UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA PORTUGUESA

SELMO RIBEIRO FIGUEIREDO JUNIOR

ATLAS LINGUÍSTICO PLURIDIMENSIONAL DO PORTUGUÊS PAULISTA

níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê

VERSÃO CORRIGIDA

TOMO I

Prolegômenos e análises

SÃO PAULO
2019

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LETRAS CLÁSSICAS E VERNÁCULAS
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOLOGIA E LÍNGUA PORTUGUESA

SELMO RIBEIRO FIGUEIREDO JUNIOR

ATLAS LINGUÍSTICO PLURIDIMENSIONAL DO PORTUGUÊS PAULISTA

níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê

VERSÃO CORRIGIDA

TOMO I

Prolegômenos e análises

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Filologia e Língua Portuguesa do Departamento de Letras Clássicas e Vernáculas, da Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, como parte dos requisitos para obtenção do título de Doutor em Letras.

Filologia e Língua Portuguesa
Área de concentração

Dr. Manoel Mourivaldo Santiago-Almeida
Orientador
Universidade de São Paulo

Dr. Harald Thun
Orientador no estágio sanduíche
Christian-Albrechts-Universität zu Kiel

SÃO PAULO

2019

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

Catalogação na Publicação
Serviço de Biblioteca e Documentação

F475a Figueiredo Junior, Selmo Ribeiro

Atlas linguístico pluridimensional do português paulista: níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê. 2018. 6 t. / Selmo Ribeiro Figueiredo Junior; orientador: Manoel Mourivaldo Santiago-Almeida; orientador no estágio sanduíche: Harald Thun. São Paulo, 2019.

2.120 pp.

Tese (Doutorado). Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Letras Clássicas e Vernáculas.

Área de concentração: Filologia e Língua Portuguesa.

1. Dialetologia Pluridimensional. 2. Geolinguística. 3. Português paulista.
4. Dialeto caipira. 5. Arquifonema /R/ em coda silábica.
I. Santiago-Almeida, Manoel Mourivaldo, orient. II. Título.

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS

ENTREGA DO EXEMPLAR CORRIGIDO DA TESE

Termo de Ciência e Concordância do orientador

Nome do aluno: Selmo Ribeiro Figueiredo Junior

Data da defesa: 30/11/2018

Nome do orientador: Dr. Manoel Mourivaldo Santiago-Almeida

Nos termos da legislação vigente, declaro **ESTAR CIENTE** do conteúdo deste **EXEMPLAR CORRIGIDO**, elaborado em atenção às sugestões dos membros da Comissão Julgadora na sessão de Defesa Púbica da Tese de Doutorado, manifestando-me plenamente favorável ao seu encaminhamento e publicação no Portal Digital de Teses da USP.

São Paulo, 25/06/2019.

Dr. Manoel Mourivaldo Santiago-Almeida

Orientador

Universidade de São Paulo

ATA DE DEFESA

Ata de defesa pública de Tese do(a) Senhor(a) Selmo Ribeiro Figueiredo Junior no Programa: Filologia e Língua Portuguesa, do(a) Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

Aos 30 dias do mês de novembro de 2018, no(a) Sala de Defesas realizou-se a Defesa da Tese do(a) Senhor(a) Selmo Ribeiro Figueiredo Junior, apresentada para a obtenção do título de Doutor intitulada:

"Atlas linguístico pluridimensional do português paulista: níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê"

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Nome dos Participantes da Banca	Função	Sigla da CPG	Resultado
Manoel Mourivaldo Santiago Almeida	Presidente	FFLCH - USP	Aprovado
Vanderci de Andrade Aguilera	Titular	UEL - Externo	Aprovado
Emilio Gozze Pagotto	Titular	UNICAMP - Externo	Aprovado
Neusa Ines Philippsen	Titular	FUNEMAT - Externo	Aprovado
Harald Thun	Titular	CAU - Externo	Aprovado

Resultado Final: Aprovado

Parecer da Comissão Julgadora *

A banca fez diversas considerações de natureza formal, teórica e metodológica a ser incorporadas na versão corrigida e julgou o trabalho relevante para a área, aprovando-o. O Prof.Dr. Harald Thun participou por meio de videoconferência.

Eu, Daiane Neres da Silva _____, lavrei a presente ata , que assino juntamente com os(as) Senhores(as) examinadores. São Paulo, aos 30 dias do mês de novembro de 2018.

Vanderci de Andrade Aguilera

Emilio Gozze Pagotto

Neusa Ines Philippsen

Harald Thun

Manoel Mourivaldo Santiago Almeida
Presidente da Comissão Julgadora

* Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

A defesa foi homologada pela Comissão de Pós-Graduação em 30/11/2018 e, portanto, o(a) aluno(a) faz jus ao título de Doutor em Letras obtido no Programa Filologia e Língua Portuguesa.

Presidente da Comissão de Pós-Graduação
Prof. Dr. Edelcio Gonçalves de Souza
Presidente da CPG
FFLCH-USP

*Ao Santiago Villela Marques (*1.m.*),*
escritor, professor, sábio, erudito, conselheiro, ético, humilde, amigo.

Aos 80 informantes paulistas voluntários
participantes da pesquisa de que se trata nesta obra.

Aos defensores da Educação, da Ciência e da Tecnologia.
Aos defensores, portanto, do avanço dos indivíduos e das sociedades.

AGRADECIMENTOS

Aos queridos Professores Doutores Manoel Mourivaldo Almeida-Santiago (Universidade de São Paulo – USP) e Harald Thun (Universidade de Kiel, orig. *Christian-Albrechts-Universität zu Kiel* – CAU), pela erudição e pela sabedoria compartilhadas ao longo da orientação, em razão das quais muito os admiro.

Às caras Professoras Doutoras Vanderci de Andrade Aguilera (Universidade Estadual de Londrina – UEL) e Neusa Inês Philippsen (Universidade do Estado de Mato Grosso – UNEMAT/Sinop), pelo rigor e competência na avaliação do trabalho e pelas importantes contribuições, tanto na ocasião da Banca de Qualificação, quanto na circunstância da Banca de Defesa Pública; bem como ao caro Professor Doutor Emilio Gozze Pagotto (Universidade Estadual de Campinas – UNICAMP), quem gentilmente integrou a última banca e, também com rigor e competência, avaliou a tese e concedeu significativas contribuições. Outrossim, aos Professores Doutores Américo Venâncio Lopes Machado Filho (Universidade Federal da Bahia – UFBA) e Silvia Figueiredo Brandão (Universidade Federal do Rio de Janeiro – UFRJ), pela cortês disponibilidade à suplência da Banca de Defesa Pública.

Aos estimados a seguir, que afavelmente me ajudaram a encontrar informantes para a pesquisa no interior paulista: Izes Bastianon Chaves Oliveira e Wagner Bastianon Chaves de Oliveira (Santana de Parnaíba); Morgana Ribeiro, Milene de Fátima Silva Fernandes Rossi, Jair de Oliveira e Olga Susana Costa Coito e Araújo (Itu); Maria de Lourdes Kerche do Amaral, Romeu Castelucci Neto, Juliana Moura Gutierrez e Raquel Antunes Galvão (Porto Feliz); Henrique Albiero Pazetti e Júlio Barros (Sorocaba); Ângelo Rosivaldo Stenico, Miriam Alves Pagotto, Ramon Pagotto e Anicleide Zequini (Capivari); Diego Fillipe Carvalho de Oliveira e Lucas Mathiazo Dante (Pirapora do Bom Jesus); Marina Azevedo Ribeiro, Gema Francisca Masetto Alonso, José Carlos Dias Bastos, Sílvia Maria Lopes de Mello e Célia Bastos (São Roque); Vandenei Dogado (Araçariguama); Renata Graziela Duarte Gava (Piracicaba); e Maria José Dal Bo Giovaneti Polastre, Mauro Beneton Alexandre Junior, Arrovani Luiz Fonseca e Pedro Henrique Macerani (Tietê).

À Comissão Coordenadora do Programa de Pós-Graduação em Filologia e Língua Portuguesa e à Secretaria do Departamento de Letras Clássicas e Vernáculas (DLCV) da Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) da USP, bem como ao Setor de Romanística (orig. *Romanisches Seminar*) da CAU na Alemanha, pelo suporte e simpatia, e, por extensão, a todos os profissionais de ambas as instituições.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), ao Serviço Alemão de Intercâmbio Acadêmico (orig. *Deutscher Akademischer Austauschdienst* – DAAD) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelos importantes apoios:

– à FAPESP, pela bolsa nacional de doutorado, processo nº 15/14038-5, e pelas diárias para a pesquisa de campo, processo nº 11/51787-5, no âmbito do Projeto de História do Português Paulista (PHPP – Projeto Caipira);

– ao DAAD, pela bolsa de doutorado sanduíche de dois anos na Alemanha, programa nº 57214225; e

– à CAPES, pelos auxílios de viagens, processo nº 0128-16-3 / 99999.000128/2016-03), em parceria com o DAAD.

Last, but not least, aos queridos colegas e amigos a seguir, que contribuíram diretamente com esta obra, através de críticas importantes, discussões dialéticas e sugestões valiosas de melhorias: Mario Ruiz Moreno (Universidade Johannes Gutenberg Mainz, orig. *Johannes Gutenberg-Universität Mainz* – JGU); Fernando Hélio Tavares de Barros (Universidade de Bremen, orig. *Universität Bremen* – UB); Lea Kerstin Paulsen (Universidade de Kiel, orig. *Christian-Albrechts-Universität zu Kiel* – CAU); Leopold Szierbowski (Universidade Humboldt de Berlim, orig. *Humboldt-Universität zu Berlin* – HU); Anni Paulsen (Escola Insular Fehmarn, orig. *Inselschule Fehmarn*); Paulo Barroso (Universidade de São Paulo – USP); Odonchimeg Tsevegmid (orig. *Одончимэг Цэвэгмид*, Universidade de Potsdam, orig. *Universität Potsdam* – UP); e Joachim Steffen (Universidade de Augsburg, orig. *Universität Augsburg* – UNIA).

A VOCÊS TODOS, MUITO OBRIGADO POR TUDO!

RESUMO

FIGUEIREDO JUNIOR, Selmo Ribeiro. *Atlas linguístico pluridimensional do português paulista:* níveis semântico-lexical e fonético-fonológico do vernáculo da região do Médio Tietê. 2018. 6 t. Tese (Doutorado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2019.

Com base teórico-metodológica da Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*), esta tese de doutorado apresenta sobretudo um atlas linguístico pluridimensional do português paulista (conhecido como "dialeto caipira") nos níveis semântico-lexical e fonético-fonológico e faz contribuições metodológicas de diferentes naturezas. Trata-se de contribuições à metodologia de coleta de dados empíricos, à metodologia de organização de *corpora,* e à metodologia de elaboração de atlas linguísticos pluridimensionais. No campo analítico, uma das contribuições deste estudo refere-se às contrapartes (a)fonéticas do arquifonema /R/ em quatro situações de coda silábica: (i) externa de desinência infinitiva (DI) seguida por pausa; (ii) externa em sílaba tônica (que não de DI) seguida por pausa; (iii) interna em sílaba tônica; e (iv) interna em sílaba átona. Quanto à situação (i), a seguinte hipótese em específico foi testada: os jovens entre 18 e 36 anos de idade do Médio Tietê (esp. aqueles com alta escolaridade, AE) estão produzindo a variante [ɻ] (a aproximante retroflexa, conhecida como "erre caipira") para o /R/ na situação (i) sob uma frequência relativa inferior àquela a ser observada entre os mais velhos a partir de 55 anos de idade (esp. aqueles com baixa escolaridade, BE). Como resultado, a fala dos jovens revelou-se na verdade mais albergadora da variante [ɻ] do que a fala dos mais velhos. A frequência relativa do fone referido em perspectiva com suas covariantes foi de 57% entre os jovens com AE e 62% entre os jovens com BE (contra 45% entre os mais velhos com AE e 30% entre os mais velhos com BE). Os dados são de 80 voluntários moradores em alguns dos municípios mais antigos do interior paulista: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari e Piracicaba. A coleta dos dados realizou-se com a aplicação dos instrumentos metodológicos do Atlas Linguístico do Brasil (ALiB, 2014) com modificações, gerando um conjunto de *corpora*, constituído por *corpora* semântico-lexicais, *corpora* fonético-fonológicos, um *corpus* metalinguístico-etnográfico e *corpora* extralinguísticos. Como base de dados, esse conjunto de *corpora* permitiu a elaboração do atlas linguístico de maneira a contemplar, além da dimensão diatópica inerente a esse empreendimento cartográfico e geolinguístico, ainda as dimensões diastrática, diagenérica e diageracional. No campo teórico-metodológico, uma das contribuições mais importantes deste trabalho reside no desenvolvimento de uma série de procedimentos ao inquérito semântico-lexical a qual denominamos 'técnica de entrevista orientada à contundência responsiva/anuente', diretamente baseada na 'técnica de entrevista de três tempos', por sua vez concebida pelo Dr. Thun (ADDU, 2000). Esta pesquisa só foi possível graças ao apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (procs. 2015/14038-5 e 2011/51787-5), do Serviço Alemão de Intercâmbio Acadêmico – DAAD (prog. 57214225) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (proc. 0128-16-3 / 99999.000128/2016-03).

Palavras-chaves: Dialetologia Pluridimensional. Geolinguística. Português paulista. Dialeto caipira. Arquifonema /R/ em coda silábica.

ZUSAMMENFASSUNG

FIGUEIREDO JUNIOR, Selmo Ribeiro. *Pluridimensionaler Sprachatlas des im Bundesstaat São Paulo in Brasilien gesprochenen Portugiesischen:* das semantisch-lexikalische und das phonetisch-phonologische Niveau der Vernakularsprache der Region Médio Tietê. 2018. 6 Bde. Doktorarbeit – Fakultät für Philosophie, Sprach- und Geisteswissenschaften der Universität von São Paulo, 2019.

Auf der Basis der Theorie und Methodik der Pluridimensionalen Dialektologie (RADTKE u. THUN, 1996; THUN, 2000, 2005 *etc.*) wurde ein pluridimensionaler Sprachatlas des im Bundesstaat São Paulo in Brasilien gesprochenen Portugiesischen (alias „Caipira"-Dialekt) erstellt. Die Sprachebenen der Kartografie sind sowohl semantisch-lexikalisch als auch phonetisch-phonologisch. Diese Dissertation leistet auch methodologische Beiträge zur Datenerhebung, zur Gliederung von Korpora und zur Ausarbeitung von pluridimensionalen Sprachatlanten. In der Analyse wurden die phonetischen Gegenstücke zum Archiphonem /R/ in der Silbenkoda entweder am Ende oder in der Mitte von Wörtern ausgewertet: (i) am Ende als Infinitivendung (IE) gefolgt von einer Pause; (ii) am Ende in betonter Silbe (aber nicht als IE) gefolgt von einer Pause; (iii) in der Mitte in betonter Silbe; und (iv) in der Mitte in unbetonter Silbe. Bezüglich des ersten Falls, (i), stellte man eine spezifische Hypothese auf die Probe: Die Jugendlichen im Alter von 18 bis 36 Jahren in der Gegend Médio Tietê (insb. die mit hohem Bildungsgrad) sprechen die Variante [ɻ] (d.h., den stimmhaften retroflexen Approximant, auch bekannt als „Caipira R") als Gegenstück zum /R/ beim Fall (i) unter einer relativen Frequenz aus, die niedriger ist als die, die bei den Älteren im Alter ab 55 Jahren aufwärts (insb. den mit niedrigem Bildungsgrad) zu beobachten ist. Es stellte sich überraschend heraus, dass die genannte Variante eigentlich häufiger bei den Jugendlichen auftritt als bei den Älteren. Die relative Frequenz der Verwendung des Phons [ɻ] im Vergleich zu seinen Kovarianten lag bei 57% bei den Jugendlichen mit hohem Bildungsgrad und bei 62% bei den Jugendlichen mit niedrigem Bildungsgrad (gegenüber 45% bei den Älteren mit hohem Bildungsgrad und 30% bei den Älteren mit niedrigem Bildungsgrad). Die Daten haben 80 Freiwillige aus zehn der ältesten Städte im Landesinneren des Bundesstaates São Paulo vor Ort geliefert, nämlich Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari und Piracicaba. Die Datenerhebung erfolgte mithilfe der von uns modifizierten methodologischen Mittel des „Atlas Linguístico do Brasil" (ALiB, 2014) und daraus ist ein Satz von Korpora erzeugt worden: semantisch-lexikalische Korpora, phonetisch-phonologische Korpora, ein metalinguistisch-ethnographisches Korpus und extralinguistische Korpora. Dieser Satz bildete die Datenbank für die Erstellung des Atlasses, welcher — abgesehen von der diatopischen Dimension, die diesem kartographischen und geolinguistischen Unterfangen inhärent ist — diastratische, diageschlechtsspezifische und diagenerationelle Dimensionen umfasst. Ebenfalls trägt diese Doktorarbeit zur Methodik von semantisch-lexikalischen Interviews bei. Es wurde, basierend auf der „Technik in drei Phasen" von Thun (ADDU, 2000), eine Befragungstechnik entwickelt, die das Ziel hat, die Sicherheit der Antworten anzutreffen. Diese Forschung wurde dank folgender Förderorganisationen ermöglicht: „Fundação de Amparo à Pesquisa do Estado de São Paulo" – FAPESP (Verfahrensnummern 2015/14038-5 u. 2011/51787-5), Deutscher Akademischer Austauschdienst – DAAD (Förderprogrammnummer 57214225) und „Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" – CAPES (Verfahrensnummer 0128-16-3 / 99999.000128/2016-03).

Stichwörter: Pluridimensionale Dialektologie. Geolinguistik.
Portugiesisch des Bundesstaates São Paulo. „Caipira"-Dialekt.
Archiphonem /R/ in der Silbenkoda.

ABSTRACT

FIGUEIREDO JUNIOR, Selmo Ribeiro. *Pluridimensional linguistic atlas of the Portuguese spoken in the Brazilian state of São Paulo:* semantic-lexical and phonetic-phonological levels of the vernacular in Médio Tietê region. 2018. 6 vols. Doctoral Thesis – Faculty of Philosophy, Language Studies, and Human Sciences, University of São Paulo, 2019.

Following the theoretical and methodological background of the Pluridimensional Dialectology (RADTKE & THUN, 1996; THUN, 2000, 2005, among others), this doctoral thesis presents mainly a pluridimensional linguistic atlas of the Portuguese spoken in the Brazilian state of São Paulo, also known as "Caipira" dialect. The linguistic levels approached are both semantic-lexical and phonetic-phonological. This study also makes methodological contributions to data collection process, to *corpora* structuring, and to development of pluridimensional atlases. Regarding the analyses, the archiphoneme /R/'s phonetic counterparts were investigated in four different syllable codas: (i) final of infinitive ending (IE) followed by a break; (ii) final in stressed syllable (but not of IE) followed by a break; (iii) medial in stressed syllable; and (iv) medial in unstressed syllable. Concerning case (i), a specific hypothesis was tested out, namely: young individuals aged 18 to 36 from the Médio Tietê region (esp. those with high education, HE) articulate the variant [ɻ] (voiced retroflex approximant, also known as "Caipira-R") for the /R/ in case (i) under a relative frequency that is lower than the one to be observed among elders from 55 years of age (esp. those with low education, LE). Surprisingly, it turned out that younger ones speak [ɻ] more than older ones do. The relative frequency of [ɻ] put in perspective with its covariants was 57% among HE young individuals and 62% among LE ones (against 45% among HE elders and 30% among LE ones). Data from 80 local speakers were collected in some of the São Paulo State's oldest inland cities: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari, and Piracicaba. For the data collection, methodological instruments made by "Atlas Linguístico do Brasil" (ALiB, 2014) were applied after modifications. Thus, a set of *corpora* was produced: semantic-lexical ones, phonetic-phonological ones, a metalinguistic-ethnographic one, and extralinguistic ones. They constitute the database for the elaboration of the atlas, which encompasses — in addition to the diatopic dimension inherent in this cartographic and geolinguistic undertaking — diastratic, diagender-specific, and diagenerational dimensions. From a both theoretical and methodological perspective, another contribution of this work lies in the development of a series of procedures — whose main goal is to obtain the informants' certitude about answers in the semantic-lexical interview — that is derived from the three-step technique created by Dr. Thun (ADDU, 2000). This research was only possible thanks to "Fundação de Amparo à Pesquisa do Estado de São Paulo" – FAPESP (process numbers 2015/14038-5 and 2011/51787-5), German Academic Exchange Service – DAAD (funding program number 57214225), and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" – CAPES (process number 0128-16-3 / 99999.000128/2016-03).

Keywords: Pluridimensional Dialectology. Geolinguistics. São Paulo State's Portuguese. "Caipira" dialect. Archiphoneme /R/ in syllable coda.

(1) FIGURAS E (2) GRÁFICOS[1]

(1)

1 Sub-bacias hidrográficas paulistas (SÃO PAULO, 2016) 94
2 Partes de S1, S2 e S3 (ordem da esquerda para a direita) (exemplo) 100
3 Componentes das cartas linguísticas (exemplo do atlas semântico-lexical) 102
4 Recorte do P1 da Figura 3 105
5 Carta /R/ da série CxGy e especificidade 1 (carta miniaturizada) 142
6 Carta /R/ da série CxWz e especificidade 1 (carta miniaturizada) 144
7 Carta /R/ da série WzGy e especificidade 1 (carta miniaturizada) 147
8 Carta /R/ da série CxGy e especificidade 2 (carta miniaturizada) 149
9 Carta /R/ da série CxWz e especificidade 2 (carta miniaturizada) 151
10 Carta /R/ da série WzGy e especificidade 2 (carta miniaturizada) 153
11 Carta /R/ da série CxGy e especificidade 3 (carta miniaturizada) 155
12 Carta /R/ da série CxWz e especificidade 3 (carta miniaturizada) 157
13 Carta /R/ da série WzGy e especificidade 3 (carta miniaturizada) 159
14 Carta /R/ da série CxGy e especificidade 4 (carta miniaturizada) 161
15 Carta /R/ da série CxWz e especificidade 4 (carta miniaturizada) 163
16 Carta /R/ da série WzGy e especificidade 4 (carta miniaturizada) 165

(2)

1 Presença das mídias na vida dos 80 informantes 218
2 Formalidade e informalidade das mídias presentes na vida dos informantes 219-220

[1] Os elementos com autoria omissa são nossas criações. As cartas linguísticas em tamanho não miniaturizado receberão listas especiais, disponíveis nos tomos nos quais os atlas estarão contidos.

TABELAS[1]

Pesquisas empíricas de base oral com enfoque...

1 ... no nível fonético/fonológico do PP (seleção) 88
2 ... no nível morfológico/(morfos)sintático do PP (seleção) 89
3 ... em diversos níveis linguísticos do PP (seleção) 90

4 Dimensões e parâmetros de análise (com destaques) 96
5 Áreas semânticas do QSL^{m} e suas questões 118
6 Lógica da simbologia cartográfica (para cartas especiais, apenas o primeiro grau) 138
7 Exemplos da lógica da simbologia cartográfica estendida (para cartas especiais) 139

Base documental...

8 ... geral (com destaques) para as "Cartas /R/ E1" 180
9 ... (com destaques) da série CxGy para a "Carta /R/ E1" correspondente 181
10 ... (com destaques) da série CxWz para a "Carta /R/ E1" correspondente 182
11 ... (com destaques) da série WzGy para a "Carta /R/ E1" correspondente 183
12 ... geral (com destaques) para as "Cartas /R/ E2" 184
13 ... (com destaques) da série CxGy para a "Carta /R/ E2" correspondente 185
14 ... (com destaques) da série CxWz para a "Carta /R/ E2" correspondente 186
15 ... (com destaques) da série WzGy para a "Carta /R/ E2" correspondente 187
16 ... geral (com destaques) para as "Cartas /R/ E3" 188
17 ... (com destaques) da série CxGy para a "Carta /R/ E3" correspondente 189
18 ... (com destaques) da série CxWz para a "Carta /R/ E3" correspondente 190
19 ... (com destaques) da série WzGy para a "Carta /R/ E3" correspondente 191
20 ... geral (com destaques) para as "Cartas /R/ E4" 192
21 ... (com destaques) da série CxGy para a "Carta /R/ E4" correspondente 193
22 ... (com destaques) da série CxWz para a "Carta /R/ E4" correspondente 194
23 ... (com destaques) da série WzGy para a "Carta /R/ E4" correspondente 195

[1] Todas as tabelas são de nossa autoria.

24 Exemplos de respostas ao QSL^{m} 210
25 Processos psicossociais e seus efeitos possíveis 211
26 Categorização simplificada do que <beija-flor> denota 213
27 Classificações das opções midiáticas preferidas 220
28 Lógica da simbologia cartográfica [tab. no tomo do atlas semântico-lexical] 234
29 Lógica da simbologia cartográfica [tab. no tomo do atlas fonético-fonológico] 702
30 Questões do QSL^{m} em resposta às quais não houve normas absolutas 1762
31 Questões do QFF^{m} que obtiveram apenas um valor 1768
32 Questões do QFF^{m} em resposta às quais não houve normas absolutas 1877

SIGLAS, ACRÔNIMOS E ABREVIAÇÕES[1]

-SDI	n.o., sílaba que não seja de desinência infinitiva;
+SDI	n.o., sílaba de desinência infinitiva;
2(G)C	n.o., duas (ou dois grupos de) covariantes^{+};
3(G)C	n.o., três (grupos de) covariantes^{+};
4(G)C	n.o., quatro (grupos de) covariantes^{+};
AFF	atlas fonético^{+}-fonológico^{+};
ALiB	Atlas Linguístico do Brasil;
ASL	atlas semântico^{+}-lexical^{+};
aux.	auxílio;
C	consoante ou, consideradas suas características distributivas, semiconsoante (em simbologia fonológica^{+} categorial);
Ca	classe escolar alta (ou nível escolar alto). Neste trabalho, a partir de estudos universitários, completos ou incompletos;
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior;
CAU	Universidade de Kiel (orig. *Christian-Albrechts-Universität zu Kiel)*;
Cb	classe escolar baixa (ou nível escolar baixo). Neste trabalho, no escopo de abrangência entre analfabetismo e ensino médio incompleto;
cf.	confronte-se, confira-se;
CFF	carta fonético^{+}-fonológica^{+};
covar.	covariante^{+};
CSL	carta semântico^{+}-lexical^{+};
Cx	expressão abstrata a ser parametrizada concretamente ou por 'Ca', ou por 'Cb';

[1] Denotando 'nesta obra', a abreviação 'n.o.' se associa a (partes de) definições de termos — segundo nosso conhecimento — próprias deste trabalho; definições essas que, se não estiverem apresentadas aqui e se cujos termos forem seguidos pelo sinal de mais sobrescrito, $^{+}$, encontrar-se-ão na seção "Glossário", no Tomo VI.

DAAD Serviço Alemão de Intercâmbio Acadêmico (orig. *Deutscher Akademischer Austauschdienst);*

DC dialeto^{+} caipira^{+};

DI desinência infinitiva;

DO doutorado;

DP Dialetologia Pluridimensional;

e.g. por exemplo (orig. *exempli gratia*);

E1 n.o., especificidade 1, cuja definição é: 'coda externa de desinência infinitiva seguida por pausa';

E2 n.o., especificidade 2, cuja definição é: 'coda externa em sílaba tônica (a qual não seja, porém, de desinência infinitiva) seguida por pausa';

E3 n.o., especificidade 3, cuja definição é: 'coda interna em sílaba tônica';

E4 n.o., especificidade 4, cuja definição é: 'coda interna em sílaba átona';

FA n.o., forma sugerida aceita^{+};

FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo;

FdI Ficha do Informante em versão de 2001 da lavra do ALiB (2014);

FdI^{m} FdI modificada por nós;

FI n.o., forma resultante de insistência^{+};

FN n.o., forma espontânea negativa^{+};

FP n.o., forma espontânea positiva^{+};

FR n.o., forma sugerida rejeitada^{+};

FS n.o., forma semasiológica^{+};

GI grupo etário mais jovem (neste trabalho, entre 18 e 36 anos);

GII grupo etário mais velho (neste trabalho, a partir de 55 anos);

Gy expressão abstrata a ser parametrizada concretamente ou por 'GI', ou por 'GII';

Hes n.o., hesitação^{+};

i.e. isto é (orig. *id est*);

IC iniciação científica;

inf. informante não especificado;

inf*x* informante determinado, sendo *x* variável de especificação;

IPA Alfabeto Fonético Internacional (orig. *International Phonetic Alphabet)*;

inst. instituição;

MS Mato Grosso do Sul;

ME mestrado;

obs. observação, -ções;

orig. original(mente);

p. página;

P1 n.o., ponto 1 da rede de pontos^{+}, *i.e.*, município de Santana de Parnaíba;

P2 n.o., ponto 2 da rede de pontos^{+}, *i.e.*, município de Pirapora do Bom Jesus;

P3 n.o., ponto 3 da rede de pontos^{+}, *i.e.*, município de Araçariguama;

P4 n.o., ponto 4 da rede de pontos^{+}, *i.e.*, município de São Roque;

P5 n.o., ponto 5 da rede de pontos^{+}, *i.e.*, município de Sorocaba;

P6 n.o., ponto 6 da rede de pontos^{+}, *i.e.*, município de Itu;

P7 n.o., ponto 7 da rede de pontos^{+}, *i.e.*, município de Porto Feliz;

P8 n.o., ponto 8 da rede de pontos^{+}, *i.e.*, município de Tietê;

P9 n.o., ponto 9 da rede de pontos^{+}, *i.e.*, município de Capivari;

P10 n.o., ponto 10 da rede de pontos^{+}, *i.e.*, município de Piracicaba;

part. particípio;

PDO pós-doutorado;

PP português paulista;

pp. páginas;

proc. processo;

proj. projeto;

Q(*x*) questão (dos questionários), sendo *x* variável de numeração;

QFF Questionário Fonético-Fonológico em versão de 2001 da lavra do ALiB (2014);

QFF^{m} QFF modificado por nós;

QSL Questionário Semântico-Lexical em versão de 2001 da lavra do ALiB (2014);

QSL^{m} QSL modificado por nós;

ref. referente(mente), com referência;

Ret n.o., retificação^{+};

S no QSL, comando "Sugira" (*v.* sugestão^{+}); nas demais partes, item sugerido;

S1 planilha 1 (orig. *spreadsheet* 1);

S2 planilha 2 (orig. *spreadsheet* 2);

S3 planilha 3 (orig. *spreadsheet* 3);

seç. seção;

SP Estado de São Paulo;

subst. substantivo;

Sus n.o., suspeição de designação^{+};

T1 primeiro tempo da TTT;

T2 segundo tempo da TTT;

T3 terceiro tempo da TTT;

tab. tabela;

TDS Temas para Discursos Semidirigidos em versão de 2001 da lavra do ALiB (2014);

TDS^{m} Temas para Discursos Semidirigidos modificados por nós;

TE n.o., técnica de entrevista orientada à contundência responsiva/anuente^{+};

TTT	técnica de entrevista de três tempos[+];
USP	Universidade de São Paulo;
V	vogal (em simbologia fonológica[+] categorial);
v.	veja(m)-se;
VA	valor absoluto;
vc/sr	você / (o)(a) senhor(a);
VR	valor relativo;
vs.	em contraste com (orig. *versus*);
Wf	indivíduos do gênero de registro/social[+] feminino;
Wm	indivíduos do gênero de registro/social[+] masculino; e
Wz	expressão abstrata a ser parametrizada concretamente ou por 'Wf', ou por 'Wm'.

SÍMBOLOS E RECURSOS TIPOGRÁFICOS[1]

' aspa simples única: em notação fonético⁺-fonológica⁺, indicação de sílaba tônica primária à direita;

' ' par de aspas simples: (i) destaca termos técnicos, conceitos, noções e valores (abreviados ou por extenso); e (ii) aplica-se como substituto de aspas duplas já encerradas entre aspas duplas;

" " par de aspas duplas: destaca: (i) enunciados; (ii) expressões figuradas; (iii) símbolos; (iv) títulos de seção ou de tomo de obras escritas (exceto em sumários e em outros componentes chamados "pré-textuais"); e (v) citações diretas;

ˌ apóstrofo subscrito: em notação fonético⁺-fonológica⁺, indicação de sílaba tônica secundária à direita;

. ponto: no âmbito fonético⁺-fonológico⁺, separação silábica;

... reticências: devem ser lidas "Como se chama(m)" pelo inquiridor nos questionários;

● ponto marcador: (i) separador de itens; ou (ii) marcador simples de listas;

* asterisco: (i) sinal de nota de rodapé em alguns casos; e (ii) comando "Mostre a respectiva imagem" no bojo dos questionários;

⁺ sinal de mais sobrescrito: n.o., encontrado à direita de termos técnicos, ele é uma indicação de existência das respectivas definições na seção "Glossário", no Tomo VI. Quando os termos técnicos tiverem extensão superior a uma palavra, o escopo do sinal de mais sobrescrito será indicado por uma sublinha;

+? sinal de mais e sinal de interrogação adjuntos: n.o., lembrete para a execução da insistência⁺;

[1] Denotando 'nesta obra', a abreviação 'n.o.' se associa a (partes de) definições de símbolos — segundo nosso conhecimento — próprias deste trabalho.

- hífen: quando isolado e não definido localmente, representa (i) questão eliminada no âmbito dos questionários, bem como (ii) "nenhum(a)" ou "sem resultado" em tabelas. Ainda, o hífen é utilizado como (iii) item de ligação entre covariantes+ que perfizerem uma configuração variacional+;

— travessão em negrito: n.o., introduz, imediatamente à sua direita, formas semasiológicas+ que não entram na ordem de relevância+;

— — par de travessões: é, em geral, uma alternativa estilística ao par de parênteses em relação a apostos e, em específico, um realçador destes;

_ *underline:* mero conector das partes do nome técnico de uma carta linguística ou de um enfoque ponderativo específico, sendo que as partes mais à esquerda são mais genéricas e as mais à direita, mais particulares;

/ barra obliqua inclinada à direita: equivalente a: (i) "ou" em geral; e, (ii) em específico, "de" entre números de páginas;

/ / par de barras obliquas inclinadas à direita: encerra notação fonológica+;

| | par de barras paralelas verticalmente: denota 'pausa';

[ ] par de colchetes retos: encerra: (i) notação fonética+; (ii) constituintes sintáticos isolados; e (iii) enunciados alternativos ou complementares de questões do QSL+;

( ) par de parênteses: encerra: (i) numeração de itens; (ii) aposto; (iii) detalhes de obras; (iv) itens contingentes/facultativos; (v) desinências morfossintáticas+ alternativas; bem como funciona como: (vi) organizador sintático; e (vii) delimitador de escopo de etiqueta no âmbito dos *corpora*;

{ } par de chaves: encerra: (i) comentários do pesquisador; e, (ii) se entre colchetes retos, [ ], segmentos fônicos adjacentes (aos ambientes fonológicos+ sob enfoque em cada questão do QFFm) atualizados sonoramente de modo tal no inquérito que eles não perfazem nenhuma das contrapartes fonéticas+ diretamente correspondentes à estrutura fonológica estritamente implicada na forma grafemática+ prototípica do respectivo vocábulo; bem como funciona como (iii) organizador/delimitador de escopo em notações técnicas;

< > par de colchetes angulares: encerra grafema(s)+ (no entanto, recurso em geral dispensado em células de tabelas nas quais os elementos grafemáticos figurem sozinhos);

= sinal de igual: equivale a: (i) "igual a"; e (ii) "significa";

hash ou cerquilha: n.o., (i) comando "Faça o gesto previamente convencionado" entre inquiridor e informante para casos de lacuna nos questionários que deve ser preenchida; e (ii) indicação de fronteira vocabular quando o símbolo estiver encerrado por um par de chaves, { }, o qual, por sua vez, esteja no interior de uma notação fonológica+ ou fonética+;

Σ letra grega sigma: significa "somatório";

Ø letra maiúscula dos alfabetos norueguês e dinamarquês (notação fonológica+: /ø:/): representa: (i) ausência de fone+ (apagamento) em notação fonética+ como contraparte de um dado (arqui)fonema+; e, em outros domínios, (ii) equivalente a "nada";

((!)) exclamação entre parênteses duplos: n.o., alerta que acompanha questões (dos questionários) que frequentemente requerem paráfrases ou atos ostensivos pragmáticos+ adicionais por parte do inquiridor no trabalho de fazer com que o inquirido entenda bem o que se esteja perguntando;

itálico aplicado a: (i) expressões de outros idiomas; (ii) termos, conceitos, noções e frases enfatizados; (iii) títulos de obras e de subseções geralmente sem marcadores; e (iv) sugestões+ no QSL^{m}. Esse recurso tipográfico é revertido quando há acúmulo (dessas aplicações) em número par;

^{m} eme sobrescrito: n.o., acompanha os nomes de instrumentos metodológicos de coleta que se apresentam em nossa versão, modificados, derivados dos originais de 2001 concebidos pelo Atlas Linguístico do Brasil (ALiB, 2014);

negrito aplicado a: (i) constituintes de perguntas dos questionários a serem enunciados sob ênfase; e (iii) ambientes fonológicos+ sob enfoque nas questões do QFF^{m} (nesse caso, o destaque do **negrito** é reforçado pela <u>sublinha</u>); e

<u>sublinha</u> aplicado: (i) n.o., a itens de extensão superior a uma palavra que estejam em relação de intercambialidade entre si, mediados por uma barra oblíqua inclinada à direita, /; (ii) a termos técnicos de extensão superior a uma palavra sobre os quais o recurso do sinal de mais sobrescrito, $^{+}$, tem escopo; e (iii) ambientes fonológicos^{+} sob enfoque nas questões do QFF^{m} (nesse caso, o destaque da <u>sublinha</u> é reforçado pelo **negrito**).

SUMÁRIO

TOMO I
Prolegômenos e análises

PREÂMBULO 55
APRESENTAÇÃO 61

1 RESUMOS 67
1.1 Proposta inicial da investigação 68
1.2 Alterações metodológicas no projeto 70
1.3 Atividades intermediárias da pesquisa no Brasil 71
1.4 Atividades intermediárias da pesquisa na Alemanha 74
1.5 Progressos realizados e resultados obtidos 75

2 INTRODUÇÃO 85
2.1 Estudos acerca do português paulista 85
2.2 Atlas linguísticos no Estado de São Paulo 91
2.3 Nossa pesquisa 92
2.3.1 Exórdio 92
2.3.2 Transcrição, tabulação e organização dos dados coletados 98
2.3.3 Treinamento no *software* de cartografação 101
2.3.4 Dificuldades surgidas/esperadas 105

3 OBJETIVOS E HIPÓTESE 107
3.1 Objetivos gerais 107
3.2 Objetivos específicos 108
3.3 Hipótese 109

4 JUSTIFICATIVAS 111
4.1 Dos objetivos gerais 111
4.2 Do *locus* 112

TOMO I (*continuação*)
Prolegômenos e análises

5 METODOLOGIA 113
5.1 Procedimentos gerais em campo e materiais 113
5.2 Perfis dos informantes 114
5.3 Instrumentos de coleta 115
5.4 Modificações nos instrumentos metodológicos 121
5.5 Técnica de entrevista ao QSL^{m} e *labels* às formas semasiológicas 123
5.6 Neutralização pragmática da problemática do "certo" *vs.* "errado" 130
6 ANÁLISES DO /R/ EM CODA 135
6.1 Externa 142
6.1.1 De desinência infinitiva (DI) seguida por pausa 142
6.1.2 Em sílaba tônica (a qual não seja, porém, de DI) seguida por pausa 148
6.2 Interna 154
6.2.1 Em sílaba tônica 154
6.2.2 Em sílaba átona 160
6.3 Por localidade 166
6.4 Normas linguísticas e influências de variáveis extralinguísticas 179
6.5 Conclusões 195
7 NOTAS 209
7.1 QSL^{m} em foco 209
7.2 Três tipos de informantes 214
7.3 Táticas responsivas ao QSL^{m} 215
7.4 Mídias 218

TOMO II
Atlas semântico-lexical

LISTA DE CARTAS SEMÂNTICO-LEXICAIS 223
INTRODUÇÃO AO ATLAS SEMÂNTICO-LEXICAL 231

1 QUATRO (GRUPOS DE) COVARIANTES 237
1.1 Série CxGy 239
1.2 Série CxWz 325
1.3 Série WzGy 411
2 TRÊS (GRUPOS DE) COVARIANTES 497
2.1 Série CxGy 499
2.2 Série CxWz 545
2.3 Série WzGy 591
3 DUAS (OU DOIS GRUPOS DE) COVARIANTES 637
3.1 Série CxGy 639
3.2 Série CxWz 657
3.3 Série WzGy 675

TOMO III
Atlas fonético-fonológico

LISTA DE CARTAS FONÉTICO-FONOLÓGICAS 693
INTRODUÇÃO AO ATLAS FONÉTICO-FONOLÓGICO 699

1 QUATRO (GRUPOS DE) COVARIANTES 705
1.1 Série CxGy 707
1.2 Série CxWz 727
1.3 Série WzGy 747

2 TRÊS (GRUPOS DE) COVARIANTES 767
2.1 Série CxGy 769
2.2 Série CxWz 805
2.3 Série WzGy 841

3 DUAS (OU DOIS GRUPOS DE) COVARIANTES 877
3.1 Série CxGy 879
3.2 Série CxWz 933
3.3 Série WzGy 987

4 CARTAS ESPECIAIS SOBRE O /R/ EM CODA 1041
4.1 Externa de desinência infinitiva seguida por pausa 1043
4.2 Externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa. 1049
4.3 Interna em sílaba tônica 1055
4.4 Interna em sílaba átona 1061

TOMO IV
Corpora (parte 1/2)

INTRODUÇÃO AOS *CORPORA* 1067

1 *CORPORA* SEMÂNTICO-LEXICAIS 1069
1.1 Formas semasiológicas com informação fonética (*corpus* primário) 1069

TOMO V
Corpora (parte 2/2)

1.2 Cabeças de relevância 1471
1.3 Contagem das lexias 1640
1.4 Normas semântico-lexicais absolutas 1762

2 *CORPORA* FONÉTICO-FONOLÓGICOS 1767
2.1 Fones (*corpus* primário) 1767
2.2 Normas fonético-fonológicas absolutas 1877

3 *CORPUS* METALINGUÍSTICO-ETNOGRÁFICO 1883

4 *CORPORA* EXTRALINGUÍSTICOS 1901
4.1 Informações básicas 1903
4.2 Domicílios e filiação 1908
4.3 Línguas 1913
4.4 Trabalho, religião e grau escolar 1914
4.5 Exposição às mídias escrita e sonora 1919
4.6 Exposição à mídia televisiva e a eventos esportivos e culturais 1924
4.7 Entrevista 1929

TOMO VI
Epílogo e adendos

CONSIDERAÇÕES FINAIS 1931
REFERÊNCIAS 1945
GLOSSÁRIO 1951

NOTAS SOBRE OS ADENDOS 1959

APÊNDICES
Instrumentos de coleta
A Ficha do Informante[m] 1969
B Temas para Discursos Semidirigidos[m] 1971
C Questionário Semântico-Lexical[m] 1973
D Questionário Fonético-Fonológico[m] 2021
Fontes das imagens para os questionários
E QSL[m] 2061
F QFF[m] 2065
Outros
G Universo das sugestões 2071
H Metaplasmos 2077
I Classificação das ocupações/profissões 2081
J Termo de Consentimento Livre e Esclarecido 2083

ANEXOS
I Ficha do Informante do ALiB 2087
II Temas para Discursos Semidirigidos do ALiB 2089
III Questionário Semântico-Lexical do ALiB 2091
IV Questionário Fonético-Fonológico do ALiB 2107
V IPA 2119

PREÂMBULO

De maneira sucinta, esta seção oferecerá um conjunto de informações importantes ao entendimento preliminar de aspectos gerais desta tese idealmente a não serem ignorados. Os aspectos aqui tratados vão desde medidas textuais e de diagramação adotadas ao longo desta obra até acautelamentos tanto conceituais quanto de ordem prática.

Isso adiantado, diga-se desde já que definições de termos técnicos poderão ser consultadas na seção "Glossário", no Tomo VI. Ao longo deste trabalho, os termos técnicos constituídos por apenas uma palavra levarão, imediatamente à sua direita, um sinal de mais sobrescrito, $^{+}$. Quando os termos, entretanto, forem compostos por mais de uma palavra, eles serão adicionalmente sinalizados com sublinha. Tal sinalização será aplicada, todavia, apenas à primeira ocorrência dos termos, em geral, no âmbito constituído por seções primárias e suas seções subordinadas e, em específico, no interior de definições; definições essas presentes nomeadamente nas seções "Siglas, acrônimos e abreviações", "Símbolos e recursos tipográficos" (no corrente tomo) e "Glossário" (no Tomo VI). Ainda, diga-se que esse recurso tipográfico será prescindido: (i) nas seções onde suas definições já se encontrem; (ii) em títulos de seções, de documentos, de figuras, de gráficos e de tabelas; (iii) no interior de cartas linguísticas, gráficos e tabelas; e, por fim, (iv) na seção "Considerações finais", no Tomo VI.

Quanto às siglas, acrônimos e abreviações, assim como os símbolos e recursos tipográficos utilizados, eles foram explicados imediatamente antes da presente seção. Sua familiarização é recomendada.

Isso dito, vejamos a seguir algumas observações preambulares de cuja ciência quer-se que o leitor possua, muitas das quais tendo sido motivadas por sugestões dadas pela Comissão Julgadora por ocasião da Defesa Púbica da Tese de Doutorado.

Terminologia e expressões:

- o termo 'carta linguística' será aplicado em detrimento de 'carta geolinguística' (e de 'mapa'), dado que a noção semântica^{+} sugerida pelo prefixo "geo-" já está implicada no vocábulo "carta" (de unidade cartográfica);

- o termo 'extralinguístico' será preferencialmente posto em uso em detrimento de 'social', como designativo geral aos dados não linguísticos coletados dos informantes, tais como os estritamente sociais (*e.g.* escolaridade), geográficos (*e.g.* domicílios), culturais (*e.g.* hábitos de consumo de mídia) e cronológico (*e.g.* idade);

- expressões semelhantes a "formas (pel)as quais o informante responde" se referirão não só aos itens que o entrevistado ativamente enuncia, mas também àqueles aos quais ele passivamente anui no processo do inquérito;

- a expressão "informação fonética"$^{+}$ será aplicada como alternativa a "transcrição fonética parcial";

- quando nos referirmos a ambos os atlas aqui presentes — *i.e.*, ao atlas semântico-lexical^{+} (Tomo II) e ao atlas fonético-fonológico^{+} (Tomo III) — sem necessidade de fazer distinções entre eles, diremos apenas "atlas linguístico", por simplicidade, a exemplo do título (sem considerar o subtítulo) desta obra;

- o vocábulo "seção" será utilizado para referir-se a qualquer divisão do presente trabalho, independentemente de seu estatuto e localização (capítulo, subseção, apêndice, símbolos *etc.*);

- os seguintes termos e vocábulos serão utilizados de maneira intercambiável sem diferença semântica fundamental:

  – 'informante', 'inquirido', 'entrevistado', 'participante' e 'falante', bem como 'indivíduo' quando a sinonímia associada aos primeiros estiver pelo menos relativamente clara no local de ocorrência no texto; e

  – 'inquérito', 'entrevista' e 'elicitação';

- os termos 'variante' e 'covariante' serão utilizados de maneira intercambiável em princípio, mas o segundo será preferido quando a intenção for destacar a relação estrita de *covariação* entre dois ou mais itens a constituir uma mesma variável (haja vista que o primeiro termo é passível de estar sob um uso que denote tanto itens associados a variáveis distintas entre si, quanto um item sem concorrentes);

- os termos 'lexia'+ e 'forma semasiológica'+ serão utilizados de maneira intercambiável em princípio, mas o segundo será preferido quando a intenção for destacar a abrangência semântica que inclui orações+ (e não só lexias propriamente ditas — e não necessariamente as chamadas 'lexias textuais') como respostas dos informantes ao Questionário Semântico-Lexical m (QSL m), de natureza onomasiológica+;

- os vocábulos "consulente" e "leitor" serão utilizados de maneira intercambiável em princípio, mas haverá preferência pelo primeiro quando a ideia for destacar a ideia de consulta (pontual);

- os termos 'diagenérico'+ e 'gênero de registro/social' substituirão os tradicionais e — no âmbito da investigação de que se trata aqui — menos adequados 'diassexual' e 'sexo', respectivamente[1]; e

- o termo 'semiconsoante' será aplicado em detrimento de *'glide'* e de 'semivogal' quando o interesse for destacar as propriedades distributivas do segmento.

Simbologia:

- os símbolos IPA (cf. seç. de mesmo nome no Tomo VI) serão os aplicados, sempre apresentados ou entre barras obliquas inclinadas à direita, / / (em notação fonológica), ou entre colchetes retos, [ ] (em notação fonética);

[1] A dimensão sexual — de natureza íntima e que pode ser parametrizada com valores para além do binarismo 'feminino' *vs.* 'masculino' convencional — não foi controlada em nossa pesquisa, mas sim a dimensão do gênero de registro/social da pessoa — de natureza social (para a qual as únicas opções de registro cível disponíveis aos indivíduos submetidos ao ordenamento jurídico brasileiro ainda são 'feminino' e 'masculino').

- a africada alveopalatal desvozeada e a africada alveopalatal vozeada serão representadas com a barra superior de atamento: [t͡ʃ] e [d͡ʒ], respectivamente; e

- os símbolos [i̯] e [u̯] representarão os *glides*/semiconsoantes/semivogais, em detrimento de [j] e [w], respectivamente.

Fraseologia e diagramação:

- optamos pelo uso da primeira pessoa do plural, ora em detrimento da primeira pessoa do singular (apenas usada na seção "Agradecimentos") vinculada ao autor, ora em prol de um "nós" constituído por autor e leitor. No entanto, recorremos também a construções sintáticas impessoais como alternativas fraseológicas;

- uma das referências temporais dominantes nesta obra — a qual lança mão, por exemplo, da conjugação verbal no futuro do presente do indicativo — tem o propósito de causar um efeito fraseológico (ou estilístico) de associação entre autor e leitor na condução da leitura, no sentido em que ambos "verão juntos" algo posicionado posteriormente a um dado ponto de leitura;

- quando sem prejuízo do que queremos comunicar com precisão, alternaremos diferentes tempos e modos verbais como medida fraseológica, mas procurando não ferir a observância da compatibilidade gramatical entre eles no domínio do período sintático;

- parágrafos antecedidos ou por uma linha em branco, ou por algum marcador isolado, serão iniciados sem recuo;

- as páginas em branco em dadas partes da obra serão necessárias para sua correta impressão no modo frente e verso e em observância às normas a esse respeito aplicáveis a este trabalho; e

- a numeração das notas de rodapé zerará na passagem entre seções.

Outras notas:

- o aporte teórico-metodológico que adotamos aqui é ocasionalmente chamado de formas diferentes — Dialetologia Pluridimensional (DP), Geolinguística Pluridimensional (GP), Dialetologia Pluridimensional e Relacional (DPR), entre outras designações possíveis —, mas optaremos por DP;

- ao leitor externo à Linguística, diga-se que as noções de 'consoante', *'glide'* (ou 'semivogal', ou 'semiconsoante') e 'vogal' não corresponderão aqui a letras, mas sim a segmentos fônicos;

- entre os tipos de transcrição fonética ampla (*broad transcription*, que se atém aos aspectos gerais) e restrita (*narrow transcription*, que inclui propriedades secundárias), far-se-á uso da primeira, e recorrer-se-á à segunda quando julgado relevante;

- salvo em casos específicos, quando houver a mera necessidade de se fazer referência às perguntas dos questionários, ela será realizada com duas informações combinadas: (i) com o número que as identificam; e (ii) com a covariante coletada mais frequente — em casos de agrupamentos de covariantes^{+}, apenas o primeiro item será utilizado; situação ref. ao Questionário-Semântico-Lexical^{m}, QSL^{m} — ou com o vocábulo cuja pronúncia por parte do informante se induziu (situação ref. ao Questionário Fonético-Fonológico^{m}, QFF^{m}). Caso contrário, as referências ficariam ou muito abstratas apenas com o número da questão (no caso de ambos os questionários), ou demasiadamente longas com o enunciado em si das perguntas (no caso específico do QSL^{m});

- os diálogos ilustrativos de que lançaremos mão em algumas seções são produções específicas para a redação estandardizada da presente obra e são baseados em conversas reais, que se deram sob variedades linguísticas^{+} populares na pesquisa de campo;

- para esta versão final, dadas as proporções do trabalho, fomos solicitados a repetir algumas informações — o que, ocasionalmente, fizemos *ipsis litteris* — em diferentes seções (especialmente naquelas distantes entre si) e assim, portanto, a reduzir a frequência dos pedidos de remissão do leitor a diferentes partes ou tomos da obra;

- às vezes, dialetólogos/geolinguistas doutorandos, em função da quantidade de trabalho e do respeito aos cronogramas de execução da pesquisa, precisam delegar algumas tarefas (*e.g.* condução dos inquéritos, transcrição dos dados, elaboração das cartas linguísticas, diagramação da tese) a terceiros, nomeadamente a outros acadêmicos (como a bolsistas de iniciação científica) e/ou a profissionais externos (como a cartógrafos). Todavia, em nosso caso, nada em absoluto foi objeto de terceirização; e

- muito importante que o leitor tenha em mente que, primariamente, o nexo entre forma expressa e conteúdo semântico estabelecido pelo informante não deverá ser julgado à luz de vocabulários de outras variedades linguísticas, mas sim tendo-se em conta as relações de significação da própria comunidade sobre a qual aqui se dá notícia através da gama de dados empíricos disponibilizada.

Por fim, gostaríamos de encerrar esta seção, por um lado, desejando ao leitor uma boa leitura/consulta e, por outro lado, encorajando-o a enviar-nos sua crítica sobre qualquer aspecto deste trabalho para selmojunior@usp.br / selmojunior@gmail.com.

APRESENTAÇÃO

Aqui, vamos abordar a organização do presente trabalho, pensada para fundamentalmente propiciar ao leitor/consulente acesso rápido aos dados, com possibilidade de rastreamento relativamente completo de sua origem. Tanto o acesso em geral quanto o rastreamento em específico serão facilitados pelo uso de índices, tal como o número do informante (inf*x*) no caso dos *corpora*.

Ou seja, através dos índices, não só o acesso a todo o material linguístico e metalinguístico^{+}-etnográfico produzido por cada entrevistado será possível, mas também o rastreamento de uma gama de atributos extralinguísticos associados ao producente, cuja extensão vai desde sua ascendência étnica, passando por seu histórico domiciliário topodinâmico^{+} e/ou topostático^{+}, até seus hábitos de interação com diferentes veículos midiáticos.

Para este e outros estudos, as informações extralinguísticas subsidiarão uma compreensão abrangente de fatos de língua produzidos pelos entrevistados do interior paulista; fatos cuja coleção proporcionará uma visão de conjunto dos dados — o que, por sua vez, promoverá condições de empiricidade para generalizações hipotético-explicativas *mediante identidades* — e cuja individualidade permitirá um aprofundamento investigativo qualitativo da realidade dialetal *mediante idiossincrasias*. Tal perspectiva é fomentada pela Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*), que aqui adotamos.

Isso dito, o presente empreendimento organiza-se espacialmente em seis tomos:

(I) Prolegômenos e análises;
(II) Atlas semântico-lexical;
(III) Atlas fonético-fonológico;
(IV) *Corpora* (parte 1/2);
(V) *Corpora* (parte 2/2); e
(VI) Epílogo e adendos.

(I)

Tirante a presente seção e "Preâmbulo", este primeiro tomo divide-se em sete partes principais, apresentadas sumariamente abaixo.

RESUMOS, concebidos para darem uma ideia intermediária entre a corrente seção e o material na íntegra distribuído pelos seis tomos, eles se concentram, porém: (i) na proposta inicial da nossa investigação; (ii) nas alterações metodológicas no projeto inicial de pesquisa; (iii) nas atividades intermediárias da pesquisa no Brasil; (iv) nas atividades intermediárias da pesquisa na Alemanha; e (v) nos progressos realizados e resultados obtidos.

INTRODUÇÃO, que expõe três partes primárias: (i) estudos acerca do português paulista; (ii) atlas linguísticos no Estado de São Paulo; e (iii) observações iniciais sobre nossa pesquisa. Estas últimas dividem-se em outras quatro partes secundárias. Elas tratam: (i) de um exórdio; (ii) da transcrição, da tabulação e da organização dos dados coletados; (iii) do treinamento no *software* de cartografação; e (iv) das dificuldades surgidas/esperadas durante a pesquisa.

OBJETIVOS E HIPÓTESE, que versam inicialmente sobre os dois objetivos gerais e sobre os sete objetivos específicos principais da nossa investigação. Como de praxe, estes últimos servem aqueles primeiros, os quais dizem respeito à elaboração de: (i) um atlas semântico^{+}-lexical^{+}; e (ii) um atlas fonético^{+}-fonológico^{+}. Adicionalmente, a seção em pauta apresenta uma hipótese específica que testamos; segundo a qual, *grosso modo*, os jovens estariam atualizando menos o fone^{+} [ɻ] (*i.e.*, a aproximante retroflexa, também conhecida como "erre caipira"$^{+}$) em coda^{+} externa de desinência infinitiva seguida por pausa do que os velhos.

JUSTIFICATIVAS, tratando das razões essenciais que nos fizeram estabelecer os nossos objetivos gerais e nos fizeram escolher o *locus* do estudo, dizem respeito ao fato (entre outros) de que inexiste um atlas linguístico do português paulista (ou "dialeto^{+} caipira") na região que tanto é um dos iniciais e principais centros de difusão da língua portuguesa trazida pelos colonizadores portugueses, quanto é considerada o berço da cultura caipira: Médio Tietê. Além disso, na seção em foco arrolam-se benefícios/aplicações decorrentes desse empreendimento dialetológico e geolinguístico que os atlas linguísticos representam.

METODOLOGIA, cujo conteúdo divide-se em seis partes: (i) procedimentos gerais em campo e materiais (abordagem aos indivíduos e dimensão prática do inquérito); (ii) perfis dos informantes (características dos indivíduos entrevistados); (iii) instrumentos de coleta (meios utilizados para o levantamento dos dados empíricos *in loco*); (iv) edição dos instrumentos metodológicos (detalhes acerca das modificações incidentes nos questionários *etc.*); (v) técnica de entrevista e etiquetas às formas semasiológicas+ (condução da entrevista no âmbito do questionário voltado ao nível semântico-lexical e valores atribuídos às respostas); e, finalmente, (vi) neutralização prática da problemática do "certo" *vs.* "errado" (medida orientada à otimização na obtenção de dados autênticos).

ANÁLISES DO /R/, ou melhor, análises das contrapartes (a)fonéticas+ do arquifonema+ /R/ em coda silábica no português paulista falado na região do Médio Tietê, com base em nosso *corpus* fonético primário e em nossas cartas fonético-fonológicas especiais. As variáveis linguísticas controladas são: (i) /R/ em coda externa de desinência infinitiva seguida por pausa; (ii) /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa; (iii) /R/ em coda interna em sílaba tônica; e (iv) /R/ em coda interna em sílaba átona.

NOTAS, que se referem a quatro tópicos: (i) Questionário Semântico-Lexical m (QSL m) (problemas relacionados e imagens ilustrativas de referentes semânticos+); (ii) três tipos de informantes (esboço de tipologia considerando idioletos+); (iii) táticas responsivas ao QSL m (expedientes (in)conscientes performados pelos informantes); e (iv) mídias (presença de veículos de comunicação de massa na vida dos entrevistados).

(II) e (III)

Os Tomos II e III compreendem exclusivamente o atlas semântico-lexical e o atlas fonético-fonológico, respectivamente. Compondo o que às vezes chamaremos simplesmente de atlas linguístico pluridimensional do português paulista, ambos estão organizados primariamente por categorias (ref. ao número de covariantes+ simultaneamente tratadas e cartografadas) e secundariamente por séries (ref. às dimensões extralinguísticas simultaneamente correlacionadas com os fatos de língua abordados). As dimensões em questão — que representam variáveis extralinguísticas sistematicamente controladas em nossa pesquisa — são a diastrática+ escolar, a diageracional+ e a

diagenérica+. Adicionalmente, o Tomo III contém as cartas fonético-fonológicas especiais às contrapartes (a)fonéticas do arquifonema /R/ em coda silábica.

(IV)

O Tomo IV abarca a primeira de duas partes dos *corpora* constituídos por nossa pesquisa. Essa primeira parte encerra o *corpus* primário dos *corpora* semântico-lexicais. Nela, trata-se de todas as formas semasiológicas (pel)as quais os informantes responderam ao Questionário Semântico-Lexical m (QSL m). Parte desse *corpus* primário também está transcrito foneticamente.

(V)

Dando sequência ao Tomo IV — o qual cobre o (i) *corpus* semântico-lexical primário —, o Tomo V inicia com as seções restantes dos *corpora* semântico-lexicais: (ii) cabeças de relevância (primeiros itens na ordem de relevância+ das respostas do informante ao QSLm estabelecida pelo próprio entrevistado ou com ajuda dele); (iii) contagem das lexias+ (pel)as quais os inquiridos responderam no bojo do QSLm (cômputo geral das ocorrências totais distribuídas ao longo de nossa rede de pontos+); (iv) normas+ semântico-lexicais absolutas (lexias observadas ao longo de toda a rede de pontos e com frequência relativa superior a cinquenta por cento, sem discretização de grupos específicos de falantes). Além disso, as outras partes do tomo sob enfoque são as arroladas abaixo.

CORPORA FONÉTICO-FONOLÓGICOS, cindidos em duas porções: (i) *corpus* primário (conjunto dos fones produzidos pelos informantes por ensejo da aplicação do Questionário Fonético-Fonológicom, QFFm); e (ii) normas fonético-fonológicas absolutas (fones observados ao longo de toda a rede de pontos e com frequência relativa superior a cinquenta por cento, sem discretização de grupos específicos de falantes).

CORPUS METALINGUÍSTICO-ETNOGRÁFICO, que consiste de comentários cuja natureza o título da coleção sugere e os quais foram espontaneamente enunciados pelos informantes por ensejo da condução do inquérito semântico-lexical. Nesse sentido então, quando observada a natureza de seu conteúdo, as declarações coletadas foram divididas em dois grupos: grupo metalinguístico e grupo etnográfico.

CORPORA EXTRALINGUÍSTICOS, compostos pelos seguintes dados referentes aos entrevistados: (i) informações básicas (ascendência étnica, classe econômica *etc.*); (ii) domicílios e filiação (histórico de domicílios, naturalidade dos genitores *etc.*); (iii) línguas (proficiência em outros idiomas, contexto em que são utilizados *etc.*); (iv) trabalho, religião e grau escolar (onde trabalha/estuda, prática religiosa *etc.*); (v) exposição às mídias escrita e sonora (hábito de leitura, preferências de rádio *etc.*); (vi) exposição à mídia televisiva e a eventos esportivos e culturais (programas de TV aos quais se assiste, esportes os quais se acompanham *etc.*); e (vii) entrevista (estado psicológico no inquérito, grau de espontaneidade *etc.*).

(VI)

Salvo as seções "Considerações finais", "Referências" e "Glossário", o último tomo engloba três partes principais, mencionadas a seguir.

NOTAS SOBRE OS ADENDOS, que apresentam explicações e sugestões relativas aos apêndices e aos anexos. Esses apontamentos podem ser importantes tanto para aqueles que venham a utilizar os adendos, tal como eles estão no momento, quanto para aqueles que venham a editá-los, visando a melhorias.

APÊNDICES, editados ou formados por nós, classificam-se em: (i) instrumentos de coleta; (ii) fontes das imagens para os questionários; entre (iii) outros. Os primeiros são os seguintes elementos: (i.a) Ficha do Informante[m] (aplicada para a construção dos *corpora* extralinguísticos); (i.b) Temas para Discursos Semidirigidos[m] (utilizados como facilitadores preliminares ao engajamento interacional entre inquirido e inquiridor); (i.c) Questionário Semântico-Lexical[m] (empregue para a criação do *corpus* semântico-lexical primário); e (i.d) Questionário Fonético-Fonológico[m] (operado para a composição do *corpus* fonético primário). O segundo grupo de apêndices, (ii), reporta os créditos e as fontes das quais as imagens ilustrativas aos questionários foram extraídas, bem como fornece os caminhos para seu acesso. Por fim, o terceiro grupo de apêndices refere-se: (iii.a) ao universo das sugestões (ref. às lexias apresentadas aos falantes para se verificar se elas lhes eram conhecidas passivamente ou não); (iii.b) a metaplasmos[+] (ref. a uma compilação de processos fonológicos); (iii.c) à classificação das ocupações/profissões (dos entrevistados); e (iii.d) ao

Termo de Consentimento Livre e Esclarecido (documento que foi assinado pelos participantes da pesquisa empírica).

ANEXOS, encerrando a obra, são constituídos por cinco partes: (i) Ficha do Informante do Atlas Linguístico do Brasil (ALiB, 2014); (ii) Temas para Discursos Semidirigidos (*ibidem*); (iii) Questionário Semântico-Lexical (*ib.*); (iv) Questionário Fonético-Fonológico (*ib.*); e (v) Alfabeto Fonético Internacional (orig. *International Phonetic Alphabet*, IPA).

1 RESUMOS

Versaremos a seguir, tão objetivamente quanto possível, sobre uma série de resumos importantes e fundamentais à nossa pesquisa agora consolidada e — nos tomos que compõem este trabalho — divulgada com seus resultados.

Isso dito, as súmulas mais adiante tematizarão:

(a) (na seção 1.1:) a proposta inicial da investigação conduzida no âmbito da Universidade de São Paulo (USP) como instituição sede e nacionalmente apresentada à, e apoiada pela, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), via bolsa de doutorado (processo nº 15/14038-5) e via diárias de campo no âmbito do Projeto de História do Português Paulista – PHPP, Projeto Caipira (processo nº 11/51787-5);

(b) (na seção 1.2:) as mudanças realizadas no projeto inicial da pesquisa com aprovação da USP e da FAPESP;

(c) (na seção 1.3:) as atividades intermediárias do estudo levadas a cabo no Brasil até o dia 31 de março de 2016;

(d) (na seção 1.4:) as tarefas também de entremeio ocorridas durante a vigência de um estágio na Alemanha, entre o dia 1º de abril de 2016 e o dia 31 de março de 2018, apoiado tanto pelo Serviço Alemão de Intercâmbio Acadêmico (orig. *Deutscher Akademischer Austauschdienst,* DAAD), via bolsa de doutorado sanduíche (programa nº 57214225), quanto pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), via auxílios de deslocamento (processo nº 0128-16-3 / 99999.000128/2016-03), em parceria com o DAAD; e

(e) (na seção 1.5:) os principais produtos da pesquisa.

Essas sínteses são oferecidas ao leitor como subsídios preliminares considerados significativos para a compreensão desde já não só do nosso ponto de partida, mas também de elementos do contexto dos esforços e da evolução da investigação. Em seções posteriores, vários aspectos aqui apenas delineados serão retomados para detalhamentos.

1.1 Proposta inicial da investigação

A seguir, expõe-se *ipsis litteris* o plano inicial do nosso estudo junto à USP e à FAPESP, nos termos de sua enunciação oficial no ano de 2015.

"A pesquisa de doutoramento de que se vai tratar aqui tem a tarefa de realizar um atlas linguístico parcial (níveis fonético[+]-fonológico[+] e léxico-semântico[+]) da variedade[+] da língua portuguesa falada na região do Médio Tietê, composta pelos municípios de Tietê, Santana de Parnaíba, Araçariguama, Itu, Sorocaba, São Roque, Porto Feliz, Pirapora do Bom Jesus, Capivari e Piracicaba, municípios paulistas cujo início de urbanização remonta ao século XVI.

"A perspectiva teórico-metodológica que o presente projeto adota é a dialetologia (ou geolinguística) pluridimensional e relacional (DP) (THUN & RADTKE, 1996[1]), uma abordagem da variação linguística que concilia a preocupação de se estudarem os fenômenos espacialmente localizados (chamado aspecto 'horizontal') em correlação com variáveis sociais (aspecto 'vertical'). Thun (1998[2], p.704) define a DP como uma ciência geral da variação linguística e das relações entre variedades e variantes[+], e variantes e falantes.

"Para a coleta de dados, haverá um ponto de inquérito em cada uma das dez localidades, e cada qual contará com quatro informantes, dois homens e duas mulheres, divididos equitativamente em duas faixas etárias (GI: 18 a 36 e GII: 55 em diante). Ainda, de maneira equitativa, os informantes serão divididos entre i) os que possuem escolaridade superior parcial ou completa e ocupação profissional autônoma/livre (Ca); e ii) os que

[1] A referência correta tem os nomes invertidos: Radtke e Thun (1996).
[2] Na seção "Referências", no Tomo VI, procure-se pelo ano seguido pela letra "b", ou seja, por Thun (1998b).

possuem escolaridade básica até o ensino médio completo[3] ou analfabetos que trabalhem na agricultura ou em atividade que não faça utilização da escrita (Cb). Todos cumprindo a condição de serem nascidos na localidade ou que viveram três quartos da vida nela (necessariamente os últimos cinco anos).

"Como instrumentos de coleta de dados, serão usados dois questionários (um semântico-lexical[+] e outro fonético-fonológico), uma conversa livre e a leitura de um etnotexto. Haverá assim um total de 40 informantes, separados segundo as dimensões de [*sic*] sexual (dimensão diassexual[4]), de idade (diageracional[+]) e de classe sociocultural (diastrática[+]).

"Outras dimensões a serem controladas/observadas são a diatópica[+] (informantes com domicílio fixo; topostática), a diatópica-cinética (contraste entre mudança de domicílio e domicílio fixo; topodinâmica), a diafásica (respostas aos questionários, conversa livre e leitura; três estilos de utilização da língua) e a diarreferencial (língua-objeto e metalíngua).

"Metodologicamente, os informantes serão considerados ainda em quatro grupos de fala, segundo os parâmetros definitórios das dimensões, em cada um dos pontos da rede[+] de inquérito, a fim de propiciar posterior confronto e cotejo dos dados produzidos por esses grupos: (a) informantes da Ca e da GI (CaGI); (b) da Ca e da GII (CaGII); (c) da Cb e da GI (CbGI); e (d) da Cb e da GII (CbGII).

"Com esse estudo, visa-se a contribuir com os esforços dos pesquisadores cujas produções, tratadas coletivamente, buscam inventariar e descrever as diversas comunidades de fala, tornando mais e mais possível o estabelecimento do cenário geral da realidade linguística dos países. Esta pesquisa está inserida no Projeto de História do Português Paulista (PHPP - Projeto Caipira), Projeto Temático/FAPESP, processo 2011/51787-5 (AU)."

Adiante, vejamos um apanhado das principais modificações metodológicas incidentes no projeto de pesquisa inicial, reportadas e aprovadas já em inícios de 2016.

[3] Retificação: "incompleto".
[4] O termo 'diassexual' deu lugar a 'diagenérico(a)'+, segundo as razões expressas na parte sobre terminologia constante da seção "Preâmbulo".

1.2 Alterações metodológicas no projeto

Em princípios do ano de 2016, foram implementadas algumas mudanças metodológicas no projeto de investigação, com aquiescência das instituições envolvidas. Os rearranjos mais relevantes, ao lado de suas respectivas justificativas, resumem-se nos seguintes itens:

(a) duplicação do número de informantes por localidade da rede de pontos. Antes, o número era quatro, passando posteriormente a oito. O total, portanto, de 40 passou a 80 participantes. Justificativa: o poder estatístico decorrente do aumento do número de entrevistados eleva-se proporcionalmente, dando mais força às conclusões de pesquisa a partir dos contrastes sistemáticos entre os diferentes grupos de inquiridos controlados na investigação;

(b) retirada, como critério para a seleção de indivíduos para o inquérito, da necessidade simétrica de, por um lado, os informantes com alta escolaridade terem ocupação profissional autônoma/livre e de, por outro lado, os menos escolarizados trabalharem na agricultura ou em atividade que não faça utilização da escrita. Justificativa: as primeiras tentativas antes dessa mudança implicaram morosidade acentuada na condução da pesquisa de campo, dada a dificuldade de localização desses perfis, especialmente dos perfis associados a trabalhos na agricultura ou a atividades que não façam utilização da escrita; e

(c) substituição da conversa livre como um dos procedimentos metodológicos pelo instrumento de coleta denominado "Temas para Discursos Semidirigidos", em sua versão do ano de 2001 concebida pelo Atlas Linguístico do Brasil (ALiB, 2014), aplicado após alterações. Justificativa: mais facilidade na uniformização do procedimento para todos os participantes. Isso dito, é pertinente acrescentar que esse instrumento foi utilizado, entretanto, apenas como facilitador ao engajamento interacional entre inquirido e inquiridor, com vistas à aplicação central de um questionário semântico-lexical e um questionário fonético-fonológico, embora essa interação preliminar também

tenha sido gravada com cuidado metodológico e a abordagem a seus dados possa ser objeto de outro trabalho.

1.3 Atividades intermediárias da pesquisa no Brasil

Até os primeiros meses do ano de 2016, as principais tarefas executadas resumem-se a: (I) uma edição incidente nos instrumentos metodológicos do ALiB (2014); (II) um conjunto de reflexões críticas sobre a aplicação do questionário semântico-lexical adotado; e (III) à condução da primeira de duas fases de coleta de dados.

(I)

Os instrumentos de coleta do ALiB (2014) que modificamos são versões do ano de 2001. São eles: a Ficha do Informante (FdI), os Temas para Discursos Semidirigidos (TDS), o Questionário Semântico-Lexical (QSL) e o Questionário Fonético-Fonológico (QFF). A nossa edição foi extensiva e detalhada sobretudo no QSL e no QFF. Sinalizamos as versões (que, via edição, assim derivamos) através da inserção da letra eme — como mnemônico para "modificado(a)(s)" — sobrescrita à direita dos nomes desses instrumentos: FdI^{m}, TDS^{m}, QSL^{m} e QFF^{m}.

A edição foi implementada na fase de preparação à pesquisa de campo. Essencialmente, as modificações buscaram melhorar enunciados e eliminar perguntas consideradas improdutivas após um teste preliminar, bem como acrescentar um conjunto de imagens representativas de referentes semânticos^{+} tematizados nas questões, para assim haver um aumento no índice de responsividade dos informantes. Acessoriamente, símbolos de orientação ao inquiridor foram adicionados nos instrumentos, visando a otimizar seu desempenho na condução da coleta de dados.

(II)

Em relação ao QSL^{m}, desenvolvemos um conjunto de ponderações referentes aos seguintes temas: (i) identificação de processos psicossociais a serem, na medida do possível, evitados na elicitação^{+}: confusão, equívoco, precipitação, insegurança, dúvida, desconhecimento, esquecimento e incompreensão; (ii) constatação, como um problema de pesquisa, de que

mesmo enunciados constituídos por palavras simples/comuns/recorrentes são não raramente objeto de um dos processos psicossociais citados; (iii) hipótese para esse problema: trata-se de dificuldade de categorização mental a partir da oitiva dos enunciados, e essa seria uma questão anterior e mais elementar do que os processos psicossociais elencados; (iv) necessidade de uso, para o maior número possível de questões do QSL^{m}, de imagens que exibam uma representação dos referentes semânticos que os enunciados tematizam, como medida de fácil implementação e adoção orientada a auxiliar a superação dos problemas mencionados.

(III)

Quanto à execução da primeira de duas fases de levantamento *in situ* de dados, conduzida em cinco das dez localidades da rede de pontos, os instrumentos metodológicos que foram utilizados são os já mencionados, ou seja: (A) a FdI^{m}; (B) os TDS^{m}; (C) o QSL^{m}; e (D) o QFF^{m}.

(A)

A FdI^{m} é formada por 62 campos de coleta de informações extralinguísticas, informações essas que são importantes ao empreendimento explicativo de influências externas a fatos linguísticos. No tocante ao preenchimento de tais campos, uma parte dele é baseada na observação empírica do pesquisador, e outra se dá por questionamento direto ao entrevistado. Relativamente à natureza dos dados, eles dividem-se em duas classes: (a) uma "situacional" (em alusão à circunstância do inquérito); e outra (b) "extrassituacional".

(a) classe situacional: (i) característica psicológica; (ii) espontaneidade; (iii) cooperação; e

(b) classe extrassituacional: (i) ascendência étnica; (ii) religião; (iii) classe social; (iv) escolaridade; (v) pessoal; (vi) profissional; (vii) proficiência em idiomas; (viii) contato linguístico; (ix) contato com mídias; e (x) topodinâmica.

(B)

Os TDS^{m}, constituídos por enunciados concebidos para incitar a produção de quatro tipos textuais orais — 'relato pessoal', 'comentário', 'descrição' e 'relato não pessoal' —, foram aplicados; entretanto, apenas como facilitadores táticos a favor do estabelecimento de um engajamento interacional produtivo entre inquirido e inquiridor, com vistas à aplicação central do QSL^{m} e do QFF^{m}, embora essa interação preliminar também tenha sido gravada com cuidado metodológico e a abordagem a seus dados possa ser objeto de um futuro trabalho.

(C)

Os dados semântico-lexicais foram coletados com a aplicação do QSL^{m}, composto por 170 questões onomasiológicas^{+} orientadas à obtenção de formas semasiológicas^{+}. Muitas dessas perguntas foram acompanhadas por imagens representativas de referentes semânticos tematizados nas indagações, com a finalidade de se potencializar o entendimento do entrevistado acerca do que se perguntava. Os campos semânticos abarcados nesse questionário são: (i) acidentes geográficos; (ii) fenômenos atmosféricos; (iii) astros e tempo; (iv) atividades agropastoris; (v) fauna; (vi) corpo humano; (vii) ciclos da vida; (viii) convívio e comportamento social; (ix) religião e crenças; (x) jogos e diversões infantis; (xi) habitação; (xii) alimentação e cozinha; (xiii) vestuário e acessórios; e (xiv) vida urbana.

(D)

Por fim, fez-se uso do QFF^{m} para a elicitação de dados fonético-fonológicos, formado por 159 perguntas indutivas cujo propósito foi observar o comportamento tanto de vogais e *glides* — itens (i) a (v), a seguir —, quanto de consoantes — itens (vi) a (xiv): (i) ditongos decrescentes; (ii) vogais baixas postônicas; (iii) vogais altas (pretônicas e tônicas); (iv) vogais médias (pretônicas, tônicas e postônicas); (v) /A/ e /O/ adjacentes a consoante nasal no padrão '(C)_.C[+nasal]'; (vi) gerúndio; (vii) lateral palatal /ʎ/; (viii) nasal palatal /ɲ/; (ix) /L/ (em coda^{+} e como C2 em *onset* complexo); (x) /S/ (em coda e intervocálico no padrão 'V._V'); (xi) /R/ (em coda, em *onset* simples e como C2 em *onset* complexo); (xii) encontros consonantais interssilábicos (/f.t/, /p.n/ e /d.v/); (xiii) palatalização (da lateral /l/ e das oclusivas /t, d/ diante da vogal alta [i]); e (xiv) alternâncias ([b] *vs.* [v] em *onset* simples;

[k] *vs.* [kʊ̯]; [ʒɪ] *vs.* [ʒɐ̃ɲ] como sílaba final átona; [i] *vs.* [ɐ̃ɲ] pós-nuclear em sílaba final átona; e [ĩ] *vs.* [ũ] em sílaba inicial).

Nessa primeira fase de registros empíricos, as cinco primeiras localidades que foram visitadas são: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque e Sorocaba. O período abrangido estendeu-se do dia 15 de fevereiro de 2016 ao dia 21 de março do mesmo ano, com 40 dos 80 informantes tendo sido entrevistados.

1.4 Atividades intermediárias da pesquisa na Alemanha

Entre o dia 1º de abril de 2016 e o dia 31 de março de 2018, a condução do nosso estudo deu-se a partir do Setor de Romanística (orig. *Romanisches Seminar*) da Universidade de Kiel (orig. *Christian-Albrechts-Universität zu Kiel* - CAU), no âmbito de um estágio. No período, entre outras atividades conectadas à pesquisa, tomaram lugar sobretudo:

(a) a transcrição, a tabulação e a organização dos dados coletados na primeira fase de levantamento empírico;

(b) a análise de uma amostra da variação semântico-lexical exibida em Santana de Parnaíba, primeira das dez localidades visitadas;

(c) a segunda e última fase de coleta, entre o dia 23 de agosto de 2017 e o dia 17 de setembro do mesmo ano, com as cinco últimas localidades da rede de pontos visitadas (Itu, Porto Feliz, Tietê, Capivari e Piracicaba) e os 40 informantes restantes entrevistados, mediante a aplicação dos seguintes instrumentos metodológicos, sob as mesmas condições metodológicas da primeira fase de pesquisa de campo: a Ficha do Informante^{m} (FdI^{m}), os Temas para Discursos Semidirigidos^{m} (TDS^{m}), o Questionário Semântico-Lexical^{m} (QSL^{m}) e o Questionário Fonético-Fonológico^{m} (QFF^{m}); e

(d) um treinamento no *software* de cartografia utilizado pelo grupo de pesquisadores da CAU.

A seção "Nossa pesquisa", neste mesmo tomo, trará mais detalhes sobre essas atividades.

1.5 Progressos realizados e resultados obtidos

Os principais produtos alcançados pela pesquisa relacionam-se com: (I) a metodologia; (II) a constituição de um conjunto de *corpora* (integralmente disponibilizado nesta obra); (III) a elaboração de um atlas linguístico (ou, dito de forma mais estrita, de um atlas semântico-lexical e de outro fonético-fonológico); e (IV) uma análise sobre as contrapartes (a)fonéticas^{+} do arquifonema^{+} /R/ em coda silábica. Complementarmente, fizemos: (V) a testagem de uma hipótese específica; (VI) uma compilação de metaplasmos^{+}; e (VII) algumas notas suplementares.

(I)

Em relação à metodologia, nosso estudo contribui sobretudo com: (A) a criação do conceito 'cabeça de relevância'; (B) uma revisão contingencial da técnica thuniana de entrevista; e (C) a elaboração de um sistema de *labels* (etiquetas) para *corpora* semântico-lexicais primários.

(A)

A noção 'cabeça de relevância' (CR) é dependente de um conjunto de outros conceitos que desenvolvemos na pesquisa com o objetivo de dar transparência ao *status* qualitativo das variantes obtidas mediante a aplicação do Questionário Semântico-Lexical^{m} (QSL^{m}). De maneira sinóptica, a CR nada mais é do que a variante que assume a primeira posição num ordenamento de relevância aplicado às formas semasiológicas covariantes^{+} enunciadas por um mesmo informante. A determinação de qual entre as covariantes é alçada à primeira posição é performada ou pelo próprio falante, espontaneamente, ou pelo inquiridor, através do levantamento de perguntas ao entrevistado para esse objetivo. A CR relaciona-se com os próximos dois itens, (B) e (C).

(B)

Durante a pesquisa, para a aplicação do QSL [m], desenvolvemos um conjunto de procedimentos que denominamos 'técnica de entrevista orientada à contundência responsiva/anuente' (TE). Trata-se de um expediente metodológico plasmado a partir diretamente da 'técnica dos três passos' da Dialetologia Pluridimensional (THUN, 2000d), cujo funcionamento nuclear consiste nas etapas "perguntar" (para obtenção de respostas espontâneas), "insistir" (para obtenção de variantes coexistentes) e "sugerir" (para verificação do conhecimento passivo do inquirido a respeito de covariantes apresentadas pelo inquiridor antes não nomeadas pelo informante). Quanto à TE, seus procedimentos preveem adicionalmente o 'controle de sugestões[+]' e a atribuição de 'estatutos qualitativos' às formas semasiológicas enunciadas pelo entrevistado, além de a técnica implicar o estabelecimento da 'cabeça de relevância' e do ordenamento lexical aludidos no item anterior, (A).

(C)

Relativamente ao sistema de etiquetas que desenvolvemos para o *corpus* semântico-lexical primário, ele identifica o estatuto qualitativo de cada forma semasiológica, graças à TE mencionada no item anterior, (B). Essas etiquetas são ou primárias, ou secundárias. As primárias chamam-se 'FP' (forma espontânea positiva[+]), 'FN' (forma espontânea negativa[+]), 'FI' (forma resultante de insistência[+]), 'FA' (forma sugerida aceita[+]) e 'FR' (forma sugerida rejeitada[+]). Já as secundárias denominam-se 'Ret' (retificação[+]), 'Hes' (hesitação[+]) e 'Sus' (suspeição de designação[+]).

(II)

Os grupos de *corpora* criados pela pesquisa estão organizada e integralmente disponibilizados neste trabalho. Quer-se com isso que os dados possam ser utilizados também por outros linguistas interessados no português paulista (ou "dialeto[+] caipira[+]") a partir de bases empíricas. Os *corpora* são formados de: (A) *corpora* semântico-lexicais; (B) *corpora* fonético-fonológicos; (C) um *corpus* metalinguístico[+]-etnográfico; e, por fim, (D) *corpora* extralinguísticos. Respectivamente, cada item segue discriminado adiante.

(A)

A coleção de dados semântico-lexicais é formada pelos seguintes *corpora:* (i) formas semasiológicas com informação fonética (*corpus* primário); (ii) cabeças de relevância; (iii) contagem das lexias^{+}; e (iv) normas^{+} semântico-lexicais absolutas.

Para a constituição dessa coleção, o Questionário Semântico-Lexical^{m} (QSL^{m}) foi utilizado como instrumento metodológico de coleta, composto por 170 questões onomasiológicas orientadas à obtenção de formas semasiológicas. Muitas dessas perguntas foram acompanhadas por imagens representativas de referentes semânticos tematizados nas indagações, com a finalidade de se potencializar o entendimento do entrevistado acerca do que se perguntava. Os campos semânticos abarcados nesse questionário são: (i) acidentes geográficos; (ii) fenômenos atmosféricos; (iii) astros e tempo; (iv) atividades agropastoris; (v) fauna; (vi) corpo humano; (vii) ciclos da vida; (viii) convívio e comportamento social; (ix) religião e crenças; (x) jogos e diversões infantis; (xi) habitação; (xii) alimentação e cozinha; (xiii) vestuário e acessórios; e (xiv) vida urbana.

(B)

O grupo de *corpora* fonético-fonológicos, por seu turno, divide-se em: (i) fones^{+} (*corpus* primário); e (ii) normas fonético-fonológicas absolutas. A constituição do *corpus* fonético-fonológico primário deu-se mediante a aplicação do Questionário Fonético-Fonológico^{m} (QFF ^{m}), composto por 159 perguntas indutivas cujo propósito é observar o comportamento tanto de vogais e *glides* — itens (a) a (e) a seguir —, quanto de consoantes — itens (f) a (n): (a) ditongos decrescentes; (b) vogais baixas postônicas; (c) vogais altas (pretônicas e tônicas); (d) vogais médias (pretônicas, tônicas e postônicas); (e) /A/ e /O/ adjacentes a consoante nasal no padrão '(C)_.C[+nasal]'; (f) gerúndio; (g) lateral palatal /ʎ/; (h) nasal palatal /ɲ/; (i) /L/ (em coda e como C2 em *onset* complexo); (j) /S/ (em coda e intervocálico no padrão 'V._V'); (k) /R/ (em coda, em *onset* simples e como C2 em *onset* complexo); (l) encontros consonantais interssilábicos (/f.t/, /p.n/ e /d.v/); (m) palatalização (da lateral /l/ e das oclusivas /t, d/ diante da vogal alta [i]); e (n) alternâncias ([b] *vs.* [v] em *onset* simples; [k] *vs.* [kʊ]; [ʒɪ] *vs.* [ʒẽɲ] como sílaba final átona; [i] *vs.* [ẽɲ] pós-nuclear em sílaba final átona; e [ĩ] *vs.* [ũ] em sílaba inicial).

(C)

Quanto ao *corpus* metalinguístico-etnográfico, ele consiste de comentários cuja natureza o título da coleção sugere e os quais foram espontaneamente enunciados pelos informantes. Os comentários metalinguísticos — itens (i) a (x) — e etnográficos — itens (xi) a (xiv) — surgiram no bojo da execução do QSL[m] e ajudam a revelar razões que respondem por determinadas escolhas lexicais, e não por outras, por parte dos indivíduos. Ambos os grupos de asserções foram assim classificados: (i) sinônimos; (ii) definições semânticas; (iii) jocosidade associada; (iv) informação diageracional; (v) *status* temporal de formas: antigas *vs.* recentes; (vi) *status* popular *vs.* técnico de formas; (vii) percepção sobre frequência; (viii) percepção de caída em desuso; (ix) juízos de valor negativo; (x) juízos de "certo" *vs.* "errado"; (xi) tabus; (xii) lendas; (xiii) crenças religiosas; e (xiv) superstições laicas.

(D)

Por fim, temos os *corpora* extralinguísticos, elicitados com a Ficha do Informante [m]. Importantes ao empreendimento explicativo das influências externas sobre fatos linguísticos, essas coleções distribuem-se sob as seguintes denominações: (i) informações básicas; (ii) domicílios e filiação; (iii) línguas; (iv) trabalho, religião e grau escolar; (v) exposição às mídias escrita e sonora; (vi) exposição à mídia televisiva e a eventos esportivos e culturais; e (vii) entrevista.

(III)

Como principal resultado obtido pela pesquisa, elaboramos um atlas linguístico, que abarca na verdade um atlas semântico-lexical e outro fonético-fonológico. Além da dimensão *diatópica* intrínseca das cartas linguísticas, elas permitem ainda a visualização de outras três dimensões associadas aos fatos de língua coletados: a *diastrática escolar*, a *diageracional* e a *diagenérica*.

As cartas linguísticas organizam-se primariamente em três *categorias*:

(a) quatro (grupos de) covariantes — 4(G)C;
(b) três (grupos de) covariantes — 3(G)C; e
(c) duas (ou dois grupos de) covariantes — 2(G)C.

Essas denominações fazem alusão ao número de covariantes de maior frequência entre as demais. Mais especificamente, pertencem à categoria 4(G)C os casos em que há pelo menos quatro covariantes como itens mais frequentes sem empates quantitativos com as covariantes a partir da quinta mais frequente. À categoria 3(G)C, pertencem os casos em que há pelo menos três covariantes como itens mais frequentes sem empates quantitativos com as covariantes a partir da quarta mais frequente. O mesmo *mutatis mutandis* em relação à última categoria, 2(G)C.

Sob um nível secundário de organização das cartas, cada uma das categorias compreende três *séries*, abstratamente denominadas "CxGy", "CxWz" e "WzGy". Tais denominações aludem conceitualmente a distintos grupos extralinguísticos exibidos nas cartas linguísticas e, ao mesmo tempo, a diferentes cruzamentos dos dados associados a esses grupos. Esses cruzamentos são visualizados tanto na cartografia propriamente dita, quanto nos gráficos estatísticos presentes nas cartas.

Na composição dos nomes das séries, 'C' refere-se a 'classe escolar'; 'G', a 'geração'; e 'W', a 'gênero de registro/social do indivíduo'+. Ainda, observa-se a presença de 'x', 'y' e 'z', as quais são variáveis lógicas. A variável 'x' é saturada ou pelo parâmetro 'a' de 'classe escolar alta', ou 'b' de 'classe escolar baixa'. Já a variável 'y' é saturada ou pelo parâmetro 'II' de 'geração mais velha', ou 'I' de 'geração mais nova'. Finalmente, a variável 'z' é saturada ou pelo parâmetro 'f' de 'indivíduo feminino', ou 'm' de 'indivíduo masculino'. Ou seja, cada série tem quatro parametrizações distintas:

(i) CxGy → CaGII, CaGI, CbGII e CbGI;
(ii) CxWz → CaWf, CaWm, CbWf e CbWm; e
(iii) WzGy → WfGII, WfGI, WmGII e WmGI.

Essas parametrizações podem ser assim lidas: (i) CaGII: classe 'a' da geração 'II'; CaGI: classe 'a' da geração 'I'; CbGII: classe 'b' da geração 'II'; e CbGI: classe 'b' da geração 'I'). ● (ii) CaWf: classe 'a' feminina; CaWm: classe 'a' masculina; CbWf: classe 'b' feminina; e CbWm: classe 'b' masculina. ● (iii) WfGII: mulheres da geração 'II'; WfGI: mulheres da geração 'I'; WmGII: homens da geração 'II'; e WmGI: homens da geração 'I'.

Uma série abarca a correlação diastrático-geracional, uma outra caracteriza-se pela correlação diastrático-diagenérica e uma última série abrange a correlação diagenérica-diageracional. Dito de outro modo, o leitor estará apto a observar nas cartas linguísticas a correlação entre fatos linguísticos e fatores extralinguísticos distribuídos no plano geográfico (dimensão *diatópica* = eixo horizontal) e no plano social (dimensões *diastrática escolar*, *diageracional* e *diagenérica* = eixo vertical).

Por esse motivo, os atlas são pluridimensionais, consoante ao programa da Dialetologia Pluridimensional (RADTKE e THUN, 1996).

(IV)

Para um engajamento analítico, escolhemos abordar as contrapartes (a)fonéticas do arquifonema /R/ no português paulista (ou "dialeto caipira"), com base em nosso *corpus* fonético primário. Para esse efeito, selecionamos quatro situações estruturais (ou variáveis fonético-fonológicas) do arquifonema em apreço: (i) coda externa de desinência infinitiva seguida por pausa; (ii) coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa; (iii) coda interna em sílaba tônica; e (iv) coda interna em sílaba átona.

A partir dos dados empíricos representativos dessas situações estruturais, montamos 12 cartas fonético-fonológicas especiais, a fim de visualizar o comportamento dos dados tanto no plano espacial, quanto no plano social, e daí iniciar as ilações. Nesse sentido, as ponderações se pautaram preliminarmente em três tópicos, os quais são acerca: (a) da menor variação (registro de uma e duas variantes); (b) da maior variação (registro de três e quatro variantes); e (c) de contrastes quantitativos. Lá, os itens (a) e (b) estão associados mais diretamente à cartografia propriamente dita, e o item (c), ao gráfico estatístico integrado às cartas.

Quanto às variantes cartografadas, são elas: (i) a aproximante retroflexa, [ɻ] (conhecida como "erre retroflexo" ou ainda como "erre caipira"); (ii) o tepe (orig. *tap*) alveolar, [ɾ] (conhecido também como "*flap*", "erre fraco", "erro brando", "erre simples" ou ainda "vibrante simples"); (iii) o apagamento, [Ø] (*i.e.*, a ausência de fone); e, finalmente, (iv) a vibrante alveolar, [r] (conhecida como "vibrante múltipla" ou ainda como "erre forte").

(V)

A hipótese específica que levantamos e testamos foi:

> *Os jovens entre 18 e 36 anos de idade do Médio Tietê (especialmente aqueles com nível escolar alto[5]) estão produzindo a covariante [ɻ] (a aproximante retroflexa[6]) como contraparte fonética do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa sob uma frequência relativa inferior àquela a ser observada entre os falantes mais velhos a partir de 55 anos de idade (especialmente aqueles com nível escolar baixo[7]).*

Como resultado, na realidade incontestável dos fatos concretos observados, a fala dos jovens do interior paulista evidenciou-se, na verdade, mais albergadora da variante [ɻ] do que a fala dos mais velhos; e o grau dessa constatação tornou-se ainda mais eloquente na mesma medida em que soubemos que a variante [ɻ] fez-se presente ao longo de toda a nossa rede de pontos, com frequência relativa de 57% entre os jovens com nível escolar alto e 62% entre os jovens com nível escolar baixo (contra 45% entre os mais velhos com nível escolar alto e 30% entre os mais velhos com nível escolar baixo).

(VI)

Outrossim, fizemos uma compilação dos metaplasmos mais comuns, ou processos fonológicos, a fim de que ela possa servir como auxílio ao leitor mais interessado na tarefa de identificar os processos fonológicos visíveis tanto no *corpus* semântico-lexical primário (uma vez que ele apresenta transcrição fonética parcial), criado via QSL^{m}, quanto no *corpus* fonético primário, formado por aplicação do QFF^{m}.

Os metaplasmos foram reunidos em cinco categorias principais; "por acréscimo" — itens (i) a (iii); "por subtração" — itens (iv) a (vii); "transformações gerais" — itens (viii) a (xxiii); "transformações via assimilação" — itens (xxiv) a (xxx); e, ainda, "deslocamentos" — itens (xxxi) a (xxxiii): (i) prótese; (ii) epêntese; (iii) paragoge; (iv) aférese; (v) síncope; (vi) apócope; (vii) crase; (viii) dissimilação; (ix) vocalização; (x) consonantização; (xi) monotongação; (xii) ditongação; (xiii) apofonia; (xiv) metafonia; (xv) assibilação;

[5] A partir de estudos universitários, completos ou incompletos.
[6] Conhecida como "erre retroflexo" ou ainda como "erre caipira".
[7] No escopo de abrangência entre analfabetismo e ensino médio incompleto.

(xvi) degeneração; (xvii) rotacismo; (xviii) lambdacismo; (xix) desnasalização; (xx) sonorização; (xxi) palatização; (xxii) despalatalização; (xxiii) nasalização não assimilativa; (xxiv) geral progressiva; (xxv) geral regressiva; (xxvi) nasalizante; (xxvii) palatalizante; (xxviii) labializante; (xxix) sonorizante; (xxx) de harmonização vocálica; (xxxi) hipértese; (xxxii) metátese; e (xxxiii) hiperbibasmo.

(VII)

Também elaboramos um conjunto de ponderações complementares acerca: (A) do QSL^{m}; (B) de três tipos de informantes observados ao longo do inquérito; (C) das táticas responsivas postas em prática pelos informantes por ensejo do QSL^{m}; e (D) dos hábitos de consumo midiático desses indivíduos.

(A)

As considerações sobre o QSL ^{m} dividem-se em quatro partes: (i) identificação e caracterização de processos psicossociais ocasionalmente atuantes na elicitação (confusão, equívoco, precipitação, insegurança, dúvida, desconhecimento, esquecimento e incompreensão); (ii) constatação de que mesmo enunciados simples (constituídos por seleção lexical e gramatical comuns às regiões brasileiras) ensejam não raramente a ocorrência de algum dos processos psicossociais mencionados no item acima, (i); (iii) apresentação de um esboço de hipótese explicativa atinente ao problema referido no item prévio, (ii), relativo à problemática da categorização mental; e (iv) reforço enfático à adoção de imagens ilustrativas dos referentes semânticos tematizados nas questões do QSL^{m}, a fim de que a problemática citada nos itens precedentes seja mitigada.

(B)

Quanto aos três tipos de entrevistados observados na investigação de campo, a tipologia levou em consideração, como critério, os idioletos^{+} dos participantes da pesquisa. A ideia de base pressupõe um *spectrum* cujos polos extremos são representados, de um lado, pelo dialeto caipira rural e, de outro lado, pela variedade *standard* da língua portuguesa. Os tipos são o (i) caipira, (ii) o transicional e (iii) o estandardizado.

(C)

Ainda, ficou claro ao longo da pesquisa de campo que certos expedientes responsivos eram frequentemente perpetrados pelos inquiridos a propósito do QSL [m]. Ao registrá-los, dividimo-los em duas categorias, "gerais" — itens (i) a (iii); e "específicos" — itens (iv) a (xiii): (i) retificação retroativa; (ii) mescla entre lexias; (iii) metáfora; (iv) "diminutivação"; (v) "aumentativação"; (vi) perífrase; (vii) eufemismo; (viii) hiperonímia; (ix) metonímia; (x) analogia; (xi) palpite/dúvida sob entoação interrogativa; (xii) resposta evasiva não seguida por resposta assertiva; e (xiii) repetição.

(D)

Outro conjunto de informações complementares refere-se à presença das mídias na vida dos entrevistados. Para isso, uma quantificação foi feita, cujos dados são oriundos da aplicação da Ficha do Informante [m] e foram agrupados em três classes, (a) "seção de jornal/revista preferida"; (b) "programa de rádio preferido"; e (c) "programa de TV preferido":

(a) itens (i) a (xi) teriam o atributo relativo '+formal', e o restante, '-formal': (i) edital; (ii) ciência; (iii) cultura; (iv) tecnologia; (v) economia; (vi) profissional; (vii) artigo; (viii) página policial; (ix) editorial; (x) política; (xi) entrevista; (xii) humorístico; (xiii) esporte; (xiv) fofoca; (xv) entretenimento; (xvi) social; (xvii) carros antigos; (xviii) cotidiano; (xix) infantil; e (xx) variedades;

(b) itens (i) a (iii) teriam o atributo relativo '+formal', e o restante, '-formal': (i) noticiário geral; (ii) política; (iii) religioso; (iv) música; (v) esporte; (vi) entretenimento; e (vii) variedades; e

(c) itens (i) a (vi) teriam o atributo relativo '+formal', e o restante, '-formal': (i) noticiário; (ii) religioso; (iii) tecnologia; (iv) cultura; (v) documentário; (vi) *talk show*; (vii) novela; (viii) auditório; (ix) esporte; (x) culinária; (xi) decoração; (xii) série; e (xiii) filme.

Com os apontamentos nas páginas precedentes, oferecemos em linhas gerais alguns dos pontos mais relevantes do presente trabalho. A seguir, lidaremos com questões propriamente introdutórias da pesquisa de que se trata aqui.

2 INTRODUÇÃO

Em termos latos, este é um trabalho linguístico de base empírica e, em termos estritos, dialetológico pluridimensional. O objeto de estudo elegido é o vernáculo paulista, *i.e.*, a variedade linguística+ falada no interior do Estado de São Paulo. Desde sobretudo Amaral (1920), essa variedade também é conhecida por estudiosos, pesquisadores e interessados como "dialeto+ caipira+". Outros, salvo imprecisões ou nuances de abrangência geográfica, chamam-na de "dialeto paulista" (*e.g.* BOHN, 2017) ou "português paulista" (*e.g.* KEWITZ, 2011). A coleção de nomes, porém, não se limita a esses citados. Em nosso caso, vamos utilizar "português paulista" primariamente e "dialeto caipira" secundariamente.

Antes de expormos os detalhes da nossa pesquisa, mencionaremos a seguir alguns estudos de base empírica sobre a variedade linguística em questão, a partir de diferentes perspectivas teórico-metodológicas e concernentes a diferentes níveis ou aspectos linguísticos.

2.1 Estudos acerca do português paulista

Desde o pioneirismo do já referido Amaral (1920), seguido por Rodrigues (1974), o dialeto caipira do interior paulista vem sendo tema de estudos de base empírica a vários grupos de pesquisa, a partir de diversas perspectivas teórico-metodológicas da Linguística e acerca de diferentes níveis linguísticos. Entre esses grupos, encontram-se:

(a) o Projeto para História do Português Paulista (PHPP – Projeto Caipira)[1], primeira e segunda edições (2006-2017[2]), do qual participamos, sob a coordenação de Ataliba Teixeira de Castilho (2006-2010), Clélia Cândida Abreu Spinardi Jubran (2011-2014) e Manoel Mourivaldo Santiago-Almeida (2014-2017), e situado na área de Linguística Histórica do Português, com os

[1] Proj. Temático de Equipe FAPESP (procs. 06/55944-0 e 11/51787-5).
[2] Esse é o período como Proj. Temático de Equipe FAPESP. Todavia, o PHPP foi lançado originalmente em 1997 pelo Programa de Pós-Graduação em Filologia e Língua Portuguesa da USP, tal como Castilho (2011) reporta. Foi a partir desse projeto que, já em 1998 (por ocasião do I Seminário do PHPP), surgiu a proposta para o Projeto para História do Português Brasileiro (PHPB).

objetivos realizados de constituir e disponibilizar *corpora* diacrônicos+, bem como de analisá-los em quatro eixos de estudos: (i) variação e mudança gramaticais; (ii) concordância e classe de palavras; (iii) formação da variedade culta e da difusão da variedade popular na região do Médio Tietê; e (iv) gêneros discursivos e processos de construção textual;

(b) o Projeto Amostra Linguística do Interior Paulista – Projeto ALIP[3] (2003-2007), sob coordenação de Sebastião Carlos Leite Gonçalves, que envolveu localidades do noroeste do interior paulista (São José do Rio Preto, Bady Bassitt, Cedral, Guapiaçu, Ipiguá, Mirassol e Onda Verde), com interesse não em fenômenos particulares, mas sim em "captar o máximo possível do dinamismo linguístico da região" (RUBIO e GONÇALVES, 2012:1005). O Projeto ALIP gerou o banco de dados Iboruna, composto por duas amostras orais: (i) a Amostra Comunidade (ou Amostra Censo), composta por material oral gravado sob metodologia sociolinguística laboviana; e (ii) a Amostra de Interação Dialógica, constituída por falas secretamente gravadas em situações interacionais livres;

(c) o Projeto Filologia Bandeirante (1997-2003)[4], coordenado por Heitor Megale e interessado em buscar traços de língua antiga (arcaísmos) nas trilhas das bandeiras paulistas dos séculos XVII e XVIII, e por isso cobriu os territórios dos Estados de São Paulo, Minas Gerais, Goiás e Mato Grosso; e

(d) o Sistema Corpora do Português Falado de São Paulo (Sistema CorPor, 1972-1973), projeto coordenado por Zilda Maria Zapparoli e interessado em contribuir para o processamento do português paulista (falado em Campinas, Itu e capital) nos âmbitos fonético+ e ortográfico, a partir de entrevistas e conversas com 216 informantes, selecionados por "sexo", escolaridade, faixa etária, região de origem e nível socioeconômico. O Sistema CorPor produziu

[3] Aux. à Pesq. FAPESP, registrado inicialmente sob o título "O português falado na região de São José do Rio Preto: constituição de um banco de dados anotado para o seu estudo" (proc. 03/08058-6).
[4] Proj. Temático de Equipe FAPESP (proc. 96/01265-0).

> recursos linguísticos computacionais (Bases de Dados Textuais, Léxico de Frequência, Léxico de Junturas e Dicionário de Junturas) e permite acesso aos *corpora* de maneira interativa (www.corpor.fflch.usp.br).

Outrossim, inúmeros estudos (vários dos quais vinculados aos grupos acima indicados) sobre o dialeto caipira estão sendo desenvolvidos, cujas instituições sedes são sobretudo a Universidade Estadual Paulista (UNESP), a Universidade de São Paulo (USP) e a Universidade Estadual de Campinas (UNICAMP).

São investigações de iniciação científica (IC), mestrado (ME), doutorado (DO) e pós-doutorado (PDO). De maneira breve e seletiva, listaremos — nas Tabelas 1-3 adiante — algumas delas de base empírica oral já finalizadas, agrupadas por nível linguístico enfocado e organizadas cronologicamente a partir da mais recente.

Em tais tabelas, apresentaremos: (i) o nome dos autores; (ii) o ano de publicação (quando pesquisa de ME ou DO) ou de conclusão (quando IC ou PDO) do estudo; (iii) o nível da investigação (IC, ME, DO ou PDO); (iv) a(s) área(s) ou teoria(s) utilizada(s); (v) a sigla ou o acrônimo da instituição sede; e, por fim, (vi) o nome do(a) orientador(a).

Note-se ainda que: (α) em caso de obra publicada (ref. a pesquisa de ME e DO), a referência bibliográfica encontrar-se-á na seção "Referências", no Tomo VI; e (β) a Sociolinguística Variacionista (também conhecida como Teoria da Variação e Mudança Linguísticas ou Sociolinguística Laboviana) será referida nas tabelas como "Sociolinguística", por simplicidade.

Isso dito, comecemos com os trabalhos que abordaram o nível fonético/fonológico^{+} do português paulista (PP), arrolados na Tabela 1, na próxima página.

Autor(a)	Ano	Pesq.	Área(s)/Teoria(s)	Inst. sede	Orientador(a)
Carreão	2018	ME	SV	UNICAMP	Emilio Gozze Pagotto
Picinato	2018	DO	SV	UNESP	Rosane de Andrade Berlinck
Carmo	2013	DO	SV	UNESP	Luciani Ester Tenani
Picinato	2013	ME	SV	UNESP	Rosane de Andrade Berlinck
Leal	2012	DO	FG & SV	USP	Raquel Santana Santos
Aronne	2010	ME	Dial. & SV	USP	Mário Eduardo Viaro
Marcato[5]	2010	IC	SV	UNESP	Luciani Ester Tenani
Carmo	2009	ME	SV & FA	UNESP	Luciani Ester Tenani
Garcia	2009	ME	Dial. (& Fil.)	USP	M. M. Santiago-Almeida
Ramos	2009	ME	SV & FM & FS & FA	UNESP	Luciani Ester Tenani
Pires	2008	ME	SV & Dial. (& Fil.)	USP	M. M. Santiago-Almeida
Leal	2006	ME	FA & GT & TP & FM	USP	Raquel Santana Santos
Leme	1994	ME	SV	UNICAMP	Luiz Carlos Cagliari

Dial. = Dialetologia; Fil. = Filologia; FA = Fonologia Autossegmental; FG = Fon. Gerativista; FS = Fon. da Sílaba; FM = Fon. Métrica; GT = Geometria de Traços; SV = Sociolinguística; TP = Teoria Prosódica

Tabela 1: Pesquisas empíricas de base oral com enfoque no nível fonético/fonológico do PP (seleção)

Os estudos citados na Tabela 1 (bem como nas Tabelas 2 e 3, a seguir) ou tomaram *corpora* anteriormente constituídos — como aqueles organizados pelos grupos mencionados mais acima, nos itens de (a) a (d) —, ou fizeram um levantamento próprio de dados orais.

Quanto aos *loci* investigados pelas pesquisas constantes da Tabela 1, eles são: Louveira (CARREÃO, 2018); Sales Oliveira (PICINATO, 2018); o noroeste paulista (São José do Rio Preto, Bady Bassitt, Cedral, Guapiaçu, Ipiguá, Mirassol e Onda Verde) (CARMO, 2013); Sales Oliveira e Orlândia (PICINATO, 2013); Capivari e Campinas (LEAL, 2012); São José do Rio Preto (ARONNE, 2010; Fernanda Marcato, 2010; CARMO, 2009; RAMOS, 2009); Capivari (GARCIA, 2009; LEAL, 2006); e Piracicaba (PIRES, 2008; LEME, 1994).

Em relação aos níveis morfológicos^{+} e (morfos)sintáticos^{+} do dialeto caipira, referimo-nos a alguns estudos na Tabela 2, na página a seguir.

[5] Proj. "A realização das vogais mediais de prefixos na variedade do noroeste paulista" (2009-2010, proc. FAPESP 09/10684-9).

Autor(a)	Ano	Pesq.	Área(s)/Teoria(s)	Inst. sede	Orientador(a)
Gomes[6]	2016	IC	SV	UNESP	Gisele Cássia de Sousa
Kuhlmann[7]	2011	IC	TG & SV	USP	Maria Célia P. L. Hernandes
Oliveira[8]	2010	IC	TG	USP	Maria Célia P. L. Hernandes
Rubio	2008	ME	SV	UNESP	Sebastião Carlos L. Gonçalves
Pereira	2004	ME	SV & LF	USP	Ângela Cecília de S. Rodrigues

LF = Linguística Funcional; SV = Sociolinguística; TG = Teoria da Gramaticalização

Tabela 2: Pesquisas empíricas de base oral com enfoque no nível morfológico/(morfos)sintático do PP (seleção)

Os *loci* das investigações arroladas na Tabela 2 são: o noroeste paulista (Aline Bianca dos Santos Gomes, 2016); Capivari (Mariana Corallo Mello de Azevedo Kuhlmann, 2011); o Estado de São Paulo (Anna Karolina Miranda Oliveira, 2010); São José do Rio Preto (RUBIO, 2008); e as trilhas das bandeiras paulistas nos Estados de São Paulo e Minas Gerais (PEREIRA, 2004).

Na Tabela 3, na página adiante, aludimos a estudos científicos que abordaram variados níveis linguísticos do português paulista.

[6] Proj. "A alternância 'mim/eu' como sujeito de orações infinitivas introduzidas por 'para' no português do interior paulista: um caso de variação estável ou de mudança em curso" (2016, proc. FAPESP 16/04771-0).
[7] Proj. "O dialeto caipira na contemporaneidade de São Paulo: desaparecimento ou descaracterização?" (2010-2011, proc. FAPESP 10/11483-4).
[8] Proj. "Antes de entrar no elevador, verifique se o mesmo encontra-se parado neste andar: a mudança gramatical do item 'mesmo' no português paulista" (2009-2010, proc. FAPESP 09/07797-6).

Autor(a)	Ano	Pesq.	Área(s)/Teoria(s)	Inst. sede	Orientador(a)
Rosignoli	2017	ME	FE & FP	USP	Flaviane R. F. Svartman
Garcia	2015	DO	Dial.	USP	Waldemar Ferreira Netto
Brito	2013	ME	SV & Geol.	USP	M. M. Santiago-Almeida
Barbosa[9]	2012	PDO	GSC	USP	Ataliba T. de Castilho
Gomes[10]	2010	PDO	SG	USP	Maria Ap.ª C. R. T. Morais
Kewitz[11]	2010	PDO	GSC & LC & SV	UNICAMP	Ataliba T. de Castilho
Mota	2007	ME	SV (& Fil.)	USP	M. M. Santiago-Almeida
Romero	2007	ME	Dial. & Geol. & SV & Lex.	USP	M. M. Santiago-Almeida

Dial. = Dialetologia; FE = Fonologia Entoacional; FP = Fonologia Prosódica; Fil. = Filologia; Geol. = Geolinguística; GSC = Gramática de Sistemas Complexos; Lex. = Lexicografia; LC = Linguística Cognitiva; SG = Semântica de Graus; SV = Sociolinguística

Tabela 3: Pesquisas empíricas de base oral com enfoque em diversos níveis linguísticos do PP (seleção)

Os estudos listados na Tabela 3 dizem respeito a aspectos: (i) semântico$^+$-lexicais$^+$ ou semânticos (BRITO, 2013; Ana Paula Quadros Gomes, 2010; ROMERO, 2007); (ii) prosódicos$^+$ (ROSIGNOLI, 2017; GARCIA, 2015); e (iii) vários em paralelo (Verena Kewitz, 2010; MOTA, 2007; Edilaine Buin Barbosa, 2012).

Os *loci* de onde os dados foram colhidos são: a região do Médio Tietê (Santana de Parnaíba, Pirapora do Bom Jesus, Porto Feliz, Tietê, Capivari, Itu e Piracicaba) e o norte de Portugal (Braga, Bragança e Vila Real) (GARCIA, 2015); Piracicaba (BRITO, 2013); Santana de Parnaíba (MOTA, 2007); e Itu (ROMERO, 2007).[12]

Até aqui, fizemos referência a pesquisas de base empírica oral que investigaram os diversos níveis linguísticos do dialeto caipira. A seguir, falaremos dos atlas linguísticos desenvolvidos no Estado de São Paulo.

[9] Proj. "Diacronia da concordância no português paulista: fatores discursivos e semânticos" (2010-2012, proc. FAPESP 09/54938-4).
[10] Proj. "Modificação por advérbios: o caminho das pedras para uma semântica basilar dos adjetivos do português falado em São Paulo" (2009-2010, proc. FAPESP 08/58145-6).
[11] Proj. "A representação do espaço no português paulista numa abordagem sociocognitiva" (2008-2010, proc. FAPESP 07/59707-5).
[12] Os demais estudos referidos na Tabela 3 não especificam claramente seus *loci* paulistas.

2.2 Atlas linguísticos no Estado de São Paulo

Na década de 1980, o Projeto para o Atlas Linguístico do Estado de São Paulo (ALESP) foi iniciado pelos pesquisadores do Atlas Linguístico do Brasil (ALiB), sob coordenação de Pedro Caruso da UNESP/Assis, chegando a concluir a fase de coleta de dados. No entanto, como Aguilera e Biolada (2015:149) reportam, os dados desapareceram, e até hoje não se sabe onde eles estão, o que provocou à época a interrupção da execução do projeto.

Quanto aos projetos de atlas paulistas já executados, as localidades linguisticamente cartografadas até o momento (2018) são: (I) a capital estadual (a cidade de São Paulo); (II) a região do Grande ABC (municípios de Santo André, São Bernardo do Campo, São Caetano do Sul, Diadema, Mauá, Ribeirão Pires e Rio Grande da Serra); (III) o litoral norte do Estado (Caraguatatuba, Ilha bela, São Sebastião e Ubatuba); e (IV) a região norte do Alto Tietê (Guarulhos, Arujá, Santa Isabel, Mairiporã e Nazaré Paulista).

(I)

A cartografia linguística respeitante à cidade de São Paulo situa-se em relação às outras capitais brasileiras[13] e é de lavra dos realizadores do ALiB, importante empreendimento dialetológico e geolinguístico nacional com um primeiro volume de cartas já publicado (ALiB, 2014). As cartas dadas a lume recobrem uma gama de fenômenos dialetais fonético-fonológicos, semântico-lexicais e morfossintáticos do português brasileiro, e informações sociolinguísticas ("diassexuais", diageracionais^{+} e diastráticas^{+}) acompanham a obra. O ALiB é atualmente o trabalho que mais influencia os estudos dialetológicos e geolinguísticos brasileiros.

(II)

A região do Grande ABC foi investigada e cartografada a propósito do Atlas Semântico-Lexical da Região do Grande ABC (ASL-GrABC, 2007). Nesse estudo, a versão de 2001 do Questionário Semântico-Lexical (QSL) do ALiB foi aplicada a indivíduos femininos e masculinos de duas faixas etárias, e os dados semântico-lexicais tiveram sua frequência de ocorrência e sua distribuição registradas.

[13] Exceto Palmas, Estado de Tocantins, e Brasília, Distrito Federal, "em virtude de serem cidades novas e, consequentemente, com uma norma linguística ainda em fase de consolidação" (ALiB, 2014:11).

(III)

O litoral norte do Estado também foi objeto de um atlas, do Atlas Semântico-Lexical de Caraguatatuba, Ilhabela, São Sebastião e Ubatuba (ASL-LNSP, 2010). Semelhantemente ao trabalho citado no item anterior, o objetivo foi abordar o nível semântico-lexical a partir do QSL do ALiB, em versão, porém, de 2000, bem como verificar a frequência de ocorrência das variantes+ e sua distribuição. Para isso, jovens e velhos, tanto homens quanto mulheres, com baixa escolaridade foram entrevistados.

(IV)

Por fim, a região norte do Alto Tietê foi geolinguisticamente pesquisada mediante inquérito junto a informantes jovens e velhos, masculinos e femininos, para o Atlas Semântico-Lexical da Região Norte do Alto Tietê (ASL-ReNAT, 2012). Sob interesses investigativos essencialmente como os da pesquisa mencionada no item III, os instrumentos metodológicos utilizados foram tanto o QSL do ALiB, quanto o QSL do Projeto Estudo Sociogeolinguístico do Município de São Paulo, do Grupo de Pesquisa em Dialetologia e Geolinguística (GPDG) da USP.

Feita essa breve introdução dividida em duas partes — "Estudos acerca do dialeto caipira (DC)" e "Atlas linguísticos no Estado de São Paulo" —, adentremos a seguir o âmbito da nossa pesquisa.

2.3 Nossa pesquisa

2.3.1 Exórdio

Em dois períodos — de 15.02 a 15.03.2016 e de 23.08 a 17.09.2017 —, percorremos dez dos municípios mais antigos do Estado de São Paulo, às margens do rio Tietê. Fizemo-lo para a realização de um inquérito linguístico e extralinguístico junto a 80 voluntários. Deles, os dados orais coletados somam 115h25m19s de gravação útil. Para isso, a nossa principal motivação era elaborar um atlas dialetal da variedade do português falado no interior do Estado de São Paulo, também conhecida como "dialeto caipira".

Com a visitação empírica apoiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), os municípios que foram investigados são: Santana de Parnaíba, Pirapora do Bom Jesus, Araçariguama, São Roque, Sorocaba, Itu, Porto Feliz, Tietê, Capivari e Piracicaba. Cada uma delas nos forneceu oito indivíduos de diferentes perfis para o nosso estudo.

Culturalmente e sob outras perspectivas, essas localidades podem ser consideradas como integrando a região denominada "Médio Tietê", para a qual o rio Tietê representou e representa muito mais do que um elemento fluvial em si. De um ponto de vista histórico, lembremo-nos, em primeiro lugar, de que grande parte das expedições exploratórias conduzidas pelos colonizadores a partir dos anos 1500 rumo ao interior do país deveu-se às águas desse rio e, em segundo lugar, de que, dado o uso aludido ao qual o rio Tietê estava submetido, vários aglomerados urbanos começaram a estrategicamente surgir — tal é o caso das localidades da nossa pesquisa —, desde as primeiras expedições.

Atualmente, apesar de toda a degradação infligida sobretudo por indústrias profundamente sem ética ambiental, o rio Tietê ainda é tido como o mais importante do Estado de São Paulo. Formado entre 10 e 15 milhões de anos atrás, ele já começou a apresentar carência de oxigênio a partir de 1970, em decorrência da poluição, iniciada por sua vez há mais de um século. Tristemente, esse curso d'água continua recebendo dejetos industriais e de esgoto em grande volume a partir do décimo quinto quilômetro da nascente, à altura de Mogi das Cruzes-SP.

A nascente se localiza em Salesópolis-SP, na Serra do Mar, a 22 quilômetros do oceano Atlântico, e deságua em Três Lagoas-MS, na barragem do Jupiá, rio Paraná. Seu nascedouro está a 1.120 metros de altitude, com jusante do sudeste ao noroeste ao longo de 1.150 quilômetros. Compreende seis sub-bacias hidrográficas paulistas, a começar pela do Alto Tietê e a finalizar na do Baixo Tietê — as informações são do Sistema Integrado de Gerenciamento de Recursos Hídricos do Estado de São Paulo (SÃO PAULO, 2016):

(1) Alto Tietê: situa-se na nascente e implica 35 municípios da região metropolitana de São Paulo, incluindo Santana de Parnaíba e Pirapora do Bom Jesus;

(2) Sorocaba/Médio Tietê: abarca 34 municípios, envolvendo Araçariguama, Itu, Porto Feliz, São Roque, Sorocaba e Tietê;

(3) Piracicaba/Capivari/Jundiaí: tem 60 municípios paulistas, abrangendo Capivari e Piracicaba;

(4) Tietê/Jacaré: no centro do Estado, envolve 34 municípios;

(5) Tietê/Batalha: relaciona-se com 36 municípios; e

(6) Baixo Tietê: no noroeste do Estado de São Paulo, vai da corredeira de Laje à desembocadura com o rio Paraná, com 42 municípios.

Mantendo-se a numeração acima em mente, de (1) a (6), as sub-bacias associadas ao rio Tietê podem ser visualizadas na carta que a Figura 1 abaixo traz.

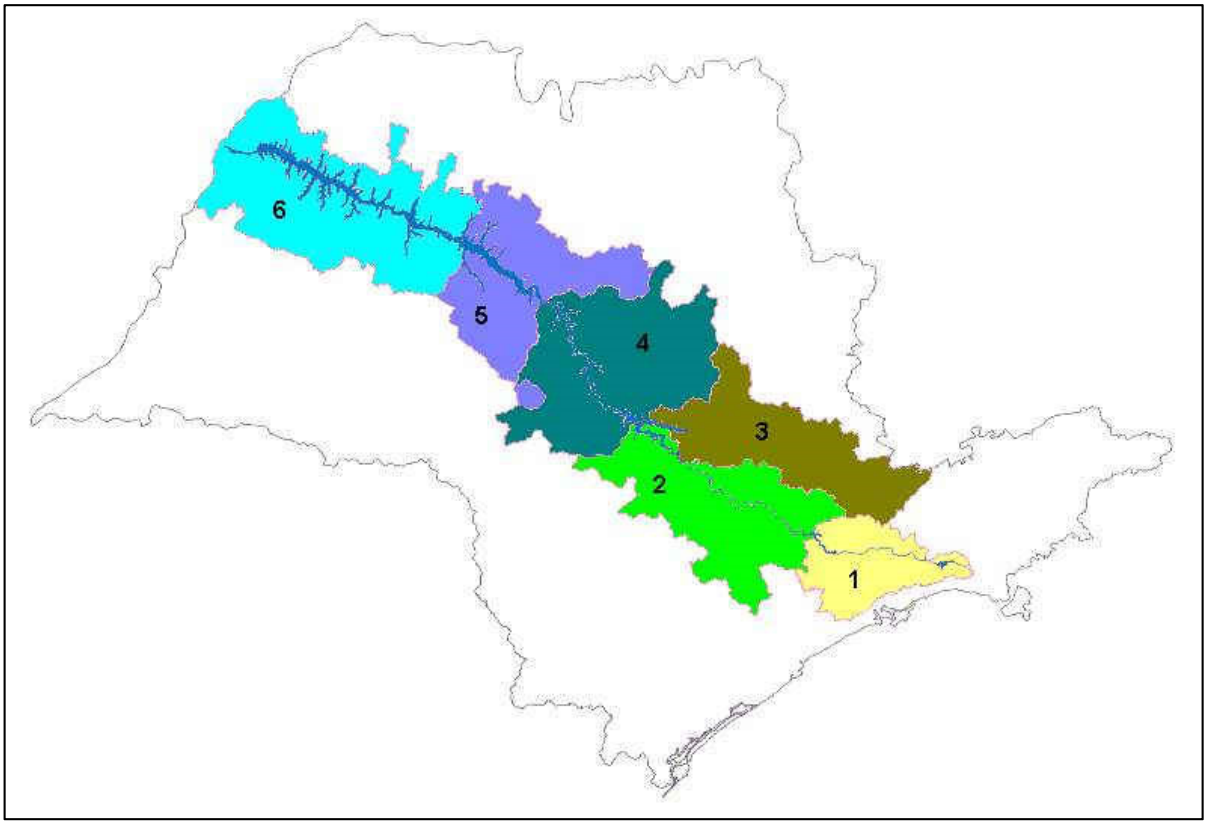


Figura 1: Sub-bacias hidrográficas paulistas (SÃO PAULO, 2016)

Como se pode observar, as dez localidades da nossa investigação distribuem-se pelas sub-bacias da seguinte forma (cidades à esquerda — sub-bacias à direita): Santana de Parnaíba e Pirapora do Bom Jesus — Alto Tietê ● Araçariguama, Itu, Porto Feliz, São Roque, Sorocaba e Tietê — Sorocaba/Médio Tietê ● Capivari e Piracicaba — Piracicaba/Capivari/Jundiaí. Consequentemente, a maioria das nossas cidades situa-se na sub-bacia de Sorocaba/Médio Tietê.

Em todo caso, a associação acima entre municípios e sub-bacias é de natureza preponderantemente hidrográfica, e, como dissemos antes, a designação "Médio Tietê" ganha diferente abrangência quando a perspectiva de consideração é eminentemente cultural, à qual as nossas dez localidades são subsumidas.

Isso posto, diga-se que a escolha dessa região por nós enquanto campo de investigação baseia-se no fato de que ela se constitui em um dos primeiros rincões do nosso país que portugueses colonizaram a partir dos anos 1500, tal como dizíamos há pouco. Por essa razão, o Médio Tietê acabou tornando-se um dos primeiros centros de difusão da língua portuguesa no Brasil, fato que atrai a atenção tanto de linguistas históricos (em perspectiva diacrônica), quanto de (pós-)estruturalistas (em perspectiva sincrônica^{+}). Entre os primeiros, além de dialetólogos, encontram-se por exemplo filólogos interessados no contraste entre diferentes sincronias da língua, baseados em determinadas evidências grafemáticas^{+} (encontradas em cartas, diários *etc.*) que testemunham, segundo hipóteses e suas testagens, a oralidade de tempos longínquos.

No que concerne à urbanização da nossa rede de pontos^{+}, Tietê seria o mais antigo, remontando a inícios do século XVI, enquanto Piracicaba seria o menos antigo, com sua história iniciando-se em 1766. Todas as localidades devem seu surgimento em parte às paradas dos bandeirantes que passavam pela região entre 1550 e 1720, saindo do sudeste do país rumo às outras regiões, entre cujos caminhos havia as águas do rio Anhembi (hoje rio Tietê), em torno do qual as nossas dez cidades foram erigidas.

Em 1718, ricas jazidas de ouro foram descobertas nas proximidades de Cuiabá, Estado de Mato Grosso, atraindo muitas gentes, que também para lá iam por vias fluviais, uma alternativa à mata virgem cerrada. As caravanas, todavia, eram frequentemente interceptadas por indígenas, fazendo com que os viajantes ocasionalmente se espalhassem, a partir do que acabavam formando acampamentos às margens dos rios (dentre eles, do rio Tietê), alguns dos quais vindo a efetivamente tornar-se cidades mais tarde.

Feitas as observações preliminares acima, sigamos este exórdio do nosso estudo fornecendo informações a respeito da orientação metodológica e teórica que adotamos para o nosso empreendimento investigativo.

Os instrumentos de coleta que aplicamos são direcionados à obtenção de dados tanto linguísticos quanto extralinguísticos dos falantes representantes do nosso *locus*. Para a elicitação+ da parte linguística, os aspectos envolvidos — embora nem todos estejam contemplados na presente obra — são de natureza semasiológica (semântico-lexical em particular), fonético-fonológica, morfossintática, prosódica e pragmática+, bem como eles dizem respeito à modalidade de leitura e a diferentes tipos discursivos (relato pessoal, comentário, descrição e relato não pessoal). Quanto à recolha extralinguística, os dados envolvidos — todos aqui disponibilizados, mas nem todos processados — são ligados à ascendência étnica, à religião, à classe social, à escolaridade, à profissão, à proficiência em idiomas, ao contato linguístico, ao contato com mídias e à topodinâmica+ do inquirido.

Pela ótica da Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*) — referida mais adiante como "DP" —, toda essa gama de aspectos extralinguísticos, mais os diafásicos e os diarreferenciais, permite que pelo menos oito dimensões de análise (postas em perspectiva com os fatos linguísticos) sejam possíveis, tal como a Tabela 4 a seguir sugere.

	Dimensão	Parâmetros
i	diagenérica	F/M
ii	diageracional	GI/GII
iii	diastrática I	Ca/Cb
iv	diastrática II	A/B

	Dimensão	Parâmetros
v	diatópica	+M/-M
vi	diafásica	Re/RO/Co/De/Lt/Di
vii	diarreferencial	L/ML
viii	diarreligiosa	*x*

Tabela 4: Dimensões e parâmetros de análise (com destaques)

Sendo:

(i) F = feminino ● M = masculino;

(ii) GI = grupo etário mais jovem (de 18 a 36 anos) ● GII = grupo etário mais velho (a partir de 55);

(iii) diastrática I = diastrática escolar ● Ca = escolaridade alta (a partir de estudos universitários, completos ou incompletos) ● Cb = escolaridade baixa (no escopo de abrangência entre analfabetismo e ensino médio incompleto);

(iv) diastrática II = diastrática econômica ● B = baixa/pobre (condições precárias de moradia/vida) ● A = alta/rica (ausência de precariedade);

(v) +M = com mudança de domicílio ● -M = sem mudança de domicílio;

(vi) Re = relato ● RO = resposta objetiva ● Co = comentário ● De = descrição ● Lt = leitura ● Di = diálogo;

(vii) L = linguagem com denotação orientada ao mundo ● ML = linguagem com denotação orientada à linguagem (metalinguagem); e

(viii) x = orientação religiosa (católica, umbandista *etc.*).

Os destaques na Tabela 4 referem-se às dimensões com que lidaremos diretamente no presente trabalho, segundo seus parâmetros, com exceção da dimensão diatópica^{+}, a qual trataremos sem distinção de parâmetros, embora os tenhamos observado e registrado (cf. seç. "*Corpora* extralinguísticos", no Tomo V). Isso porque selecionamos os indivíduos a partir das dimensões em destaque, *i.e.*, a partir das dimensões diagenérica^{+}, diageracional, diastrática I (escolar) e diatópica, conforme veremos com mais detalhes na seção "Perfis dos informantes", no presente tomo. Por consequência, as dimensões mencionadas foram sistematicamente monitoradas em nossa pesquisa, razão pela qual elas se constituem aqui em variáveis extralinguísticas controladas. Porém, afora a dimensão diafásica, todas as demais citadas na tabela em foco podem ter seus dados acessados nesta obra[14].

[14] Para aspectos diastráticos econômicos e diarreligiosos, v. seç. "*Corpora* extralinguísticos" e, para aspectos diarreferenciais, v. seç. "*Corpus* metalinguístico-etnográfico", ambas no Tomo V.

Até aqui, já se pode perceber que a DP se apresenta como sendo uma abordagem da variação e mudança linguísticas que concilia a preocupação de se estudarem os fenômenos geograficamente localizados (perspectiva "horizontal") e socialmente situados (perspectiva "vertical").

Afinal, a variação e a mudança, às quais as línguas estão constitutiva e inescapavelmente submetidas em graus distintos, estão atreladas às diferenciações extralinguísticas (sociais, geográficas *etc.*) que caracterizam uma comunidade linguística ou segmentos dela. Por seu turno, essas diferenciações extralinguísticas comumente correlacionam-se com formas linguísticas correspondentes, de ordem sonora, lexical ou gramatical.

Mediante a ótica pluridimensional, o entendimento das variantes linguísticas é aumentado, já que há a preocupação de verificarem-se as motivações, os condicionamentos linguísticos e as causas externas dos fatos de língua, além da descrição deles tendo-se em conta seu *locus* de ocorrência. *Obviamente, no entanto, esse programa de investigação não é exclusividade da DP*. Todavia, é importante dizer que ela tem o mérito de controlar as várias dimensões de análise e cartografá-las simultaneamente, gerando, assim, a perspectiva pluridimensional geolinguística propriamente dita.

2.3.2 Transcrição, tabulação e organização dos dados coletados

A transcrição, a tabulação e a organização dos nossos dados foram feitas em Zoho Sheet e Excel, *softwares* que possibilitaram tanto a comunicação entre planilhas à maneira de gestão de banco de dados, quanto a criação dos gráficos estatísticos incorporados nas cartas linguísticas.

Inicialmente, criamos um conjunto de planilhas para cada questão dos questionários (Questionário Semântico-Lexical [m], QSL [m]; e Questionário Fonético-Fonológico[m], QFF[m]). Nesse conjunto, uma das planilhas — que chamaremos S1 (*spreadsheet* 1) — manualmente recebia os dados da oitiva das gravações, enquanto uma outra, S2, preparada anteriormente com códigos de lógica de programação, simultânea e automaticamente organizava os *inputs* de acordo com as localidades geográficas visitadas e de acordo com as dimensões diastrática, diageracional e diagenérica. Outra planilha

automatizada, S3, calculava as ocorrências de cada variante performadas por cada grupo controlado de inquiridos e as dispunha em quadros dedicados, exibindo-as aí quantificadas em termos absolutos e relativos.

Por grupos controlados de entrevistados, entendem-se as três combinações (ou cruzamentos) de variáveis extralinguísticas sistematicamente controladas em nossa pesquisa. Numa notação abstrata que derivamos da DP, essas combinações — que também chamaremos de 'séries' — podem ser assim expressas: CxGy, CxWz e WzGy; sendo 'C' para 'classe escolar', 'G' para 'grupo etário' e 'W' para 'gênero de registro/social'+; bem como sendo 'x' uma variável lógica saturável ou por 'a' (de escolaridade alta), ou por 'b' (de escolaridade baixa); 'y' saturável ou por 'II' (de geração mais velha), ou por 'I' (de geração mais nova); e, finalmente, 'z' saturável ou por 'f' (de feminino), ou por 'm' (de masculino). Note-se que um grupo controlado específico é uma de três parametrizações possíveis de uma dada série.

Isso dito, e para ilustrar um dos quadros dedicados mencionados no parágrafo prévio ao anterior, tomemos um exemplo referente ao grupo CaGII — que é uma de três parametrizações da série CxGy — com relação à produtividade absoluta de uma variante linguística *A* (ou *1*), que é representada na simbologia polifórmica da DP com o símbolo ◔ (círculo com o quadrante superior direito preenchido), mas que é representada como *A* em nível de programação/codificação.

No quadro dedicado aos grupos, na S3, há uma interseção formada pela coluna "CaGII" e pela linha "A", interseção essa na qual será exibida posteriormente a produtividade em termos de frequência absoluta da variante *A* no grupo CaGII. Em tal interseção, cria-se uma célula, em que a linha de comando "=COUNTIF(w:h;“A”)" é inserida. Nessa fórmula, "=COUNTIF" conceitualmente se traduz como "conte se [determinada condição se verificar]"; "w", uma variável, é saturada pelo endereço da célula inicial onde a coleção de dados está registrada na S2; "h", outra variável, é por sua vez saturada pelo endereço da célula final da coleção de dados na S2. Portanto, o comando em tela pode ser assim lido: "Conte o número de células no *spectrum* indicado, de *w* a *h*, que contenham o valor *A*". Dessa forma, a interseção formada pelo encontro da coluna "CaGII" com a linha "A" mostrará o número de ocorrências absolutas da variante *A* realizadas pelo grupo CaGII. O mesmo, *mutatis mutandis*, é feito para os outros grupos (CaGI, ...) e para as covariantes+ de *A* (*B*, ...).

Se seguirmos lidando apenas com o grupo CaGII, e se estivermos considerando as quatro covariantes mais frequentes no universo de itens coletados, então as variáveis correspondentes serão *A*, *B*, *C* e *D* em nível de codificação, que equivalem, respectivamente, a ◔, ◔, ◕ e ◕ na simbologia polifórmica da DP. Ainda, dois outros símbolos são importantes aqui: *E* (○) e *V* (▮). Na lógica e na simbologia, eles representam respectivamente "outras formas" — *i.e.*, demais covariantes que não sejam *A*, *B*, *C* ou *D* — e "resultado inválido", em função, entre outras razões, de abstenção por parte do informante.

Paralelamente à coluna "CaGII" na S3, estabelecem-se duas colunas auxiliares dedicadas à conversão dos termos absolutos em termos relativos: uma contendo o percentual de todos os resultados, e outra comportando o percentual dos resultados só válidos.

Isso posto, a Figura 2 abaixo exibe exemplos de partes das S1, S2 e S3 (ordem da esquerda para a direita), extraídos da categoria correspondente às quatro covariantes mais frequentes e da série CxGy, com enfoque no grupo CaGII.

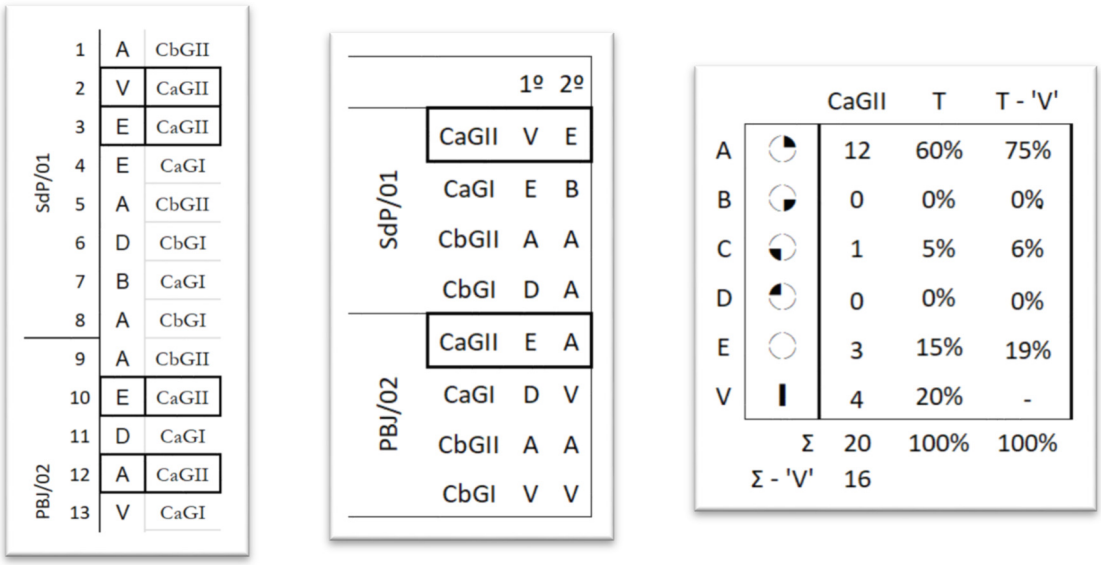


Figura 2: Partes de S1, S2 e S3 (ordem da esquerda para a direita) (exemplo)

As partes referentes às S1 e S2 contêm apenas uma pequena amostra dos resultados produzidos pelo grupo CaGII, enquanto a parte referente à S3 encerra uma quantificação completa dos resultados produzidos por esse mesmo grupo.

Os dados presentes na Figura 2 foram retirados do nosso *corpus* semântico-lexical e são resultados advindos da aplicação da questão 8 do QSL [m]. Nomeadamente, as lexias[+] nesse caso registradas em relação de covariação são: <mandioca brava> / <mandioca braba> (◔), <mandioca> (◔), <aipim> (◔) e <macaxeira brava> (◔).

2.3.3 Treinamento no *software* de cartografação

O treinamento que realizamos no estágio sanduíche, junto ao Dr. Harald Thun e seu grupo de pesquisa, forneceu os elementos necessários e finais para a cartografação dos resultados da nossa pesquisa. O *software* empregado por eles na Universidade de Kiel (orig. *Christian-Albrechts-Universität zu Kiel*, CAU) é uma customização do Excel.

Uma vantagem no uso dessa ferramenta é a capacidade que ela tem de incorporar uma base cartográfica — frequentemente plasmada em (e exportada de) um programa de geoprocessamento específico (como Philcarto ou QGIS) — e de permitir que haja interação entre essa base cartográfica e a base de dados (o *corpus* linguístico estruturado, no caso) através de linhas de comando.

O modelo de carta linguística que concebemos para a geração dos atlas consiste de oito (conjuntos de) elementos: (I) título comum; (II) título técnico; (III) mapa administrativo do Estado de São Paulo; (IV) enunciado da questão (no caso do atlas semântico-lexical) ou ausência de elemento (no caso do atlas fonético-fonológico); (V) relação de covariantes linguísticas; (VI) ampliação da área geográfica em que a rede de pontos investigada se encontra e a cartografia linguística propriamente dita; e (VII) gráfico estatístico.

A fim de exemplificar os itens de (I) a (VII), tomemos a Figura 3, presente na próxima página, que consiste de uma carta linguística extraída do nosso atlas semântico-lexical.

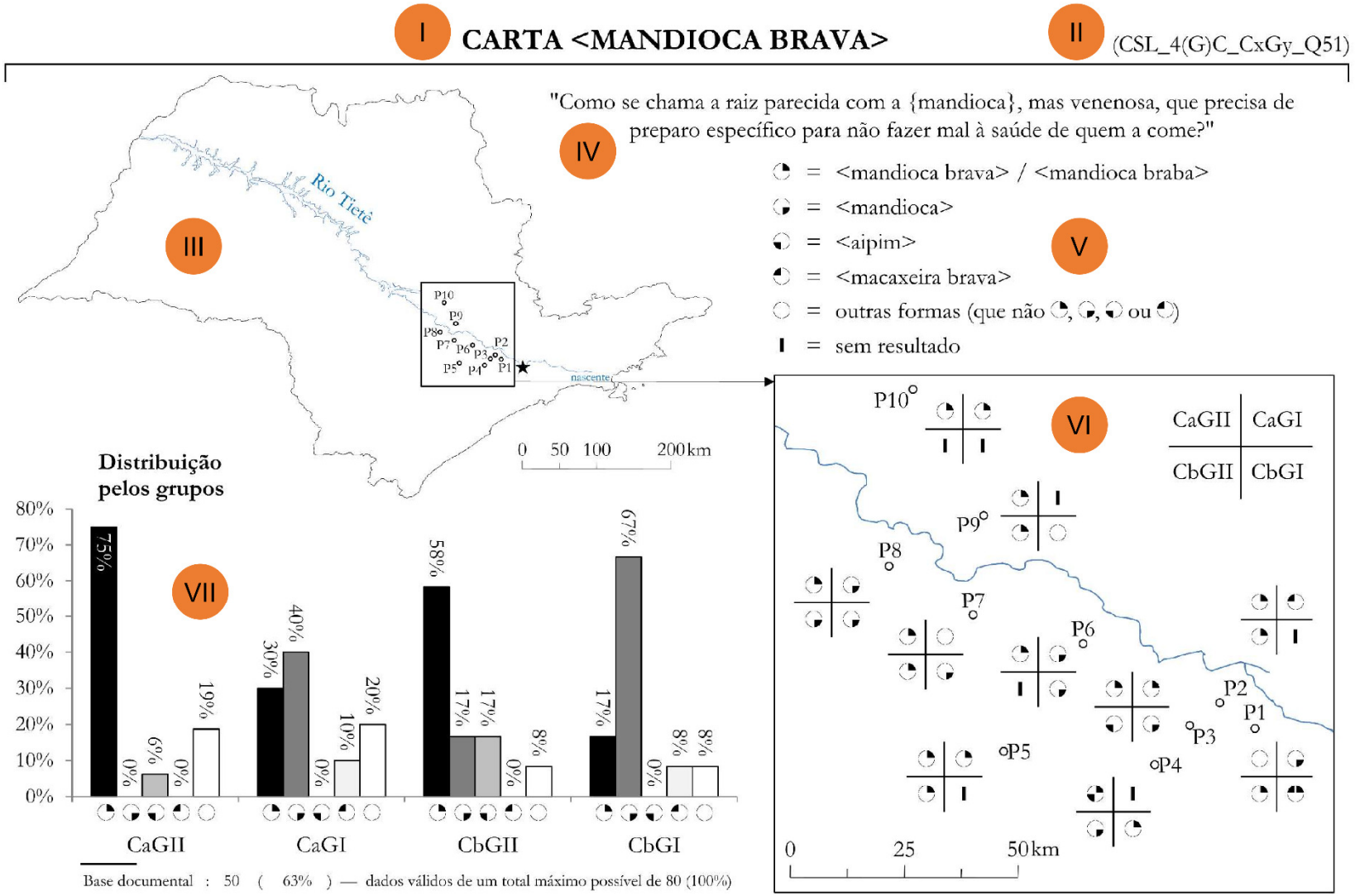


Figura 3: Componentes das cartas linguísticas (exemplo do atlas semântico-lexical)

(I)

O título comum da carta é uma solução simples e amigável de referência. Ele é constituído pela palavra "carta" seguida pela lexia representada pelo símbolo ◔ (no caso do atlas semântico-lexical) ou pelo vocábulo cuja pronúncia por parte dos informantes foi oral e/ou imageticamente induzida pelo inquiridor (no caso do atlas fonético-fonológico).

(II)

Já o título técnico oferece as informações completas de identificação da carta linguística. Ele contém quatro variáveis. Na ordem da esquerda para a direita, elas são: o tipo (CSL ou CFF), a categoria (4(G)C, 3(G)C ou 2(G)C), a série (CxGy, CxWz ou WzGy) e o número da questão (Q?).

A sigla 'CSL' designa 'carta semântico-lexical'; e 'CFF', 'carta fonético-fonológica'. Por seu turno, a expressão '4(G)C' refere-se a 'quatro (grupos de) covariantes'; '3(G)C', 'três (grupos de) covariantes; e '2(G)C', 'duas (ou dois grupos de) covariantes'. Quanto às expressões que denotam as diferentes séries, elas já foram explicadas na seção anterior, "Transcrição, tabulação e organização dos dados coletados". E 'Q?' remete ao número da questão do respectivo questionário (QSL^{m} ou QFF^{m}).

(III)

O mapa administrativo do Estado de São Paulo em escala consta da carta linguística pela função principal de indicar ao leitor a posição relativa das localidades integrantes da nossa rede de pontos investigada no interior da circunscrição estadual. Além disso, o mapa exibe o posicionamento da capital administrativa e do trajeto do rio Tietê, rio esse cujas existência e características físicas (infelizmente em processo de deterioramento crônico por ações humanas) estão direta ou indiretamente ligadas à formação dos centros urbanos situados em seus arredores.

(IV)

Nas cartas semântico-lexicais, há o enunciado da questão do QSL^{m} à qual os informantes responderam. Nas cartas fonético-fonológicas, essa parte fica omissa, uma vez que seu equivalente — o vocábulo cuja pronúncia por parte dos entrevistados foi induzida a propósito do QFF^{m} — já aparece no título comum da carta, o qual foi objeto do item (I).

(V)

Também como componente da carta linguística, uma relação das covariantes mais frequentes é apresentada, segundo a simbologia polimórfica da DP já introduzida antes. Ou seja, numa carta de categoria 4(G)C, de quatro (grupos de) covariantes, estas últimas são arroladas de cima para baixo em ordem de maior para menor frequência respectivamente com os símbolos ◔, ◔, ◔ e ◔. As demais covariantes são representadas em conjunto pelo símbolo ◌, e os resultados inválidos, pelo símbolo ❙.

(VI)

Na carta, a cartografia propriamente dita concentra-se na parte ampliada do mapa administrativo de São Paulo em escala. Nessa ampliação, as localidades geográficas visitadas estão codificadas de P1 a P10: Santana de Parnaíba (P1), Pirapora do Bom Jesus (P2), Araçariguama (P3), São Roque (P4), Sorocaba (P5), Itu (P6), Porto Feliz (P7), Tietê (P8), Capivari (P9) e Piracicaba (P10).

Na parte superior direita da ampliação, acha-se a "chave de intepretação" da cartografia, constituída por uma cruz com quatro quadrantes simétricos entre si. Em cada um desses quadrantes, um grupo particular de inquiridos está representado. No caso da série CxGy, os grupos particulares, em ordem horária, são: CaGI, CbGI, CbGII e CaGII.

Como visto antes, nessas expressões a letra 'C' é usada para 'classe escolar', e 'G', para 'grupo etário', enquanto 'a' denota 'escolaridade alta' (a partir do ensino superior não necessariamente completo), e 'b', 'escolaridade baixa' (espectro a partir do analfabetismo até o ensino médio incompleto), bem como 'II' designa 'geração mais velha' (a partir dos 55 anos), e 'I', 'geração mais nova' (entre 18 e 36). Ou seja, a expressão "CaGI", por exemplo, pode ser lida como "grupo mais velho com escolaridade alta".

Isso dito, tomemos a configuração simbólica exibida no P1 da Figura 3 (localizada duas páginas atrás), reproduzida por praticidade na próxima página.

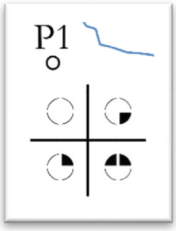


Figura 4: Recorte do P1 da Figura 3

Com utilização da chave de intepretação, a configuração simbólica mostrada na Figura 4 implica que, na localidade em questão, houve pelo menos uma ocorrência das seguintes covariantes nos seguintes grupos: ◔ em CaGI; ◔ e ◖ em CbGI; ◔ em CbGII; e ○ em CaGII.

(VII)

Para finalizar a exposição exemplificada da relação de componentes das cartas linguísticas, referimo-nos agora ao gráfico estatístico. Seu intuito é fornecer um entendimento quantitativo de conjunto a propósito dos resultados registrados. Ele os distribui pelas quatro parametrizações de cada série. Ou seja, se a série tematizada na carta for CxGy, suas parametrizações correspondentes serão CaGI, CbGI, CbGII e CaGII. Cada parametrização equivale a um grupo controlado particular de informantes. No gráfico, os grupos apresentam a frequência relativa dos resultados válidos.

Esse componente da carta permite, por conseguinte, que se visualizem a concorrência e a produtividade das covariantes linguísticas são só em relação a elas mesmas, mas também em relação aos diferentes grupos de falantes. Por fim, na parte inferior esquerda encontra-se a informação sobre a base documental do gráfico.

2.3.4 Dificuldades surgidas/esperadas

Como é de praxe nos estudos que envolvem pesquisa de campo, a fase de coleta de dados empíricos é uma das que mais apresenta desafios, dificuldades e problemas a serem superados. No nosso caso, visitamos dez municípios do interior paulista e entrevistamos 80 indivíduos, e a tarefa de encontrá-los, de acordo com os perfis metodológicos estipulados pela investigação, demandou paciência e calma, até porque eles haviam de

dispor de tempo na agenda social em geral e laboral em específico para serem entrevistados num inquérito relativamente longo.

Também digno de nota é um dilema concernente à fase de constituição do grupo de *corpora* e à fase de elaboração das cartas constituintes dos atlas linguísticos. Embora a nossa proposta junto à FAPESP tenha declarado, desde o princípio, que o objetivo a perseguir era elaborar um atlas linguístico (nomeada e especificamente, um atlas semântico-lexical e outro fonético-fonológico), nutríamos o interesse de realizar adicionalmente análises extensivas dos dados obtidos. No entanto, percebemos nas fases mencionadas que, realmente, teríamos de nos focar no objetivo inicialmente proposto, caso o desejo fosse constituir *corpora* e atlas de qualidade e sem atrasos. Assim, pois, resolvemos cumprir com o prometido antes de tudo, e essa decisão, além de já inicialmente prevista, ainda nos parece ter sido de fato acertada: a partir daqui, a partir deste trabalho, também outros pesquisadores interessados no português paulista (ou "dialeto caipira") poderão utilizar os *corpora* e os atlas, sob variados propósitos de estudo.

Ademais, outra dificuldade com que lidamos relaciona-se com o fato de que os *corpora*, os atlas e demais elementos produzidos pela pesquisa somam, na íntegra, mais de duas mil páginas. Tal fato torna a tarefa de organizar tudo numa edição bem organizada de vários tomos, com o menor número possível de erros, um desafio geralmente subestimado. Apesar dessas e outras dificuldades, menores ou maiores, a satisfação em se contribuir diretamente para o conhecimento científico e — tendo-se em conta os benefícios e aplicações decorrentes possíveis — indiretamente para a sociedade através da consecução dos objetivos de exploração/investigação faz com que o trabalho de pesquisa valha a pena.

3 OBJETIVOS E HIPÓTESE

3.1 Objetivos gerais

A partir do aporte teórico-metodológico da Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*) e com instrumentos de coleta de dados empíricos do Atlas Linguístico do Brasil (ALiB, 2014) modificados (cf. seç. "Modificações nos instrumentos metodológicos"), os objetivos gerais atingidos por esta tese de doutorado dizem respeito à elaboração de dois atlas linguísticos pluridimensionais (um semântico^{+}-lexical^{+} e outro fonético^{+}-fonológico^{+}) do português paulista (ou "dialeto^{+} caipira^{+}") falado na região do Médio Tietê. Na cartografia de ambos, além da dimensão diatópica^{+} inerente a esse tipo de empreendimento geolinguístico, há ainda as dimensões diastrática^{+}, diagenérica^{+} e diageracional^{+}. Cada carta compreende também um gráfico estatístico. Os atlas, com suas características mais gerais, estão descritos a seguir.

(a) O atlas semântico-lexical (ASL) é formado por 432 cartas linguísticas, cujos dados foram levantados através da aplicação do Questionário Semântico-Lexical^{m}, versão por nós modificada da versão original de 2001 do ALiB. O ASL abarca 14 áreas semânticas: (i) acidentes geográficos; (ii) fenômenos atmosféricos; (iii) astros e tempo; (iv) atividades agropastoris; (v) fauna; (vi) corpo humano; (vii) ciclos da vida; (viii) convívio e comportamento social; (ix) religião e crenças; (x) jogos e diversões infantis; (xi) habitação; (xii) alimentação e cozinha; (xiii) vestuário e acessórios; e (xiv) vida urbana. Mais detalhes encontram-se no Tomo II.

(b) O atlas fonético-fonológico (AFF), por seu turno, é composto por 324 cartas linguísticas, cujos dados foram obtidos pelo emprego do Questionário Fonético-Fonológico^{m} (versão por nós modificada da versão original de 2001 do ALiB). o AFF abrange o registro do comportamento tanto de vogais e *glides* — itens (i) a (v), a seguir —, quanto de consoantes — itens (vi) a (xiv): (i) ditongos decrescentes; (ii) vogais baixas postônicas; (iii) vogais altas (pretônicas e tônicas); (iv) vogais médias (pretônicas, tônicas e postônicas);

(v) /A/ e /O/ adjacentes a consoante nasal no padrão '(C)_.C[+nasal]'; (vi) gerúndio; (vii) lateral palatal /ʎ/; (viii) nasal palatal /ɲ/; (ix) /L/ (em coda+ e como C2 em *onset* complexo); (x) /S/ (em coda e intervocálico no padrão 'V._V'); (xi) /R/ (em coda, em *onset* simples e como C2 em *onset* complexo); (xii) encontros consonantais interssilábicos (/ f.t /, / p.n / e / d.v /); (xiii) palatalização (da lateral /l/ e das oclusivas /t, d/ diante da vogal alta [i]); e (xiv) alternâncias ([b] *vs.* [v] em *onset* simples; [k] *vs.* [kʊ̯]; [ʒɪ] *vs.* [ʒẽɲ] como sílaba final átona; [i] *vs.* [ẽɲ] pós-nuclear em sílaba final átona; e [ĩ] *vs.* [ũ] em sílaba inicial). Além disso, o AFF compreende cartas especiais sobre as contrapartes (a)fonéticas+ do arquifonema+ /R/ em quatro situações estruturais distintas: (a) em coda externa de desinência infinitiva seguida por pausa; (b) em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa; (c) em coda interna em sílaba tônica; e (d) em coda interna em sílaba átona. Mais detalhes podem ser consultados no Tomo III.

3.2 Objetivos específicos

A fim de alcançarmos os objetivos gerais mencionados há pouco, os principais objetivos específicos que tivemos de realizar são:

(α) constituição de *corpora* semântico-lexicais;
(β) formação de *corpora* fonético-fonológicos;
(γ) criação de *corpora* extralinguísticos;
(δ) transcrição fonética e grafemática+ do material oral;
(ε) tabulação e organização dos dados empíricos;
(ζ) formulação de gráficos estatísticos; e
(η) cartografação dos resultados.

3.3 Hipótese

Adicionalmente aos objetivos gerais referidos anteriormente, testamos uma hipótese específica, qual seja:

> *Os jovens entre 18 e 36 anos de idade do Médio Tietê (especialmente aqueles com nível escolar alto[1]) estão produzindo a covariante+ [ɻ] (a aproximante retroflexa[2]) como contraparte fonética do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa sob uma frequência relativa inferior àquela a ser observada entre os falantes mais velhos a partir de 55 anos de idade (especialmente aqueles com nível escolar baixo[3]).*

Como resultado, na realidade incontestável dos fatos concretos observados, a fala dos jovens evidenciou-se ser, na verdade, mais albergadora da variante+ [ɻ] do que a fala dos mais velhos. Portanto, a hipótese foi refutada, e o grau dessa constatação tornou-se ainda mais eloquente na mesma medida em que soubemos que a variante [ɻ] fez-se presente ao longo de toda a nossa rede de pontos+, com frequência relativa de 57% entre os jovens com nível escolar alto e 62% entre os jovens com nível escolar baixo (contra 45% entre os mais velhos com nível escolar alto e 30% entre os mais velhos com nível escolar baixo). Mais detalhes sobre a gama de dados e seus resultados acerca dessa questão estão disponíveis na seção "Análises do /R/ em coda".

Com base nessa verificação, nota-se que há uma tendência de a região do Médio Tietê — considerada o berço da cultura caipira (para mais, v. seç. "Justificativas") — ser atualmente até mais produtiva no tocante à realização da variante [ɻ] em coda externa de desinência infinitiva seguida por pausa do que outrora, dado que os mais velhos atualizam-na menos do que os mais jovens hoje em dia.

[1] A partir de estudos universitários, completos ou incompletos.
[2] Conhecida como "erre retroflexo" ou ainda como "erre caipira".
[3] No escopo de abrangência entre analfabetismo e ensino médio incompleto.

Ademais, outra inferência decorrente da eloquência dos dados obtidos diz respeito à evidência de que a variante em apreço não se mostra no presente tão estigmatizada quanto à época dos estudos de Amaral (1920), desde quem se sabe também o quão caracterizadora a variante [ɻ] tem sido do vernáculo da região (conhecido como "dialeto caipira") em particular, bem como da cultura caipira por extensão.

O levantamento da nossa hipótese baseou-se na ideia de nivelamento dialetal (orig. *dialect levelling*), segundo a qual dialetos tendem a perder (parte de) suas características linguísticas em favor, via contato, da incorporação de características próprias de variedades^{+} estandardizadas. Tal foi a base da hipótese porque a variedade paulista (ou o dialeto caipira) possui o fone^{+} em apreço como característica bastante conhecida (embora não exclusiva), em contraste com a variedade paulistana, comumente entendida, por sua vez, como tendencialmente estandardizada, na qual o som em questão (apesar de existir como variante possível na situação fonológica sob enfoque) não é considerado (tão) característico.

4 JUSTIFICATIVAS

4.1 Dos objetivos gerais

Como introduzidos antes, a nossa pesquisa estabeleceu os objetivos gerais de elaboração de dois atlas linguísticos pluridimensionais (um semântico^{+}-lexical^{+} e outro fonético^{+}-fonológico^{+}) do português paulista falado na região do Médio Tietê. A seguir, vamos justificá-los em termos essenciais.

Partindo-se do fato, por um lado, de que a região do Médio Tietê ainda não foi objeto de cartografia dialetológica pluridimensional, há que se saber, por outro lado, que são várias as vantagens decorrentes da existência de um atlas linguístico do vernáculo português falado no interior paulista.

Em primeiro lugar, esse engajamento técnico, geolinguístico e científico: (i) foca-se na descrição linguística; (ii) oportuniza o descobrimento de fatos externos responsáveis pela variação dialetal; e (iii) historiografa a língua em uso em dadas coordenadas geográficas, temporais e sociais.

Em segundo lugar, esse empenho dialetológico produz subsídios de base empírica: (iv) à pedagogia de língua (seja enquanto língua materna, seja enquanto língua estrangeira); e (v) aos estudos comparativos entre estados de língua (estudos tanto sincrônicos^{+} quanto diacrônicos^{+}).

Em terceiro lugar, as aplicações baseadas nos resultados da pesquisa ainda incluem: (vi) referências empiricamente abonadas tanto para gramáticas descritivas, quanto para obras lexicográficas; e (vii) conhecimentos factuais à Linguística Computacional em relação ao desenvolvimento de assistentes virtuais e inteligências artificiais de interação oral com humanos.

Por fim, diga-se que as vantagens se estendem ainda à retroalimentação acadêmica potencial da área do conhecimento mobilizada — e até de outras áreas, como a Sociolinguística Variacionista —, no sentido em que cada investigação de base empírica implica, em diferentes graus, desde revisões metodológicas e conceituais até reavaliações de princípios epistemológicos. Como se sabe, esses fazeres é que levam a Ciência, segundo suas diversas disciplinas, a avançar.

4.2 Do *locus*

Há duas grandes razões a justificarem a escolha da região do Médio Tietê do interior paulista como *locus* da pesquisa. A primeira delas diz respeito ao fato de que o português ali falado representa uma das primeiras lusitanizações do Brasil, desde 1532 e a partir da cidade de São Vicente. A segunda grande razão liga-se à consideração de que o Médio Tietê é o berço da cultura e do dialeto+ caipiras+ do Estado de São Paulo. Tal condição tem a ver com o fato de que municípios dessa região são legatários das passagens das bandeiras, das moções e dos tropeiros rumo ao interior do país — especialmente a partir de Porto Feliz — ocorridas entre os séculos XVI e XVIII; fato que conferiu características culturais (e linguísticas) próprias à região (cf. CÂNDIDO, 2001).

Isso posto, é de se esperar que uma gama de linguistas se interesse em estudar o português paulista mediante diferentes aportes teórico-metodológicos, com envolvimento de várias universidades paulistas, tal como evidenciamos nas seções "Estudos acerca do português paulista" e "Atlas linguísticos no Estado de São Paulo", presentes neste mesmo tomo.

Representando, desse modo, um dos primeiros centros de difusão nacional da língua portuguesa trazida originalmente pelos colonizadores portugueses do século XVI, a região sob enfoque foi selecionada, portanto, devido à sua gênese e às suas características culturais e dialetais — fatos, aliás, que conferem importância ao nosso *locus* não só sob uma perspectiva estadual, mas também nacional.

5 METODOLOGIA

5.1 Procedimentos gerais em campo e materiais

Assim que chegávamos às localidades para conduzir a pesquisa de campo, o primeiro passo era localizar uma pessoa conhecida por locais, quem iniciava então a mediação entre inquiridor e inquiridos, de acordo com os perfis procurados (cf. seç. "Perfis dos informantes").

Feita a referida mediação, a regra era interagir com o informante em local escolhido por ele, de sua preferência. Na maioria das vezes, a interação tomava lugar na casa do entrevistado, mas também ela acontecia em seu local de trabalho ou na casa de amigos ou familiares.

Uma vez que estávamos acomodados, o Termo de Consentimento Livre e Esclarecido — disponível integralmente neste trabalho como apêndice, no Tomo VI — era apresentado. Ele é um documento de natureza ética que elaboramos em observância:

(a) à Declaração Universal dos Direitos Humanos (Resolução 217 A (III) da Assembleia Geral das Nações Unidas, de 10.12.1948);

(b) ao Código de Nuremberg de 1947;

(c) às diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos da Comissão Nacional de Ética em Pesquisa, Conselho Nacional de Saúde, Ministério da Saúde, aprovadas em sua 240ª Reunião Ordinária (Resolução nº 466, de 12.12.2012); e

(d) ao Código de Ética da Universidade de São Paulo (Resolução nº 4871, de 22.10.2001).

Ato contínuo, era feita uma explicação extra de que não se trataria de nenhum tipo de teste (psicológico, de conhecimento *etc.*), tampouco de formas "certas" *vs.* "erradas" (cf. seç. "Neutralização da problemática do 'certo' *vs.* 'errado'").

Em seguida, iniciava-se a elicitação+, começando com a Ficha do Informantem, preparada para levantar informações extralinguísticas. Na sequência, o instrumento denominado Temas para Discursos Semidirigidosm era aplicado; servindo, no entanto, apenas como facilitador ao engajamento interacional entre inquirido e inquiridor, com vistas à aplicação central subsequente do Questionário Semântico-Lexicalm (QSLm) e do Questionário Fonético-Fonológicom (QFFm), respectivamente.

Quanto a tais questionários, durante a sua aplicação, um conjunto de ilustrações de referentes semânticos+ tematizados em questões, reunido especificamente para a ocasião, era sequencialmente exibido ao entrevistado. Essa exibição era realizada na tela *full HD* 13.3 polegadas de um *laptop*, na qual as imagens eram expostas em modo "tela cheia" e de maneira sincronizada com a enunciação das respectivas perguntas. Adicionalmente, um *laser pointer* foi utilizado, tanto para claramente indicar partes específicas das ilustrações quando necessário, quanto para remota e mais agilmente avançá-las umas após as outras.

Toda a interação era registrada por um gravador de voz digital, com microfone *stereo*, taxa de áudio de 128kbps, alto falante 25/32" e com 75-17 mil Hz de resposta de frequência. Imediatamente após cada entrevista, dela se fazia dois *backups*, um no *hard disk* do *laptop* e outro num *flash drive*.

5.2 Perfis dos informantes

Com base na metodologia da Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*), oito informantes para cada uma das localidades da rede de pontos+ foram aleatoriamente convidados e entrevistados *in situ* como voluntários, quatro mulheres e quatro homens. No entanto, era imprescindível que todos eles satisfizessem quatro critérios. Eles precisavam ter:

(a) ou entre 18 e 36 anos de idade (GI), ou a partir de 55 (GII);
(b) ou baixa escolaridade (Cb; no escopo de abrangência entre analfabetismo e ensino médio incompleto), ou alta (Ca; a partir de estudos universitários, completos ou incompletos);
(c) moradia na localidade por no mínimo três quartos da vida; e

(d) residência ininterrupta na localidade ao longo pelo menos dos últimos cinco anos.

Entre as quatro mulheres por cidade, tínhamos necessariamente uma do perfil CaGII, uma do CaGI, uma do CbGII e outra do CbGI. O mesmo em relação aos quatro homens. No tocante à rede de pontos como um todo, tivemos, portanto, 80 informantes no cômputo geral.

5.3 Instrumentos de coleta

Para o levantamento dos dados empíricos, fizemos uso de quatro instrumentos metodológicos de coleta (eles encontram-se na íntegra entre os apêndices, no Tomo VI):

(I) Ficha do Informante^{m} (FdI^{m});
(II) Temas para Discursos Semidirigidos^{m} (TDS^{m});
(III) Questionário Semântico-Lexical^{m} (QSL^{m}); e
(IV) um Questionário Fonético-Fonológico^{m} (QFF^{m}).

(I)
A FdI^{m} — que modificamos da FdI em versão de 2001 da lavra do ALiB (2014) — que utilizamos é formada por 62 campos de coleta de informações extralinguísticas, informações essas que representam a potencial importância de auxílio ao empreendimento explicativo de possíveis influências externas a fatos linguísticos. No tocante ao preenchimento de tais campos, uma parte dele é baseada na observação do pesquisador, e outra se dá por questionamento direto ao entrevistado. Relativamente à natureza dos dados, eles dividem-se em duas classes: (a) uma "situacional" (em alusão à circunstância do inquérito e a características momentâneas do informante); e outra (b) "extrassituacional" (ref. ao *background* e a características tendencialmente constantes do entrevistado).

(a) classe situacional: (i) característica psicológica; (ii) espontaneidade; e (iii) cooperação; e

(b) classe extrassituacional: (i) ascendência étnica; (ii) religião; (iii) classe social; (iv) escolaridade; (v) pessoal; (vi) profissional; (vii) proficiência em idiomas; (viii) contato linguístico; (ix) contato com mídias; e (x) topodinâmica^{+}.

A FdI^{m}, apesar de não se estender por mais de duas páginas, permite o levantamento de números e valiosos dados extralinguísticos. Nela, duas das nossas alterações são bastante importantes, ambas contidas no item 1.1 do documento. Elas referem-se aos campos "GR" (em substituição a "sexo") e "estado conjugal" (no lugar de "estado civil").

Em relação ao campo "GR" (de 'gênero de registro/social'), é necessário dizer que a dimensão sexual — de natureza íntima e que pode ser parametrizada com valores para além do binarismo 'feminino' *vs.* 'masculino' convencional — não foi controlada em nossa pesquisa, mas sim a dimensão do gênero de registro/social da pessoa — de natureza social para a qual as únicas opções de registro cível disponíveis aos indivíduos submetidos ao ordenamento jurídico brasileiro ainda são 'feminino' e 'masculino'.

Quanto ao campo "estado conjugal" — diferentemente de "estado civil" —, ele refere-se ao estado cujas opções são factuais ('sozinho', 'acompanhado', 'viúvo' e 'separado'), e não jurídicas ('solteiro', 'divorciado', 'casado' *etc.*). Isso porque o interesse reside não em se saber qual é o estado civil do entrevistado (quem pode estar oficialmente casado, por exemplo, mas sem estar mais em contato com o marido ou com a esposa), e sim em se saber qual é a situação conjugal prática do inquirido. A essa ideia subjaz a premissa segundo a qual a convivência conjugal real é uma fonte de influência linguística presumida, *independentemente de qualquer estatuto legal.*

(II)

Em relação aos TDS^{m}, trata-se de uma versão moderadamente modificada dos TDS originários do ALiB (2014) em versão do ano de 2001. Esse instrumento metodológico é constituído por enunciados concebidos para incitar a produção de diferentes tipos textuais orais, mas ele apenas foi aplicado por nós como facilitador tático a favor do estabelecimento

de um engajamento interacional produtivo entre inquirido e inquiridor, com vistas à aplicação central do Questionário Semântico-Lexical^{m} (QSL^{m}) e do Questionário Fonético-Fonológico^{m} (QFF^{m}), embora essa interação preliminar também tenha sido gravada com cuidado metodológico e a abordagem a seus dados possa ser objeto de outro trabalho.

Os enunciados dos TDS^{m} são quatro, de (a) a (d) adiante. Neles: (i) "vc/sr" é uma abreviação para "você / (o)(a) senhor(a)"; (ii) o par de chaves, { }, encerra instrução/comentário ao inquiridor; e (iii) a <u>sublinha</u> sinaliza itens em relação de intercambialidade entre si.

(a) (Q1, ref. a 'relato pessoal':) "Relate um acontecimento marcante em sua vida (casamento, namoro...).";

(b) (Q2, ref. a 'comentário':) "Por que vc/sr gosta de assistir a X {*X = informação extraída da Ficha do Informante^{m} sobre hábito relacionado à televisão*}?";

(c) (Q3, ref. a 'descrição':) "Fale um pouco sobre <u>seu trabalho</u> / <u>suas atividades diárias</u>."; e

(d) (Q4, ref. a 'relato não pessoal':) "Conte uma lenda, história de assombração, qualquer causo de que talvez todos aqui já ouviram falar.".

(III)

Os dados sobretudo para a elaboração de um atlas semântico^{+}-lexical^{+} do português paulista (ou "dialeto^{+} caipira^{+}") falado na região do Médio Tietê foram colhidos a partir da aplicação do QSL^{m}. Esse instrumento metodológico é uma versão que estabelecemos após realizarmos mudanças no QSL de 2001 da lavra do ALiB (2014), que o concebeu para uso em território nacional brasileiro, após versões preliminares desde 1998, visando à comparabilidade metodológica dos dados coletados em diferentes regiões. As alterações implementadas nesse questionário encontram-se reportadas na seção "Modificações nos instrumentos metodológicos".

O QSL^{m} contém 170 questões onomasiológicas^{+} como *input* orientadas à obtenção de formas semasiológicas^{+} como *output*. Para potencializar esse objetivo, fizemos uso metodológico de imagens representativas de referentes semânticos tematizados nas questões. Tais ilustrações acompanham 102 perguntas, *i.e.*, 60% do questionário. Esse percentual só não foi maior em função do fato de que nem todos os referentes semânticos tematizados por esse instrumento de coleta se prestam à uma ilustração clara ou sem margens para imprecisões de ordem denotativa.

As imagens podem ser visualizadas integradas ao QSL^{m}, o qual se encontra disponível entre os apêndices, no Tomo VI. Lá, também há uma seção em que as fontes e os créditos das ilustrações são informados.

O QSL^{m}, seguindo o original, implica 14 áreas semânticas. A Tabela 5 adiante as exibe na ordem em que elas aparecem no questionário e também fornece a informação sobre quais e quantas questões estão associadas a cada uma dessas áreas.

	Área semântica	Questões	Quantidade de questões	
			Σ	%
1	Acidentes geográficos	1, 2, 4	3	1,76
2	Fenômenos atmosféricos	7-11, 13-21	14	8,24
3	Astros e tempo	22-33, 37, 38	14	8,24
4	Atividades agropastoris	42-58, 61-63	20	11,76
5	Fauna	64-71, 73-76, 78-80, 82-88	22	12,94
6	Corpo humano	89-102, 105-110, 113-120	28	16,47
7	Ciclos da vida	121, 125, 128-134	9	5,29
8	Convívio e comportamento social	136-139, 141-143, 145, 146	9	5,29
9	Religião e crenças	147-150, 153	5	2,94
10	Jogos e diversões infantis	155-167	13	7,65
11	Habitação	168, 170-175	7	4,12
12	Alimentação e cozinha	177-187	11	6,47
13	Vestuário e acessórios	188-193	6	3,53
14	Vida urbana	194-202	9	5,29
			170	100
			Σ	

Tabela 5: Áreas semânticas do QSL^{m} e suas questões

Os números não mencionados na coluna "Questões" tiveram suas respectivas perguntas retiradas do questionário, de acordo com o que se diz na já referida seção "Modificações nos instrumentos metodológicos".

Outro aspecto importante do QSL^{m} são as sugestões lexicais que se seguem às perguntas. Essas sugestões são, em determinada etapa do inquérito, apresentadas ao informante sob o intuito de se verificar se elas lhe são passivamente conhecidas ou não. Eventualmente, elas podem denotar referentes semânticos não equivalentes tanto àqueles de sugestões concorrentes, quanto àqueles encontrados nos respectivos verbetes de dicionários, especialmente dos tradicionais/normativos, *i.e.*, dos dicionários cuja abonação costuma não advir de base oral, e, por isso, eles tendem a negligenciar regionalismos vernaculares.

Ainda a respeito das sugestões, muitas delas já estão presentes no QSL original, mas fizemos acréscimos em dois momentos. O primeiro momento foi prévio à visita à rede de pontos, *i.e.*, ainda na fase de preparação à elicitação, e outro ocorreu durante a pesquisa de campo. As sugestões surgidas no segundo momento, sob estatuto diferenciado, provieram de circunstantes presentes no *set* de entrevista, as quais eram igualmente testadas e, assim, aceitas ou rejeitadas pelo informante. Todas as sugestões utilizadas, com a diferenciação de estatuto aludida, podem ser consultadas na seção "Universo das sugestões", no Tomo VI.

(IV)

Os dados máxime para a elaboração de um atlas fonético^{+}-fonológico^{+} do português paulista (ou dialeto caipira) falado na região do Médio Tietê foram colhidos a partir da aplicação do QFF ^{m}. Esse instrumento metodológico também é uma versão que estabelecemos após realizarmos mudanças no QFF de 2001 da lavra do ALiB (2014), que, assim como os demais instrumentos, o concebeu para uso em território nacional brasileiro, após versões preliminares desde 1998, visando à comparabilidade metodológica dos dados coletados em diferentes regiões. As alterações implementadas nesse questionário encontram-se reportadas na seção "Modificações nos instrumentos metodológicos".

O QFF^{m} contém 159 questões indutivas como *input* cujo propósito é produzir pronúncias genuínas (por parte dos falantes entrevistados) de certos vocábulos como *output*. Essa indução de pronúncia é realizada oral e/ou imageticamente pelo inquiridor, e para

potencializar a consecução do objetivo, fizemos uso metodológico de imagens representativas de referentes semânticos tematizados nas questões. Tais ilustrações acompanham 89 perguntas, *i.e.*, praticamente 56% do questionário. Esse percentual só não foi maior em função do fato de que nem todos os referentes semânticos implicados no instrumento de coleta se prestam à uma ilustração clara ou sem margens para imprecisões de ordem denotativa.

As imagens podem ser visualizadas integradas ao QFF^{m}, disponível entre os apêndices, no Tomo VI. Lá, também há uma seção em que as fontes e os créditos das ilustrações empregues são reportados.

Com o QFF^{m}, foi possível observar o comportamento tanto de vogais e *glides* — itens (a) a (e), a seguir —, quanto de consoantes — itens (f) a (n):

(a) ditongos decrescentes;
(b) vogais baixas postônicas;
(c) vogais altas (pretônicas e tônicas);
(d) vogais médias (pretônicas, tônicas e postônicas);
(e) /A/ e /O/ adjacentes a consoante nasal no padrão '(C)_.C[+nasal]';
(f) gerúndio;
(g) lateral palatal /ʎ/;
(h) nasal palatal /ɲ/;
(i) /L/ (em coda^{+} e como C2 em *onset* complexo);
(j) /S/ (em coda e intervocálico no padrão 'V._V');
(k) /R/ (em coda, em *onset* simples e como C2 em *onset* complexo);
(l) encontros consonantais interssilábicos (/f.t/, /p.n/ e /d.v/);
(m) palatalização (da lateral /l/ e das oclusivas /t, d/ diante da vogal alta [i]); e
(n) alternâncias ([b] *vs.* [v] em *onset* simples; [k] *vs.* [kʊ̯]; [ʒɪ] *vs.* [ʒẽɲ] como sílaba final átona; [i] *vs.* [ẽɲ] pós-nuclear em sílaba final átona; e [ĩ] *vs.* [ũ] em sílaba inicial).

5.4 Modificações nos instrumentos metodológicos

Como aludido anteriormente, realizamos uma edição na Ficha do Informante (FdI), nos Temas para Discursos Semidirigidos (TDS), no Questionário Semântico-Lexical (QSL) e no Questionário Fonético-Fonológico (QFF) do ALiB em versão do ano de 2001. As mudanças (que produziram as versões revisadas/modificadas FdI^{m}, TDS^{m}, QSL^{m} e QFF^{m}) foram implementadas após um teste preliminar anterior à pesquisa de campo. A edição incidente nas questões deu-se essencialmente por:

(I) mudança de campos/enunciados em geral e acréscimo de imagens ilustrativas de referentes semânticos tematizados nos questionários em específico;

(II) eliminação de perguntas e acréscimo de sugestões no caso particular do QSL^{m};

(III) aumento do número de explicitações dos nomes das classes de palavra associadas aos vocábulos a serem indutivamente pronunciadas pelo informante no caso particular do QFF^{m}; e

(IV) acréscimo de signos auxiliares nos questionários a favor da tarefa inquisitiva do pesquisador.

(I)

A respeito da FdI^{m}, duas das nossas alterações são bastante importantes, ambas contidas no item 1.1 do documento. Elas referem-se aos campos "GR" (em substituição a "sexo") e "estado conjugal" (no lugar de "estado civil").

Quanto ao campo "GR" (de 'gênero de registro/social'), é necessário dizer que a dimensão sexual — de natureza íntima e que pode ser parametrizada com valores para além do binarismo 'feminino' *vs.* 'masculino' convencional — não foi controlada em nossa pesquisa, mas sim a dimensão do gênero de registro/social da pessoa — de natureza social para a qual as únicas opções de registro cível disponíveis aos indivíduos submetidos ao ordenamento jurídico brasileiro ainda são 'feminino' e 'masculino'.

No tocante ao campo "estado conjugal" — diferentemente de "estado civil" —, ele refere-se ao estado cujas opções são factuais ('sozinho', 'acompanhado', 'viúvo' e 'separado'), e não jurídicas ('solteiro', 'divorciado', 'casado' *etc.*). Isso porque o interesse reside não em se saber qual é o estado civil do entrevistado (quem pode estar oficialmente casado, por exemplo, mas sem estar mais em contato com o marido ou com a esposa), e sim em se saber qual é a situação conjugal prática do inquirido. A essa ideia subjaz a premissa segundo a qual a convivência conjugal real é uma fonte de influência linguística presumida, *independentemente de qualquer estatuto legal*.

Já em relação às alterações de enunciados em geral, elas visaram à simplificação e/ou à adaptação de frases às imagens que utilizamos (no caso do QSL^{m} e do QFF^{m}). Essas ilustrações, com a função de potencializar o entendimento do informante em relação ao que se pergunta no inquérito, acompanham apenas aquelas questões que tematizam referentes semânticos que se prestam mais a uma ilustração clara ou sem margens para imprecisões de ordem denotativa.

(II)

A exclusão de questões do QSL^{m} incidiu naquelas preliminarmente tidas como não produtivas[1]. As remanescentes, no entanto, preservam o mesmo referente semântico (exceto no caso das questões 32, 86 e, de maneira parcial, 171 e 185), bem como conservam a mesma numeração que aquela presente no QSL original. Já o aditamento de mais sugestões no QSL serve ao propósito de ampliar os testes de léxico.

(III)

No QFF^{m} em particular, com atinência especificamente aos vocábulos a serem indutivamente pronunciados pelos informantes, foi promovido adicionalmente um aumento do número de explicitações dos nomes das classes de palavra associadas aos vocábulos que são *per se* categorialmente mais sujeitos a ambiguidades.

[1] A saber: Q3, 5, 6, 12, 34-36, 39-41, 59, 60, 72, 77, 81, 103, 104, 111, 112, 122-124, 126, 127, 135, 140, 144, 151, 152, 154, 169 e 176.

Acerca das vantagens do aumento do número dessas explicitações (de três para 16[2]), além destas oportunizarem mais clareza ao inquiridor a respeito do tipo de dado a ser esperado na ocasião de coleta, elas também ajudam o leitor a saber mais rapidamente (e com mais independência em relação aos enunciados das questões) do que se trata. Ainda, esse recurso releva-se especialmente útil na compreensão categorial dos vocábulos tematizados nas cartas fonético-fonológicas.

(IV)

Ademais, alguns signos foram acrescentados na qualidade de pequenos detalhes. O asterisco, *, é o comando "Mostre a respectiva imagem". O sinal de mais junto ao de interrogação, +?, é um lembrete para a execução da insistência (quando o caso). A abreviação "S" é o comando "Sugira" (quando o caso). O símbolo *hashtag* ou cerquilha, #, é o comando "Faça o gesto previamente convencionado" entre inquiridor e informante para casos de lacuna que deve ser preenchida. A exclamação entre parênteses duplos, ((!)), alerta para o fato de que a respectiva pergunta frequentemente requer paráfrases ou atos ostensivos pragmáticos^{+} adicionais por parte do inquiridor no trabalho de fazer com que o inquirido entenda bem a questão. Por fim, o recurso tipográfico **negrito** aplica-se a constituintes de perguntas a serem enunciados sob ênfase pelo pesquisador.

5.5 Técnica de entrevista ao QSL^{m} e *labels* às formas semasiológicas

Para o inquérito semântico-lexical, plasmamos e utilizamos um conjunto de procedimentos o qual estamos chamando de 'técnica de entrevista orientada à contundência responsiva/anuente' (simplesmente TE doravante). Ela é diretamente baseada na *técnica de três tempos* (TTT) desenvolvida por Thun (ADDU, 2000), no âmbito da Dialetologia Pluridimensional (DP).

[2] Dado que a Q7, 25 e 112 já contavam com explicitação de classe no QFF original, as novas explicitações envolvem as seguintes questões: 1, 14, 19, 23, 24, 40, 69, 74, 81, 103, 125, 126 e 149.

Essencialmente, a técnica thuniana consiste em:

(a) perguntar - T1;
(b) insistir - T2; e
(c) sugerir - T3.

No tempo 1 (T1), procura-se obter respostas espontâneas do entrevistado. No T2, o informante é estimulado a fornecer variantes^{+} coexistentes. No T3, apresenta-se-lhe uma relação de variantes não mencionadas nos passos anteriores, sob a intenção de se saber se elas são passivamente conhecidas pelo inquirido ou não.

Um postulado importante que subjaz a esse método da DP refere-se ao funcionamento da memória. Nela, existem itens ativamente disponíveis ao acesso e itens que o são passivamente. Essa premissa traz o benefício de subsidiar a reflexão se um item sugerido e aceito seria uma *forma velha/regressiva*, se seria uma *forma nova/progressiva*, ou ainda se seria uma *forma estável* na comunidade de fala.

Em sua estrutura elementar, os três tempos (ou passos, ou fases) podem ser assim ilustrados:

(a') "Como se chama *x*?";
(b') "Conhece outros nomes (alternativos a *y*)?"; e
(c') "Já ouviu z_n?".

Aí, *x* é uma variável que representa uma questão onomasiológica que serve de *input* com o objetivo de obter uma ou mais formas semasiológicas (FSs) como *output*. A variável *y* é uma eventual FS introduzida pelo inquirido como resposta no tempo 1 da técnica de entrevista. Já a expressão z_n simboliza uma ou mais covariantes^{+} possíveis de *y*.

Esses procedimentos são pensados para serem postos em prática o mais sistematicamente possível. Havendo ou não resposta espontânea, executa-se (na sequência) a insistência, após a qual se realizam sugestões. Esse seria o funcionamento padrão da técnica.

Todavia, a ordem dos três tempos pode espontaneamente mudar em função da situação particular. Por exemplo, uma forma sugerida pode subjetivamente despertar no informante a lembrança de outra lexia+, não dita antes (e não necessariamente semelhante ao item sugerido), ensejando, portanto, o surgimento de uma resposta elementarmente espontânea após o T3.

Um dispositivo metodológico auxiliar adicional da DP ao inquérito chama-se 'pluralidade de informantes', que pode ser tanto consecutiva quanto simultânea quando performada. Por meio desse recurso, cada perfil de inquirido é duplicado (ou triplicado *etc.*), gerando-se, pois, mais de um representante por perfil, o que potencializa a fidedignidade dos dados em perspectiva não só quantitativa e contrastiva, mas também qualitativa. Se indivíduos forem entrevistados juntos, ter-se-á a pluralidade simultânea. Se forem inquiridos à parte, tratar-se-á de pluralidade consecutiva.

A técnica que aplicamos, a TE, foi uma saída exequível às condições cronogramáticas das fases do levantamento empírico no interior das quais estivemos circunscritos, relativas ao número de diárias disponibilizadas para a estadia em campo.

A esse propósito, prioridades tiveram de ser estabelecidas. Uma delas foi a de atingir a 'contundência responsiva/anuente' do informante (noção aludida antes e retomada mais adiante), e então seguir às questões subsequentes do questionário. Amiúde, essa contundência foi encontrada nas respostas espontâneas.

Outra medida tomada foi a adaptação do recurso das sugestões, medida essa que se baseou na interação espontânea entre informante e circunstantes (familiares, colegas de trabalho ou amigos) presentes no *set* de inquérito, sem a pretensão, no entanto, de que ela fosse qualquer substituto à pluralidade de informantes *stricto sensu* da TTT.

Essa interação — estimulada uma vez que o *set* de trabalho se punha e observada em maior ou menor grau em todos os casos — dava liberdade aos circunstantes para responderem as questões. Todavia, suas respostas eram metodologicamente tratadas como variantes sob o *status* de sugestões — as quais, não raramente, coincidiam com as pré-organizadas — perante o informante, quem então as aceitava ou as rejeitava. Esse processo ocasionalmente punha variedades linguísticas+ de diferentes grupos (*e.g.* os etários) em contraste, sobre cujos comentários feitos pelo próprio informante a esse respeito se dará notícia na seção "*Corpus* metalinguístico-etnográfico", no Tomo V. Mais do que isso: esse

processo não raramente fornecia informação útil ao estabelecimento da 'ordem de relevância' (da qual se tratará mais adiante) das lexias.

As sugestões recorrentes assim surgidas revelavam-se inesperadamente variantes contextualmente relevantes, *status* esse justamente pretendido pelas sugestões pré-estipuladas. Esse fato implicou, por conseguinte, a substituição da "linearidade" e da "incondicionalidade" da apresentação das variantes pré-selecionadas (ambas características da TTT) pela "relevância contextual" da enunciação de variantes performada por circunstantes (perspectiva da TE). Tal substituição conferiu relativa celeridade à elicitação.

Mesmo tendo-se isso em conta, o foco principal do inquérito estava, entretanto, ajustado à obtenção de dados espontâneos contundentes, e os dados adicionais frutos do expediente acima mencionado com relação às sugestões devem ser considerados como informação extra neste trabalho, e não, portanto, como resultantes de uma aplicação *stricto sensu* da TTT.

Isso dito, tratemos de algumas noções constituintes da TE, quais sejam:

(I) contundência responsiva/anuente;
(II) controle de sugestões;
(III) estatutos qualitativos; e
(IV) ordem de relevância dos estatutos qualitativos.

Como aludido antes, essa técnica é pensada para alcançar a *contundência responsiva/anuente* do entrevistado. Ainda, prevê um *controle de sugestões* e preocupa-se com a *ordem de relevância* dos *estatutos qualitativos* das formas (pel)as quais o informante responde.

(I)

Por contundência responsiva/anuente, entende-se explicitamente a postura performada pelo entrevistado pela qual ele expõe convicção acerca de determinada forma semasiológica, seja esta ativamente dita por ele (*postura responsiva*), seja ele passivamente anuindo ao item (*postura anuente*). Todavia, houve casos em que se foi além da obtenção da contundência; extrapolação, monitorada, como reflexo sobretudo de continuidade da interação espontânea entre informante e circunstantes por ensejo das questões.

(II)

Já no tocante ao controle de sugestões, ele destina-se a verificar se a aceitação delas por parte do entrevistado dar-se-ia em função da existência prévia das formas no léxico mental passivo do falante ou, ao contrário, se a aceitação de sugestões ocorreria em função do efeito segundo o qual elas simplesmente lhe "fazem sentido" na ocasião.

Tal efeito pode ser observado por azo de questões *verbi gratia* "Como se chamam duas bananas que nascem grudadas?" (Q43) e "Quando a mãe não tem leite, e outra mulher amamenta a criança, como chamam essa mulher de quem a criança então mama?" (Q128), para as quais as sugestões <gêmeas> e <mãe de leite>, respectivamente, podem ser aceitas ocasionalmente porque seu valor semântico de face em princípio o permite, e não porque necessariamente estejam presentes no léxico mental do inquirido como respostas *a priori* adequadas às Q43 e 128.

Por isso, em casos como esses, perguntas de controle — como "Você/O(A) senhor(a) já falava assim (ou ouvia isso) antes?" e variações — foram utilizadas como auxílio na tentativa de identificar o real *status* da forma semasiológica, *i.e.*, se ela já pertencia ao léxico passivo do informante ou se ela foi aceita porque simplesmente "fez sentido".

Outras noções importantes a explicitar são os *estatutos qualitativos* das lexias e a sua *ordem de relevância*.

(III)

Cada forma (pel)a qual o entrevistado responde pode ter um de cinco estatutos qualitativos primários distintos:

(a) forma espontânea positiva - FP(*x*);
(b) forma espontânea negativa - FN;
(c) forma resultante de insistência - FI(*x*);
(d) forma sugerida aceita - FA(*x*); e
(e) forma sugerida rejeitada - FR.

Como respostas espontaneamente proferidas pelo entrevistado — sendo *w* a seguir uma variante qualquer —, prototipicamente (a) uma FP é dada na forma "É *w*" (e variações) e (b) uma FN, na forma "Não é *w*" (e variações). Quando o informante fornece alguma variante após a pergunta "Conhece outros nomes?", tem-se (c) uma FI.

Quando o entrevistado aceita uma sugestão, trata-se de (d) uma FA, e, quando ele nega alguma, o caso é de (e) uma FR. A reação do indivíduo em relação a uma forma que ele rejeita dá-se tipicamente via formas que estas duas frases ilustram: "Nunca ouvi *w*" ou "Já ouvi *w*, mas eu não associava a isso {que se pergunta}".

No tocante ao índice *x* visto nos itens (a), (c) e (d) na enumeração mais acima, ele assume o valor 1, 2, e assim por diante, quando houver mais de uma forma semasiológica sob o mesmo estatuto, respeitando-se a sequência de aparecimento por princípio inicial. Por outro lado, quando houver apenas uma forma com um determinado estatuto, prescindir-se-á de índices. Por exemplo, quando se pergunta "Como se chamam duas bananas que nascem grudadas" (Q43), e o falante responde apenas <felipe>, essa resposta tem estatuto FP. Caso, porém, ele responda <felipe> e ainda <inconho>, o estatuto é FP1 para <felipe> e FP2 para <inconho>. O mesmo procedimento aplica-se *mutatis mutandis* aos demais estatutos.

Além dos estatutos qualitativos primários acima tratados, há ainda três secundários:

(f) retificação - Ret(*Yx*);
(g) hesitação - Hes; e
(h) suspeição de designação - Sus.

(f) No curso do inquérito, se o entrevistado enunciar uma forma que retifique lexical ou foneticamente uma outra anterior, de qualquer estatuto qualitativo, esse evento será sinalizado com 'Ret'. (g) Quando houver forte hesitação na enunciação ou na rejeição de determinada forma, ou ainda na anuência a ela, a sinalização será 'Hes'. (h) Por fim, quando uma forma de *ocorrência única* for considerada dubitável no universo de respostas, essa situação será marcada por 'Sus'.

A variável *Y* de (f) será saturada por um dos estatutos qualitativos primários sobre o qual os estatutos qualitativos secundários terão escopo (*e.g.* "Ret(FP1)").

Todos os estatutos qualitativos, de (a) a (h), serão *labels* (ou etiquetas) a acompanhar as respostas ao QSL^m.

(IV)
Com relação à ordem de relevância de estatutos primários, sua configuração *a priori* é:

(a) FP(*x*);
(b) FI(*x*);
(c) FA(*x*);
(d) FN; e
(e) FR.

Quanto aos estatutos secundários, a ordem *a priori* é:

(f) Ret(*Yx*); e
(g) Hes.

Importante dizer e enfatizar que a etiqueta 'Sus' não interferirá no ordenamento de relevância do estatuto primário do item etiquetado.

Em todos os casos *a priori*, o índice 1 terá precedência sobre o índice 2, e assim por diante.

O ordenamento *a priori*, no entanto, pode mudar *a posteriori*. Por exemplo, uma FA pode ser mais relevante do que uma FP, como o diálogo na próxima página ilustra (α = inquiridor; β = informante).

Diálogo ilustrativo (ref. à Q168)

α — Como se chama aquela pecinha de madeira que gira ao redor de um prego para fechar porta, janela?

β — Tem um nome... Uma "trinca", uma "trava"... Mas tem outro nome... {Vários segundos se passam.}

α — Já ouviu "tramela"?

β — Isso! "Tramela" que eu estava tentando lembrar. Isso é uma "tramela".

Dado o *cotexto* concreto em específico (e o *contexto* no qual o cotexto se insere e sobre o qual ele dá pistas), nota-se no diálogo acima que a FA <tramela> figura como mais relevante do que as FP1 <trinca> e FP2 <trava>, correspondentemente à performance do próprio entrevistado. Veja-se, portanto, que a observação da performance do indivíduo é a chave para o estabelecimento da ordem de relevância das formas (pel)as quais ele responde. No caso em tela, essa ordem seria respectivamente: FA, FP1 e FP2.

Em generalidade, a ordem *a priori* de relevância manter-se-á desde que não haja informação concreta para rearranjos. Casos incertos serão resolvidos no bojo do inquérito via perguntas ao informante a esse respeito.

Complementarmente, outro controle é constituinte da TE; a saber, pedidos ao inquirido para confirmar aquelas respostas que, entre outras possibilidades, sejam ditas em voz muito baixa, pareçam inusitadas num primeiro momento, impliquem aparente contrassenso *etc.*, como, por exemplo <arroz-doce> como resposta a "Como se chama este alimento {*mostrar imagem*}, feito com **grãos de milho branco**, **coco** e **canela**?" (Q181).

5.6 Neutralização pragmática da problemática do "certo" *vs.* "errado"

Via de regra, antes de que o inquérito propriamente fosse iniciado, de maneira simples e breve informava-se oralmente ao informante — em adição ao teor do documento Termo de Consentimento Livre e Esclarecido, arrolado entre os apêndices, no Tomo VI — que a

entrevista tinha por objetivo registrar a fala popular/natural/espontânea/autêntica da localidade. Eventualmente, não obstante, ocorria a necessidade de que a explicação acerca da natureza da interação fosse retomada, por ensejo de comentários do falante como "Não sei se está certo assim", "Está certo falar assim?" e variações (*i.e.*, ref. à problemática do "certo" *vs.* "errado"), que se seguiam às respostas ao questionário.

Ocasionalmente, durante tal retomada explicativa, surgia a demanda de se dizer que não se tratava de nenhum tipo de teste de conhecimentos gerais, tampouco de medição de tempo de reação responsiva, mas sim apenas de registrar pura e simplesmente, como já dito, a fala popular/natural/espontânea/autêntica — explicação essa que, sempre parafraseada em função do perfil do entrevistado, podia aprofundar-se consoante o interesse ou a demanda direta ou indireta do próprio entrevistado.

Em todo caso, a emersão da problemática do "certo" *vs.* "errado" era frequente, e o empenho em neutralizá-la respondia ao intento pragmático de montar-se um *corpus* espontâneo e autenticamente representativo da localidade. Em caso de ausência dessa medida, o risco era a permanência do entrevistado numa posição de fala regida, ou perpassada, por juízos de valor cuja fonte remonta à milenar tradição gramatical no Ocidente penetrada por *doxas*, atualmente ainda pouco sensível aos estudos linguísticos científicos modernos de base empírica movidos por *epistemes*. Segundo o observado, a tentação à permanência na aludida posição de fala acometia pessoas virtualmente de todos os perfis de informantes, independentemente das diferentes influências diretas ou indiretas oriundas do saber gramatical e independentemente das contrapartes concretas dessas influências de fundo manifestadas oralmente.

Nesse sentido, quando pontualmente necessário, falava-se da problemática do "certo" *vs.* "errado", sob o intuito de caracterizá-la prática e compreensivelmente mediante demonstrações concretas, nos moldes dos diálogos ilustrativos 1 e 2 na próxima página (α = inquiridor; β = entrevistado).

Diálogo 1 (ref. à Q69)

α — Como se chama uma galinha sem rabo?

β — "Corve" {pronunciado ['kɔɻ.vɪ]}, mas não sei se essa palavra existe.

α — Como assim?

β — Não sei se ela está no dicionário.

α — Mas os dicionários não registram necessariamente todas as palavras que existem. O senhor utiliza essa palavra, não é?

β — Sim.

α — Seus amigos, familiares também?

β — Sim, o pessoal do sítio.

α — Então, tudo indica que "corve", pelo menos no meio em que o senhor vive, existe, independentemente de livros, não é verdade?

β — É.

α — Nossa conversa aqui trata justamente disso. Fique à vontade, portanto, para dizer as palavras que o senhor fala no dia a dia. É exatamente isso de que preciso.

β — Tudo bem.

Diálogo 2 (ref. à Q114)

α — Como se chama quem não tem uma perna?

β — "Saci", mas é mais gíria e tem um aspecto pejorativo. Deve ter um termo correto ou mais adequado, ou ainda algum que a medicina use.

α — Você tem razão, mas o interesse nesta nossa interação não é por palavras consideradas *corretas* ou *certas*, ou por aquelas mais ou menos *adequadas*, muito menos *técnicas*, mas sim simplesmente pelo *que se usa* aqui, independentemente de qualquer juízo de valor ou classificação. Se "saci" é uma palavra usada aqui para referir-se a "quem não tem uma perna", então ela é uma resposta não só possível, mas também válida, e é isso que procuro. Se você conhecer mais palavras além de "saci", elas também serão superbem-vindas. Isso vale a todas as perguntas que lhe farei.

β — Entendi.

Entretanto, mesmo assim ocasionalmente houve comentários acerca do "certo" *vs.* "errado" por parte dos entrevistados, que foram objeto de registro e estão reportados na seção "*Corpus* metalinguístico-etnográfico", no Tomo V.

6 ANÁLISES DO /R/ EM CODA[1]

Pela perspectiva da Dialetologia Pluridimensional (RADTKE e THUN, 1996; THUN, 2000, 2005 *etc.*), a partir de agora vamos abordar as diferentes realizações fonéticas+ e de apagamento do arquifonema+ /R/ em quatro posições estruturais de coda silábica+, as quais perfazem quatro variáveis. Trataremos de tais variáveis linguísticas em perspectiva com quatro dimensões de análise (ou quatro variáveis extralinguísticas) — diatópica+, diastrática+, diageracional+ e diagenérica+ —, as quais foram sistematicamente controladas em nossa pesquisa.

Com isso, além de apresentarmos dados estatísticos, poderemos verificar a distribuição de variantes+ tanto pelo espaço físico (eixo horizontal), quanto pelo espaço social (eixo vertical) da região paulista do Médio Tietê. Ainda, poderemos obter subsídios para identificar quais são as variáveis extralinguísticas a desempenharem papel de fatores externos na determinação do (ou de influência sobre o) comportamento das variáveis linguísticas.

As variáveis linguísticas (ou, mais precisamente, fonético-fonológicas+) com as quais nos ocuparemos são:

(a) /R/ em coda externa de desinência infinitiva (DI) seguida por pausa;
(b) /R/ em coda externa em sílaba tônica (que não de DI) seguida por pausa;
(c) /R/ em coda interna em sílaba tônica; e
(d) /R/ em coda interna em sílaba átona.

Para nosso engajamento analítico, contaremos com uma base de dados empíricos que coletamos mediante a aplicação, a 80 informantes, de 13 questões integrantes do Questionário Fonético-Fonológico m (QFF m), desenhadas originalmente pelo Atlas Linguístico do Brasil (ALIB, 2014) para a observação de contrapartes (a)fonéticas+ de vários (arqui)fonemas além do /R/. Tais questões seguem na próxima página apenas com as variáveis relativas ao arquifonema /R/ em coda sinalizadas, com sublinha e em **negrito**).

[1] A referência temporal dominante nesta seção — a qual lança mão, por exemplo, da conjugação verbal no futuro do presente do indicativo — tem o propósito fraseológico (ou estilístico) de associar autor e leitor na condução da leitura. Nesse sentido, o "agora" estará semanticamente ancorado no tempo de leitura do leitor.

(a') Q18 <VARRE**R**>, 36 <BOTA**R**> e 153 <SAI**R**>;
(b') Q25 <COLHE**R**> (subst.), 61 <CALO**R**> e 129 <MULHE**R**>;
(c') Q105 <CE**R**TO> e 158 <ESQUE**R**DO>; e
(d') Q12 <TO**R**NEIRA>, 27 <FE**R**VENDO>, 110 <PE**R**DÃO>, 144 <PE**R**FUME> e 152 <PE**R**GUNTAR>.

Todos os dados coletados referentes a tais questões (e às demais do QFF^{m}) podem ser examinados na seção "Fones (*corpus* primário)", constante do Tomo V.

As respostas às questões mencionadas nos itens de (a') a (d') foram organizadas com vistas à constituição de grupos representantes das situações fonológicas do arquifonema /R/ referidas nos itens de (a) a (d), das quais uma é também morfológica^{+}, (a). Tais grupos foram cartografados, o que originou 12 cartas linguísticas especiais sobre o /R/ em coda. Essas cartas especiais foram estruturadas em:

(A) uma única categoria: 4C;
(B) três séries: CxGy, CxWz e WzGy; e
(C) quatro especificidades: E1, E2, E3 e E4.

(A)

'4C' denota 'quatro covariantes^{+}' como contrapartes (a)fonéticas que foram cartografadas; nomeadamente: (i) a aproximante retroflexa, [ɻ] (conhecida como "erre retroflexo" ou ainda como "erre caipira^{+}"); (ii) o tepe (orig. *tap*) alveolar, [ɾ] (conhecido também como "*flap*", "erre fraco", "erro brando", "erre simples" ou ainda "vibrante simples"); (iii) o apagamento, [Ø] (*i.e.*, a ausência de fone^{+}); e, finalmente, (iv) a vibrante alveolar, [r] (conhecida como "vibrante múltipla" ou ainda como "erre forte").

(B)

Os nomes das séries CxGy, CxWz e WzGy abstratamente aludem a diferentes dimensões (ou variáveis) extralinguísticas que foram controladas em nossa pesquisa e com as quais os dados linguísticos serão cruzados. Tais cruzamentos (ou correlações) serão exibidos tanto na cartografia propriamente dita, quanto nos respectivos gráficos quantitativos inseridos nas cartas linguísticas.

Na composição dos nomes dessas séries, 'C' refere-se a 'classe escolar' (ou 'nível escolar'); 'G', 'geração' (ou 'grupo etário'); e 'W', 'gênero de registro/social do indivíduo^{+}' (feminino *vs.* masculino). Observa-se também a presença de 'x', 'y' e 'z', as quais são variáveis lógicas. A variável *x* é saturada ou pelo parâmetro 'a' de 'classe escolar alta' (a partir de estudos universitários, completos ou incompletos), ou 'b' de 'classe escolar baixa' (no escopo de abrangência entre analfabetismo e ensino médio incompleto). Já a variável *y* é saturada ou pelo parâmetro 'II' de 'geração mais velha' (a partir de 55 anos de idade), ou 'I' de 'geração mais nova' (entre 18 e 36). Finalmente, a variável 'z' é saturada ou pelo parâmetro 'f' de 'feminino' (de, mais exatamente, 'gênero de registro/social feminino'), ou 'm' de 'masculino' (de 'gênero de registro/social masculino').

Ou seja, cada série tem quatro parametrizações possíveis:

(α) CxGy → CaGII, CaGI, CbGII e CbGI;
(β) CxWz → CaWf, CaWm, CbWf e CbWm; e
(γ) WzGy → WfGII, WfGI, WmGII e WmGI.

Essas parametrizações podem ser assim lidas:

(α') CaGII: classe 'a' da geração 'II'; CaGI: classe 'a' da geração 'I'; CbGII: classe 'b' da geração 'II'; e CbGI: classe 'b' da geração 'I';

(β') CaWf: classe 'a' feminina; CaWm: classe 'a' masculina; CbWf: classe 'b' feminina; e CbWm: classe 'b' masculina; e

(γ') WfGII: mulheres da geração 'II'; WfGI: mulheres da geração 'I'; WmGII: homens da geração 'II'; e WmGI: homens da geração 'I'.

Nas cartas linguísticas, cada um desses grupos (ou parametrizações) estará representado, ocupando uma de quatro posições de uma cruz, ┼. Numa legenda que acompanhará cada carta, as posições específicas de cada grupo serão exibidas na cruz. *Essa será a chave para a interpretação das cartas linguísticas.*

É importante salientar ainda que oito pessoas por localidade geográfica visitada participaram do inquérito mediante o qual os dados com que ora estamos trabalhando foram gerados. Isso significa que cada posição de cada uma das cruzes na cartografia será então representada por dois indivíduos, A e B, que, em cada série, compartilham entre si dois atributos por um lado, e, por outro, diferenciam-se entre si por um atributo.

Por exemplo, com respeito ao grupo CaGII (da série CxGy), os participantes A e B correspondentes possuem alta escolaridade e são velhos, mas um deles é feminino, e outro, masculino. Relativamente ao grupo CaWf (da série CxWz), trata-se de duas mulheres com alta escolaridade, mas de gerações distintas. E assim por diante. Desse modo, as combinações representadas pelas séries implicam diferentes cruzamentos de dados. Esses cruzamentos serão visíveis na cartografia e nos gráficos estatísticos que acompanharão as cartas.

Mencione-se ainda, como se verá nas cartas, que apenas um símbolo aparecerá em cada uma das posições da cruz, apesar de cada posição representar dois informantes, A e B, como já dito. Para entendê-lo, vejamos, na Tabela 6 abaixo (em que Ø = nada), um conjunto de primeiro grau de símbolos polifórmicos cartográficos e sua lógica.

Se A responder ... ,	e se B responder ... ,	o símbolo será ... .
{ou vice-versa}		
◔	◔	◔
◔	◔	◑
◔	◕	◕
◔	◕	◕
◔	○	◔
◔	Ø	◔
◔	◔	◔
◔	◕	◕
◔	◕	◕
◔	○	○
◔	Ø	◔

Se A responder ... ,	e se B responder ... ,	o símbolo será ... .
{ou vice-versa}		
◕	◕	◕
◕	◕	◕
◕	○	◕
◕	Ø	◕
◕	◕	◕
◕	○	◕
◕	Ø	◕
○	○	○
○	Ø	○
Ø	Ø	▮
-	-	-

Tabela 6: Lógica da simbologia cartográfica (para cartas especiais, apenas o primeiro grau)

Como se infere dessa lógica, e lembrando-se do dito antes, a prioridade nas cartas será mostrar cartograficamente a ocorrência das covariantes mais frequentes (◔, ◶, ◵, ◴) — em detrimento das demais (◌) ou da ausência delas (▮) — distribuídas pelas séries CxGy, CxWz e WzGy. Porém, a quantificação exibida nos gráficos que acompanharão as cartas computará absolutamente todos os dados válidos, incluindo assim, portanto, aqueles que não estarão visíveis na cartografia. Por seu turno, a visibilidade completa dos resultados oriundos da base de dados será oferecida na seção "Fones (*corpus* primário)", no Tomo V.

Ainda, é preciso enfatizar o fato de que, para a elaboração das cartas especiais sobre o /R/ em coda, resultados oriundos de mais de uma questão a respeito de uma mesma situação fonológica (e eventualmente também morfológica) comporão a base de dados de cada carta linguística, como já aludido no início da presente seção. Daí ser a Tabela 6 continente de uma simbologia e lógica de primeiro grau, *i.e.*, ela cobre apenas os resultados de uma única questão. Logo, havendo resultados oriundos de mais questões do QFF^{m} sendo considerados para uma abordagem referente a uma mesma e única situação estrutural (ou variável linguística), uma simbologia e uma lógica estendidas serão necessárias.

Para introduzi-lo, seja uma circunstância na qual a cartografia ampare-se nos resultados de três questões, tal como é o caso das cartas sobre o /R/ em coda externa de desinência infinitiva seguida por pausa. Isso significa que ambos os informantes representantes por perfil por localidade (indivíduos A e B) contribuem, no total, com três turnos responsivos cada. Essa contribuição é exemplificada na Tabela 7 abaixo, que exibe cinco cenários (de vários outros) possíveis.

Turnos responsivos	n. 1		n. 2		n. 3			Símbolo resultante
Informantes	A	B	A	B	A	B		
Cenário 1	◔	◵	◶	◌	◔	◔	=	◓
Cenário 2	◵	◶	◔	◴	▮	◶	=	◓
Cenário 3	◴	◵	◶	◶	◴	◶	=	◓
Cenário 4	◌	◔	◔	◴	◶	◔	=	◓
Cenário 5	◔	◵	▮	◔	◌	◴	=	◓

Tabela 7: Exemplos da lógica da simbologia cartográfica estendida (para cartas especiais)

Com isso, note-se que a base documental das cartas especiais é multiplicada. Enquanto cada uma das nossas outras cartas contará com um total máximo possível de 80 dados válidos, uma carta especial montada com dados oriundos de três questões, por exemplo, contará com um total máximo possível de 240 dados válidos (3 x 80).

(C)

Cada uma das quatro diferentes situações estruturais do arquifonema /R/ em coda já antes referidas estará associada, nas cartas linguísticas, ao termo 'especificidade' (E), conforme as equivalências abaixo.

(i) coda externa de desinência infinitiva (DI) seguida por pausa = E1;
(ii) coda externa em sílaba tônica (que não de DI) seguida por pausa = E2;
(iii) coda interna em sílaba tônica = E3; e
(iv) coda interna em sílaba átona = E4.

Além de se encontrar, na parte de cima das cartas, a explicitação da situação estrutural que elas representarão, elas também incluirão um título técnico equivalente, posicionado no canto superior direito e entre parênteses. De maneira codificada, esse título fornecerá a identidade detalhada das cartas. No interior da expressão (encabeçada pela sigla "CFF", de 'carta fonético-fonológica') que denotará essa identidade, as quatro especificidades serão assim referidas:

(i') E1 = "+SDI{'(C)V/R/||}";
(ii') E2 = "-SDI{'(C)V/R/||}";
(iii') E3 = " '(C)V/R/."; e
(iv') E4 = "(C)V/R/.".

O termo '+SDI' exprime 'trata-se de sílaba de desinência infinitiva', enquanto '-SDI' é 'não se trata de sílaba de desinência infinitiva'; o par de chaves, { }, demarca/organiza escopo lógico; o apóstrofo, ', significa que, à sua direita, trata-se de uma sílaba tônica; o par de parênteses, ( ), encerra elemento facultativo; as letras 'C' e 'V' representam '(semi)consoante' e 'vogal', respectivamente; o símbolo '/R/' denota o arquifonema para os róticos[+] (e para

o apagamento) possíveis; o par de barras paralelas verticalmente, | |, designa 'pausa'; e, por fim, o ponto, ., indica separação silábica, sugerindo que, à sua direita, se encontra(m) outra(s) sílaba(s) da mesma palavra.

Isso posto, queremos ressaltar que as análises adiante ancorar-se-ão nas cartas especiais sobre o /R/ em coda que elaboramos (as quais terão seu tamanho reduzido a seguir, mas estarão em tamanho normal e com alta qualidade no Tomo III), cuja sequência de abordagem iniciará com a série CxGy, passando por CxWz e finalizando com WzGy. Nesse sentido, e a fim de que haja linhas condutoras para as ponderações, os tópicos a serem privilegiados são acerca:

(I) da menor variação (registro de uma e duas variantes);
(II) da maior variação (registro de três e quatro variantes); e
(III) de contrastes quantitativos.

Os itens (I) e (II) estarão baseados especificamente na cartografia propriamente dita, e o item (III) estará fundamentado no gráfico inserido nas cartas; gráfico esse que estará voltado à distribuição dos fatos linguísticos através dos diferentes grupos de informantes.

Nas seções subsequentes, esses tópicos continuarão a ser identificados pela mesma numeração romana em letras maiúsculas — *i.e.*, (I) estará associado ao tópico da menor variação; (II), ao da maior variação; e (III), ao de contrastes quantitativos. A essa numeração, porém, será adicionado ".I" para referir-se à série CxGy, ".II" para CxWz e ".III" para WzGy.

Por fim, lembramos que as localidades da nossa rede de pontos^{+} estarão codificadas nas cartas: P1 = Santana de Parnaíba; P2 = Pirapora do Bom Jesus; P3 = Araçariguama; P4 = São Roque; P5 = Sorocaba; P6 = Itu; P7 = Porto Feliz; P8 = Tietê; P9 = Capivari; e P10 = Piracicaba. Para mais detalhes sobre os componentes das cartas em geral, veja-se a seção "Treinamento no *software* de cartografação", no presente tomo.

6.1 Externa

6.1.1 De desinência infinitiva (DI) seguida por pausa

Seguindo o que introduzimos há pouco, abordaremos o arquifonema /R/ inicialmente em coda externa de desinência infinitiva seguida por pausa em relação primeiramente à série CxGy, cuja carta a servir de base é a que a Figura 5 abaixo apresenta.

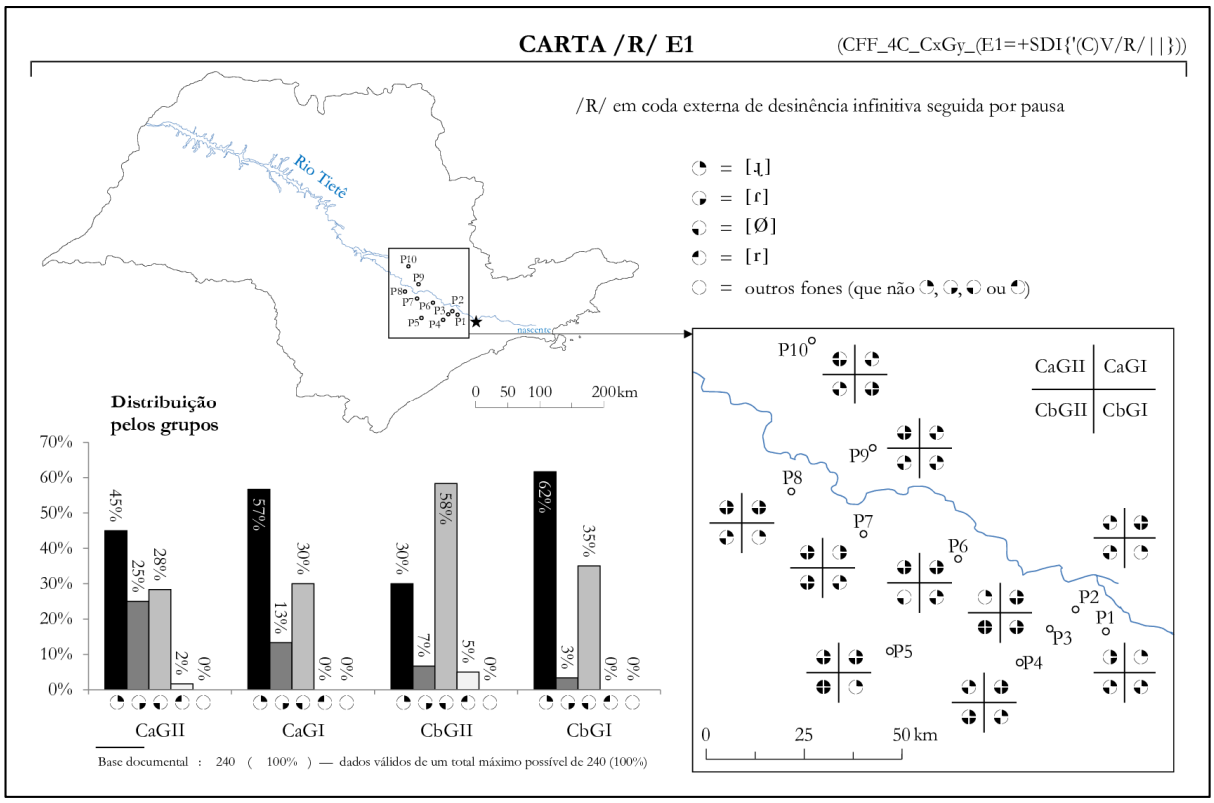


Figura 5: Carta /R/ da série CxGy e especificidade 1 (carta miniaturizada)

(I.I)

Na parte inferior direita da carta, encontra-se a parte geográfica ampliada que contém as localidades da nossa pesquisa. Cada localidade (ou ponto) tem uma cruz, constituída por quatro quadrantes. Cada quadrante representa um grupo de informantes, de acordo com a legenda posicionada no canto superior direito da ampliação. Na carta trazida pela Figura 5, vê-se, então, que alguns quadrantes denotam ausência de variação, *i.e.*, de registro de apenas uma variante, com seis instâncias⁺, distribuídas por seis das dez localidades da nossa rede pontos: Santana de Parnaíba (P1), Pirapora do Bom Jesus (P2), Araçariguama (P3),

Sorocaba (P5), Itu (P6) e Tietê (P8). Exceto no P6, em todos esses pontos a variante única em questão é [ɻ]. No P6, trata-se de [Ø], de autoria enunciativa do grupo CbGII. O grupo que foi predominante na produção de apenas uma variante é CbGI, responsável pelas instâncias de [ɻ] no P2, no P5 e no P8.

Quanto à variação constituída por duas covariantes nos quadrantes, surgiram duas configurações variacionais⁺: [ɻ]-[ɾ] (com duas instâncias) e [ɻ]-[Ø] (predominante, com 14 instâncias). A primeira configuração foi perpetrada pela Ca, *i.e.*, por informantes da classe alta, em duas localidades: Santana de Parnaíba (P1) e Porto Feliz (P7). Especificamente, no P1 pelo grupo CaGII, e no P7 pelo grupo CaGI. Já a segunda configuração, [ɻ]-[Ø], revelou-se diatopicamente mais difusa. Considerando-se todas as suas instâncias (que somaram 14), essa segunda configuração ocorreu pelo menos duas vezes em cada um dos quadrantes possíveis por localidade. A esse propósito, um destaque que queremos fazer reside no fato de que 10 das 14 instâncias foram articuladas por indivíduos da Cb, *i.e.*, por falantes pertencentes à classe escolar baixa.

(II.I)

Já com relação à variação constituída por três covariantes, houve duas configurações: [ɾ]-[Ø]-[r] (com uma instância) e [ɻ]-[ɾ]-[Ø] (com 15 instâncias). Observou-se a primeira apenas em Piracicaba (P10), no grupo CaGII, e a segunda em todos os pontos da rede, exceto em Santana de Parnaíba (P1). Das 15 instâncias de [ɻ]-[ɾ]-[Ø], 12 foram performadas pela Ca, *i.e.*, por falantes com nível escolar baixo.

A respeito da variação plasmada por quatro covariantes, ela foi percebida em duas localidades: Araçariguama (P3) e Sorocaba (P5). Nelas, as covariantes em questão são [ɻ], [ɾ], [Ø] e [r], e o grupo de informantes que foi responsável por essa produção variacional, em ambos os locais, é CbGII, *i.e.*, são os indivíduos mais velhos com escolaridade baixa.

(III.I)

Quanto aos aspectos quantitativos, chamamos a atenção para um padrão quantitativo verificado nos grupos CaGII, CaGI e CbGI; padrão esse segundo o qual houve prevalência da covariante [ɻ], seguida por [Ø] e então por [ɾ]. No grupo CbGII, por seu turno, a covariante [Ø] superou a produtividade de [ɻ]: foram 58% de frequência relativa da

primeira contra 30% da segunda. Entre os quatro grupos, o CbGI é o que mais produziu [ɻ] e o que menos produziu [ɾ].

Em relação à covariante [r], sua produção não foi observada na fala da geração mais nova, GI. Ainda, uma simetria invertida entre os grupos CaGI e CbGII — que são, do ponto de vista de seu perfil, completamente opostos entre si — atrai a atenção: o primeiro realizou [ɻ] (57%) e [Ø] (30%) quase exatamente na mesma proporção em que o segundo realizou [Ø] (58%) e [ɻ] (30%).

Agora, tratemos dos resultados fonéticos e de apagamento do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa pelo ângulo da série CxWz, a partir da carta linguística que a Figura 6 adiante expõe.

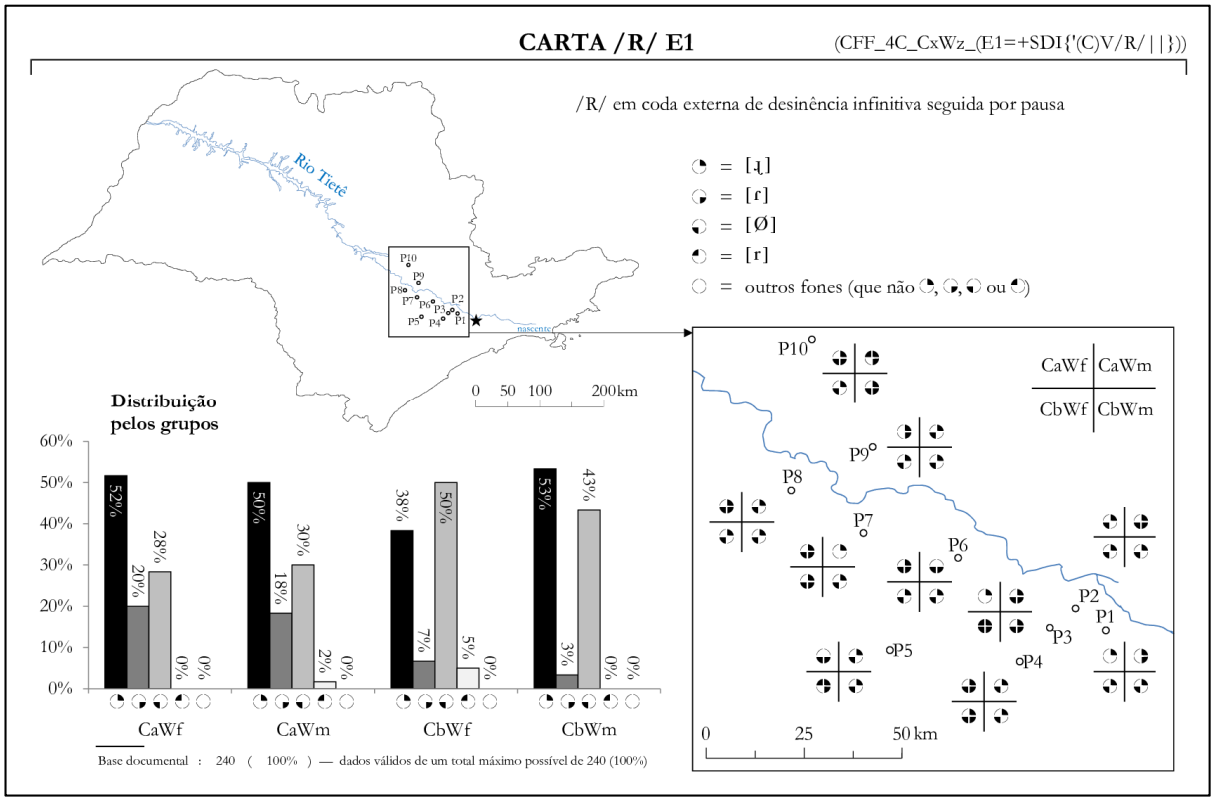


Figura 6: Carta /R/ da série CxWz e especificidade 1 (carta miniaturizada)

(I.II)

Como se percebe na carta acima, os dois quadrantes superiores de Santana de Parnaíba (P1), Araçariguama (P3) e Porto Feliz (P7) apresentam um grupo cada em que não houve variação. Trata-se da configuração única [ɻ], com maior produção pelo grupo CaWf (P1 e

P3), seguido pelo grupo CaWm (P7). Note-se que ambos esses grupos de inquiridos compartilham entre si o fato de pertencerem à classe escolar alta.

Também é visível que a variação estabelecida entre duas covariantes aconteceu mediante três configurações: [ɻ]-[ɾ], [ɾ]-[Ø] e [ɻ]-[Ø]. As duas primeiras obtiveram duas instâncias cada, e a terceira, mais comum, 19. A configuração [ɻ]-[ɾ] adveio em dois pontos distantes entre si, Santana de Parnaíba (P1) e Capivari (P9), nos grupos CaWm e CaWf, respectivamente; portanto, ambos da classe escolar alta. Com referência à configuração [ɾ]-[Ø], por sua vez, suas duas instâncias estiveram distribuídas em localidades vizinhas, Sorocaba (P5) e Itu (P6), nos grupos CaWf e CaWm, respectivamente; por consequência, ambos também da classe escolar alta. Já no tocante à terceira e última configuração, [ɻ]-[Ø], ela foi realizada em toda a rede de pontos, exceto em Araçariguama (P3), de modo a ter figurado pelo menos uma vez em cada um dos quadrantes possíveis. Indivíduos da Cb, *i.e.*, da classe escolar baixa, responderam por 14 das 19 instâncias de [ɻ]-[Ø]. No grupo CaWf, essa configuração surgiu uma única vez.

(II.II)

A variação composta por três covariantes, por seu turno, foi verificada na forma de duas configurações: [ɻ]-[ɾ]-[r] (com duas instâncias) e [ɻ]-[ɾ]-[Ø] (com 11). A primeira ocorreu em duas localidades afastadas entre si, Sorocaba (P5) e Piracicaba (P10), e em dois grupos, CbWf e CaWm, respectivamente, os quais não guardam em comum qualquer característica extralinguística no momento em foco, quais sejam: diastrática e diagenérica. A segunda configuração, [ɻ]-[ɾ]-[Ø], teve cinco das suas 11 instâncias ocorridas no grupo CaWf de cinco localidades: São Roque (P4), Itu (P6), Porto Feliz (P7), Tietê (P8) e Piracicaba (P10); logo, de maneira mais consistente na parte alta (em relação ao rio Tietê) da porção da região do Médio Tietê sob enfoque.

No que concerne à variação envolvendo quatro covariantes, ela tomou lugar somente em Araçariguama (P3), uma única vez, no grupo CbWf.

(III.II)

Do gráfico estatístico, infere-se a existência de um padrão entre ambos os grupos da classe escolar alta. Nomeadamente, nos grupos CaWf e CaWm, os valores percentuais de todas as covariantes mostraram-se bastante próximos uns dos outros. Entre as três covariantes com dados mais significativos, [ɻ] obteve frequência relativa de 52% no primeiro grupo e de 50% no segundo. O fone [ɾ], por sua vez, apareceu 20% das vezes no primeiro grupo, CaWf, e 18% no segundo, CaWm. E [Ø] foi o caso em 28% no primeiro grupo e 30% no segundo.

Outro fato quantitativo que chama a atenção é a discrepância do grupo CbWf em relação aos demais. Enquanto nos outros a covariante [ɻ] é a que se manifestou a mais frequente, no grupo CbWf esse lugar foi ocupado pelo concorrente [Ø], com 50% de aparecimento contra 38% de [ɻ].

Também a covariante [Ø], em ambos os grupos da classe escolar baixa (CbWf e CbWm), logrou um valor significativamente mais alto do que ela o fez nos grupos de classe escolar alta (CaWf e CaWm). E o contrário se diz acerca de [ɾ]: valores, em comparação, mais altos na classe escolar alta (CaWf: 20%; e CaWm: 18%) e baixos na classe escolar baixa (CbWf: 7%; e CbWm: 3%).

Para analisarmos um pouco mais os resultados advindos da pesquisa sobre o /R/ em coda externa de desinência infinitiva seguida por pausa, tomemos ainda em consideração a Figura 7 na próxima página, que contém uma carta fonético-fonológica a respeito da série WzGy.

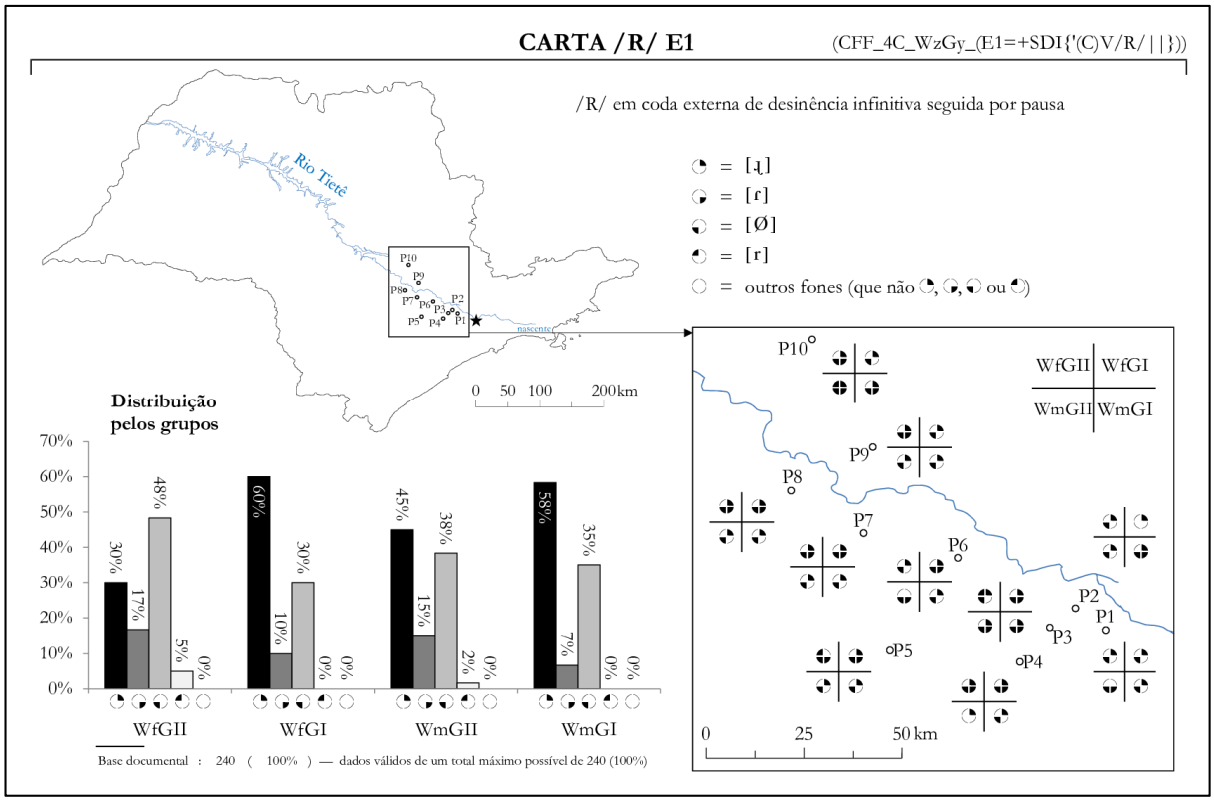


Figura 7: Carta /R/ da série WzGy e especificidade 1 (carta miniaturizada)

(I.III)

Conforme se pode notar na Figura 7, não houve concorrência entre variantes nem no grupo WfGI de Pirapora do Bom Jesus (P2), nem no grupo WmGII de São Roque (P4). Em ambos, que não compartilham propriedades extralinguísticas entre si (ref. àquelas controladas em nossa pesquisa sob a condição de variáveis), trata-se apenas da realização do fone [ɻ].

Relativamente à variação constituída por duas covariantes, houve duas configurações: [ɾ]-[Ø] e [ɻ]-[Ø]. A primeira contou com duas instâncias: uma em Santana de Parnaíba (P1) e outra em Itu (P6); ambas observadas no grupo WmGII. A segunda configuração obteve 18 instâncias, cobrindo toda a rede de pontos, exceto Araçariguama (P3), e figurou em todos os quadrantes possíveis. Também vale destacar que os grupos masculinos somados (WmGII e WmGI) produziram 12 das 18 instâncias, ao longo de toda a rede de pontos, afora Araçariguama (P3) e Piracicaba (P10).

(II.III)

No que diz respeito à variação entre três covariantes, houve três configurações: [ɻ]-[Ø]-[r] (com uma instância), [ɾ]-[Ø]-[r] (também com uma) e [ɻ]-[ɾ]-[Ø] (com 15). As duas primeiras ocorreram em Araçariguama (P3) e em Sorocaba (P5), respectivamente, e em ambos os casos via grupo WfGII. Já a configuração [ɻ]-[ɾ]-[Ø] apareceu em toda a rede de pontos, tirante Santana de Parnaíba (P1), sendo ambos os grupos femininos (WfGII e WfGI) os autores de 11 das 15 instâncias, ao longo de todas as localidades, salvo — além do P1 neste caso — Pirapora do Bom Jesus (P2).

Ainda, em relação à variação constituída por quatro covariantes, ela só figurou uma vez, em Piracicaba (P10), no grupo WmGI.

(III.III)

Por fim, tomando-se o gráfico da Figura 7 em consideração, um padrão revela-se entre os grupos WfGI, WmGII e WmGI: a covariante [ɻ] foi a que mais ocorreu (com 60%, 45% e 58% de frequência relativa por grupo, respectivamente), seguida por [Ø] (com 30%, 38% e 35%) e depois por [ɾ] (com 10%, 15% e 7%). Além desse padrão, nota-se também uma discrepância entre tais grupos e o grupo WfGII. Neste último, a covariante mais produtiva não foi [ɻ] (30%), mas sim [Ø] (48%), o que, portanto, sugere que as mulheres mais velhas são os indivíduos que mais realizam o apagamento como contraparte concreta do /R/ em coda externa de desinência infinitiva seguida por pausa.

6.1.2 Em sílaba tônica (a qual não seja, porém, de DI) seguida por pausa

Para seguir agora com inferências sobre os resultados relacionados com o /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa, partiremos de cartas fonético-fonológicas que elaboramos acerca do assunto, com abordagem inicial à série CxGy, para depois voltarmo-nos às séries CxWz e WzGy.

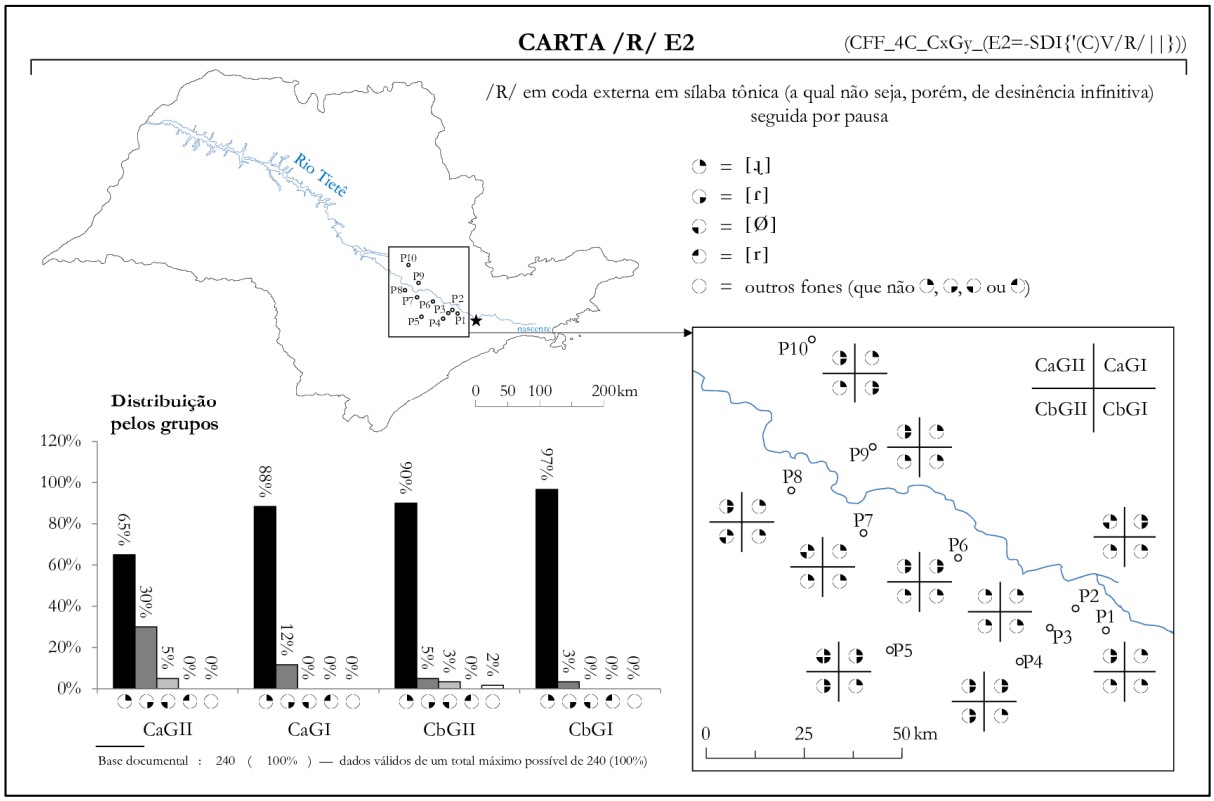


Figura 8: Carta /R/ da série CxGy e especificidade 2 (carta miniaturizada)

(I.I)

Em toda a rede de pontos, não houve variação em pelo menos um dos quatro grupos de informantes por localidade; ausência de variação essa que foi representada pelo surgimento exclusivo da configuração [ɻ]. Foram 23 instâncias suas, especialmente nos grupos CaGI (responsáveis por seis instâncias), CbGI (nove) e CbGII (sete). Já no grupo CaGII, essa não variação aconteceu somente em Araçariguama (P3). Outro destaque relativo ao P3 reside no fato de que ele foi o único no qual a não variação em foco foi observada em todos os quatro grupos de entrevistados. Ainda, em Pirapora do Bom Jesus (P2) e em Itu (P6), apenas os indivíduos de ambos os respectivos grupos da classe escolar baixa (CbGII e CbGI) realizaram exclusivamente [ɻ]. Ademais, Tietê (P8) é a única localidade na qual a configuração [ɻ] foi observada em ambos os grupos de falantes jovens (CaGI e CbGI) de maneira exclusiva.

A variação entre duas covariantes deu-se mediante duas configurações: [ɻ]-[Ø] (com três instâncias) e [ɻ]-[ɾ] (com 13). A primeira ocorreu em Pirapora do Bom Jesus (P2), Porto Feliz (P7) e Tietê (P8), apenas entre os falantes mais velhos, com prevalência da classe escolar alta (grupo CaGII no P2 e no P7). A segunda configuração, [ɻ]-[ɾ],

predominou nos grupos de classe escolar alta, com 10 das 13 instâncias totais: seis no grupo CaGII (de maneira diatopicamente dispersa pela rede de pontos) e quatro no grupo CaGI (com concentração de incidência na parte baixa — em relação ao rio Tietê — da porção da região do Médio Tietê em foco).

(II.I)

Quanto à variação entre três covariantes, houve apenas uma instância em toda a rede de pontos, sob a configuração [ɻ]-[ɾ]-[Ø], em Sorocaba (P5), grupo CaGII. Ademais, não houve variação composta por quatro covariantes.

(III.I)

Quantitativamente, é clara a prevalência da covariante [ɻ] em todos os grupos, sobretudo no grupo CbGI, no qual o fone em tela obteve 97% de frequência relativa. O segundo grupo que mais o produziu é o outro representante da classe escolar baixa, CbGII, responsável pela enunciação de [ɻ] com 90% de frequência relativa.

Em relação a ambos os grupos da classe escolar alta, eles são os únicos a terem produzido a covariante [ɾ] de maneira significativa: 30% em CaGII e 12% em CaGI (contra 5% em CbGII e 3% em CbGI). No tocante às demais covariantes, elas não obtiveram valores significativos.

Com os principais destaques sobre o objeto em apreciação — a variação fonética atinente ao /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa — com respeito à série CxGy tendo sido mencionados nas linhas anteriores, chega o turno da série CxWz, com base na carta linguística que a Figura 9 na próxima página traz.

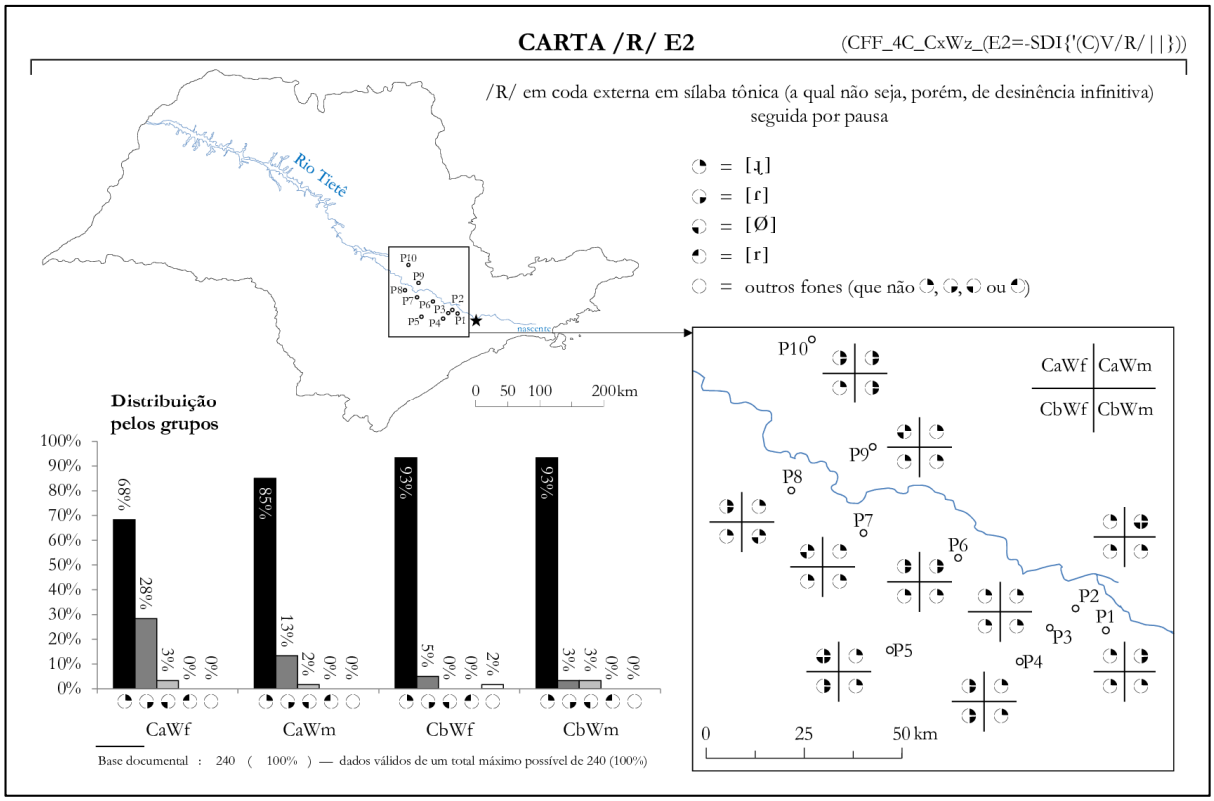


Figura 9: Carta /R/ da série CxWz e especificidade 2 (carta miniaturizada)

(I.II)

Iniciando-se com a abordagem sobre ausência de variação, apenas houve uma configuração, [ɻ], presente em toda a rede de pontos, com 25 instâncias; sendo 16 delas de responsabilidade dos informantes da classe escolar baixa (CbWm e CbWf), de modo diatopicamente difuso pela região do Médio Tietê. Especificamente em Araçariguama (P3), a ausência de variação representada pela ocorrência unicamente da variante [ɻ] foi o caso de todos os quatro grupos de indivíduos.

Relativamente à variação composta por duas covariantes, houve duas configurações: [ɻ]-[Ø] (com três instâncias) e [ɻ]-[ɾ] (com dez). A primeira predominou no grupo CaWf, em Porto Feliz (P7) e Capivari (P9). A segunda configuração, [ɻ]-[ɾ], encontrou-se em toda a rede de pontos, salvo em Pirapora do Bom Jesus (P2), Araçariguama (P3), Porto Feliz (P7) e Capivari (P9), e majoritariamente na classe escolar alta (grupos CaWf e CaWm), a qual respondeu por sete das dez instâncias.

(II.II)

A variação entre três covariantes só foi constatada em Pirapora do Bom Jesus (P2) e em Sorocaba (P5), sob a configuração única [ɻ]-[ɾ]-[Ø], que se manifestou mediante duas instâncias, uma em cada localidade, ambas atualizadas por membros da classe escolar alta (CaWm no P2; CaWf no P5). Já a respeito da variação envolvendo quatro covariantes, ela não se verificou em nenhuma localidade.

(III.II)

Em relação aos resultados do ponto de vista estatístico, a produção da covariante [ɻ] é, de longe, a mais alta, em todos os quatro grupos de entrevistados. Essa superioridade é bastante elevada no que concerne aos grupos pertencentes à classe escolar baixa (CbWf e CbWf), em que o percentual de ocorrência de [ɻ] é de 93%.

Uma outra covariante que atingiu resultados significativos é [ɾ], mas apenas entre pessoas com nível escolar alto. Nomeadamente, nos grupos CaWf e CaWm, o fone [ɾ] surgiu com 28% e com 13% de frequência relativa, respectivamente.

Feitas essas considerações acerca dos fones (e do apagamento) do /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa com referência à série CxWz, examinemos os resultados da variável linguística em pauta relativamente à série WzGy, a partir da carta fonético-fonológica apresentada pela Figura 10 na página a seguir.

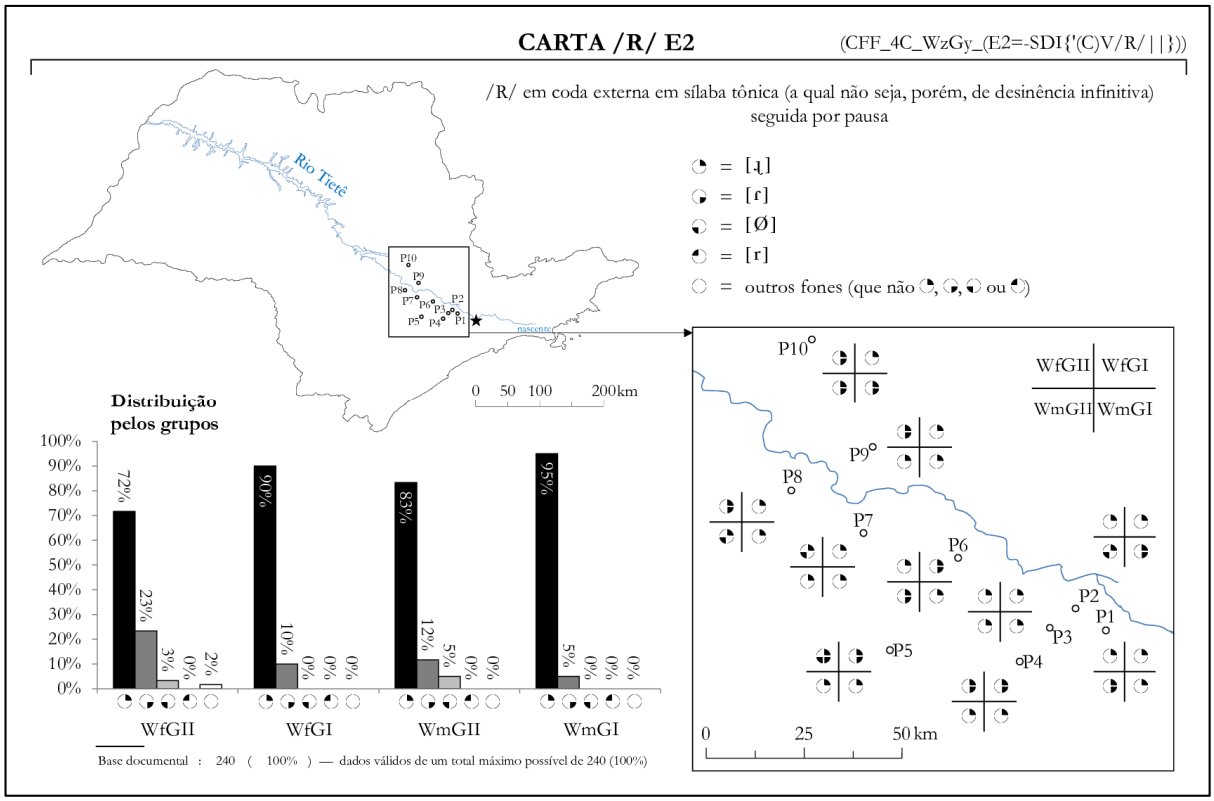


Figura 10: Carta /R/ da série WzGy e especificidade 2 (carta miniaturizada)

(I.III)

Como se vê na cartografia acima, toda a rede de pontos albergou, em pelo menos um de seus quadrantes por localidade, a variação nula, representada no caso exclusivamente pela ocorrência do fone [ɻ], cujas instâncias somam 24. Em Araçariguama (P3) em específico, todos os seus quadrantes apresentam apenas [ɻ]. Ainda, considerando-se toda a rede, os indivíduos mais novos (dos grupos WfGI e WmGI) são responsáveis por 15 das 24 instâncias registradas de [ɻ], com distribuição geográfica difusa.

No tocante à variação constituída por duas covariantes, houve duas configurações: [ɻ]-[Ø] (com três instâncias) e [ɻ]-[ɾ] (com 12). Todas as instâncias da primeira configuração são de autoria dos falantes mais velhos, predominantemente masculinos de Pirapora do Bom Jesus (P2) e Tietê (P8). A segunda configuração, [ɻ]-[ɾ], emergiu em toda a rede de pontos, exceto em Araçariguama (P3) e Porto Feliz (P7), com suas 12 instâncias distribuídas praticamente de maneira homogênea pelos quatro grupos de inquiridos. No entanto, a configuração em apreço, quando produzida pelo grupo WmGI, apareceu apenas ao lado direito do rio Tietê, bem como, quando produzida pelo grupo WfGI, ela surgiu somente ao lado esquerdo desse rio.

(II.III)

Quanto à variação entre três covariantes, houve uma única instância, sob a configuração [ɻ]-[ɾ]-[Ø], no grupo WfGII em Sorocaba (P5). Já em relação à variação constituída por quatro covariantes, não houve qualquer registro.

(III.III)

No tocante à quantificação dos dados distribuída pelos quatro grupos de falantes, nota-se uma vitalidade bastante significativa de [ɻ] na situação estrutural em foco, *i.e.*, em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa. Essa vitalidade foi especialmente grande entre os indivíduos, curiosamente, mais jovens. Foram 90% de frequência relativa de [ɻ] no grupo WfGI e 95% no grupo WmGI (contra 72% no grupo WfGII e 83% no grupo WmGII).

As outras covariantes não obtiveram resultados significativos, salvo [ɾ], com ocorrência sutilmente maior entre as pessoas mais velhas (WfGII: 23%; e WmGII: 12%) em contraste com as mais jovens (WfGI: 10%; e WmGI: 5%).

6.2 Interna

6.2.1 Em sílaba tônica

Agora, vamos considerar os róticos (bem como sua ausência) em coda interna em sílaba tônica. Assim como fizemos anteriormente acerca do arquifonema /R/ em um distinto ambiente fonológico, tomaremos a seguir três cartas linguísticas que elaboramos para iniciarmos as depreensões de natureza preeminentemente descritiva acerca do tema. Essas cartas relacionam-se às séries CxGy (com a qual vamos iniciar), CxWz e WzGy.

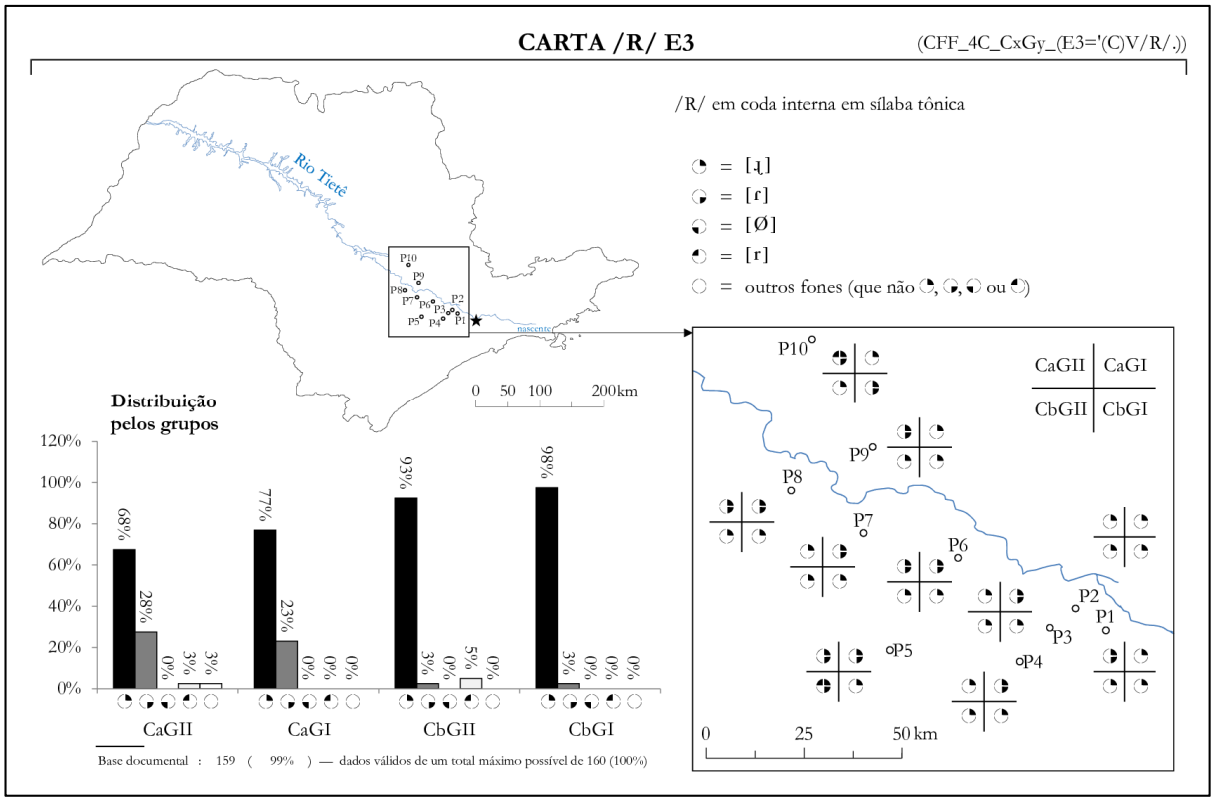


Figura 11: Carta /R/ da série CxGy e especificidade 3 (carta miniaturizada)

(I.I)

Tal como é possível observar na cartografia trazida pela Figura 11, todos os quadrantes juntos somam 40, e em mais da metade, 26, não houve variação, ao longo de toda a rede de pontos. Essa ausência de concorrência variacional é representada pela presença apenas da variante [ɻ], e a distribuição de suas instâncias foi difusa diatopicamente e simétrica entre, de um lado, os grupos da classe escolar alta (Ca) e, de outro, entre os grupos da classe escolar baixa (Cb): foram quatro instâncias em cada grupo da Ca (CaGII e CaGI) e nove instâncias em cada grupo da Cb (CbGII e CbGI). Assim, a Ca é responsável por ter gerado oito instâncias no total, e a Cb é autora de 18.

Em relação à variação constituída por duas covariantes, houve o registro somente da configuração [ɻ]-[ɾ] (com 12 instâncias), ao longo de toda a rede de pontos, afora Pirapora do Bom Jesus (P2). No entanto, alguns padrões emergiram. Um deles reside no fato de que quase a totalidade das instâncias (11 delas) teve lugar na classe escolar alta. Outro aspecto visível na cartografia mostra que a ocorrência da configuração [ɻ]-[ɾ] entre falantes do grupo CaGI concentrou-se exclusivamente ao lado esquerdo do rio Tietê, em todas as localidades do referido lado, exceto em Santana de Parnaíba (P1). Ainda, um último

destaque a ser feito tem a ver com a constatação de que o grupo CbGII, ao longo de absolutamente toda a rede de pontos, é o único que não produziu qualquer instância da configuração em tela, [ɻ]-[ɾ].

(II.I)

Relativamente à variação entre três covariantes, ela apresentou-se sob a configuração [ɻ]-[ɾ]-[r], com duas instâncias, uma em Sorocaba (P5) e outra em Piracicaba (P10), ambas produzidas por informantes mais velhos (pelo grupo CbGII no P5 e pelo grupo CaGII no P10). Com respeito à variação entre quatro covariantes, ela não foi observada.

(III.I)

Considerando-se o gráfico na Figura 11, e desprezados os valores percentuais iguais ou abaixo de 5% obtidos por outras covariantes, notam-se frequências relativas bastante altas dos fones [ɾ] e, sobretudo, [ɻ].

O fone [ɾ] foi atualizado de maneira significativa por membros da classe escolar alta, especificamente com 28% de ocorrência no grupo CaGII e com 23% no grupo CaGI (contra 3% de cada grupo da classe escolar baixa, CbGII e CbGI).

A covariante [ɻ], por seu turno, mostrou-se estatisticamente significativa em todos os grupos, especialmente naqueles ligados à classe escolar baixa: foram 93% de frequência relativa no grupo CbGII e 98% no grupo CbGI. Esses valores são sensivelmente maiores do que aqueles dos grupos da classe escolar alta — 68% de ocorrência de [ɻ] em CaGII e 77% em CaGI —, os quais também já são suasórios.

Vejamos agora como os resultados fonéticos e de apagamento do /R/ em coda interna em sílaba tônica situam-se na série CxWz, a partir da carta linguística que a Figura 12 na página adiante expressa.

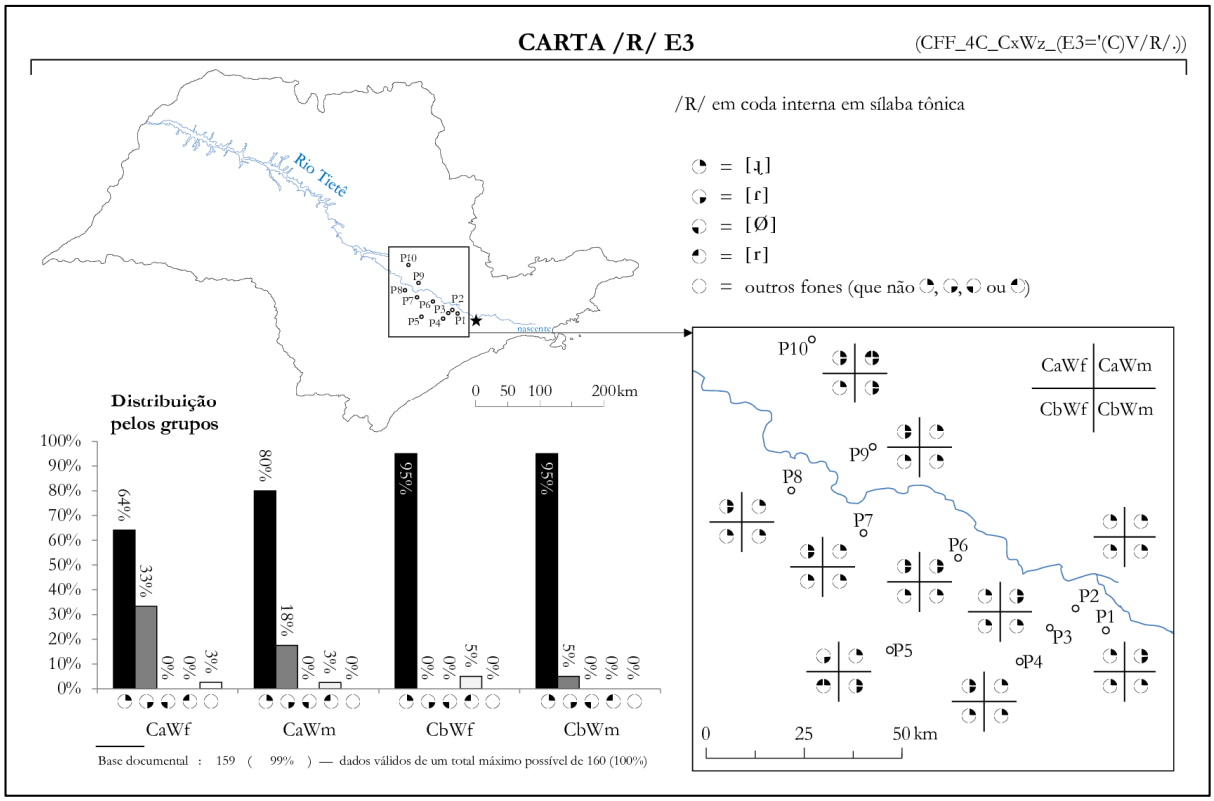


Figura 12: Carta /R/ da série CxWz e especificidade 3 (carta miniaturizada)

(I.II)

Abordando-se inicialmente os casos de ausência de variação por grupo de inquiridos demonstrados na carta fonético-fonológica disponibilizada na Figura 12, houve duas configurações variacionais: [ɾ] (com apenas uma instância) e [ɻ] (com 26). A primeira foi observada em Sorocaba (P5), no quadrante superior esquerdo, correspondente ao grupo CaWf. A segunda configuração, [ɻ], surgiu em toda a rede de pontos e figurou pelo menos uma vez em cada um dos quadrantes por localidade. Em Pirapora do Bom Jesus (P2), [ɻ] foi a regra em todos os grupos de entrevistados. Outro destaque fica por conta do fato de que a classe escolar baixa (grupos CbWf e CbWm) é responsável pela maioria das instâncias de [ɻ]: 17 (contra nove da classe escolar alta, grupos CaWf e CaWm). Ainda, chama a atenção que as ocorrências de responsabilidade do grupo CaWf concentraram-se na parte baixa (em relação ao rio Tietê) da porção da região do Médio Tietê em consideração. Relativamente aos demais grupos, suas produções foram geograficamente dispersas pela região.

Em se tratando da variação formada pela concorrência de duas covariantes, registramos duas configurações: [ɻ]-[r] (com apenas uma instância) e [ɻ]-[ɾ] (com 11). A primeira manifestou-se em Sorocaba (P5), no grupo CbWf. A segunda, [ɻ]-[ɾ], ocorreu por toda a rede de pontos, salvo em Pirapora do Bom Jesus (P2). Das 11 instâncias, nove foram geradas pela classe escolar alta, especialmente pela ala feminina, que realizou seis instâncias, contra três dos homens. Ainda, em nenhuma localidade, a configuração em foco apareceu no grupo CbWf. E um último realce refere-se ao fato de que a contribuição de três instâncias de [ɻ]-[ɾ] por parte do grupo CaWm concentrou-se somente ao lado esquerdo do rio Tietê e, adicionalmente, apenas na parte baixa (em relação ao rio mencionado) da porção da região do Médio Tietê sob enfoque.

(II.II)
No tocante à variação entre três covariantes, houve apenas uma instância, sob a configuração [ɻ]-[ɾ]-[r], em Piracicaba (P10), no grupo CaWm. Quanto à variação consistente de quatro covariantes, não houve aparecimento.

(III.II)
Voltando-se a atenção por um momento ao gráfico estatístico integrado à carta linguística que a Figura 12 exibe, percebe-se que [ɻ] como contraparte fonética possível de /R/ em coda interna em sílaba tônica é a mais corrente, e isso entre todos os quatro grupos de entrevistados, sobretudo entre ambos os grupos da classe escolar baixa, CbWf e CbWm: enquanto cada um destes obteve 95% de frequência do fone [ɻ], os grupos CaWf e CaWm obtiveram 64% e 80%, respectivamente, o que também não deixa de representar valores percentuais altos. Por outro lado, são os grupos da classe escolar alta aqueles que produziram a covariante [ɾ] de maneira significativa, especialmente entre as mulheres: foram 33% em CaWf e 18% em CaWm (contra insignificativos 0% em CbWf e 5% em CbWm).

A seguir, será a vez de se verificarem as relações entre, de um lado, os resultados concretos da abstração /R/ no ambiente fonológico em tela (*i.e.*, em coda interna em sílaba tônica) e, de outro lado, as dimensões (ou variáveis) extralinguísticas que a série WzGy simboliza.

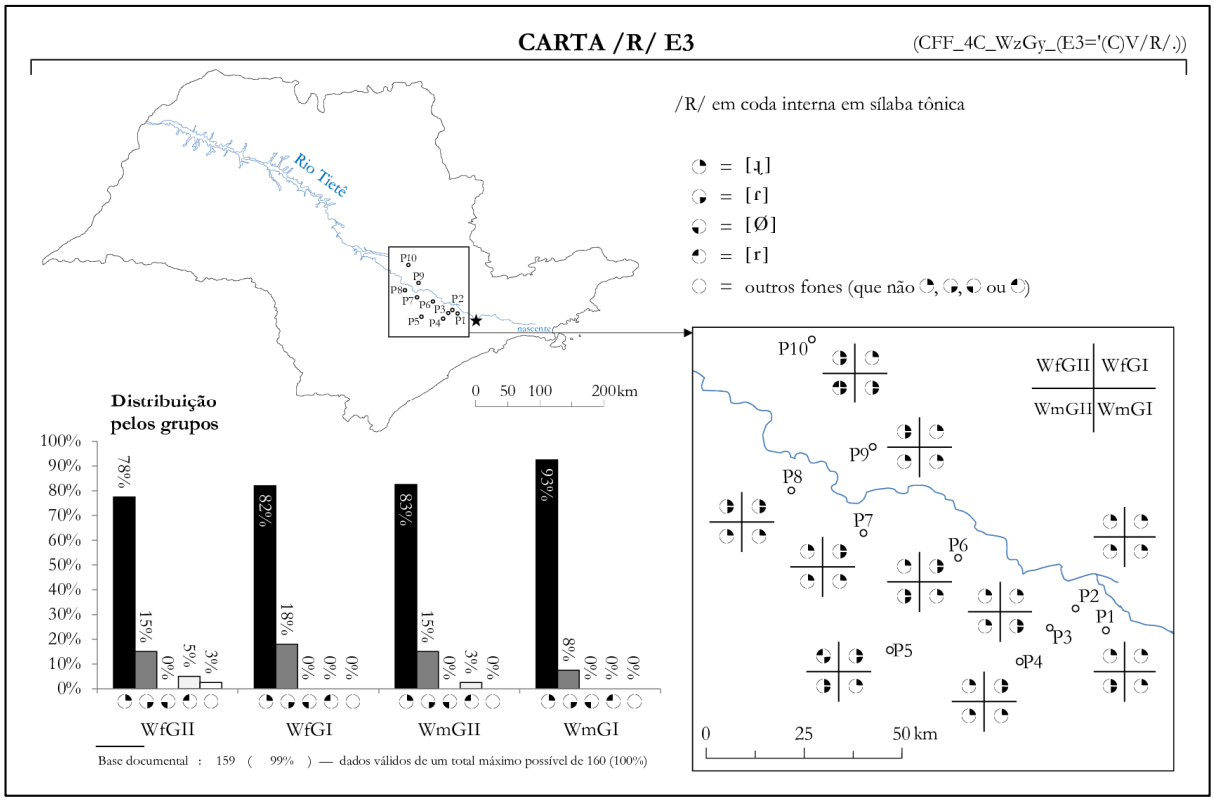


Figura 13: Carta /R/ da série WzGy e especificidade 3 (carta miniaturizada)

(I.III)

Conforme a cartografia mostra, o fone [ɻ] sozinho, implicando ausência de variação, e sendo aqui a única configuração que implique ausência de variação, contou com 25 instâncias, distribuídas difusamente ao longo de toda a rede de pontos. Pirapora do Bom Jesus (P2) em específico apresentou essa configuração em todos os seus quadrantes, e o grupo WmGI, considerando-se toda a rede, é responsável pela maioria das instâncias: oito (contra seis de WfGII, cinco de WfGI e seis de WmGII).

Relativamente à variação constituída por duas covariantes, tomaram lugar as configurações [ɾ]-[r] (com uma instância) e [ɻ]-[ɾ] (com 13). A primeira emergiu apenas em Sorocaba (P5), no grupo WfGII. A segunda configuração, [ɻ]-[ɾ], foi registrada ao longo de toda a rede de pontos, tirante Pirapora do Bom Jesus (P2). De todas as suas instâncias, o grupo WfGI contribuiu com a maioria, nomeadamente com cinco (contra três de WfGII, duas de WmGI e três de WmGII), as quais se concentraram unicamente ao lado esquerdo do rio Tietê. Ainda, as instâncias de responsabilidade do grupo WfGII localizaram-se exclusivamente na parte alta (em relação ao rio mencionado) da porção da

região do Médio Tietê em apreço, enquanto as instâncias de autoria do grupo WmGII apareceram apenas na parte baixa.

(II.III)
No tocante à variação entre três covariantes, houve apenas uma instância em toda a rede de pontos, sob a configuração [ɻ]-[ɾ]-[r], em Piracicaba (P10), no grupo WmGII. Em relação à variação formada por quatro covariantes, não houve qualquer caso.

(III.III)
Com os resultados arranjados no gráfico trazido pela Figura 13, observa-se mais uma vez o predomínio da covariante [ɻ] sobre as demais, em todos os quatro grupos de falantes, especialmente no grupo WmGI, o qual respondeu por 93% de frequência do fone em questão, contra 78% em WfGII, 82% em WfGI e 83% em WmGII.

Uma segunda covariante com resultados significativos é [ɾ], com valores percentuais semelhantes nos grupos WfGII, WfGI e WmGII (15%, 18% e 15%, respectivamente), por um lado, e, por outro, com valor particularmente diferente e mais baixo no grupo WmGI (8%).

6.2.2 Em sílaba átona

Feitas as considerações sobre as correlações entre, de um lado, a variável linguística relativa às contrapartes (a)fonéticas do arquifonema /R/ em coda interna em sílaba tônica no vernáculo da região paulista do Médio Tietê e, de outro lado, as variáveis extralinguísticas (diastrática, diagenérica e diageracional situadas no plano diatópico), vamos tratar a seguir dos resultados relativos ao /R/ agora em coda interna em sílaba átona, a começar com a série CxGy, a partir da carta fonético-fonológica que a Figura 14 na página subsequente apresenta.

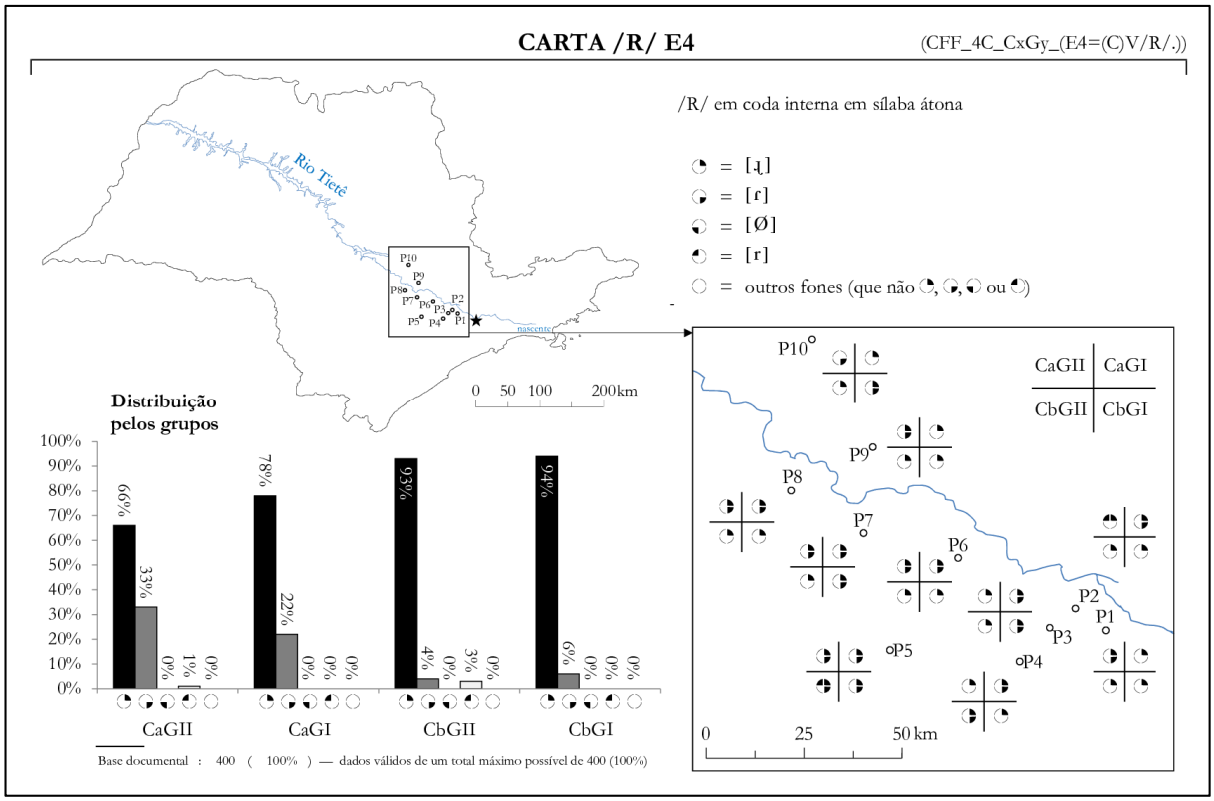


Figura 14: Carta /R/ da série CxGy e especificidade 4 (carta miniaturizada)

(I.I)

Os casos sem variação, sem concorrência entre covariantes, resumiram-se a duas configurações: [ɾ] (com apenas uma instância) e [ɻ] (com 19). A primeira surgiu somente em Piracicaba (P10), no grupo CaGII. A segunda esteve presente em toda a rede de pontos, exceto em Sorocaba (P5), figurando pelo menos duas vezes em cada um dos quadrantes possíveis. Das 19 instâncias totais, 14 foram de responsabilidade da classe escolar baixa (grupos CbGII e CbGI). Outro fato que se revelou é que as instâncias produzidas pelo grupo CaGII concentraram-se ao lado esquerdo do rio Tietê na parte baixa (em relação a esse mesmo rio) da porção da região do Médio Tietê sob enfoque. As demais instâncias de [ɻ] distribuíram-se difusamente pela rede de pontos.

Considerando-se a variação composta por dois fones concorrentes, houve duas configurações variacionais: [ɻ]-[r] (com somente uma instância) e [ɻ]-[ɾ] (com 18). A primeira ocorreu em Pirapora do Bom Jesus (P2), no grupo CaGII. A segunda, [ɻ]-[ɾ], emergiu ao longo de toda a rede de pontos, especialmente por contribuição da classe escolar alta (grupos CaGII e CaGI), autora de 13 das 18 instâncias totais. Outra constatação refere-

se ao fato de que o grupo CbGII fez uma contribuição mínima no estabelecimento da configuração [ɻ]-[ɾ], respondendo por uma única instância, em São Roque (P4).

(II.I)

A respeito da variação entre três covariantes, apresentou-se apenas a configuração [ɻ]-[ɾ]-[ɾ], com uma única instância, em Sorocaba (P5), no grupo CbGII. Quanto à variação constituída por quatro covariantes, não houve registro.

(III.I)

Atendo-se aos resultados distribuídos pelos quatro grupos de informantes, da maneira como o gráfico presente na Figura 14 os exibe, [ɻ] e [ɾ] foram as únicas covariantes que obtiveram valores estatísticos significativos, em termos de frequência relativa.

A primeira, [ɻ], foi a mais produtiva entre ambas, sobretudo na classe escolar baixa (93% em CbGII e 94% em CbGI, contra 66% em CaGII e 78% em CaGI). A segunda, [ɾ], por outro lado, ocorreu mais entre indivíduos pertencentes à classe escolar alta (33% em CaGII e 22% em CaGI, contra 4% em CbGII e 6% em CbGI).

Ou seja, enquanto os valores percentuais de ambas as covariantes significativas em questão mostram-se semelhantes entre pessoas velhas e novas da classe escolar mais baixa, os valores das mesmas covariantes revelam-se sensivelmente distintos entre as duas faixas etárias da classe escolar alta. No caso desta última classe, a diferença de preferência dos entrevistados mais velhos pelo fone [ɻ] ou pelo fone [ɾ] foi de 33% a favor do primeiro fone. Já a diferença de preferência dos inquiridos mais jovens, por seu turno, foi de 56% a favor também do primeiro fone, *i.e.*, [ɻ].

A seguir, os resultados fonéticos e de apagamento obtidos concernentes ao /R/ em coda interna em sílaba átona serão postos em relação com as variáveis extralinguísticas que a série CxWz representa, a partir da carta linguística exposta na Figura 15, na página seguinte.

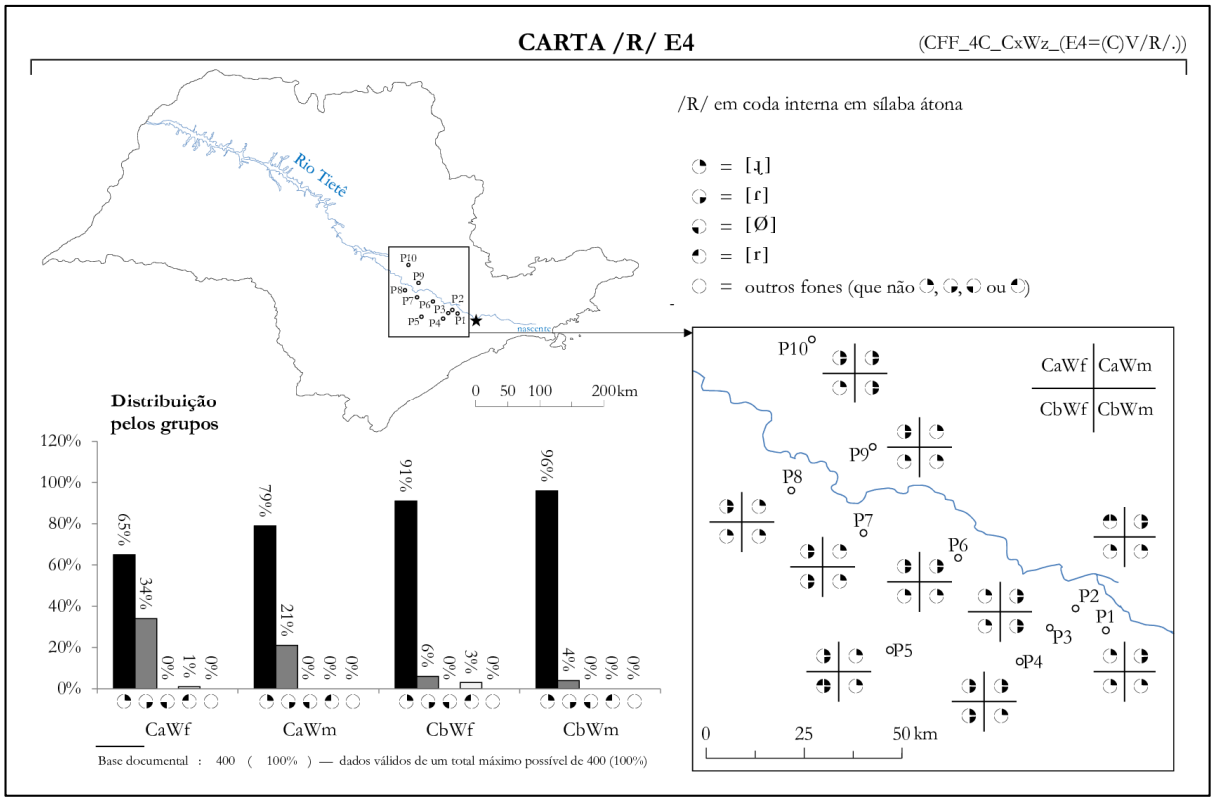


Figura 15: Carta /R/ da série CxWz e especificidade 4 (carta miniaturizada)

(I.II)

Em matéria de ausência de variação, houve unicamente a configuração [ɻ], ao longo de toda a rede de pontos. De suas 21 instâncias totais, 15 foram de autoria da classe escolar baixa (sete em CbWf e oito em CbWm, contra duas em CaWf e quatro em CaWm). Também, diga-se que as instâncias produzidas pelo grupo CaWf concentraram-se ao lado esquerdo do rio Tietê na parte baixa (em relação ao rio) da porção da região do Médio Tietê em tela.

Quanto à variação entre duas covariantes, surgiram duas configurações: [ɻ]-[r] (com uma única instância) e [ɻ]-[ɾ] (com 17). A primeira apareceu em Pirapora do Bom Jesus (P2), no grupo CaWf, e a segunda, ao longo da rede de pontos, figurando pelo menos duas vezes em cada um dos quadrantes possíveis. Das 17 instâncias da configuração [ɻ]-[ɾ], 13 foram geradas pela classe escolar alta (sete em CaWf e seis em CaWm, contra duas em CbWf e outras duas em CbWm). Ademais, as duas instâncias do grupo CbWf concentraram-se ao lado esquerdo do rio Tietê.

(II.II)

Em relação à variação formada por três covariantes, apareceu somente uma configuração, [ɻ]-[ɾ]-[r], com apenas uma instância, em Sorocaba (P5), no grupo CbWf. Já com atinência à variação composta por quatro covariantes, não houve surgimento.

(III.II)

Relativamente à distribuição dos resultados variacionais em termos percentuais pelos quatro grupos de falantes, fica claro na carta fonético-fonológica trazida pela Figura 15 que a frequência relativa do fone [ɻ] predominou sobre os demais em todos os grupos, principalmente naqueles pertencentes à classe escolar baixa: foram 91% em CbWf e 96% em CbWm (contra 65% em CaWf e 79% em CaWm).

A segunda covariante que foi mais frequente entre os quatro grupos de informantes é [ɾ], com resultados significativos nos grupos da classe escolar alta: foram 34% em CaWf e 21% em CaWm (contra 6% em CbWf e apenas 4% em CbWm). De resto, as demais covariantes não obtiveram valores superiores a cinco por cento.

Vejamos como os resultados concernentes às manifestações fonéticas e de apagamento do arquifonema /R/ em coda interna em sílaba átona com relação à série WzGy se deram, a partir da carta linguística oferecida pela Figura 16, na próxima página.

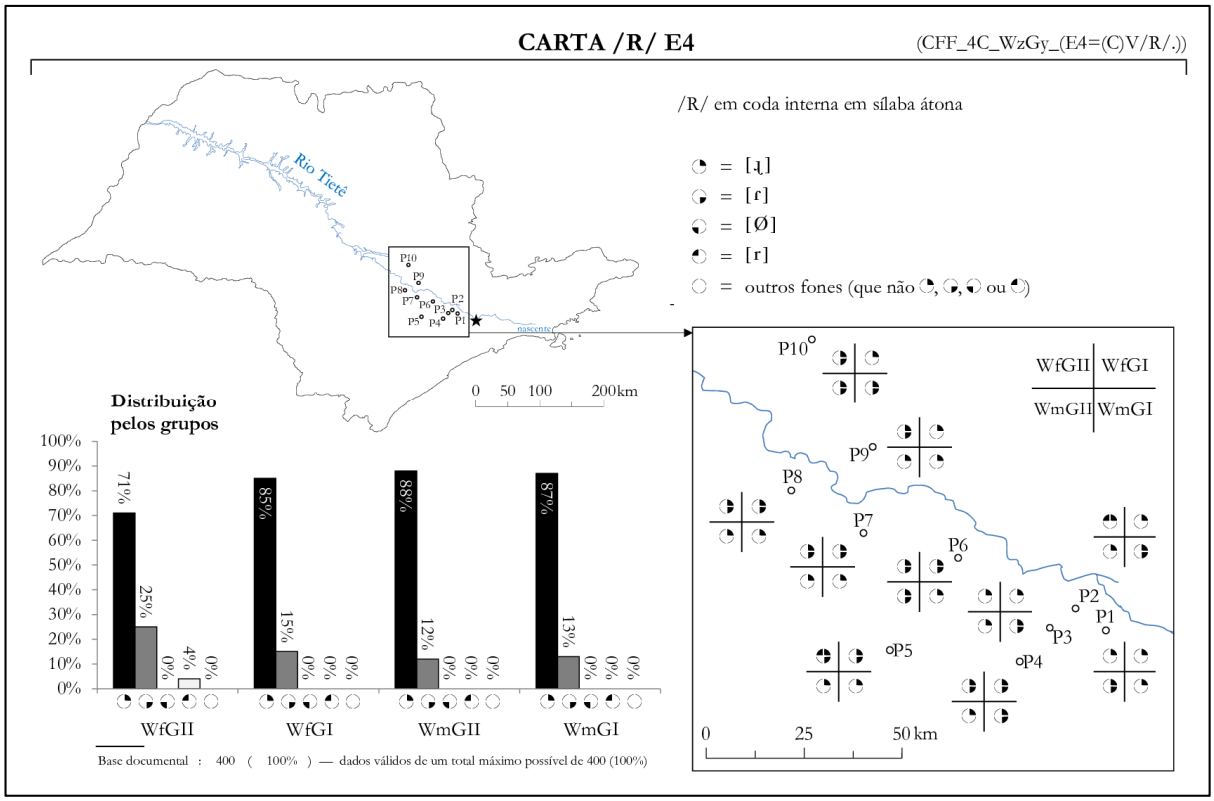


Figura 16: Carta /R/ da série WzGy e especificidade 4 (carta miniaturizada)

(I.III)

Ao longo de toda a rede de pontos, houve bastantes instâncias da configuração [ɻ], sendo a única a representar ausência de variação. Em cada um dos quadrantes possíveis, ela se mostrou pelo menos duas vezes. De suas 20 instâncias totais, 13 emergiram dos indivíduos masculinos (sete em WmGII e seis em WmGI, contra duas em WfGII e cinco em WfGI). A correlação 'produção de [ɻ]'-'perfis (ou grupos) de inquiridos produtores de [ɻ]' verificou-se difusamente pela rede de pontos, com exceção do grupo WfGII, cuja produção concentrou-se ao lado esquerdo do rio Tietê na porção baixa (em relação ao rio indicado) da região paulista do Médio Tietê ora focalizada.

A respeito da variação entre duas covariantes, assomaram-se duas configurações: [ɻ]-[r] (com somente uma instância) e [ɻ]-[ɾ] (com 18). A primeira foi registrada em Pirapora do Bom Jesus (P2), no grupo WfGII. A segunda, [ɻ]-[ɾ], foi observada ao longo de toda a rede de pontos, figurando pelo menos três vezes em cada quadrante possível. Entre ambos os gêneros de registro/social, o feminino (grupos WfGII e WfGI) é responsável pela maioria das instâncias: 11 (contra sete dos indivíduos masculinos, grupos

WmGII e WmGI). Também se nota que as instâncias concebidas pelo grupo WfGI concentraram-se ao lado esquerdo do rio Tietê.

(II.III)
Relativamente à variação composta por três covariantes, houve apenas a configuração [ɻ]-[ɾ]-[r], com uma instância, em Sorocaba (P5), no grupo WfGII. No tocante à variação entre quatro covariantes, ela não foi observada em campo.

(III.III)
Voltando-se o olhar à distribuição quantitativa dos resultados empíricos atinentes à variável '/R/ em coda interna em sílaba átona' entre os diferentes grupos de entrevistados, percebe-se que a força da covariante [ɻ] foi expressiva, com três grupos em especial obtendo valores mais semelhantes entre si: 85% em WfGI, 88% em WmGII e 87% em WmGI, enquanto foram 71% em WfGII.

Com referência à segunda e última covariante com valores significativos, *i.e.*, [ɾ], é no grupo WfGII que ela se apresentou mais frequente: foram 25% de ocorrência, contra 15% em WfGI, 12% em WmGII e 13% em WmGII. Com isso, vê-se ainda que os três grupos cujos valores concernentes à covariante [ɻ] assemelharam-se (grupos WfGI, WmGII e WmGI) também foram similares relativamente à produção da covariante [ɾ].

6.3 Por localidade

Alguns dos resultados sobre as diferentes contrapartes (a)fonéticas do arquifonema /R/ em coda, tais como foram demonstrados nas cartas linguísticas exibidas anteriormente, estiveram particularmente associados a uma ou outra localidade em específico. Na presente seção, o intuito é seletivamente expor destaques atinentes a tais associações particulares, levando-se em conta tanto as três séries (CxGy, CxWz, WzGy), quanto as quatro situações estruturais específicas do /R/ sob enfoque (E1 = em coda externa de desinência infinitiva seguida por pausa; E2 = em coda externa em sílaba tônica (que não de desinência infinitiva)

seguida por pausa; E3 = em coda interna em sílaba tônica; e E4 = em coda interna em sílaba átona).

Santana de Parnaíba (P1)

CxGy_E1: Entre os destaques relacionados ao P1 envolvendo a série CxGy acerca da E1, encontra-se o fato de que o grupo CaGI dessa localidade foi o único entre seus pares na rede de pontos a ter apresentado ausência de variação, fato representado neste caso pela presença unicamente do fone [ɻ]. Sobre a concorrência entre duas covariantes, o P1 é a única localidade a ter exibido a configuração [ɻ]-[ɾ] no grupo CaGII. Ademais, no tocante à variação constituída por três covariantes, o P1 é o único lugar de toda a rede de pontos no qual a coexistência variacional entre [ɻ], [ɾ] e [Ø] não foi observada, em nenhum de seus quadrantes, *i.e.*, em nenhum de seus quatro grupos representativos de falantes (CaGII, CaGI, CbGII e CbGI).

CxWz_E1: Uma ausência de variação representada pela configuração [ɻ] só foi registrada no grupo CaWf em duas localidades ao longo de toda a rede de pontos: no P1 e em Araçariguama (P3). Em se tratando de variação composta por duas covariantes, o P1 é a única localidade em que a configuração [ɻ]-[ɾ] surgiu no grupo CaWm.

WzGy_E1: Da perspectiva dos resultados de pesquisa referentes à série WzGy a respeito da especificidade 1 (*i.e.*, acerca dos fones e do apagamento do /R/ em coda externa de desinência infinitiva seguida por pausa), a configuração [ɾ]-[Ø] é uma de duas a ter expressado a variação entre duas covariantes. Tal configuração revelou-se mediante duas instâncias ao longo de toda a rede de pontos, uma no P1 e outra em Itu (P6), ambas no âmbito do grupo WmGII. Outro destaque envolvendo particularmente o P1 diz respeito ao fato de que ele é o único lugar, entre todos os demais da rede de pontos, onde a configuração variacional [ɻ]-[ɾ]-[Ø] não figurou.

CxGy_E3: Também foi observada uma particularidade envolvendo o P1 relativamente à série CxGy e à especificidade 3, *i.e.*, às contrapartes (a)fonéticas do /R/ em coda interna em sílaba tônica. Essa particularidade relaciona-se com a constatação segundo a qual as instâncias da configuração [ɻ]-[ɾ] entre indivíduos pertencentes ao grupo CaGI concentraram-se em todas as localidades posicionadas à esquerda do rio Tietê, com exclusão do P1.

Pirapora do Bom Jesus (P2)

CxGy_E1: No P2 — bem como em Santana de Parnaíba (P1), Araçariguama (P3), Sorocaba (P5), Itu (P6) e Tietê (P8), onde, no entanto, quadrantes diversos estiveram implicados —, uma de duas configurações variacionais representantes da variação nula, [ɻ], foi observada no grupo CbGI. Tal grupo é o principal responsável pelas instâncias de [ɻ], com colaboração — além do P2 — do P5 e do P8.

WzGy_E1: O P2 é uma de duas entre as dez localidades de nossa rede de pontos (a outra é São Roque, P4) que exibiu caso de variação nula, mediante o aparecimento da configuração [ɻ]. No P2, essa situação foi registrada no grupo WfGI. E sob o tema da variação constituída por três covariantes, a configuração variacional mais comum, [ɻ]-[ɾ]-[Ø] — com 15 instâncias, contra duas divididas entre as configurações restantes, [ɻ]-[Ø]-[r] e [ɾ]-[Ø]-[r] — apareceu em toda a rede de pontos, afora Santana de Parnaíba (P1), e 11 de suas 15 instâncias totais foram produzidas por mulheres (grupos WfGII e WfGI). Ademais, o P2 é uma de duas localidades (a outra é Santana de Parnaíba, P1) cujas mulheres não contribuíram com o engendramento da configuração [ɻ]-[ɾ]-[Ø].

CxGy_E2: No âmbito dos resultados ligados à série CxGy e à especificidade 2 — *i.e.*, aos sons (e ao apagamento) do /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa —, a configuração [ɻ], representando ausência de variação, foi exclusivamente enunciada por informantes da classe escolar baixa no P2 e em Itu (P6). No tocante à variação formada por duas covariantes, o P2 também merece destaque, o qual participou da produção de uma das configurações variacionais, [ɻ]-[Ø], só observada no P2, em Porto Feliz (P7) e em Tietê (P8), e apenas realizada por entrevistados da geração mais velha, com preponderância daqueles com escolaridade alta.

CxWz_E2: Acerca da variação entre duas covariantes, a configuração mais comum, [ɻ]-[ɾ] (com dez instâncias, contra três de uma segunda e última configuração, [ɻ]-[Ø]), emergiu ao longo de toda a rede de pontos, exceto no P2, em Araçariguama (P3), em Porto Feliz (P7) e em Capivari (P9). Porém, há outro destaque envolvendo o P2 que é sensivelmente mais particular, que se refere à variação composta por três covariantes. A esse respeito, houve a configuração [ɻ]-[ɾ]-[Ø], que só foi registrada no P2 e em Sorocaba (P5), a partir de falantes com nível escolar baixo.

WzGy_E2: Um destaque relativo à série WzGy e à E2 tem a ver com a variação entre duas covariantes. Das duas configurações registradas, [ɻ]-[Ø] e [ɻ]-[ɾ], a primeira foi objeto de enunciação exclusivamente por parte dos indivíduos mais velhos no P2 e em Tietê (P8).

CxGy_E3: Assim como no item imediatamente anterior, o relevo a ser explicitado aqui concerne à variação constituída por duas covariantes. Ao longo de toda a rede de pontos, houve apenas uma configuração, [ɻ]-[ɾ], expressa através de 12 instâncias, surgidas em todas as localidades, menos no P2.

CxWz_E3: De toda a rede de pontos, o P2 foi o único a ter manifestado a configuração [ɻ] em todos os seus quadrantes, *i.e.*, em todos os seus quatro grupos de inquiridos. Vinculada a isso, outra particularidade importante envolvendo o P2 refere-se ao fato de que essa localidade é a única, entre as demais, a não ter engendrado qualquer instância da configuração [ɻ]-[ɾ].

WzGy_E3: Como expressão de ausência de concorrência entre covariantes, a configuração [ɻ] manifestou-se abundantemente ao longo da rede de pontos, mas o P2 é o único município no qual ela figurou em todos os quadrantes, *i.e.*, em todos os quatro diferentes grupos de informantes (WfGII, WfGI, WmGII e WmGI). Associado a esse fato, há um segundo, que diz respeito à variação entre duas covariantes. Enquanto nove das dez cidades integrantes da nossa rede de pontos foram cenário para a realização da configuração [ɻ]-[ɾ], o P2 é o único ponto em que tal configuração não se apresentou.

CxGy_E4: Sob o tema da variação constituída por duas covariantes, o P2 ainda se destaca por ser a única localidade a ter sediado o surgimento da configuração [ɻ]-[r] (sob o ângulo, sublinhe-se, da E4 em conjunção com a série CxGy), com uma instância (contra a segunda e última configuração [ɻ]-[ɾ], que contou com 18 instâncias ao longo da rede de pontos).

CxWz_E4: Quanto à série CxWz e à especificidade 4 (*i.e.*, aos róticos e ao apagamento do /R/ em coda interna em sílaba átona), um realce a ser feito diz respeito à variação composta por duas covariantes: a configuração [ɻ]-[r] — que competiu com a mais frequente, [ɻ]-[ɾ] — obteve uma única instância, e o P2 é onde ela se manifestou.

WzGy_E4: Uma última peculiaridade envolvendo o P2, além de também implicar a E4 e a variação plasmada por duas covariantes, reporta-se agora à série WzGy. Considerando-se toda a rede de pontos, tomou-se nota de duas configurações, e uma delas, [ɻ]-[r], surgiu apenas no P2, enquanto a segunda, [ɻ]-[ɾ], foi registrada ao longo de toda a rede de pontos.

Araçariguama (P3)

CxGy_E1: Quando olhamos os resultados referentes às contrapartes (a)fonéticas do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa com respeito à série CxGy e ao P3, vemos, por exemplo, que houve casos nessa localidade — bem como em Santana de Parnaíba (P1), Pirapora do Bom Jesus (P2), Sorocaba (P5), Itu (P6) e Tietê (P8) — de não variação, fato representado neste caso pela emergência de uma única configuração: [ɻ]. No P3, isso foi observado no grupo CaGII. Além disso, o P3 é uma de duas localidades (a outra é Sorocaba, P5) de toda a rede de pontos a ter presenciado, no grupo CbGII, uma variação formada pela concorrência entre quatro covariantes: [ɻ], [ɾ], [Ø] e [r].

CxWz_E1: O P3, e junto com ele Santana de Parnaíba (P1) e Porto Feliz (P7), testemunhou ausência de variação em um de seus quadrantes; ausência essa representada neste caso pela configuração [ɻ]; e os grupos de informantes responsáveis por esse fato são da classe escolar alta, nomeadamente do grupo CaWf tanto no P1 quanto no P3 e, ainda, do grupo CaWm no P7. Também, o P3 é a única cidade de toda a rede de pontos onde uma de três configurações de variação entre duas covariantes, [ɻ]-[Ø] (as outras duas são [ɻ]-[ɾ] e [ɾ]-[Ø]), não se manifestou. Por outro lado, é o P3 o lugar em que ocorreu uma única instância de variação composta por quatro covariantes — por [ɻ], [ɾ], [Ø] e [r] —, registrada no grupo CbWf.

WzGy_E1: Ainda sobre a especificidade 1, mas agora com referência à série WzGy, um outro aspecto sobre o qual se pode lançar luz tange à variação composta por duas covariantes. O P3 é a única localidade, de toda a rede de pontos, em que uma das duas configurações variacionais observadas, [ɻ]-[Ø] (a outra é [ɾ]-[Ø]), não apareceu. Outro relevo a ser explicitado envolvendo o P3 diz respeito ao fato de que esse é o único ponto a ter assistido à atualização de uma de três configurações de variação entre três covariantes,

[ɻ]-[Ø]-[ɾ] (as outras duas são [ɾ]-[Ø]-[ɾ] e [ɻ]-[ɾ]-[Ø]), e isso se deu especificamente no grupo WfGII.

CxGy_E2: Em se tratando da série CxGy em conjunção com a especificidade 2 — *i.e.*, com as contrapartes (a)fonéticas do /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa —, uma peculiaridade envolvendo o P3 relaciona-se com os casos registrados de não variação; não variação essa que se expressou unicamente pela configuração [ɻ], com número de instâncias abundante. Ao longo de toda a rede de pontos, essa configuração foi observada, figurando em todos os quadrantes possíveis. No entanto, o grupo CaGII, diferentemente dos demais, a enunciou apenas uma vez, o que se deu no P3. Adicionalmente, o P3 é a única localidade a ter exibido a configuração em tela em todos os seus quadrantes, *i.e.*, em todos os seus quatro grupos de entrevistados.

CxWz_E2: Ao longo de toda a rede de pontos, surgiu ostensivamente uma só configuração a implicar ausência de variação, [ɻ], mas o P3 é a única localidade entre as demais a ter apresentado essa configuração em todos os seus grupos de falantes (CaWf, CaWm, CbWf e CbWm). Ainda, no tocante à variação plasmada pela atuação de duas covariantes, uma de duas configurações, [ɻ]-[ɾ] (a outra é [ɻ]-[Ø]), atualizou-se ao longo de toda a rede de pontos, salvo no P3, em Pirapora do Bom Jesus (P2), Porto Feliz (P7) e Capivari (P9).

WzGy_E2: Um realce envolvendo o P3 em relação à série WzGy e à especificidade 2 diz respeito à variação inexistente. Nesse sentido, a configuração [ɻ] representa, sozinha, a variação inexistente e apresentou-se ao longo de toda a rede de pontos. Entretanto, nenhuma localidade exibiu essa configuração em todos os seus quatro distintos grupos de inquiridos como o P3, por seu turno, o fez. Complementarmente, quanto à variação fundada na competição entre duas covariantes, uma de duas configurações registradas, [ɻ]-[ɾ] (a segunda é [ɻ]-[Ø]), pululou ao longo de toda a rede de pontos, com exclusão do P3 (e também de Porto Feliz, P7).

São Roque (P4)

CxWz_E1: Acerca da série CxWz e da especificidade 1 (*i.e.*, dos sons e da ausência de som do /R/ em coda externa de desinência infinitiva seguida por pausa), o P4 envolveu-se no estabelecimento de um fato linguístico relativo à variação constituída por três

covariantes. Uma de duas configurações variacionais observadas, [ɻ]-[ɾ]-[Ø] (a outra é [ɻ]-[ɾ]-[r]), obteve cinco de suas 11 instâncias totais advindas do grupo CaWf, manifestadas em cinco locais da nossa rede de pontos, entre os quais o P4 (os outros quatro são: Itu, P6; Porto Feliz, P7; Tietê, P8; e Piracicaba, P10).

WzGy_E1: Pela perspectiva que correlaciona a série e a especificidade em tela (série WzGy e E1), revelou-se que o P4 compartilhara com Pirapora do Bom Jesus (P2) um fato concernente à ausência de variação, haja vista que ambos sediaram uma instância cada da configuração [ɻ]. No entanto, o P4 é a única localidade na qual isso se deu especificamente no grupo WmGII.

CxGy_E4: Com a atenção voltada à série CxGy e à especificidade 4 (*i.e.*, aos róticos e ao apagamento em coda interna em sílaba átona), um aspecto a ser realçado atinente ao P4 refere-se à variação composta por duas covariantes. A esse respeito, uma de duas configurações registradas, [ɻ]-[ɾ] (a outra é [ɻ]-[r]), embora ela tenha surgido em abundância ao longo de toda a rede de pontos, obteve uma única instância entre os entrevistados integrantes do grupo CbGII, e isso tomou lugar no P4.

Sorocaba (P5)

CxGy_E1: O P5 esteve implicado num fato envolvendo a variação nula através de uma de duas configurações observadas, [ɻ] (a outra é [Ø]), juntamente com Santana de Parnaíba (P1), Pirapora do Bom Jesus (P2), Araçariguama (P3) e Tietê (P8); sendo o P2, o P5 e o P8 aqueles cujo grupo CbGI foi o principal entre os demais na produção da configuração [ɻ]. Além disso, o P5 em parceria com Araçariguama (P3) foi palco para a variação estabelecida por quatro covariantes — [ɻ], [ɾ], [Ø] e [r] —, gerada exclusivamente pelo grupo de inquiridos CbGII.

CxWz_E1: O P5 também criou resultados sobre os quais luz pode ser lançada relativamente à variação entre duas covariantes. Nesse sentido, uma de três configurações variacionais registradas, [ɾ]-[Ø] (as outras duas são [ɻ]-[ɾ] e [ɻ]-[Ø]), foi observada somente no P5 e em seu ponto vizinho Itu (P6). Todavia, entre todos os membros do grupo CaWf, o P5 é a única localidade onde a configuração [ɾ]-[Ø] emergiu. Ainda, outro realce a se fazer tem a ver com a variação formada por três covariantes: o P5 é o único município

em que uma de duas configurações observadas, [ɻ]-[ɾ]-[r] (a outra é [ɻ]-[ɾ]-[Ø]), foi gerada pelo grupo CbWf, entre todos seus pares distribuídos pela rede de pontos.

WzGy_E1: Quanto à correlação entre a série WzGy e a especificidade 1 (*i.e.*, como já se sabe, as contrapartes (a)fonéticas do /R/ em coda externa de desinência infinitiva seguida por pausa) a respeito da variação composta por três covariantes, vale salientar que somente o grupo WfGII (ou seja, somente as mulheres da faixa etária mais velha) do P5 realizou a configuração [ɾ]-[Ø]-[r] na região.

CxGy_E2: Com atinência à série CxGy em conjunção com a especificidade 2 — *i.e.*, com as realizações fonéticas do /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa —, o P5 mostrou uma peculiaridade concernente à variação formada por três covariantes. A esse respeito, houve somente uma configuração, [ɻ]-[ɾ]-[Ø], e ela só se verificou no P5, através de seu grupo CaGII.

CxWz_E2: Ainda sobre a E2, mas agora pela perspectiva da série CxWz, um destaque envolvendo o P5 pode ser feito quanto à variação fundada em três covariantes em concorrência. A esse propósito, a única configuração que foi observada é [ɻ]-[ɾ]-[Ø], e a única instância de autoria do grupo CaWf deu-se no P5.

WzGy_E2: Em se tratando da série WzGy, o grupo WfGII do P5 é o único — tanto entre seus pares, quanto entre os demais perfis de informantes — a ter produzido uma única instância de uma só variação entre três covariantes observada em toda a rede de pontos, e essa instância adveio sob a configuração [ɻ]-[ɾ]-[Ø].

CxGy_E3: Ainda sob o tema da variação composta por três covariantes, mas agora com relação à série CxGy e à especificidade 3 (*i.e.*, aos róticos e ao apagamento em coda interna em sílaba tônica), o P5 é a única cidade da nossa rede de pontos a ter sido palco para o aparecimento da configuração [ɻ]-[ɾ]-[r], cuja singular instância veio a lume mediante o grupo CbGII.

CxWz_E3: No que concerne à série CxWz, um acontecimento envolvendo o P5 tange à variação nula: essa localidade é a única integrante da nossa rede de pontos a ter gerado uma de duas configurações registradas, [ɾ] (a outra é [ɻ]), e o grupo autor dessa realização é CaWf. Outro aspecto peculiar implicando o P5 tem a ver com a variação entre duas covariantes: o P5 é o único da rede a ter emitido uma de duas configurações variacionais, [ɻ]-[r] (a outra é [ɻ]-[ɾ]), e o fez por intermédio de seu grupo CbWf.

WzGy_E3: A propósito da série WzGy, uma menção ao P5 a se fazer relaciona-se à variação constituída por duas covariantes. Uma de duas configurações, [ɾ]-[r] (a outra é [ɻ]-[ɾ]), foi concebida exclusivamente pelo grupo das mulheres da faixa etária mais elevada (WfGII) do P5.

CxGy_E4: Vem à baila agora a especificidade 4 (*i.e.*, os fones e o apagamento do /R/ em coda interna em sílaba átona) em perspectiva com a série CxGy a propósito do tema da ausência de variação no P5, mas em sentido inverso. Ou seja, enquanto nove dos dez pontos da rede assistiram ao surgimento da configuração [ɻ], o P5 é a única localidade onde essa configuração não figurou. Além disso, outra particularidade relacionada ao P5 é sobre a variação composta por três covariantes: foi nele que a única configuração registrada, [ɻ]-[ɾ]-[r], emergiu, por contribuição do grupo CbGII.

CxWz_E4: Em se tratando da série CxWz (*i.e.*, da série que diz respeito à classe escolar e ao gênero de registro/social dos indivíduos) ao lado da especificidade 4, um fato com envolvimento do P5 tem a ver com a variação entre três covariantes. Isso porque o P5 é o único ponto em que a configuração [ɻ]-[ɾ]-[r] apareceu, e isso por conta do grupo CbWf.

WzGy_E4: A constatação apontada no parágrafo anterior se preserva quando a série em questão é WzGy (em vez de CxWz). Ou seja, a mudança da variável "série" na corrente análise (de CxWz para WzGy) não implica uma inferência essencialmente distinta daquela já enunciada no parágrafo anterior. Isso nada mais significa que o perfil exato responsável pela enunciação única da configuração [ɻ]-[ɾ]-[r] é feminino de mais idade e da classe escolar baixa.

Itu (P6)

CxGy_E1: A correlação entre a série CxGy e a especificidade 1 (*i.e.*, os sons e o apagamento do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa), em perspectiva com o P6, mostrou que essa localidade é a única, de toda a rede de pontos, a ter produzido a configuração [Ø] como expressão de variação nula, através do grupo representativo da classe escolar baixa e faixa etária velha (CbGII).

CxWz_E1: Ainda sobre a E1, mas agora pondo em jogo a série CxWz, um fato singulariza o P6 com referência à variação fundada em duas covariantes. Isso porque, entre todos seus pares, apenas o grupo CaWm do P6 gerou a configuração [ɾ]-[Ø]. Outro acontecimento a ser posto em apreço liga-se à variação constituída por três covariantes, especificamente à configuração [ɻ]-[ɾ]-[Ø]. Entre as 11 instâncias totais dessa configuração, cinco foram performadas exclusivamente por membros do grupo CaWf, sediados no P6 e também em São Roque (P4), Porto Feliz (P7), Tietê (P8) e Piracicaba (P10).

WzGy_E1: Em relação à série WzGy, um realce a ser feito envolvendo o P6 reporta-se à variação plasmada por duas covariantes. A esse respeito, o município em questão é um de dois (o outro é Santana de Parnaíba, P1) a ter observado o surgimento da configuração [ɾ]-[Ø] por meio do grupo de informantes WmGII.

CxGy_E2: Para se encerrarem os turnos das inferências focalizadas nas particularidades envolvendo o P6, resta um aspecto ligado à série CxGy e à especificidade 2, bem como à ausência de variação. No quadro dessa variação nula, a configuração [ɻ] foi pródiga ao longo de toda a rede de pontos, e ela ocupou ambos os quadrantes inferiores (ref. à classe escolar baixa) apenas do P6 e de Pirapora do Bom Jesus (P2).

Porto Feliz (P7)

CxGy_E1: Sob o ângulo da série CxGy e da E1, um aspecto particular relacionado ao P7 concerne à variação entre duas covariantes. O grupo CaGI da localidade em pauta é o único entre seus pares distribuídos pela rede de pontos em que a configuração [ɻ]-[ɾ] se manifestou.

CxWz_E1: Ainda acerca da especificidade 1, mas agora com relação à série CxWz (ou seja, à classe escolar e ao gênero de registro/social), há o que se dizer sobre a variação nula. Isso por conta da constatação de que o grupo CaWm do P7 é o único, entre seus pares espalhados pela rede de pontos, a ter feito a configuração [ɻ] advir. Adicionalmente, existe uma peculiaridade envolvendo o P7 que tem a ver com a variação formada por três covariantes. A localidade em questão é uma de cinco (as outras quatro são: São Roque, P4; Itu, P6; Tietê, P8; e Piracicaba, P10) cujo grupo de falantes CaWf participou ativamente da emergência da configuração [ɻ]-[ɾ]-[Ø].

CxGy_E2: No tocante à série CxGy em conjunção com a especificidade 2, a atenção ora se volta à variação constituída por duas covariantes. O P7 juntou-se a Pirapora do Bom Jesus (P2) no engendramento da configuração [ɻ]-[Ø], uma de duas (a outra é [ɻ]-[ɾ]) exibidas na rede de pontos, e as instâncias dessa configuração variacional em ambas as localidades foram realizadas por falantes do grupo CaGII.

CxWz_E2: Preservados o tema da E2 e o da variação entre duas covariantes, a série a ser abordada agora é CxWz. Uma de duas configurações, [ɻ]-[Ø] (a outra é [ɻ]-[ɾ]), foi concebida pelo P7 e por Capivari (P9). Em ambos os locais, o grupo que enunciou as instâncias de [ɻ]-[Ø] é CaWf. Além disso, o P7 também se envolveu com a segunda configuração, [ɻ]-[ɾ] — de maneira inversa, porém. Dito com detalhes, a segunda configuração figurou ao longo da rede de pontos, exceto no P7 e em outros três municípios (Pirapora do Bom Jesus, P2; Araçariguama, P3; e Capivari, P9).

WzGy_E2: Quanto à série WzGy em associação com a especificidade 2, um último fato relacionado (inversamente) com o P7 refere-se também à variação constituída por duas covariantes. Uma de duas configurações, [ɻ]-[ɾ] (a outra é [ɻ]-[Ø]), foi registrada ao longo de toda a rede de pontos, salvo no P7 e no P3.

Tietê (P8)

CxGy_E1: Quando fazemos o cruzamento entre, de um lado, as variáveis extralinguísticas 'classe escolar' e 'grupo etário' (ref. à série CxGy) e, de outro lado, os resultados linguísticos concernentes aos fones e ao apagamento do /R/ em coda externa de desinência infinitiva seguida por pausa (ref. à E1), vemos que o P8 é uma de três localidades (as outras duas são Pirapora do Bom Jesus, P2, e Sorocaba, P5) onde se observou uma ausência de variação expressa por instâncias da configuração [ɻ] realizadas pelo grupo CbGI, o qual se revelou, entre os quatro possíveis, ser o predominante na geração da configuração mencionada quando os demais grupos de falantes também são considerados.

CxWz_E1: O P8 também se juntou a outros municípios na constituição de um aspecto peculiar envolvendo agora a série CxWz em conjunção com a especificidade 1 a propósito da variação entre três covariantes. O grupo de informantes que mais atualizou uma de duas configurações, [ɻ]-[ɾ]-[Ø] (a outra é [ɻ]-[ɾ]-[ɾ]), é CaWf, autor de cinco de

11 instâncias totais abrangendo outros grupos; e esse fato esteve mais concentrado diatopicamente na parte alta (em relação ao rio Tietê) da área da região do Médio Tietê em pauta. Os locais que estiveram implicados na formação do evento em tela, além do P8, são: São Roque (P4), Itu (P6), Porto Feliz (P7) e Piracicaba (P10).

CxGy_E2: Relativamente à série CxGy e à E2, um aspecto relacionado ao P8 diz respeito à variação nula. Isso porque a configuração única [ɻ] — embora com instâncias abundantes, ao longo de toda a rede de pontos — alojou-se exclusivamente entre os entrevistados de ambos os grupos jovens (CaGI e CbGI) no P8. Outro realce a ser feito atinente ao P8 tem a ver com a variação constituída por duas covariantes: essa localidade é a única a ter produzido instância da configuração [ɻ]-[Ø] através do grupo CbGII.

WzGy_E2: Em se tratando da série WzGy e especificidade 2, em perspectiva com o P8 e com a variação fundada em duas covariantes, a localidade em foco foi parceira de Pirapora do Bom Jesus (P2) acerca de um dado aspecto: seus informantes masculinos mais velhos são os únicos, de toda a rede de pontos, a guardarem em comum o fato de terem produzido uma de duas configurações variacionais registradas, [ɻ]-[Ø] (a outra é [ɻ]-[ɾ]).

Capivari (P9)

CxWz_E1: O P9, no tocante à série CxWz e à E1, propiciou um acontecimento que, tendo em conta toda a rede de pontos, foi próprio de Capivari, ligado à variação composta por duas covariantes. Trata-se de uma de três configurações, [ɻ]-[ɾ] (as outras duas são [ɾ]-[Ø] e [ɻ]-[Ø]), que, quando cunhada por indivíduos representantes do grupo CaWf, só se apresentou no P9.

CxWz_E2: Ainda, com a mesma relação à série CxWz, mas agora implicando a especificidade 2, o P9 envolveu-se com outros dois aspectos, ambos com referência à variação constituída por duas covariantes. A esse respeito, houve duas configurações: [ɻ]-[Ø] e [ɻ]-[ɾ]. Por um lado, o P9 e Porto Feliz (P7) são os únicos municípios da rede de pontos a terem se alinhado na geração da primeira configuração por apenas falantes do grupo CaWf. Por outro lado, o P9 uniu-se a Pirapora do Bom Jesus (P2), Araçariguama (P3) e Porto Feliz (P7) na situação em que tais localidades são as únicas onde a segunda configuração, [ɻ]-[ɾ], não foi observada.

Piracicaba (P10)

CxGy_E1: Entre outras, uma peculiaridade que surgiu relativamente ao P10 diz respeito à variação composta por três covariantes, no âmbito da série CxGy (*i.e.*, da classe escolar e da idade) e concernente à especificidade 1 (ou seja, ao apagamento e aos róticos em coda externa de desinência infinitiva seguida por pausa): de toda a rede de pontos, o P10 é a única cidade a ter engendrado uma de duas configurações, [ɾ]-[Ø]-[r] (a outra é [ɻ]-[ɾ]-[Ø]), e isso se deu via membros do grupo CaGII.

CxWz_E1: Com atinência à série CxWz e à E1, um fato singular respeitante ao P10 remete-se também ao tema da variação entre três covariantes. Isso porque a localidade em foco é a única a ter tido seus informantes do grupo CaWm a produzir a configuração [ɻ]-[ɾ]-[r]. Paralelamente à configuração variacional indicada, houve também uma segunda, [ɻ]-[ɾ]-[Ø], cuja emergência, no particular do grupo CaWf, liga-se ao P10, mas também a São Roque (P4), Itu (P6), Porto Feliz (P7) e Tietê (P8), fazendo com que esse fato se vincule especialmente à parte alta (em relação ao rio Tietê) da área da região do Médio Tietê em pauta.

WzGy_E1: Com referência à série WzGy e à especificidade 1, há o que se dizer sobre o P10 em perspectiva com a variação plasmada por duas covariantes. Uma de duas covariações, [ɻ]-[Ø] (a outra é [ɾ]-[Ø]), foi significativamente produtiva ao longo da rede de pontos. Entretanto, ela só não foi observada em Araçariguama (P3) e no P10. Além disso, no tocante à variação entre quatro covariantes, o P10 é o único *locus* da rede a ter oportunizado o aparecimento da configuração [ɻ]-[ɾ]-[Ø]-[r].

CxGy_E3: Neste instante, vejamos algo relativo à série CxGy e à E3 (*i.e.*, aos fones e ao apagamento do /R/ em coda interna em sílaba tônica) em conjunção com a variação constituída por três covariantes. Esse algo se refere simplesmente ao fato de que o P10 é o único município da nossa rede de pontos a ter tido seus falantes do grupo CaGII a atualizarem a configuração [ɻ]-[ɾ]-[r].

CxWz_E3: Outro aspecto especial envolvendo o P10 e ainda a especificidade 3, mas agora acerca da série CxWz, tem a ver com a variação consistente de três covariantes: o P10 é o único ponto, de toda a rede, a ter concebido a configuração [ɻ]-[ɾ]-[r], e o fez mediante seus indivíduos representantes do grupo CaWm.

WzGy_E3: Quanto à série WzGy (*i.e.*, ao gênero de registro/social e à idade) e à especificidade 3, uma particularidade a ser destacada tange à variação formada por três covariantes. Nesse escopo, a única configuração que foi registrada é [ɻ]-[ɾ]-[r], enunciada exclusivamente pelo grupo WmGII do P10.

CxGy_E4: Para se encerrarem os turnos das abordagens a respeito dos fatos ligados ao P10, resta um último destaque, associado à série CxGy e à E4 (*i.e.*, às contrapartes (a)fonéticas do /R/ em coda interna em sílaba átona) em perspectiva com o tópico da variação nula. Houve duas configurações nesse sentido, [ɾ] e [ɻ]. Enquanto a segunda foi observada em praticamente toda a rede de pontos, a primeira só se manifestou no P10, através de seu grupo CaGII.

6.4 Normas linguísticas e influências de variáveis extralinguísticas

Nesta seção, traremos à explicitude tanto as covariantes que se estabeleceram como normas⁺, quanto fatores extralinguísticos que possam ter influído em certos fatos linguísticos. Para isso, vamos continuar contando com as cartas fonético-fonológicas apresentadas anteriormente, bem como vamos usar tabelas estatísticas simples. As ponderações seguem agrupadas por especificidade (E1, E2, E3 e E4). Ademais, é importante frisar que os valores percentuais exibidos nas tabelas estão calculados exclusivamente com base nos resultados válidos coletados por nossa pesquisa.

/R/ em coda externa de desinência infinitiva seguida por pausa (E1)

Os resultados quantitativos relativos aos fones e ao apagamento do arquifonema /R/ na E1 que obtivemos na região do Médio Tietê são aqueles que a Tabela 8 na próxima página exibe, a qual adicionalmente informa as questões do Questionário Fonético-Fonológicoᵐ (QFFᵐ) utilizadas em campo para a formação da base de dados empíricos. Essa base de dados é a base documental geral que empregamos para elaborar os gráficos estatísticos integrantes das "Cartas /R/ E1" (já introduzidas anteriormente).

Questões do QFF^{m} utilizadas		Covariantes coletadas e seus valores válidos auferidos				Σ
		[ɻ]	[ɾ]	[Ø]	[r]	
Q18 <VARRE**R**>		36	6	37	1	80
Q36 <BOTA**R**>		26	10	44	0	80
Q153 <SAI**R**>		54	13	10	3	80
Σ	valor absoluto	116	29	91	4	240
	valor relativo	48%	12%	38%	2%	100%

QFF^{m} = Questionário Fonético-Fonológico modificado

Tabela 8: Base documental geral (com destaques) para as "Cartas /R/ E1"

O contraste entre os dados faz ver que o apagamento fonético em coda externa no caso do vocábulo <SAI**R**> pouco aconteceu. Isso porque, por hipótese, a realização [sa'i] não se ligaria muito fortemente à forma infinitiva do verbo em apreço, mas sim à sua forma conjugada na primeira pessoa do singular no pretérito perfeito do indicativo, grafematicamente^{+} representada como <saí>.

De modo global, sendo o caso de mais preenchimento fonético do que de apagamento, o rótico que acabou sendo mais bem-sucedido entre os falantes em geral é [ɻ]. Importante destacar, nesse sentido, que a competição que a covariante [ɻ] venceu foi, de fato, em concorrência mais diretamente com o apagamento, [Ø]; constatação essa à qual se chega se os resultados advindos pela aplicação das questões 18 e 36 são postos em perspectiva com os resultados oriundos da elicitação^{+} mediante a questão 153.

A respeito de norma linguística (ou, no caso, de norma fonético-fonológica, mais precisamente), considerando-se os resultados das três questões (Q18, 36 e 153) em bloco, nenhuma das covariantes em foco alçou a posição de norma absoluta; posição essa, entretanto, quase atingida pelo fone [ɻ], com seus 48% de frequência relativa e sua distribuição diatópica regular^{+}.

Note-se que, se apenas os resultados da Q153 fossem considerados, a covariante [ɻ] seria norma absoluta, dado que ela atingiu 68% de frequência e foi registrada ao longo de toda a rede de pontos. Mediante esse mesmo exercício abstrativo, mas agora com escopo sobre a covariante [Ø], o apagamento também seria norma absoluta se apenas os resultados da Q36 fossem considerados, uma vez que a covariante apontada obteve, nessa situação, 55% de frequência, além de ter emergido ao longo de toda a rede de pontos.

Vejamos os resultados envolvendo os róticos e o apagamento do /R/ em coda externa de desinência infinitiva seguida por pausa (E1) organizados por grupos de informantes.

Vamos abordar primeiramente a série de grupos CxGy, com apoio da Tabela 9 abaixo, para seguirmos com as demais séries posteriormente.

Covar.	CaGII		CaGI		CbGII		CbGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	27	45%	34	57%	18	30%	37	62%
[ɾ]	15	25%	8	13%	4	7%	2	3%
[Ø]	17	28%	18	30%	35	58%	21	35%
[r]	1	2%	0	0%	3	5%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● GII/I = grupo etário velho/jovem

Tabela 9: Base documental (com destaques) da série CxGy para a "Carta /R/ E1" correspondente

Com os dados distribuídos pelas diferentes parametrizações da série CxGy (CaGII, CaGI, CbGII e CbGI) em mãos, exibidos na Tabela 9, e tendo-se ainda a "Carta /R/ E1"[2] correspondente em mente, verifica-se que a covariante [ɻ] se estabeleceu como norma relativa entre os jovens, e não entre os velhos, e isso em ambas as classes escolares, cujos valores percentuais se mostraram semelhantes entre si: 57% de frequência entre os jovens com nível alto de escolaridade e 62% entre os jovens com nível baixo de escolaridade. Em todo caso, fato é que o valor percentual do fone [ɻ] na E1 se evidenciou sensivelmente mais alto entre os falantes mais novos com baixa escolaridade.

Já entre os indivíduos mais velhos, o destaque fica para a covariante [Ø] com relação aos inquiridos com nível escolar baixo, no sentido em que o apagamento obteve 58% de frequência relativa e distribuição diatópica homogênea ao longo de toda a rede de pontos, o que faz dele uma norma relativa no bojo dos falantes representantes do grupo CbGII.

[2] Nome técnico: CFF_4C_CxGy_(E1=+SDI{'(C)V/R/||}).

Outra informação relevante inferível da Tabela 9 diz respeito aos resultados da covariante [ɾ]. Se, por um momento, considerarmos a sequência 'CaGII, CaGI, CbGII e CbGI' como sendo decrescente, veremos que os valores obtidos pela covariante [ɾ] quase configuraram uma progressão geométrica crescente (em relação à sequência) com razão aproximada dois (q ≅ 2). Assumindo essa perspectiva, a constatação em tela acarreta o estado segundo o qual quanto mais novo se é e menos escolaridade se tem, menor é a frequência relativa do fone [ɾ].

Para discutirmos os resultados linguísticos associados à série CxWz, vamos tomar a Tabela 10.

Covar.	CaWf		CaWm		CbWf		CbWm	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	31	52%	30	50%	23	38%	32	53%
[ɾ]	12	20%	11	18%	4	7%	2	3%
[Ø]	17	28%	18	30%	30	50%	26	43%
[ɾ]	0	0%	1	2%	3	5%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● Wf/m = gênero de registro/social feminino/masculino

Tabela 10: Base documental (com destaques) da série CxWz para a "Carta /R/ E1"[3] correspondente

No tocante à covariante [ɻ] entre os informantes do grupo CaWf, ela por muito pouco quase se constituiu como norma relativa. Isso porque, apesar de ter aparecido com 52% de frequência, ela não figurou em uma das dez localidades da rede de pontos, a saber, Sorocaba (P5). Todavia, o fone em consideração atingiu o *status* de norma relativa no tocante a outro grupo, CbWm, com 53% de frequência; o qual, aliás, não guarda em comum qualquer característica dimensional ora sob enfoque (*i.e.*, diastrática e diageracional) com o grupo há pouco mencionado, ou seja, com o grupo CaWf.

[3] Nome técnico: CFF_4C_CxWz_(E1=+SDI{'(C)V/R/||}).

Outrossim, observa-se que ambas as classes escolares se colocaram em lados opostos quando o assunto é variação entre [ɾ] e [Ø]. Ou seja, de um lado, os grupos CaWf e CaWm apresentaram, respectivamente, os valores (próximos) 20% e 18% para a covariante [ɾ] e, de outro lado, os grupos CbWf e CbWm mostraram 7% e 3% para a mesma covariante. Paralela e analogamente, quanto à covariante [Ø], os primeiros expuseram, também respectivamente, os valores (próximos) 28% e 30%, enquanto os segundos expressaram 50% e 43%.

Relativamente à terceira e última série, WzGy, miremos antes de mais nada a Tabela 11 abaixo.

Covar.	WfGII		WfGI		WmGII		WmGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	18	30%	36	60%	27	45%	35	58%
[ɾ]	10	17%	6	10%	9	15%	4	7%
[Ø]	29	48%	18	30%	23	38%	21	35%
[r]	3	5%	0	0%	1	2%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Wf/m = gênero de registro/social feminino/masculino ● GII/I = grupo etário velho/jovem

Tabela 11: Base documental (com destaques) da série WzGy para a "Carta /R/ E1"[4] correspondente

Um primeiro fato a ser explicitado a partir da Tabela 11 relaciona-se com a covariante [ɻ]. Entre os indivíduos pertencentes à geração mais nova, esse fone foi registrado tanto ao longo de toda a rede de pontos, quanto com um percentual relativo superior a cinquenta por cento, o que faz do som sob enfoque uma norma relativa em ambos os grupos da faixa etária aludida, WfGI e WmGI, os quais obtiveram, respectivamente, os valores (próximos) 60% e 58%.

Com atinência ao grupo feminino mais velho, WfGII, a covariante [Ø] quase se erigiu em norma relativa, com seus 48% de frequência e distribuição regular. Em todo caso, essa covariante, no grupo apontado, obteve 18 pontos percentuais a mais do que o fone

[4] Nome técnico: CFF_4C_WzGy_(E1=+SDI{'(C)V/R/||}).

[ɻ], o qual, por seu turno, considerados todos os grupos juntos, se manifestou o mais comum, tal como a Tabela 8 anteriormente o sugeriu.

/R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa (E2)

Analisemos agora os resultados das realizações fonéticas e do apagamento do /R/ na E2. Para isso, vamos partir da Tabela 12 abaixo, que abarca os valores das covariantes que estiveram em concorrência para a saturação da variável linguística em questão. Usamos essas informações para conceber os gráficos quantitativos presentes nas "Cartas /R/ E2"; com as quais, aliás, também vamos contar para nossos apontamentos mais adiante.

Questões do QFF^{m} utilizadas		Covariantes coletadas e seus valores válidos auferidos					Σ
		[ɻ]	[ɾ]	[Ø]	[r]	[h]	
Q25 <COLHE**R**> (subst.)		71	7	2	0	0	80
Q61 <CALO**R**>		66	12	2	0	0	80
Q129 <MULHE**R**>		67	11	1	0	1	80
Σ	valor absoluto	204	30	5	0	1	240
	valor relativo	85%	13%	2%	0%	0%	100%

QFF^{m} = Questionário Fonético-Fonológico modificado

Tabela 12: Base documental geral (com destaques) para as "Cartas /R/ E2"

Desconsiderando-se os resultados expressos na Tabela 12 não significativos ligados a outras covariantes (especialmente a [r] e a [h]), um dos primeiros fatos a serem explorados remete-se à uma certa constância dos valores das covariantes [ɻ] e [ɾ] em suas respectivas colunas. Isso indica que as diferentes opções em concorrência, de acordo com previsões, não causaram conflitos semânticos^{+} ou categoriais (diferentemente do que vimos acerca — e a partir — da Tabela 8 anteriormente). O poder da covariante [ɻ] subjugou a principal concorrente, [ɾ], e todas as demais, com seus enérgicos 85% de frequência relativa e presença ostensiva em todas as localidades da rede; fato que, por consequência, fez dela uma norma absoluta na região do Médio Tietê.

Ademais, para se aventarem algumas palavras sobre a covariante [Ø], seus baixos valores percentuais aqui se contrastam drasticamente com os valores que ela obteve na E1 — *i.e.*, em coda externa de desinência infinitiva seguida por pausa (cf. Tabela 8) —, o que

deixa clara a influência de questões formais/estruturais atuantes sobre este e aquele caso, provocando efeitos fonéticos sensivelmente distintos.

Aumentando o *zoom* das análises, vamos em direção à discretização dos valores gerais em valores especificamente associados aos diferentes grupos de informantes. É esse processo que permite que se veja se há possíveis influências extralinguísticas específicas sobre os fatos linguísticos em foco. Em nosso caso — para oferecermos um lembrete rápido —, os aspectos extralinguísticos controlados são diatópicos, diastráticos, diageracionais e diagenéricos.

Como o leitor já deve estar acostumado a esta altura, vamos iniciar as ilações com os fenômenos pertinentes à série CxGy, a partir da Tabela 13 abaixo, bem como com base na carta fonético-fonológica /R/ correspondente, já trazida páginas antes.

Covar.	CaGII		CaGI		CbGII		CbGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	39	65%	53	88%	54	90%	58	97%
[ɾ]	18	30%	7	12%	3	5%	2	3%
[Ø]	3	5%	0	0%	2	3%	0	0%
[r]	0	0%	0	0%	0	0%	0	0%
[h]	0	0%	0	0%	1	2%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● GII/I = grupo etário velho/jovem

Tabela 13: Base documental (com destaques) da série CxGy para a "Carta /R/ E2"[5] correspondente

Além de o som [ɻ] ter se revelado uma norma absoluta — tal como vimos por ensejo da Tabela 12 mais atrás —, ele também se estabeleceu como norma relativa — segundo podemos ver agora na Tabela 13 imediatamente acima —, e isso em relação a todos os distintos grupos, especialmente entre os falantes do grupo CbGI, responsáveis por imódicos 97% de frequência da covariante em consideração.

[5] Nome técnico: CFF_4C_CxGy_(E2=-SDI{'(C)V/R/||}).

Outro aspecto concernente à covariante [ɻ] que a Tabela 13 permite enxergar é o aumento decrescente (em relação à sequência inframencionada) da frequência de ocorrência desse fone entre os grupos de inquiridos, caso consideremos 'CaGII, CaGI, CbGII e CbGI' como sendo uma sequência decrescente de termos (em conformidade com o que já fizemos antes por oportunidade das inferências baseadas na Tabela 9). Admitindo-se então tais assunções, isso significa dizer que, quanto menores são o nível escolar e a idade, maior é a frequência de aparecimento do rótico em apreço.

Ademais, observa-se ainda que a concorrência real no caso da especificidade 2 alinhada com a série CxGy desenvolveu-se na verdade entre as covariantes [ɻ] e [ɾ]. Por esse ângulo, nota-se que os valores percentuais de ambas cresceram ou diminuíram inversamente em relação a si mesmas nos grupos de informantes. Ou seja, tomando-se 'CaGII, CaGI, CbGII e CbGI' como uma sequência decrescente, enquanto os valores de [ɻ] aumentaram decrescentemente (em relação à sequência), os valores de [ɾ] aumentaram, pelo contrário, crescentemente (em relação à sequência).

Concentremo-nos agora nos resultados postos em perspectiva com a série CxWy. Para fazê-lo, vamos tomar a Tabela 14 a seguir como referência inicial.

Covar.	CaWf		CaWm		CbWf		CbWm	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	41	68%	51	85%	56	93%	56	93%
[ɾ]	17	28%	8	13%	3	5%	2	3%
[Ø]	2	3%	1	2%	0	0%	2	3%
[r]	0	0%	0	0%	0	0%	0	0%
[h]	0	0%	0	0%	1	2%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● Wf/m = gênero de registro/social feminino/masculino

Tabela 14: Base documental (com destaques) da série CxWz para a "Carta /R/ E2"[6] correspondente

[6] Nome técnico: CFF_4C_CxWz_(E2=-SDI{'(C)V/R/||}).

Se sistematicamente compararmos os valores da Tabela 13 precedente (a qual exibiu os resultados do cruzamento entre os dados linguísticos e as variáveis extralinguísticas diastrática e diageracional) com os valores da Tabela 14 (que apresenta, na página antecedente, os resultados do cruzamento entre os dados linguísticos e as variáveis extralinguísticas diastrática e diagenérica), observaremos que as diferenças percentuais não são significativas. Essa condição sugere pelo menos que as variáveis diageracional e diagenérica não exerceram influência significativa nos resultados linguísticos sob enfoque. Portanto, isso implica que o que dissemos antes sobre os valores da Tabela 13 aplica-se, *mutatis mutandis*, aos valores da Tabela 14[7].

Isso dito, dirijamo-nos às correlações entre a série WzGy e os resultados fonéticos e de apagamento concernente ao /R/ em coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa. Para isso, vamos inicialmente contar com a Tabela 15.

Covar.	WfGII		WfGI		WmGII		WmGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	43	72%	54	90%	50	83%	57	95%
[ɾ]	14	23%	6	10%	7	12%	3	5%
[Ø]	2	3%	0	0%	3	5%	0	0%
[r]	0	0%	0	0%	0	0%	0	0%
[h]	1	2%	0	0%	0	0%	0	0%
Σ	60	100%	60	100%	60	100%	60	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Wf/m = gênero de registro/social feminino/masculino ● GII/I = grupo etário velho/jovem

Tabela 15: Base documental (com destaques) da série WzGy para a "Carta /R/ E2"[8] correspondente

Com valores bastante altos, a covariante [ɻ] mostrou-se norma relativa mais uma vez, e isso em todos os grupos de informantes, especialmente no grupo masculino jovem, WmGI, em que o fone em pauta atingiu 95% de frequência relativa. O grupo com o menor valor é o completamente oposto ao WmGI, *i.e.*, WfGII, que é o grupo das mulheres mais velhas, no

[7] Diferenças em pontos percentuais da Tabela 14 em relação à Tabela 13 (em ordem das colunas): [ɻ]: +3 | -3 | +3 | -4 ● [ɾ]: -2 | +1 | 0 | 0 ● [Ø]: -2 | +2 | -3 | +3 ● [r]: 0 | 0 | 0 | 0 ● [h]: 0 | 0 | 0 | 0.
[8] Nome técnico: CFF_4C_WzGy_(E2=-SDI{'(C)V/R/||}).

qual [ɻ] obteve 72%; e é neste último grupo, aliás, que a concorrência real — performada entre [ɻ] e [ɾ] no presente caso — estabeleceu-se mais acirradamente, com o som [ɾ] contando com 23% de ocorrência relativa.

/R/ em coda interna em sílaba tônica (E3)

É chegada a hora de lidarmos com algumas informações referentes à especificidade 3. A Tabela 16 a seguir nos ajudará com isso. Nela, estão reunidos os resultados coletados em campo sobre as ocorrências das covariantes que concorreram entre si para a saturação da variável linguística que a E3 representa. Assim como as tabelas do mesmo tipo que vimos antes, esta a seguir também informa mediante quais questões do QFF^{m} os dados foram levantados.

Questões do QFF^{m} utilizadas		Covariantes coletadas e seus valores válidos auferidos					Σ
		[ɻ]	[ɾ]	[Ø]	[r]	[h]	
Q105 <CE**R**TO>		64	12	0	2	1	79
Q158 <ESQUE**R**DO>		69	10	0	1	0	80
Σ	valor absoluto	133	22	0	3	1	159
	valor relativo[9]	84%	14%	0%	2%	1%	100%

QFF^{m} = Questionário Fonético-Fonológico modificado

Tabela 16: Base documental geral (com destaques) para as "Cartas /R/ E3"

Como se vê, o predomínio contundente da covariante [ɻ] na E3 foi bastante veemente na região do Médio Tietê, a julgar por seus 84% de frequência relativa. Essa constatação associada ao fato de que o som em apreço foi registrado ao longo de toda a rede de pontos acarreta, por consequência, que seu *status* variacional é de norma absoluta. Em segundo lugar (e último, se desconsideradas as demais covariantes cujos valores não foram significativos), figurou o fone [ɾ], com seus não tão contundentes 14%.

[9] Somatório aparente aqui é de 101% por conta dos arredondamentos envolvendo frações.

Agora, vejamos se os resultados linguísticos tiveram alguma relação estreita ou não com alguma variável extralinguística. Para iniciá-lo, vamos começar com a série CxGy, a partir da Tabela 17.

Covar.	CaGII		CaGI		CbGII		CbGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	27	68%	30	77%	37	93%	39	98%
[ɾ]	11	28%	9	23%	1	3%	1	3%
[Ø]	0	0%	0	0%	0	0%	0	0%
[r]	1	3%	0	0%	2	5%	0	0%
[h]	1	3%	0	0%	0	0%	0	0%
Σ	40	100%	39	100%	40	100%	40	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● GII/I = grupo etário velho/jovem

Tabela 17: Base documental (com destaques) da série CxGy para a "Carta /R/ E3"[10] correspondente

Queremos destacar dois fatos a partir da Tabela 17. O primeiro deles diz respeito à constatação de que a covariante [ɻ] consolidou-se como norma relativa em cada um dos quatro grupos de informantes, com especial força naqueles integrantes da classe escolar baixa, CbGII e CbGI, os quais contaram com 93% e 98% de frequência relativa, respectivamente.

O segundo fato a ser salientado refere-se à observação de que uma concorrência real só se deu entre [ɻ] e [ɾ], e isso só nos grupos representantes da classe escolar alta, CaGII e CaGI. Sob um ângulo bastante simplificado e meramente ilustrativo, é possível dizer que, de dez indivíduos do grupo CaGII, praticamente sete realizaram a covariante [ɻ], enquanto praticamente três atualizaram o fone [ɾ]. Sob esse mesmo ângulo de abordagem, de dez falantes do grupo CaGI, praticamente oito enunciaram o som [ɻ], enquanto praticamente dois produziram [ɾ].

[10] Nome técnico: CFF_4C_CxGy_(E3='(C)V/R/.).

Ajustemos nosso foco de atenção ao cruzamento entre os resultados linguísticos e as variáveis extralinguísticas que a série CxWz representa. Para isso, vamos utilizar a Tabela 18 como ponto de partida.

Covar.	CaWf		CaWm		CbWf		CbWm	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	25	64%	32	80%	38	95%	38	95%
[ɾ]	13	33%	7	18%	0	0%	2	5%
[Ø]	0	0%	0	0%	0	0%	0	0%
[r]	0	0%	1	3%	2	5%	0	0%
[h]	1	3%	0	0%	0	0%	0	0%
Σ	39	100%	40	100%	40	100%	40	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● Wf/m = gênero de registro/social feminino/masculino

Tabela 18: Base documental (com destaques) da série CxWz para a "Carta /R/ E3"[11] correspondente

Os resultados exibidos na Tabela 18 não discrepam significativamente daqueles expostos na Tabela 17. Isso significa que as variáveis extralinguísticas diagenérica e diageracional não funcionaram como fatores de influência importante sobre os resultados linguísticos. No entanto, no caso da série CxWz, a covariante [ɻ] não se assomou em norma relativa no grupo feminino da classe escolar alta (diferentemente do que vimos por ensejo da série CxGy há pouco), posto que ela não foi registrada no quadrante correspondente da localidade de Sorocaba (P5). As demais ilações incidentes nos resultados manifestos na Tabela 17 essencialmente aplicam-se, *mutatis mutandis*, aos resultados expressos na Tabela 18.

Feitos os exames voltados às séries CxGy e CxWz, resta incluirmos apontamentos sobre a série WzGy em perspectiva com as covariantes (róticos e apagamento) coletadas em coda interna em sílaba tônica. Para esse desiderato, vamos partir da Tabela 19 a seguir.

[11] Nome técnico: CFF_4C_CxWz_(E3='(C)V/R/.).

Covar.	WfGII		WfGI		WmGII		WmGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	31	78%	32	82%	33	83%	37	93%
[ɾ]	6	15%	7	18%	6	15%	3	8%
[Ø]	0	0%	0	0%	0	0%	0	0%
[r]	2	5%	0	0%	1	3%	0	0%
[h]	1	3%	0	0%	0	0%	0	0%
Σ	40	100%	39	100%	40	100%	40	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Wf/m = gênero de registro/social feminino/masculino ● GII/I = grupo etário velho/jovem

Tabela 19: Base documental (com destaques) da série WzGy para a "Carta /R/ E3"[12] correspondente

Depreende-se tanto da tabela acima quanto da carta fonético-fonológica correspondente que a covariante [ɻ] obteve o *status* de norma relativa em três grupos de entrevistados — WfGI, WmGII e WmGI —, mormente no grupo masculino jovem (WmGI), no qual o fone em questão atingiu eloquentes 93% de frequência relativa.

No tocante à concorrência real entre as covariantes, ela se verificou substancialmente apenas entre [ɻ] e [ɾ], em todos os grupos de informantes — mas ela foi bem pouco acirrada no grupo dos homens jovens, WmGI, no qual [ɾ] contou com apenas 8% contra 93% da concorrente [ɻ] —, com sutil destaque no grupo das mulheres jovens, WfGI, em que os valores da emulação foram 18% para [ɾ] e 82% para [ɻ].

Para encerrarmos esta investida analítica (sem pretensão de esgotamento) acerca dos resultados colhidos em campo referentes às quatro especificidades (ou situações estruturais) escolhidas para este momento (E1, E2, E3 e E4), precisamos dissertar sobre a última delas, E4.

/R/ em coda interna em sílaba átona (E4)

Para versarmos sobre os resultados válidos produzidos por nossos 80 informantes e auferidos por nós a propósito da E4, vamos, como já previsto a esta altura, lançar mão de uma tabela que nos comunica as informações mais gerais a esse respeito. Nela, encontra-se

[12] Nome técnico: CFF_4C_WzGy_(E3='(C)V/R/.).

a indicação de quais questões do QFFᵐ foram mobilizadas e quais foram seus respectivos valores variacionais registrados na pesquisa.

Questões do QFFᵐ utilizadas		Covariantes coletadas e seus valores válidos auferidos				Σ
		[ɻ]	[ɾ]	[Ø]	[r]	
Q12 <TO**R**NEIRA>		69	10	0	1	80
Q27 <FE**R**VENDO>		67	13	0	0	80
Q110 <PE**R**DÃO>		63	15	0	2	80
Q144 <PE**R**FUME>		64	15	0	1	80
Q152 <PE**R**GUNTAR>		68	12	0	0	80
Σ	valor absoluto	331	65	0	4	400
	valor relativo	83%	16%	0%	1%	100%

QFFᵐ = Questionário Fonético-Fonológico modificado

Tabela 20: Base documental geral (com destaques) para as "Cartas /R/ E4"

De um universo de 400 (100%) respostas, 331 (83%) contiveram a covariante [ɻ], em detrimento das demais. Isso, somado ao fato de que o fone mencionado foi registrado ao longo de toda a rede de pontos, conferiu ao som [ɻ] a condição de norma absoluta. Isso posto, outro aspecto, a ser apontado de passagem, é a semelhança entre os valores dos resultados obtidos com ajuda das cinco questões do QFFᵐ mencionadas na Tabela 20.

A ver que informações emergem com a correlação desses dados linguísticos com as dimensões extralinguísticas que as séries representam; a começar com a série CxGy, por azo da Tabela 21 na página adiante.

Covar.	CaGII		CaGI		CbGII		CbGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	66	66%	78	78%	93	93%	94	94%
[ɾ]	33	33%	22	22%	4	4%	6	6%
[Ø]	0	0%	0	0%	0	0%	0	0%
[r]	1	1%	0	0%	3	3%	0	0%
Σ	100	100%	100	100%	100	100%	100	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● GII/I = grupo etário velho/jovem

Tabela 21: Base documental (com destaques) da série CxGy para a "Carta /R/ E4"[13] correspondente

Além de gozar do *status* de norma absoluta, como visto mais atrás, a covariante [ɻ] também desfruta da condição de norma relativa, e isso em relação a três dos quatro grupos de falantes — CaGI, CbGII e CbGI —, máxime relativamente a estes dois últimos, ambos pertencentes à classe escolar baixa, com seus suasórios 93% e 94% de frequência relativa, respectivamente. Quanto ao grupo dos informantes mais velhos com escolaridade alta (CaGII), embora a covariante em pauta tenha alcançado não desprezíveis 66% de frequência relativa, ela não figurou no quadrante correspondente da localidade de Piracicaba (P10); razão pela qual a posição de norma relativa no grupo mencionado lhe escapou.

Outra menção específica cabe fazer, agora com referência à covariante [ɾ], a qual obteve valores mais propriamente significativos só entre os inquiridos com nível escolar alto, *i.e.*, entre os indivíduos dos grupos CaGII e CaGI. É nesses grupos, portanto, que uma concorrência real concernente à variação linguística em tela foi registrada, a saber, entre [ɻ] e [ɾ]; e essa concorrência foi mais acirrada entre os falantes mais velhos com escolaridade alta, posto que lá a distância entre os percentuais de [ɻ] (66%) e [ɾ] (33%) foi a menor observada, em contraste com os outros três grupos.

Olhemos as relações a serem explicitadas a partir da Tabela 22 adiante, atinente à série CxWz.

[13] Nome técnico: CFF_4C_CxGy_(E4=(C)V/R/.).

Covar.	CaWf		CaWm		CbWf		CbWm	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	65	65%	79	79%	91	91%	96	96%
[ɾ]	34	34%	21	21%	6	6%	4	4%
[Ø]	0	0%	0	0%	0	0%	0	0%
[ɾ]	1	1%	0	0%	3	3%	0	0%
Σ	100	100%	100	100%	100	100%	100	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Ca/b = classe escolar alta/baixa ● Wf/m = gênero de registro/social feminino/masculino

Tabela 22: Base documental (com destaques) da série CxWz para a "Carta /R/ E4"[14] correspondente

A covariante [ɻ] estabeleceu-se como norma relativa em três dos quatro grupos de participantes da investigação — CaWm, CbWf e CbWm —, com 79%, 91% e 96% de frequência relativa, respectivamente. Quanto ao grupo CaWf, embora a covariante em tela tenha obtido um valor percentual superior a cinquenta por cento, ela não figurou no quadrante correspondente de Sorocaba (P5).

Como se vê, a força do fone [ɻ] entre os indivíduos com nível escolar baixo foi bastante considerável. Outro fato envolvendo esse som que atrai a atenção é a progressão do valor percentual a partir, de um lado, do grupo das mulheres com alta escolaridade até, de outro lado, os homens com baixa escolaridade. Se lançarmos mão da ideia de que 'CaWf, CaWm, CbWf e CbWm' pode ser encarado como uma sequência decrescente, isso significará dizer que seus valores aumentaram decrescentemente (em relação à sequência).

Também, infere-se que uma concorrência real teve lugar praticamente apenas nos grupos de falantes representantes da classe escolar alta, a partir da variação constituída por [ɻ] e [ɾ], especialmente entre as mulheres, que produziram a competição mais acirrada: foram não poucos 34% obtidos pela covariante [ɾ] contra 65% de frequência relativa da concorrente [ɻ].

Lancemos um olhar agora aos resultados linguísticos concernentes às diferentes realizações fonéticas e ao apagamento do /R/ em coda interna em sílaba átona em perspectiva com as

[14] Nome técnico: CFF_4C_CxWz_(E4=(C)V/R/.).

dimensões diagenérica e diageracional, *i.e.*, com a série WzGy. Para fazê-lo, vamos contar com a Tabela 23 a seguir.

Covar.	WfGII		WfGI		WmGII		WmGI	
	VA	VR	VA	VR	VA	VR	VA	VR
[ɻ]	71	71%	85	85%	88	88%	87	87%
[ɾ]	25	25%	15	15%	12	12%	13	13%
[Ø]	0	0%	0	0%	0	0%	0	0%
[r]	4	4%	0	0%	0	0%	0	0%
Σ	100	100%	100	100%	100	100%	100	100%

Covar. = covariante ● VA = valor absoluto ● VR = valor relativo
Wf/m = gênero de registro/social feminino/masculino ● GII/I = grupo etário velho/jovem

Tabela 23: Base documental (com destaques) da série WzGy para a "Carta /R/ E4"[15] correspondente

Sob o ângulo da série WzGy, a covariante [ɻ] apresentou-se como norma relativa nos grupos WfGI, WmGII e WmGI, com 85%, 88% e 87% de frequência relativa, respectivamente. No grupo WfGII, a despeito de ter atingido 71% de ocorrência, a covariante em questão não foi registrada no quadrante correspondente de Sorocaba (P5).

Por fim, com referência à concorrência real verificada, entre [ɻ] e [ɾ], nota-se que ela foi significativa em todos os diferentes grupos de informantes, com destaque à sua manifestação entre as mulheres mais velhas, WfGII, com 25% de frequência relativa de [ɾ] e 71% de [ɻ].

6.5 Conclusões

Nesta seção, vamos sumariamente trazer para o primeiro plano conclusões e comentários a respeito das quatro situações estruturais do arquifonema /R/ que elegemos, para observar suas respectivas variações fonéticas na região paulista do Médio Tietê; postas, por sua vez, oportunamente em perspectiva com variáveis extralinguísticas.

[15] Nome técnico: CFF_4C_WzGy_(E4=(C)V/R/.).

Como já feito antes, as especificidades (ou situações estruturais) do /R/ serão frequentemente referidas de modo abreviado: E1 = coda externa de desinência infinitiva seguida por pausa; E2 = coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa; E3 = coda interna em sílaba tônica; e E4 = coda interna em sílaba átona.

Sob um ângulo geral:

(1) a E1 foi saturada sobretudo por [ɻ] (48%), depois por [Ø] (38%), por [ɾ] (12%) e, finalmente, por [r] (2%);

(2) a E2 foi saturada principalmente por [ɻ] (85%), em seguida por [ɾ] (13%), por [Ø] (2%) ([r] e [h] não pontuaram percentualmente);

(3) a E3 foi saturada especialmente por [ɻ] (84%) e então por [ɾ] (14%), por [r] (2%) e ainda por [h] (1%) ([Ø] não pontuou percentualmente)[16]; e

(4) a E4 foi saturada máxime por [ɻ] (83%), bem como por [ɾ] (16%) e por [r] (1%) ([Ø] não pontuou percentualmente).

Nota-se, pois, que os valores percentuais obtidos por todas as covariantes nas especificidades 2, 3 e 4 foram extremamente próximos uns aos outros. Já em relação à E1, duas de suas quatro covariantes registradas, [ɾ] e [r], obtiveram valores bastante próximos àqueles que elas atingiram nas outras especificidades (2, 3 e 4). Todavia, os valores de [ɻ] e [Ø] na E1, comparados com aqueles observados nas outras especificidades, manifestaram-se de maneira quantitativamente particular.

Ademais:

(5) o fone [r] não ocorreu, nenhuma vez, na E2; e

(6) o apagamento, [Ø], não ocorreu, nenhuma vez, nem na E3, nem na E4.

[16] Somatório aparente aqui é de 101% por conta dos arredondamentos envolvendo frações.

Sob o ângulo da série CxGy em perspectiva com a E1:

(7) a covariante [ɻ] revelou-se norma relativa apenas entre os jovens, tanto com alta escolaridade (57%) quanto com baixa escolaridade (62%);

(8) a covariante [Ø] apresentou-se como norma relativa (58%) apenas entre os falantes mais velhos com nível escolar baixo;

(9) constatou-se que, quanto mais novo se é e menos escolaridade se tem, menor é a frequência relativa do fone [ɾ] — CaGII: 25%; CaGI: 13%; CbGII: 7%; e CbGI: 3% —, cujos valores quase configuraram, como se vê, uma progressão geométrica com razão aproximada dois ($q \cong 2$);

(10) a covariante [r] não apareceu na fala da geração mais nova; e

(11) verificou-se uma simetria invertida quase perfeita entre os grupos CaGI e CbGII — os quais, aliás, são completamente opostos em relação às variáveis extralinguísticas que eles representam: o primeiro produziu [ɻ] (57%) e [Ø] (30%) quase precisamente na mesma proporção em que o segundo enunciou [Ø] (58%) e [ɻ] (30%).

O item (7) diz respeito a uma hipótese específica que levantamos na seção "Objetivos e hipótese", qual seja:

> *Os jovens entre 18 e 36 anos de idade do Médio Tietê (especialmente aqueles com nível escolar alto[17]) estão produzindo a covariante [ɻ] (a aproximante retroflexa[18]) como contraparte fonética do arquifonema /R/ em coda externa de desinência infinitiva seguida por pausa sob uma frequência relativa inferior àquela a ser observada entre os falantes mais velhos a partir de 55 anos de idade (especialmente aqueles com nível escolar baixo[19]).*

[17] A partir de estudos universitários, completos ou incompletos.
[18] Conhecida como "erre retroflexo" ou ainda como "erre caipira".
[19] No escopo de abrangência entre analfabetismo e ensino médio incompleto.

Ou seja, lembrando-se do que o item (7) expõe: hipótese testada e surpreendentemente refutada pelos dados empíricos. Na realidade incontestável dos fatos concretos, a fala dos jovens do Médio Tietê evidenciou-se, na verdade, mais albergadora da covariante [ɻ] do que a fala dos mais velhos no tocante à E1; e o grau dessa constatação torna-se mais eloquente na mesma medida em que se sabe que a covariante [ɻ] na E1 fez-se presente ao longo de toda a rede de pontos, com frequência relativa, como já visto antes (cf. Tabela 9), de 57% entre os jovens com nível escolar alto e 62% entre os jovens com nível escolar baixo (contra 45% entre os mais velhos com nível escolar alto e 30% entre os mais velhos com nível escolar baixo).

Agora, ainda sobre a E1, teçamos considerações atinentes à série CxWz:

(12) a covariante [ɻ] quase se constituiu em norma relativa entre as mulheres com alta escolaridade. Isso porque, embora com 52% de frequência relativa, ela não logrou figurar no quadrante correspondente de Sorocaba (P5);

(13) o som [ɻ], no entanto, adveio como norma relativa (53%) entre os homens com baixa escolaridade;

(14) acerca da concorrência entre [ɾ] e [Ø], ambas as classes escolares colocaram-se em flancos opostos. De um lado, os grupos CaWf e CaWm apresentaram, respectivamente, os valores (próximos) 20% e 18% para a covariante [ɾ] e, de outro lado, os grupos CbWf e CbWm mostraram 7% e 3% para a mesma covariante. Paralela e analogamente, quanto à covariante [Ø], os primeiros expuseram, também respectivamente, os valores (próximos) 28% e 30%, enquanto os segundos expressaram 50% e 43%;

(15) 74% das instâncias da configuração variacional [ɻ]-[Ø] é de responsabilidade dos indivíduos com nível escolar baixo; e

(16) as instâncias da configuração [ɻ]-[ɾ]-[Ø] produzidas por mulheres com alta escolaridade concentraram-se diatopicamente na parte alta (em relação ao rio Tietê) da porção da região do Médio Tietê sob enfoque.

No que tange aos itens (12) e (13), observe-se que, não obstante o fato de que ambos tenham obtido valores percentuais próximos na atualização do som [ɻ], seus perfis são completamente opostos (de um lado, mulheres com nível escolar alto; de outro lado, homens com nível escolar baixo). Quanto ao item (14), veja-se o corolário segundo o qual as mulheres em geral demonstraram maior predileção do que os homens relativamente às covariantes [ɾ] e [Ø].

Doravante, adotemos o ângulo da série WzGy em perspectiva com a E1.

(17) os jovens femininos e masculinos são os responsáveis por terem produzido a covariante [ɻ] sob *status* de norma relativa (WfGI: 60%; e WmGI: 58%);

(18) a covariante [Ø], por sua vez, quase se tornou norma relativa entre as mulheres mais velhas; não por falta de consistência a propósito da distribuição diatópica regular, mas sim por insuficiência percentual (48%);

(19) os homens geraram a maioria (67%) das ocorrências da configuração [ɻ]-[Ø], presente em toda a rede de pontos, exceto em Araçariguama (P3) e Piracicaba (P10); e

(20) as mulheres engendraram a maioria (73%) das ocorrências da configuração [ɻ]-[ɾ]-[Ø], presentes ao longo de toda a rede de pontos, salvo em Santana de Parnaíba (P1) e em Pirapora do Bom Jesus (P2).

Quanto ao item (17), se recuperarmos também o que dissemos mais atrás sobre o item (7) — em particular no que concerne à faixa etária, a qual é a variável extralinguística em comum entre ambos os itens ora postos em relação — e compararmos os valores

percentuais de um e outro item; bem como se tivermos em mente a distribuição regular do fone [ɻ]; perceberemos que aparentemente nem a variável diastrática, nem a variável diagenérica, nem a variável diatópica atuou como fator de determinação relevante no estabelecimento dos valores percentuais obtidos pela covariante [ɻ] a partir de 57% (valor desde o qual as séries envolvidas aqui quantitativamente assemelharam-se entre si), mas sim a variável diageracional. Essa conclusão agrega informações adicionais àquelas versadas anteriormente com referência à nossa hipótese (enunciada três páginas atrás).

Agora, seja o ângulo sob o qual vamos apresentar informações relacionadas aos resultados ligados à E2 em conjunção com a série CxGy:

(21) a covariante [ɻ] granjeou a condição de norma relativa, e isso em todos os quatro diferentes grupos de entrevistados, com maior destaque entre os jovens com baixa escolaridade, com quem o fone em foco obteve sobejos 97% de frequência relativa;

(22) ainda com referência à covariante [ɻ], constatou-se que, quanto menores são o nível escolar e a idade, maior é a frequência de aparecimento do rótico em apreço (CaGII: 65%; CaGI: 88%; CbGII: 90%; e CbGI: 97%);

(23) houve apenas uma concorrência real, entre [ɻ] e [ɾ], com predomínio bastante superior do primeiro (valores percentuais de [ɻ] já enunciados no item imediatamente acima; quanto aos de [ɾ], CaGII: 30%; CaGI: 12%; CbGII: 5%; e CbGI: 3%); e

(24) 77% das instâncias da configuração [ɻ]-[ɾ] tiveram como autores os falantes com alto nível escolar. Desse percentual, 60% foram construídos especificamente por indivíduos velhos (espalhados diatopicamente), e 40%, por indivíduos novos (concentrados na parte baixa — em relação ao rio Tietê — da porção da região do Médio Tietê em foco).

Veja-se que a concorrência a que o item (23) se refere concentrou-se mais acirradamente entre os informantes com alta escolaridade, principalmente entre aqueles mais velhos.

Isso dito, abordemos a seguir aspectos conclusivos a partir da correlação entre as covariantes surgidas para a E2 — *i.e.*, para a coda externa em sílaba tônica (que não de desinência infinitiva) seguida por pausa — e a série CxWz (ou seja, as dimensões diastrática e diagenérica):

(25) um rápido contraste entre a base documental da série CxWz[20] e a base documental da série CxGy[21] mostra que seus valores são muito próximos;

(26) a constatação acima implica que as variáveis diageracional e diagenérica não influíram significativamente na determinação dos resultados ligados à série CxWz (e à série CxGy);

(27) os entrevistados com nível escolar baixo responderam por 64% das ocorrências da configuração [ɻ], registradas difusamente ao longo da rede de pontos; e

(28) 70% das instâncias da configuração variacional [ɻ]-[ɾ] foram concebidos pela classe escolar alta, de maneira geograficamente dispersa.

Um corolário decorrente dos itens (25) e (26) é que o estado de coisas referido foi influenciado ou pela dimensão diastrática, ou pela dimensão diatópica, ou ainda por ambas. Para que se o conheça assertivamente, há que se investigar mais.

[20] V. a Tabela 14.
[21] V. a Tabela 13.

Para encerrarmos o elenco dos apontamentos sobre a E2, precisamos nos remeter à série WzGy:

(29) a covariante [ɻ] posicionou-se como norma relativa, e isso em todos os grupos de inquiridos, sobretudo entre os jovens masculinos, com quem o fone em tela obteve 95% de frequência relativa;

(30) houve apenas uma concorrência real, entre [ɻ] e [ɾ], e ela se deu mais propriamente entre as mulheres mais velhas, que produziram [ɻ] com 72% de frequência relativa e [ɾ] com 23%;

(31) os jovens contribuíram com 63% de todas as ocorrências da configuração [ɻ], com distribuição geográfica difusa; e

(32) a configuração [ɻ]-[ɾ] — manifestada, em geral, ao longo de toda a rede de pontos, salvo em Araçariguama (P3) e em Porto Feliz (P7) —, quando especificamente enunciada por jovens masculinos, surgiu somente ao lado direito do rio Tietê; e quando enunciada por jovens femininos, surgiu apenas ao lado esquerdo do rio referido.

A respeito do item (29), se recobrarmos os valores contidos na Tabela 15, veremos que o segundo grupo que demostrou maior predileção pela covariante [ɻ] é também de jovens, WfGI, que gerou o som em questão com 90% de frequência relativa.

Também, tragamos ao primeiro plano algumas das considerações mais relevantes relativamente aos resultados registrados a partir do cruzamento das contrapartes (a)fonéticas do /R/ na E3 (*i.e.*, em coda interna em sílaba tônica), a começar pela série CxGy:

(33) a covariante [ɻ] firmou-se como norma relativa em todos os grupos de falantes, máxime entre os indivíduos com baixa escolaridade, com maior

saliência por parte dos mais jovens, que enunciaram o fone em apreço com 98% de frequência relativa;

(34) a concorrência real entre covariantes só se verificou entre [ɻ] e [ɾ] na classe escolar alta. O grupo CaGII gerou a covariante [ɻ] com 68% de frequência relativa e a covariante [ɾ] com 28%, enquanto o grupo CaGI realizou o primeiro fone com 77% de frequência e o segundo com 23%;

(35) 65% de todos os quadrantes cartografados, ao longo de toda a rede de pontos, representaram a variação nula, sob a configuração [ɻ], de modo difuso diatopicamente;

(36) 92% das instâncias da configuração [ɻ]-[ɾ] foram promovidas por membros da classe escolar alta, por toda a rede de pontos, exceto em Pirapora do Bom Jesus (P2) e em Piracicaba (P10);

(37) quando a configuração mencionada no item acima, [ɻ]-[ɾ], foi perpetrada por jovens com nível escolar alto, ela se concentrou exclusivamente ao lado esquerdo do rio Tietê, abarcando todas as localidades desse lado, afora Santana de Parnaíba (P1); e

(38) ainda quanto a essa mesma configuração, [ɻ]-[ɾ], absolutamente nenhum de todos os informantes mais velhos com baixa escolaridade atualizou qualquer instância dela.

Quanto ao item (33), vê-se que a vitalidade observada da covariante [ɻ] foi notável, especialmente entre jovens. Já em relação ao item (34) — e também ainda em relação ao item (33) —, pode-se levantar a pergunta se a covariante [ɾ] na E3 (e em outras situações estruturais) no Médio Tietê estaria indubitavelmente em processo de regresso a favor da covariante [ɻ], mais ouvida da fala dos jovens. Se olharmos os dados aqui apresentados, eles indicarão um "sim" promissor.

A ver agora sinteticamente o que adveio de concreto do encontro dos resultados linguísticos do /R/ na E3 em perspectiva com a série CxWz:

(39) a covariante [ɻ] revelou-se norma relativa em três grupos de entrevistados (CaWm: 80; CbWf: 95%; e CbWm: 95%), exceto naquele constituído por mulheres com nível escolar alto, porque, embora com 64% de frequência relativa, a covariante em foco não logrou distribuição regular (ausente em Sorocaba, P5);

(40) ademais, as considerações antes feitas em relação aos resultados associados à série CxGy aplicam-se essencialmente àqueles associados à série CxWz;

(41) 65% das instâncias de variação nula representada pela configuração [ɻ] — configuração essa observada em geral ao longo de toda a rede de pontos — foram produzidos pela classe escolar baixa;

(42) todas as instâncias da configuração supramencionada, [ɻ], quando realizadas por mulheres com nível escolar alto, concentraram-se na parte baixa (em relação ao rio Tietê) do Médio Tietê;

(43) 82% das instâncias da configuração [ɻ]-[ɾ] — observada em geral em toda a rede de pontos — foram de autoria da classe escolar alta;

(44) a configuração acima indicada, [ɻ]-[ɾ], não foi, em nenhuma localidade, objeto de enunciação por falantes femininos com baixa escolaridade; e

(45) ainda sobre essa mesma configuração, [ɻ]-[ɾ], todas as instâncias de responsabilidade dos homens da classe escolar alta estiveram concentradas apenas ao lado esquerdo do rio Tietê, mais especificamente na parte baixa (em relação a esse rio) da porção da região do Médio Tietê sob enfoque.

A asserção presente no item (40) sugere que as variáveis extralinguísticas diageracional e diagenérica não exerceram papel influenciador relevante nos resultados da variação linguística constituída pelas diferentes covariantes fonéticas.

Sejam agora as observações resumidas sobre a correlação entre, de um lado, os resultados linguísticos referentes à E3 e, de outro lado, a série WzGy:

(46) a covariante [ɻ] alçou-se à posição de norma relativa em três grupos de inquiridos (WfGI, WmGII e WmGI), especialmente entre os jovens masculinos, com 93% de frequência relativa;

(47) uma concorrência real aconteceu somente entre [ɻ] e [ɾ], e mais significativamente apenas entre as mulheres mais novas (82% para [ɻ] e 18% para [ɾ]);

(48) todas as instâncias da configuração [ɻ]-[ɾ] — observada em geral em toda a rede de pontos, exceto em Pirapora do Bom Jesus (P2) —, quando ditas por mulheres jovens, concentraram-se apenas ao lado esquerdo do rio Tietê;

(49) todas as instâncias da configuração citada acima, [ɻ]-[ɾ], quando oportunizadas por mulheres mais velhas, situaram-se exclusivamente na parte alta (em relação ao rio Tietê) do Médio Tietê; e

(50) ainda a propósito da configuração [ɻ]-[ɾ], quando surgida através dos homens mais velhos, suas instâncias concentraram-se somente na parte baixa (em relação ao rio Tietê) do Médio Tietê.

Um rápido contraste entre os itens (46) e (47) faz ver que, enquanto os homens jovens foram os principais responsáveis pela produtividade da covariante [ɻ], as mulheres jovens foram as principais responsáveis pela única concorrência real verificada, entre [ɻ] e [ɾ].

Sendo a E4 a última situação estrutural observada a propósito das contrapartes (a)fonéticas do /R/, vejamos agora o que se extrai dela como essencial em perspectiva com a primeira das três séries, CxGy:

(51) a covariante [ɻ] estabeleceu-se como norma relativa em três grupos de informantes — CaGI, CbGII e CbGI —, com 78%, 93% e 94% de frequência relativa, respectivamente;

(52) a concorrência real deu-se mediante a variação formada pelas covariantes [ɻ] e [ɾ] somente entre os falantes com nível escolar alto, sobretudo entre os mais velhos ([ɻ]: 66%; e [ɾ]: 33%);

(53) 74% das instâncias da configuração [ɻ], representante de ausência de variação — e observada em geral em toda a rede de pontos, afora Sorocaba (P5) —, foram articuladas pela classe escolar baixa;

(54) todas as instâncias de tal configuração, [ɻ], quando enunciadas por inquiridos velhos com escolaridade alta, concentraram-se ao lado esquerdo do rio Tietê na parte baixa (em relação a esse mesmo rio) da porção da região do Médio Tietê sob enfoque; e

(55) 72% das instâncias da configuração [ɻ]-[ɾ] — registrada em geral ao longo de toda a rede de pontos — foram de autoria da classe escolar alta.

Também quanto à coda interna em sílaba átona, percebe-se que houve uma forte correlação entre a covariante [ɻ] e os jovens, a julgar pelo que o item (51) expressa. Outrossim, tendo-se em mente o item (52), mais uma vez vem à tona a importância dos indivíduos mais velhos na produção da covariante [ɾ].

Acerca da série CxWz:

(56) a covariante [ɻ] revelou-se norma relativa em três grupos de entrevistados (CaWm: 79%; CbWf: 91%; e CbWm e 96%);

(57) acrescentando-se às informações acima o valor percentual obtido pelo fone [ɻ] no grupo CaWf, 65%, observa-se que esses valores foram progressivos, com ponto de partida o grupo CaWf e de chegada o grupo CbWm;

(58) o principal grupo responsável pela concorrência real, entre [ɻ] e [ɾ], é o das mulheres com alta escolaridade ([ɻ]: 65%; e [ɾ]: 34%);

(59) 71% das instâncias da configuração [ɻ], representante de variação nula — e registrada em geral ao longo de toda a rede de pontos —, foram causados pela classe escolar baixa; e

(60) 76% das instâncias da configuração [ɻ]-[ɾ] — registrada em geral ao longo de toda a rede de pontos — foram contribuição dos entrevistados com nível escolar alto.

Atenção ao fato de que a propriedade 'nível escolar baixo' foi decisiva nos elevados valores da covariante [ɻ] em coda interna em sílaba átona, segundo se vê no item (56). Já no tocante ao item (58), se ainda incluirmos a esse respeito a informação segundo a qual a concorrência entre [ɻ] e [ɾ] também foi não desprezível no âmbito do grupo CaWm ([ɻ]: 79%; e [ɾ]: 21%), concluir-se-á que a vitalidade (ainda que relativamente baixa) da frequência do som [ɾ] na E4 deveu-se sobretudo à classe escolar alta.

Feitos alguns realces e conclusões até aqui, e restando apenas tratar da última série (WzGy) em relação à última especificidade (E4), avancemos, pois, por mais algumas linhas:

(61) a covariante [ɻ] apresentou-se como norma relativa nos grupos WfGI (85%), WmGII (88%) e WmGI (87%);

(62) a concorrência real constatada, entre [ɻ] e [ɾ], teve lugar em todos os quatro grupos de falantes, especialmente entre as mulheres mais velhas ([ɻ]: 71%; e [ɾ]: 25%); e

(63) as instâncias da configuração [ɻ]-[ɾ] perpetradas pelas mulheres jovens concentraram-se exclusivamente ao lado esquerdo do rio Tietê, em todas as localidades, salvo em Santana de Parnaíba (P1) e Araçariguama (P3).

Por ensejo do item (58), há pouco aventamos a conclusão segundo a qual a vitalidade da covariante [ɾ] em variação com a covariante [ɻ] deveu-se aos informantes com nível escolar alto. Agora, graças ao cruzamento de dados oportunizado pela série WzGy — e ao que o item (62) indica —, vê-se que, entre tais informantes, as mulheres mais velhas são as principais responsáveis pela vitalidade referida.

7 NOTAS

Já em direção ao fim do presente tomo, vamos compartilhar agora algumas notas, umas mais breves do que outras, que aludem à superfície mais apreensível de fenômenos complexos que nossa pesquisa colateralmente tocou (com exceção feita às observações sobre as mídias), os quais demandarão eventualmente pesquisas próprias futuras devotadas a partir de aportes teóricos específicos.

7.1 QSL^{m} em foco

Dos instrumentos metodológicos de coleta de dados originados pela lavra do ALiB (2014), o Questionário Semântico-Lexical (QSL) — considerado aqui em sua versão do ano de 2001 — revela-se o mais complexo. Algo importante inerente a essa condição é o fato de que sua dimensão semântica^{+} implica uma relação "língua/fala" *vs.* "mundo" ligada a combinações virtualmente infinitas de conteúdos, enquanto os outros componentes linguísticos — tais como o fonético^{+}-fonológico^{+}, o morfossintático^{+} e o prosódico^{+} — teoricamente podem ser, em perspectiva sincrônica^{+}, finitamente delimitados.

Isso posto, seguem-se algumas observações sobre:

(a) a identificação de processos psicossociais a serem evitados na elicitação^{+}: confusão, equívoco, precipitação, insegurança, dúvida, desconhecimento, esquecimento e incompreensão;

(b) a constatação, como um problema não insólito na pesquisa, de que mesmo enunciados constituídos por palavras simples/comuns/recorrentes são não raramente objeto de um dos processos psicossociais supracitados;

(c) uma hipótese explicativa para esse problema: trata-se de dificuldade de categorização mental a partir da oitiva dos enunciados, e esta seria uma questão anterior e mais elementar do que os processos psicossociais elencados no item (a); e

(d) a necessidade de uso, para o maior número possível de questões do QSL^{m}, de imagens que exibam uma representação dos referentes semânticos^{+} que os enunciados tematizam, como medida auxiliar de fácil adoção orientada à minimização dos problemas mencionados.

Sendo colocados à demanda de responderem as perguntas do inquiridor, os informantes são instados a reagirem a ela, ainda que com respostas precipitadas, inseguras ou duvidosas *etc.*, mais ou menos confiáveis. Para ilustrá-lo, alguns exemplos na Tabela 24 abaixo, extraídos do nosso *corpus* semântico-lexical^{+} primário, disponível na íntegra no Tomo IV.

Q	Enunciado	Lexia obtida
71	... o bicho que solta um cheiro ruim quando se sente ameaçado?	<tamanduá>
139	... quem deixa as contas penduradas e não paga?	<agiota>

... = *"Como se chama"*

Tabela 24: Exemplos de respostas ao QSL^{m}

Outra maneira de se reagir à demanda das perguntas é por meio de abstenções. As abstenções decorrem de esquecimento, desconhecimento consciente, de não compreensão da questão, e não de confusão ou equívoco, de cujo acontecimento tendencialmente o informante não se dá conta.

Na próxima página, a Tabela 25 faz uma correspondência entre processos psicossociais presumidos e seus efeitos possíveis. Os termos 'tomar' e 'escolha' referem-se à atribuição de sentido (intensional ou extensionalmente) a um enunciado (o tipo específico sob consideração é interrogativo).

	Processo	Efeito possível
(α)	confusão	tomar X por Y
(β)	equívoco	tomar *como certo* X por Y
(γ)	precipitação	tomar X *sem confirmação* da escolha de X
(δ)	insegurança	tomar X com *grau de certeza baixo*
(ε)	dúvida	tomar X com *grau de certeza baixo*, mas superior ao da insegurança
(ζ)	desconhecimento	em geral, *abstenção*
(η)	esquecimento	
(θ)	incompreensão consciente	

Tabela 25: Processos psicossociais e seus efeitos possíveis

Esses processos são exemplificados a seguir, exceto nos casos (ζ), (η) e (θ), em que se trata geralmente de simples abstenção, como já aludido na Tabela 25.

(α') *confusão:* <mãe seca> por <mãe/ama de leite> como resposta a "Quando a mãe não tem leite, e outra mulher amamenta a criança, como chamam essa mulher de quem a criança então mama?" (Q128);

(β') *equívoco:* <pernilongo> (Y podendo ser ou <libélula>, ou <lava-bunda>, ou <helicóptero> *etc.*) como resposta a "Como se chama o inseto de corpo comprido e fino com quatro asas bem transparentes que voa e bate a parte traseira na água?" (Q85);

(γ') *precipitação:* <pavão> como resposta (seguida ou não por retificação) realizada antes do término da enunciação da pergunta "Como se chama a ave de penas coloridas que, quando presa, pode 'aprender a falar'?" (Q68);

(δ') *insegurança:* enunciação hesitante sob entoação interrogativa de <vaca> como resposta a "Como se chama o boi sem chifres?" (Q78); e

(ε') *dúvida:* enunciação hesitante sob entoação afirmativa de <estrela d'alva> como resposta a "Como se chama a estrela que, à noite, se move e deixa um risco de luz?" (Q31).

A confirmação citada no item (γ) da Tabela 25 viria de um movimento mental (simultâneo ou imediatamente posterior à escolha de X) complementar de procura — por análise do cotexto e do contexto — a corroborar a escolha de X ou a descartá-la. E por incompreensão consciente no item (θ), entende-se a tendência de ausência total de entendimento da pergunta, a despeito do foco de atenção do sujeito estar concentrado na indagação.

Em todo caso, respostas como as exemplificadas eventualmente podem ser mantidas pelo inquirido após questionamentos de controle do inquiridor, mediante comentários do entrevistado de que tais respostas representariam mesmo uma significação local corrente.

Para darmos mais um exemplo, referimo-nos à pesquisa de campo de Soares (ASL-ReNAT, 2012:189) sob o objetivo de elaborar o *Atlas semântico-lexical da região norte do Alto Tietê*. Na ocasião, a pesquisadora deparou-se com dois informantes que se abstiveram de responder a Q65 — à época assim enunciada: "[Como se chama o passarinho bem pequeno]$_{S1}$ [que bate muito rápido as asas]$_{S2}$, [tem o bico comprido]$_{S3}$ e [fica parado no ar]$_{S4}$?" —, porque eles não a teriam compreendido. Caso, então, provavelmente do processo (θ).

Voltando-se o olhar a esse enunciado da Q65, parece-nos ponto plausível e pacífico que os seguintes elementos façam parte do repertório linguístico comum de indivíduos nativos vivendo no Brasil cuja língua materna seja a portuguesa brasileira, independentemente de seu grau de instrução/educação formal:

(1) as palavras tanto funcionais (pron. <se>, arts. <o> e <as>, rel. <que>, conj. <e>, prep.), quanto lexicais;

(2) as desinências verbais do pres. do ind. na 3ª pess. do sing. <-a> (sufixadas a <cham-> e <fic->), <-e> (<bat->) e -m (<te->); e

(3) a análise sintático-semântica, *i.e.*, o processo de significação das sentenças do enunciado (aqui, em especial, à competência linguístico-mental de recuperar a referência do rel. <que> e de preencher as categorias vazias anafóricas na cabeça das sentenças S3 e S4 com o suj. <o passarinho bem pequeno>).

Isso dito pressupondo-se uma prosódia e uma realização fonética não estranhas ao previsível nas variedades+ nacionais de fala, as quais não implicam, como se sabe, interincompreensão significativa entre si.

Evidentemente, esses pontos são relativizados sob certas condições, tais como aquelas associadas a um indivíduo que sofra de um grau considerado acentuado de esquecimento de informações a curtíssimo prazo (tendo-se em conta o comprimento do enunciado que ora estamos considerando). Sob essa condição, o falante pode ter dificuldades ao tentar recuperar um antecedente anafórico antes pronunciado, o que afetaria uma categorização cognitiva correspondente.

Nossa hipótese — a qual apenas aventamos aqui — é de que tais dificuldades são uma contraparte fenomênica de um problema na categorização mental, mesmo em casos nos quais se suponha como normalmente operante a capacidade de se atribuir significado aos elementos componentes do enunciado.

Entendida como um processo por meio do qual o indivíduo mentalmente situa entidades semelhantes em classes específicas após o processamento de *inputs* linguísticos, e cuja contraparte linguística revela-se como predicação ou predicações, a categorização, se falhar em algum estágio, deixará de subsidiar o indivíduo na tarefa de formar a ideia completa correspondente ao conteúdo semântico da questão que ouve.

Na Tabela 26, uma representação simplificada da categorização possível relativa ao referente semântico tematizado na questão em apreço (*i.e.*, "Como se chama o passarinho bem pequeno que bate muito rápido as asas, tem o bico comprido e fica parado no ar?") é a que segue:

Entidade	pássaro			
Classe	entidades que voam			
Propriedades	pequenino	com movimento veloz de asas	com bico comprido	pára no ar

Tabela 26: Categorização simplificada do que <beija-flor> denota

O processo de categorização conceitual do referente semântico iniciar-se-ia selecionando, do grupo de entidades que voam, uma que seja bem pequena, que movimente rapidamente as asas, possua bico comprido e tenha a capacidade de parar no ar. Como sentido, o indivíduo é suposto a chegar à conclusão de que, segundo seu mundo conceitual (formado

por percepções externas e internas), se trata da entidade designada, entre outras variantes+, pela lexia+ <beija-flor>.

Os informantes mencionados de Soares (ASL-ReNAT, 2012) teriam falhado nessa categorização. Se ambos não tivessem falhado, e então tivessem compreendido a questão, mas se não tivessem tido uma representação conceitual do referente semântico a partir de sua experiência de mundo — o que implicaria, portanto, um desconhecimento —, não faria sentido, embora não fosse impossível, falar de incompreensão.

Isso posto, e considerando que:

(i) o QSL 2001 do ALiB já faz recomendações ao inquiridor para realizar ostensão pragmática+ (para 15 questões), mímica (para 18) e apresentação de imagens (4) para facilitar a compreensão de questões; e

(ii) o procedimento que consiste em se dizer apenas "Como se chama isto?", acompanhado pela respectiva ostensão pragmática, já é adotado a propósito das questões 104, 118, 119, 186 e 187 do QSL;

é, então, de se fomentar — sob o intuito de evitarem-se ao máximo problemas de categorização e de processos psicossociais tais como os explicitados — a utilização de imagens para o maior número possível de questões que ilustrem claramente o referente semântico tematizado.

7.2 Três tipos de informantes

Pelo que observamos no campo de investigação, houve a manifestação de três tipos gerais de entrevistados. Como critério para a tipologização, levamos em conta os idioletos+ dos participantes da pesquisa. A ideia de base pressupõe um *spectrum* cujos polos extremos são representados, de um lado, pelo dialeto+ caipira+ rural e, de outro lado, pela variedade *standard* da língua portuguesa. Assim, os tipos são como seguem na página seguinte.

(a) o caipira;
(b) o transicional; e
(c) o estandardizado.

(a) Aqui, entende-se o caipira como aquele entrevistado cujo idioleto é a atualização virtualmente apenas do dialeto caipira. (b) O transicional é o indivíduo que alternadamente atualiza elementos linguísticos tanto do dialeto caipira quanto de outras variedades linguísticas, especialmente das estandardizantes. (c) Por fim, o estandardizado é o falante do interior que virtualmente já não conta com um idioleto caipira, mas sim com algum outro (ou alguns outros) que seja a atualização de uma variedade linguística que goze do atributo estandardizante, tal como uma variedade da capital ou alguma que seja padronizadamente ensinada em escolas com o objetivo de neutralizar marcas dialetais locais.

7.3 Táticas responsivas ao QSL^{m}

Ao longo da investigação empírica, alguns expedientes responsivos perpetrados (in)conscientemente pelos inquiridos tomavam lugar. Essas táticas foram observadas mais de uma vez no decurso da aplicação do QSL^{m}, as quais dividimos em (1) gerais, por um lado, e em (2) específicas, por outro. Muitas dentre estas últimas, da classe (2), foram postas em uso aparentemente quando faltava ao indivíduo denotações específicas. Ambos os grupos de táticas seguem arrolados com exemplos extraídos do nosso *corpus* semântico-lexical primário, disponível no Tomo IV.

(1) Gerais:

(1.1) retificação retroativa. *E.g.* <ônibus> como resposta a "Como se chama a condução que leva mais ou menos quarenta passageiros e faz o percurso dentro da cidade?" (Q200) substituído por <(ônibus) circular> após audição da pergunta seguinte, *i.e.*, "E quando o percurso é entre cidades, como se chama?" (Q201);

(1.2) mescla entre lexias. *E.g.* <bitoco> — que, hipoteticamente, pode ser o resultado da fusão entre <bituca> e <toco>, duas formas comuns na região[1] — como resposta a "Como se chama o resto do cigarro que se joga fora?" (Q146); e

(1.3) metáfora. *E.g.* <trono do rei> como resposta a "Como se chama isto? {*Mostrar imagem*} [Quando se vai ao banheiro, onde a pessoa se senta para fazer as necessidades?]" (Q170).

(2) Específicas:

(2.1) "diminutivação" de designação já presente na questão. *E.g.* <riozinho> como resposta a "Como se chama um rio pequeno, estreito, de uns dois metros de largura?" (Q1);

(2.2) "aumentativação" de designação já presente na questão. *E.g.* <chuvão> como resposta a "Como se chama uma chuva de pouca duração, mas muito forte e pesada?" (Q13);

(2.3) perífrase. *E.g.* <o sol aparece> como resposta a "O que acontece no céu de manhã cedo quando começa a clarear?" (Q23);

(2.4) eufemismo. *E.g.* <lento> como resposta a "Como se chama quem tem dificuldade de aprender as coisas? (Q137);

(2.5) hiperonímia. *E.g.* <autônomo> como resposta a "Como se chama o homem que é contratado para trabalhar na roça de outro e recebe por dia de trabalho?" (Q61);

(2.6) metonímia. *E.g.* <ramela> como resposta a "Como se chama a inflamação no olho que faz com que ele fique vermelho e amanheça grudado?" (Q95);

[1] Cf. seç. "Cabeças de relevância", no Tomo V.

(2.7) analogia. *E.g.* <triângulo> como resposta a "Como se chama a armação de madeira que se coloca no pescoço de animais (porco, terneiro/bezerro, carneiro, vaca) para não atravessarem a cerca?" {*Mostrar imagem*} (Q54);

(2.8) palpite sob entoação interrogativa. *E.g.* "cruzeiro do sul?" como resposta a "Como se chama este conjunto ou faixa de estrelas, bem juntas umas às outras, que fica no céu de fora a fora?" (Q33);

(2.9) resposta evasiva não seguida por resposta assertiva. *E.g.* "olha, isso aí é uma coisa bonita da natureza" como resposta a "Como se chama a estrela que aparece de manhã cedo, brilha mais do que as outras e é a última a desaparecer?" (Q29); e

(2.10) repetição de:

(2.10.1) designação já presente na questão. *E.g.* <mosca> como resposta a "Como se chama um tipo de mosca grande e esverdeada que faz um barulhão quando voa?" (Q83);

(2.10.2) uma mesma resposta a perguntas que tematizam referentes semânticos...

(2.10.2.1) ... distintos. *E.g.* <pôr do sol> como resposta tanto a "Como se chama a parte do dia quando começa a clarear?" (Q22), quanto a "O que acontece no céu de manhã cedo quando começa a clarear?" (Q23); e

(2.10.2.2) ... semelhantes. *E.g.* <trilha> como resposta tanto a "Como se chama o caminho feito num mato fechado, aberto com facão ou foice?" (Q62), quanto a "Como se chama o caminho no pasto onde não cresce mais grama, de tanto animais ou homens passarem por ali?" (Q63).

7.4 Mídias

Para encerrar o turno das nossas notas complementares de pesquisa, resta apenas compartilharmos algumas observações sobre a relação dos indivíduos entrevistados com as mídias tradicionais (TV, jornal/revista e rádio).

Contemporaneamente, as mídias fazem cada vez mais parte da vida das pessoas e as influenciam ativamente, sobretudo no âmbito do repertório linguístico passivo dos indivíduos, mas também na esfera de seu repertório ativo.

Isso como premissa em plano de fundo, vejamos a seguir como a presença das mídias na vida dos nossos informantes quantitativamente se dá de uma perspectiva global.

Os dados foram levantados com a Ficha do Informante[m], disponibilizada entre os apêndices, no Tomo VI, e a matriz de dados utilizada para a quantificação encontra-se distribuída nas seções "Exposição às mídias escrita e sonora" e "Exposição à mídia televisiva e a eventos esportivos e culturais", no mesmo tomo referido.

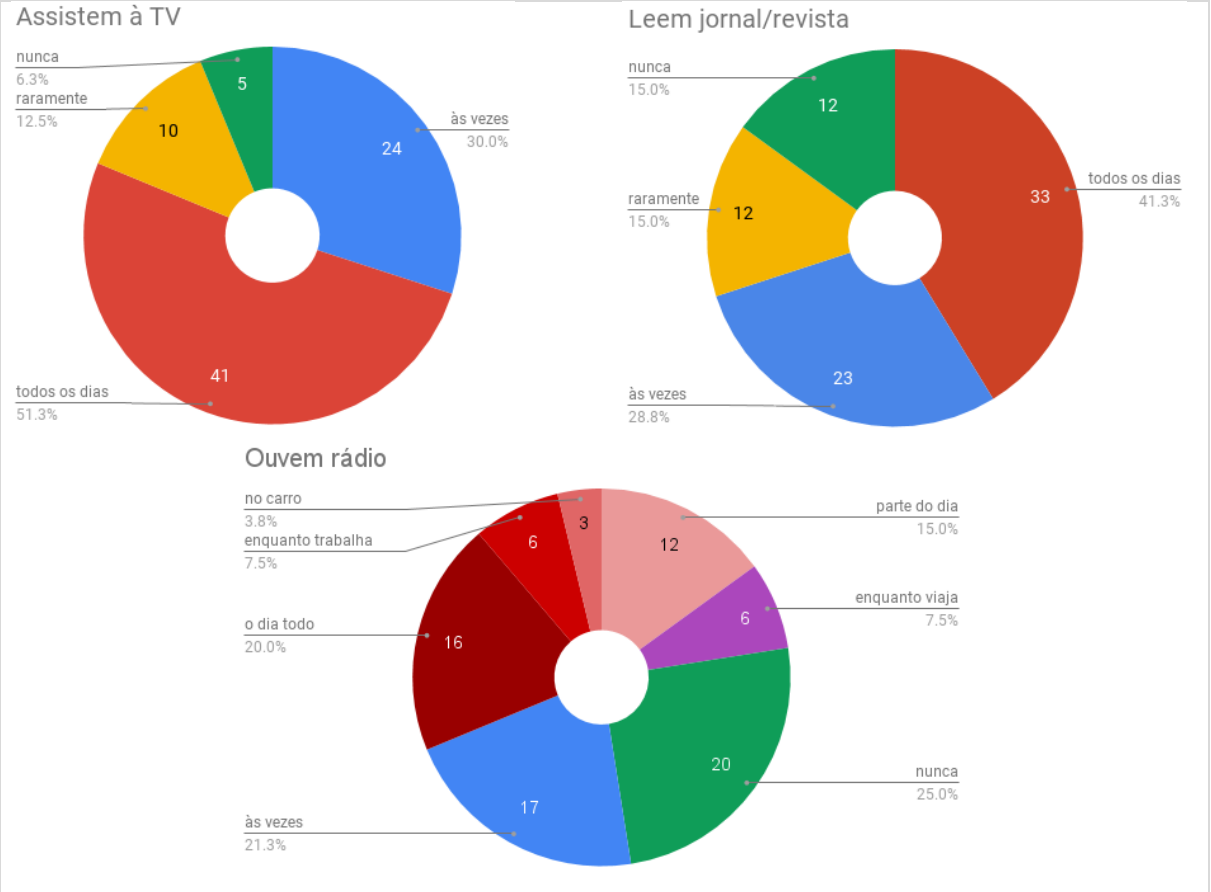


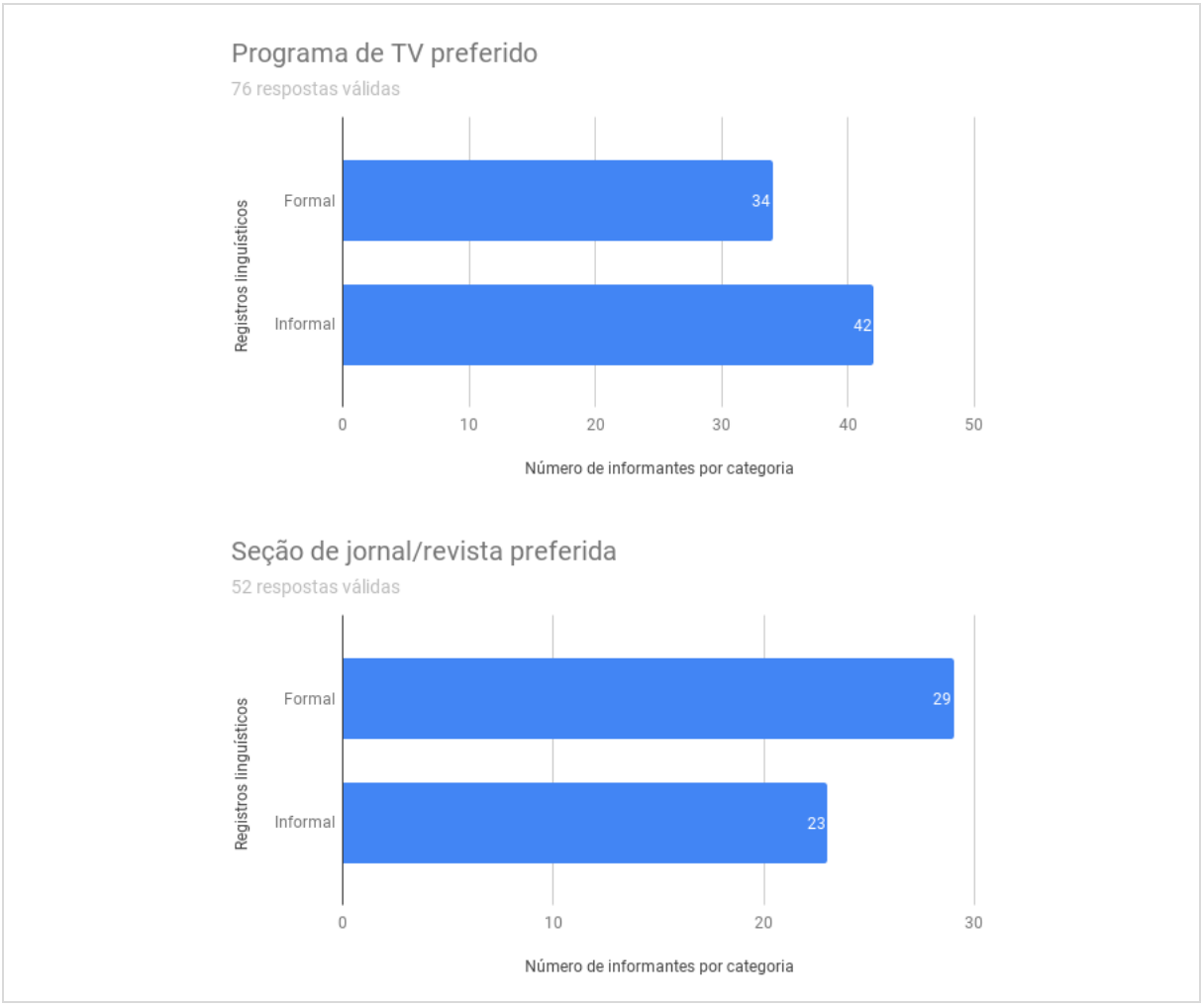
Gráfico 1: Presença das mídias na vida dos 80 informantes[2]

[2] O arredondamento para uma casa decimal teve como efeito o total de 100,1% em cada um dos três casos no gráfico.

Diariamente, 51,3% dos indivíduos veem televisão, 41,3% leem e 46,3% ouvem rádio (este último incluindo ambas as opções "enquanto trabalha" e "no carro" juntamente com "o dia todo" e "parte do dia").

Um número significativamente alto dos entrevistados está exposto a manifestações verbais de tendência supradialetal, o que, por sua vez, pode ligar-se a processos de convergência gradual entre variedades (fenômeno conhecido como *dialect levelling*), pelos quais as variedades mais tradicionais tendem a desaparecer.

Baseando-se nas opções midiáticas preferidas dos inquiridos, o gráfico adiante agrupa os dados válidos nas categorias de registro formal e registro informal.



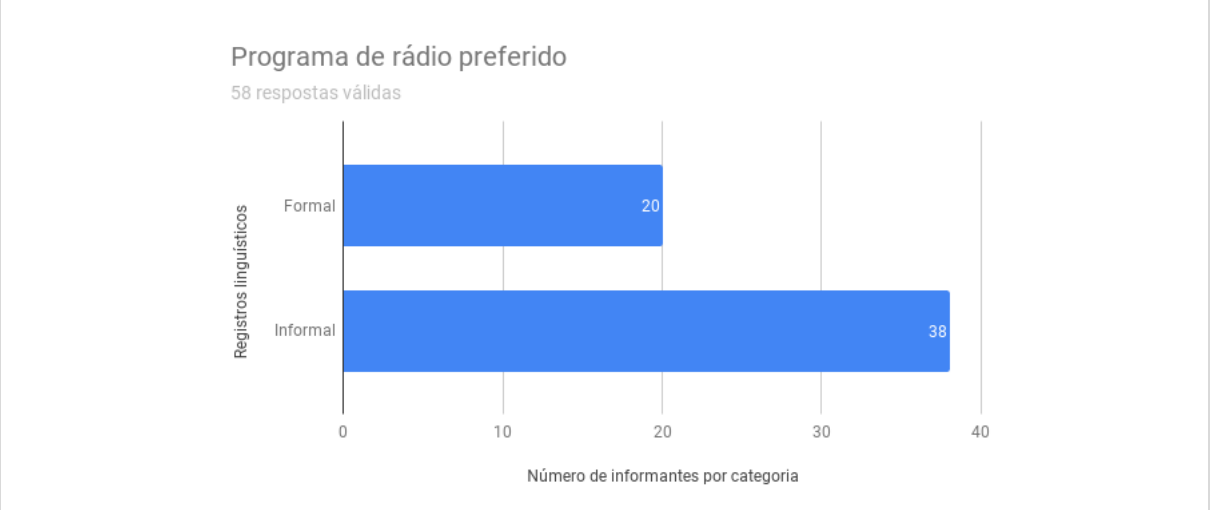


Gráfico 2: Formalidade e informalidade das mídias presentes na vida dos informantes

Para o Gráfico 2, o agrupamento das respostas se deu segundo as classificações que a Tabela 27 abaixo apresenta.

Seção de jornal/revista preferida:		Programa de rádio preferido	
F	I	F	I
edital	humorístico	noticiário geral	música
ciência	esporte	política	esporte
cultura	fofoca	religioso	entretenimento
tecnologia	entretenimento	-	variedades
economia	social	Programa de TV preferido	
profissional	carros antigos	F	I
artigo	cotidiano	noticiário	novela
página policial	infantil	religioso	auditório
editorial	variedades	tecnologia	esporte
política	-	cultura	culinária
entrevista	-	documentário	decoração
-	-	*talk show*	série
-	-	-	filme

F = formal ● I = informal

Tabela 27: Classificações das opções midiáticas preferidas

Considerando-se os três meios de comunicação de massa em tela, o registro informal é superior em relação à TV e ao rádio, mas inferior quanto à leitura. Em todo caso, o índice de formalidade não deixa de ser expressivo até mesmo no tocante à TV e ao rádio. São 45% de preferência na TV, 56% na leitura e 34% no rádio.