

Variação semântico-lexical em Santana de Parnaíba: amostra dialetal contrastiva

Semantic-lexical variation in Santana de Parnaíba:
a contrastive dialectal sample

DOI: <https://doi.org/10.24206/lh.v5i2.27079>

Selmo Ribeiro Figueiredo Junior

É pós-doutorando na Universidade de São Paulo. Atua nos seguintes temas: Dialetoлогия/Geolinguística, Língua Portuguesa, Língua Alemã e Tradução.

E-mail: selmojunior@usp.br

ORCID: <https://orcid.org/0000-0001-8367-0306>

Manoel Mourivaldo Santiago-Almeida

É professor titular e pesquisador da Universidade de São Paulo. Atua nos seguintes temas: Linguística Histórica, Crítica Textual e Filologia, Dialetoлогия e História da Língua Portuguesa.

E-mail: msantiago@usp.br

ORCID: <http://orcid.org/0000-0003-0680-1151>

RESUMO

Este artigo traz um recorte semântico-lexical do projeto de atlas linguístico da região do Médio Tietê, no Estado de São Paulo, Brasil. Dentre as dez localidades que perfazem a rede de pontos de investigação, apresentamos aqui dados do município de Santana de Parnaíba (SdP). Trata-se de dez questões onomasiológicas respondidas por informantes de perfis sociais distintos. As perguntas foram extraídas do questionário semântico-lexical do *Atlas linguístico do Brasil* (CARDOSO, 2014) e aplicadas após modificações. Elas tematizam características pessoais, fenômenos e entes no mundo físico, cujas respostas mais frequentes em SdP foram RAIZ, URUBU, JOÃO-DE-BARRO, VAREJEIRA, TERÇOL, CATARRO, TAGARELA, MÃO-DE-VACA, BALANÇO e INTERRUPTOR. Para a eliciação, aplicou-se a *técnica de entrevista em três passos*, que basicamente consiste em (i) perguntar, (ii) insistir e (iii) sugerir (THUN, 2000). O objetivo do presente texto é (a) sistematicamente expor e brevemente comentar as (primeiras) respostas espontâneas dos parnaibanos. Adicionalmente, apresentamos: (b) três critérios de invalidação de respostas; (c) correlações internas linguístico-sociais; (d) contrastes das normas lexicais e formas mais frequentes apuradas em SdP com resultados de seis atlas linguísticos fora do perímetro da região do Médio Tietê; e (e) observações linguísticas complementares acerca das lexias. No Brasil, a pesquisa é financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), no âmbito do Projeto para a História do Português Paulista (PHPP), e, na Alemanha, pelo Serviço Alemão de Intercâmbio Acadêmico (DAAD).

Palavras-chave: Médio Tietê. Dialeto Caipira. Dialectologia. Sincronia. Comparabilidade.

ABSTRACT

A semantic-lexical excerpt from our project of a linguistic atlas for the region of Médio Tietê, São Paulo state (Brazil), is here presented. The research field is formed by ten cities, which one defined as being a point of inquiry. Santana de Parnaíba (SdP), one of them, is taken into account for this paper. Specifically, spontaneous answers given by local residents with different social profiles to ten onomasiological questions. The questions were taken with modifications from the semantic-lexical questionnaire of the Linguistic Atlas of Brazil (CARDOSO, 2014). They are related to personal characteristics, as well as physical entities and phenomena, whose more frequent answers in SdP were

RAIZ, URUBU, JOÃO-DE-BARRO, VAREJEIRA, TERÇOL, CATARRO, TAGARELA, MÃO-DE-VACA, BALANÇO e INTERRUPTOR. For the elicitation, the *three-step interview technique* was employed, consisting basically in (i) asking; (ii) insisting; and (iii) suggesting (THUN, 2000). This paper aims to (a) expose systematically and offer brief comments on the (first) spontaneous answers of our SdP participants. Additionally, we (b) present three criteria for invalidating answers; (c) establish internal sociolinguistic correlations; (d) report a contrast between, on the one hand, lexical items that occurred as the most frequent ones or rather as norms in SdP, and, on the other hand, results of six other Brazilian linguistic atlases outside of Médio Tietê region; (e) and, finally, provide complementary linguistic observations on certain lexical occurrences. Within the framework of the Project for the History of Portuguese in São Paulo state (PHPP), this research receives funding from the São Paulo Research Foundation (FAPESP); in Germany, the German Academic Exchange Service (DAAD) also kindly finances this investigation.

Keywords: Médio Tietê region. 'Dialeto Caipira'. Dialectology. Synchrony. Comparability.

Introdução

Neste artigo, o propósito é apresentar um recorte da pesquisa dialetológica em curso que objetiva a elaboração do atlas linguístico do Médio Tietê, Estado de São Paulo, Brasil.

A escolha dessa área enquanto campo de investigação se baseia no fato de que ela se constitui em um dos primeiros rincões do país que os portugueses, desde os anos 1500, colonizaram. Por essa razão, a região em questão acabou se tornando um dos primeiros centros de difusão da língua portuguesa no Brasil. Isso em vista, muitos linguistas estão se interessando inclusive em comparar estados mais modernos da língua com aqueles cujas cartas, diários e outros documentos antigos dão notícia, muitas vezes até evidenciando marcas orais de tempos longínquos.

Em nosso trabalho, procuramos contribuir com os estudos sobre uma das variedades do português falado no interior do Estado de São Paulo, ou *dialeto caipira*, investigada desde Amaral (1920) até pesquisadores contemporâneos, como Pires (2008); Leopoldino (2009); Brito (2013); Romero (2008); Garcia (2009); Leal (2006, 2012); Almeida (2007); Mota (2007); Levado (2009); entre outros.

Geograficamente, limitamo-nos ao que identificamos como Médio Tietê. Os municípios que perfazem nossa rede de pontos de investigação oral são dez e seguem nomeados juntamente com a data de surgimento enquanto povoados: Tietê (± 1500), Santana de Parnaíba (1561), Araçariguama (1590), Itu (1610), Sorocaba (1654), São Roque (± 1665), Porto Feliz (1721), Pirapora do Bom Jesus (1725), Capivari (± 1760) e Piracicaba (1766) (v. apêndice 1).

Selecionamos aleatoriamente oito informantes para cada ponto da rede, quatro mulheres e quatro homens, que pudessem, no entanto, satisfazer quatro critérios. Eles precisavam necessariamente de ter:

- I. entre 18 e 36 anos (GI) ou mais de 55 (GII);
- II. baixa escolaridade (Cb; no raio de abrangência entre analfabetismo e ensino médio incompleto) ou alta (Ca; a partir de estudos universitários, completos ou incompletos);
- III. moradia na localidade por no mínimo três quartos da vida;
- IV. residência ininterrupta na localidade ao longo dos últimos cinco anos.

Entre as quatro mulheres, tínhamos uma da CaGII, uma da CaGI, uma da CbGII e outra da CbGII. O mesmo em relação aos quatro homens.

Para a coleta de dados, utilizamos os instrumentos do *Atlas linguístico do Brasil* (ALIB, 2014) com modificações. A comparabilidade que eles permitem entre estudos foi um aspecto decisivo para

essa escolha. Eles são nomeadamente oito: (i) questões para discursos semidirigidos; (ii) questionário semântico-lexical (doravante, QSL); (iii) questionário fonético-fonológico; (iv) questionário morfossintático; (v) questões de pragmática; (vi) texto para leitura; (vii) perguntas metalinguísticas; e (viii) questões de prosódia.

Em relação ao QSL, procuramos aplicar o procedimento de inquérito desenvolvido por Thun (2000), denominado *técnica de entrevista em três passos*, no âmbito da Dialetologia Relacional e Pluridimensional. Essa técnica consiste em (1) perguntar, (2) insistir e (3) sugerir. Na fase 1, procura-se obter respostas espontâneas dos entrevistados. No momento 2, o informante é estimulado a fornecer variantes lexicais coexistentes. No passo 3, apresenta-se a ele uma relação de variantes não mencionadas antes, sob a intenção de se saber se elas são conhecidas passivamente ou não¹.

A visita a nossa rede de pontos se deu em duas fases: a primeira entre os dias 15 de fevereiro e 15 de março de 2016, envolvendo cinco das dez cidades, e a segunda etapa entre os dias 23 de agosto e 17 de setembro de 2017, nos outros cinco pontos de inquérito. Cada município foi definido como sendo um ponto da rede de nossa pesquisa.

No Brasil, a pesquisa é financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), no âmbito do Projeto para a História do Português Paulista (PHPP), e pelo Serviço Alemão de Intercâmbio Acadêmico (DAAD) na Alemanha.

Nas seções a seguir, informaremos e discutiremos dados empíricos semântico-lexicais produzidos entre os dias 15 e 19 de fevereiro de 2016 no município de Santana de Parnaíba (doravante, SdP), localidade onde a pesquisa de campo de modo geral teve início. Nomeadamente, o objetivo do presente texto é (a) sistematicamente expor e brevemente comentar as (primeiras) respostas espontâneas dos parnaibanos a dez questões onomasiológicas. Adicionalmente, apresentaremos: (b) três critérios de invalidação de respostas; (c) correlações internas linguístico-sociais; (d) contrastes das normas lexicais e formas mais frequentes apuradas na localidade com resultados de seis atlas linguísticos fora do perímetro da região do Médio Tietê; e (e) observações linguísticas complementares acerca das lexias.

Os itens *a-e* foram assim distribuídos nas seções seguintes:

1. *Apresentação das questões (Qs) e dos critérios de invalidação de respostas espontâneas (REs)*
2. *Amostra semântico-lexical*

¹ A aplicação dos passos 2 e 3 da técnica não foi sistemática em nosso caso. Ela só o foi quando o informante dispunha de mais de duas horas para a entrevista e não demonstrava impaciência, indisposição etc. no decorrer do inquérito. O prazo concedido pela agência financiadora não permitia trocas constantes de informantes em campo no intuito de se encontrarem os estritamente ideais.

3. *Correlações internas linguístico-sociais*
4. *Contrastes das normas e formas mais frequentes em SdP com resultados externos*
5. *Observações esparsas complementares*

Antes de irmos a elas, arrolamos na tabela 1 as siglas, abreviações e acrônimos que utilizaremos com frequência ao longo do artigo a partir daqui.

Tabela 1 – Siglas, abreviações e acrônimos em ordem alfabética.

ALAP	<i>Atlas linguístico do Amapá</i>
ALERS	<i>Atlas linguístico-etnográfico da Região Sul do Brasil</i>
ALiB	<i>Atlas linguístico do Brasil</i>
ALPR I	<i>Atlas linguístico do Paraná I</i>
ALPR II	<i>Atlas linguístico do Paraná II</i>
ASL-ABC	<i>Atlas semântico-lexical da região do Grande ABC</i>
ASL-GO	<i>Atlas semântico-lexical do Estado de Goiás</i>
ASL-LN-SP	<i>Atlas semântico-lexical de Caraguatatuba, Ilhabela, São Sebastião e Ubatuba - municípios do litoral norte de São Paulo</i>
Ca, Cb	classe de alta escolaridade, classe de baixa escolaridade
F	feminino
GII, GI	grupo etário mais velho, grupo etário mais jovem
Inf _x	informante. Por exemplo, 'Inf1' para 'informante número 1'
M	masculino
P _x	pode ser 'P1', 'P2' ou 'P3' para 'passo 1', 'passo 2' ou 'passo 3' da <i>técnica de entrevista em três passos</i>
Q _x	refere-se a questões do QSL
RE(<i>x</i>)	RE está para 'resposta espontânea' (para os casos em que só houve uma por entrevistado), da mesma forma que RE1 está para 'primeira resposta espontânea' e RE2 para 'segunda resposta espontânea' por inquerido
SdP	Santana de Parnaíba

Fonte: Elaboração própria.

1. Apresentação das Qs e dos critérios de invalidação de REs

As questões que vamos abordar envolvem seis áreas semânticas: (1) atividades agropastoris (Q47); (2) fauna (Q64, Q66 e Q83); (3) corpo humano (Q94 e Q102); (4) convívio e comportamento social (Q136 e Q138); (5) jogos e diversões infantis (Q166); e (6) habitação (Q175).

A tabela 2 traz os enunciados das perguntas acompanhadas pelas sugestões, utilizadas por ensejo do P3. O asterisco indica uso de imagem para a respectiva pergunta, para efeito ilustrativo do referente semântico tematizado, e as reticências simbolizam a enunciação de "Como se chama(m)".

Tabela 2 – Questões.

	Enunciado	Sugestões
Q47	Depois que se corta o pé de arroz ou de fumo, ainda fica uma pequena parte no chão. Como se chama essa parte?	soca, touceira, cepo
Q64	*... a ave preta que come animal morto, podre?	urubu, corvo
Q66	*... a ave que faz a casa com terra, nos postes, nas árvores e até nos cantos da casa?	joão-de-barro
Q83	*... um tipo de mosca grande, esverdeada, que faz um barulhão quando voa?	varejeira, beronha, bironha, mutuca, butuca
Q94	*... a bolinha que nasce n[a pálpebra], fica vermelha e incha?	terçol, viúva, três sol, viuvinha
Q102	... a sujeirinha dura que se tira do nariz com o dedo?	meleca, tatu, caca, catota, cera, ceroto, ranho
Q136	... quem fala demais?	tagarela, falador, papagaio, matraca
Q138	... quem não gosta de gastar dinheiro e, às vezes, até passa dificuldades para não gastar?	sovina, mão-de-vaca, pão-duro, miserável, muquirana, murruga, mão fechada
Q166	* ... uma tábua, pendurada por meio de cordas, onde uma criança se senta e se move para frente e para trás?	balanço
Q175	* ... o objeto que fica na parede e serve para acender a lâmpada?	interruptor (de luz)

Fonte: Elaboração própria.

É preciso introduzir também os dois critérios de invalidação de REs que estipulamos e aplicamos nos casos isolados (ocorrências únicas) de:

- A. clara discrepância entre resposta dada e referente tematizado pela questão;
- B. repetição na resposta de lexia ou expressão já presente na questão enunciada;
- C. designação vaga ou falta de termo específico.

Exemplos:

- A. "Trigo" como resposta a "Depois que se corta o pé de arroz ou de fumo, ainda fica uma pequena parte no chão. Como se chama essa parte?";
- B. "Rio" como resposta a "Como se chama um rio pequeno, de uns dois metros de largura?";
- C. "Deficiente visual" como resposta a "Como se chama quem só enxerga com um olho?".

Isso se relaciona ao P1. Nos P2 e P3, era comum que os mesmos informantes trouxessem respostas (no P2) ou confirmassem conhecimento passivo de lexias (no P3) fora do escopo dos critérios A, B e C.

2. Amostra semântico-lexical

As primeiras respostas espontâneas obtidas em SdP são exibidas nas tabelas 3a-b. O hífen (-) representa as abstenções realizadas por entrevistados no P1. As respostas invalidadas pelo critério A estão marcadas com um asterisco (*); pelo critério B, com dois (**). Entre chaves ({x}) destacamos as ocorrências lexicais que qualitativamente chamaram à atenção.

Tabela 3a – Respostas espontâneas (parte 1 de 2).

Inf	Q47	Q64	Q66	Q83	Q94
1	raiz	urubu	joão-de-barro	{mamangaba}	terçol
2	raiz	urubu	joão-de-barro	varejeira	terçol
3	-	urubu	joão-de-barro	varejeira	terçol
4	*	abutre	joão-de-barro	varejeira	{treçol}
5	-	{corvo}	joão-de-barro	butuca	três sol
6	-	urubu	joão-de-barro	**	três sol

7	-	urubu	{baitixó}	-	terçol
8	talo	urubu	joão-de-barro	varejeira	{cisto}

- = *abstenção*; * = *invalidação pelo critério A*; ** = *invalidação pelo critério B*; {x} = *variante inusitada*

Fonte: Elaboração própria.

Tabela 3b – Respostas espontâneas (parte 2 de 2).

Inf	Q102	Q136	Q138	Q166	Q175
1	cera	faladeiro	sovina	balanço	{tomada}
2	catarro	falante	mão-de-vaca	balança	*
3	ranho	tagarela	avarento	balanço	interruptor
4	{cacota}	falador	pão-duro	balanço	interruptor
5	-	-	miserável	{puleu}	-
6	caca	papagaio	mão-de-vaca	balanço	acendedor
7	catota	tagarela	pão-duro	balanço	-
8	catarro	{talarico}	{miquinha}	balança	interruptor

- = *abstenção*; * = *invalidação pelo critério A*; ** = *invalidação pelo critério B*; {x} = *variante inusitada*

Fonte: Elaboração própria.

As normas e formas mais frequentes são apresentadas na tabela 4, com os valores de frequência relativa (FR) e frequência absoluta (FA) em relação às respostas válidas. As variantes que se estabeleceram em norma lexical, i.e., com mais de cinquenta por cento de frequência, estão grafadas em itálico.

Tabela 4 – Normas e lexias mais frequentes.

Q	Lexia	FR (%)	FA	Q	Lexia	FR (%)	FA
47	<i>raiz</i>	66,66	2/3	102	catarro	28,57	2/7
64	<i>urubu</i>	75,00	6/8	136	tagarela	28,57	2/7
66	<i>joão-de-barro</i>	87,50	7/8	138	mão-de-vaca, pão-duro	25,00 (cada)	2/8

83	<i>varejeira</i>	66,66	4/6	166	<i>balanço</i>	62,50	5/8
94	<i>terçol</i>	50,00	4/8	175	<i>interruptor</i>	57,14	3/5

FR = frequência relativa; FA = frequência absoluta; as normas estão grafadas em itálico

Fonte: Elaboração própria.

Q47

RAIZ é a norma, seguida por TALO. No P3, em que a forma SOCA foi sugerida, o Inf3 foi o único sujeito a dizer SOQUEIRA. Também foi ouvido o item TRIGO como resposta, que foi, no entanto, invalidada pelo critério A, embora ela tenha sido dita com segurança pelo entrevistado após perguntas de controle (para certificação) do inquiridor. Quanto à designação RAIZ, o critério C de invalidação poderia ser aplicado, o que não fizemos por não se tratar de ocorrência única. Ademais, chama à atenção o fato de que metade das respostas são abstenções.

Q64

Apurou-se que URUBU é a norma local. Adicionalmente, houve ocorrência de ABUTRE e CORVO. Embora o Inf4 tenha realizado ABUTRE como RE, ele disse que URUBU seria a forma mais comum quando estávamos no P2. A variante ABUTRE foi mencionada, no passo da insistência, também pelo Inf3, que havia respondido URUBU no P1. Os Inf1 e Inf2, que haviam dito URUBU como RE1, proferiram CORVO como RE2. Isso indica a vitalidade da variante CORVO em SdP na condição de segunda forma mais comum.

Q66

JOÃO-DE-BARRO é categoricamente a norma lexical. O Inf7 é o único indivíduo que proferiu uma RE diferente, BAITIXÓ. Indagado a respeito do lugar onde ele teria ouvido essa forma, ele disse que a escutou na Bahia. Quando estávamos no P3, no entanto, ele mencionou que JOÃO-DE-BARRO seria a designação mais comum na cidade. Em todo caso, os dados apontam que a variação a respeito dessa questão não é produtiva.

Q83

Como norma, a ocorrência de VAREJEIRA predominou. As outras lexias que ocorreram são BUTUCA e MAMANGABA. O Inf7, sem RE, disse no P3, todavia, que conhece VAREJEIRA (conhecimento passivo). Foi registrada ainda a resposta MOSCA, que foi invalidada pelo critério B.

Q94

O item TERÇOL foi o mais produtivo, seguido por TRÊS SOL. A forma TREÇOL, dita pelo Inf4, poderia ser subsumida à realização TERÇOL (se se considerar caso de metátese) ou à variante TRÊS SOL (como caso de síncope). Quanto à forma CISTO, utilizada uma única vez, consideramo-la de estatuto duvidoso. Outra alternativa seria invalidá-la pelo critério C, baseando-se na definição de CISTO como uma espécie de tumor ou bexiga (cf. FERREIRA, 1975, p. 332 e p. 1177). Isso evocaria a noção *umbrella term*, ou melhor, uma expressão que pode referir-se a diferentes tipos de coisas. No entanto, um *umbrella term* pode tornar-se um termo específico em determinados contextos.

Q102

CATARRO é a realização mais frequente. Cada uma das outras REs consiste em uma variante distinta: CERA, RANHO, CACA, CATOTA e CACOTA. Esta última, dita pelo Inf4, poderia ser interpretada como a fusão entre CACA e CATOTA. O Inf5 confirmou no P3 o conhecimento passivo de CERA e CEROTO. O Inf7, embora tenha enunciado CATOTA como RE1, disse que RAMELA, proferida como RE2, seria a variante local mais comum; afirmação que, todavia, não encontra amparo na nossa base empírica.

Q136

Como sendo a variante mais frequente, temos TAGARELA. As demais apareceram apenas uma vez cada, o que torna a variação a essa questão bastante intensa: FALADEIRO, FALANTE, FALADOR, PAPAGAIO e TALARICO.

Q138

Grande variação se encontra também entre as REs a essa questão. Duas formas compartilham a posição de maior frequência: MÃO-DE-VACA e PÃO-DURO. As demais ocorrências surgem uma única vez: SOVINA, AVARENTO, MISERÁVEL e MIQUINHA. O Inf4, que disse PÃO-DURO como RE1, ainda deu MUQUIRANO como RE2, mas reforçou que sua RE1 seria a forma mais usual na comunidade.

Q166

A norma é BALANÇO, seguida por BALANÇA; formas que metodologicamente estamos distinguindo uma da outra. A outra resposta é curiosa, PULEU, dada pelo Inf5.

Q175

A variante INTERRUPTOR foi apurada como norma. As formas ACENDEDOR e TOMADA apareceram uma vez cada. O Inf4, que disse INTERRUPTOR como RE, chegou a mencionar a forma ACENDEDOR também, mas ele a descartou instantes depois de tê-la realizado, e acrescentou ainda que esse objeto seria "popularmente conhecido como TOMADA", corroborando a resposta do Inf1, quem enunciou esta última forma como RE.

As respostas ESPELHO e CONDUTOR foram colhidas também, como ocorrências únicas, invalidadas pelo critério A. Se elas não tivessem sido ocorrências únicas e se o processo da metonímia fosse aceito aqui, teríamos de repensar a invalidação, uma vez que, na linguagem popular, a parte normalmente feita de plástico envolvendo o botão na parede que acende a lâmpada pode ser conhecido como espelho, bem como o cano comumente dentro da parede com os fios do dispositivo aqui tematizado ou os próprios fios podem chamar-se condutor.

3. Correlações internas linguístico-sociais

Verificamos agora a influência de fatores extralinguísticos nas respostas coletadas, especificamente os fatores relacionados às variações sociais diastrática (Ca e Cb), diasssexual (F e M) e diageracional (GII e GI) em tempo aparente. Mediante esse expediente, pode-se visualizar, em certos casos, se há manutenção, tendência ao desuso ou à inovação de variantes.

Antes, utilizemos a tabela 5 como auxílio, que mostra os perfis sociais dos inqueridos. O termo 'classe' está para escolaridade.

Tabela 5 – Perfis dos informantes.

	Sexo	Classe	Idade		Sexo	Classe	Idade
Inf1	F	Cb	GII	Inf5	M	Cb	GII
Inf2	F	Ca	GII	Inf6	F	Cb	GI
Inf3	M	Ca	GII	Inf7	M	Ca	GI
Inf4	F	Ca	GI	Inf8	M	Cb	GI

Ca = classe de alta escolaridade; Cb = classe de baixa escolaridade; GII = velhos; GI = jovens

Fonte: Elaboração própria.

Q47

O que as duas ocorrências de RAIZ têm em comum do ponto de vista extralinguístico é o fato de que ambas foram produzidas por mulheres do grupo etário mais velho. Por razões inferíveis, como o alto índice de respostas invalidadas, não se sabe ao certo se esse é um caso de manutenção ou tendência ao desuso ou à inovação da lexia em relação às variáveis sociais.

Q64

Como esperado, a variante CORVO foi dada como resposta à Q64 por uma pessoa mais velha de baixa escolaridade. Isso corrobora a tendência, entre os jovens, ao desuso de CORVO, por um lado, e, de outro, à expansão da forma URUBU, que já goza por seu turno do estatuto de norma lexical na localidade.

Q66

Recorrendo-se à ficha de informações pessoais do jovem Inf7, estudante de Educação Física, de família humilde, que enunciou a inesperada forma BAITIXÓ, a qual disse ter ouvido na Bahia, verificou-se que nasceu no *locus* da pesquisa, mas seu pai é da cidade baiana de Seabra, fato que torna o informante não ideal (para a dialetologia tradicional) em relação ao dialeto caipira. Em relação à lexia JOÃO-DE-BARRO, evidenciou-se sua manutenção relativamente a todas as variáveis sociais.

Q83

A entrevistada que disse MAMANGABA pertence à faixa etária mais velha e cursou apenas os anos iniciais do ensino fundamental. No tocante às respostas inválidas dos Inf6 e 7 — invalidação de MOSCA pelo critério B (repetição de termo já presente na pergunta) no primeiro caso; abstenção no segundo —, ambos os inquiridos são jovens.

Q94

Três das quatro ocorrências de TERÇOL foram proferidas pelas pessoas de mais idade, dentre as quais ambas as mulheres mais velhas, o que indica estabilidade dessa forma no grupo etário equivalente. As ocorrências da lexia TRÊS SOL foram realizadas por informantes de baixa escolaridade.

Q102, Q136 e Q175

Os dados linguísticos e sociais aqui expostos referentes às Q102 e Q175 não favorecem correlações, dada sobretudo a alta variação entre ocorrências lexicais (caso da Q102) e entre variantes sociais (Q175). Tratando-se da Q136, as ocorrências de TAGARELA foram ditas por homens de alta escolaridade.

Q138

A lexia MÃO-DE-VACA como resposta possível à Q138 foi proferida por mulheres e PÃO-DURO, por jovens de alta escolaridade. SOVINA não apareceu na boca de jovens; conjectura-se que possa existir uma tendência ao desuso dessa designação por parte das gerações mais novas.

Q166

A variante BALANÇO se mantém estável entre os indivíduos jovens e velhos, homens e mulheres, inclusive em relação aos diferentes níveis de escolaridade. A inusitada ocorrência de PULEU foi realizada por um senhor com baixa escolaridade, com ensino fundamental incompleto.

Para visualizar a influência, em *zoom* ampliado, dos fatores diastrático, diageracional e diassexual nas normas e variantes mais frequentes apuradas, preparamos a tabela 6. O círculo meio a meio preenchido (◐) representa que participantes de ambos os sexos disseram a lexia; o círculo totalmente hachurado (●), só pessoas do sexo feminino; o círculo vazio (○), só sujeitos masculinos; e o traço (|), nenhum indivíduo.

Tabela 6 – Lexias e fatores sociais.

Lexia	CaGII	CaGI	CbGII	CbGI	Σ			
					◐	●	○	
raiz	●		●		0	2	0	2
urubu	◐	○	●	◐	2	1	1	0
joão-de-barro	◐	●	◐	◐	3	1	0	0
varejeira	◐	●	○		1	1	1	1
terçol	◐	○			1	0	1	2

catarro	●	I	I	○	0	1	1	2
tagarela	○	○	I	I	0	0	2	2
mão-de-vacalpão-duro	● I	I ●	I I	● I	0 1	2 0	0 0	2 3
balanço	○	●	●	●	1	2	1	0
interruptor	○	●	I	○	0	1	2	1

Σ	●	4 4	1 2	1 1	2 2	8 9	11 9	9 9	12 13
	●	3 2	3 3	3 3	2 1	<i>Ca = classe alta; Cb = classe baixa</i> <i>GII = velhos; GI = jovens</i>			
	○	3 3	3 3	1 1	2 2				
	I	0 1	3 2	5 5	4 5				

● = ambos os sexos disseram a lexia; ● = só mulheres disseram a lexia;
○ = só homens disseram a lexia; I = nenhum indivíduo disse a lexia

Fonte: Elaboração própria.

Na tabela 6, vê-se que a coincidência entre os sexos no tocante à enunciação das formas URUBU, JOÃO-DE-BARRO, VAREJEIRA e TERÇOL é a maior que se apresenta no grupo de indivíduos com mais escolaridade e mais velhos.

Também se nota que o índice de ausência de resposta (seja por simples abstenção, seja por invalidação) é maior entre sujeitos de baixa escolaridade. Por outro lado, se considerarmos MÃO-DE-VACA (e não PÃO-DURO) no conjunto das demais lexias, o grupo de indivíduos mais velhos da classe alta teve participação total na enunciação dos itens lexicais mais frequentes.

A norma JOÃO-DE-BARRO é a forma mais diassexualmente representada entre as demais, com alto índice de envolvimento de ambos os sexos.

No conjunto das pessoas pertencentes à classe com alta escolaridade, não há diferença diassexual concernente ao pronunciamento das formas. Já quanto ao grupo dos participantes com baixa escolaridade, as mulheres foram mais produtivas, especialmente as mais velhas.

4. Contrastes das normas e formas mais frequentes em SdP com resultados externos

Nesta seção, faremos comparações entre nossos dados e os resultados de seis atlas, escolhidos aleatoriamente, dois do Sul do Brasil, dois paulistas (porém, fora do domínio do Médio Tietê) e dois

de outras regiões. A tabela 7 apresenta esses estudos. Nela, quando o número de localidades for muito grande, apenas o total de pontos, sem os nomes, é informado. As siglas 'IP' e 'TI' simbolizam 'informantes por ponto' e 'total de informantes'. 'ALPR' é uma simplificação para a junção entre 'ALPR I' e 'ALPR II'. Ainda, 'R' é para 'região'; 'SP', Estado de São Paulo; 'ORs', 'outras regiões'.

Tabela 7 – Atlas para comparações.

R	Obra	Rede de pontos	IP	TI
Sul	ALERS (2011)	275 localidades dos Estados do Paraná, Santa Catarina e Rio Grande do Sul	1	275
	ALPR (1994, 2007)	65 pontos paranaenses	2	130
SP	ASL-LN-SP (2010)	Costa norte paulista: Ilhabela, São Sebastião, Ubatuba e Caraguatatuba (4)	4	16
	ASL-ABC (2007)	Região do Grande ABC: Ribeirão Pires, São Caetano do Sul, Santo André, Paranapiacaba (distrito de Santo André), Mauá, Rio Grande da Serra, São Bernardo do Campo (um ponto urbano e outro rural) e Diadema (9)	4	36
ORs	ALAP (2017)	Cidades de Amapá, Macapá, Santana, Calçoene, Oiapoque, Porto Grande, Mazagão, Tartarugalzinho, Laranjal do Jarí e Pedra Branca do Amapari (10)	4	40
	ASL-GO (2012)	Municípios de Goiás, Goiandira, Ipameri, Trindade, Morrinhos, Aruanã, Mineiros, Pirenópolis e Caldas Novas (9)	4	36

R = região; SP = São Paulo; ORs = outras regiões; IP = informantes por ponto; TI = total de informantes

Fonte: Elaboração própria.

Queremos verificar se as lexias que se revelaram ser as normas ou as formas mais frequentes em SdP são ocorrentes nas pesquisas mencionadas acima. Em caso afirmativo, marcaremos-las com a letra *e* (N) na tabela 8 quando forem normas lexicais na respectiva rede de pontos; com '+F' quando forem as mais frequentes sem serem normas; com a expressão '>1' quando sua ocorrência for maior do que uma e não for a forma mais frequente; com o número 1 para ocorrência única; e, finalmente, com o número 0 para zero ocorrências. Utiliza-se um hífen (-) caso: (a) a pergunta correspondente não tenha sido aplicada; (b) o resultado não tenha sido divulgado²; ou (c) não se saiba se a lexia figurou ou

² A propósito deste item *b*, os resultados do ALiB (2014), envolvendo 200 informantes de 25 capitais brasileiras, não serão comparados aqui por conta dos dados relativos às Qs sob consideração neste artigo não terem sido publicados.

não como resposta nos trabalhos. A possibilidade *c* se deve ao fato de que algumas publicações utilizam o rótulo 'outras' para variantes de baixa frequência e não mencionam quais são.

Tabela 8 – Designações mais frequentes ou normas em SdP contrastadas com resultados de outras regiões.

Lexia (norma em <i>itálico</i>)	Atlas linguístico											
	Sul		São Paulo		Outras regiões							
	ALERS	ALPR	ASL-LN-SP	ASL-ABC	ALAP	ASL-GO						
<i>raiz</i>	>1	-	+F	+F	-	0						
<i>urubu</i>	>1	+F	N	N	-	N						
<i>joão-de-barro</i>	+F	N	+F	N	-	N						
<i>varejeira</i>	+F	-	N	+F	-	>1						
<i>terçol</i>	>1	>1	>1	N	N	N						
<i>catarro</i>	>1	1	0	>1	-	>1						
<i>tagarela</i>	>1	-	+F	+F	+F	+F						
<i>mão-de-vaca</i> <i>pão-duro</i>	>1	+F	-	-	N	>1	+F	>1	+F	>1	+F	>1
<i>balanço</i>	-	+F	N	N	-	N						
<i>interruptor</i>	-	-	>1	+F	-	>1						

□	N	0 0	1 1	4 3	4 4	1 1	4 4
	+F	2 3	2 2	3 3	5 4	2 1	2 1
	>1	6 5	1 1	2 3	1 2	0 1	3 4
	1	0 0	1 1	0 0	0 0	0 0	0 0
	0	0 0	0 0	1 1	0 0	0 0	1 1
	-	2 2	5 5	0 0	0 0	7 7	0 0

N = norma; +F = item mais frequente sem ser norma; >1 = item com mais de uma ocorrência sem ser a mais frequente; 1 = ocorrência única; 0 = zero ocorrências; - = pergunta não aplicada ou resultados não publicados

Fonte: Elaboração própria.

Antes de irmos às inferências baseadas na tabela 8, é preciso dizer que: (i) em alguns dos trabalhos comparados, fala-se de norma lexical quando a forma tem a partir de cinquenta por cento de frequência relativa (bem como distribuição regular, quando se trata de mais de uma localidade). Neste artigo, no entanto, só chamamos de norma os itens com mais de cinquenta por cento de frequência relativa (e distribuição regular pela rede de pontos); (ii) no ALAP (2017), TAGARELA é a forma mais frequente juntamente com FALADEIRA (FALADOR) (cf. *ib.*, p. 158); (iii) a formulação da Q47 no ALERS (2011) — "Depois que se corta o pé de arroz, ainda fica uma pequena parte, que é..." (ALERS,

2011, p. 41) —, bem como da Q166 no ALPR — "Como se chama aquele brinquedo que as crianças fazem amarrando uma corda numa árvore e, sentadas embaixo, fazem assim? (gesto)" (ALPR II, 2007, p. 234) — não é inteiramente idêntica à aplicada por nós em campo (v. tabela 2).

Tomando a tabela 8 em consideração, nota-se que a região do Grande ABC apresenta o maior número de normas (Ns) e formas mais frequentes (+Fs) que também são Ns e +Fs em SdP. Como antevisto, o fato de que essa região, entre as demais, é a mais próxima a SdP reforça o papel que fatores geográficos exercem na variação/manutenção linguística.

Outra inferência possível relaciona-se com a ausência de Ns na rede de pontos do ALERS (2011) que são Ns em SdP. O ALPR, no entanto, apresenta pelo menos uma N (JOÃO-DE-BARRO) que também é N em SdP. Por outro lado, excetuando-se o território paulista, o Estado de Goiás é a área em que mais Ns apresenta que também são Ns em SdP.

O ALERS (2011) exibe o maior índice de ocorrência de variantes lexicais secundárias (nem Ns, nem +Fs) que são Ns ou +Fs em SdP.

Com o *zoom* ajustado à lexia JOÃO-DE-BARRO, vê-se que sua ocorrência é a mais forte entre os demais itens sob análise nos atlas comparados: N tanto no Paraná (segundo o ALPR), quanto na região do Grande ABC e em Goiás; +F no Sul do Brasil (conforme o ALERS, 2011) e no litoral norte de São Paulo. De outro lado, a ocorrência de CATARRO, +F em SdP, é a forma lexical menos representativa nos locais comparados; não sendo nem N, nem +F.

Quanto ao par MÃO-DE-VACA e PÃO-DURO, +Fs em SdP, a primeira forma tem mais representatividade do que a segunda no quadro da presente comparação. Enquanto PÃO-DURO é +F no ALERS (2011) e tem ocorrência >1 em São Paulo e Goiás (ASL-LN-SP, 2010; ASL-ABC, 2007; e ASL-GO, 2012), MÃO-DE-VACA é N no ASL-LN-SP (2010) e +F no ASL-ABC (2007), no ALAP (2017) e no ASL-GO (2012).

5. Observações esparsas complementares

Q47

A variante SOQUEIRA, que foi proferida uma vez em SdP depois de a forma SOCA haver sido sugerida ao entrevistado (Inf3) no P3, apareceu também uma vez no município paulista de Santo André (cf. ASL-ABC, 2007, p. 441).

No Paraná, SOQUEIRA é a resposta mais frequente entre oito formas para duas perguntas distintas: (a) "Depois que se corta o pé de arroz, ainda fica uma pequena parte enterrada. Como se chama essa parte que fica na terra?", com 43,07% de ocorrência; e (b) "Como se chama o arroz de

segunda colheita?", com 43,59% de surgimento. Para a questão *b*, a variante SOQUEIRA ocorreria virtualmente ao longo do território do Estado, enquanto que, para a pergunta *a*, ela só não seria o caso na região do Paraná oficialmente denominada metropolitana (cf. ALPR II, 2007, cartas 298 e 299).

Como variante que aparenta fundir SOCA e TOUCEIRA, em Ferreira (1975, p. 1324) há duas entradas para SOQUEIRA, a primeira com duas acepções e a segunda com uma. O primeiro verbete assevera que se trata de uma composição entre 'soca' e '-eira', que denotaria ou o raizame de plantas, especialmente de cana, que resta após o corte, ou um engenho em que se produz muita soca, entendida como a segunda produção da cada após a primeira ter sido cortada; esta última definição seria um regionalismo pernambucano. Na segunda entrada, lê-se que a forma seria uma combinação entre 'soco' e '-eira', que se referiria, nos Estados do Rio Grande do Sul e de São Paulo, ao objeto de metal que se ajusta à mão para o propósito de desferir golpes.

Q64

Na pesquisa de Soares (2012, p. 188 et seq., bem como p. 438), com 24 informantes nos municípios paulistas de Guarulhos, Arujá, Santa Isabel, Mairiporã e Nazaré Paulista, estes quatro últimos pertencendo à região metropolitana da cidade de São Paulo, URUBU também é a norma, com 91,7% de frequência relativa e distribuição regular, enquanto CORVO aparece uma única vez, em Nazaré Paulista.

Q66

Em Pernambuco, MARIA-DE-BARRO seria uma variante de JOÃO-DE-BARRO (cf. FILIPAK, 2002, p. 218). Quando essa pergunta foi aplicada no contexto do ASL-LN-SP (2010), só na localidade de Caraguatatuba a lexia JOÃO-DE-BARRO não foi ouvida, onde ao invés disso as formas JURITI e PICA-PAU apareceram (cf. *ib.*, p. 261 et seq., p. 257 e p. 274 et seq.). Apesar desse registro, é possível porém aventar a hipótese segundo a qual essas duas últimas designações têm referência semântica distinta daquela que a Q66 tematiza.

Q83

Não se ouviu a expressão BIRU em Santana de Parnaíba, item lexical que seria usado por caipiras de São Paulo (cf. IHERING, 1968; verbete BIRÚ). Como se vê no ALERS (2011), essa lexia teve ocorrência única nas áreas rurais das cidades paranaenses de Guaraqueçaba e Paranaguá. Em relação à rede de pontos do atlas mencionado, a forma VAREJEIRA predominou, ocorrendo 104 vezes, face a outras 26 variantes (cf. *ib.*, p. 244 et seq.). Nesse mesmo estudo, em relação a BUTUCA e MAMANGABA (formas que surgiram em SdP, além de VAREJEIRA), a primeira se apresentou em

três pontos catarinenses (Joaçaba, Lajes e Urubici), enquanto a segunda não foi ouvida em nenhuma das localidades investigadas.

Q94

Em nosso campo empírico, as variantes TERÇOL (lexia mais frequente), TRÊS SOL (segunda mais frequente), TREÇOL (ocorrência única) e CISTO (uma vez) foram ouvidas e foram consideradas formas distintas. No ASL-GO (2012), as ocorrências de TREIÇOL, TRIÇOL e TRUÇOL foram registradas e definidas como realizações de TERÇOL, operação da qual decorre a consequência de que tal forma seria, no caso, a norma lexical, com distribuição regular e ocorrência de 77,78%, face às variantes BUNITINHA, ESPINHA, FIGUEIRINHA e abstenções (cf. *ib.*, p. 103 e p. 240). Quanto à forma CISTO, que foi dita em SdP, no ALERS (2011, p. 486), entre as 21 formas distintas registradas em concorrência variacional, ela não figura nenhuma vez.

Q102

Em Iguape, ao sul da costa paulista, onde se encontra um dialeto caiçara do português, 12 pessoas foram entrevistadas, e a lexia TATU foi a mais frequente, com 75% de incidência relativa em concorrência com outras duas variantes, CERA (com duas ocorrências) e PELANCA (com uma) (cf. SILVEIRA, 2009, p. 123).

Q136

Segundo Viola (2010, p. 662 *et seq.*, bem como p. 700), na cidade de Guiratinga e em seus distritos de Alcantilado e Vale Rico no Estado do Mato Grosso, foi constatado que CONVERSADOR é a norma municipal, com 81,12% de frequência relativa, e TAGARELA ocorreu apenas 5%, num conjunto de quatro variantes, incluindo LINGUARUDO e FUXIQUEIRO; as quatro perfazendo 80 ocorrências totais (uma por inquirido) em frequência absoluta. Aí, mesmo LINGUARUDO (10%) foi mais frequente do que TAGARELA, designação mais frequente em SdP.

Q138

No ALERS (2011), a lexia PÃO-DURO é o caso em 91 das 267 respostas válidas, num cenário de 15 variantes com mais de uma ocorrência (contando com as de ocorrências únicas, 30) registradas. MÃO-DE-VACA ocupa o sexto lugar entre as formas mais frequentes, com dez realizações. Entre todos concorrentes lexicais, MIQUINHA, forma com a qual nos deparamos em Santana de Parnaíba, não ocorreria no Sul do país. Em compensação, as interessantes realizações MÍSIKO (3 vezes) e

SUMÍTICO (1), sim. A primeira no Estado do Paraná, nas cidades de Pinhão, São José dos Pinhais e Tijucas do Sul; estas duas últimas vizinhas bem próximas uma da outra. A segunda forma foi coletada em Imbituba, Santa Catarina (cf. *ib.*, p. 584 *et seq.*).

Q166

Verificou-se que a forma BALANÇO é a norma entre quatro lexias, com 36 ocorrências, seguida pela variante BALANGO, com 20, no norte do Estado de Mato Grosso, nomeadamente nos municípios de Sinop, Cláudia, Vera e Santa Carmem, numa investigação que contou com 40 informantes (cf. PHILIPPSSEN, 2013, p. 540 *et seq.*).

Q175

Em Goiás (v. ASL-GO, 2012), a designação APAGADOR foi a mais frequente (36,11%), seguida por INTERRUPTOR (33,33%). Em relação à lexia TOMADA, ela foi proferida quatro vezes, com 11,11% de frequência (cf. *ib.*, p. 325).

Quanto à lexia APAGADOR, o fato é curioso. Isso porque o objeto, o dispositivo que apaga também acende a luz, mas seu nome só alude ao ato de apagar. O inverso acontece relativamente à forma ACENDEDOR, ouvida em SdP.

Considerações finais

O estudo para o atlas da comunidade de fala da região do Médio Tietê se debruça no momento nos demais dados semântico-lexicais, bem como de outros níveis linguísticos, produzidos no município de Santana de Parnaíba, e segue em direção às informações coletadas nos outros nove pontos da rede de investigação, para então serem expostas, avaliadas, analisadas, fornecendo-se assim registros do vernáculo dessa área do Estado de São Paulo, os quais podem ser comparados ao longo do território nacional brasileiro a partir de outras pesquisas acerca dos estados da língua; estados passados, presentes e futuros.

Ainda, compartilhamos o nosso entendimento de que análises que tomem em consideração respostas do tipo A (com discrepância em relação ao referente tematizado pela questão), B (repetição de lexia já presente na indagação) ou C (designação vaga ou falta de termo específico), deliberadamente ou não, para efeito de cálculo de frequência relativa ou qualquer outro efeito, podem enviesar as conclusões de pesquisa dialetológica, em todo caso já delicadas por definição (desde um

ponto de vista da representatividade estatística); nomeadamente por envolver via de regra poucos informantes por ponto de inquérito. A esse respeito, parece-nos que muitas das ocorrências únicas registradas e analiticamente tratadas em trabalhos dialetológicos são, na verdade, casos do tipo A, B ou C, aos quais às vezes fortes dispêndios estatísticos são gastos.

Também gostaríamos de dizer que a medida, quando possível/viável, de se publicarem todos os resultados nos atlas e em outros estudos, ainda que implicando resultados que não chamem à atenção, nem qualitativa, nem quantitativamente, contribui de maneira importante aos trabalhos comparativos. A favor dessa observação, basta lembrarmos as lacunas observadas na tabela 8, sinalizadas com hífen (-), no tocante aos casos *b* (resultados não publicados) ou *c* (ausência de explicitação de variantes de baixa frequência). Às vezes, o que ocorre pouco na localidade/região X pode ser muito frequente na localidade/região Y.

Aspecto a ser lembrado, o qual não aprofundamos aqui, é a potencialização da análise linguística relativa às razões e consequências de aparecimento ou não de determinadas lexias que não se estabeleceram em normas ou em formas mais frequentes às dez questões abordadas. Esse tipo de avaliação é tema de relevância na Dialetologia Relacional e Pluridimensional, pela qual se amplia qualitativamente o entendimento sobre as variantes, tanto sincrônica quanto diacronicamente.

Para finalizar, mencionamos a relevância da correta aplicação da técnica de entrevista de três passos na pesquisa dialetológica. No bojo dos nossos dados empíricos relacionados às questões semântico-lexicais, existem lacunas importantes originadas pela assistemática do uso da técnica. Com ela, não apenas há condições de se registrarem variantes do domínio do conhecimento passivo do informante ou de hierarquizá-las conforme os comentários metalinguísticos do entrevistado, senão que também se pode controlar mais eficientemente a condução do inquérito, no sentido de esclarecer o entrevistado sobre do que realmente se trata, sobre qual referente semântico está em questão, mediante os passos 2 (insistência) e 3 (sugestões).

Referências bibliográficas

- RAZKY, Abdelhak; RIBEIRO, Celeste Maria da Rocha; SANCHES, Romário Duarte. **Atlas linguístico do Amapá**. São Paulo: Labrador, 2017.
- ALTENHOFEN, Cléo Vilson; KLASSMANN, Mário Silfredo (Orgs.). **Atlas linguístico-etnográfico da Região Sul do Brasil – ALERS: cartas semântico-lexicais**. Porto Alegre: UFRGS, 2011. Florianópolis: UFSC, 2011.
- CARDOSO, Suzana Alice Marcelino da Silva *et al.* **Atlas linguístico do Brasil**. Vols. 1 e 2. Londrina: EDUEL, 2014.
- ALMEIDA, Michelle Viana. **Edição de documentos do século XIX para o estudo da variedade linguística em Porto Feliz**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.
- AGUILERA, Vanderci de Andrade. **Atlas linguístico do Paraná**. Curitiba: Imprensa Oficial do Estado, 1994.
- ALTINO, Fabiane Cristina. **Atlas linguístico do Paraná – ALPR II**. Tese (Doutorado em Estudos da Linguagem). Centro de Letras e Ciências Humanas, Universidade Estadual de Londrina, Londrina, 2007.
- AMARAL, Amadeu. **O dialeto caipira: gramática, vocabulário**. São Paulo: O Livro, 1920.
- CRISTIANINI, Adriana Cristina. **Atlas semântico-lexical da região do Grande ABC**. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.
- AUGUSTO, Vera Lúcia Dias dos Santos. **Atlas semântico-lexical do Estado de Goiás**. Vols. I, II e III. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2012.
- ENCARNAÇÃO, Márcia Regina Teixeira da. **Atlas semântico-lexical de Caraguatatuba, Ilhabela, São Sebastião e Ubatuba – municípios do litoral norte de São Paulo**. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2010.
- BRITO, Mariana Santiago de. **Estudo do vocabulário do Cururu de Piracicaba**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.
- FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário da língua portuguesa**. Rio de Janeiro: Nova Fronteira, 1975.
- FILIPAK, Francisco. **Dicionário sociolinguístico paranaense**. Curitiba: Imprensa Oficial, 2002.

- GARCIA, Rosicleide Rodrigues. **Para o estudo da formação e expansão do dialeto caipira na região de Capivari**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.
- IHERING, Rodolpho von. **Dicionário dos animais do Brasil**. Brasília: Universidade de Brasília, 1968.
- MOTA, Camila. **Edição de documentos oitocentistas e estudo da variedade linguística em Santana de Parnaíba**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.
- PHILIPPSEN, Neusa Inês. **A constituição do léxico norte mato-grossense na perspectiva geolinguística: abordagens sócio-semântico-lexicais**. Tese (Doutorado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.
- PIRES, Cibélia Renata Silva. **Formação e expansão da cultura e do dialeto caipira na região de Piracicaba**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.
- LEAL, Eneida de Goes. **Elisão silábica e Haplologia: aspectos fonológicos do falar da cidade paulista de Capivari**. Dissertação (Mestrado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006.
- LEAL, Eneida de Goes. **Teoria fonológica e variação: a queda de sílaba em Capivari e em Campinas**. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2012.
- LEOPOLDINO, Everton Altmayer. **A fala dos tirolezes de Piracicaba: um perfil linguístico dos bairros Santana e Santa Olímpia**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.
- LEVADO, Rosimeire Firão. **Para o estudo da formação e expansão da cultura e do dialeto caipira na região de Tietê**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.
- ROMERO, Renata Maran Longuini. **Estudo semântico-lexical na região de Itu**. Dissertação (Mestrado Mestre em Filologia Românica e Língua Portuguesa). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.
- SANTIAGO-ALMEIDA, Manoel Mourivaldo. **História e variedade do português paulista na região do Médio Tietê**. Projeto Pq/CNPq. São Paulo: Mimeo, 2013.
- SILVEIRA, Roseli da. **Estudo sociogeolinguístico do município de Iguape: aspectos semânticos-lexicais**. Dissertação (Mestrado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2009.

- SOARES, Rita de Cássia da Silva. **Atlas semântico-lexical da região norte do Alto Tietê (ReNAT)** – São Paulo. Tese (Doutorado em Língua Portuguesa). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2012.
- THUN, Harald (Coord.). **Atlas lingüístico diatópico y diastrático del Uruguay – Norte (ADDU-Norte)**. tomo I. Kiel: Westensee-Verl, 2000.
- VIOLA, Wanderleia Silva Carvalho. **O léxico guiratinguense na perspectiva dialetológica: aspectos semânticos-lexicais**. Dissertação (Mestrado em Letras). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo,