e à Helen, pelo tempo longo de estrada... Obrigada pela companhia sempre tão calorosa e alegre e pelo tempo em que dividiram não só a casa, mas os dias comigo.

Ao Duda, pelo carinho e por me inspirar com sua empolgação e amor pelo seu trabalho e por cada projeto novo.

Ao meu irmão, Leonardo, por estar ao meu lado mesmo lá da Austrália e pelas tardes mais epicuristas que se pode ter enquanto se desenvolve uma pesquisa acadêmica.

À minha mãe, Geny, e ao meu pai, Lauro, por nem sempre me entenderem, mas sempre me apoiarem e confiarem em minhas decisões. Obrigada por me darem a segurança suficiente para que eu não me entregasse ao conforto e aceitasse novos riscos.

Ao meu par, Renan, pelo amor cotidiano e por andar de mãos dadas comigo. Obrigada por me fazer perguntas sobre estatística (sem nem saber o que) para me ajudar a pensar, pelo café e o chocolate nas noites de escrita e por ser mais gentil comigo do que eu mesma.

A todos os paulistanos que dividiram comigo suas impressões a respeito da língua, durante as entrevistas do ProjetoSP2010, por me provocarem inquietações que por fim desembocariam nesta pesquisa de mestrado.

E, por fim, a todos os meus amigos e colegas que me indicaram muitos informantes durante esses anos, por responderem e divulgarem o experimento de percepção que compõe esta pesquisa. Trabalho de campo é trabalhadeira!

Resumo

A presente pesquisa de mestrado investiga como as características sociais dos ouvintes, bem como o significado social das variantes linguísticas, afetam o modo como são atribuídas diferenças acústicas a pares formados entre cinco variantes de (-r) em coda: (i) vibrante com três batidas, (ii) vibrante com duas batidas, (iii) tepe (vibrante simples), (iv) aproximante alveolar, (v) aproximante retroflexa.

Em entrevistas sociolinguísticas (MENDES; OUSHIRO, 2012), informantes paulistanos comentam acerca de dois “graus de retroflexão” do (-r): um “mais forte” (que corresponderia ao retroflexo), típico de um falar “caipira” e um “mais fraco” (a aproximante alveolar) que seria falado na cidade de São Paulo. Essas duas variantes ainda se oporiam ao tepe, tipicamente paulistano. Tais discursos revelam que os falantes parecem perceber diferenças fonéticas bastante sutis e, ao mesmo tempo, associar significados sociais a elas.

No intuito de avaliar quão salientes são essas sutilezas acústicas, desenvolveu-se um experimento de percepção sociofonética (CAMPBELL-KIBLER, 2006, 2007, 2009; HAY; DRAGER, 2007) no qual diferentes grupos de moradores da cidade de São Paulo deveriam atribuir diferenças a pares formados pelo mesmo item lexical mas com pronúncias de /-r/ diferentes: p.e, a palavra *artéria* primeiro pronunciada com o tepe e depois com a aproximante alveolar.

Os resultados indicam que as respostas dos ouvintes variam de acordo com seu Sexo/Gênero, Região de Nascimento, Local de Residência na Cidade e Variante que Reconhece na Própria Fala. Isto confirma que fatores extralinguísticos de fato influenciam até mesmo a percepção fonética – um

fenômeno que, *a priori*, variaria de forma menos sistemática do que a produção linguística.

Palavras-chave: Percepção sociofonética. /-r/ em coda. Português paulistano. Sociolinguística. Significados sociais.

Abstract

This study investigates how listeners' social characteristics and the social meanings of linguistic variants impact how acoustic differences are assigned to pairs formed between five variants of post-vocalic (-r): (i) trill with three tongue hits, (ii) trill with two tongue hits, (iii) tap, (iv) alveolar approximant and (v) retroflex.

In sociolinguistic interviews (MENDES; OUSHIRO, 2012), *Paulistanos* (people born and raised in the city of São Paulo) comment on two "degrees of retroflexion" of post-vocalic (-r) in Paulistano speech: one that is perceived as "stronger" (which corresponds to the retroflex), typical of the state's countryside, and one that is perceived as "weaker" (the alveolar approximant). These two variants, on their turn, are opposed to the tap, the prototypical Paulistano variant. These evaluations reveal that speakers seem to perceive quite subtle phonetic differences and, at the same time, to associate social meanings to them.

In order to measure how salient these acoustic subtleties are to different groups of speakers, a sociophonetic experiment was developed (CAMPBELL-KIBLER, 2006, 2007, 2009; HAY; DRAGER, 2007) in which residents of São Paulo were asked to assign differences to pairs formed by the same lexical item but with different pronunciations of /-r/.

The results show that the listeners' perceptions vary according to their gender, region of birth, region of residence in the city and the variants they recognize in their own speech. This reveals how extra-linguistic factors influence even the phonetic perception, a phenomenon that, *a priori*, would be more "objective" than linguistic production.

Keywords: Sociophonetic perception. Post-vocalic /-r/. Paulistano Portuguese. Sociolinguistics. Social meanings.

Sumário

INTRODUÇÃO	- 1 -
CAPÍTULO 1	- 8 -
1.1 AS TRÊS ONDAS DA SOCIOLINGUÍSTICA	- 8 -
1.2 ESTUDOS DE PERCEPÇÃO	- 16 -
1.3 SOCIOFONÉTICA	- 21 -
CAPÍTULO 2	- 29 -
2.1 O /-R/ EM CODA NO PORTUGUÊS BRASILEIRO	- 29 -
2.2 A VARIÁVEL (-R) EM SÃO PAULO: PRODUÇÃO	- 34 -
2.3 A VARIÁVEL (-R) EM SÃO PAULO: PERCEPÇÃO	- 38 -
2.4 OUTRAS PERCEPÇÕES	- 45 -
CAPÍTULO 3	- 56 -
3.1 DEFINIÇÕES DO EXPERIMENTO DE PERCEPÇÃO	- 56 -
3.2 PALAVRAS-ALVO	- 58 -
3.3 GRAVAÇÃO DOS ESTÍMULOS	- 60 -
3.4 ANÁLISE ACÚSTICA E MANIPULAÇÃO DIGITAL DOS ESTÍMULOS	- 61 -
3.5 MONTAGEM DOS ESTÍMULOS	- 73 -
3.6 EXPERIMENTO-PILOTO	- 75 -
3.7 MODIFICAÇÕES PARA O EXPERIMENTO PRINCIPAL	- 82 -
CAPÍTULO 4	- 88 -
4.1 VARIÁVEIS E HIPÓTESES	- 88 -
4.2 AS DIFERENÇAS ATRIBUÍDAS PARA CADA PAR	- 96 -
4.3 EM BUSCA DE CORRELAÇÕES	- 100 -
4.4 VARIÁVEIS NÃO SIGNIFICATIVAS	- 102 -
4.5 VARIÁVEIS ESTATISTICAMENTE SIGNIFICATIVAS	- 102 -
4.6 SÍNTESE	- 113 -

Lista de Tabelas

Tabela 2.1 – Pronúncia dos documentadores nos estímulos

Tabela 3.1 – Itens lexicais utilizados como estímulos

Tabela 3.2 – Distribuição dos participantes por características sociais

Tabela 3.3 – Variáveis linguísticas e aleatórias

Tabela 3.4 – Médias dos Graus de Diferença Atribuídos a cada par de variantes

Tabela 4.1 – Distribuição dos participantes de acordo com suas características sociais

Tabela 4.2.1 – Participantes que deram a mesma resposta para as duas perguntas “Qual dessas variantes você fala?”, distribuídos de acordo com a Região de Nascimento e Local de Residência

Tabela 4.2.2 – Participantes que deram respostas diferentes para as duas perguntas “Qual dessas variantes você fala?”, distribuídos de acordo com a Região de Nascimento e Local de Residência

Tabela 4.3 – Variáveis Linguísticas

Tabela 4.4 – Graus de Diferença Atribuídos por Par (médias)

Tabela 4.5 – Valores de significância das variáveis no modelo de efeitos mistos

Tabela 4.6 – Reorganização da variável “Variante que reconhece em sua própria fala” de acordo com a Região de Nascimento e Local de Residência do participante

Tabela 4.7 – Significância das diferenças entre os grupos da variável “Variante que reconhece em sua própria fala”

Tabela 4.8 – Significância das variáveis no modelo de efeitos mistos (APENAS PAULISTANOS)

Tabela 4.9 – Distribuição dos participantes paulistanos de acordo com “variantes que reconhece na própria fala” e Local de Residência

Lista de Figuras

Figura 1.1 – Campo indicial de /t/

Figura 2.1 – Pesos relativos referentes ao uso de retroflexo no cruzamento de Faixa Etária e (a) Região de Residência; (b) Nível de Escolaridade e (c) Classe Social

Figura 2.2 – Questionário de percepção aplicado por Oushiro (2015) a partir de Campbell-Kibler (2006)

Figura 2.3 – Campo indicial de (-r) em São Paulo

Figura 2.4 – Campos Indiciais de (-r) de acordo com Sexo/Gênero, Região de Residência e Origem dos ouvintes

Figura 3.1 – Produção da palavra “cerda” com tepe

Figura 3.2 – Produção da palavra “pardal” com vibrante de três batidas

Figura 3.3 – Produção da palavra “aborto” com vibrante de duas batidas

Figura 3.4 – Produção da palavra “português” com aproximante alveolar

Figura 3.5 – Produção da palavra “português” com retroflexo

Figura 3.6 – Detalhe da aproximante alveolar e do retroflexo nas produções da palavra “português”

Figura 3.7 – Curva de entonação na palavra “urtiga” produzida com a aproximante alveolar

Figura 3.8 – Curva de entonação na palavra “urtiga” produzida com o retroflexo

Figura 3.9 – Curva de entonação na palavra “urtiga” produzida com a aproximante alveolar, após manipulação digital

Figura 3.10 – Curva de entonação na palavra “urtiga” produzida com o retroflexo, após a manipulação digital

Figura 3.11 – Estímulo manipulado de “iogurte” pronunciado com vibrante 3

Figura 3.12 – Estímulo de “iogurte” com a pronúncia vibrante 2

Figura 3.13 – Estímulo manipulado de “iogurte” com a pronúncia tepe

Figura 3.14 – Estímulo manipulado de “iogurte” com a pronúncia aproximante alveolar

Figura 3.15 – Estímulo manipulado de “iogurte” com a pronúncia retroflexa

Figura 3.16 – Formação dos dez pares estímulos, de acordo com a distância na “escala fonética”

Figura 3.17 – Tela de respostas do experimento desenvolvido no programa PsychoPy

Figura 3.18 – Tela do programa PsychoPy em que os ouvintes deveriam responder a respeito de suas características sociais

Figura 3.19 – Tela de respostas da pesquisa na plataforma *online* SurveyGizmo

Figura 3.20 – Página de instruções do experimento

Figura 3.21 – Páginas-teste

Figura 3.22 – Tela da pergunta “Entre esses, qual mais se parece com o modo como você fala?”

Figura 4.1 – Médias dos Graus de Diferença Atribuídos para cada Par, de acordo com a Hierarquia de Distância Fonética

Figura 4.2 – Médias das Diferenças Atribuídas a cada Par

Figura 4.3 – Interação entre Região de Nascimento e Par para Grau de Diferença Atribuído

Figura 4.4 – Graus de Diferença Atribuídos para cada Par, pelos grupos de falantes classificados de acordo com “Variante que reconhecem em sua própria fala” (SOMENTE PAULISTANOS)

Figura 4.5 – Médias dos Graus de Diferença Atribuídos por homens e mulheres para cada Par (AMOSTRA GERAL)

Figura 4.6 – Média da Diferença Atribuída por homens e mulheres paulistanos e não paulistanos

Figura 4.7 – Médias dos Graus de Diferença Atribuídos para cada Par por homens e mulheres (na amostra geral e apenas paulistanos)

Introdução

Grande parte dos estudos sociolinguísticos se concentra na análise de *produção* de variáveis linguísticas (MENDES; OUSHIRO, 2011) e suas correlações com fatores sociais. Já os estudos das *percepções* buscam elucidar o papel do ouvinte na construção do espaço social – organizado também pela língua –, uma vez que tão importante quanto o que se diz é o que se entende a respeito do que foi dito. Nesse sentido, esta pesquisa de mestrado investiga, por meio de um experimento de percepção fonética, como moradores da cidade de São Paulo atribuem diferenças acústicas entre cinco variantes de (-r)¹ em coda, a saber: (i) vibrante com três “batidas” [r-3]², (ii) vibrante com duas “batidas” [r-2], (iii) tepe (vibrante simples) [r], (iv) aproximante alveolar [ɹ] e (v) aproximante retroflexa [ɻ]³.

Seria plausível imaginar que não há diferenças entre as percepções dos ouvintes uma vez que, *a priori*, todos os indivíduos que não possuem distúrbios auditivos têm a mesma capacidade de ouvir os sons. Entretanto, mais do que um processo meramente físico, em que as ondas sonoras vibram nos órgãos auditivos de um indivíduo, ouvir é interpretar os sons e por isso tal processo está intimamente ligado ao nosso sistema cognitivo. Teorias fonológicas baseadas no uso (BYBEE, 2001; PIERREHUMBERT, 2001), por exemplo, encaram o conhecimento linguístico como parte de um conhecimento mais geral, em que as categorias estabelecidas no mundo (entre elas, as unidades linguísticas) são construídas através da nossa experiência social. Dessa maneira, há diferenças

¹ Labov (1969) propõe a notação para variável (-r) em contraposição às notações fonética [r] e fonológica /r/.

² Embora não haja os símbolos [r-3] e [r-2] no alfabeto fonético, nesta dissertação optou-se por usar tal notação com a intenção de distinguir as vibrantes dois e três, uma vez que não há símbolos específicos para essa diferenciação.

³ No capítulo 3 serão apresentadas graficamente as diferenças acústicas entre essas variantes.

sistemáticas entre indivíduos não só no ato de *falar* (como trabalhos tradicionais de variação já demonstraram extensivamente), mas também no de *ouvir* os sons.

A pesquisa sobre as percepções dos ouvintes, portanto, parte da premissa de que o grau de diferença atribuído a pares de variantes difere a depender de seu repertório sociolinguístico. Ou seja, as avaliações que os indivíduos fazem a respeito das variáveis linguísticas relacionam-se com os padrões sociolinguísticos que adquirem (e de cujos processos de variação e mudança participam), em uma contínua e dinâmica construção de um sistema sociosemiótico de referências. Esse repertório é constituído por suas experiências, desde os grupos sociais a que se filiam, até o seu modo de falar e as variantes linguísticas com que têm mais contato.

Na ocasião do Projeto SP2010⁴ (MENDES; OUSHIRO, 2012), foram realizadas entrevistas sociolinguísticas com 60 paulistanos de diferentes grupos sociais que, ao final, eram questionados sobre suas impressões a respeito de certos “modos de falar”. As perguntas focalizavam diferentes variáveis linguísticas: a ditongação do /e/ nasal (“o que você acha do modo de falar: ‘você (es)tá ent[ẽj]d[ẽj]do o que eu (es)tou diz[ẽj]do?’” – que se mostrou como uma das menos salientes para os informantes, sobre a qual muitos não conseguiram tecer qualquer comentário); a concordância nominal (“o que você acha do modo de falar: ‘me vê dois pastel e um chopps’?” – que foi alvo de muitos julgamentos); e as pronúncias tepe e retroflexa do /r/⁵. A depender da variante de /r/ que a pergunta focalizava, grupos diferentes demonstraram reações diferentes: alguns informantes não distinguiram tepe de retroflexo, outros disseram considerar o tepe “normal” e o retroflexo “caipira”, outros ainda disseram perceber diferenças

⁴ Participei da gravação e transcrição das entrevistas realizadas no Projeto como bolsista de Iniciação Científica (FAPESP processo: 2011/22855-2), entre 2011 e 2012. O material completo está disponível no site: <http://projetosp2010.fflch.usp.br/>

⁵ Oushiro (2015) analisa as avaliações dos informantes a respeito de cada uma dessas variáveis, bem como descreve os usos e encaixamentos de todas elas na comunidade.

fonéticas tão sutis que sequer haviam sido incluídas no roteiro de perguntas, como no exemplo (1):

- (1) S1⁶: *às vezes eu vamos dizer assim estou lendo algum trecho da de Evangelho de livros da igreja às vezes eu tenho que ir lá ler então quando a gente vai ler aqui eu eu quero ler de uma maneira que eu consiga ser entendido assim que as pessoas consigam entender então eu procuro ler da/ da maneira mais sem sotaque nenhum mas a gente falando às vezes informalmente aí sai de qualquer jeito né*

D1: *e e... falando sem sotaque que você diz aí era diferente o erre seria diferente?*

S1: *ah sim... é*

D1: *e aí col/ era como?*

S1: *ah o erre assim do do jeito que eu li né... 'a porta [ɾ] (es)tá aberta' [ɾ] agora se está no corre-corre aí "fecha ali que a 'porta [ɹ] (es)tá aberta'" [ɹ] quer dizer já sai não assim daquele jeito igual Sabrina Sato 'aberta' [ɹ] mais... né mas sai mais assim (Rodolfo R., M3MP)*

Neste trecho da entrevista, Rodolfo R. diferencia dois graus de retroflexão, um que ele (paulistano) pronunciaria quando estivesse “na correria” e outro mais “acentuado”, como o da apresentadora Sabrina Sato, natural do interior de São Paulo e conhecida por manter o seu sotaque “original” proeminente. Esses dois “tipos de retroflexo” ainda se oporiam ao tepe, ao mesmo tempo mais formal e mais “ausente de sotaque”, que ele usaria quando estivesse monitorando mais a sua fala, ao proferir leituras na igreja. Ou seja, detalhes fonéticos seriam capazes de revelar diferentes graus de formalidade, bem como de diferenciar dois sotaques.

É provável que as pronúncias do /r/ sejam tão salientes para algumas pessoas porque essa variável é um importante índice de diferenciação dialetal no português brasileiro. Por exemplo, só na região Sul do país, há uma grande variedade de formas: no Rio Grande do Sul, as vibrantes e o tepe são as variantes

⁶ “D1” refere-se ao documentador e “S1” ao informante.

predominantes, em Santa Catarina, há um predomínio do uso da fricativa na capital, e de vibrantes em regiões do interior, no Paraná, a variante retroflexa ganha mais força (MONARETTO, 2000; MONGUILHOTT, 2007). No Nordeste, no Rio de Janeiro e em Brasília, há um predomínio das variantes fricativas (CALLOU, MORAES; LEITE, 2002; OLIVEIRA, 1999; SKEETE, 1997; BUENO, 2013); enquanto no interior de São Paulo, Minas Gerais, Mato Grosso do Sul, Mato Grosso e Goiás, as variantes retroflexas são mais frequentes (GUIOTTI, 2002; LEITE, 2004, 2010; DORNELLES; CERQUEIRA, 2011).

Bieler da Silva (2015) investigou as correlações de uso das variantes fricativa e tepe na cidade sul-mineira de Itanhandu, cuja variante predominante é a retroflexa. Apesar de apresentarem índices de ocorrência bastante baixos (8,4% e 0,8%, respectivamente), as duas variantes “estranhas” à variedade revelam significação social na comunidade e aparecem nas falas dos informantes principalmente nos momentos de leitura, quando eles mais monitoram sua fala (o que é um forte indício de que eles as consideram mais prestigiadas que o retroflexo). O tempo de contato que os informantes tiveram com essas duas variantes (seja por terem morado nas cidades de São Paulo ou Rio de Janeiro, ou por terem parentes residindo nesses lugares) bem como sua identificação com certos aspectos do estilo de vida nessas duas capitais (mais urbano ou até mesmo uma personalidade de mulher “mais forte e independente” (BIELER DA SILVA, 2015: 134)) são fatores importantes para a produção dessas formas. Em outras palavras, os usos das variantes de (-r) apresentam distribuições e correlações diferentes em cada região do país, e também adquirem significações sociais diferentes em cada lugar. Isto é, pronunciar uma fricativa em Florianópolis não significa a mesma coisa que pronunciá-la em Itanhandu, por exemplo.

Na cidade de São Paulo, embora provavelmente todas essas variantes estejam em contato, o tepe e o retroflexo são as mais produtivas na fala das

pessoas nascidas na cidade⁷. O tepe é a pronúncia mais frequente em todas as regiões da cidade, em todas as faixas etárias e graus de escolaridade⁸, enquanto o retroflexo é favorecido por homens menos escolarizados, moradores de regiões mais periféricas, com menor mobilidade geográfica dentro de São Paulo e cujos pais não nasceram na cidade (MENDES, 2010; OUSHIRO, 2015). Assim, além de a fala paulistana estar longe de ser homogênea no que diz respeito à produção de /-r/, a classe social e a região de residência na cidade são fatores fundamentais para a diferenciação linguística, bem como para a indicação de menor prestígio do retroflexo.

A presente pesquisa investiga de que modo os significados sociais associados ao (-r), os padrões de uso das variantes na comunidade, bem como as características sociais do ouvinte (seu sexo/gênero, o lugar em que mora na cidade etc), afetam o modo como são atribuídas diferenças acústicas entre essas cinco variantes ((i)-(v)⁹). Será que pessoas que moram na periferia e aquelas que moram no centro atribuem o mesmo grau de diferença fonética entre o tepe e o retroflexo? Pessoas que nasceram fora de São Paulo (e, portanto, tiveram maior contato com outras variantes de (-r)) atribuem diferenças a esses pares de variantes no mesmo grau que os paulistanos? O fato de o /r/ estar em uma sílaba tônica ou em uma sílaba átona, diante de /t/ ou /d/ etc, afetaria o modo como as

⁷ Dados do Projeto NURC indicam que também eram produzidas outras variantes no falar culto paulistano da década de 1970 (mesmo que em quantidades menos expressivas), principalmente a vibrante alveolar, que representou 3% dos /-r/ produzidos em posição de coda medial e 7% na coda final (CALLOU et al., 1996, 2002). Trabalhos mais recentes, no entanto, não levaram em conta diferenças acústicas entre tepe e vibrantes, por um lado, e aproximante alveolar e retroflexa, por outro.

⁸ Essas amostras foram estratificadas de acordo com os critérios estabelecidos pelo projeto SP2010: Sexo/gênero (F/M); Faixa etária (1: 20-34 anos; 2: 35-59 anos; 3: 60 anos ou mais); Escolaridade (até o Ensino Médio/Ensino Superior); Região de residência na cidade (Centro/Periferia).

⁹ A fim de garantir maior fluência no texto, são conferidas abreviações aos nomes das variantes ao longo desta dissertação. Desta forma, (i) a vibrante com três batidas também será chamada de “vibrante três”, “vibrante 3” ou “3”; (ii) a vibrante com duas batidas, de “vibrante dois”, “vibrante 2” ou “2”, (iii) o tepe pode ser referido como “1”, (iv) a aproximante alveolar será também denominada apenas “aproximante”, ou “A” e (v) a aproximante retroflexa será referida como “retroflexo” ou “R”.

variantes são percebidas? A fim de explorar essas questões, elaborou-se um experimento que permitisse aos participantes responder de forma rápida e objetiva quão diferentes entre si as cinco variantes lhes parecem.

Foram então criados estímulos que contrastam duas variantes, em um mesmo item lexical: por exemplo, a palavra *carta* pronunciada com a vibrante de três batidas, e depois a mesma palavra pronunciada com o retroflexo. Para cada item lexical foram formados 10 pares (todas as combinações possíveis entre as cinco pronúncias), para os quais os participantes foram convidados a atribuir diferenças em uma escala que vai de “Igual” a “Muito diferente” (cujos valores de zero a 100 definem a escala, mas são invisíveis para eles). Desta maneira, buscou-se uma medida que pudesse quantificar até que ponto as diferenças fonéticas entre as variantes são salientes para esses ouvintes.

As análises das respostas dos 109 participantes, desenvolvidas com a plataforma R (R Core Team, 2016), mostram que o sexo/gênero do ouvinte, sua região de nascimento, o local em que reside na cidade e o reconhecimento que faz de certas variantes em sua própria fala influenciam significativamente o grau de diferença atribuído às variantes. As mulheres consideraram o retroflexo mais diferente do tepe do que a aproximante alveolar. Os homens, por outro lado, não fizeram tal distinção; bem como os paulistanos que dizem pronunciá-las. Já os paulistanos que reconhecem tanto aproximantes como vibrantes na própria fala foram os que menos atribuíram diferenças fonéticas entre esses “tipos” de variantes e os que consideraram a vibrante 3 mais diferente do tepe. O modo como o ouvinte classifica sua própria fala, por sua vez, tem forte relação com a região em que ele nasceu e com o local em que reside na cidade, ambos fatores também importantes para a variação nas respostas. Assim, se desvendam quais aspectos do repertório sociolinguístico dos ouvintes influenciam o grau de diferença que eles atribuem às variantes de (-r) em São Paulo.

O primeiro capítulo discute os pressupostos teóricos da presente pesquisa e a localiza dentro do campo de investigação da sociolinguística, enquanto revisa

alguns trabalhos que discutem a percepção linguística. O capítulo 2 trata da variável (-r), desde seus aspectos fonológicos até os padrões de produção de suas variantes na cidade de São Paulo. Relatam-se ainda os valores sociais que adquirem na comunidade, por meio da revisão de trabalhos de percepção e de uma análise qualitativa dos metacomentários dos informantes do SP2010. O capítulo 3 especifica as diferenças fonéticas entre as cinco variantes e descreve a metodologia para a elaboração do experimento. O capítulo 4 apresenta as hipóteses que são investigadas e as análises quantitativas a respeito das respostas dos ouvintes.

Capítulo 1

Para uma sociofonética do /-r/ em coda em São Paulo

As premissas de que a variação linguística carrega consigo significado social e que o comportamento linguístico do falante se relaciona com outros aspectos de seu comportamento social são parte essencial das bases teóricas da sociolinguística e, portanto, dos interesses gerais de todos os trabalhos desta área (CAMPBELL-KIBLER, 2009), a partir das quais se elaboram interesses mais específicos. No entanto, diferentes pesquisas se distinguem de acordo com o ponto de vista a partir do qual olham para essa relação entre língua e sociedade. Este capítulo tem por objetivo contextualizar a presente dissertação de mestrado entre as pesquisas sociolinguísticas, à medida que elucida os pressupostos teóricos dos quais parte.

1.1 As três ondas da Sociolinguística

Em uma discussão epistemológica, Eckert (2008, 2012) divide os trabalhos sociolinguísticos em três “ondas”, que se diferenciam entre si pelos seus interesses principais e por sua metodologia.

A “primeira onda” diz respeito aos trabalhos que se iniciaram com o estudo de Labov em Nova Iorque (2006[1966]), cujo principal enfoque é a descrição de padrões regulares de variação e de mudança linguística. Nos trabalhos que se enquadram nessa “onda”, a estratificação socioeconômica é de grande importância para explicar esses padrões e a variação é vista em correlação com macrocategorias sociais (como faixa etária, sexo/gênero, grau de escolaridade, classe social, entre outros). Nessa perspectiva, a ocorrência de uma ou outra variante se correlacionaria ao pertencimento do falante a essas grandes

categorias e o prestígio das formas linguísticas é fundamental para articular essa correlação (uma vez que certos grupos de falantes tenderiam a favorecer o uso das variantes consideradas de maior prestígio na comunidade – p.e., as mulheres (CHAMBERS, 1995) e aqueles que estejam inseridos no mercado de trabalho (LABOV, 2008 [1972])). Em muitos trabalhos dessa primeira onda, as variáveis sociais são impostas pelo pesquisador (já no momento da estratificação e construção da sua amostra) e usadas como instrumentos para descrever os usos e encaixamentos das variáveis linguísticas. Assim, muitas vezes, as macrocategorias passaram a ser confundidas com as identidades dos falantes (MENDOZA-DENTON, 2002).

Na segunda onda, embora a busca por padrões de mudança continue sendo o principal interesse, os métodos de análise se encaminharam no sentido de uma aproximação etnográfica em relação aos falantes, a fim de investigar os padrões mais localmente. Com essa aproximação, são definidas categorias sociais mais relevantes para cada comunidade (que, por sua vez, relacionam-se com as macrocategorias sociais). Um exemplo de trabalho que se encaixa nessa “onda” é o da própria Eckert (2000), no colégio Belten High, em Detroit. Ela descobriu que os diferentes usos de algumas variantes (o alçamento e a posteriorização de vogais médias e baixas, além da variável sintática “negação”) ajudavam a constituir estilos associados a duas comunidades de práticas distintas que mais se destacavam entre os alunos: os *jocks* (atletas, com valores mais orientados na direção daqueles do colégio, da relação com os professores e com a comunidade escolar) e os *burnouts* (geralmente de classes sociais mais baixas, com orientação mais urbana e menos voltada para as normas da escola). Ela identificou os membros mais prototípicos desses grupos e seus usos linguísticos, e descobriu que os usos por parte dos demais alunos iam no sentido de se aproximar ou se afastar (ideologicamente) desses dois grupos. No caso de Belten High, essas duas

categorias tinham uma influência maior para a variação nos usos linguísticos dos falantes do que as macrocategorias sociais, como a classe social.

Já na chamada terceira onda, a direção de análise se desvia das duas primeiras na medida em que a variação não é mais vista como produto do pertencimento do falante a categorias (sejam elas macro ou mais localmente definidas), mas como práticas através das quais ele se posiciona no mundo social. Nessa vertente, o foco não é a mudança linguística, mas a capacidade de as pessoas comunicarem informações sociais a respeito de sua identidade e sobre sua “tomada de posição”¹⁰ – isto é, a posição discursiva na qual o indivíduo se coloca no momento de uma interação social. Nesta vertente, são de extrema importância os significados sociais das variantes e o conceito de *estilo*, entendido como o *locus* de criação de significados sociais, ou seja, da internalização pelos falantes de distribuições sociais mais amplas da variação (ECKERT; RICKFORD, 2001: 1) e sempre mediado por ideologias e crenças. Segundo Eckert (2012):

A variação constitui um sistema sociossemiótico capaz de expressar toda a gama de preocupações sociais de uma comunidade. E, como estas preocupações mudam continuamente, as variáveis não podem ser marcadores consensuais de significados fixos; pelo contrário, a sua propriedade central deve ser a mutabilidade indicial. Esta mutabilidade é alcançada na prática estilística, enquanto os falantes fazem movimentos sociossemióticos, reinterpretando variáveis e combinando e recombina-as em um processo contínuo de bricolagem [Hebdige 1984]¹¹ (ECKERT, 2012: 94).

¹⁰ Uma das traduções propostas para o termo *stance* (Jaffe, 2009).

¹¹ Tradução própria do trecho original: “Variation constitutes a social semiotic system capable of expressing the full range of a community’s social concerns. And as these concerns continually change, variables cannot be consensual markers of fixed meanings; on the contrary, their central property must be indexical mutability. This mutability is achieved in stylistic practice, as speakers make social-semiotic moves, reinterpreting variables and combining and recombining them in a continual process of bricolage (Hebdige 1984).”

Um exemplo de trabalho considerado típico da terceira onda de estudos sociolinguísticos é o de Podesva (2007), que analisou o emprego do falsete na construção do estilo de fala de Heath, um médico gay norte-americano. Podesva pediu a ele que gravasse suas conversas em três situações de interação social diferentes: no trabalho (com os seus pacientes em um consultório médico), em um churrasco com amigos e numa conversa ao telefone com seu pai. Posteriormente, Podesva mediu o *pitch* da fala de Heath nas gravações, a fim de observar com que frequência o falsete era usado por ele em cada uma das situações e em que momentos o seu uso era favorecido. Chegou ao resultado de que, na conversa com amigos, os falsetes eram mais frequentes, mais longos e com a frequência (F0) mais alta; mas também apareciam expressivamente nas interações no consultório.

Uma vez que a frequência (F0) mais alta é geralmente associada à fala das mulheres, nas duas situações o uso do falsete é correlacionado a características “mais femininas”. O mais interessante (e fundamental para um trabalho de terceira onda) é que soar “mais feminino” tinha significados diferentes nas duas situações. Ao analisar a expressividade de Heath em ambas (ou seja, quando falava com falsete, que emoções ele expressava e com quais conotações), Podesva observou dois significados associados a essa variante suprasegmental. No consultório, contribui para a construção de um estilo mais cuidadoso, de um médico que se importa verdadeiramente com o bem-estar do seu paciente. Já no churrasco com os amigos, o falsete é parte de um estilo “diva”, grandioso e socialmente eloquente. Ou seja, uma mesma variante – socialmente correlacionada a traços de feminilidade – veicula significados diferentes e participa da criação de diferentes estilos.

Esses significados potenciais relacionam-se entre si e formam um *campo indicial* dinâmico e fluido (ECKERT, 2008, 2012). Neste campo, as variantes linguísticas (por exemplo, o falsete) não se relacionariam diretamente com as

categorias sociais (amplas), mas com características mais específicas (“cuidadoso”, “educado”, “diva”) que só fazem sentido em contextos específicos. Já essas características se relacionam com as categorias sociais de maneira indireta. Ou seja, uma mesma variante linguística liga-se a um conjunto de significados gerais, que podem se tornar mais específicos a depender do contexto social e do estilo de cada falante (“jeito de falar mais feminino” significando “cuidadoso” ou “diva”).

Assim, os significados das variantes não são vistos como fixos, mas mutáveis, relacionados ideologicamente entre si neste campo indicial. Cada um desses significados pode ser ativado situacionalmente, no momento do uso. Este campo, por sua vez, também é fluido, já que em cada interação há a possibilidade de reforçar certas conexões de significados, que são criadas ideologicamente. Nas palavras de Eckert (2008: 454), “a variação constitui um sistema indicial que incorpora a ideologia na linguagem e que é, também, parte essencial da construção da ideologia”¹².

Eckert (2008) propõe um campo indicial para a pronúncia hiperarticulada de /t/, em palavras como *but*. Estudos diferentes mostraram a associação dessa variante tanto a figuras sociais, como jovens judeus ortodoxos, meninas *nerds* e homens gays (respectivamente, BENOR 2001, BUCHOLTZ 2001 e PODESVA et al. 2002, apud ECKERT 2008) quanto a significados sociais como “cuidadoso”, “enfático” e “irritado” (Figura 1.1).

¹² Original: “Thus variation constitutes an indexical system that embeds ideology in language and that is in turn part and parcel of the construction of ideology.”

Figura 1.1 – Campo indicial de /t/. Fonte: Eckert, 2008: 469

Esse campo proposto por ela mostra como significados tão diferentes se articulam ideologicamente, uma vez que estão dispostos de modo que se relacionam entre si pela proximidade dos conceitos. Em um lado do campo, os adjetivos “elegante”, “formal”, “claro” e “articulado” estão próximos uns dos outros, bem como das *personas* sociais “professor”, “menina *nerd*” e “britânico”. Em outro lado, “fresco”, “enfático” e “exasperado” se aproximam mais da *persona* “diva gay”. O campo mostra que todos os significados sociais de /t/ estão disponíveis no momento em que essa variante é pronunciada, mas que há certa probabilidade de alguns deles serem ativados juntos. Não é impossível que em alguma situação de comunicação, o uso dessa variante seja interpretado como “formal” e “irritado” ao mesmo tempo, mas a distância desses significados no campo mostra que tal associação parece menos provável.

Tal como a mutabilidade dos significados das variantes, a *agentividade* do indivíduo no construir (e no expressar) de sua identidade é um ponto de diferenciação importante entre as “ondas”. Uma das principais críticas aos trabalhos de primeira onda diz respeito à imposição das macrocategorias sociais

pelo pesquisador que, ao mesmo tempo, se baseia nelas para explicar as diferenças de produção, sem levar em conta situações de fala ou mesmo outras associações ideológicas. Essa abordagem pode levar a uma interpretação essencialista do processo de variação, dos usos, e até mesmo das identidades dos falantes (MENDOZA-DENTON, 2002: 476). Já na terceira, não há um esforço em explicar a variação por meio de macrocategorias sociais estáticas, e muitas pesquisas se concentram em um único indivíduo ou em grupos reduzidos de falantes, no intuito de oferecer uma explicação detalhada sobre a dinamicidade dos significados das variantes nas diferentes situações e sobre como os falantes os ressignificam a cada interação. Assim, enquanto na primeira onda não se coloca no centro a identidade do falante, mas sim *identidades sociais*, na terceira, a análise é mais localizada no papel do indivíduo e na construção de estilos e *personas*. Por esse ponto de vista, a identidade social do falante é encarada como uma consequência inferencial de seus atos e tomadas de posição, codificados linguisticamente, e não como algo preestabelecido na forma de variáveis independentes (OCHS, 1993: 296).

Parece necessário, no entanto, relativizar essa agentividade e “poder de escolha” do falante em uma situação. Em primeiro lugar, parece bastante improvável que uma pessoa consiga controlar sua fala o tempo todo a fim de criar certa identidade (OUSHIRO, 2015: 25), e, mesmo que conseguisse, também parece improvável que um falante queira, intencionalmente, expressar valores como “fresco” ou “exasperado”. Mas mais do que isso, os construtos sociais que pré-existem ao indivíduo restringem suas escolhas linguísticas, de modo que o falante fica limitado ao que faz parte de seu repertório social e cognitivo (OCHS, 1993). Nesse sentido, é possível encarar as macrocategorias sociais como o “lugar” de onde cada indivíduo parte para suas interações e interpretações do mundo; por isso, o uso de tais categorias na análise permite generalizações importantes a respeito dos construtos sociais a que o indivíduo tem acesso. Ou

seja, nenhum falante é exposto a todas as variantes de variáveis, e mesmo no caso das variantes que ele conhece, é possível que não tenha acesso a todos os significados sociais “disponíveis”.

A *saliência* das formas linguísticas, aqui entendida como “a relativa capacidade de uma variante linguística evocar significado social” (LEVON; FOX, 2014: 193) parece fundamental para explicar essas relações entre os falantes e as variáveis. Essa noção de *saliência* se relaciona diretamente com os conceitos de indicadores, marcadores e estereótipos de Labov (2008[1972]), que se distinguem por graus de consciência social e de diferenciação estilística na concepção laboviana (LABOV, 2006[1966]), ou seja, os graus de atenção que um indivíduo presta à própria fala. Os indicadores são variáveis cujas variantes, mesmo quando distribuídas diferentemente entre grupos de uma comunidade, têm sua significação social “abaixo do nível de consciência dos falantes” (LABOV, 2001: 196). Para essas variáveis, é difícil encontrar padrões de variação, inclusive na fala de um mesmo indivíduo (uma vez que os falantes tendem, mesmo que inconscientemente, a empregar as variantes de maior prestígio quando monitoram sua fala). Os marcadores, além de revelarem variação nos usos de acordo com uma estratificação social, são reconhecidos pelos membros da comunidade, de modo que os usos variam sistematicamente também na fala de um mesmo indivíduo. Já os estereótipos são variáveis que atingiram um grau de consciência evidente nos falantes, e que por isso são sujeitas a metacomentários por parte deles, e suas variantes são passíveis de estigmatização. Esse é o caso da variável (-r) na cidade de São Paulo.

Dialogando com o que Preston (2010, 2011) chamou de “atenção (*regard*) linguística”, Levon e Fox (2014) propõem que a *saliência* de uma variável não é fixa, no sentido que não há um “consenso social” a respeito da *saliência* de uma forma linguística. Isso porque, para um indivíduo, o “grau de *saliência*” dependeria de fatores sociais e contextuais, além de cognitivos. Preston (2010,

2011) apresenta um modelo processual da percepção dos falantes que mostra como eles passariam de “experimental passivamente” uma variante linguística até produzirem uma reação a ela. O primeiro passo seria “notar” a variante (p.e, a hiperarticulação do /t/), que em seguida seria “classificada” pelo falante de acordo com critérios sociais, linguísticos e contextuais (p.e, “modo de falar de um professor”, ou “modo de falar mais formal”). Posteriormente, essa forma seria dotada de informações avaliativas (p.e, “formal”, “claro”, “cuidadoso”), que, finalmente, dão origem à reação do indivíduo (consciente ou não) àquela variante. Seria, portanto, um processo bastante dinâmico, no qual a saliência tem papel fundamental principalmente nas primeiras etapas. O que Levon e Fox (2014: 193) indicam, portanto, é que assim como esses fatores contextuais (sociais e individuais) afetariam o modo como um indivíduo “nota” uma forma linguística, eles por fim determinariam o modo como essa mesma forma é interpretada e classificada socialmente.

Dessa perspectiva, os significados sociais estão sempre em construção na cooperação entre falantes e ouvintes, nos momentos de interação (OCHS, 1993: 291). De forma que tão importante quanto as tomadas de posição por parte do falante são as interpretações do ouvinte, uma vez que o estilo só vai ser reconhecido como “legítimo” – tal como o falante pretende que o seja – se tais interpretações também estiverem disponíveis em seu repertório sociosemiótico (CAMPBELL-KIBLER, 2006).

1.2 Estudos de percepção

Uma vez que também é calcado nas experiências cognitivas, sociais e interativas do indivíduo (PRESTON, 2010, 2011; LEVON; FOX, 2014; ROSENBLUM, 2008; BYBEE, 2001), o processo de *ouvir os sons* (assim como o de falar) também estaria sujeito a variação de acordo com influências do ambiente,

tais como a situação de fala, as expectativas do ouvinte a respeito do que vai ouvir e até mesmo o interlocutor (BELL, 1984).

Desde o nascimento, somos inundados com inúmeros estímulos sonoros, de diferentes naturezas. Para que possamos nos orientar em meio ao caos, estruturamos esses sinais concentrando-nos naqueles que têm mais informações (JEKOSCH, 2005: 55). Uma vez que fonemas são abstrações linguísticas, e não entidades físicas (CLEARY; PISONI, 2001: 508), a interpretação deles inclui nossa capacidade de ouvir assim como nossas habilidades pragmáticas e semânticas (CLEARY; PISONI, 2001: 500). A percepção da fala acontece, portanto, “de acordo com a bagagem do que já foi experienciado, aprendido e reconhecido. Assim, é claramente dependente de representações [Kohonen 1989]” (JEKOSCH, 2005: 55)¹³.

Não por acaso, a investigação teórica a respeito da percepção da fala hoje se concentra em um esforço multidisciplinar, que inclui não só áreas da linguística (como a psicolinguística e a sociolinguística) mas outras áreas do conhecimento, como a psicofísica, a neurofisiologia, e até mesmo a inteligência artificial (MASSARO, 2001: 14870). São inúmeros os estudos que se dedicam a entender como acontece a percepção, assim como inúmeras são as teorias a respeito dos mecanismos de processamento dos sons.

Alguns desses trabalhos, como o dos psicólogos McGurk e MacDonald (1976 apud CLEARY; PISONI, 2001; ROSENBLUM, 2008), concentram-se em demonstrar como mesmo o processo “básico” de identificação de sílabas não se baseia exclusivamente no “reconhecimento” de frequências (F1, F2...), e são afetados por elementos que, *a priori*, não se relacionariam diretamente à audição.

¹³ Tradução de: Perception occurs therefore against the background of what has already been experienced, learnt and recognized. Thus perception is clearly dependent on representations [Kohonen 1989]

Em seu experimento, eles apresentaram aos participantes áudios gravados da sílaba “ba” e, ao mesmo tempo, um vídeo de uma pessoa articulando a sílaba “ga”. Como resultado, os participantes marcaram que ouviram a sílaba “da”, que seria resultado de uma “fusão” dos dois sons, com um ponto de articulação intermediário que não foi apresentado. Essa “interferência” do estímulo visual na audição ficou conhecida como Efeito McGurk e embora seja resultado de uma ilusão criada em uma situação experimental, baseada na coocorrência não natural de *inputs* distintos (CLEARY; PISONI, 2001), foi uma das primeiras evidências científicas de que os estímulos visuais também afetam a percepção dos sons.

Já com enfoque mais “sociolinguístico”, alguns trabalhos se debruçam sobre as conexões entre a percepção da fala e a categorização de dialetos, como os que se baseiam em tarefas de desenho de mapas (“map-drawing tasks”). Este tipo de experimento consiste em apresentar um mapa de um país (ou região) para os participantes e pedir que eles apontem para os lugares em que as pessoas “falam de um jeito diferente”. Preston (1986) realizou um teste desse tipo em que pediu que adultos de diferentes regiões dos Estados Unidos (Indiana, Havaí, Nova Iorque e Michigan) indicassem em um mapa do país as regiões em que, na “opinião” deles, há “sotaques diferentes”. De maneira geral, os limites dialetais indicados pelos participantes não corresponderam àqueles obtidos em pesquisas anteriores. A maioria indicou boa parte do país como um falar típico do “sul” e Nova Iorque como uma região de sotaque único e próprio. Além disso, os participantes tenderam a marcar os lugares mais próximos de onde viviam com uma maior variedade dialetal, enquanto nos lugares mais distantes foram marcados como mais homogêneos. Esse resultado é evidência de que as pessoas são mais sensíveis à variação que experienciam, a que são expostas. Há outros tipos de testes que vão no mesmo sentido, e que podem responder como os ouvintes usam seu conhecimento a respeito da variação para determinar de onde um certo falante é. Com o uso de ferramentas de edição, por exemplo, é possível

evidenciar uma ou outra variável linguística, e assim ter acesso a quais características da fala são mais relevantes para identificar determinado dialeto (CLOPPER; PISONI, 2008).

Evidentemente, é comum que façamos inferências a respeito dos falantes que ouvimos. Ao ouvir uma voz, sem nenhum esforço consciente imaginamos se é de um homem ou uma mulher, se a pessoa é jovem ou velha, “mano” ou “playboy” ou se é proveniente de determinada região (OUSHIRO, 2015: 265). O trabalho de Pear (1931, apud CAMPBELL-KIBLER, 2006; OUSHIRO, 2015) foi um dos primeiros a tratar da sistematicidade das percepções linguísticas. Ele pediu que ouvintes de uma rádio fornecessem perfis sobre a personalidade de diversas pessoas que eles ouviam no canal e, posteriormente, comparou com os perfis feitos por indivíduos que conheciam pessoalmente os donos das vozes. De modo geral, os ouvintes forneceram perfis bastante diferentes daqueles oferecidos pelos conhecidos, mas suas impressões foram bastante consistentes entre si, o que mostra o poder dos estereótipos na organização social (OUSHIRO, 2015: 267).

Parte dos trabalhos que se enquadram na terceira onda da sociolinguística buscam verificar as percepções dos ouvintes a respeito de variantes de variáveis linguísticas e em quais “pistas” fonéticas eles se baseiam para fazer inferências a respeito do dialeto ou de características pessoais do falante (CAMPBELL-KIBLER, 2006, 2009; PODESVA, 2007; DRAGER, 2010; HAY; DRAGER; THOMAS, 2013; LEVON, 2007; MENDES, 2014; OUSHIRO, 2015, entre outros).

Campbell-Kibler (2006, 2007, 2009) realizou um dos trabalhos expoentes sobre percepções sociolinguísticas, no qual investiga os significados sociais das variantes de (ING) no inglês norte-americano. Em seu estudo, ela verificou como ouvintes percebiam “sotaques” em relação a duas pronúncias de (ING): [ɪŋ] (velar, tradicionalmente tomada como padrão no inglês norte-americano) e [ɪn] (alveolar, tradicionalmente associada à região dialetal sul dos EUA), utilizando a técnica *matched-guise* (LAMBERT et al., 1960, apud CAMPBELL-KIBLER, 2006).

Desenvolvida para estudos da psicologia social, esta técnica parte da ideia de que, em pesquisas nas quais se pede aos participantes que comentem acerca de formas linguísticas, clara e abertamente mencionadas, as informações fornecidas por eles não correspondem necessariamente às suas opiniões pessoais a respeito das variantes de variáveis. Com a técnica *matched-guise*, diferentemente, acessam-se reações inconscientes dos participantes, através de estímulos pareados, produzidos por um mesmo falante: idealmente iguais em tudo (mesma sentença, mesma entonação etc) a não ser pela variável em foco. No caso de Campbell-Kibler 2006, por exemplo, um estímulo do par deveria conter apenas ocorrências de [ɪŋ] ao passo que o outro conteria apenas [ɪn]. Cada participante ouve apenas um dos estímulos e então responde sobre as impressões que teve a respeito do falante – p.e., sobre sua inteligência, confiabilidade e simpatia, masculinidade/feminilidade, classe socioeconômica ou o que quer que o pesquisador queira enfocar. Na análise das respostas, comparam-se todas as aquelas que foram dadas com base em um e outro estímulo, e como ambos só se diferenciam em relação às variantes, segue que as diferenças nas respostas dizem respeito aos significados sociais relacionados a cada uma delas.

Em seus resultados, a autora verificou que, de maneira geral, os ouvintes julgaram os falantes como mais educados, articulados e inteligentes diante do estímulo com a variante velar. Já diante da forma alveolar, os falantes foram percebidos como casuais, descuidados e preguiçosos. Além disso, seus resultados mostram que as percepções a respeito de cada variante dependem também do contexto de uso e, quando elas são pronunciadas por pessoas de diferentes regiões, há diferentes expectativas (a respeito do uso) que, por sua vez, levam a uma mudança no significado social. Por exemplo, quando os falantes não eram identificados como pertencentes a uma região específica, não houve diferença para as respostas a respeito da “inteligência” mesmo quando o estímulo apresentado era com [ɪn]. No entanto, quando os falantes eram

identificados como da região sul (graças a outras variáveis, como a monotongação de /aj/ como em *time*), não eram vistos como mais inteligentes diante da variante velar [ɪŋ], mas sim como “pessoas tentando soar mais inteligentes”.

1.3 Sociofonética

Trabalhos mais recentes (FOULKES; SCOBIE; WATT, 2010; HAY; DRAGER, 2007, DRAGER, 2010; BARANOWSKI, 2013; THOMAS, 2002; MENDES, 2014; OUSHIRO, 2015, entre outros), que indicam que mesmo diferenças fonéticas mais sutis podem carregar informação social de forma sistemática, dialogam diretamente com Campbell-Kibler (2006, 2007, 2009). Hay e Drager (2007) argumentam que, assim como as análises etnográficas são capazes de estabelecer as categorias sociais mais relevantes para determinados contextos, análises acústicas detalhadas (e não somente de oitiva) podem identificar sutilezas fonéticas que são relevantes para a comunicação e para a variação linguística e sua significação social (HAY; DRAGER, 2007:90).

Estudos que procedem a uma análise acústica sistemática em busca de elucidar processos de variação têm sido enquadrados em um campo de estudos denominado sociofonética. O termo foi mencionado pela primeira vez por Deshaies-Lafontaine (1974, apud FOULKES; SCOBIE; WATT, 2010), em uma tese a respeito da variação do (-r) no francês canadense. No entanto, a configuração como área de estudo é relativamente nova, de modo que ainda não há uma definição e delimitação clara da disciplina (FOULKES; SCOBIE; WATT, 2010: 703). De maneira geral, a sociofonética se vale de princípios e técnicas da sociolinguística e da fonética, para analisar como a variação no sistema fonológico é aprendida, armazenada, avaliada subjetivamente e processada ao

ouvir e falar (FOULKES; SCOBIE; WATT, 2010) e, no “caminho contrário”, como a bagagem social do ouvinte influencia a forma como ele ouve e classifica os sons.

Trabalhos como o de Ladefoged e Broadbent (1957), por exemplo, indicam que as variantes observadas na fala de um indivíduo refletem a forma como ele percebe os sons (p.e., a sua região de origem pode prever não só como o falante produz os sons, mas também como ele os percebe). Em seu experimento, eles pediram que os participantes ouvissem uma sentença e, em seguida, marcassem em um formulário se a palavra que ouviram era *bit*, *bet*, *bat* ou *but*. Eles tinham três subgrupos de ouvintes: um cujo dialeto correspondia ao “sistema vocálico básico escocês”; um de escoceses que tiveram seu sistema vocálico modificado graças a influência do inglês da Inglaterra; e um último, composto de ingleses. Seus resultados mostraram que os subgrupos tenderam a classificar as vogais de acordo com o seu próprio sistema vocálico, principalmente os escoceses, que assinalaram majoritariamente *bat*. Segundo os autores, provavelmente isso ocorreu porque em sua variedade, a posição relativa da vogal em *bat* é similar à da vogal sintetizada apresentada no teste. No entanto, outros trabalhos evidenciam que essa relação entre produção e percepção pode não ser direta, como Hay et al. (2006, apud DRAGER, 2010), que descobriram que ouvintes eram bastante precisos em identificar alguns sons, ainda que não os distinguíssem em sua produção.

Já Williams et al. (1999, apud DRAGER, 2010) pediram para que ouvintes classificassem os lugares de origem dos falantes a partir de excertos de sua fala e constataram que os ouvintes tendiam a classificá-los como provenientes da mesma região que eles quando também os julgavam como mais “simpáticos”. Esse resultado levanta a questão sobre quais julgamentos sociais afetam outros: até que ponto o fato de uma pessoa avaliar que a outra “soa simpática” faz com que também a perceba como proveniente de sua mesma região? Ou ainda: de que

modo a avaliação de que um falante soa como proveniente da mesma região de origem do ouvinte leva-o a perceber tal falante também como “simpático”?

Outros resultados, como os de Niedzielski (1999) e Levon (2014) demonstram ainda que os estereótipos, além de ter um papel na organização social, influenciam a percepção; e ter certas informações sociais podem determinar o modo como as variáveis fonéticas são classificadas pelos ouvintes. Em Niedzielski (1999), moradores da cidade de Detroit, nos EUA, ouviram algumas sentenças lidas por uma informante (também de Detroit) e foram instruídos a selecionar uma de seis vogais sintetizadas que correspondesse a uma vogal que ouviram na sentença. A pesquisadora preparou dois formulários diferentes: metade dos participantes recebeu um em que havia “Detroit” escrito no cabeçalho, enquanto a outra metade recebeu um em que estava escrito “Windsor, Ontario”, cidade canadense vizinha. Em seus resultados, ela constata que aqueles que receberam o formulário que indicava que a informante era do Canadá tenderam a selecionar vogais mais alçadas, refletindo o estereótipo de que os canadenses tendem a pronunciar *about*, por exemplo, com a variante [əw] e não [aw]. Já o grupo “Detroit” tendeu a selecionar a vogal “mais neutra”, no sentido de menos marcada socialmente. Ela concluiu, portanto, que os moradores de Detroit tendem a se perceber como falando um tipo de inglês “padrão”, enquanto percebem os canadenses como “falantes que têm sotaque”.

Hay e Drager (2010) também descrevem como a expectativa criada a partir de certos estereótipos influencia a percepção. Em um experimento inspirado em Niedzielski (1999), elas pediram que os participantes classificassem a vogal /ɪ/ em palavras como *near*, e os dividiram em dois grupos. Para o primeiro grupo, as pesquisadoras propositalmente deixaram à mostra, quando abriram um armário para supostamente pegar os formulários, um coala de pelúcia (animal típico da Austrália) enquanto, para o outro grupo, deixaram à mostra um kiwi de pelúcia (típico da Nova Zelândia), sem dar qualquer pista de que os objetos se

relacionavam com o estudo. O grupo que foi exposto ao coala tendeu a classificar a vogal mais próxima ao “padrão” australiano.

Levon (2014), por sua vez, investigou a relação entre certos estereótipos de gênero e sexualidade e a atribuição de características como “masculino” ou “gay” a partir de excertos de fala de um homem britânico, além do quanto um significado social (p.e., “pertencente a classes baixas”) pode bloquear o surgimento de outro significado (p.e., “gay”). Para isso, ele avaliou, através da técnica *matched-guise*, as reações dos ouvintes a respeito de três variantes linguísticas (isoladas e combinadas entre si): elevação do *pitch*, que geralmente é associada a características de gênero; sibilância de /s/, que trabalhos de percepção mostraram se relacionar à identificação de um homem como gay; e realizações frontais de TH (p.e., realizações lábio-dentais [f] da fricativa interdental /θ/) que são consideradas típicas de classes sociais mais baixas no Reino Unido. Assim como nos experimentos realizados por Campbell-Kilber (2006, 2007, 2009), os participantes ouviram os estímulos e responderam um questionário a respeito das características que eles atribuíam ao falante, como sua inteligência, educação, honestidade, além do seu gênero (masculino/não masculino) e sexualidade (gay/não gay).

Para medir a relação dos estereótipos com as percepções linguísticas, ele pediu que os participantes ainda respondessem a um questionário (MRAS – Male Role Attitudes Survey, “pesquisa de atitudes a respeito de papéis masculinos”, PLECK; SONENSTEIN; KU 1993, 1994, apud LEVON, 2014) a respeito do quanto concordavam com certas afirmações a respeito da masculinidade e dos “papéis” que os homens teriam na sociedade. As perguntas evocam, por exemplo, crenças de que os homens ocupam uma posição social dominante (“é importante para um homem que ele seja respeitado pelos outros”; “o homem é responsável por ganhar um bom salário e prover sua família”), crenças a respeito da “dureza masculina” (“um homem não deveria falar sobre os seus problemas”) e sobre a

separação entre papéis masculinos e femininos (“eu não acho que homens devam realizar tarefas domésticas”). Posteriormente, ele comparou o grau de identificação do ouvinte com essas afirmações e suas respostas a respeito da “masculinidade” e “homossexualidade” dos falantes.

Em seus resultados, ele detectou que embora a sibilância tenha sido fortemente associada a “homossexualidade”, quando essa variante era combinada com a frontalidade do TH (“típica” das classes mais baixas) houve uma grande diminuição nas avaliações de homossexualidade e um aumento para masculinidade. Ele interpreta os resultados a partir do fato de os estereótipos de que homens da classe trabalhadora são “masculinos e heterossexuais”. Além disso, ele concluiu que os ouvintes que mais se identificavam com aqueles estereótipos masculinos do MRAS também associaram um *pitch* mais alto e a sibilância como “mais gay” e “menos masculino”, enquanto os que concordam menos com a conceitualização normativa de gênero não demonstraram tal associação.

Por fim, ele define os estereótipos como estruturas cognitivas que ligam conceitos a respeito de grupos (p.e., homens) com atributos (p.e., atlético, dominante) e papéis sociais (p.e., pai). Eles seriam, portanto, uma forma associativa de conhecimento social, que codifica ideologias e liga, ao mesmo tempo um grupo social (que, por sua vez, também é uma ideia construída por meio de conceitos) com atributos e papéis relevantes. Nesse sentido, o processo de ativação dos estereótipos seria bi-direcional: tanto um grupo “homem” pode ativar o traço “atlético”, quanto vice-versa. Essas associações podem ainda ter forças diferentes, de tal forma que certas formas linguísticas podem ativar certas atitudes e papéis mais fortemente do que outras (LEVON, 2014: 544).

Já Labov et al. (2006, 2011) procuram desvendar como essas informações sociolinguísticas seriam “localizadas, armazenadas e processadas” (LABOV et al. 2011: 435) por meio do que eles chamaram de “monitor sociolinguístico”. Esse

monitor seria um componente do processamento linguístico dos indivíduos que controlaria os efeitos das frequências das variáveis sociolinguísticas (não apenas “no geral”, mas em um certo trecho de fala) no julgamento a respeito de comportamentos sociais.

Os autores partem da proposição de que a variação na fala não é categórica (ou seja, certos grupos somente produziram uma variante de uma variável, enquanto outros grupos nunca a produziram), e que, por isso, os julgamentos sociais não dependeriam apenas da presença ou ausência de certa forma linguística, mas seriam sensíveis à frequência de determinada forma em um trecho de fala. Sendo assim, a percepção linguística estaria diretamente ligada à saliência social da variável, de forma que quanto mais frequente, mais saliente a forma seria. Esse monitor, portanto, seria a ligação necessária para explicar como acontece a “relação cognitiva” entre as formas linguísticas e os julgamentos sociais.

Eles buscaram averiguar principalmente três questões a respeito do funcionamento desse monitor: sua “janela temporal”, ou seja, o período de tempo em que ele opera; sua sensibilidade, isto é, a frequência necessária para que ele seja ativado; e os padrões de respostas gerados a partir dele (LABOV et al. 2006: 4). Para testar tais mecanismos de funcionamento, eles elaboraram um experimento em que apresentaram para os ouvintes sete excertos de fala, de cerca de 45 segundos, que uma suposta estudante de jornalismo submeteria a um jornal televisivo, na tentativa de conseguir um estágio. Eles manipularam os excertos de forma que tivessem taxas diferentes de uso da variante alveolar [ɪn] (que estudos anteriores, p.e., CAMPBELL-KIBLER, 2006, 2007, 2009, mostraram ser mais estigmatizada no contexto norte-americano). Assim, um excerto tinha 0% de [ɪn], ou seja, era inteiramente com a variante velar [ɪŋ], outro 10% alveolar, 20%, 30%, 50%, 70% e 100%. A partir desses estímulos, os participantes deveriam julgar quão profissional a falante soava em cada um deles.

Seus resultados mostram que a janela temporal do monitor é grande, uma vez que acomodou as informações sociolinguísticas no tempo dos excertos. Além disso, os ouvintes de fato se mostraram sensíveis à pronúncia de (ING), de modo que eles julgaram como “menos profissionais” os trechos que continham mais ocorrências da forma alveolar. Em relação à frequência, eles descobriram que a percepção em relação ao profissionalismo não foi perfeitamente proporcional à ocorrência da variante, já que depois de certa frequência (por volta de 30%, no seu experimento), não houve uma alteração consistente na percepção dos ouvintes. Segundo os autores, isso demonstraria que o monitor é extremamente sensível (uma vez que os ouvintes já foram capazes de julgar a falante como menos profissional a partir da taxa de 10% de uso da variante alveolar), e, ao mesmo tempo, mais atento a ocorrências menos frequentes.

Combinados, todos esses estudos mostram como a percepção da fala está longe de ser um processo trivial e, por ser calcado em nossas experiências sociais, sofre influência de inúmeros fatores (inclusive extralinguísticos). De maneira geral, o estudo das percepções nos ajuda a entender como as negociações são feitas no espaço social, uma vez que permite identificar os significados disponíveis para os falantes, quais desses conceitos se associam a outros, de que maneira eles são considerados “aparentados” e como podem coocorrer (CAMPBELL-KIBLER, 2006). Conforme ressalta Drager (2010)

Estudos de produção e percepção se complementam, auxiliando a entender como a mudança linguística acontece, como os estereótipos são formados, como a variação linguística é armazenada na mente e que tipos de processos cognitivos subjazem o uso da linguagem¹⁴ (DRAGER, 2010: 473)

¹⁴ Tradução do trecho: “Production and perception studies can complement each other to aid our understanding of how language change occurs, how stereotypes are formed, how linguistic variation is stored in the mind, and what kinds of cognitive processes underlie language use”.

Tal como se descreve brevemente na introdução à presente dissertação, aqui se investiga quão diferentes de fato são cinco variantes de (-r), na percepção de informantes moradores da cidade de São Paulo: (i) vibrante 3, (ii) vibrante 2, (iii) tepe, (iv) aproximante alveolar e (v) retroflexo. Em outras palavras: os paulistanos de fato ouvem essas diferenças acústicas? Ou ainda: haveria subgrupos de paulistanos que “ouvem” essas diferenças, enquanto outros não? Se é este o caso, o que define socialmente esses subgrupos de ouvintes? Metodologicamente, o trabalho que aqui se apresenta enquadra-se no campo da sociofonética; por um lado, vale-se de ferramentas digitais (os programas Praat e Audacity) para controlar acusticamente os estímulos, garantindo que as diferenças entre as variantes não sejam meramente impressionísticas, mas baseadas em padrões dos formantes (F1, F2 e F3) e imagens acústicas das produções (que serão detalhadas no capítulo 3). Por outro lado, parte da hipótese de que certas características sociais dos ouvintes interagem com sua percepção dessas diferenças acústicas.

Essas variantes são alvo de metacomentários (descritos a seguir, no capítulo 2) por parte de informantes paulistanos. Os significados sociais ligados a elas são, portanto, acessíveis a esses informantes (mas não homogeneamente) e, por sua vez, podem influenciar o modo como eles classificam as diferenças acústicas entre elas. Análises quantitativas permitem verificar quais fatores (linguísticos e extralinguísticos) influenciam a avaliação dos informantes – por exemplo, seu local de nascimento (e, conseqüentemente, o dialeto a que foram expostos desde o nascimento), seu sexo/gênero, sua idade, o bairro onde moram na cidade, etc.

Capítulo 2

A variável (-r)

No capítulo anterior, discutiram-se os preceitos teóricos que dão base a esta pesquisa. No presente capítulo, a discussão concentra-se na variável (-r) em posição de coda silábica, desde processos fonológicos que podem atingir a classe dos róticos no português brasileiro até trabalhos sociolinguísticos e dialetológicos que abordam a distribuição das variantes dessa variável no Brasil. Finalmente, enfoca-se (-r) em São Paulo, no que toca às variáveis linguísticas e sociais que se correlacionam à ocorrência das diferentes pronúncias variantes, além das percepções e atitudes linguísticas acerca delas.

2.1 O /-r/ em coda no português brasileiro

Os róticos são sons difíceis de classificar. Enquanto a maioria das classes referidas nas teorias fonéticas são definidas por propriedades articulatórias ou auditórias, “a unidade desse grupo parece residir principalmente nas relações históricas entre os subgrupos e na escolha da representação pelo caractere ‘r’” (LADEFOGED; MADDIESON, 1996: 245)¹⁵, uma vez que se enquadram nesse mesmo grupo sons bastante diferentes como as vibrantes, tepe, fricativas e aproximantes. Os róticos não compartilham o mesmo modo de articulação (uma vez que há fricativas, aproximantes e vibrantes), e tampouco há um ponto de articulação específico envolvido, já que esse grupo inclui tanto sons uvulares, alveolares, palatais; tanto coronais como dorsais. Na teoria dos Traços Distintivos, Chomsky e Halle (1968) propõem um conjunto de 22 traços de base

¹⁵ Trecho original: “the overall unity of the group seems to rest mostly on the historical connections between these subgroups, and on the choice of the letter ‘r’ to represent them all”.

articulatória coerentes com a realidade fonética e capazes de descrever as classes fonológicas. Segundo a sua proposta, as consoantes líquidas (classe que inclui os róticos e laterais) formam o único grupo que possui os três tipos de traços de classes principais: soante, consonantal e vocálico. Ou seja, são agregados em um mesmo grupo segmentos com um amplo espectro de variação fonética que não possuem uma propriedade em comum.

Fonologicamente, no entanto, os fones dessa classe tendem a se comportar de maneira semelhante, uma vez que os róticos “ocupam lugares privilegiados na estrutura silábica de diferentes línguas. Não é incomum que sejam a única consoante permitida como o segundo membro no ataque silábico, ou os primeiros membros num conjunto de sons em coda silábica”¹⁶ (LADEFOGED; MADDIESON, 1996: 216). Esse comportamento pode ser justificado pelo valor de soância¹⁷ das líquidas. Em Clements (1985), essa classe é intermediária entre os segmentos de maior valor de sonância (as vogais) e os de menor (obstruintes). De natureza “camaleônica” (VAN DE VELDE; VAN HOUT, 2011), pode-se dizer que uma das características principais desse grupo é o fato de variarem entre si – isto é, constituem variantes de variáveis.

No português brasileiro (doravante PB), são possíveis as seguintes realizações dos róticos: [x] fricativa velar; [h] fricativa glotal ou aspirada; [ɣ] fricativa velar sonora; [χ] fricativa uvular surda; [ʁ] fricativa uvular sonora; [r] tepe (vibrante simples); [r̥] vibrante múltipla; [R] vibrante uvular; [ɾ] vibrante simples retroflexa; [ɹ] aproximante alveolar e [ɻ] aproximante retroflexa (CALLOU et al., 1996; LUCENTE, 2008). Elas compõem um quadro fonológico no qual (i) em posição pré-vocálica inicial, ocorre o “/r/ forte” (vibrante múltipla,

¹⁶ Tradução própria de: “Phonologically, rhotics tend to behave in similar ways. In particular, rhotics often occupy privileged places in the syllable structure of different languages. They are not uncommonly the only consonants allowed as second members of clusters in the syllable onset, or as first members of clusters in coda position.”

¹⁷ Tradicionalmente, o termo utilizado é “sonoridade”, no entanto, optou-se por soância para evitar ambiguidade com “vozeamento”.

fricativa ou aspirada); (ii) em posição intervocálica, ocorrem o tepe e o /r/ forte em oposição fonológica; (iii) em ataques complexos, apenas o tepe; e (iv) em posição pós-vocálica, alternam-se o tepe, a aproximante alveolar, o retroflexo, a vibrante múltipla, a fricativa e a aspirada (LUCENTE, 2008: 1).

Como já se mencionou na Introdução desta dissertação, a variação do (-r) em posição de coda silábica é um dos índices mais expressivos e salientes na diferenciação dialetal no PB (CALLOU et al., 1996; MENDES, 2010; OUSHIRO, 2015) e há inúmeros trabalhos que focalizam essa variável, com interesses fonéticos, fonológicos ou sociolinguísticos (ver, p.e., BRESCANCINI; MONARETTO, 2005, 2008; CALLOU et al., 1996, 1998; LEITE, 2004, 2010; OLIVEIRA, 1999; PIMENTEL, 2003).

No intuito de descrever o sistema fonológico do PB, Mattoso Câmara Junior (1953 apud LEITE, 2004) propõe, no seu trabalho “Para o estudo da fonêmica portuguesa”, que existe apenas um fonema de /r/ na forma subjacente, que seria a vibrante. O tepe seria, portanto, uma variante posicional intervocálica enfraquecida. Abaurre e Sandalo (2003) também defendem a vibrante como a forma subjacente no PB, baseando-se no Princípio do Contorno Obrigatório (OCP – MCCARTHY, 1986, apud ABAURRE; SANDALO, 2003), em que dois elementos adjacentes idênticos são proibidos. Defendem, portanto, a hipótese de que o /r/ forte é um epifenômeno de dois fonemas idênticos adjacentes, que não se superficializam. Argumentam ainda que os alofones de /r/ (em posição inicial, intervocálico, em coda) podem ser gerados a partir dessa vibrante e mostram que essa posição tem poder preditivo na direção da mudança linguística.

Os estudos de Callou et al. (1996) também indicam que vem acontecendo um processo de enfraquecimento do /r/ em posição de coda no PB e, em seu trabalho, indicam o caminho que a mudança linguística estaria seguindo. Eles analisam a produção de /r/ em coda nas cidades de Porto Alegre, São Paulo, Rio de Janeiro, Salvador e Recife a partir de dados coletados pelo Projeto NURC.

Nessa amostra, encontraram sete realizações do /r/ nessa posição: vibrante simples (tepe) [r], vibrante apical múltipla [r], fricativa velar [x ɣ], fricativa velar aspirada [h fɪ], vibrante uvular [R], aproximante retroflexa [ɻ] e o zero fonético. Com base nas distribuições, apontam para uma divisão dialetológica em que Porto Alegre e São Paulo são caracterizados pela maior produção do tepe, ([+anterior]) em comparação às outras realizações, enquanto Rio de Janeiro, Salvador e Recife são marcadas linguisticamente pela produção das fricativas velares e aspiradas (com o traço [-anterior]). A partir desses resultados, bem como de outros processos de posteriorização e enfraquecimento de consoantes (como a palatalização [d] → [ɟ], por exemplo), postulam que a variação do /-r/ no Brasil seria resultado do que eles chamaram de um processo de enfraquecimento a partir do /r/ forte subjacente, que seguiria a trajetória: RR → R → x → h → ∅ (CALLOU et al. 2002: 544).

Como se pode observar, essa sequência não leva em conta o retroflexo, possivelmente deixado de lado devido à sua baixa produtividade nessa amostra (1% do total). No entanto, não se pode dizer que o retroflexo seja pouco produzido no Brasil. Brandão (2007) faz um levantamento dos Atlas Linguísticos Brasileiros e encontra ocorrências documentadas de pronúncias retroflexas em todas as regiões do país, em maior ou menor grau, sendo em São Paulo, Paraná, Mato Grosso do Sul, Mato Grosso, Goiás e Minas Gerais os lugares com maior proporção dessas variantes (que incluem o tepe retroflexo e a aproximante alveolar, além da aproximante retroflexa). Pensando na importância das pronúncias retroflexas para a distinção da variedade brasileira do português, Rennie (2011) busca as origens desse segmento no PB. Ela adota o ponto de vista de São Paulo como um polo “irradiador” dessas variantes para outros lugares do país e argumenta contra as hipóteses de o retroflexo ter surgido a partir de contato com línguas indígenas ou africanas, uma vez que nenhuma língua do grupo tupi-guarani ou bantu (que tiveram maior contato com o

português) têm esse som em seus sistemas. A autora defende que a origem do retroflexo está ligada possivelmente à posteriorização do /l/ em coda e ao rotacismo (ou seja, à “substituição” do /l/ pelo /r/ como em *caldo* > *cardo*) (HEAD, 1987, apud RENNICKE, 2011). Como o rotacismo é ligado principalmente a baixa escolaridade – e restrito a alguns grupos – nem todo mundo que produz o retroflexo realiza rotacismo. Ela aproveita ainda de Cohen (2006 apud RENNICKE, 2011), que baseia suas análises em amostras de fala, o argumento de que o retroflexo teria se originado tanto do /l/ quanto do /r/, em dois processos diferentes, que aconteceram em Minas Gerais e em São Paulo. No primeiro, a mudança teria acontecido na sequência: /l/ > velarização > retroflexão > vocalização > Ø; enquanto que em São Paulo não houve a fase de velarização: /l/ > retroflexão > vocalização > Ø. A partir de análises dos Atlas Linguísticos, ela observa que quanto mais ao norte do país, menores as taxas de ocorrência do retroflexo.

Documentado desde o início do século XX, o retroflexo foi primeiro descrito como parte do “Dialeto Caipira” por Amadeu Amaral (1920), que o tratou como parte de “vícios feios de linguagem”, que “contaminava” o falar de uma “minoría culta” da capital (AMARAL, 1920: 1). À época, o autor ainda postulou que, por ser típica de moradores do interior do estado de São Paulo, marginalizados e restritos à zona rural, essa variante estaria fadada a desaparecer em um breve período de tempo. A previsão de Amaral não se concretizou, mas é possível dizer que, ainda hoje, as pronúncias retroflexas sofrem estigmatização. Leite (2004), ao estudar atitudes linguísticas de estudantes de São José do Rio Preto residentes em Campinas (ambas cidades no interior de São Paulo), mostra que há maior emprego do retroflexo por alunos ingressantes, enquanto nos últimos anos de faculdade eles tendem a usar mais a forma de aproximante alveolar e realizações vocálicas (que ela chama de “vogal colorida”; LEITE, 2004: 19). Os próprios falantes de sua amostra indicaram que, por terem sido alvo de

provocações diante do uso do retroflexo (que seria avaliado como “feio” e “puxado”) tenderam a modificar sua pronúncia para se aproximar do padrão campineiro (que seria “neutro” e “ideal”). No entanto, em uma análise acústica das pronúncias, Leite (2010) chega à conclusão de que tanto a pronúncia em Campinas quanto a dos estudantes vindos de São José do Rio Preto teriam o mesmo “grau” de retroflexão. Esse resultado é mais um indício das “interferências” extralinguísticas nas avaliações dos ouvintes a respeito de formas linguísticas. Nesse caso, da ideologia a respeito dos falares campineiro e rio-pretense – que provavelmente se intersecciona com ideologias a respeito das duas cidades.

2.2 A variável (-r) em São Paulo: produção

Excetuando-se o trabalho de Callou et al. (1996), que trataram com dados do NURC, coletados nos anos 1970, os estudos sociolinguísticos sobre /-r/ na cidade de São Paulo, principalmente no que toca à distribuição e padrões de ocorrências das variantes de (-r) em coda, são recentes (MENDES, 2010; MENDES; OUSHIRO, 2011; OUSHIRO, 2015). De maneira geral, essas pesquisas mostram que a variante tepe¹⁸ é a mais produtiva no dialeto paulistano; entretanto, não se pode dizer que a pronúncia retroflexa seja rarefeita. Embora os trabalhos tragam diferentes resultados a respeito da taxa de uso do retroflexo, já em um primeiro momento é possível verificar que a fala paulistana não é tão homogênea em relação à pronúncia do /r/ como se poderia supor.

Mendes (2010) analisou a fala de 24 informantes paulistanos considerados “prototípicos”, ou seja, residentes em regiões mais centrais de São Paulo e que se identificavam com um certo tipo de “paulistanidade” e estilo de vida da

¹⁸ Como já se indicou na Introdução dessa dissertação, os trabalhos sobre produção e percepção sociolinguística do /r/ em São Paulo não realizaram análises acústicas das pronúncias. Assim, tepe e vibrantes (em conjunto) se opõem a retroflexo e aproximante alveolar (também em conjunto).

metrópole. Em sua amostra, estratificada de acordo com Sexo/Gênero, três Faixas Etárias e dois níveis de Escolaridade, encontrou uma taxa de emprego de 12% do retroflexo e um indício de mudança aparente em favor do tepe – ou seja, os informantes mais jovens apresentaram maior taxa de tepe do que os informantes mais velhos.¹⁹ Mendes e Oushiro (2011) analisaram uma amostra mais diversificada, de 48 falantes paulistanos, estratificados do mesmo modo como no estudo anterior, mas sem levar em conta o critério de grau de identificação com a cidade. Com tal amostra, encontraram uma taxa de 31% de emprego do retroflexo, com indicativo de variação estável, sem uma diferença significativa quanto ao emprego das variantes em relação às faixas etárias. Oushiro (2015), por sua vez, analisou uma amostra mais robusta e representativa da cidade: 118 entrevistas sociolinguísticas, estratificadas da mesma maneira que as anteriores, mas que incluía também como fator de estratificação o Local de Residência na cidade²⁰. A taxa de emprego do retroflexo encontrada foi de 28,3%, também com indicativo de variação estável. As diferenças de estratificação da amostra provavelmente têm impacto direto nesses resultados, apenas aparentemente contraditórios.

A fim de analisar o encaixamento da variável, no sentido de explicar essas incongruências, Oushiro (2015) incluiu em suas análises outros fatores sociais (que não estratificaram a amostra) como a Mobilidade Geográfica do informante

¹⁹ O construto da mudança em tempo aparente (Labov, 2001) pressupõe que as diferenças no comportamento linguístico de gerações diferentes de falantes em determinado momento refletem diferentes estágios de desenvolvimento da língua. “Sua validade depende crucialmente da hipótese de que a fala das pessoas de 40 anos reflete diretamente a fala das pessoas de 20 anos há 20 anos atrás e pode, portanto, ser comparada com a fala das pessoas de 20 anos de hoje, para uma pesquisa da difusão da mudança linguística. As diferenças entre a fala das pessoas de 40 e 20 anos são atribuídas ao progresso da inovação linguística nos vinte anos que separam os dois grupos” (CHAMBERS & TRUDGILL, 1980: 151 – tradução própria).

²⁰ Essas amostras foram estratificadas de acordo com os critérios estabelecidos pelo projeto SP2010 (<http://projetosp2010.fflch.usp.br>): Sexo/gênero (F/M); Faixa etária (1: 20-34 anos; 2: 35-59 anos; 3: 60 anos ou mais); Escolaridade (até o Ensino Médio/Ensino Superior); Região de residência na cidade (Centro/Periferia). Os critérios para a classificação dos bairros como “centrais” ou “periféricos” se baseiam não só na localização geográfica, mas também na história de ocupação, grau de verticalização, serviços e IDH dos bairros.

na cidade²¹, Local de Origem dos Pais (São Paulo, interior/MG/PR, exterior, Norte/Nordeste) e Classe Social.

A partir dessas análises, a autora constatou que o uso do retroflexo era favorecido por moradores de bairros mais periféricos, enquanto os moradores de bairros mais centrais tendem a evitá-lo. O Local de Residência na cidade tem forte correlação com Classe Social, que se revelou como o critério mais importante para o emprego do retroflexo, o que explicaria as contradições em relação aos resultados de Mendes (2010). O retroflexo também é favorecido por falantes com menor mobilidade geográfica (o que indica que essa variante pode ser a de menor prestígio na comunidade) e por falantes com menor nível de escolaridade. No que diz respeito à classe social dos falantes, o retroflexo é favorecido pelas classes C1 e C2/D (as mais baixas, na sua estratificação), desfavorecido pela classe B2 e ainda mais evitado pelas classes A/B1 (as mais altas).

Quanto à Origem dos Pais, diferentemente do que se poderia imaginar, não são os filhos de pessoas vindas das regiões de maior incidência do retroflexo (interior de SP/MG/PR) que privilegiam essa variante, mas os filhos de nortistas/nordestinos. Segundo a autora, há duas explicações possíveis para que esse grupo não retenha as variantes aspiradas típicas dos dialetos dos seus pais: a primeira, de natureza linguística, advém do fato de que o retroflexo se aproxima mais do aspirado por ser um segmento [-anterior] em relação ao tepe; a outra, de natureza social, está na possibilidade de que a variante seja mais empregada por eles de acordo com a sua posição na cidade. Ela cita um levantamento do IPEA (2011 apud OUSHIRO, 2015) que constata que esses migrantes atingem em média os menores níveis de escolaridade, têm menos

²¹ Em sua tese, Oushiro (2015) classifica os falantes como sendo de maior ou menor mobilidade geográfica de acordo com o lugar em que já viveram: (i) sempre no mesmo bairro, (ii) sempre na mesma zona, (iii) em diferentes zonas ou cidades. “A hipótese é que falantes com menor mobilidade tendem a se orientar por normas locais, ao passo que aqueles com maior mobilidade, menos enraizados, adotam valores externos e das classes mais altas. (Milroy, 1987[1980])” (OUSHIRO, 2015: 46).

acesso à internet e rendas mensais relativamente mais baixas, o que caracterizaria pertencimento a classes sociais mais baixas.

Comparando esses resultados com os de Mendes (2010) e Mendes e Oushiro (2011), há indícios de que o retroflexo e o tepe têm movimentos divergentes na comunidade, uma vez que classes mais altas favorecem o tepe e as mais baixas tendem a favorecer o retroflexo. Cruzando-se os dados de produção com faixa etária, ela descobriu que, assim como os jovens de áreas mais centrais tendem a favorecer o tepe, os jovens de áreas mais periféricas tendem a privilegiar fortemente o retroflexo. Essas relações podem ser vistas na Figura 2.1.

Figura 2.1: Pesos relativos referentes ao uso de retroflexo no cruzamento de Faixa Etária e (a) Região de Residência; (b) Nível de Escolaridade e (c) Classe Social.
(Fonte: Oushiro, 2015:115).

Nos gráficos, as linhas contínuas dizem respeito aos fatores que mais favorecem o retroflexo: habitantes de bairros mais periféricos, menos escolarizados e pertencentes às classes C1-C2. Os habitantes de bairros mais centrais, mais escolarizados e de classe social mais alta (linhas tracejadas) também esboçam tendência ao retroflexo quando se considera a diferença entre segunda e terceira faixas etárias. No entanto, na primeira faixa etária, os movimentos são opostos: enquanto aqueles mais escolarizados, moradores de regiões centrais e de classes mais altas desfavorecem fortemente o uso do retroflexo, os menos escolarizados, moradores de regiões periféricas e de classes

mais baixas tendem a favorecê-lo. Essas tendências opostas podem indicar que as duas variantes potencialmente têm significados sociais distintos para grupos sociais diferentes; e, pelos resultados dos gráficos, a divisão socioeconômica parece ser a mais relevante nesse caso.

2.3 A variável (-r) em São Paulo: percepção

Crenças e atitudes dos falantes em relação a variáveis linguísticas podem ser avaliadas de diferentes maneiras. Mendes e Oushiro (2011a) e Oushiro (2015) fizeram testes de percepção para tentar desvendar os significados sociais dessas variantes na cidade e como os falantes se comportam em relação a elas. Para isso, elaboraram um experimento com a técnica *matched-guise* (LAMBERT et al., 1960; CAMPBELL-KIBLER 2006).

Eles recolheram excertos de fala de quatro informantes paulistanos (dois homens e duas mulheres com cerca de 30 anos, moradores da zona oeste da cidade). Cada um dos trechos foi manipulado digitalmente de forma que resultasse em dois estímulos: um em que só havia a pronúncia tepe e outro em que só havia a pronúncia retroflexa. Cada participante ouviu um trecho por falante (ou seja, um mesmo participante só ouviu um trecho com tepe ou um com retroflexo de cada um dos quatro falantes). A partir daí, o participante preencheu um questionário (Figura 2.2) de acordo com a imagem do falante construída a partir do trecho ouvido.

Falante 1
Para você, esta pessoa parece...

falar muito devagar	☐	☐	☐	☐	☐	falar muito rápido
muito tímida	☐	☐	☐	☐	☐	bastante extrovertida
pouco escolarizada	☐	☐	☐	☐	☐	bastante escolarizada
pouco inteligente	☐	☐	☐	☐	☐	bastante inteligente
nem um pouco masculino	☐	☐	☐	☐	☐	bastante masculino
mais casual	☐	☐	☐	☐	☐	mais formal
mais hostil	☐	☐	☐	☐	☐	mais amigável
nem um pouco paulistana	☐	☐	☐	☐	☐	bastante paulistana
não ter sotaque	☐	☐	☐	☐	☐	ter muito sotaque
ter poucos amigos	☐	☐	☐	☐	☐	ter muitos amigos
morar num bairro mais periférico	☐	☐	☐	☐	☐	morar num bairro mais central

Em que zona de São Paulo esta pessoa deve morar?

☐ Norte	☐ Oeste
☐ Sul	☐ Centro
☐ Leste	☐ Não sei

Você acha que esta pessoa está em que faixa etária?

☐ Na adolescência	☐ Por volta dos 30
☐ 20 e poucos anos	☐ Por volta dos 40

Você acha que esta pessoa estudou...

☐ Até o ensino fundamental I	☐ Até o curso superior
☐ Até o ensino fundamental II	☐ Até a pós-graduação
☐ Até o ensino médio	

Você acha que esta pessoa pertence à...

☐ Classe baixa	☐ Classe média alta
☐ Classe média baixa	☐ Classe alta
☐ Classe média	

Qual deve ser o estado civil desta pessoa?

☐ Solteira	☐ Divorciada
☐ Casada	☐ Não sei

Do que você ouviu, você acha que esta pessoa deve ser... (assinale todas que achar pertinente)

☐ Gorda	☐ Simples	☐ Mimada
☐ Alta	☐ Trabalhadora	☐ Profissional liberal
☐ Branca	☐ Desonesta	☐ Conservadora
☐ Tímida	☐ Confiante	☐ Caipira
☐ Descolada	☐ Solidária	☐ Deprimida
☐ Maconheira	☐ Prática	☐ Nerd
☐ Religiosa	☐ Preguiçosa	☐ Sofisticada
☐ Ligada à família	☐ Mal-educada	☐ Independente
☐ Articulada	☐ Engraçada	☐ Patricinha/ Mauricinho
☐ Irritante	☐ Desencanada	
☐ Metida	☐ Sincera	
☐ Confiável	☐ Gay/lésbica	

☐ Outro(s): _____

Você teve alguma outra percepção sobre esta pessoa?

Figura 2.2: **Questionário de percepção aplicado por Oushiro (2015) a partir de Campbell-Kibler (2006).** Fonte: Oushiro 2015:360

Na primeira parte do questionário, as respostas do participante eram dadas em escalas: a respeito da velocidade de fala, timidez, escolarização, inteligência, masculinidade (para vozes masculinas)/feminilidade (para vozes femininas), casualidade, hostilidade, paulistanidade, presença/ausência de sotaque e região de residência na cidade. Depois, deveriam responder a respeito de sua zona de residência, escolaridade, classe social e estado civil. Já na última parte do questionário, os informantes poderiam escolher os adjetivos que

achassem pertinentes àquele falante (ou seja, nem todos esses seriam sempre selecionados).

Mendes e Oushiro (2011b) aplicaram o questionário a 24 paulistanos (13 homens e 11 mulheres, todos universitários, de 20 a 30 anos), ou seja, um grupo bastante homogêneo em relação a faixa etária e grau de escolarização. Esses ouvintes ressaltaram diferenças quanto a paulistanidade, sotaque e centralidade do bairro, a depender do estímulo. Os falantes, quando ouvidos com o retroflexo, foram percebidos como “menos paulistanos”, como tendo “mais sotaque”, e como moradores de regiões “menos centrais” em relação a quando foram escutados nos excertos que continham tepe. Oushiro (2015) ampliou o teste para uma amostra de 185 ouvintes moradores da cidade de São Paulo (129 deles paulistanos), mais diversificada em relação às variáveis sociais. Mais uma vez, essas três escalas – que se relacionam a identidades geográficas – foram as que apresentaram as diferenças mais significativas a depender da variante de (-r). Esses resultados indicam que, nessa amostra, o principal significado social das variantes diz respeito a uma oposição entre um “paulistano prototípico” ou não. Já os resultados em relação à centralidade do bairro indicam uma diferenciação de *status* das variantes. Uma vez que as classes mais altas se concentram em bairros mais centrais e as mais baixas em bairros mais periféricos, a associação do tepe com os primeiros também indica uma associação com classe social. Por outro lado, para as escalas de extroversão, amigabilidade, inteligência e feminilidade/masculinidade não foram observadas diferenças significativas a depender da variante.

A partir desses resultados, a autora elaborou diferentes campos indiciais de significados sociais dessas variantes de (-r) na cidade, utilizando o método de Árvore de Distâncias Mínimas. Este método consiste em um cálculo de correlações entre todos os pares possíveis de variáveis, realizado na plataforma R (R Core Team, 2016) através da função `Dist ()` do pacote `amap` (Lucas, 2014), e

na computação da árvore de distâncias com o auxílio da função `spantree()` do pacote `vegan` (Oksanen et al., 2013) (OUSHIRO, 2015: 312). Os resultados gerais (Figura 2.3) mostram essas principais correlações de significados com as duas variantes na comunidade.

Figura 2.3: **Campo indicial de (-r) em São Paulo.** Fonte: Oushiro 2015:313

Nessa representação, dispõem-se todos os termos que ouvintes paulistanos atribuíram a essas variantes. Os pontos mais próximos entre si mostram os fatores que coocorreram mais frequentemente. A variante retroflexa, por exemplo, aparece fortemente correlacionada aos termos “sotaque”, “caipira” e “simples”, que por sua vez se correlaciona às características “trabalhadora” e “classe baixa”. Já o tepe está mais próximo a “paulistanidade”, “bairro central” e “classe alta”, termo que foi frequentemente ligado a “escolarização”, “inteligência” e “articulada”. Embora esta não seja uma representação do mapa

mental dos participantes, o campo mostra quais relações foram feitas por eles entre as variantes e os conceitos, bem como os possíveis caminhos de associação entre conceitos.

Como todos os termos estão de algum modo conectados na árvore, existe a possibilidade de que qualquer um desses significados seja inferido pelos paulistanos ao ouvir um tepe ou um retroflexo, mas com diferentes probabilidades de associação. Por exemplo, o termo “sincera” está equidistante das duas variantes, o que quer dizer que, a depender da situação, esse termo pode ser relacionado com qualquer uma delas. Já o termo “paulistanidade” está muito mais próximo do tepe, o que mostra que, na comunidade, tal significado se associa mais provavelmente a essa variante do que ao retroflexo (embora essa segunda associação não seja impossível). Dessa forma, em uma certa situação, para determinado ouvinte, o tepe pode ter um valor de “formalidade”, enquanto que, em outra, pode apontar para a característica “desencanada”.

No entanto, uma vez que a teoria dos campos indiciais prevê justamente sua dinamicidade (ECKERT, 2008, 2012), Oushiro (2015) elaborou ainda diferentes campos, a depender do Sexo/Gênero, Região de Residência e Origem dos ouvintes (divididos entre paulistanos e não paulistanos).

Figura 2.4: Campos Indiciais de (-r) de acordo com Sexo/Gênero, Região de Residência e Origem dos ouvintes. Fonte: Oushiro, 2015:315-316

Os resultados (na Figura 2.4) mostram que certos *clusters* de associação são recorrentes, enquanto outros variam a depender do grupo social a que o ouvinte se filia. Entre aqueles que aparecem em todas as árvores estão “paulistanidade-bairro central-classe alta-escolarização” (sempre próximos ao tepe) e “retroflexo-caipira-simples”. Há ainda outros grupos de significados que, apesar de não se associarem exatamente a uma das variantes, são recorrentemente ligados entre si: “simples-trabalhadora-classe baixa-Zona Leste”, “patricinha/mauricinho-mimada-metida-irritante” e “amigos-amicabilidade-extroversão”.

Já em relação às principais diferenças entre as árvores, a autora destaca, por exemplo, que enquanto na amostra geral o termo “sotaque” fica próximo a “retroflexo-caipira-simples”, na árvore dos moradores de regiões mais periféricas ele aparece mais indiretamente ligado ao retroflexo, mediado pelo nó “classe baixa”. Já para os ouvintes não paulistanos, esse termo não só aparece mais distante na árvore, mas como mais próximo à variante tepe do que à retroflexa, o que indica que, para aqueles que nasceram em outras cidades, tanto o tepe quanto o retroflexo são “sotaques”. Já o termo “patricinha” está mais próximo à variante tepe nas árvores das mulheres, da região periférica e de paulistanos, mas em uma distância intermediária entre as duas variantes nas árvores dos homens, da região central e dos não paulistanos.

Esses achados evidenciam como os significados sociais das variantes não são estáticos, mas por serem intermediados por ideologias, variam a depender do grupo social a que o ouvinte se filia. Para o presente trabalho de mestrado, essa dinamicidade dos significados dessas variantes em São Paulo é de fundamental importância. Uma vez que se pretende observar como diferentes grupos atribuem graus de diferenças entre cinco variantes de (-r), é de se esperar que certas diferenças fonéticas sejam mais relevantes para alguns grupos do que para outros. Aventa-se a hipótese de que, nesse jogo de associações, o significado social das variantes seja fundamental na “decisão” (muito provavelmente,

inconsciente) por parte do ouvinte de atribuir maior ou menor diferença fonética entre elas.

2.4 Outras percepções

Conforme já se indicou na anteriormente, tanto as análises de produção quanto de percepção feitas por Mendes (2010), Mendes e Oushiro (2011) e Oushiro (2015) não levaram em conta diferenças acústicas entre o tepe e as vibrantes, nem entre o retroflexo e a aproximante alveolar.

Oushiro (2015: 291), no entanto, chama atenção para o fato de que, para um dos falantes de sua amostra, as percepções dos ouvintes não mudaram a depender da variante ouvida, ou seja, parece que os excertos deste falante eram “imunes” ao impacto do tepe e do retroflexo. Para esse falante, ainda que o tepe tenha influenciado a percepção dos participantes a respeito da sua paulistanidade, os estímulos com retroflexo foram avaliados como uma forma mais neutra em relação aos resultados para os outros falantes. Uma das razões possíveis para isso, segundo a autora, pode ser o fato de que, em comparação aos outros falantes, o retroflexo deste parecia menos acentuado (ou seja, com um “grau” de retroflexão menor). Se essa hipótese for verdadeira, este pode ser um indício de que ouvintes paulistanos associam esse tipo de (-r) (provavelmente, uma aproximante alveolar) a uma fala mais paulistana, em oposição àqueles “retroflexos mais marcados”, que remeteriam a um “sotaque caipira”. Embora não se possa afirmar com certeza que seja esse o caso, esses resultados parecem dar indícios de que os ouvintes paulistanos são realmente sensíveis a esse tipo de “gradação” de retroflexão.

No intuito de examinar as avaliações dos paulistanos a respeito do seu próprio dialeto, Soriano (2014) fez uma análise qualitativa do discurso

metapragmático de 60 informantes do Projeto SP2010 (entrevistados por membros do GESOL-USP entre 2011 e 2012).

O roteiro para a gravação das entrevistas do SP2010 foi elaborado de maneira que começasse com perguntas mais pessoais (a respeito da infância, família, educação, redes sociais do informante), a fim “relaxá-lo” tanto quanto possível, ou seja, no sentido de que ele monitorasse o menos possível a sua própria fala no decorrer da entrevista sociolinguística (e que permitisse ao pesquisador o acesso ao vernáculo – Labov 2006 [1966]).²² Na segunda parte do questionário, as perguntas foram mais direcionadas à relação do informante com a cidade, suas avaliações a respeito de certas variáveis linguísticas e identidades sociais. Pediu-se ainda para que o informante lesse uma lista de palavras (parte da entrevista em que se imaginava encontrar o maior grau de monitoramento da fala), uma notícia de jornal e, por último, um trecho que se chamou de ‘depoimento’: uma narrativa em primeira pessoa com características próprias da fala. As análises de Soriano (2014) concentraram-se nessa segunda parte, principalmente naquelas perguntas relacionadas à avaliação e percepção²³ sociolinguística.²⁴ Transcrevem-se a seguir as principais perguntas analisadas:

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

²² Labov (2006 [1966]), a respeito de entrevistas sociolinguísticas, chama a atenção para o “paradoxo do observador”. A ferramenta da entrevista visa a obter a fala do informante do modo como ele a usa cotidianamente, mas a própria situação de entrevista provoca o distanciamento em relação a um modo mais espontâneo de fala.

²³ Oushiro (2015) diferencia conceitualmente os dois termos: *avaliação* diria respeito ao discurso metalinguístico dos informantes sobre variantes; enquanto *percepção* seriam as inferências feitas pelos usuários de uma língua a partir de um estímulo de outro falante (e que podem ou não ser conscientes, ou seja, podem não ser alvo de um comentário metalinguístico). (p.32)

²⁴ Parte destas entrevistas também foram analisadas qualitativamente por Oushiro (2015). A análise que se apresenta aqui, portanto, não pretende esgotar as possibilidades de interpretação das respostas dos informantes, mas chamar a atenção para pontos diferentes que não foram ressaltados na pesquisa anterior.

- (B) E como é que as pessoas falam na cidade de São Paulo? (evitar usar a palavra “sotaque”)
- (C) (Após a leitura da lista de palavras) A gente estava falando sobre sotaque agora há pouco... Como você acha que uma pessoa do interior falaria algumas dessas palavras? Tem diferenças?
- (D) E como um carioca falaria algumas dessas palavras?
- (E) Tem mais algum sotaque que você conhece?
- (F) (Após a leitura dos outros dois textos) O que você acha do jeito de falar ‘a porta [r] (es)tá aberta [r]’? Quem você acha que fala assim? Você acha que alguém em São Paulo fala assim? Que tipo de pessoa? Você fala assim?
- (G) O que você acha do jeito de falar ‘a porta [ɹ] (es)tá aberta [ɹ]’? Quem você acha que fala assim? Você acha que alguém em São Paulo fala assim? Que tipo de pessoa? Você fala assim?
- (H) E tem mais algum modo de falar que você acha assim bem paulistano?

Vale ressaltar que a orientação dada aos documentadores do projeto SP2010 era de que se evitasse mencionar diretamente que as perguntas focalizavam a variável (-r) e que “exagerassem” as pronúncias, tanto em (F) quanto em (G), a fim de chamar a atenção dos informantes. Sendo assim, seria de se esperar que nem sempre as pronúncias dos documentadores fossem iguais.

É possível que os informantes tenham mencionado as pronúncias tepe e retroflexa em vários momentos da entrevista (como nas perguntas (A)-(E),(H) listadas acima), no entanto, como as perguntas (F) e (G) eram os principais estímulos para a percepção dos informantes, analisou-se, com o auxílio do programa Praat (BOERSMA; WEENIK, 2016)²⁵, as variantes realmente utilizadas

²⁵ A análise foi baseada na estrutura dos formantes (F1, F2 e F3) e na análise visual do sinal acústico das variantes. Seguiram-se os mesmos padrões utilizados para diferenciar as variantes

nesses estímulos. Das 60 entrevistas analisadas, em 12 delas o documentador não fez a pergunta (F) a respeito da pronúncia do tepe e em 9 não fez a pergunta (G) a respeito da pronúncia retroflexa.

Tabela 2.1 – **Pronúncia dos documentadores nos estímulos**

Pergunta	Pronúncia do Documentador
(F) “a porta [ɾ] (es)tá aberta” [ɾ]	Tepe [ɾ] – 37 Vibrante Dois [ɾ-2] – 7 Vibrante Três [ɾ-3] – 4
(G) “a porta [ɻ] (es)tá aberta” [ɻ]	Retroflexo [ɻ] – 39 Aproximante Alveolar [ɻ] – 12

Na pergunta (G), na maior parte das vezes, os documentadores “exageraram” a pronúncia alongando as vogais precedentes [ɔ] (em ‘porta’) e [ɛ] (em ‘aberta’), fossem elas seguidas da aproximante alveolar ou do retroflexo. Em 12 casos, os documentadores produziram a aproximante alveolar. Dos ouvintes que tiveram esse tipo de estímulo, dois disseram ouvir essa pronúncia tanto no interior quanto em São Paulo, sendo que na cidade ela seria mais típica de regiões periféricas, além de produzida por pessoas vindas do interior ou que convivem com nativos dessa região.

(1) D1: e sobre... ‘a porta [ɻ] (es)tá aberta’ [ɻ] ? que que...

S1: não isso não sou eu

D1: (risos) você não fala assim?

S1: não a ‘porta’ (vogal bastante alongada) [ɻ] não (risos) ‘a porta’ (vogal breve) [ɻ]... ‘porta’ [ɾ]... o erre/ normal né

D1: entendi você acha que pol os paulistanos falam assim?

S1: olha... poucos... tsc... eh... não são poucos não

dos estímulos do experimento de percepção realizado nesse trabalho de mestrado. As análises serão mais detalhadas no capítulo 3, que trata da metodologia.

D1: *entendi você acha que tem alguma região assim que... que fala mais uma que fala menos*

S1: *eu acho... de paulistano é porque... ah aqui por exemplo no no nosso universo aqui da subprefeitura... m/ bem pouquinho pessoa então...*

D1: *aham*

S1: *é que são pessoas que moram nessa redondeza é Guaianazes de Itaquera de mesmo São Miguel mas você tem um que dá uma puxada no erre... esse veio do interior né (Gilvan S., M3SP)*

Já dos 39 informantes que ouviram a pergunta sendo pronunciada com o retroflexo, a grande maioria disse se tratar de uma pronúncia típica do interior; quatro disseram também perceber a variante como parte do dialeto paulistano, e apenas um deles a associou a regiões mais periféricas da cidade.

(2) D1: *e... sobre 'a porta [ɾ] está aberta' [ɾ] ?*

S1: *olha... uma pergunta... (risos) não tenho muito o que nem sei*

D1: *você acha que... que é o modo de falar paulistano assim?*

S1: *não... não sei depende se for... ah... eu puxo bastante se eu falar 'a porta [ɾ] está aberta [ɾ]' a minha língua até trava... (risos) mas depende da pessoa... tem pessoas que falam direitinho e não puxa nada então... acho que vai... de paulistano pra paulistano né?*

D1: *ah entendi... você fala de qual jeito?*

S1: *eu falo 'a porta [ɹ] está aberta' [ɹ] (risos)*

D1: *mas você acha que por exemplo 'a porta [ɹ] está aberta' [ɹ] é típico de São Paulo?*

S1: *eu acho que... é eu acho que é sim... é porque tem muitas raça junto né?*

D1: *(risos)*

S1: *o paulistano só tem vários então acaba surgindo um sotaque diferente do outro (Milena F., F1MC)*

Em (2), a pergunta se direcionava ao tepe, sobre o qual a informante não conseguiu tecer algum comentário. Em seguida, Milena F. diz espontaneamente

que ela própria (paulistana, filha e neta de paulistanos) tende a usar a aproximante. Quando o documentador faz a pergunta direcionando-se para a percepção a respeito do retroflexo, suas impressões continuam as mesmas, o que indica que ela aparentemente não diferenciou a aproximante alveolar (que ela produziu) do retroflexo (produzido pelo documentador).

De maneira geral, um dos resultados mais interessantes dessas análises está na constatação de que alguns informantes parecem não distinguir não só as aproximantes, mas também a variante tepe das aproximantes e, ao mesmo tempo, salientar uma diferença entre “dois graus de retroflexão” – que, por sua vez, distinguiriam uma pronúncia tipicamente paulistana de outra tipicamente interiorana.

(3) D1: *e e 'a porta [ɾ] (es)tá aberta' [ɾ]?*

S1: *'a porta [ɹ] (es)tá aber[ɹ]' / eu acho que a al/ ess/ certos tipo de palavra em qualquer lugar você ouve mais naturalmente assim... acho que algumas outras palavras você entende mais o sotaque entende... acho que se fosse mais pro interior ia ficar 'porta [ɹ] (es)tá aberta' [ɹ] sabe ia puxar mas aqui em São Paulo 'a porta [ɹ] (es)tá aberta'[ɹ] (Janaína B., F1MP)*

(4) D1: *e 'a porta [ɾ] (es)tá aberta' [ɾ]?*

S1: *(es)tá ok mas ainda (es)tá bem popular né... Porque o certo seria 'a porta está aberta' (verbo 'estar' pleno)*

D1: *uhum*

S1: *mas a gente fa/ popularmente falando pode (Madalena M., F2SC)*

Na pergunta (F), muitos informantes pareceram não entender que o foco era a variante tepe, como em (2). Em outros casos, como em (4), a informante salientou o uso da forma “tá” no lugar de “está”. Esse fato pode indicar que, para essas pessoas, a pronúncia tepe é “normal” a ponto de a única coisa que pudesse causar estranheza e motivar uma pergunta seria o uso da forma “tá” (OUSHIRO, 2015). Já em (3), apesar de Janaína B. não ter respondido a respeito da variante

tepe, mas ter repetido a sentença-alvo utilizando a aproximante alveolar e continuar respondendo a respeito do /-r/, indica que ela entendeu que a pergunta focalizava a pronúncia de /-r/, mas que a diferença entre as aproximantes alveolar e retroflexa parece mais importante para ela do que a oposição entre o tepe e a aproximante alveolar. Essa diferença pode ser de maior importância, principalmente, porque esses “graus de retroflexão” distinguiriam um modo de falar paulistano e outro que seria tipicamente do interior.

(5) D1: *e... e o... "a porta [r-2] (es)tá aberta [r-2]"?*

S1: *"a porta [ɹ] (es)tá aberta" [ɹ]... eu acho que o paulistano também "a porta [ɹ] (es)tá aberta [ɹ]"*

D1: *e "a porta [ɹ] (es)tá aberta" [ɹ] ?*

S1: *puxando o /e/ né "está entendendo [ẽ]" "aberta" [ɹ] é... paulistano... puxa um pouco o /e/... O /e/ e o erre*

D1: *o erre?*

S1: *é... acho que por isso que eu puxo o erre (Douglas A., M1MP)*

(6) D1: *e [o que você acha] do eh 'a porta [r] (es)tá aberta' [r]?*

S1: *'a porta [ɹ] (es)tá aberta' [ɹ] então isso é que é aquela coisa como/ como em São Paulo tem muita/... muita gente de todo lugar vem pra cá... São Paulo mesmo eu acho que ele foi construído a maior parte dos interiores que vêm então essa coisa do interior de falar 'porta' [ɹ] (com a vogal bastante alongada)*

D1: *uhum*

S1: *a gente tem só que diminuiu então ficou 'porta' [ɹ] (com a vogal breve) (Lucas S., M1MC)*

Em (5), primeiro a pergunta foi apresentada com a vibrante dois e, nesse caso, o informante também pareceu não diferenciá-la das aproximantes alveolar e retroflexa, uma vez que repetiu a sentença-alvo utilizando as duas pronúncias. Quando a pergunta é repetida pelo documentador com a pronúncia retroflexa,

Douglas A. primeiro chama a atenção para a ditongação do /e/ nasal (que, embora não apareça na sentença, tinha sido alvo de uma pergunta anterior), o que faz parecer que, a princípio, ele não captou a diferença entre as duas perguntas. Mais tarde, no entanto, o informante chama a atenção para o “/-r/ puxado”, que ele considera paulistano (e como a forma que ele próprio fala). Em (6), Lucas S. também confunde a pronúncia tepe com a retroflexa, o que parece indicar que ambas são para ele tipicamente paulistanas, mas novamente reforça os dois “graus de retroflexão” como um diferenciador dos dialetos paulistano e interiorano (mesmo que em sua pronúncia ele não faça essa diferenciação).

Embora esses casos possam parecer atípicos, os resultados do teste de percepção desenvolvido por Oushiro (2015), principalmente aqueles em relação ao falante para o qual as respostas têm um padrão diferente, sugerem que “graus de retroflexão” diferentes provocam percepções diferentes na comunidade. Uma vez que a análise das entrevistas não mostrou regularidades quanto ao tipo de estímulo que provoca essa percepção, ela indica que a concepção de que “no interior se fala um /-r/ mais forte do que em São Paulo” já é preconcebida e figura no imaginário de certos falantes da comunidade. Pode-se deduzir que, para esses informantes, diferenciar esses graus de retroflexão seria uma maneira de diferenciar as duas identidades sociais (paulistana e “caipira”, a última mais fortemente correlacionada ao uso do retroflexo). Nesse caso, portanto, não seria o estímulo fonético a suscitar certos significados, mas um tipo de imagem preconcebida que os falantes procuram justificar por meio dos usos linguísticos.

Em comum, os informantes Janaína B., Douglas A. e Lucas S. são pertencentes a classes sociais mais baixas e cujos pais não são paulistanos (os dos dois primeiros são nordestinos e os de Lucas S., mineiros), ou seja, pertencem a

um grupo que tende a favorecer o uso das aproximantes²⁶. Cabe verificar, portanto, se essas diferenças entre os graus de retroflexão são salientes também para diferentes grupos sociais e quais são eles.

(7) D1: *tem algum sotaque do Brasil que você goste mais assim?*

S1: *eu gosto do sotaque paulistano (risos) (...) uma vez eu saía com uma menina que ela que ela falava tudo... ela falava bem assim... ela falava até os erres assim "ah eu vou ver" (pronuncia todas as palavras pausadamente e com o tepe no final de 'ver') não sei acha/ porque também ela fazia teatro então ela marcava bem mas ela fazia um sotaque bem paulistano eu achava engraçado... assim paulistano esse né porque tem paulistano também do do in/ da periferia né que é um pouco diferente*

D1: *como é que é da periferia?*

S1: *não é uma coisa mal/ eh o erre já é um o erre já é um pouco/ pouco parecido com o erre do interior né aquela coisa assim... aí tinha a minha namora/ a minha ex-namorada ela falava igual São Paulo só que ela fazia às vezes ela puxava o erre pouco um pouco diferente assim um pouco mais acho que da periferia... não sei mas acho sim... acho que era por aí porque ela viajava pro interior também e ela morou também na na Zona Sul não nas quebradas né (Rodrigo R., M1SC)*

Em (7), Rodrigo R. não só menciona um /-r/ tipicamente paulistano e “parecido com o do interior”, mas também associa tal pronúncia a um grupo social específico: paulistanos residentes na periferia. Esse é mais um indício de que os falantes de uma variedade podem relacionar grupos diferentes a pronúncias acusticamente bem próximas e ter variadas percepções – a respeito desse grupo e das pessoas que usam cada uma dessas pronúncias – de maneira complexa.

Na pergunta (F), não houve diferença entre as respostas dos informantes quando o documentador usou a vibrante dois ou o tepe. Em ambos os casos, quando alguma identidade social foi relacionada a essas variantes, tratava-se de

²⁶ Oushiro (2015) notou a taxa de retroflexo/aproximante em relação a tepe/vibrante de alguns informantes, entre eles Lucas S., cuja taxa de produção de retroflexo foi de 80% e Milena F., que foi de 74% (p.362)

“paulistano”. Nessa pergunta, somente a vibrante três (que foi pronunciada quatro vezes) causou uma impressão diferente em dois informantes, que a dissociaram de um falar típico da cidade.

(8) D1: *e 'a porta [r-3] (es)tá aberta [r-3]'*?

S1: *ah bem ah com bastante realce [r-3] assim no erre?*

D1: *aham*

S1: *ah isso são bem poucos né não tem*

D1: *é o modo de falar paulistano você acha?*

S1: *não é não é muito arrastado eu acho que é (Gilvan S., M3SP)*

(9) D1: *e 'a porta [r-3] (es)tá aberta [r-3]'*?

S1: *então aí já é mais pro Sul né... 'a porta [r-3] tchê' (José N., M2SP)*

Em (9), a vibração chamou a atenção para outro grupo social distinto, de outra região do país e com outro “sotaque”, mesmo que os dados do Projeto NURC (CALLOU et al., 2002) mostrem que a vibrante também é produzida em São Paulo. Já em (8), embora o informante não associe a pronúncia a algum grupo social, ele rejeita a pronúncia como um modo de falar paulistano e a chama de “arrastada”. Infelizmente, o número de ocorrências da vibrante três no *corpus* é bastante pequeno, o que não permite que se recolham mais indícios a respeito da forma como ela é percebida (seja seu reconhecimento como parte do sotaque paulistano ou não, seja da sua ligação a outro grupo social). No entanto, esses excertos mostram como sutilezas fonéticas entre as variantes de (-r) podem carregar significados sociais em São Paulo, uma vez que a vibrante dois foi, de maneira geral, mais “aceita” como paulistana (provavelmente até confundida com o tepe) enquanto a vibrante três já causou certo estranhamento.

De forma geral, essas análises levam a questões a respeito das percepções fonéticas e sociais por parte dos moradores de São Paulo, que motivaram o experimento que compõe essa pesquisa de mestrado. Cabe verificar, portanto, quais desses detalhes são mais salientes para os falantes dessa comunidade, e se essas percepções relacionadas a tais detalhes são estruturadas entre diferentes grupos sociais.

Parte-se da hipótese de que os informantes tenderão a atribuir um maior grau de diferença fonética entre variantes que, para eles, diferenciem certos grupos ou identidades sociais. Portanto, procura-se verificar, por meio das atribuições de diferença fonética, se algumas pessoas “prestam mais atenção” a certas variantes do que a outras, e se essas percepções distintas se correlacionam a características sociais do ouvinte. O próximo capítulo se debruça sobre os métodos utilizados na elaboração de tal experimento de percepção fonética no intuito de responder a essas perguntas.

Capítulo 3

Metodologia

No capítulo anterior, discutiram-se trabalhos de produção e percepção sociolinguística a respeito das pronúncias tepe e retroflexa do (-r) na cidade de São Paulo. A análise qualitativa do discurso metapragmático dos paulistanos do Projeto SP2010 obteve indícios de que eles distinguem (socialmente) ainda outras variantes que ocorrem na cidade. A fim de verificar, sistematicamente, até que ponto ouvintes paulistanos diferenciam as seguintes variantes de (-r) – (i) vibrante três (daqui em diante, também “r 3”), (ii) vibrante dois (“r 2”), (iii) tepe, (iv) aproximante alveolar e (v) retroflexo –, elaborou-se um experimento de percepção sociofonética, que se descreve a seguir.

3.1 Definições do experimento de percepção

Por um lado, é preciso ir além dos indícios oferecidos pelo discurso metapragmático de alguns dos informantes do Projeto SP2010. Conforme vimos no capítulo 2, a percepção de diferenças acústicas, por parte dos falantes/ouvintes que residem na cidade de São Paulo, vai além do par tepe/retroflexo (as duas pronúncias que constam na literatura como as mais produtivas na comunidade). Lembre-se, por exemplo, dos comentários sobre “diferentes graus de retroflexão”, feitos por Janaína B. (exemplo 3), que nos levam a crer que as diferenças acústicas entre a aproximante alveolar e o retroflexo podem ser percebidas (e tal diferença fonético-fonológica pode ser acompanhada de significação social, ou vice versa). Por outro lado, é necessário verificar se a percepção das diferenças entre as variantes de (-r) são sistemáticas na comunidade: se diferentes paulistanos percebem todas essas variantes; se

características sociais do falante/ouvinte influenciam a percepção acústica dessas variantes (por exemplo, sua idade, a região em que mora na cidade, etc) e se tais percepções são diferentes em certos contextos linguísticos (por exemplo, a depender da tonicidade da sílaba com /-r/). Dessa forma, pensou-se em um experimento que pudesse medir o quanto os ouvintes percebem, como distintos, diferentes pares de variantes de (-r). Note-se, então, que tal experimento não visa a elucidar os significados sociais associados às cinco variantes mencionadas. Aqui, interessa verificar sobretudo se informações de natureza social que caracterizam os ouvintes influenciam suas percepções fonéticas. Em outras palavras, o experimento que se descreve a seguir não incita respostas a respeito dos significados sociais das variantes.

O teste consiste em apresentar aos participantes diferentes pares formados por um mesmo item lexical, nos quais a única diferença é o segmento rótico. Ao ouvir cada par, o ouvinte deve indicar quão semelhantes ou diferentes as pronúncias lhe parecem. Por exemplo: ao ouvir duas vezes seguidas a palavra *oferta* – na primeira, com uma vibrante 3 (*oferta* [r-3]) e, em seguida, com a aproximante alveolar (*oferta* [ɹ]) – o participante indica como percebe as duas pronúncias, numa escala que vai de “Igual” a “Muito diferente” (uma escala numérica, cujos valores são invisíveis ao participante). Evidentemente, os participantes não são informados de que o teste focaliza o (-r), pois interessam suas reações inconscientes a respeito da semelhança ou diferença fonética entre as duas variantes apresentadas. Por outro lado, não se pode afirmar, num experimento como este, que o participante “não ouviu a diferença”, mas sim que ele “não atribuiu um certo grau de diferença”.

3.2 Palavras-alvo

A escolha dos itens lexicais usados como palavras-alvo no experimento foi baseada em resultados a respeito da pronúncia das variantes tepe e retroflexa em São Paulo. Segundo Oushiro (2015), os fatores linguísticos que se revelaram mais fortemente correlacionadas com a variação da pronúncia do /-r/ foram o Contexto Fônico Antecedente e o Seguinte, a Classe de Palavra e a Tonicidade da Sílabas.

O Contexto Fônico Antecedente e o Contexto Fônico Seguinte indicam uma tendência de assimilação com a vogal antecedente (sendo que a pronúncia retroflexa é favorecida quando a vogal antecedente tem o traço [-alto]) e, em menor medida, com a consoante seguinte (sendo que o retroflexo foi privilegiado diante de consoantes que têm o traço [+sonoro] e também o traço [+coronal] – um resultado interessante, já que o tepe também tem o traço [+coronal]). O retroflexo ainda foi favorecido por palavras mais lexicais (verbos, substantivos e adjetivos), em sílaba tônica e em final de palavra. Tomaram-se esses resultados como base na intenção de controlar os fatores linguísticos que pudessem influenciar a percepção das variantes. Portanto, os itens lexicais escolhidos foram balanceados segundo os seguintes critérios:

- O /-r/ precedido pelas sete vogais orais do PB (em caso de posição tônica: i, e, ε, a, ɔ, o, u), e cinco na posição átona, uma vez que não há sílabas átonas com /-ε/ e /-ɔ/ no dialeto paulistano;
- Seguido de /-t/ e /-d/ (consoantes que se diferenciam pelo traço de sonoridade, mas têm o mesmo ponto e modo de articulação,²⁷ e que possuem o traço [+coronal], assim como todas as variantes de (-r)

²⁷ Optou-se por essas duas porque, como se verá adiante, foi necessário manipular digitalmente os estímulos, “recortando” o /-r/ produzido. As oclusivas facilitam essa manipulação, por apresentarem um breve silêncio antes da liberação do ar, enquanto no caso de assimilação seria bem mais difícil determinar no espectrograma onde termina o segmento de /-r/ e começa a consoante seguinte.

em foco neste experimento, possibilitando o controle de uma possível interferência de processos de assimilação);

- Com a sílaba que contém o /-r/ tanto em posição átona quanto tônica;
- Substantivos;
- Itens lexicais que fossem “comuns” (ou seja, que fossem costumeiros na língua falada e que não causassem estranheza a ponto de “distrair” os ouvintes).

No total, formou-se uma lista com 23 palavras, conforme a Tabela 3.1:

Tabela 3.1 – **Itens lexicais utilizados como estímulos**

		t	d
tônica	i	divirto/Airton	_____
	e	aperto	perda
	ε	oferta	cerda
	a	carta	farda
	o	horta	corda
	o	aborto	gordo
	u	iogurte	surdo
átona	i	virtude	_____
	e	certeza	verdade
	a	artéria	pardal
	o	português	mordomo
	u	urtiga	surdez

Trabalhou-se com duas palavras na célula tônica-i-t: divirto/Airton – em razão de suas incongruências relativamente às outras: um verbo e um nome próprio. Optou-se por utilizar as duas e posteriormente avaliar se as percepções

a respeito delas apresentam alguma diferença significativa. Além disso, duas células ficaram em branco, ambas com /-ird/. Nesses casos, houve dificuldade em encontrar palavras que fossem “comuns”; por exemplo, para a posição tônica há verbos da terceira conjugação, para a segunda pessoa do plural, no futuro do subjuntivo ou infinitivo pessoal (*sorrirdes, unirdes, cairdes* etc). Por serem pouco comuns no uso cotidiano, pensou-se que essas palavras pudessem distrair os ouvintes, que poderiam dispor de mais atenção para entender a palavra que foi dita, do que para comparar as duas pronúncias.

Para que os ouvintes não “automatizassem” suas respostas (DRAGER, 2013), foram incluídos no experimento pares distratores formados por itens lexicais que não contivessem /-r/ em posição de coda silábica, mas que também variassem entre si, foram eles: bi[s]coito x bi[ʃ]coito; cad[ej]ra x cad[e]ra; c[o]meço x c[u]meço; co[s]ta x co[ʃ]ta; [d]ia x [dʒ]ia; dinh[ej]ro x dinh[e]ro; [e]scola x [i]scola; [e]stilo x [i]stilo; faz[ẽ]da x faz[ẽj]da; jus[t]iça x jus[tʃ]iça; mensag[ẽ] x mensag[I]; re[s]peito x re[ʃ]peito e t[ẽ]po x t[ẽj]po. A partir desses contrastes, foram formados vários pares, alguns idênticos e outros que de fato variavam.

3.3 Gravação dos estímulos

A principal preocupação para a gravação dos estímulos era a de manter “controladas” quaisquer características que pudessem desviar a atenção dos ouvintes em relação às pronúncias de /-r/. Uma vez que não é possível saber quais características do falante podem suscitar diferentes percepções, sejam elas em relação ao perfil social ou a qualquer particularidade da fala (CAMPBELL-KIBLER, 2006), decidiu-se por usar somente uma “voz” para o experimento. O informante foi um homem, de 40 anos, nascido em São José dos Campos e residente na cidade de São Paulo há mais de 20 anos, professor universitário e linguista (ou seja, ciente da importância de produzir as variantes necessárias ao

experimento e de manter certas regularidades na pronúncia; mas não um profissional com treinamento vocal que provavelmente teria um controle ainda maior sobre a produção). Para cada item lexical, ele realizou as cinco variantes de (-r).

Os estímulos foram gravados em uma cabine acústica na Universidade de São Paulo. Utilizou-se um gravador Tascam modelo DR-100mkII, com microfone Sennheiser HMD 26. O falante leu os itens lexicais (com todas as variantes necessárias para o trabalho) quatro vezes. As palavras lhe foram apresentadas em uma lista, o que mais tarde se verificou como um problema, visto que esta disposição incitou que algumas fossem lidas com entonações diferentes. Em razão da disponibilidade do informante, toda a gravação foi realizada em um mesmo dia, sem que se tivesse o tempo necessário para fazer uma análise acústica, no momento da gravação, que pudesse detectar problemas que não poderiam ser corrigidos digitalmente (tal como a não realização de algumas das variantes segundo padrões definidos, que serão apresentados a seguir).

3.4 Análise acústica e manipulação digital dos estímulos

Em trabalhos como este, cujo foco está no detalhe fonético, não se pode contrastar duas variantes apenas de acordo com o que se percebe auditivamente (uma vez que se supõe que essa percepção varia); por isso, exige-se que todas as características acústicas estejam controladas (DRAGER, 2013; CAMPBELL-KIBLER, 2006, 2009). Com essa finalidade, foram realizadas análises visuais do sinal acústico dessas gravações no Praat (BOERSMA; WEENINK, 2016), bem como estabelecidos padrões de classificação para cada uma das variantes, baseados nas estruturas dos formantes. A seguir, descrevem-se os padrões adotados para cada uma delas. Segundo Ladefoged e Maddieson (1996), “um tepe é um som em que um breve contato entre os articuladores é feito pelo

movimento do articulador ativo diretamente em direção ao céu da boca”²⁸ (LADEFOGED; MADDIESON, 1996: 231), o que resulta em um breve espaço em branco no espectrograma, causado pela breve interrupção do ar (SILVA; CLEMENTE; NISHIDA, 2006).

Visualmente, os tepes (intervocálico, final e em grupos) se caracterizam pela descontinuidade espectral, sendo possível observar dois momentos bastante distintos: o início do fechamento (durante o qual a energia de vozeamento é muito baixa, chegando, inclusive, a ser praticamente nula) e o final do fechamento, que coincide no espectrograma com um estouro muito breve (SILVA, 1996: 66-67).

Além disso, os tepes em posição de coda se caracterizam por um elemento vocálico, que pode apresentar até uma estrutura formântica, mais breve que o da vogal central e com menos energia (CLEMENTE; NISHIDA, 2007). Na maioria das produções do informante gravado, o tepe foi facilmente identificável de acordo com essas características.

As figuras a seguir são recortes de telas do programa Praat (BOERSMA; WEENINK, 2016), nos quais é possível visualizar uma imagem do sinal acústico das gravações. As primeiras linhas são os formatos das ondas acústicas (oscilograma), e a caixa abaixo é o espectrograma, em que se verifica a frequência das ondas, medidas em *hertz*. No espectrograma, as linhas horizontais mais escuras representam uma grande concentração de energia, que se agrupam em diferentes frequências, a depender do fone produzido – determinados de acordo com as disposições no trato vocal (como a maior ou menor abertura dos lábios, posição da língua, etc). Para a análise das variantes, faz-se uma comparação do formato das ondas e também uma medição desses pontos (chamados de formantes), a fim de se encontrar padrões para cada variante. A linha mais baixa

²⁸ Tradução própria: “(...) a tap is a sound in which a brief contact between the articulators is made by moving the active articulator directly towards the roof of the mouth.”

é o primeiro formante, F1, relacionado à conformação da faringe e que também sofre influências da altura da língua e abertura da boca (quanto mais alta a posição da língua, mais baixo o F1). O segundo é o F2, que está relacionado à cavidade oral e sofre influências de variações no eixo ântero-posterior (quanto mais anterior, mais alto o F2). O terceiro formante, F3, é afetado pela conformação do trato vocal, sendo que o arredondamento dos lábios provoca um alongamento do trato e abaixa F3 em direção a F2 (REEZT; JONGMAN, 2009: 184). Nas imagens, as partes destacadas em amarelo correspondem às variantes de /-r/.

Figura 3.1 – Produção da palavra “cerda” com tepe

Na Figura 3.1, é possível notar que, após a vogal [ε], há um breve espaço em branco no espectrograma (que indica a interrupção da passagem do ar) seguido de um estouro breve de energia (indicado pela seta vermelha), o que caracteriza um tepe “típico”. Em todas as produções do tepe o sinal acústico deveria ser parecido com este.

As vibrantes têm uma estrutura bastante parecida com a do tepe, diferenciando-se pelo número de vezes em que a língua toca o céu da boca.

Cada período consiste numa fase de fechamento, durante a qual os articuladores estão em contato, sucedida por uma fase de

abertura na qual eles estão brevemente separados. No espectrograma, as fases de fechamento aparecem como áreas mais claras, em que a energia formântica é nula ou muito fraca. A fase de abertura entre os fechamentos, que tem características de vogal em sua estrutura acústica, aparece como uma área mais escura, com concentração de energia em regiões características dos formantes (LADEFOGED; MADDIESON, 1996: 218).²⁹

Figura 3.2 – Produção da palavra “pardal” com vibrante de três batidas

(As setas indicam as fases de fechamento)

²⁹ Tradução própria: “Each period consists of a closed phase during which the articulators are in contact, succeeded by an open phase in which they are slightly apart. On the spectrograms, the closed phases appear as light areas, as the formant energy is absent or weak. The open phase between the two closures, which is vowel-like in its acoustic structure, appears as a dark areas with concentrations of energy in characteristic formant regions.”

Figura 3.3 – Produção da palavra “aborto” com vibrante de duas batidas

As Figuras 3.2 e 3.3 correspondem às produções de vibrantes. É possível notar que, assim como no tepe, o som é produzido por uma rápida obstrução do ar seguida de um breve estouro de energia. No caso das vibrantes, no entanto, a língua encosta mais de uma vez no céu da boca, o que provoca o efeito de “vibrações”. Na produção de uma vibrante 3 (Figura 3.2) a língua bate três vezes – como é possível verificar com os pequenos espaços em branco seguidos de explosões (indicados pelas setas vermelhas) – enquanto na vibrante 2 isso acontece duas vezes.

As aproximantes alveolar e retroflexa também têm propriedades acústicas similares entre si, sendo que, no geral, elas se caracterizam pelo abaixamento de F3 em direção a F2 e o arredondamento dos lábios.

O mais notável a respeito desse som [as aproximantes, no geral, tanto a aproximante alveolar como o retroflexo] é a frequência muito baixa do terceiro formante (...) Quando há constrição nas regiões faringal ou palatal, modelos acústicos de produção de vogais preveem um terceiro formante relativamente baixo, próximo ao segundo formante (Fant, 1968). (...) falantes do inglês americano usam mais ou menos uma articulação retroflexa, que é combinada com uma constrição na faringe, assim como um arredondamento dos lábios (Delattre & Freeman, 1968). Essa

combinação articulatória também produz um terceiro formante baixo (LADEFOGED; MADDIESON, 1996: 234).³⁰

Tanto a aproximante alveolar quanto o retroflexo têm características de vogal e, no espectrograma, parecem “unidas” à vogal antecedente. Para diferenciar as duas, fez-se uma análise visual do sinal acústico, contrastando as produções de cada item lexical. Verificou-se o padrão de que a aproximante alveolar possui F3 baixo em direção a F2, enquanto o retroflexo, após o abaixamento de F3, é caracterizado por uma breve curva ascendente, indício do movimento de retroflexão da língua. Fazendo-se as medições dos formantes (tomando como critério o momento de transição da vogal precedente para o /-r/), encontrou-se um padrão de distância entre F2 e F3: enquanto nas aproximantes alveolares a média de distância entre os formantes foi de 550 Hz ou mais, nos retroflexos essa distância diminui para cerca de 300 Hz, o que indica um fechamento maior da cavidade oral. Esses critérios de comparação foram adotados para todos os itens lexicais.

³⁰ Tradução própria: “Most striking about this sound is the very low frequency of the third formant (...) When there are constrictions in the lower pharyngeal region or in the palatal region, acoustic models of vowel production predict a relatively low third formant, close to the second formant (Fant, 1968) (...) Speakers of American English use a more or less retroflex articulation for ɻ, which is also combined with a constriction in the lower pharynx, as well a lip rounding (Delattre & Freeman, 1968). This articulatory combination also produces a low third formant.”

Figura 3.4 – Produção da palavra “português” com aproximante alveolar

Figura 3.5 – Produção da palavra “português” com retroflexo

Figura 3.6 – Detalhe da aproximante alveolar (à esquerda) e do retroflexo (à direita) nas produções da palavra “português”

Na Figura 3.4, está destacada a produção da aproximante alveolar na palavra “português”; note-se, no espectrograma, que o terceiro formante (terceira linha mais escura, de baixo para cima) se abaixa em direção ao segundo formante, o que indica a aproximação da língua em direção ao céu da boca. Já na Figura 3.5, que corresponde à produção do retroflexo na mesma palavra, é possível verificar que os padrões dos formantes são parecidos para as duas aproximantes, mas no retroflexo F3 é mais baixo e, depois de se aproximar de F2, ascende novamente.

Ainda que todas as variantes de (-r) houvessem sido produzidas de maneira satisfatória de acordo com esses critérios, atentou-se também para diferenças de entonação entre duas “versões” do mesmo item lexical, bem como diferenças de silabação e de ênfase nas vogais. Constatou-se que não poderiam ser obtidos resultados confiáveis no experimento com esse tipo de diferenças entre as pronúncias, uma vez que tais incongruências parecem chamar mais a atenção do que as diferenças entre as variantes do (-r) em alguns casos.

Nas Figuras 3.7 e 3.8 é possível visualizar as diferentes curvas entonacionais (representadas pelas linhas azuis) produzidas no mesmo item lexical, “urtiga”. Na primeira (3.7), em que a palavra é produzida com a aproximante alveolar, a entonação está crescente (a linha azul sobe no final da palavra), enquanto na segunda, com o retroflexo, a entonação é descendente (a linha azul sobe na primeira sílaba e desce do meio para o fim da palavra). Se essas pronúncias, em que a diferença acústica entre as duas variantes é tão sutil, fossem utilizadas como estímulos, seria possível que a diferença de entonação fosse mais saliente para o ouvinte do que propriamente as diferenças entre essas duas variantes de /-r/.

Figura 3.7 – Curva de entonação na palavra “urtiga” produzida com a aproximante alveolar

Figura 3.8 – Curva de entonação na palavra “urtiga” produzida com o retroflexo

Nesses casos, fez-se necessária a manipulação digital dos estímulos, no sentido de “corrigir” essas diferenças, com o auxílio dos programas Praat (para análise acústica) e Audacity (para a edição de fato: como “recortar” e “colar” os sons).

Figura 3.9 – Curva de entonação na palavra “urtiga” produzida com a aproximante alveolar, após manipulação digital

Figura 3.10 – Curva de entonação na palavra “urtiga” produzida com o retroflexo, após a manipulação digital

Para garantir uniformidade, todos os estímulos produzidos sofreram o mesmo processo de edição, mesmo aqueles que não apresentavam diferenças tão salientes. Para cada item lexical, foi escolhida como base uma pronúncia que tivesse mais características em comum com a maioria delas (ou seja, curva entoacional semelhante, silabação – ou ausência dela –, mesma amplitude do som etc). Depois, foi recortado o segmento /-r/ (cada variante) de cada versão para o mesmo item lexical e colada sobre essa base comum. Por exemplo, na palavra “iogurte”, foi escolhida como base aquela que tinha sido produzida com a vibrante 2. A vibrante foi recortada da produção e, em seu lugar, foi colada uma

vibrante 3 que havia sido produzida em outra versão da mesma palavra, dando origem a um novo estímulo. O mesmo foi feito com o tepe, a aproximante alveolar e o retroflexo. No caso das aproximantes, como ocorre assimilação com a vogal antecedente, foi necessário recortar também essa porção da vogal para que o estímulo soasse natural – ou ainda parte da consoante antecedente, como no caso de “iogurte”, em que a vogal se assimila também com /-g/. Em todos os casos, a consoante seguinte inicia-se com um silêncio (total obstrução da passagem do ar), razão pela qual escolheu-se sempre o mesmo ponto de silêncio já “estabilizado” antes do começo da explosão para a marcação final dos /-r/. Como todas tinham a mesma base comum, todas as características permaneceram idênticas, exceto a variante de (-r). Nas Figuras 3.11 a 3.15, é possível ver a imagem acústica dos estímulos de “iogurte” já manipulados.

Figura 3.11 – Estímulo manipulado de “iogurte” pronunciado com vibrante 3

Figura 3.12 – Estímulo de “iogurte” com a pronúncia vibrante 2

(Este estímulo serviu de base para as outras pronúncias no mesmo item lexical)

Figura 3.13 – Estímulo manipulado de “iogurte” com a pronúncia tepe

Figura 3.14 – Estímulo manipulado de “iogurte” com a pronúncia aproximante alveolar

Figura 3.15 – Estímulo manipulado de “iogurte” com a pronúncia retroflexa

3.5 Montagem dos estímulos

Cada item lexical deu origem a dez pares de estímulos, resultantes de todas as combinações possíveis entre as cinco variantes, uma vez que cada estímulo diz respeito a um item lexical sendo pronunciado duas vezes. O esquema a seguir mostra todos os pares possíveis para cada item lexical. Para o perfeito entendimento da figura, convém explicar que, para fins de simplificação, são adotados os seguintes códigos: a vibrante 3 é representada pelo número “3” (ou seja, a quantidade de batidas); a vibrante 2 é representada por “2”; o tepe, por “1”; a aproximante alveolar, por “A” e o retroflexo por “R”. Além disso, como os estímulos se referem à comparação entre uma variante e outra, optou-se por representar as variantes de cada par separadas por x (“versus”).

Figura 3.16 – Formação dos dez pares estímulos, de acordo com a distância na “escala fonética”

A escala “3 2 1 A R” coloca as variantes lado a lado de acordo com a “distância fonética” entres elas; sendo que em uma ponta está a vibrante 3 e no outro extremo, o retroflexo. A depender da distância entre as variantes na escala,

imaginou-se um “grau de diferença fonética” para cada par. Aqueles que têm a linha azul são os de menor grau (e que se encontram lado a lado na escala) e abarcam os pares formados por vibrante 3 e vibrante 2 (3x2), vibrante 2 e tepe (2x1), tepe e aproximante alveolar (1xA) e aproximante alveolar e retroflexo (AxR). Os de “grau 2” estão em verde e abarcam os pares formados pela vibrante 3 e o tepe (3x1), vibrante 2 e aproximante alveolar (2xA) e tepe e retroflexo (1xR). Os de “grau 3”, em lilás, abarcam os pares vibrante 3 e aproximante alveolar (3xA) e vibrante 2 e retroflexo (2xR). Já o par de maior “grau de diferença fonética” é o que está em vermelho, formado pela vibrante 3 e o retroflexo (3xR).

Vale mencionar que esta escala de “distância fonética” é uma abstração feita para fins de “medição” das percepções fonéticas dos informantes. A hierarquia proposta aqui não pretende representar “graus” reais de diferenças acústicas ou articulatórias entre as variantes, de acordo com um suposto critério. Com efeito, a medição de tais diferenças é *justamente* o propósito deste experimento. Dessa forma, prevê-se que aos pares com a maior “distância” nesta hierarquia sejam atribuídos os maiores graus de diferença pelos ouvintes, enquanto àqueles com a menor “distância” sejam atribuídos os menores. Assim, tão interessante quanto saber se as diferenças que os ouvintes percebem (ou indicam como relevantes) seguem essa hierarquia (ou seja, se para eles, a diferença entre a vibrante 3 e o retroflexo (3xR) é muito maior do que a diferença entre as vibrantes 3 e 2 (3x2), por exemplo), é saber se aos pares que estão classificados aqui com o mesmo “grau de diferença” “se comportam” de maneira semelhante. Por exemplo, se para o par formado pela vibrante 3 e o tepe (3x1) será atribuído o mesmo grau de diferença que para o par formado pela vibrante 2 e a aproximante alveolar (2xA), que são equivalentes na hierarquia aqui proposta. Assim, o intuito é desvendar quais dessas diferenças fonéticas são mais e menos relevantes para os ouvintes paulistanos.

3.6 Experimento-piloto

A fim de verificar as condições de aplicação do experimento (tais como a dificuldade da tarefa destinada aos ouvintes, o tempo de execução, bem como resultados que pudessem indicar possíveis erros ou novos *insights*) foi realizado inicialmente um experimento piloto. Para esse experimento, foram inseridos 40 pares de distratores (alguns com produções idênticas entre si) e 214 pares de palavras-alvo (nenhum com produções idênticas), totalizando 254 pares a serem avaliados pelos ouvintes.

Devido ao grande número de condições, previu-se que o experimento pudesse ficar longo e cansativo, o que poderia enviesar as respostas dos participantes. Por isso, ele foi dividido em quatro versões, cada uma delas com os 40 pares de distratores e cerca de 50 pares-alvo. Cada participante ouviu somente uma dessas versões. Garantiu-se que cada uma das quatro versões contivessem todos os itens lexicais, e que fossem equilibradas de forma que aparecessem todos os “níveis” de diferença. Nessa configuração, a duração total do experimento é de cerca de 10 minutos, considerada razoável (DRAGER, 2013).

Para a aplicação, utilizou-se o programa PsychoPy (PEIRCE 2007, 2009), por oferecer recursos que permitem que os estímulos sejam apresentados para os ouvintes em ordem aleatória, o que exclui a possibilidade de os resultados serem enviesados a partir de um estímulo específico e evita os efeitos da aprendizagem (ou seja, o tempo que o participante demora para “se acostumar” com determinado tipo de tarefa requerida) e do cansaço dos participantes; fatores estes que, em uma ordem de apresentação fixa, podem determinar as notas atribuídas a certos estímulos. Esse programa permite que se mantenha uma escala numérica “embutida” àquela apresentada aos participantes e fornece, no *output*, suas respostas organizadas em uma planilha.

Figura 3.17 – Tela de respostas do experimento desenvolvido no programa PsychoPy

A Figura 3.17 mostra a tela de respostas criada para o experimento no programa PsychoPy. Para cada par ouvido (p.e., “pardal-2xA” – ou seja, “pardal” primeiro pronunciada com a vibrante 2 e em seguida “pardal” pronunciada com a aproximante alveolar), os participantes deveriam responder quão diferentes entre si as pronúncias lhe soavam, movendo a seta azul no ponto escolhido da escala. Após a resposta, o participante deveria pressionar a barra de espaço do teclado para que o próximo par de estímulos fosse tocado (p.e., “cerda-3xR”). Ao se mudar de tela, o novo estímulo sonoro tocava automaticamente, portanto o informante só deveria responder na escala e pressionar a barra de espaço. Esta escala que aparece na tela contém os seguintes valores numéricos embutidos: Igual corresponde a 0 (zero) e Muito diferente corresponde a 4 (quatro). Esses valores correspondem à diferença que os participantes atribuíram a cada par, e a partir deles se realiza a análise estatística das respostas.

The image shows a window titled "Experimento" with a close button (X) in the top right corner. Inside the window, there is a form with the following fields and labels:

bairro de residência	<input type="text"/>
escolaridade (EF, EM, ES, PG)	<input type="text"/>
idade (p.ex. 20)	<input type="text"/>
participant	<input type="text"/>
session	001
sexo (F, M)	<input type="text"/>

At the bottom right of the window, there are two buttons: "OK" and "Cancel".

Figura 3.18 – Tela do programa PsychoPy em que os ouvintes deveriam responder a respeito de suas características sociais

Foram ainda recolhidas informações a respeito das características sociais dos ouvintes (como o seu sexo, idade, escolaridade, bairro de residência e local de nascimento dos pais – fatores que se correlacionaram com a produção e a percepção do /-r/ em São Paulo, segundo análises revistas no capítulo anterior). Uma vez que o PsychoPy se refere a um programa (e não uma página na internet), é preciso que esteja instalado no computador em que o teste é realizado. Ponderou-se que para essa fase de testes a aplicação presencial seria ideal, uma vez que permitiria que o pesquisador acompanhasse as reações dos participantes (com que facilidade respondiam aos estímulos, se o teste estava claro para eles, quanto tempo demoravam para responder etc).

De forma geral, os participantes não demonstraram dificuldade em utilizar o programa ou responder aos estímulos. No entanto, alguns disseram ter assinalado menores graus de diferença no começo do teste, esperando que em seguida fosse apresentada alguma diferença “mais saliente”. Principalmente devido a um número desequilibrado de pares distratores, alguns informantes também espontaneamente comentaram perceber se tratar de um teste a respeito da pronúncia do (-r).

Dado seu tempo de realização e a necessidade do uso de fones de ouvido, os participantes não poderiam ser abordados em locais de passagem (na rua, no

metrô e afins), mas em um local em que pudessem se sentar sem pressa. Devido a essas condições, optou-se por realizar o experimento nas dependências da Universidade de São Paulo e do Centro Cultural São Paulo, o que acabou por determinar que a amostra fosse relativamente pequena (40 participantes) e que a faixa etária e graus de escolaridade dos participantes fossem bastante homogêneos (entre 20 e 30 anos, a maioria com ensino superior ou pós-graduação). A amostra foi balanceada para Sexo/Gênero (foram 20 participantes homens e 20 mulheres) e para o Local de Residência na cidade (20 moradores de regiões centrais e 20 de regiões mais periféricas)³¹.

A partir dessas respostas, foram realizadas análises estatísticas na plataforma R (R Core Team, 2016), a fim de investigar se as diferenças encontradas nas percepções dos ouvintes foram aleatórias ou sistemáticas dentro da amostra coletada e se alguma (e qual) característica social ou linguística as influenciam significativamente.

A Tabela 3.2 mostra a distribuição dos 40 participantes de acordo com suas características sociais. Além dos fatores que a estratificam, a amostra também foi organizada de acordo com três níveis de escolaridade (Ensino Médio/Superior/Pós-Graduação) e Local de Origem dos Pais.

³¹ Os participantes foram instruídos a responder o seu *bairro* de residência, para que depois fossem classificados como bairros “mais centrais” ou “mais periféricos” de acordo com os critérios do Projeto SP2010. Assim, os resultados a respeito dessa variável são comparáveis e podem ser discutidos em relação a resultados de Oushiro & Mendes, 2013 e Oushiro, 2015.

Tabela 3.2 – Distribuição dos participantes por características sociais

		Feminino (20)		Masculino (20)	
		Centro (11)	Periferia (9)	Centro (9)	Periferia (11)
Escolaridade					
Médio	(7)	0	5	0	2
Superior	(23)	9	2	5	7
Pós-Graduação	(10)	2	2	4	2
Local de Origem dos Pais					
Ambos de São Paulo	(20)	6	4	3	7
Ambos do interior	(5)	1	0	3	1
Ambos do Nordeste	(2)	0	2	0	0
Ambos de outros	(2)	1	0	0	1
São Paulo e interior	(3)	1	0	1	1
São Paulo e Nordeste	(5)	1	3	1	0
São Paulo e outros	(1)	1	0	0	0
Interior e Nordeste	(1)	0	0	0	1
Interior e outros	(1)	0	0	1	0
Local de Origem dos Pais (reorganizado)					
Ambos de São Paulo	(20)	6	4	3	7
Outras Origens	(20)	5	5	6	4

Como é possível observar na Tabela, a amostra ficou desequilibrada em relação a alguns perfis (p.e., não há nenhuma mulher que more na região central e tenha estudado até o Ensino Médio) por isso, foram feitas reorganizações. A variável Escolaridade ficou bastante desbalanceada (dos 40 participantes, 23 têm nível superior de escolarização, enquanto 10 têm nível de pós-graduação), por esse motivo, ela foi excluída das análises. Já a variável Local de Origem dos Pais³² foi reorganizada em dois níveis (filhos de paulistanos x filhos de não paulistanos).

Por se tratar de um experimento controlado e não de produções espontâneas de fala, as variáveis linguísticas incluídas na análise dizem respeito

³² Na tela final do experimento, os informantes deveriam responder à pergunta “Onde seus pais nasceram?” em duas escalas que tinham as opções “São Paulo”, “interior”, “Nordeste” e “Outros”. Em uma escala eles deveriam responder sobre o local de nascimento da mãe e, na outra, sobre o pai.

àquelas que determinaram a escolha dos estímulos (conforme lista na Tabela 3.1) e estão explicitadas na forma de variáveis na Tabela 3.3.

Tabela 3.3 – Variáveis linguísticas e aleatórias

Variável Dependente	Grau de Diferença Atribuído	0, 1, 2, 3, 4
Variáveis Linguísticas Independentes	Tonicidade	Átona Tônica
	Vogal Precedente	[i, e, ε, a, o, u]
	Consoante Seguinte	[t,d]
	Par de variantes de (-r)	(1) 3x2, 2x1, 1xA, AxR
	(de acordo com a Hierarquia das Diferenças Fonéticas)	(2) 3x1, 2xA, 1xR (3) 3xA, 2xR (4) 3xR

A Tabela 3.4 mostra, de um lado, as médias dos Graus de Diferença Atribuídos pelos participantes a cada par (organizadas da menor para a maior) e, de outro, o lugar do par na Hierarquia. A menor Diferença Atribuída foi para o par formado pelas vibrantes 3 e 2 (3x2), com valor de 0,47 e a maior para o contraste entre a vibrante 3 e o retroflexo (3xR), com 1,92.

Tabela 3.4 – Médias dos Graus de Diferença Atribuídos a cada par de variantes

Par	Grau de Diferença Atribuído (médias)	Grau na Hierarquia Proposta
3x2	0,47	1
AxR	0,72	1
2x1	0,82	1
1xA	1,18	1
3x1	1,25	2
2xA	1,44	2
2xR	1,63	3
3xA	1,71	3
1xR	1,77	2
3xR	1,92	4

O que primeiro chama a atenção nesses resultados é que, no geral, as médias das diferenças foram baixas. Embora o par “mais diferente entre si” tenha obtido a maior média, esta é ainda muito distante de 4 (o maior valor possível na escala). Este pode ser um indício de que as diferenças entre as variantes de /-r/ são menos salientes para os ouvintes do que se imaginava (embora não se possa afirmar se foram menos salientes por razões acústicas, pela expectativa dos participantes em ouvir diferenças mais “evidentes”, ou por avaliações sociais que perpassariam a Atribuição de Diferença Fonética).

De maneira geral, as respostas dos participantes seguiram a Hierarquia proposta: enquanto os pares com as diferenças fonéticas mais sutis (vibrantes 3 e 2 – 3x2; aproximantes alveolar e retroflexa – AxR; e vibrante 2 e tepe – 2x1) tiveram as menores médias, o par que parecia ser mais distante foneticamente (3xR) obteve a maior delas. No entanto, como se poderia imaginar, pares com o mesmo Grau de Hierarquia tiveram médias diferentes. Para o Grau 1, o par formado pelo tepe e a aproximante alveolar (1xA) teve uma média significativamente maior que os outros pares (3x2, 2x1 e AxR)³³.

Foram realizadas também análises multivariadas, em que são testados os efeitos de mais de uma variável (social ou linguística) ao mesmo tempo. Uma vez que a produção da fala acontece em múltiplos contextos (em diferentes situações, por diferentes indivíduos de características sociais diversas, com as variáveis linguísticas em diferentes contextos fônicos, etc) é pouco provável que apenas um fator seja determinante para a variação.

As variáveis linguísticas Vogal Precedente, Consoante Seguinte e Tonicidade *não* se mostraram correlacionadas, assim como as variáveis sociais Local de Origem dos Pais (já reorganizada) e Sexo/Gênero. Já as variáveis Local

³³ $p=0$ (resultado do teste Tukey para o modelo (GRAU DE DIFERENÇA ATRIBUÍDO ~ PAR*LOCAL DE RESIDÊNCIA))

de Residência (social) e Par (linguística) foram consideradas estatisticamente relevantes para a variação nas respostas dos ouvintes.

A função principal dessas análises foi testar possíveis problemas na aplicação do experimento. Por isso, esses resultados não serão largamente explorados nesse momento, mas apenas servem de modelo para ajustes no experimento principal. Por um lado, as análises mostraram resultados satisfatórios (no sentido de aplicação do teste) em relação a algumas hipóteses aventadas, principalmente no que diz respeito à Hierarquia de Distância Fonética. Por outro lado, o fato de somente uma variável linguística e uma social terem sido significativas motivou a inclusão de novas variáveis para o experimento principal, na tentativa de aprofundar os conhecimentos a respeito dos fenômenos que envolvem a percepção linguística.

3.7 Modificações para o experimento principal

A primeira alteração para o experimento principal diz respeito à substituição de alguns dos itens lexicais que formaram os estímulos, por diferentes razões. O item *curto* substituiu o item *iogurte*, na tentativa de eliminar uma possível influência da palatalização do /t/ na percepção³⁴. Análises também mostraram não ter havido qualquer disparidade nas respostas para os itens *divirto* e *Airton*, que divergiam dos demais por serem um verbo e um nome próprio, respectivamente. Sendo assim, optou-se por utilizar apenas *Airton* no experimento, mantendo somente a classe de substantivos.

A principal mudança entre os experimentos piloto e principal, no entanto, diz respeito à plataforma do teste, que foi transferido do programa PsychoPy

³⁴ Embora na palavra *urtiga* também aconteça a palatalização, não foi encontrada uma palavra substituta que obedecesse a todos os critérios estabelecidos (ser um substantivo, de relativa frequência na língua, com /urt-/ na sílaba átona), portanto, esse item foi mantido.

para o site especializado em pesquisas *online* SurveyGizmo. Com esta alteração, foi possível divulgar o experimento pela internet (uma vez que não requer a instalação de um *software* no computador), o que facilitou a participação de um maior número de ouvintes, de perfis sociais mais variados.

O SurveyGizmo oferece as mesmas funções que eram essenciais para esse experimento, tais como a apresentação dos estímulos de forma aleatória (fundamental para evitar que um possível padrão de respostas eventualmente enviesado), o fornecimento das respostas dos participantes organizadas em planilhas e a possibilidade do uso de uma escala numérica “embutida” invisível aos participantes. Na nova plataforma, também foi possível alterar a exibição da escala de cinco pontos delimitados para uma no formato de *slider*, em que os participantes poderiam deslizar o marcador em qualquer ponto (Figura 3.19).

Figura 3.19 – Tela de respostas da pesquisa na plataforma *online* SurveyGizmo

Na tela de respostas, a barra preta é o *player* dos áudios, que tocavam automaticamente quando o ouvinte mudava de página (clitando no quadrado azul com a seta >>); nessa plataforma, ainda é possível controlar o volume e ouvir cada estímulo quantas vezes se quisesse (o que não é possível com o PsychoPy).

A escala, ao meio, continua com os pontos “Iguais” e “Muito Diferentes” como referência, mas o participante pôde deslizar a bolinha azul (cuja posição *default* era o centro, também para evitar possível enviesamento das respostas) até qualquer ponto desejado. Para que essa movimentação fosse possível, foi necessário aumentar os valores da escala, de forma que “Iguais” continua com o valor numérico 0, mas “Muito Diferentes” passa a ter o valor 100.

Também diferentemente do que ocorreu no experimento piloto, o pesquisador não estaria presente no momento da coleta das respostas (e, na maior parte das vezes, sequer conversou com os participantes); por isso, fez-se necessário criar uma página de instruções que explicasse os principais interesses do projeto (sem mencionar a variável (-r)) e, explicitamente, pedisse o uso de fones de ouvido (fundamentais para o bom entendimento do teste). Infelizmente, contudo, não houve um modo de garantir que todos os participantes o tenham feito efetivamente.

Pesquisa de Linguística

Olá! Obrigada por participar.
Meu nome é Larissa Soriano e essa pesquisa faz parte do meu projeto de mestrado em Linguística na FFLCH-USP.
O propósito deste experimento é saber como as pessoas percebem a pronúncia de algumas palavras.
Por isso, **é muito importante que você use fones de ouvido** para realizar a pesquisa.
No momento, só estou coletando respostas de **pessoas que moram na cidade de São Paulo** (não vale para moradores da Grande São Paulo), sejam paulistanos ou não.

O teste consiste em ouvir pares compostos pela mesma palavra e indicar, em uma escala, o quanto as pronúncias foram parecidas ou diferentes na sua opinião. Preste atenção no MODO como as palavras são pronunciadas.
Algumas respostas vão parecer óbvias para você; outras vezes, você pode sentir que está chutando.
Tudo bem! Aqui não há respostas certas ou erradas. **O importante é saber como você percebeu.**
O tempo total do experimento é de cerca de 20 a 25 minutos.

Primeiro, faremos um breve treinamento para você se familiarizar...
Sinta-se à vontade para entrar em contato comigo caso tenha qualquer dúvida em relação ao experimento. Meu e-mail é larissa.soriano@gmail.com

>>

Figura 3.20 – Página de instruções do experimento

Uma vez que alguns participantes disseram ter esperado por “diferenças maiores” (e os resultados de fato mostraram médias de respostas bastante baixas)

incluiram-se três páginas-teste, que também explicavam o funcionamento do experimento; nelas foram apresentados um par (de distratores) idêntico entre si e dois diferentes, sendo que em um deles a diferença era justamente o /-r/ em coda.

Pesquisa de Linguística

▶ 00:00 00:04 🔊

Vamos começar com um treinamento?
Os áudios tocam automaticamente, mas você pode ouvir cada par quantas vezes quiser.

Para responder, deslize a bolinha pela barra até a posição que você acha que representa a diferença entre as pronúncias.

Depois, clique em >> para continuar.

O quanto as duas pronúncias soaram parecidas para você? *

Iguais Muito diferentes

Pesquisa de Linguística

▶ 00:00 00:04 🔊

Ao longo do teste, algumas palavras vão se repetir.
Ou seja, você vai ouvir DOIS pares formados pela mesma palavra, em momentos diferentes.

Mesmo que a bolinha já esteja em uma posição que você considera a ideal, clique na bolinha.
As respostas só são computadas quando a barra ficar cinza escuro.

O quanto as duas pronúncias soaram parecidas para você? *

Iguais Muito diferentes

Figura 3.21 – Páginas-teste

Outros participantes também mencionaram um desbalanceamento entre o número de pares distratores e de outros com /-r/, o que acabava “revelando” a verdadeira intenção do teste. Para que se minimizasse esse problema, incluíram-se novos estímulos distratores (cozim[ê]to x cozim[ěj]to; mo[s]ca x mo[l]ca; prim[ej]ro x prim[e]ro; m[e]nino x m[i]nino; s[ê]tido x s[ĩ]tido; [ê]tão x [ĩ]tão; q[e]rido x q[l]rido e gran[de] x gran [dʒI]), fazendo com que o número ficasse mais próximo. Sendo assim, dividiu-se o teste em cinco “versões”, cada uma delas com dois pares de cada item lexical (seja palavra-alvo ou distrator). Para que o participante fosse “encaminhado” para uma das versões, antes do teste começar, ele deveria responder “Qual o dia do mês do seu aniversário”, com as opções: “De 01 a 06”, que encaminhava para a primeira versão; “De 07 a 12”, que encaminhava para a segunda, e assim por diante. Essa foi a maneira encontrada de aleatorizar a atribuição de grupos de estímulos aos participantes, ao mesmo tempo, com a expectativa de que houvesse um equilíbrio entre os grupos. Apesar dessa divisão, a nova versão do experimento ficou mais longa (os participantes demoraram em média 25 minutos para concluí-lo), o que acabou gerando um grande número de desistências (que não foram computadas no resultado final).

Os participantes continuaram tendo que fornecer alguns dados demográficos (seu sexo/gênero, idade, escolaridade, bairro de residência, local de nascimento e local de nascimento dos pais) ao final do experimento. Além dessas perguntas, foi ainda incluída uma a respeito de qual variante de (-r) ele identificava em sua própria fala.

Pesquisa de Linguística

E entre esses, qual mais se parece com o modo como você fala? *

--Selecione-- ▾

1.

2.

3.

4.

Figura 3.22 – Tela da pergunta “Entre esses, qual mais se parece com o modo como você fala?”

Eles deveriam escolher entre seis pronúncias possíveis do (-r): as cinco variantes de interesse da pesquisa e também uma variante fricativa [x]. A inclusão dessa variante se deve ao fato de que, para essa nova fase, também foram coletadas respostas de moradores da cidade de São Paulo que nasceram em outras regiões do país.

O detalhamento da amostra coletada é feito no próximo capítulo, que se concentra nas análises e nas correlações encontradas.

Capítulo 4

As percepções sociofonéticas a respeito do (-r)

O capítulo anterior apresentou a configuração do experimento a que se dedica este trabalho de mestrado: os critérios adotados para a escolha deste modelo de teste, as análises acústicas e manipulações que permitiram controlar os estímulos, os programas utilizados para a aplicação dos experimentos piloto e principal, bem como breves análises das respostas ao primeiro deles. No presente capítulo, apresentam-se os resultados de análises quantitativas multivariadas desenvolvidas na plataforma R (R Core Team, 2016) a partir das respostas dos participantes ao experimento principal, no sentido de evidenciar as correlações entre, de um lado, as percepções acerca das cinco variantes de (-r) em São Paulo e, de outro, variáveis sociais que caracterizam os participantes do experimento e variáveis linguísticas.

4.1 Variáveis e hipóteses

O Grau de Diferença Atribuído pelos participantes aos pares de variantes de (-r) corresponde à variável dependente das análises apresentadas neste capítulo, uma vez que se espera essas respostas variem a depender de características sociais do participante (“traduzidas” na forma de variáveis independentes, como Sexo/Gênero, Local de Residência, etc), bem como de características linguísticas dos estímulos (Tonicidade da Sílabas, Vogal Precedente, etc). As variantes desta variável são, portanto, de 0 a 100 (todos os valores possíveis na escala).

O experimento de percepção foi aplicado a 109 moradores da cidade de São Paulo. Ele foi dividido em cinco partes (cada uma delas com 44 estímulos,

desconsiderando-se os distratores), o que resultou em 4796 avaliações. A Tabela 4.1 resume as variáveis sociais que caracterizam os participantes e indica sua distribuição.

Tabela 4.1 – Distribuição dos participantes de acordo com suas características sociais

	Feminino (65)		Masculino (44)		TOTAIS
	Centro (36)	Periferia (29)	Centro (25)	Periferia (19)	109
Escolaridade					
Ensino Médio	2	3	2	2	9
Ensino Superior	23	16	11	12	62
Pós-Graduação	11	10	12	5	38
Região de Nascimento					
São Paulo	18	20	13	16	67
Interior de SP, PR, sul de MG e GO	13	3	7	2	25
Grande SP	2	4	2	1	9
NE, Brasília, RJ, norte de MG	3	2	3	0	8
Região de Nascimento dos Pais					
São Paulo	3	6	4	7	20
Grande SP	0	1	0	0	1
Interior de SP, PR, MG e GO	12	5	7	1	25
NE, Brasília, RJ, MG	3	3	5	4	15
Outro	1	0	0	0	1
SP + Interior, PR, MG e GO	8	5	3	3	19
SP + NE, Brasília, RJ, MG	1	2	0	1	4
SP + Grande SP	0	2	0	1	3
SP + Outro	3	0	2	1	6
Interior, PR, MG, GO + Grande SP	2	1	0	0	3
Interior + NE, Brasília, RJ, MG	3	2	3	0	8
Interior + Outro	0	0	1	1	2
Grande SP + NE, Brasília, RJ, MG	0	0	0	0	0
Grande SP + Outro	0	0	0	0	0
NE, Brasília, RJ, MG + Outro	0	2	0	0	2
Região de Nascimento dos Pais (reorganizado)					
Pelo menos um dos pais paulistano	15	15	9	13	52
Todos os outros	21	14	16	6	57

De maneira geral, a inclusão dessas variáveis sociais busca analisar um possível paralelismo entre a produção e a percepção linguísticas. Um contato

maior com certas formas linguísticas mudaria principalmente a saliência (*salience*) e a marcação (*markedness*) delas para o ouvinte, o que, por sua vez, influenciaria o grau que ele atribui às diferenças fonéticas. Campbell-Kibler (2006) explicita o papel da marcação de uma variante para a percepção sociolinguística.

Enquanto variáveis podem ser mais ou menos salientes, em uma única variável as variantes podem se estruturar em termos de marcação. O conceito de marcação se refere a qual membro de determinado paradigma é considerado mais natural, e menos notável em geral ou em relação a fatores contextuais. (...) Em um paradigma de variação, o membro não marcado pode ser aquele que aparece mais frequentemente, que aparece na ausência de fatores condicionantes e que carrega um peso mais leve em relação a significados sociais ou situacionais³⁵ (CAMPBELL-KIBLER, 2006: 9).

De acordo com essa definição, a variante menos marcada em São Paulo seria o tepe, uma vez que é aquela considerada mais prototípica do dialeto, sobre a qual muitos informantes do SP2010 disseram ser o jeito “normal” de falar, enquanto outros sequer teceram comentários, por provavelmente não acharem relevante mencioná-lo³⁶. Em diferentes comunidades, no entanto, onde outra variante de (-r) é a “mais comum”, o tepe teria um efeito de marcação maior.

Os participantes nascidos fora da cidade de São Paulo foram agrupados de acordo com as variantes de (-r) prototípicas de sua região de origem. Por isso, fazem parte do mesmo grupo aqueles que nasceram no interior de São Paulo e

³⁵ Tradução própria do trecho: “While variables may be more or less salient, within a single variable, variants may be structured in terms of markedness. The concept of markedness refers to which member of a given paradigm is considered more natural, and less noticeable, either in general or with respect to contextual factors. (...) In a variable paradigm, the unmarked member might similarly be said to be the one which appears most often, which appears in absence of conditioning factors and which carries a lighter burden of social or situational meanings.”

³⁶ Lembre-se, por exemplo, do excerto (4) no capítulo 2, em que ao ser perguntada sobre o que achava do modo de falar “a po[r]ta (es)tá abe[r]ta”, a informante respondeu a respeito da forma “tá” e não mencionou o /r/.

nos estados do Paraná, Goiás, e sul de Minas Gerais, porque nessas regiões predominam os usos das aproximantes. Em outro grupo estão aqueles que nasceram em estados do Nordeste, Rio de Janeiro, Brasília e norte de Minas Gerais, regiões em que predominam as variantes fricativas de (-r). Já os participantes nascidos na Grande São Paulo foram agrupados à parte, dada a maior incidência de tepe e vibrantes nessa região do que no interior do estado (ao mesmo tempo em que não se pode supor que as aproximantes tenham os mesmos valores sociais que na capital). Essas foram as regiões contempladas pela amostra, de forma que não houve participantes nascidos nas regiões sul (onde há proeminência de vibrantes, p.e.) e norte do país. Com essa classificação, a Região de Nascimento é uma forma de avaliar o quanto a exposição dos ouvintes a certas variantes de (-r) afetaria a forma como eles atribuem as diferenças.

Ainda sobre o conceito de marcação, Campbell-Kibler (2006) elucida:

É claro que, mesmo quando há um sistema de marcação presente em uma variável, não significa que todos os falantes e ouvintes concordarão com essa distribuição de marcação. Da mesma forma, diferentes variáveis serão mais ou menos salientes em cada comunidade. Uma das coisas que influencia a distribuição de marcação e saliência é o modo habitual de falar de um falante/ouvinte, os padrões linguísticos que eles estão mais acostumados a usar e a ouvir. Embora as avaliações de marcação não dependam somente das expectativas que derivam do hábito, é provável que aquelas variantes que se afastam mais fortemente ou inesperadamente de uma experiência costumeira do ouvinte são mais aptas a serem notadas e terem significados atribuídos do que aquelas que se diferenciam apenas ligeiramente³⁷ (CAMPBELL-KIBLER, 2006: 9).

³⁷ Tradução de: "Of course, a system of markedness being present in a variable does not mean that all speakers and listeners will agree on the distribution of markedness. Likewise, different variables will be more or less salient within different communities. One of the things which influences the distributions of markedness and salience is the habitual mode of speaking for a given speaker /listener, the patterns of language they are most accustomed to using and hearing. Although assessments of markedness do not depend solely on expectations derive from habit, it is likely that those variants which depart more strongly or unexpectedly from a listener's customary experience are more apt to be noticed and assigned meaning than those which differ only slightly."

Uma vez que a marcação de certa variante depende do hábito de uso, é bastante provável que, para as pessoas nascidas no Nordeste, Rio de Janeiro, Brasília e norte de Minas Gerais, as variantes fricativas sejam as menos marcadas, enquanto que para aquelas nascidas no Interior de SP, Paraná, sul de Minas Gerais e Goiás, as variantes aproximantes o sejam. A hipótese que se aventa, portanto, é que os participantes tendam a atribuir maiores diferenças a variantes mais marcadas para eles, e essa saliência seria influenciada por suas Regiões de Nascimento.

Nesse mesmo sentido, a variável Região de Nascimento dos Pais pode responder se, a depender dos *inputs* linguísticos que os informantes receberam dos pais ao adquirir sua língua materna, eles atribuiriam maior ou menor grau de diferença fonética a certos pares de variantes. Agruparam-se os filhos de ambos os pais paulistanos; ambos da Grande São Paulo, ambos da “zona do retroflexo”, ambos da “zona das fricativas” e “outros” (no caso da amostra, da região sul do país ou do exterior), além da combinação possível entre todos esses níveis. Como é possível verificar na Tabela 4.1, a distribuição dos participantes de acordo com essa variável ficou bastante desbalanceada: quase metade dos ouvintes são filhos de paulistanos (ambos os pais, ou apenas um deles), enquanto a outra metade se divide entre filhos de pessoas nascidas em outros lugares do país. A variável foi então reorganizada numa diferenciação binária de indivíduos: aqueles cujo pai ou cuja mãe seja natural de São Paulo e aqueles cujos pais sejam ambos de fora da cidade. Esta reorganização não foi feita somente para balanceamento da amostra, mas também porque se considerou plausível imaginar que se a Origem dos Pais fosse um fator relevante, seria possível que os filhos de paulistanos percebessem as variantes de modo diferente do que daqueles cujos pais nasceram em outras comunidades.

Em relação ao Local de Residência, seria plausível considerar que, em São Paulo, os Graus de Diferença Atribuídos às variantes variariam entre moradores de regiões centrais e de regiões mais periféricas – já que, de acordo com os resultados de produção do /-r/ (MENDES, 2010; OUSHIRO, 2015) explicitados no capítulo 2, esta é uma das variáveis que mais se correlacionam com o emprego de uma ou outra variante. A hipótese que se aventava, neste caso, também se relaciona com os conceitos de saliência e marcação. Uma vez que as variantes aproximantes são mais produzidas por moradores de regiões mais periféricas da cidade, espera-se que os participantes residentes na periferia atribuam menores diferenças entre as variantes (tanto aos pares de aproximantes quanto àqueles que contrastam as aproximantes ao tepe).

Uma gama de trabalhos sociolinguísticos que avaliam a *produção* demonstra a importância do Sexo/Gênero para a variação linguística em inúmeras comunidades. Estes trabalhos indicam que as mulheres tendem a produzir as variantes de maior prestígio (LABOV, 2001) na comunidade, fato para o qual as interpretações divergem: enquanto alguns argumentam que essa seria uma forma de elas superarem sua posição desprivilegiada na sociedade, outros afirmam que, na verdade, as formas empregadas por elas tendem a ser consideradas “mais corretas” (OUSHIRO, 2015: 108). Partindo da pressuposição de que os valores sociais das variantes têm grande importância para as atribuições de diferenças fonéticas entre elas, as expectativas a respeito dessa variável são de que as mulheres assinalem diferenças maiores do que os homens àquelas que se distinguiriam por tais valores.

A Escolaridade permitiria responder se, a depender do seu grau de escolarização, os ouvintes tenderiam a atribuir maior ou menor grau de diferença a determinados pares de variantes. A hipótese é de que quanto mais escolarizadas, mais as pessoas se atentam aos valores sociais das formas linguísticas. No entanto, a amostra ficou desbalanceada em relação a essa

variável (100 dos 109 participantes têm nível Superior ou de Pós-Graduação), de modo que a Escolaridade foi desconsiderada nas análises mais uma vez.

Na etapa final do experimento, foram apresentados aos participantes estímulos com diferentes variantes de (-r) – as cinco de interesse desta pesquisa e uma fricativa – para que respondessem qual delas eles reconhecem em sua própria fala. A pergunta deveria ser respondida duas vezes (a ordem de apresentação das variantes mudava na segunda), tanto para notar a consistência de suas respostas, quanto para lhes dar a possibilidade de assinalar duas variantes, caso considerassem que falam mais de uma delas. As análises das entrevistas do Projeto SP2010 mostram que alguns dos participantes que reconheceram as aproximantes como típicas da cidade de São Paulo disseram que eles próprios produzem essas variantes. Assim, aventa-se a hipótese de que as “imagens” que o falante faz a respeito das variantes correlaciona-se com as “imagens” que ele constrói acerca da própria fala, e suas percepções a respeito delas variariam a depender de qual ou quais eles reconhecem como “sua(s)”.

Se uma pessoa nascida no interior de São Paulo, cujos pais também são do interior, marcasse que fala tanto tepe quanto fricativa, seria mais provável que ela não tivesse prestado atenção na segunda pergunta do que de fato produzisse a fricativa. Em casos como esse (que foram raros), a segunda resposta foi desconsiderada. Uma análise qualitativa mostra que embora muitos participantes tenham marcado variantes diferentes nas duas perguntas, na maior parte das vezes elas eram “próximas” uma da outra, como retroflexo e aproximante alveolar, ou tepe e vibrante dois. As Tabelas 4.2.1 e 4.2.2 mostram a distribuição de suas respostas, em relação à Região de Nascimento e Local de Residência.

**Tabela 4.2.1 – Participantes que deram a mesma resposta para as duas perguntas
“Qual dessas variantes você fala?”, distribuídos de acordo com a Região de
Nascimento e Local de Residência**

	São Paulo		Grande SP		Interior de SP, PR, MG e GO		NE, RJ, Brasília e MG		TOTAL
	C	P	C	P	C	P	C	P	
Tepe	11	7	0	3	5	1	0	1	28
Vibrante 2	0	0	1	0	0	0	0	0	1
Vibrante 3	0	1	0	0	0	0	0	0	1
Ap. Alveolar	1	3	0	1	7	0	0	0	12
Retroflexo	0	6	0	0	2	0	0	0	8
Fricativa	0	0	0	0	0	0	3	1	4
TOTAL	12	17	1	4	14	1	3	2	54

C – Centro; P – Periferia

**Tabela 4.2.2 – Participantes que deram respostas diferentes para as duas perguntas
“Qual dessas variantes você fala?”, distribuídos de acordo com a Região de
Nascimento e Local de Residência**

	São Paulo		Grande SP		Interior de SP, PR, MG e GO		NE, RJ, Brasília e MG		TOTAL
	C	P	C	P	C	P	C	P	
1-2	6	2	0	1	1	0	0	0	10
1-3	0	2	0	0	0	0	0	0	2
1-A	4	3	1	0	0	0	0	0	8
1-R	1	3	0	0	0	1	0	0	5
1-X	1	0	0	0	1	0	1	0	3
2-3	1	0	0	0	0	1	0	0	2
2-A	1	1	0	0	0	1	0	0	3
2-R	0	0	0	0	0	1	0	0	1
2-X	0	0	0	0	0	0	2	0	2
3-A	1	1	0	0	1	0	0	0	3
3-R	0	1	0	0	0	0	0	0	1
3-X	0	0	0	0	0	0	0	0	0
A-R	4	6	2	0	2	1	0	0	15
A-X	0	0	0	0	0	0	0	0	0
R-X	0	0	0	0	0	0	0	0	0
TOTAL	19	19	3	1	5	5	3	0	55

As tabelas explicitam as inúmeras combinações possíveis para as respostas e a relação (já esperada) entre a Região de Nascimento e as variantes que o

participante reconhece em sua fala. Por exemplo, entre as pessoas que nasceram na “zona do retroflexo”, nenhuma afirmou consistentemente reconhecer a variante fricativa em sua fala. Para as análises de efeitos mistos (que serão apresentadas mais adiante), foram feitas reorganizações para que se pudessem balancear os perfis e buscar padrões de correlações entre essa variável a atribuição de diferença fonética.

Já as variáveis linguísticas, como mencionado no item 3.6 do capítulo anterior, dizem respeito àquelas que determinaram as escolhas dos estímulos.

Tabela 4.3 – Variáveis Linguísticas

Variável Linguística	Variantes
Tonicidade	Átona Tônica
Vogal Precedente	[i, e, ε, a, ɔ, o, u]
Consoante Seguinte	[t,d]
Par de variantes de (-r)	(1) 3x2, 2x1, 1xA, AxR
(de acordo com a Hierarquia das	(2) 3x1, 2xA, 1xR
Diferenças Fonéticas)	(3) 3xA, 2xR
	(4) 3xR

A partir dos resultados do experimento piloto, espera-se que o Par seja a variável linguística mais importante para a variação (uma vez que é o próprio contraste entre as variantes de (-r)).

4.2 As diferenças atribuídas para cada par

Os resultados para Par são gerais, na medida em que nos permitem observar como se organizaram as percepções “como um todo”. A Tabela 4.4 mostra as médias dos Graus de Diferença Atribuídos a cada Par, organizadas da menor para a maior, juntamente de seu lugar na Hierarquia de Diferença Fonética.

Tabela 4.4 – Graus de Diferença Atribuídos por Par (médias)

Par	Grau de Diferença Atribuído (médias)	Grau na Hierarquia Proposta
3x2	22,9	1
2x1	28,1	1
AxR	32,7	1
3x1	46,8	2
1xA	52,4	1
2xA	54,1	2
2xR	66,8	3
3xA	66,8	3
1xR	67,9	2
3xR	72,7	4

Assim como no experimento piloto (item 3.6 do capítulo 3), a nota mais baixa foi atribuída a um par de Grau 1 (3x2, média: 22,9) e a mais alta ao par de maior distância na escala (3xR, média: 72,7). Entretanto, as médias foram mais altas no experimento principal que no piloto: p.e., no primeiro, a média para o par 3xR foi 1,92, o que corresponderia a uma média 48 na escala de 0 a 100.

A Figura 4.1 mostra as médias das diferenças atribuídas aos pares, organizadas de acordo com os graus da hierarquia. As barras mais claras (à esquerda) dizem respeito aos pares de Grau 1 e a mais escura corresponde ao par de Grau 4. As linhas no centro de cada barra mostram o intervalo de confiança para as respostas, ou seja, a variabilidade prevista estatisticamente para cada média. Ao se comparar as respostas em duas ou mais barras, se os intervalos de confiança se sobrepuserem, há maior probabilidade de que os resultados sejam estatisticamente equivalentes.

Figura 4.1 – Médias dos Graus de Diferença Atribuídos para cada Par, de acordo com a Hierarquia de Distância Fonética

O par 1xA recebeu notas muito mais altas do que os outros pares de mesmo Grau 1 (3x2, 2x1, AxR), e do que o par 3x1, de Grau 2. Como é possível observar, a barra correspondente ao par 1xA é maior que as outras e a diferença entre elas é significativa. Esse resultado não é surpreendente, uma vez que tanto para os outros pares do Grau 1 quanto para o par 3x1 a diferença fonética parece ser mais sutil (número de batidas da língua nos alvéolos – 3x2, 2x1 e 3x1 – ou grau de retroflexão – AxR), enquanto o tepe e a aproximante alveolar se diferenciam também pelo modo de articulação, o que pode fazer com que essas variantes soem “mais diferentes entre si” do que aquelas que formam os outros pares.

Os pares 3x2 e 2x1, nos quais as variantes se diferenciam por uma batida, receberam as menores notas (22,9 e 28,1 respectivamente) e não apresentam diferença estatisticamente significativa ($p^{38}=0,12$)³⁹. Já a diferença entre a vibrante

³⁸ O valor de p (ou de significância) mede a probabilidade de se obter determinado resultado no caso de a hipótese nula ser verdadeira. Convencionou-se que o valor de confiança para p é de 5% (ou seja, p menor que 0,05 indica que os resultados são estatisticamente significativos, enquanto p maior que 0,05 indica que não há diferença estatística significativa entre os resultados).

³⁹ Os valores de p apresentados nessa seção dizem respeito aos resultados do teste Tukey para o modelo (GRAU DE DIFERENÇA ATRIBUÍDO~PAR*LOCAL DE RESIDÊNCIA)

três e o tepe (3x1) ou seja, de duas batidas, foi mais saliente (3x1: 46,8 e 2x1: 28,1; $p=0$). Em seus metacomentários, os informantes do SP2010 disseram perceber diferenças de graus de retroflexão entre as aproximantes; no entanto, o par que compara as duas também recebeu notas baixas (AxR: média 32,7). Quando ambas foram contrastadas às vibrantes, contudo, sempre foram atribuídas maiores notas ao retroflexo. A Figura 4.2 mostra o padrão das respostas.

Figura 4.2 – Médias das Diferenças Atribuídas a cada Par

Na figura, os pares estão dispostos da menor para a maior média atribuída, e a linha preta representa o padrão das respostas. Entre os pares 3x2, 2x1 e AxR, a inclinação da linha é pouco acentuada (exceto do par 2xA para 3xA), enquanto para o par 3x1, o ângulo de inclinação é maior. Já do par 1xA até o par 3xR (aqueles que comparam vibrantes a aproximantes), as diferenças parecem menos marcadas, exceto de 2xA para 3xA. O tepe e a vibrante 2 não se diferenciam em comparação à aproximante alveolar (1xA e 2xA, $p=0,99$), enquanto o par 3xA é significativamente diferente de ambos (3xA vs. 1xA; $p=0$ / 3xA vs. 2xA; $p=0$). Já os pares que compararam as vibrantes ao retroflexo (1xR, 2xR e 3xR) são estatisticamente equivalentes.

Esses padrões indicam que as variantes mais salientes para os participantes foram o retroflexo, por um lado, e a vibrante 3, por outro. Isso porque sempre foram atribuídas maiores notas aos pares com retroflexo (1xR, 2xR, 3xR, com exceção de AxR), ao mesmo tempo em que os pares 3xA e 3x1 revelam os maiores ângulos de inclinação na linha do gráfico, ou seja, uma diferença acentuada em relação aos outros.

4.3 Em busca de correlações

Uma vez que se entendeu, no geral, como foram atribuídas as diferenças aos pares, realizaram-se análises multivariadas no sentido de investigar as possíveis correlações entre a atribuição de diferença e as variáveis linguísticas e sociais que pudessem afetá-la. Neste tipo de análise, testam-se os efeitos de mais de uma variável ao mesmo tempo. Concentram-se nos resultados de análises de modelos de efeitos mistos (BAAYEN, 2008), que incluem dois tipos de variáveis independentes: as fixas e as aleatórias. As primeiras dizem respeito às características sociais e linguísticas que são representativas da população⁴⁰ (como o Sexo/Gênero, o Local de Residência na cidade, Tonicidade da Sílabas, Vogal Precedente, etc) e que podem ser replicados em estudos futuros. Já as variáveis aleatórias são aquelas que dizem respeito à amostra coletada, ou seja, os Participantes e os Estímulos apresentados. A inclusão dessas duas variáveis como efeitos aleatórios nos modelos estatísticos permite medir se as correlações encontradas de fato dizem respeito às variáveis sociais ou linguísticas ou à contribuição de algum(s) participante(s) ou estímulo(s) em específico, que se comportaria(m) de maneira muito diferente dos demais, enviesando os resultados (OUSHIRO, 2015).

⁴⁰ Em estatística, o termo *população* se refere ao conjunto de todos os elementos sob investigação. No caso desta pesquisa, portanto, o termo não se refere apenas ao conjunto de participantes do experimento, mas também aos itens lexicais apresentados a eles.

Os modelos de efeitos mistos avaliam também a significância de interações entre as variáveis. As interações mostram que as variáveis não são independentes uma da outra; nesses casos, uma delas pode não influenciar isoladamente a variação, mas, quando cruzadas entre si, ter efeito significativo. Isto equivale a dizer, por exemplo, que é possível que o Sexo/Gênero por si só não explique as diferenças entre as respostas dos informantes (ou seja, as percepções de homens e mulheres não são diferentes como um todo), mas, se houver uma interação positiva com o Par, isso significa que a depender do Par de variantes que o participante ouviu, ser homem ou mulher tem um efeito no modo como a diferença é atribuída. A Tabela 4.5 mostra o valor de significância das variáveis e de possíveis interações entre elas. As interações são marcadas por “___:___”, p.e., “Sexo/Gênero : Par”. Todos os resultados estatisticamente significativos foram destacados em negrito.

Tabela 4.5 – Valores de significância das variáveis no modelo de efeitos mistos

	Soma Quadr.	Quadr. Médio	Graus Liber.	Des Gr.L	Valor de F	p
Local de Residência	67	66,9	1	95,5	0,13	0,71
Par	126715	14079,4	9	181,5	27,8	<0,001
Sexo/Gênero	1829	1829,5	1	96,8	3,62	0,06
Vogal Precedente	3064	510,7	6	126,9	1,01	0,42
Consoante Seguinte	681	681,2	1	126,3	1,34	0,24
Variante que fala	1928	642,7	3	95,5	1,27	0,28
Reg.Nascimento (Pais)	35	35	1	94,9	0,06	0,79
Reg.Nascimento (Ouv.)	1002	333,9	3	95,5	0,66	0,57
Tonicidade	179	178,6	1	126,6	0,35	0,55
Local de Residência : Par	10487	1165,2	9	4428,5	2,3	0,01
Sexo/Gênero : Par	17813	1979,2	9	4440,1	3,91	<0,001
Vogal : Par	23927	443,1	54	128,4	0,87	0,7
Consoante : Par	3698	410,9	9	126,5	0,81	0,6
Variante que fala : Par	48053	1779,8	27	4419,1	3,52	<0,001
Reg,Nascimento (Pais) : Par	8853	983,6	9	4406,3	1,94	0,06
Residência : Var. que fala	1498	499,4	3	96,7	0,98	0,4
Tonicidade : Par	1557	173	9	128,5	0,34	0,95
Reg.Nascimento (ouv.) : Par	24907	922,5	27	4409,8	1,82	0,005

Teste ANOVA para o modelo: (GRAU DE DIFERENÇA ATRIBUÍDO ~ VOGAL*PAR + CONSOANTE*PAR + TONICIDADE*PAR + LOCAL DE RESIDÊNCIA*PAR + SEXO/GÊNERO*PAR + VARIANTE QUE FALA*PAR + VARIANTE QUE FALA*LOCAL DE RESIDÊNCIA + REGIÃO DE NASCIMENTO*PAR + REGIÃO DE NASCIMENTO DOS PAIS *PAR+ (1|PARTICIPANTE) + (1|ESTIMULO)

4.4 Variáveis não significativas

Assim como no experimento piloto, as variáveis linguísticas Tonicidade da Sílabas, Vogal Precedente e Consoante Seguinte não foram selecionadas como significativas nos modelos de efeitos mistos, bem como suas interações com o Par, como mostra a Tabela 4.5. Também não foi selecionada a variável social Região de Nascimento dos Pais, mesmo quando reorganizada (pelo menos um dos pais é paulistano *vs.* ambos os pais não são paulistanos).

4.5 Variáveis estatisticamente significativas

Embora a Região de Nascimento dos Pais não tenha sido significativa, a outra variável regional analisada – Região de Nascimento do ouvinte – mostrou resultados importantes, quando em interação com Par. Assim como explicita a Tabela 4.1, os ouvintes foram agrupados de acordo com a variante de /-r/ mais proeminente em sua região, de modo que se pudesse avaliar se o contato presumível com determinadas variantes afetaria o modo como eles atribuem as diferenças aos pares.

A Figura 4.3 mostra a interação entre Região de Nascimento do ouvinte e Par de variantes ouvidas. No eixo horizontal estão os pares de estímulos, organizados de acordo com a hierarquia das diferenças: os de menor grau estão à esquerda e os de maior à direita. No eixo vertical estão as médias das Diferenças Atribuídas. A linha rosa diz respeito aos informantes nascidos nas regiões típicas das variantes fricativas; a linha azul corresponde àqueles nascidos em regiões cuja variante típica é aproximante, a linha verde, aos nascidos na cidade de São Paulo e, por fim, a linha laranja diz respeito às respostas dos participantes da Grande São Paulo.

Figura 4.3 – Interação entre Região de Nascimento e Par para Grau de Diferença Atribuído

O padrão de respostas dos subgrupos foi parecido, na medida em que todos atribuíram menor diferença para o par 3x2 e maior para 3xR. No geral, os participantes nascidos na “região das fricativas” atribuíram maiores notas a todos os pares, embora a diferença só seja significativa para um par (3x1, $p=0,002$)⁴¹. Levando em conta que essas duas variantes são pouco produtivas nos dialetos dessas regiões, faz sentido, de acordo com os conceitos de marcação e saliência, que essas variantes sejam mais marcadas por esses participantes.

A outra diferença significativa entre os grupos foi justamente para o par que contrasta as duas aproximantes (AxR). Nesse caso, os participantes nascidos na “região do retroflexo” atribuíram uma nota significativamente menor a esse par do que os paulistanos e da Grande São Paulo ($p=0,008$) e também do que aqueles da região das fricativas ($p=0,03$). Uma vez que as aproximantes são frequentes para esses ouvintes, é possível que para eles a diferença entre graus de retroflexão seja percebida como parte de uma variabilidade natural da fala.

Por outro lado, considerando-se que a amostra não é balanceada para essa variável (Tabela 4.1) e que há muito mais participantes paulistanos do que das

⁴¹ Resultados do teste Tukey para o modelo (GRAU DE DIFERENÇA ATRIBUÍDO ~PAR*REGIÃO DE NASCIMENTO)

outras regiões, trata-se de um resultado que é antes um indicativo de algum outro mecanismo que pode influenciar a percepção linguística do que efetivamente uma generalização cabal.

Da mesma maneira que Região de Nascimento do ouvinte, a variável “Variante que o ouvinte reconhece em sua própria fala” busca entender o papel da marcação das variantes. É importante ressaltar, no entanto, que a análise dessa variável não pretende correlacionar usos linguísticos e percepção, uma vez que não é possível saber se o informante *realmente* pronuncia as variantes que indicou. A hipótese por detrás dessa variável está na consideração de que a presença de um discurso ao redor das variantes influencia na sua percepção, na medida em que é possível avaliar sua identificação (ou distanciamento) de certas formas linguísticas, a partir daquilo que ele *diz* pronunciar. A Tabela 4.6 mostra a distribuição das respostas dos ouvintes a essa variável, de acordo com sua Região de Nascimento e Local de Residência na cidade de São Paulo.

Tabela 4.6 – Reorganização da variável “Variante que reconhece em sua própria fala” de acordo com a Região de Nascimento e Local de Residência do participante

	São Paulo		Grande SP		Interior de SP, PR, MG e GO		NE, RJ, Brasília e MG		TOTAL
	C	P	C	P	C	P	C	P	
A, R	5	15	2	1	11	1	0	0	35
1, 2, 3	18	12	1	4	7	2	0	1	45
1, 2, 3 + A, R	6	9	1	0	1	3	0	0	21
X	0	0	0	0	0	0	3	1	4
1, 2, 3 + X	1	0	0	0	0	0	3	0	4
TOTAL	30	36	4	5	19	6	6	2	109

Uma vez que os participantes poderiam escolher entre seis variantes de (-r) – duas vezes – havia 21 combinações possíveis de respostas. Mesmo que nem todas elas fossem assinaladas, ainda assim seriam muitas variantes para uma mesma variável (portanto, generalizações seriam dificultadas, a não ser que

houvesse muitas centenas de participantes no estudo). Por isso, após algumas análises exploratórias, reorganizaram-se os níveis da variável para aqueles expostos na Tabela 4.6. Na primeira linha, estão os participantes que disseram reconhecer em sua fala as aproximantes alveolar e retroflexa (aqui, constam os 12 que responderam “A” para as duas perguntas, os 8 que responderam “R” para as duas perguntas, e os 15 que alternaram entre as duas – Tabelas 4.2.1 e 4.2.2). Na segunda linha, estão os que disseram perceber em sua fala apenas as vibrantes (tepe, vibrante 2 e 3), novamente incluindo-se tanto os que deram respostas consistentes quanto os que alternaram entre as três. Na terceira linha, estão os que disseram se reconhecer como falantes de uma das vibrantes (1, 2 ou 3) e uma das aproximantes (A ou R). Na quarta linha estão os que disseram consistentemente se reconhecer como falantes de uma variante fricativa. Por fim, na última linha estão aqueles que disseram reconhecer em sua fala uma das vibrantes e a fricativa. Essa reorganização, além de mais concisa, se mostrou eficiente para a identificação de padrões de variação nas atribuições de diferenças fonéticas.

Merece destaque a relação dessa variável com o Local de Residência na cidade, principalmente para os participantes paulistanos: dos 20 que disseram reconhecer, em sua fala, as variantes aproximantes (seja alveolar ou retroflexa), 15 são moradores de regiões periféricas e, dos 15 que se disseram falantes tanto de vibrantes quanto de aproximantes (1, 2, 3 + A, R), 9 também são moradores dessas regiões. Embora a interação dessa variável (organizada nesses níveis) com o Par tenha sido selecionada como significativa no modelo de efeitos mistos, um teste de comparação múltipla revelou não haver nenhuma diferença significativa entre os subgrupos em relação aos pares. É possível, portanto, que essa significância tenha sido “mascarada” pelas correlações já mencionadas com a Região de Nascimento e o Local de Residência.

A fim de investigar mais profundamente os efeitos dessa variável, decidiu-se por analisar as respostas apenas dos participantes paulistanos, tanto porque a amostra ficaria mais balanceada, quanto porque assim se excluiria a relação com a Região de Nascimento. Testes exploratórios com esses dados sugerem que a divisão significativa entre os paulistanos corresponde àquela do grupo geral, de modo que os participantes ficaram reorganizados em três subgrupos (correspondentes às três primeiras linhas da Tabela 4.6): os que se reconhecem como falantes de (i) aproximantes, (ii) vibrantes, (iii) tanto vibrantes quanto aproximantes.

De maneira geral, os testes mostraram que as principais diferenças residem entre, de um lado, aqueles que dizem pronunciar apenas vibrantes e, de outro, os grupos de paulistanos que reconhecem alguma aproximante na sua fala, conforme Tabela 4.7.

Tabela 4.7 – Significância das diferenças entre os grupos da variável “Variante que reconhece em sua própria fala”

	diff	lwr	upr	p adj
A, R <i>vs.</i> 1, 2, 3 + A, R	-2,89	-6,26	0,47	0,10
1, 2, 3 <i>vs.</i> A, R	5,80	2,91	8,68	0,00
1, 2, 3 <i>vs.</i> 1, 2, 3 + A, R	8,69	5,59	11,79	0,00

Resultado do teste Tukey do modelo (GRAU DE DIFERENÇA ATRIBUÍDO~PAR* VARIANTE QUE RECONHECE EM SUA FALA)

A Figura 4.4 mostra a variação nas respostas dos grupos para cada Par. A linha verde se refere ao grupo que disse pronunciar apenas aproximantes (doravante A); a linha laranja, ao grupo que reconhece apenas vibrantes (V) e a azul, os que reconhecem tanto vibrantes quanto aproximantes (T). Os triângulos pretos, por sua vez, indicam as diferenças significativas entre os grupos, para os determinados pares.

Figura 4.4 – Graus de Diferença Atribuídos para cada Par, pelos grupos de falantes classificados de acordo com “Variante que reconhecem em sua própria fala” (SOMENTE PAULISTANOS)

Não há diferença significativa nas respostas dos três subgrupos para os pares de menor grau de diferença (3x2, 2x1 e AxR), portanto, eles se distinguem de acordo com o modo como percebem o par 3x1 ou os contrastes entre aproximantes e vibrantes.

O grupo V atribuiu maiores graus de diferenças a todos os pares que comparam vibrantes e aproximantes, enquanto o grupo T atribuiu as menores diferenças a todos eles. Ou seja, o grupo que disse reconhecer vibrantes e aproximantes na sua fala foi o que menos atribuiu diferenças entre elas. Tanto é possível que, por pronunciar os dois “tipos” de variantes, esses participantes não considerem a diferença entre elas tão salientes, quanto o contrário: eles talvez não se atentem para as diferenças e, por isso, marcam ambos como “seu modo de falar”. Ao mesmo tempo, esse foi o grupo que atribuiu a maior diferença à vibrante 3 e o tepe (3x1: média 54,1).

As respostas do grupo A ficaram sempre “intermediárias” entre os outros dois, mas se diferenciam deles a depender do par: em relação ao grupo T, eles atribuíram menor nota a 3x1 (3x1_A: 40,4 e 3x1_T: 54,1); e em relação ao grupo V, menores notas os pares 2xA e 2xR (2xA_A: 47,9, 2xA_V: 61,8 / 2xR_A: 60,7 e 2xR_V: 72). De fato, a distinção reside no modo como o grupo A percebe as vibrantes: eles não as diferenciam entre si (3x2=2x1=3x1), nem quando comparadas às aproximantes (1xA=2xA=3xA e 1xR=2xR=3xR). Eles também não diferenciaram o retroflexo e a aproximante em nenhum dos pares (1xA=1xR, 2xA=2xR, 3xA=3xR). Ou seja, para esse subgrupo de falantes A, tanto a diferença entre o número de batidas da língua nos alvéolos quanto a de grau de retroflexão não são salientes, de modo que só assinalaram as diferenças entre os “tipos” de variantes (vibrantes de um lado, aproximantes de outro). Já os outros dois subgrupos (T e V), por sua vez, consideraram o tepe mais próximo da aproximante do que do retroflexo (1xR_V: 74,1 e 1xA_V: 56,8 / 1xR_T: 59,4 e 1xA_T: 40,7).

Na amostra geral (que inclui tanto paulistanos quanto os “de fora”), a interação entre Local de Residência e Par é significativa, conforme indica a nona linha da Tabela 4.5. Um teste de comparação múltipla revelou que a única diferença entre moradores das regiões centrais e periféricas reside no par 2xA, ao qual os moradores da região central atribuíram uma diferença maior. No entanto, quando se exclui do modelo a Região de Nascimento (ou seja, considerando-se apenas as respostas dos paulistanos), essa interação “perde” sua significância (Tabela 4.8), ao mesmo tempo em que a variável “Variante que reconhece em sua própria fala” continua sendo significativa em interação com o Par.

Tabela 4.8 – **Significância das variáveis no modelo de efeitos mistos (APENAS PAULISTANOS)**

	Sum Sq	Mean Sq	Num DF	Den DF	F. Value	<i>p</i>
Local de Residência	1218	1218,1	1	49,6	2,55	0,11
Par	155087	17231,9	9	454,8	36	<0,001
Sexo/Gênero	6165	6165,4	1	48,83	12,9	<0,001
Variante que fala	2450	272,2	1	49,12	0,57	0,81
Local de Residência : Par	5533	614,7	9	2624,8	1,28	0,23
Sexo/Gênero : Par	19506	2167,3	9	2634,4	4,53	<0,001
Variante que fala : Par	91060	1124,2	81	2631,1	2,35	<0,001
Residência : Variante que fala	2687	537,4	5	49,59	1,12	0,35

Teste ANOVA para o modelo (GRAU DE DIFERENÇA ATRIBUÍDO ~ LOCAL DE RESIDÊNCIA*PAR + SEXO/GÊNERO*PAR + VARIANTE QUE FALA*PAR + VARIANTE QUE FALA*LOCAL DE RESIDÊNCIA + (1|PARTICIPANTE) + (1|ESTIMULO))

Uma vez que a maioria dos paulistanos que moram em regiões centrais relata que pronunciam vibrantes (18 de 30, conforme Tabela 4.9), e os de região periférica reconhecem pelo menos uma aproximante na sua fala (35, de 65) é possível que os efeitos dessas variáveis tenham “se confundido”.

Tabela 4.9 – **Distribuição dos participantes paulistanos de acordo com “variantes que reconhece na própria fala” e Local de Residência**

	C	P	TOTAL
A, R	5	15	20
1, 2, 3 + A, R	6	9	15
1, 2, 3	18	12	30
TOTAL	29	36	65

Esse modelo (apenas com as respostas dos paulistanos) parece mais revelador, uma vez que esclarece que o que realmente influencia o modo como as pessoas atribuem diferenças fonéticas é a sua relação com as próprias variantes linguísticas. Dessa maneira, tanto Região de Nascimento quanto Local de Residência são variáveis importantes para a variação nas respostas, em virtude da ligação indireta que há entre os locais e as variantes de (-r) ali proeminentes –

que é justamente o que o último modelo explicita. Portanto, a atribuição de diferença – ou seja, a avaliação que o falante faz a partir dos construtos sociais a que tem acesso – é intermediada pelas variantes que mais costuma ouvir (que, por sua vez, se correlacionam com o Local de Residência) e de sua consideração a respeito de sua própria fala.

Por fim, a última variável revelada como significativa pelo modelo de efeitos mistos foi Sexo/Gênero em interação com Par na amostra geral (Tabela 4.5) e tanto em interação com Par quanto isoladamente, na amostra apenas de paulistanos (Tabela 4.8).

A Figura 4.5 mostra as médias dos graus de diferença atribuídos por homens e mulheres na amostra geral (que inclui todos os participantes). As barras mais escuras correspondem às médias das mulheres, e as mais claras às dos homens.

Figura 4.5 – Médias dos Graus de Diferença Atribuídos por Homens e Mulheres para cada Par (AMOSTRA GERAL)

As respostas dos dois subgrupos são equivalentes em todos os pares, com exceção daqueles que contrastam o retroflexo às vibrantes, 1xR ($p<0,001$), 2xR ($p=0,007$) e 3xR ($p=0,01$), para os quais as mulheres atribuíram notas maiores. As mulheres consideraram uma diferença maior entre tepe *vs.* retroflexo do que entre tepe *vs.* aproximante (1xR_F: 72,8, 1xA_F: 53,1; $p=0$)⁴², enquanto que para os homens não há diferença estatística entre esses pares (1xR_M:60,2 e 1xA_M:51,4 $p=0,22$).

Ambos os grupos não diferenciam o tepe e a vibrante 2, uma vez que os pares que contêm essas duas variantes são sempre equivalentes entre si: 1xA_F:53,1 = 2xA_F:57, 1xR_F:72,8 = 2xR_F:71,2 e 1xA_M:51,4 = 2xA_M:49,7, 1xR_M:60,2 = 2xR_M:60,4. Ao mesmo tempo, diferenciam as vibrantes 2 e 3 quando opostas à aproximante (3xA>2xA). Uma vez que tanto os homens quanto as mulheres não diferenciam 2xR e 3xR (2xR_F:71,2=3xR_F:76,8 e 2xR_M:60,4=3xR_M:66,4) pode-se aferir que a razão para esses pares terem notas mais altas é o retroflexo, que portanto, seria a aproximante mais marcada para ambos os subgrupos (embora as mulheres atribuam diferenças ainda maiores aos pares em que ele aparece).

De maneira geral, os homens paulistanos foram os que menos atribuíram diferenças aos pares (Figura 4.6). Provavelmente, esta é a razão pela qual, no modelo que inclui apenas respostas de paulistanos, Sexo/Gênero foi uma variável selecionada como significativa, além de sua interação com Par.

⁴² Resultados do teste Tukey para o modelo (GRAU DE DIFERENÇA ATRIBUÍDO~ PAR*SEXO/GÊNERO)

Figura 4.6 – Média da Diferença Atribuída por Homens e Mulheres Paulistanos e Não Paulistanos

A Figura 4.7 compara as respostas de homens e mulheres nas duas amostras. As barras vermelhas dizem respeito às respostas das mulheres e as azuis, dos homens. As barras mais escuras correspondem às médias na amostra geral e as mais claras, na amostra apenas de paulistanos.

Figura 4.7 – Médias dos Graus de Diferença Atribuídos para cada Par por Homens e Mulheres (na amostra geral e apenas paulistanos)

Em comparação aos homens na amostra geral, os paulistanos atribuíram notas mais baixas aos pares 2xA e 3xA, de forma que não diferenciam as vibrantes nesse contexto ($3x_{AMP}=2x_{AMP}$). Na amostra paulistana, os homens também não diferenciaram as aproximantes quando em contraste com qualquer uma das vibrantes ($1x_{AMP}=1x_{RMP}$, $2x_{AMP}=2x_{RMP}$ e $3x_{AMP}=3x_{RMP}$). Ou seja, os homens paulistanos não atribuíram diferenças nem ao número de batidas da língua nos alvéolos, nem aos graus de retroflexão, mas apenas aos “tipos” de variantes. Portanto, as diferenças entre as percepções dos homens se correlacionam à Região de Nascimento, e a provável saliência da vibrante 3 e do retroflexo para aqueles que nasceram em outras comunidades.

As respostas das mulheres nas duas amostras não demonstraram diferenças significativas, sendo que em ambas elas não diferenciam tepe e vibrante 2, mas diferenciam a vibrante 3 de ambas ($3x_{AF}>2x_{AF}$, $3x_{RF}>2x_{RF}$). Também seguindo a tendência da amostra geral, elas atribuem maiores notas aos pares com retroflexo, do que com a aproximante ($1x_{AFP}>1x_{RFP}$ e $2x_{AFP}>2x_{RFP}$).

4.6 Síntese

No seu conjunto, estas análises revelam que a única variável linguística selecionada como significativa para a variação nas respostas dos ouvintes é o Par, que corresponde ao próprio contraste entre as variantes de (-r) e abarca diferenças fonético-fonológicas de modos de articulação (vibrantes e aproximantes), número de batidas da língua nos alvéolos (3, 2 e 1) e graus de retroflexão (A e R). As respostas em relação ao tepe e à vibrante 2 (quando comparadas às aproximantes) não apresentaram diferenças, sendo possível aferir que são elas reconhecidas como “quase a mesma coisa” em São Paulo, enquanto a vibrante 3 é mais saliente em comparação às duas. Já em relação às aproximantes alveolar e retroflexa (que também se diferenciam acusticamente de

maneira sutil), alguns grupos de informantes não atribuíram diferenças a elas (em contraste às vibrantes) e outros, sim.

O modo como os ouvintes percebem e atribuem as diferenças varia a depender de seu repertório sociolinguístico: as variantes linguísticas com que têm mais contato (o que se correlaciona com a Região de Nascimento e Local de Residência na cidade), os grupos sociais aos quais se filia (nesse caso, homens e mulheres), os construtos sociais a que tem acesso, como os valores sociais das variantes na comunidade, bem como a avaliação que fazem a respeito de sua própria fala.

Conclusão

Esta pesquisa de mestrado verificou como as características sociais dos ouvintes, bem como o significado social das variantes linguísticas, afetam o grau de diferença acústica atribuída a cinco variantes de (-r) em coda: (i) vibrante com três batidas, (ii) vibrante com duas batidas, (iii) tepe (vibrante simples), (iv) aproximante alveolar e (v) aproximante retroflexa.

Em um experimento desenvolvido com o objetivo de medir suas percepções, 109 moradores da cidade de São Paulo foram convidados a atribuir diferenças a pares formados pelo mesmo item lexical e que contrastavam duas pronúncias de /-r/ diferentes: por exemplo, a palavra *virtude* primeiro pronunciada com o tepe e depois com a aproximante alveolar. Para cada item lexical foram formados 10 pares (todas as combinações possíveis entre as cinco variantes), para os quais os participantes deveriam responder em uma escala contínua que ia de “Igual” a “Muito diferente” (definida por valores de 0 a 100 que eram invisíveis a eles) o quanto julgavam as pronúncias parecidas entre si.

De modo a garantir que as percepções recaíssem sobre a pronúncia do /-r/, os estímulos foram manipulados digitalmente com o auxílio dos programas Praat (BOERSMA; WEENINK, 2016) e Audacity, com base na metodologia desenvolvida por Campbell-Kibler (2006, 2007, 2009). Para cada item lexical, escolheu-se uma pronúncia-base (p.e., para o substantivo *pardal* foi escolhida aquela pronunciada com o tepe) sobre a qual foram “coladas” as outras realizações de /-r/. Dessa forma, excluíram-se diferenças de entonação, ênfase em determinada sílaba, ou outras idiossincrasias que pudessem chamar mais a atenção do ouvinte do que a própria diferença entre as variantes. Além disso, estabeleceram-se padrões de comparação entre elas, baseados nas estruturas dos formantes (F1, F2 e F3), bem como na análise do sinal acústico, de modo a se

garantir que todas as variantes apresentadas tivessem sido pronunciadas igualmente.

O experimento foi aplicado por meio de uma plataforma de pesquisas *online*, e divulgado pela internet, com o objetivo de alcançar um maior número possível de participantes, de grupos sociais distintos. A partir de suas respostas para cada par de estímulos, foi obtido um valor numérico (de 0 a 100), que corresponde ao grau de diferença que eles atribuem a cada par de variantes.

A partir desses valores, foram realizadas análises multivariadas na plataforma R (R Core Team, 2016), que objetivavam investigar as correlações entre a variação nas atribuições de diferença fonética, por um lado, e variáveis sociais e linguísticas, por outro. As variáveis linguísticas analisadas dizem respeito aos critérios que definiram as escolhas dos estímulos: Tonicidade da sílaba com /-r/ (átona ou tônica); Vogal Precedente (as sete vogais orais do PB); Consoante Seguinte (/t/ e /d/); e o Par, que é o contraste entre as próprias variantes e inclui diferenças de modos de articulação (vibrantes e aproximantes), número de batidas da língua nos alvéolos (3, 2 e 1) e graus de retroflexão (aproximante e retroflexo). Já as variáveis sociais dizem respeito às características do ouvinte: seu Sexo/Gênero, sua Região de Nascimento, a Região de Nascimento dos seus Pais, seu Local de Residência na cidade, bem como as Variantes que ele Reconhece em sua Própria Fala.

As respostas dos ouvintes para a variável Par, quando analisadas separadamente (ou seja, sem se verificar suas correlações com outras variáveis linguísticas ou sociais), mostraram que o retroflexo por um lado, e a vibrante 3, por outro, são as variantes de (-r) mais salientes em São Paulo, de um modo geral. Já o tepe e a vibrante 2 (quando comparadas às aproximantes) obtiveram o mesmo valor. As aproximantes alveolar e retroflexa se diferenciam quando comparadas ao tepe, mas não quando à vibrante três.

As análises multivariadas revelaram que as variáveis positivamente correlacionadas à variação foram o Par e a Região de Nascimento, o Local de Residência, a Variante que o ele reconhece em sua própria Fala e o Sexo/Gênero (todas em interação com o Par).

Para a Região de Nascimento, os ouvintes foram agrupados de acordo com as variantes de (-r) proeminentes em sua região de origem, o que revelou como a saliência de determinadas variantes influencia o modo como os grupos atribuem diferenças entre elas. Os participantes que nasceram na “região das fricativas” (Nordeste, Rio de Janeiro, norte de Minas Gerais e Brasília) atribuíram maior diferença ao par 3x1 (o contraste entre a vibrante 3 e o tepe) do que os outros grupos. Os que nasceram na “região do retroflexo” (interior de São Paulo, Paraná, Goiás e sul de Minas Gerais) por outro lado, foram os menos diferenciaram as aproximantes (AxR).

Já a variável Variante que Reconhece em sua própria Fala revelou as relações entre o julgamento que o ouvinte faz a respeito do seu modo de falar e a forma como considera as diferenças acústicas entre as variantes. Por ter forte correlação com a Região de Nascimento, foram avaliadas apenas as respostas dos ouvintes paulistanos para essa variável. Os informantes que disseram perceber apenas aproximantes em sua fala não diferenciaram o número de batidas nos alvéolos nem os graus de retroflexão das variantes. O grupo que disse pronunciar tanto aproximantes quanto vibrantes foi o que atribuiu as menores diferenças aos pares que contrastavam esses dois tipos de variantes e, ao mesmo tempo, o que atribuiu maior diferença a vibrante 3 *vs.* tepe. Esse subgrupo também considerou uma diferença maior entre tepe *vs.* retroflexo (1xR) do que entre tepe *vs.* aproximante (1xA), assim como os participantes que dizem pronunciar apenas vibrantes. Esse subgrupo, por sua vez, foi o que se mostrou mais atento às diferenças entre os pares que contrastavam aproximantes e vibrantes, ao mesmo tempo em que foi o que atribuiu uma menor nota ao par 3x1.

Embora o Local de Residência na cidade apresente importante correlação com a produção do (-r) em São Paulo (MENDES; OUSHIRO, 2011; OUSHIRO, 2015), essa variável “perdeu” sua significância no modelo de efeitos mistos quando se considerou apenas as respostas dos paulistanos. Ao mesmo tempo, a Variante que Reconhece em sua Fala continuou sendo selecionada como significativa, o que demonstrou que a importância do Local de Residência para a percepção fonética dos ouvintes consiste na ligação indireta entre o lugar que ele mora e as variantes que mais costuma ouvir, bem como as considerações que faz a respeito do seu modo de falar.

De modo geral, as mulheres não atribuíram diferenças entre o tepe e a vibrante 2, mas consideraram a vibrante 3 diferente delas. Ao mesmo tempo, consideraram o contraste entre tepe *vs.* retroflexo maior do que tepe *vs.* aproximante alveolar, corroborando com a hipótese de que as duas últimas se assemelhem por um certo “grau de paulistanidade” e gozem de relativo prestígio na comunidade, enquanto o retroflexo seria considerado um modo de falar mais “caipira”. Já as percepções dos homens mudaram também a depender de sua região de nascimento. Os paulistanos só consideraram as diferenças entre vibrantes de um lado, e aproximantes de outro, enquanto na amostra que inclui todos os participantes, os homens também atribuíram uma diferença maior à vibrante 3 relativamente ao tepe e à vibrante 2. Nas duas amostras, os homens não diferenciaram as aproximantes entre si.

Em suma, esta pesquisa de mestrado ajuda a esclarecer quais aspectos das experiências sociais dos ouvintes, ligadas a sua posição na comunidade, aos estereótipos e ideologias a que tem acesso, vinculam-se ao modo como são atribuídas as diferenças fonéticas entre as variantes de (-r). Uma vez que a percepção é um fenômeno complexo e multifacetado, as análises aqui apresentadas apontam para um caminho que ainda tem muito a ser explorado.

Referências Bibliográficas

ABAURRE, M. B. M.; SANDALO, M F.S. Os róticos revisitados. In: DA HORA, D.; COLLISCHON, G. (Orgs.) **Teoria Linguística: fonologia e outros temas**. João Pessoa/Editora da UFPB, 2003.

AMARAL, Amadeu. **O dialeto caipira**. (Domínio Público), 1920.

BARANOWSKI, M. Sociophonetics. In BAYLEY, R.; CAMERON, R.; LUCAS, C. (eds.) **The Oxford Handbook of Sociolinguistics**. Oxford: Oxford University Press, 2013, p. 403-424.

BELL, A. Language style as audience design. **Language in Society** vol. I3, p.145-204, 1984.

BIELER DA SILVA, M.E. **Entre duas metrópoles: (-r) em Itanhandu**. Dissertação (Mestrado em Linguística). Universidade de São Paulo, 2015.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. "Evaluational reactions to spoken languages." Último acesso em 20 jul. 2016. <http://www.fon.hum.uva.nl/praat/>

BRANDAO, S. Nas trilhas do -R retroflexo. **Signum: Estudos Linguísticos**, Londrina, n.10(2), p.265-283, 2007

BRESCANCINI, C.; MONARETTO, V. N. O. Os róticos no sul do Brasil: panorama e generalizações. **Signum: Estudos Linguísticos**, Londrina, n.11(2), p.51-66, 2008.

BUENO, L. F. **Os róticos do português falado em Brasília por crianças de 03 a 07 anos de idade**. Dissertação (Mestrado em Linguística). Brasília: Universidade de Brasília, 2013.

BYBEE, J. **Phonology and Language Use**. Cambridge: Cambridge University Press, 2001.

CALLOU, D., LEITE, Y.; MORAES, J. Processo(s) de enfraquecimento consonantal no português do Brasil. In: ABAURRE, B.; RODRIGUES, A. (Eds.), **Gramática do português falado: novos estudos descritivos**, vol. VIII. Campinas: Editora da Unicamp, 2002, p. 537-555.

_____. Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil. In: KOCH, I. (ed.). **Gramática do Português Falado**. Campinas: Unicamp, 1996.

_____. Apagamento do R final no dialeto carioca: um estudo em tempo aparente e tempo real. **DELTA [online]**, vol. 14, 00–00, 1998.

CAMPBELL-KIBLER, K. Accent, (ING), and the social logic of listener perceptions. **American Speech**, vol. 82, p. 32-61, 2007.

_____. **Listener perceptions of sociolinguistic variables: the case of (ing)**. Tese de Doutorado. Stanford University, 2006.

_____. The nature of sociolinguistic perception. **Language Variation and Change**, vol. 21, p. 135-156, 2009.

CHAMBERS, J.; TRUDGILL, P. **Dialectology**. Cambridge: Cambridge University Press, 1980.

CHAMBERS, J.K. **Sociolinguistic Theory. Linguistic variation and its social significance**. Oxford: Blackwell, 1995.

CHOMSKY, N.; HALLE, M. **The sound pattern of English**. Nova York, 1968.

CLEARY, M.; PISONI, D. B. Speech Perception and Spoken Word Recognition: Research and Theory. In: GOLDSTEIN, B. E. (ed.) **Blackwell Handbook of Sensation and Perception**. Oxford: Blackwell Publishing, 2001, p. 499-534.

CLEMENTE, F. C.; NISHIDA, G. Características acústicas do TAP em coda: Dados do português de Curitiba e do espanhol de Buenos Aires. **Revista Letras**. N. 73, p. 73-88, 2007.

CLEMENTS, G.N. The geometry of phonological features. **Phonology Yearbook** 2, p.225-252, 1985.

CLOPPER, C. G.; PISONI, D. B. Perception of Dialect Variation. In: PISONI, D. B.; REMEZ, R. E. (eds.) **The Handbook of Speech Perception** Oxford: Blackwell Publishing, 2008, p. 313-337.

DRAGER, K. Sociophonetic variation in speech perception. **Language and Linguistics Compass**. Vol. 4(7), p.473-480, 2010.

ECKERT, P. **Linguistic Variation as Social Practice**. Blackwell Publishers, 2000.

_____. Three waves of variation study: the emergence of meaning in the study of variation. **Annual Review of Anthropology**, vol. 41, p. 87–100, 2012.

_____. Variation and the indexical field. **Journal of Sociolinguistics**, vol. 12(4), 453-478, 2008.

ECKERT, P.; RICKFORD, J. Introduction. In: ECKERT, P.; RICKFORD, J. **Style and the sociolinguistic variation**. Cambridge: Cambridge University Press, 2001, p.1-18.

FOULKES, P.; SCOBIE, J. M.; WATT, D. Sociophonetics. In: HARDCASTLE, W., LAVER, J.; GIBBON, F. (eds.) **Handbook of Phonetic Sciences**. Oxford: Blackwell, 2010, p. 703-754.

HAY, J.; DRAGER, K. Sociophonetics. **Annu. Rev. Anthropol.** Vol. 36, p. 89-103, 2007.

HAY, J.; DRAGER, K.; THOMAS, B. Using nonsense words to investigate vogal merger. **English Language and Linguistics**. Vol. 17(2), p. 241-269, 2013.

JAFFE, A. **Stance**. Oxford: Oxford University Press, 2009.

JEKOSCH, U. **Voice and Speech Quality Perception**. Assessment and Evaluation. New York: Springer, 2005.

LABOV, W. **Padrões Sociolinguísticos**. São Paulo: Editora Parábola, 2008 [1972]. Tradução Marcos Bagno, Maria Marta Pereira Scherre e Caroline R. Cardoso.

_____. The anatomy of style-shifting. In: ECKERT, P.; RICKFORD, J. **Style and the sociolinguistic variation**. Cambridge: Cambridge University Press, 2001, p. 85-108.

_____. **The social stratification of English in New York City**. São Paulo: Cambridge University Press, 2006 [1966].

LABOV, W.; ASH, S.; RAVINDRANATH, M.; WELDON, T.; BARANOWSKI, M.; NAGY, N. Properties of the sociolinguistic monitor. **Journal of Sociolinguistics** vol. 15(4), p. 431-463. 2011

_____. Listeners' sensitivity to the frequency of sociolinguistic variables. **University of Pennsylvania Working Papers in Linguistics**, vol. 12(2), p. 105-129, 2006

LADEFOGED, P.; BROADBENT, D.E. Information Conveyed by Vowels. **The journal of the acoustical society of America** vol. 29(1), p.98-104, 1957.

LADEFOGED, P.; MADDIESON, I. **The sounds of the world's languages**. Cambridge: Blackwell, 1996.

LEITE, C. M. B. **Atitudes linguísticas: a variante retroflexa em foco**. Dissertação (Mestrado em Linguística). Campinas: Unicamp, 2004.

_____. **O /R/ em posição de coda silábica no falar campineiro**. 201f. Tese (Doutorado em Linguística) Unicamp, Campinas. 2010.

LEVON, E. Sexuality in context: Variation and the sociolinguistic perception of identity. **Language in Society**, vol. 36, p. 533-554, 2007.

LEVON, E.; FOX, S. Social Salience and the Sociolinguistic Monitor: A Case Study of ING and TH-fronting in Britain. **Journal of English Linguistics**, vol. 42(3), p.185-216, 2014.

LUCENTE, L. **Ainda sobre os róticos no PB: o tepe e o R forte**. Apresentação de trabalho/comunicação, 2008.

MASSARO, D. W. Speech Perception. In: SMELSER, N.M.; BALTES, P.B. (eds.); KINTSCH, W. (section ed.), **International Encyclopaedia of Social and Behavioural Sciences**. Amsterdam: Elsevier, 2001, p. 14870-14875.

MENDES, R. B. **Sounding Paulistano: Variation and correlation in São Paulo**. Trabalho apresentado no NWAV39, San Antonio, Texas, 2010.

_____. Gendered perceptions of noun agreement in Brazilian Portuguese. **Revista Internacional de Lingüística Iberoamericana (RILI)** 23(1), p.93-108, 2014.

MENDES, R. B.; OUSHIRO, L. A pronúncia retroflexa do /-r/ na fala paulistana. In: HORA, D.; NEGRÃO, E. V. (Eds.), **Estudos da Linguagem. Casamento entre temas e perspectivas**. João Pessoa: Ideia, 2011, p. 282-299.

_____. Percepções sociolinguísticas sobre as variantes tepe e retroflexa na cidade de São Paulo. In: HORA, D.; NEGRÃO, E. V. (Eds.), **Estudos da Linguagem**. Casamento entre temas e perspectivas. João Pessoa: Ideia, 2011b, p. 229–245.

MENDOZA-DENTON, N. Language and identity. In CHAMBERS, T.; SCHILLING-ESTES, N. **The Handbook of Language Variation and Change**. Cambridge: Blackwell, 2002, p. 475–499.

NIEDZIELSKI, N. The effect of social information on the perception of sociolinguistic variables. **Journal of Language and Social Psychology**, vol. 18 n.1, p. 62–85, 1999.

OCHS, E. Constructing Social Identity: A Language Socialization Perspective. **Research on Language and Social Interaction**, vol. 26(3), p. 287–306, 1993.

OLIVEIRA, J. M. **O apagamento do /R/ implosivo na norma culta de Salvador**. Dissertação (Mestrado em Letras). Universidade Federal da Bahia, Salvador – BA. 1999.

OUSHIRO, L. **Identidade na pluralidade. Avaliação, produção e percepção linguística na cidade de São Paulo**. Tese (Doutorado em Linguística). Universidade de São Paulo, 2015.

PEIRCE, JW. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2:10. doi:10.3389/neuro.11.010.2008, 2009.

_____. PsychoPy - Psychophysics software in Python. *J Neurosci Methods*, 162(1-2):8-13, 2007.

PIERREHUMBERT, J. Exemplar dynamics: Word frequency, lenition, and contrast. In BYBEE, J.; HOPPER, P. (eds.). **Frequency effects and the emergence of lexical structure**. Amsterdam: John Benjamins, 2001, p. 137–157.

PIMENTEL, R. M. **A variação linguística do fonema /r/ na posição pós-vocálica**. 105f. Dissertação (Mestrado em Letras). Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS. 2003.

PODESVA, R.J. Phonation type as a stylistic variable: the use of falsetto in constructing a persona. **Journal of Sociolinguistics**, vol. 11(4), p. 478–504, 2007.

PRESTON, D. Five Visions of America. **Language in Society**, vol. 15, p. 221–240, 1986.

_____. The power of language regard: Discrimination, classification, comprehension and production. **Dialectologia**. Special Issue II, p. 9-33, 2011.

_____. Variation in language regard. In: GILLES, P.; SCHARLOTH, J.; ZEIGLER, E. (eds.), **Empirische Evidenzen und theoretische Passungen sprachlicher Variation**. Frankfurt: Peter Lang, 2010, p. 7-27.

R CORE TEAM. R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria, 2016. Último acesso em 20 jul. 2016. <http://www.R-project.org/>

REETZ, H.; JONGMAN, A. Acoustic characteristics of speech sound. In: REETZ, H.; JONGMAN, A. **Phonetics: Transcription, Production, Acoustics, and Perception**. Malden, MA: Blackwell, 2009, p.182-207.

RENNICKE, I. The retroflex r of Brazilian Portuguese: theories of origin and a case study of language attitudes in Minas Gerais. **Revista de Estudos Linguísticos da Universidade do Porto**. Vol. 6, n.1 p. 149-170, 2011.

ROSENBLUM, L. D. Primacy of Multimodal Speech Perception. In: PISONI, D. B.; REMEZ, R. E. (eds.) **The Hadbook of Speech Perception**, Oxford: Blackwell Publishing, 2008, p. 51-78.

SILVA, A. P. H; CLEMENTE, F. C.; NISHIDA, G. Para a representação dinâmica do tap em grupos e codas: evidências acústicas. **Revista Virtual de Estudos da Linguagem – ReVEL**. V. 4, n. 7, 2006.

SORIANO, L. G. M. Percepções Sociolinguísticas: o /-r/ em São Paulo. Trabalho apresentado no 3°CIDS, Londrina, 2014.

SKEETE, N. A. O uso variável da vibrante na cidade de João Pessoa. **Graphos**, vol. 1, n. 2, p. 77-96, 1997.

THOMAS, E. R. Sociophonetics applications of speech perception experiments. **American Speech**, vol. 77, p. 115-147, 2002.

VAN HOUT, R.; VAN de VELDE, H. Patterns of /r/ variation. In: VAN de VELDE, H.; VAN H. R. (Eds.), **'r-atics: Sociolinguistic, phonetic and phonological characteristics of /r/**, vol. Études & Travaux. Bruxelas: Université Libre de Bruxelles, 2011, p. 1-9.