

ALFA
Revista de Linguística

UNIVERSIDADE ESTADUAL PAULISTA

Vice-reitor no exercício da reitoria

Julio Cesar Durigan

Pró-Reitora de Pesquisa

Maria José Soares Mendes Giannini

Apoio:

PROPe
Pró-Reitoria de Pesquisa



UNIVERSIDADE ESTADUAL PAULISTA

ALFA
Revista de Linguística

ISSN 1981-5794 (ON LINE)
ISSN 0002-5216 (IMPRESSA)

Alfa	São Paulo	v.56	n.3	p.719-1163	2012
------	-----------	------	-----	------------	------

Alfa: Revista de Linguística
UNESP – Univ Estadual Paulista,
Pro-Reitoria de Pesquisa
Rua Quirino de Andrade, 215
01049-010 – São Paulo – SP
alfa@unesp.br

<i>Editor responsável</i>	Anna Christina Bentes
Roberto Gomes Camacho	
<i>Co-Editor</i>	<i>Revisão Geral</i>
Gládis Massini Cagliari	CCLi Consultoria Linguística
<i>Editoria Executiva</i>	<i>Assessoria Técnica</i>
Erotilde Goreti Pezatti	Ana Paula Menezes Alves
Luciani de Paula	
Odilon Helou Fleury Curado	<i>Capa</i>
Rosane de Andrade Berlínck	Adriana Bessa Damman
<i>Organizadores deste volume temático</i>	<i>Diagramação</i>
Sebastião Carlos Leite Gonçalves	Eron Pedroso Januskeivitz
	Gianfrancesco Afonso Cervelin

Conselho Editorial

Ângela Cecília Souza Rodrigues (USP), Ataliba Teixeira de Castilho (USP), Bento Carlos Dias da Silva (UNESP), Christian Hudelot (CNRS), Claudia Maria Xatara (UNESP), Clélia Cândida Abreu Spinardi Jubran (UNESP), Daniel Leonard Everett (University of Manchester), Dermeval da Hora (UFPB), Diana Luz Pessoa de Barros (USP), Edair Gorski (UFSC), Eduardo Calil (UFAL), Esmeralda Vailati Negrão (USP), Ester Miriam Scarpa (UNICAMP), Fábio Lopes da Silva (UFSC), Freda Indurski (UFRS), Helena Hatsue Nagamine Brandão (USP), Ieda Maria Alves (USP), Ingêdore G.V. Koch (UNICAMP), Jacques Fontanille (Université de Limoges), Jacyntho Luís Brandão (UFMG), João Azenha Júnior (USP), João Wanderlei Geraldi (UNICAMP), John Lachlan Mackenzie (ILTEC), John Robert Schmitz (UNICAMP), José Borges Neto (UFRJ), Kanavillai Rajagopalan (UNICAMP), Kees Hengeveld (Universidade de Amsterdã), Laurent Danon-Boileau (Paris V – CNRS), Leda Bisol (PUC-RS), Leonor Sciar Cabral (UFSC), Lúcia Teixeira (UFF), Luís Antônio Marcuschi (UFPE), Luiz Carlos Travaglia (UFU), Maria Beatriz Nascimento Decat (UFMG), Maria Bernadete Marques Abaurre (UNICAMP), Maria Helena de Moura Neves (UNESP), Maria Helena Vieira Abrahão (UNESP), Maria Luiza Braga (UFRJ), Maria Margarida Martins Salomão (UFJF), Maria Marta Pereira Scherer (UnB), Mariangela Rios de Oliveira (UFF), Renata Coelho Marchezan (UNESP), Roberta Pires de Oliveira (UFSC), Roberto Gomes Camacho (UNESP), Rosemary Arrojo (State University of New York), Sérgio de Moura Menuzzi (UFRGS), Seung Hwa Lee (UFMG), Sírio Possenti (UNICAMP), Vera Lúcia Paredes Pereira da Silva (UFRJ), Zélia de Almeida Cardoso (USP).

Publicação semestral/*Bi-annual publication*

Alfa: Revista de Linguística / UNESP – Universidade Estadual Paulista – v.1 (1962) – v.23 (1977); v.24 (1980) – São Paulo, UNESP, 1962-1977; 1980 –

Semestral
Publicação suspensa de 1978-1979

ISSN 0002-5216 (Impressa) 1981-5794 (On line)

Os artigos publicados na Alfa: Revista de Linguística são indexados por:
The articles published in Alfa: Revista de Linguística are indexed by:

BLL – Bibliography of Linguistic Literature

CLASE – Cich-Unam – Citas Latinoamericanas en Ciencias Sociales y Humanidades

Francis Database

IBZ – International Bibliography of Periodical Literature in the Humanities and Science Galé

LLBA – Linguistic and Language Behavior Abstracts

MLA – International Bibliography.

APRESENTANDO RETRATOS SOCIOLINGUÍSTICOS E DIALETOLÓGICOS DO BRASIL

Sebastião Carlos Leite Gonçalves (UNESP, CNPq)
Anna Christina Bentes (UNICAMP)

Quando nos dispusemos a organizar o presente volume temático da *Alfa – Revista de Linguística*, tínhamos em mente um volume que pudesse apresentar à comunidade científica uma fotografia do atual estado da arte dos estudos sociolinguísticos e dialetológicos brasileiros. Tínhamos em mente um volume que fosse representativo dos diversos grupos de pesquisa espalhados pelo território nacional, um volume que reunisse e colocasse em relação os mais importantes resultados de pesquisas sobre a realidade sociolinguística do português brasileiro (PB). Superando nossas expectativas, essa ideia-selho se concretizou, e, agora, presentearmos o público-leitor com 15 excelentes artigos que, emanados dos principais grupos de pesquisa em Sociolinguística e Dialetologia, revelam a contínua e incessante tarefa de coleta, registro, organização e interpretação de dados linguísticos, e compartilham certa vocação historiográfica, à medida que buscam também situar os leitores sobre o modo como foram sendo construídas as trajetórias de cada grupo.

Nessa apresentação dos artigos aos leitores, nos impusemos a tarefa de também tentar mostrar em que pontos se aproximam e se distanciam as pesquisas e os pesquisadores aqui retratados. Para tanto, o agrupamento dos artigos no volume obedece aos seguintes critérios: primeiramente dispõem-se aqueles que apresentam resultados gerais de pesquisas no interior dos grupos de pesquisa, depois aqueles voltados para questões metodológicas e, por último, os que trazem resultados de fenômenos variáveis específicos.

No primeiro conjunto, estão os artigos representantes dos grupos de pesquisadores do PEUL, do VARSUL, do PORTVIX e do VALCO,¹ os quais, a partir dos pressupostos centrais da Sociolinguística Variacionista, apresentam resultados de pesquisa ordem mais geral, com base em bancos de fala: no caso do PEUL, ao longo de mais de 30 anos de trabalho, no caso do VARSUL, ao longo de mais de 25 anos de pesquisa; no caso do PORTVIX, há mais de dez anos e, no

¹ Programa de Estudos sobre o Uso da Língua, Projeto Variação Linguística Urbana do Sul, Projeto Português Falado na Cidade de Vitória, Projeto Variação Linguística no Centro-Oeste, respectivamente.

caso do VALCO, oficialmente, ao longo dos últimos 4 anos. Assim, esses primeiros artigos retratam a longevidade dos três primeiros grupos e o fato de todos eles compartilharem os pressupostos de construção de bancos de dados com base em uma metodologia laboviana, visando a oferecer subsídios para descrição linguística, para testes e desenvolvimentos de teorias linguísticas, para formação de novos pesquisadores e para promoção do conhecimento do PB e do respeito a suas variedades. Ainda, como parte desse primeiro conjunto, inserem-se três outros artigos complementares entre si, os quais fornecem um panorama do Projeto ALiB², com informações sobre seus antecedentes, objetivos, metodologia e alguns resultados que permitem construir não só uma visão do estágio atual do desenvolvimento do projeto como também de sua dimensão sociolinguística.

Dois são os artigos que enfocam **questões metodológicas**, relacionadas principalmente à constituição de amostras representativas de dados linguísticos por meio de entrevistas sociolinguísticas. No primeiro deles, ao se discutirem procedimentos que deveriam ser adotados para a organização de novos bancos de dados, faz-se um breve retrospecto dos bancos de dados já constituídos, para, então, se apresentarem sugestões relativas à coleta e à expansão de *corpora* de diferentes comunidades de fala e de diferentes comunidades de prática. Além disso, o artigo também apresenta os critérios de constituição dos seguintes bancos de dados de fala do Nordeste: o Banco de dados da Fala Culta de Itabaiana, o banco de dados “Falares Sergipanos” e o banco de dados “Fala – Natal”. O segundo artigo, além de discutir critérios que definem a construção da amostra do GESOL-USP,³ traça um panorama dos trabalhos que se têm desenvolvido a partir dos dados já coletados pelo Grupo, que tem como objetivo geral inserir a cidade de São Paulo e o paulistano no mapa sociolinguístico do Brasil.

Menos voltados para discussões de caráter metodológico, os seis próximos artigos cuidam mais de expor resultados de análises de **fenômenos específicos**, investigados de acordo com a metodologia da sociolinguística variaacionista. A partir de banco de dados produzido no interior do PROBRAVO⁴, o primeiro artigo apresenta uma descrição sociolinguística que prioriza a variação de vogais médias pretônicas, de vogais postônicas mediais e da nasalidade alofônica presentes na variedade do português falado na Amazônia paraense. Os resultados reforçam a hipótese de que o Pará compreenderia uma ilha dialetal entre os dialetos do Norte do Brasil. Focando o português falado no extremo oposto do país, mas ainda tratando de variação fonológica, o próximo artigo analisa a palatalização variável das plosivas alveolares em amostras de fala da comunidade de Flores da Cunha, armazenadas no BDSer.⁵ A análise revela a progressão da palatalização na

² Atlas Linguístico do Brasil.

³ Grupo de Estudos e Pesquisas em Sociolinguística da USP.

⁴ Grupo de Pesquisa de Descrição Sócio-Histórica das Vogais (do Brasil) <<http://relin.letras.ufmg.br/probravo>>.

⁵ Banco de Dados de Fala da Serra Gaúcha, da Universidade de Caxias do Sul.

comunidade em função de condicionantes de natureza tanto linguística quanto social (*jovens, vogal alta fonológica, habitantes de zona urbana e consoante-alvo desvozeada*), evidenciando que a alta densidade da rede social, nucleada por informantes de grupos etários mais velhos, refreia a palatalização.

Ainda nessa linha de descrição e análise *de fenômenos específicos*, mas com objetivos e ênfases diferenciados, os três próximos artigos se debruçam sobre fenômenos variáveis de ordem morfossintática. O primeiro, baseado em amostras de fala do Banco de Dados Iboruna,⁶ apresenta comparações interdialetais para três diferentes fenômenos variáveis, e constata, na fala do interior paulista: (i) a acentuada frequência de uso de *a gente* em detrimento de *nós*, com influência apenas da variável faixa etária; (ii) índices bastante elevados de concordância verbal com *nós*; (iii) regra semicategórica na concordância verbal com *a gente*; (iv) índices acima dos 70%, na concordância verbal de 3PP. Para essa mesma variedade do português, discute-se, no artigo seguinte, a natureza de motivações funcionais e formais na explanação de fenômenos variáveis, a partir da investigação da marcação de número em SNs e SAs em posição de predicativo. As análises mostram que princípios funcionais (como o de Economia e de Condição de Distintividade) atuam na marcação de pluralidade, mas não com força suficiente para reger todo o processo; por outro lado, motivações formais, gerenciadas pelo Princípio de Paralelismo Formal, também influenciam significativamente a marcação explícita de pluralidade nos dois fenômenos. Assim como ocorre com as motivações funcionais, as formais não atuam categoricamente de modo a governar sozinhas o fenômeno, como têm mostrado os estudos de outras variedades.

No terceiro artigo desse último agrupamento, são analisadas a concordância nominal de número e a concordância verbal de terceira pessoa do plural em variedades urbanas do Português Europeu, do Português do Brasil e do Português de São Tomé. Para o Português Europeu, os dados revelam tratar-se, em termos laboviano, de regra categórica, no que se refere à concordância nominal, e semicategórica, no que se refere à verbal; no Português do Brasil e no Português de São Tomé, ambas as regras são variáveis.

Fecha o grupo dos **fenômenos variáveis específicos** e, por fim, o volume como um todo, artigo que, advindo do VALPB⁷, enfoca três processos fonológicos na oralização do texto escrito e sua interferência no processo de leitura: a ditongação, a monotongação e o apagamento da oclusiva ‘d’ no grupo –ndo. Dentre os fenômenos variáveis analisados, as formas não estigmatizadas apresentaram maior

⁶ O Banco de Dados Iboruna (= rio preto, em Tupi), constituído pelo Projeto ALIP (Amostra Linguística do Interior Paulista) entre 2004 e 2007, compõe-se de dois tipos de amostras de fala: a primeira, coletada de acordo com os critérios da Sociolinguística laboviana constitui a Amostra Censo, e a segunda, coletada secretamente em contextos interacionais livres, sem controle de qualquer variável social, constitui a Amostra de Intereração. Cf. Rúbia e Gonçalves, neste volume

⁷ Variação Linguística no Estado da Paraíba.

índice de ocorrência, e sua realização não traz, em geral, influência negativa no desempenho da leitura por parte dos alunos. As formas estigmatizadas, como mais propensas à correção pelo professor, ocorrem com menos frequência e tendem a ser menos usadas com o avanço da escolaridade.

Feitas as apresentações, tentemos, a partir de agora, estabelecer relações possíveis entre os conteúdos dos artigos, considerando três eixos: (i) os pressupostos teóricos compartilhados pelos diferentes grupos de pesquisa; (ii) os seus principais resultados e (iii) as preocupações metodológicas dos autores em relação ao fazer científico nos campos da Sociolinguística e da Dialetologia.

Uma característica importante deste volume temático é o fato de que todos os trabalhos compartilham pressupostos teórico-metodológicos gerais da sociolinguística variacionista, tal como o conceito de regra variável que emerge dos estudos variacionistas e que:

representou um avanço teórico importante para a ciência linguística, na medida em que incorpora à gramática a sistematicidade estrutural e social do componente variável da língua. O investimento constante nessa vertente de estudos [...] tem acumulado evidências consideráveis sobre (a) a configuração de vários fenômenos variáveis no português falado [...], (b) as restrições que operam sobre cada um deles, (c) a ação regular dos diferentes princípios que subjazem ao uso das formas variáveis e que regulam a implementação da mudança na comunidade de fala e no indivíduo.⁸

Além disso, é possível dizer que muitos trabalhos do volume compartilham dos princípios elaborados a partir dos resultados das pesquisas pioneiras do Projeto PEUL, no que diz respeito às trajetórias de algumas variantes linguísticas do PB: “(a) contínua implementação de variantes linguísticas; (b) estabilidade da variação e (c) retração de uma variante em contextos particulares”.⁹ Questão importante discutida ao longo de vários dos artigos é a sistematicidade no controle estrutural da variação. No entanto, variados fenômenos têm ressaltado a importância de princípios de natureza funcional. Na direção desse posicionamento, temos a apresentação de uma perspectiva na qual as gramáticas das línguas apresentam uma natureza parcialmente sistemática, o que colabora para uma compreensão da natureza da variação linguística:

As gramáticas são, por um lado, parcialmente autônomas, e daí *sistemas*, e, por outro, parcialmente suscetíveis a pressões externas ao sistema, daí *adaptativas*. Segundo Du Bois (1985), essa distinção é metodologicamente útil apenas se dermos pleno reconhecimento

⁸ Cf. Paiva e Paredes e Silva, neste volume.

⁹ Cf. Paiva e Paredes e Silva, neste volume.

à existência de motivações em competição e desenvolvêrmos um arcabouço teórico que, inicialmente, descreva e analise a interação das motivações com contextos específicos e, posteriormente, seja capaz de prever a resolução da competição entre elas.¹⁰

Essa concepção favorece um tipo de interpretação sobre as relações entre os fatores internos e os fatores externos que pressionam/constituem/formatam a variação linguística:

Dizer que a língua é adaptativa por responder a pressões do contexto externo, implica de pronto rejeitar a teoria do estruturalismo autônomo, que é limitada na tentativa de explicar a estruturação de uma língua, por reconhecer apenas forças motivadoras internas. Dizer, por outro lado, que a língua é um sistema, e que, como tal, tem certo grau de continuidade de existência, implica rejeitar o arcabouço teórico do funcionalismo transparente, que falha em explicar o fato de que categorias gramaticais podem ser preservadas, por assim dizer, numa forma mais ou menos congelada. Mais importante ainda que isso talvez, ele falha em construir apropriadamente o aspecto interno mais fundamental da gramática, o processo de gramaticalização em si mesmo.¹¹

Vemos, então, que esse número temático possibilita a leitura de autores que, ao mesmo tempo em que compartilham de alguns pressupostos mais gerais, também podem divergir sobre a maneira de ver a interação entre fatores internos e externos na explicação de fenômenos linguísticos variáveis.

Ao longo do volume, é possível observar ainda a complementariedade entre os resultados de pesquisa apresentados. Não poderíamos deixar de trazer aqui os resultados das várias pesquisas sobre o clássico fenômeno da concordância, cuja importância para a compreensão das diferenças entre variedades do Português pode ser assim resumida:

Um dos temas mais focalizados no âmbito do Português, sobretudo no que se refere à variedade brasileira, a concordância nominal e verbal suscita, ainda, grande interesse em virtude não só de fatores linguísticos que se vêm mostrando recorrentes em diversos estudos, mas também de implicações sócio-histórico-culturais para a caracterização das variedades do Português.¹²

São exatamente as implicações de natureza sócio-histórico-cultural para a caracterização da variedade do PB que podem ser observadas na formulação dos resultados de pesquisa abaixo:

¹⁰ Cf. Camacho e Salomão, neste volume.

¹¹ Cf. Camacho e Salomão, neste volume.

¹² Cf. Brandão e Vieira, neste volume.

Os primeiros estudos em tempo real de curta duração fornecem indícios claros de aumento das marcas de concordância verbal e nominal tanto no indivíduo como na comunidade de fala carioca (NARO; SCHERRE, 2003, 2010). Reexaminando essa tendência mais recentemente (SCHERRE; NARO, 2011), os dois autores retomam a questão do efeito de variáveis sociais sobre a presença das marcas de concordância entre verbo e sujeito e em todos os elementos do SN, na comunidade de fala carioca, com o intuito de evidenciar como os grupos de falantes mudam no tempo.

(...) Usando os agrupamentos de idade projetados no lugar das faixas etárias originalmente estabelecidas, Naro & Scherre (2011) puderam, então, ver mais claramente as mudanças dentro da comunidade de fala e confirmar tanto o modelo de fluxos e contrafluxos, proposto inicialmente por Naro (1981) e retomado por Naro (1999) e Naro & Scherre (2010), como a ampliação da aquisição da concordância do plano individual para a comunidade como um todo. Testemunha-se, assim, a previsão de Naro (*op. cit.*), baseada na análise da Amostra MOBRAL (NARO, 1981), no que diz respeito ao efeito da variável orientação cultural sobre a concordância verbal. Os falantes que se identificavam com valores da classe média, os de orientação vicária, favoreciam mais a concordância verbal. Este traço foi mensurado por meio do tipo de programa de televisão visto, no caso, as novelas. Assim, Naro (1981) levanta a hipótese de que este efeito no plano individual poderia ser ampliado para a comunidade por meio das faixas etárias mais jovens, que é o que estamos observando nos dados do PEUL da década de 2000.¹³

Um outro trabalho procura comparar achados mais recentes quando da comparação entre três variedades do português (PB, português europeu e português de São Tomé e Príncipe) com os achados anteriores sobre o fenômeno da concordância, fazendo uma breve discussão sobre o papel desse fenômeno na formulação de parâmetros gramaticais que caracterizem cada variedade do Português:

Embora não constitua objetivo do presente artigo¹⁴, é fundamental que se comparem os dados da presente pesquisa com os de outras investigações sociolinguísticas com amostras do PE, sobretudo no âmbito do sintagma verbal. Trabalhos como os de Varejão (2006) e Monguilhot (2010) registram percentuais gerais de não marcação mais altos do que os verificados nesta pesquisa [...] que permitiriam postular a existência de uma regra variável em determinadas variedades do PE. Por ora, uma observação preliminar desses trabalhos sugere que tenha havido diferença tanto na coleta de dados considerados variáveis, quanto na interpretação de determinadas estruturas como reveladoras de falta de concordância.

Nesse sentido, considera-se apropriada a hipótese de Naro & Scherre (2007, p. 85) da existência, mesmo no Português Europeu contemporâneo, de padrões variáveis em diversas regiões de Portugal: [...] É preciso ter

¹³ Cf. Paiva e Paredes Silva, neste volume.

¹⁴ Cf. Brandão e Vieira, neste volume.

claro, por exemplo, que não se dispõe de trabalhos variacionistas baseados em amostras do PE.

(...) Voltando aos dados da presente pesquisa, há indícios de que é a atuação de determinados fatores externos que acelera ou detém a implementação das restrições linguísticas. Nas realidades brasileira e são-tomense, ao que tudo indica, o valor social que se atribui à não concordância seria compatível com um estereótipo (nos termos de Labov (1972)) ou, na melhor das hipóteses, com um marcador. Por essa razão, o fenômeno variável funciona como nítido identificador do perfil social do usuário da língua. No PST – que configura um caso de variedade do português em formação, com normas objetivas ainda não totalmente definidas e assumidas pelos diferentes grupos de falantes –, sabe-se que dominar a concordância segundo a norma culta que lhe serve de modelo, a do Português Europeu, seria um dos índices de identificação de pleno domínio da Língua Portuguesa. No PB, ao contrário, o uso ou não da concordância canônica é uma opção que está claramente no nível de consciência dos indivíduos, de modo que é interpretado e funciona como forte traço caracterizador de classes sociais. Na realidade europeia, os dados desta pesquisa não permitem inferir diferença de comportamento vinculada a qualquer variável social, o que faz crer que a concordância pode constituir um parâmetro gramatical naturalmente assumido pelos falantes.

Se nos voltamos para as descrições recentes do português paulista, vislumbramos resultados interessantes: o primeiro mostra a interação entre motivações internas e externas, fundamentais para a compreensão da marcação de número em estruturas predicativas:

A atuação do Princípio de Paralelismo Formal, na variedade estudada, sofre severa restrição da condição social dos usuários da variedade, especialmente, a formação escolar; desse modo, ao menos no que tange à variedade investigada, não é possível dar a esse princípio o poder explanatório que lhe atribui Labov (1994), isto é, o de que a pluralidade tem uma motivação essencialmente interna, derivada da própria configuração formal do sistema gramatical, que governa, neste caso específico, a variedade riopretense. Por essa razão, a explicação mais plausível para a marcação de pluralidade nos predicativos é a de que há motivações em competição, nos termos de Du Bois (1985) e é a marcação de pluralidade o “bem limitado”, pelo qual forças múltiplas, as motivações formais ou internas e funcionais ou externas, competem entre si.¹⁵

O segundo, sumarizado no quadro reproduzido abaixo, revela o comportamento da concordância verbal e da alternância pronominal na fala do interior paulista em consonância com resultados de pesquisa sobre outras variedades do PB.

¹⁵ Cf. Camacho e Salomão, neste volume.

Quadro 7: Características dos fenômenos variáveis relacionados à concordância verbal e à alternância pronominal no português brasileiro do interior paulista.¹⁶

VARIEDADE FENÔMENOS	PORUGUES BRASILEIRO DO INTERIOR PAULISTA BANCO DE DADOS IBORUNA
	COMPORTAMENTO / VARIAVEIS ATUANTES
ALTERNANCIA PRONOMINAL DE 1^a PESSOA DO PLURAL	Variável parallelismo discursivo > saliência fônica > grau de determinação do sujeito > tempo e modo verbal > escolaridade > faixa etária.
CONCORDANCIA VERBAL COM NÓS	Variável saliência fônica > parallelismo discursivo > explicitude do sujeito > escolaridade > faixa etária.
CONCORDANCIA VERBAL COM A GENTE	Semicategórica parallelismo discursivo > saliência fônica > grau de determinação do sujeito > explicitude do sujeito > faixa etária.
CONCORDANCIA VERBAL DE 3^a PESSOA DO PLURAL	Variável parallelismo discursivo > saliência fônica > parallelismo oracional > traço semântico do sujeito > posição do sujeito > escolaridade > faixa etária > gênero.

Como vimos, o debate instaurado pelos diferentes artigos sobre esse tema específico da concordância reforça a tese do início dessa apresentação: a de que o campo dos estudos sociolinguísticos e dialetológicos no Brasil encontra-se em “efervescência”,¹⁷ produzindo múltiplas e panorâmicas fotografias da realidade sociolinguística no Brasil e de outras comunidades de fala de Língua Portuguesa.

Não é nosso objetivo nesta apresentação elencar todos os temas abordados ao longo dos artigos, mas vale a pena registrar a recorrência de alguns deles na agenda da pesquisa sociolinguística e dialetológica brasileira. No plano fonológico, destacam-se: a supressão da semivogal [y] no ditongo decrescente [ey]; o alcantamento de vogais pretônicas, de vogais postônicas finais e não finais; o apagamento da vogal postônica não final; a redução de ditongos nasais átonos; a

¹⁶ Cf. Rúbio e Gonçalves, neste volume.

¹⁷ Ver, a esse respeito, caderno de resumos do II Congresso Internacional de Dialetologia e Sociolinguística, realizado em setembro último, na Universidade Federal do Pará, sob o tema “Diversidade linguística e Políticas de ensino”, tendo como homenageada a Profa. Dra. Vandersi de Andrade Aguilera, coautora neste volume temático. O congresso teve 1.000 (hum mil) inscritos. (Cf. <http://www.cids.ufpa.br>. Acesso em out. 2012).

epêntese vocálica em grupos consonantais; fenômenos de juntura vocabular, como degeminação, elisão e ditongação; a variação de líquidas laterais, de vibrantes prevocálicas, intervocálicas e pós-vocálicas, de oclusivas dentais /t/ e /d/, diante de [i], de /s/ posvocálico; a monotongação de ditongos decrescentes; a ditongação em contextos seguidos de sibilantes; a haplologia sintática em sequência de sílabas com /t/ e /d/ subjacentes; a distribuição do <s> em coda silábica. No nível morfossintático, destacam-se: a realização variável da preposição núcleo dos complementos dativos; a variação no preenchimento da posição de sujeito e de objeto direto anafórico; a alternância pronominal entre *nós* e *a gente*, entre *você/cê/ocê* e entre *tu* e *você*; a alternância entre futuro do pretérito e pretérito imperfeito; a expressão variável do futuro do presente; a redução de gerúndio; a variação sintática das orações adverbiais finais; a expressão gramatical do imperativo; a alternância indicativo/subjuntivo; a ausência/presença de artigo diante de antropônimos e de possessivos; o uso variável da concordância verbal e nominal, dentre tantos outros.

É também importante ressaltar que há, ao longo de todo este volume temático, uma constante discussão sobre os métodos de trabalho de campo e sobre as relações entre abordagens quantitativas e qualitativas no tratamento da variação linguística. Os convergentes e/ou divergentes posicionamentos em relação ao tema da concordância já indiciam essa condição inerente aos sociolinguistas: a de busca incansável do “melhor método” de conduzir e de formatar um estudo sobre determinado fenômeno variável. É nesse sentido, por exemplo, que todos os artigos acabam, de uma forma ou de outra, tematizando questões metodológicas. Uma primeira questão tematizada é a própria validade da construção de bancos de dados:

A documentação da língua falada é fundamental para que os estudiosos da linguagem tenham onde buscar comprovação empírica para suas hipóteses e teorias. Nesse sentido, parece-nos que um banco de dados de fala tenha relevância implícita, independentemente da metodologia de coleta e de organização dos dados. Entretanto, dado que as abordagens em termos de coleta de dados e de foco da investigação linguística vão se aperfeiçoando e se modificando continuamente, cremos que cabe a discussão sobre a questão da relevância de um banco (...).¹⁸

Uma segunda questão tematizada é a relação entre os métodos de coleta dialetológicos e os métodos de coleta sociolinguísticos:

A “nouvelle géolinguistique”, a que se refere Thun, busca unir ao princípio da diatopia, pelo qual se vem regendo a Dialetologia desde os seus primórdios, princípios sociolinguísticos que, juntando-se ao primeiro, vão ensejar cartas linguísticas que permitem ao leitor não só saber onde se diz tal coisa, mas que tipo de falante — homem-mulher, jovem-velho, escolarizado-não escolarizado — é responsável por aquele enunciado.

¹⁸ Cf. Collischonn e Monaretti, neste volume.

Para alcançar a abrangência desejada, de acordo com os parâmetros metodológicos da Geolinguística Pluridimensional, o Projeto ALiB insere, ao lado da pesquisa da variação diatópica, também a de outras dimensões variacionais, como a diageracional, a diagenérica, a diastrática, a diafásica e a diarreferencial.

Do ponto de vista diatópico, a rede, constituída de 250 pontos, distribui-se, geograficamente, por todo o território nacional, levando em conta a densidade demográfica de cada região e de cada Estado, com os ajustes necessários nos casos de áreas com densidade muito baixa, como a região norte e centro-oeste, ou demasiadamente elevada, como na região sudeste, principalmente em São Paulo. E, na seleção das localidades, além da distribuição espacial, consideram-se a importância de cada uma delas na região e no Estado e dados de sua formação, além de limites interestaduais e internacionais. Incluem-se cidades de grande e médio porte, inclusive as capitais de Estado — à exceção do Distrito Federal e de Palmas, no Tocantins, cidades de formação relativamente recente nas quais ainda não se pode dispor de três gerações nascidas e criadas na área e com idade requerida para a seleção de informantes.”

Para atender à variação social, os informantes, em número de quatro em cada ponto — exceto nas capitais de Estado, onde são oito —, estratificam-se quanto ao gênero — quinhentos e cinquenta homens e quinhentas e cinquenta mulheres —, à faixa etária — a primeira de 18 a 30 anos e a segunda, de 50 a 65 anos — e, nas capitais, ao nível de escolaridade — fundamental e universitário.¹⁹

Uma terceira questão tematizada é a da inserção de uma parte significativa dos trabalhos sociolinguísticos brasileiros no interior da chamada primeira onda de estudos linguísticos (“a era dos inquéritos”):

No Brasil, os estudos quantitativos com bancos de dados estratificados de acordo com características sociodemográficas amplas têm se consolidado como modelo hegemônico, com os bancos de dados do Programa de Estudos sobre o Uso da Língua (PEUL) [...] A partir deste, foram replicados projetos em diferentes regiões do Brasil, com adaptações em sua metodologia [...] Esse tipo de banco de dados possibilita captar tendências amplas de variação e mudança em uma comunidade de fala. Implica, entretanto, a homogeneização da amostra [...] É importante destacar que a elaboração desses *corpora* permitiu a descrição do português brasileiro em diferentes aspectos linguísticos e considerando distintas variedades. De algum modo, têm-se uma descrição da variação na(s) gramática(s) do português do Brasil envolvendo diferentes fenômenos e a correlação destes com variáveis sociais.²⁰

¹⁹ Cf. Cardoso e Mota, neste volume.

²⁰ Cf. Freitag, Martins e Tavares, neste volume.

Na comparação entre os estudos das primeira e segunda ondas com os da terceira onda, chama-se atenção para o fato de que os primeiros “têm como foco a descrição da estrutura – um retrato estático”. Já os estudos de terceira onda “incorporam a dinamicidade da estrutura”; assim, não se estaria “negando a estrutura, mas sim enfatizando o papel da estrutura no condicionamento da prática paralelamente ao papel da prática na produção e reprodução da estrutura, a fim de captar com mais detalhes a dinâmica do valor social das variáveis”.²¹ Vale a pena acompanhar a argumentação abaixo no que diz respeito à necessidade de ser revista a prática brasileira de constituição de banco de dados sociolinguísticos:

Como vimos destacando, bancos de dados linguísticos têm sido fonte privilegiada para a descrição do português brasileiro e a tendência recente dos estudos de terceira onda ratifica sua importância para apontarem tendências linguísticas na comunidade. Seguindo as premissas da confiabilidade e da intersubjetividade (BAILEY; TILLERY, 2004), para dar continuidade a essa prática produtiva, novos bancos de dados precisam conservar minimamente as estratificações dos bancos de dados já existentes, pois a comparação de dados em tempo real permite análises mais acuradas com estudos de painel e de tendência (LABOV, 2001). Assim, as variáveis demográficas amplas – sexo, idade, escolarização etc. – precisam continuar a ser controladas nas novas coletas e nas novas amostras constituídas; é desejável, entretanto, que as novas coletas aprimorem o controle do falante, suas características individuais e de práticas, de modo a permitir que se construa um perfil social que contemple indicadores sociodemográficos mais amplos e abstratos.²²

Como vimos até aqui, muitas e complexas são as questões envolvidas na produção de um conhecimento sociolinguístico e/ou dialetológico. Nessa apresentação, procuramos fazer nosso leitor reconhecer aquelas que estão na base de nossas preocupações: Quais fenômenos variáveis estudar? De que forma fazer isso? Com que hipóteses? Com que ferramentas? A nosso ver, ao fim e ao cabo, o mais importante é o fato de que este volume conseguiu reunir um conjunto bastante representativo dos trabalhos dos dois campos acima referidos. Talvez sejam os campos de estudos, junto com o campo da antropologia, que mais coloque em relação direta e momentaneamente próxima (com todas as injunções éticas que disso decorrem) o pesquisador e o pesquisado. Por isso mesmo é que talvez tenham se desenvolvido, justamente nesses campos, tantos instrumentos específicos de coleta e registro de dados, tantos cuidados com a forma de se apresentar os resultados das pesquisas, tanta discussão a respeito da melhor maneira de “tratar” os dados. No entanto, não deixa de ser encantador e

²¹ Cf. Freitag, Martins e Tavares, neste volume.

²² Cf. Freitag, Martins e Tavares, neste volume.

rico para nós, pesquisadores, o carinho e a atenção que podemos receber, mesmo que apenas momentaneamente, daqueles que são nossos sujeitos de pesquisa, cujas falas se transformarão em *loci* de observação e de rigoroso controle. É com o depoimento abaixo que já vamos nos encaminhando para o fim desta apresentação, numa homenagem a todos aqueles que contribuem tanto para que o nosso trabalho possa ser feito e para que as fotografias, uma vez reveladas, possam se revestir de algum significado social mais concreto para eles e para toda a comunidade.

Na linha da compreensão do projeto, outro inquérito vem para ilustração. Fazia-se a documentação do informante faixa etária II, portanto dos 50 anos para cima, homem de escolaridade fundamental. Foi entrevistado na sala da sua casa, casa simples nos arredores de Recife, ato presenciado, além dos partícipes — o informante, a inquiridora e a auxiliar de pesquisa —, pela sua esposa. A documentação dos dados corria muito bem, acabava-se de fazer a aplicação do OFF, do QSL e do QMS — como se sabe, uma espécie de pingue-pongue em que o inquiridor pergunta e o informante responde — e passava-se às questões semidirigidas, nas quais se pede ao informante o relato de fatos sobre temas orientados. Naquele exato momento, se desgruda da sala a esposa do informante que volta, incontinente, com um sanduíche quente e um suco de laranja, e sussurra ao ouvido da inquiridora: "Estou vendo que agora é ele que fala mais e aí a senhora pode comer uma coisinha". Compreensão da pesquisa, respeito ao método, entendimento perfeito da mecânica do interrogatório linguístico, uma solidariedade ilimitada ao trabalho dialetal. É preciso que se diga: acabado o inquérito, ao entrarem no táxi inquiridora e auxiliar, aquela não pode conter as lágrimas da emoção seguradas no final da gravação. São essas e outras lições que o trabalho de campo permite ao pesquisador aprender ou reafirmar o seu aprendizado".²³

E, para finalizar, então, não podemos deixar de registrar aqui o ano do jubileu de ouro da *Alfa – Revista de Linguística*, ocasião, para nós, de alegria e honra imensas por nos ter sido concedida a oportunidade de reunir neste volume temático a contribuição de 43 pesquisadores e de 23 pareceristas anônimos das mais diferentes partes do Brasil, números que, seguramente, refletem a intensa movimentação científica do campo dos estudos sociolinguísticos e dialetológicos do/no Brasil.

Parabéns à *Alfa – Revista de Linguística*! Parabéns a todos os autores reunidos neste volume!

²³ Cardoso e Mota, neste volume.

SUMÁRIO / CONTENTS

■ Apresentando retratos sociolinguísticos e dialetológicos do Brasil <i>Sebastião Carlos Leite Gonçalves e Anna Christina Bentes</i>	723
■ Cumprindo uma pauta de trabalho: contribuições recentes do Peul <i>Maria da Conceição de Paiva e Vera Lúcia Paredes Silva</i>	739
■ Projeto portvix: a fala de Vitória/ES em cena <i>Lilian Coutinho Yacovenco, Maria Marta Pereira Scherre,</i> <i>Leila Maria Tesch, Marcela Langa L. Bragança,</i> <i>Elaine Meireles Evangelista, Alexandre Kronemberger de Mendonça</i> <i>Elba Nusa Calmon, Heitor da Silva Campos Júnior</i> <i>Astrid Franco Barbosa, Jucilene Oliveira Sousa Basílio</i> <i>Carlos Eduardo Deoclécio, Janaína Biancardi da Silva</i> <i>Aline Tomaz Fonseca Berbert, Samine de Almeida Benfica</i>	771
■ Projeto de variação linguística no centro-oeste (VALCO) <i>Rachel do Valle Dettoni, Cíntia da Silva Pacheco,</i> <i>Carolina Queiroz Andrade, Maria Marta Pereira Scherre</i>	807
■ Banco de Dados VARSUL: a relevância de suas características e a abrangência de seus resultados <i>Gisela Collischonn, Valéria de Oliveira Monaretto</i>	835
■ Projeto atlas linguístico do Brasil: antecedentes e estágio atual <i>Suzana Alice Cardoso e Jacyra Andrade Mota</i>	855
■ Para um atlas pluridimensional: pesquisas e pesquisadores <i>Vanderci de Andrade Aguilera e Fabiane Cristina Altino</i>	871
■ DIscutindo a dimensão sociolinguística do projeto alib: uma reflexão a partir do perfil dos informantes <i>Aparecida Negri Isquierdo e Valter Pereira Romano</i>	891

■ Bancos de dados sociolinguísticos do português brasileiro e os estudos de terceira onda: potencialidades e limitações <i>Raquel Meister Ko. Freitag, Marco Antonio Martins e</i>	917
<i>Maria Alice Tavares.....</i>	
■ Vogais na amazônia paraense <i>Regina Cruz</i>	945
■ O paulistano no mapa sociolinguístico Brasileiro <i>Ronald Beline Mendes e Livia Oushiro</i>	973
■ A fala do interior paulista no cenário da sociolinguística brasileira: panorama da concordância verbal e da alternância pronominal <i>Cássio Florêncio Rubio e Sebastião Carlos Leite Gonçalves</i>	1003
■ Concordância nominal e verbal: contribuições para o debate sobre o estatuto da variação em três variedades urbanas do português <i>Silvia Figueiredo Brandão e Silvia Rodrigues Vieira.....</i>	1035
■ Motivações em competição na variação sociolinguística: o plural dos predicativos na variedade de São José do Rio Preto <i>Roberto Gomes Camacho e Mircia Hermenegildo Salomão.....</i>	1065
■ Da fala para a leitura: análise variacionista <i>Dermeval da Hora e Maria de Fátima S. Aquino</i>	1099
■ Palatalização das plosivas alveolares em flores da cunha (rs): variação linguística e práticas sociais <i>Elisa Battisti e Adalberto Ayjara Dornelles Filho.....</i>	1117
■ ÍNDICE DE ASSUNTOS.....	1151
■ SUBJECTS INDEX.....	1153
■ ÍNDICE DE AUTORES / AUTHORS INDEX.....	1155
■ NORMAS PARA APRESENTAÇÃO DE ORIGINAIS.....	1157

ARTIGOS ORIGINAIS /
ORIGINAL ARTICLES

CUMPRINDO UMA PAUTA DE TRABALHO: CONTRIBUIÇÕES RECENTES DO PEUL

Maria da Conceição de PAIVA *
Vera Lúcia PAREDES SILVA **

- RESUMO: O grupo PEUL (Programa de Estudos sobre o Uso da Língua) constituiu-se nos anos 80 do século passado, em torno de uma meta comum: a de desvendar a sistematicidade da variação e mudança atestadas na variedade carioca. Inicialmente, mais concentrados na modalidade falada, os estudos ganharam novo fôlego, particularmente do ponto de vista da mudança, com a volta à comunidade de fala cerca de 20 anos mais tarde, o que permitiu a realização de análises da mudança em tempo real de curta duração. Posteriormente, os estudos foram estendidos à escrita monitorada, representada por diversos gêneros do domínio jornalístico. Esta ampliação do *corpus* permitiu comparar os princípios que regem a variação nas duas modalidades e testar hipóteses relativas à migração/incorporação de mudanças em curso na fala para a escrita. Neste artigo, mostramos de que forma os trabalhos realizados pelos pesquisadores ligados ao grupo contribuíram para alimentar e aprofundar discussões acerca de aspectos relevantes, como a sistematicidade de princípios subjacentes ao uso de formas variáveis, e desvendar o caminho trilhado pelas variantes linguísticas em tempo real e sua incorporação na escrita. Destacamos, ainda, a forma como a colocação contínua de novas questões abriu o espaço para trabalhos na interface com outras áreas do conhecimento.
- PALAVRAS-CHAVE: Sociolinguística. Língua falada. Língua escrita. Gêneros textuais. Aplicações.

Histórico

O grupo de pesquisa Programa de Estudos sobre o Uso da Língua (PEUL) foi criado no final dos anos 70 por pesquisadores interessados, num primeiro momento, em desvendar a heterogeneidade observada na variedade carioca e depreender os padrões de variação linguística e possíveis mudanças em curso nessa comunidade. Naquele momento, esse objetivo exigia, antes de mais nada, a constituição de uma amostra estratificada da comunidade visada, de forma a identificar a contraparte social da variação e mudanças em curso. Assim, foi

* UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras – Departamento de Lingüística. Rio de Janeiro – RJ – Brasil. 21941-917 - paiva@club-internet.fr

** UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras – Departamento de Lingüística. Rio de Janeiro – RJ – Brasil. 21941-917 - veraparedes@terra.com.br

organizado, de forma pioneira no Brasil, um *corpus* constituído de 64 entrevistas realizadas com falantes cariocas homens e mulheres, distribuídos por quatro faixas etárias e três níveis de escolaridade. Essa amostra representa, ainda, a configuração social diversificada do Rio de Janeiro, ao incluir representantes dos diferentes bairros da cidade e de diversos estratos sociais.¹

A análise sistemática da Amostra Censo 1980, a partir dos pressupostos centrais da Sociolinguística Variacionista (WEINREICH; LABOV; HERZOG, 2006; LABOV, 1972, 1981, 1982, 1994), tornou possível identificar e explicar fenômenos variáveis nos diversos níveis da língua e levantar hipóteses acerca de mudanças em curso na fala carioca. Processos bem conhecidos, como a variação entre **nós/a gente**, a variação na concordância de número nos elementos do Sintagma Nominal (doravante SN), a supressão das semivogais nos ditongos decrescentes, a regência variável do verbo **ir**, a alternância entre **seu** e **dele** ou o uso de artigo antes de nomes próprios foram escrutinados, a partir de uma metodologia quantitativa rigorosa, que permitiu depreender a sistematicidade na heterogeneidade, a partir da identificação dos fatores linguísticos associados às diferentes formas de expressar o mesmo conteúdo referencial assim como os padrões de correlação social das variantes. Tais estudos vieram somar evidências favoráveis ou desfavoráveis a princípios mais gerais de variação tanto no eixo linguístico como social.²

O *corpus*, que ficou conhecido como Amostra Censo 1980, foi apenas o começo de outros empreendimentos que permitiram alargar a amostragem da fala carioca. Considerando os limites inerentes às amostras estratificadas para o estudo de aspectos interacionais, entre 1989 e 1990, foi organizado o Banco de Dados Interacionais (BDI), coordenado pela pesquisadora Cláudia Roncarati. Entre 1999 e 2000, com vistas a verificar, em tempo real, hipóteses de mudança indicadas pelos estudos em tempo aparente realizados com a Amostra Censo 1980, empreendeu-se a tarefa de revisitar a comunidade carioca. Duas novas amostras vieram enriquecer o acervo do PEUL: (a) a Amostra Censo 2000, um conjunto de 32 entrevistas realizadas com falantes cariocas, organizada segundo os mesmos parâmetros da Amostra Censo 1980, e (b) uma amostra da fala de 16 indivíduos que já integravam a primeira amostra e foram re-entrevistados.³ Associadas, as amostras Censo 1980, Censo 2000 e as entrevistas dos falantes recontactados garantiram (e garantem) o estudo do que Labov (1981, 1994) denomina estudo da mudança em tempo real de curta duração, criando as condições ótimas para a verificação mais controlada da direcionalidade de alguns processos variáveis tanto na comunidade de fala (estudo tendêncial), como no indivíduo (estudo painel).

¹ Um maior detalhamento sobre a Amostra Censo 1980 pode ser encontrado em Oliveira Silva e Scherre (1996).

² Essa primeira garimpagem se encontra reunida em Oliveira Silva e Scherre (1996).

³ Maiores detalhes sobre a Amostra Censo 2000 podem se encontrados em Paiva e Duarte (2003).

Uma consequência natural do interesse pela variação e pela implementação da mudança foi o surgimento de questões acerca da forma como novas variantes detectadas na fala são incorporadas pela escrita. Assim, no período de 2002 a 2004, os pesquisadores do grupo organizaram uma amostra da escrita monitorada, composta de textos representativos de diferentes gêneros, publicados em jornais cariocas destinados a um espectro de leitores diferenciado do ponto de vista social (*O Globo*, *Jornal do Brasil*, *Extra* e *Povo*). Novas hipóteses puderam ser exploradas e tendências mais gerais vieram à luz.⁴

Uma vez delineado um perfil sociolinguístico da variedade carioca, membros do grupo se orientaram para questões concernentes à aquisição da variação estruturada, o que, por si mesmo, requer *corpora* mais específicos, constituídos de forma a representar diferentes estágios da aquisição da linguagem. O grupo já contava com uma amostra reunida, entre 1979 e 1981, pela professora Alzira V. T. de Macedo, composta de gravações com 32 crianças de 4 a 11 anos. Mais recentemente, sob a iniciativa da pesquisadora Christina Abreu, o grupo se enriqueceu com a organização de outra amostra de dados para o estudo da aquisição da variação socialmente estruturada (AMOSTRA AQUIVAR) representativo da fala de 35 crianças com idade entre 1 ano e 9 meses e 5 anos, de diferentes estratos sociais.

Dadas as limitações inerentes às amostras aleatórias e ao próprio método de coleta de dados utilizado na maioria das amostras, entrevistas sociolinguísticas, diversos pesquisadores vêm lançando mão de amostras mais específicas que possibilitem o estudo de fatos de variação mais particulares, como o caso da amostra constituída por Paredes para o estudo do retorno do pronome **tu** na variedade carioca (PAREDES SILVA, 2003a, 2011) ou a amostra de Gryner (1990), mais centrada no discurso argumentativo.

Ao longo dos mais de trinta anos em que vem atuando no cenário da Sociolinguística brasileira, o grupo PEUL contribuiu com inúmeros trabalhos voltados para a variação inerente à gramática graças à dedicação de muitos pesquisadores, para quem o uso e a dinâmica da língua estão no centro dos estudos linguísticos. Detalhes relativos à atuação do grupo podem ser encontrados em Paiva e Scherre (1999) e em Scherre e Roncarati (2008).

Procurando estimular a expansão da Sociolinguística no Brasil, os membros do grupo atuam incessantemente na formação de novos pesquisadores, seja no nível de Iniciação Científica, seja nos níveis de Mestrado e Doutorado, que, inseridos no grupo, despertaram ou consolidaram seu interesse pela Sociolinguística Variacionista. A composição do grupo foi se alterando ao longo do tempo, o que

⁴ As amostras Censo 1980, Censo 2000 e a amostra de textos midiáticos já estão disponibilizadas *on-line* no site <www.letras.ufrj.br/PEUL>. Atualmente, sob coordenação da professora Conceição Paiva, está sendo preparada a disponibilização dos registros sonoros correspondentes às amostras de fala.

permitiu não só alargar o círculo de curiosos em relação aos padrões de variação como também abrir novas perspectivas e explorar diferentes interfaces, tais como variação e discurso, variação e aquisição, variação e ensino do português como L1. Os últimos anos se caracterizaram pela abertura de novas frentes de trabalho, explorando a interface com áreas como a da Matemática (LEAL; MOLLICA, 2009, 2010, [20--?]; MOLLICA; LEAL, 2007, 2008, 2010, 2012) ou da Ciência da Informação (MOLLICA; GONÇALEZ, 2012; MOLLICA; LISBOA, 2011), possibilitada pela parceira com outros pesquisadores, especialmente das áreas de Matemática e de Ciência da Informação. Destacam-se, ainda, os esforços envidados no âmbito da Sociolinguística Educacional, liderado pela pesquisadora Cecília Mollica, no sentido de traduzir os resultados das pesquisas variacionistas em subsídios para a qualificação de professores alfabetizadores e docentes do português como L1.

Naturalmente, o acúmulo de evidências acerca dos padrões de variação e mudança na fala carioca contribuiu de forma significativa para a discussão de diversos pontos que sempre estiveram e continuam na pauta da Sociolinguística Variacionista. Dentre eles, destacamos:

- a) Restrições sobre variantes linguísticas podem ser traduzidas em princípios mais gerais?
- b) De que formas se inter-relacionam diferentes processos de mudança em curso numa mesma comunidade de fala?
- c) Como estabelecer a relação entre mudança na comunidade e mudança no indivíduo? Em outros termos, em que medida a gramática do indivíduo reflete a gramática da comunidade?
- d) Qual a trajetória de incorporação pela escrita de mudanças em curso na comunidade de fala? Que princípios norteiam essa incorporação?
- e) De que forma os estudos sociolinguísticos podem contribuir para o desenvolvimento de métodos de ensino do português?

Ao longo deste artigo, procuramos sintetizar e discutir, tarefa nada simples, algumas evidências obtidas no trabalho desenvolvido nas diferentes frentes de trabalho, as perguntas que guiaram muitos estudos e algumas respostas encontradas, ao longo desses últimos 12 anos.⁵

Variação e mudança na fala

O conceito de regra variável, que emerge com os estudos de Labov (1972), representou um avanço teórico importante para a ciência linguística, na medida

⁵ Para uma retrospectiva de trabalhos desenvolvidos até o ano de 1999, ver Paiva e Scherer (1999).

em que incorpora à gramática a sistematicidade estrutural e social do componente variável da língua. O investimento constante nessa vertente de estudos por diferentes pesquisadores do PEUL tem acumulado evidências consideráveis sobre (a) a configuração de vários fenômenos variáveis no português falado no Rio de Janeiro, (b) as restrições que operam sobre cada um deles, (c) a ação regular dos diferentes princípios que subjazem ao uso das formas variáveis e regulam a implementação da mudança na comunidade de fala e no indivíduo.

Nesta seção nos detemos em aspectos relativos à mudança, a partir dos resultados obtidos com estudos em tempo real de curta duração. A análise detalhada de fenômenos variáveis distintos na comunidade de fala carioca, através da comparação das amostras Censo 1980 e Censo 2000, e da mudança no indivíduo, através da comparação das entrevistas realizadas com o mesmo falante nos anos oitenta e no começo deste século, tornou possível depreender três trajetórias distintas no percurso de algumas variantes linguísticas já identificadas com base na amostra Censo 1980: (a) contínua implementação de variantes linguísticas; (b) estabilidade da variação e (c) retração de uma variante em contextos particulares. A primeira possibilidade pode ser ilustrada pela realização variável da preposição núcleo dos complementos dativos, instanciada como: **a**, **para** ou **zero**, focalizada por Gomes (2003). A autora comprova o aumento da preposição **para**, na comunidade de fala, em todas as faixas etárias, ocupando mesmo contextos que, no início da década de 80, eram exclusivos da preposição **a**. Tal generalização se deve, segundo a autora, ao caráter mais neutro dessa preposição em oposição, por um lado, à formalidade da preposição **a** e, por outro, a um certo estigma associado à ausência da preposição.

A tendência de estabilidade encontra ilustração na variação entre variante plena e variante nula na realização do sujeito, fenômeno focalizado por Duarte (1995, 2003a) e Paredes Silva (2003b). Embora partindo de perspectivas teóricas distintas, as duas pesquisadoras mostram índices equivalentes de sujeito preenchido nas duas amostras, sem alterações significativas ao longo do tempo. Situação similar é encontrada por Omena (2003) para a alternância entre **nós** e **a gente**, no estudo do tipo tendência. Destaque-se que essa tendência de estabilidade é, segundo as indicações depreendidas em tempo real de curta duração, independente do possível estigma social da variante focalizada, como mostra o trabalho de Mollica (2003b) para as relativas copiadoras.

A terceira possibilidade, regressão de uma variante, é detectada por Paiva (2003) para a supressão da semivogal [y] no ditongo decrescente **[ey]**. O processo é particularmente favorecido em dois contextos: o de flap (**feira/fera**) e de consoantes palatais (**beijo/bejo**). Analisando os dois contextos separadamente, a autora atesta direções contrárias para cada um deles: no contexto de flap, o processo se configura como mudança concluída; no contexto de palatais, ao

contrário, a taxa de monotongação de [ey] se reduz sensivelmente. Tal situação sinaliza a necessidade de uma análise mais fina dos contextos de implementação da mudança e permite colocar em causa o princípio de que uma mudança opera unicamente por um aumento constante de taxas em todos os contextos em princípio favoráveis.

Ademais, esses estudos trazem evidências robustas para a discussão das relações entre a mudança na comunidade e no indivíduo, uma questão central no estudo da mudança. Assim, Paiva (2003), por exemplo, mostra que a tendência à redução do processo de supressão de [y] no contexto de palatal se reflete de forma diferenciada no comportamento dos 16 indivíduos recontactados no final de década de noventa e no início de 2000. Se a grande maioria dos falantes se conforma ao recuo na redução de [ey] nesse contexto, uma pequena parte tende a se manter estável ou aumenta o índice da variante monotongada. Falantes mais inovadores convivem, portanto, com falantes mais conservadores.

Uma situação semelhante é verificada por Omena (2003) para a alternância entre **nós** e a **gente**, um processo para o qual as evidências diacrônicas de mudança são mais seguras. Alguns falantes manifestam larga implementação de **a gente**, alcançando uso categórico dessa variante. Observa-se, no entanto, nítida divergência na trajetória de **a gente** quando se consideram os indivíduos que variam. Apenas seis, do total de dezesseis falantes recontactados, apresentam, como se poderia esperar, aumento nas taxas de **a gente** e alguns falantes, surpreendentemente, aumentam o uso de **nós**. Tais evidências permitem problematizar o princípio sustentador da análise da mudança em tempo aparente, ou seja, o de estabilidade do sistema linguístico adquirido até a adolescência (LIGHTFOOT, 1999). Se tomarmos como base a alternância entre **nós** e **a gente** para a expressão da 1^a pessoa do plural, essa hipótese poderia ainda se manter, com base no argumento de que o uso de uma ou outra variante afeta unicamente os dados primários a que são expostos os indivíduos sem alterar a gramática nuclear. A maior resistência a mudanças que envolvam a gramática nuclear encontra evidências nos resultados encontrados por Duarte (2003a) para o preenchimento do sujeito. No caso em questão, uma possível mudança na marcação do parâmetro do sujeito nulo, observa-se convergência acentuada no comportamento do indivíduo e da comunidade. A questão permanece, portanto, em aberto e requer evidências de outros fenômenos relacionados a essa mudança em progresso, que serão brevemente referidos ainda nesta seção.

A própria metodologia e as conclusões autorizadas pelos agrupamentos etários, em estudos do tipo tendência, podem obscurecer aspectos da relação entre mudança no indivíduo e na comunidade. Os estudos desenvolvidos mais recentemente pelos pesquisadores Anthony Naro e Marta Scherre sobre a concordância nominal e a concordância verbal vão nesta direção. Os primeiros

estudos em tempo real de curta duração fornecem indícios claros de aumento das marcas de concordância verbal e nominal tanto no indivíduo como na comunidade de fala carioca (NARO; SCHERRE, 2003, 2010). Reexaminando essa tendência mais recentemente (SCHERRE; NARO, 2011), os dois autores retomam a questão do efeito de variáveis sociais sobre a presença das marcas de concordância entre verbo e sujeito e em todos os elementos do SN, na comunidade de fala carioca, com o intuito de evidenciar como os grupos de falantes mudam no tempo. Os autores ponderaram que a própria metodologia de análise do efeito da variável tempo, baseada na comparação direta de grupos semelhantes, pode não ser a forma mais adequada de apreender padrões de mudança no interior de uma comunidade, porque os indivíduos que compõem os conjuntos (grupos etários, por exemplo) não são os mesmos. Segundo a análise dos autores, uma reorganização dos grupos etários permite identificar, de forma mais clara, os padrões de distribuição de variantes linguísticas por diferentes grupos etários.

As faixas etárias das amostras Censo 1980 e Censo 2000 foram previamente organizadas em quatro grupos: 7-14 anos, 15-25 anos, 26-49 anos e acima de 50 anos. Considerando o intervalo de tempo que separa as duas amostras, os autores redistribuíram os falantes, deslocando a amostra Censo 1980 dezoito anos, para comparar grupos etários mais diretamente comparáveis (*age cohorts*), extraídos dos mesmos universos etários.

Essa reorganização fornece indícios mais claros de que, embora todos os grupos etários tenham aumentado sua frequência de marcas de concordância, o padrão de correlação já observado, na amostra Censo 1980, para cada um dos fenômenos de concordância, se mantém na segunda sincronia: graduação etária para a concordância nominal e perda na concordância verbal. Essa similaridade é indicativa de que os falantes de todas as idades se movem, mas sujeitos às restrições impostas por padrões sociais do grupo em que estão inseridos. O padrão é claramente modificado somente para os falantes que ainda não haviam nascido na década de 80, que revertem seu comportamento em direção a maior uso de concordância, com percentuais e pesos relativos mais altos.

Usando os agrupamentos de idade projetados no lugar das faixas etárias originalmente estabelecidas, Naro e Scherre (2011) puderam, então, ver mais claramente as mudanças dentro da comunidade de fala e confirmar tanto o modelo de fluxos e contrafluxos, proposto inicialmente por Naro (1981) e retomado por Naro (1999) e Naro e Scherre (2010), como a ampliação da aquisição da concordância do plano individual para a comunidade como um todo. Testemunha-se, assim, a previsão de Naro (1981), baseada na análise da Amostra MOBRAL⁶ (NARO, 1981), no que diz respeito ao efeito da variável

⁶ A amostra denominada MOBRAL é um conjunto de entrevistas gravadas com 20 falantes em processo de alfabetização no início da década de 70 e que participavam dos cursos oferecidos pelo Movimento Brasileiro de

orientação cultural sobre a concordância verbal. Os falantes que se identificavam com valores da classe média, os de orientação vicária, favoreciam mais a concordância verbal. Este traço foi mensurado por meio do tipo de programa de televisão visto, no caso, as novelas. Assim, Naro (1981) levanta a hipótese de que este efeito no plano individual poderia ser ampliado para a comunidade por meio das faixas etárias mais jovens, que é o que estamos observando nos dados do PEUL da década de 2000.

Um estudo do tipo painel confirma a tendência identificada no estudo da comunidade (NARO; SCHERRE, 2011). Considerando cada um dos momentos de tempo (década de 1980 e década de 2000) como uma variável independente para cada falante, os autores concluem que, num intervalo entre 16 e 18 anos, todos os falantes aumentaram os seus índices de concordância, embora em magnitudes distintas. Essa tendência é, em parte, independente de o falante ter aumentado ou não o seu nível de escolaridade. No entanto, aqueles que o aumentaram, apresentam um ganho sistemático e mais contundente de concordância. Constatam, ainda, que o aumento da concordância nominal entre falantes que não mudaram o nível de escolarização era menos saliente do que o aumento de concordância verbal, interpretado pelos autores como possível consequência da configuração linguística de cada um dos fenômenos e das restrições resultantes da saliência fônica: itens de oposição singular/plural mais salientes são mais frequentes na concordância verbal, o que propicia a sua maior percepção.

A concordância nominal, por sua vez, um tanto paradoxalmente, separa mais nitidamente os indivíduos em função da escolarização. Os autores interpretam este fato como fruto de a concordância nominal apresentar uma relação sintagmática menos complexa, entre elementos de um mesmo sintagma, e facilitar, assim, sua maior generalização em áreas urbanas, quando percebida. Tudo leva a crer, portanto, que, na variedade urbana carioca, aspectos cognitivos operam juntamente com aspectos sociais, intensificando seu efeito por meio de dimensões estruturais que operam sobre a variação e a mudança. Interpretações à parte, é fato indubitável que a análise em tempo real de curta duração revela o aumento da concordância de número, variável que envolve estigma social.

O que se destaca nos diversos estudos em tempo real de curta duração é a sistematicidade no controle estrutural da variação. O já citado estudo de Paredes Silva (2003b), por exemplo, demonstra claramente a primazia da variável conexão discursiva nas chances de preenchimento do sujeito, nos dois momentos de tempo considerados. O preenchimento do sujeito, regularmente nas duas sincronias, está nitidamente correlacionado a uma escala no grau de conexão entre as duas

Alfabetização. Neste estudo, Naro (1981) verificou que os falantes que se identificavam mais com os valores da classe média, os de orientação vicária, revelavam maior uso de concordância verbal. Este fato foi mensurado por meio do tipo de programa visto, no caso, as novelas.

orações: quando maior o estreitamento sintático-discursivo entre as orações (caso de persistência de um referente na mesma função sintática e sem mudança de tópico ou de plano discursivo), maior a probabilidade de apagamento do SN sujeito.

Essa regularidade no efeito de fatores estruturais pôde ser constatada igualmente em fenômenos socialmente estigmatizados, como é o caso das relativas copiadoras ou da concordância nominal e verbal. Como confirma Mollica (2003b), a retomada do constituinte relativizado por um pronome é controlada pelo traço [+humano] do referente: referentes com o traço [+ humano] aumentam as chances de uma relativa copiadora. No caso da concordância nominal e verbal destaca-se, por exemplo, o efeito regular de variáveis como saliência fônica, posição relativa, paralelismo discursivo e paralelismo oracional, além do traço humano para a concordância verbal (NARO, 1999; NARO; SCHERRE, 2003; SCHERRE, 2010). Essa sistematicidade é notável igualmente em fenômenos que envolvem uma nítida contraparte discursiva, como é o caso da variação entre futuro do subjuntivo e o presente do indicativo na construção condicional. Como mostra Gryner (2003), de forma regular nas duas amostras, destaca-se nitidamente a correlação entre uma escala epistêmica e o uso do futuro do subjuntivo ou do presente do indicativo: a primeira variante é significativamente mais recorrente em orações condicionais que codificam um fato possível ou eventual.

Esta constância no efeito de variáveis estruturais conduz a uma discussão mais aprofundada da forma como princípios mais gerais e de natureza distinta operam sobre a variação e a implementação de mudanças linguísticas. Retomemos, a título de exemplo, a variável traço de animacidade do referente que, como vimos acima, possui um efeito saliente tanto sobre as relativas copiadoras como sobre a concordância verbal. Diferentes trabalhos fazem sobressair o alcance explicativo desta propriedade na implementação de mudanças linguísticas e permitem, inclusive, explicar um dos pontos centrais de uma teoria da mudança, qual seja, o do encaixamento entre fenômenos aparentemente distintos, como mostram os estudos de Duarte, inseridos em uma perspectiva da Sociolinguística Paramétrica. Procurando identificar os efeitos colaterais da mudança que envolve a forma de realização do sujeito no português brasileiro, ou seja, a mudança na marcação do valor do parâmetro de sujeito nulo, a autora persegue a hipótese de que diversas mudanças em curso no PB podem ser explicadas em termos de uma hierarquia referencial, construída a partir de evidências empíricas (CYRINO, 1997; DUARTE, 2000; CYRINO; DUARTE; KATO, 2000). Considerando a expressão do sujeito, a autora assume que a mudança se implementa com maior velocidade a partir dos itens mais referenciais, aqueles com o traço inherentemente [+humano] (1^a. e 2^a. pessoas), prossegue mais lentamente naqueles em que interagem os traços [+/-humano], [+/-específico] (a 3^a. pessoa) e naqueles em que o sujeito retoma uma proposição (DUARTE, 2003a, 2012). Trata-se de sujeitos neutros,

nulos ou retomados por um demonstrativo, ou de “referência estendida” nos termos de Paredes Silva (1985)⁷. Por remeterem a porções maiores do texto, se situam em um ponto de menor referencialidade. No extremo do *continuum* de referencialidade, aparecem os sujeitos não referenciais, das chamadas sentenças impessoais, que permanecem nulos no português, embora sejam preenchidos por um expletivo em outras línguas que passaram pelo mesmo processo de mudança (DUARTE, 2003b, 2007b).

O estudo de outros fenômenos tem permitido ressaltar a importância de princípios de natureza mais funcional, como os princípios de marcação ou de iconicidade, ou de princípios relativos ao processamento na retração/expansão de mudanças linguísticas na fala, e mesmo na escrita, como veremos na seção seguinte.

A generalidade de princípios ligados a processamento é comprovada pela sua ação sobre fenômenos bastante diferenciados, como o de concordância verbal e nominal, o uso de relativas copiadoras, a posição de constituintes circunstanciais, a retomada do sujeito por um pronome, entre outros. O efeito de paralelismo discursivo, por exemplo, é amplamente comprovado e discutido por Scherre (1998) e por Scherre e Naro (1992) para os fenômenos de concordância nominal e verbal. Destaca-se, ainda, em trabalhos que focalizam alternâncias entre formas verbais, como o de Costa (2003), Tesch (2011) e Martins (2010).

Paiva et al. (2009) e Paiva (2008b) demonstram, de forma clara, as restrições impostas por um princípio de peso final (PAIVA, 2008b, 2012) sobre a disposição sintagmática de circunstanciais locativos e temporais: tendem a ocupar a margem direita da oração os constituintes circunstanciais de maior peso, ou seja, mais extensos. A restrição de que constituintes menores precedem constituintes mais extensos é mostrada ainda mais claramente nas orações em que concorrem circunstanciais de categorias semânticas distintas, como as de tempo, modo e lugar. O princípio de peso final explica uma quantidade maior da variação do que exigências decorrentes de dependência sintático-semântica. A ação desse princípio sobre constituintes argumentais pode ser atestada, por exemplo, no estudo de Gomes et al. (2003) acerca da variação na ordem dos complementos dos verbos de três argumentos.

O efeito regular de um fator como distância em diferentes fenômenos sinaliza igualmente a relevância de princípios cognitivos. O estudo das relativas copiadoras, por exemplo, mostra claramente que o referente relativizado é mais frequentemente copiado quando há maior distância entre o referente retomado e o pronome relativo. No caso dos sujeitos duplos, ou deslocamento à esquerda do sujeito, observa-se tendência similar (BRAGA; MOLLICA, 2010). É significativo igualmente na

⁷ Sujeitos que retomam uma extensão maior de texto, ou seja, uma oração ou mesmo um parágrafo e que podem ser expressos pelo pronome neutro *isso* ou pela anáfora zero, ou mesmo por um nome que funciona como rótulo ou encapsulador.

alternância entre formas nominais/pronominais nos processos de referenciação que envolvem continuidade tópica (PAREDES SILVA, 2007a, 2007b, 2008a, 2008b).

As questões colocadas até este ponto envolvem fenômenos que se conformam, pelo menos aparentemente, a uma definição de variável linguística no seu sentido laboviano clássico, isto é, a alternância entre duas ou mais formas referencialmente equivalentes. Alargando o conceito de variação, podemos recobrir igualmente casos em que uma “mesma” forma pode corresponder a significados distintos, mais importante ainda, estatutos categoriais diferentes. Desse modo, é possível tratar nos moldes da Sociolinguística Variacionista os processos de gramaticalização, isto é, mudanças em que uma mesma palavra ou construção pode sofrer, a par das mudanças semânticas e gramaticais, um desgaste em sua forma fônica. Esse tipo de variação tem despertado atenção particular de alguns pesquisadores do grupo PEUL, sendo estudado tanto na sua dimensão diacrônica como sincrônica. Ilustrações da aplicabilidade de métodos variacionistas a fenômenos dessa natureza são fornecidas pelos estudos de Braga (2003), acerca da trajetória dos itens **aí** e **então** no indivíduo; de Roncarati (2003, 2010), sobre as formas demonstrativas, especialmente a discursivização da forma **isso**; de Paiva (2001), sobre a gramaticalização da locução **por causa de que** como um conector causal, e o de Paiva e Barros (2011), sobre as construções modais **ter que + infinitivo** e **ter de + infinitivo**. Esses estudos procuram identificar as propriedades distintivas deste tipo de mudança e têm contribuído para a discussão de princípios como o da unidirecionalidade. Alguns trabalhos que procuram explicar o *continuum* sincrônico nos usos de uma mesma forma como resultado de um longo processo diacrônico (BRAGA, 2003) têm apontado que nem sempre há uma correspondência exata entre a pressuposta unidirecionalidade sincrônica e a verificada unidirecionalidade diacrônica.

Mais recentemente, esses estudos têm procurado fomentar o diálogo entre pressupostos teóricos da gramaticalização e da gramática de construções, enfatizando as restrições e motivações cognitivas capazes de explicar os processos de mudança categorial. Uma exemplificação desse diálogo pode ser encontrada nos trabalhos de Braga e Paiva (2011a, 2011b), em que as autoras traçam a diacronia das construções causais, mostrando as inter-relações entre elas ao longo dos três últimos séculos. Nessa abordagem, ganham destaque, ainda, questões ligadas à frequência de tipo (*type*) e de ocorrência (*token*) e seus efeitos sobre a representação das categorias gramaticais.

As tendências ressaltadas até este ponto são depreendidas a partir da análise de amostras estratificadas, que, embora permitam conclusões seguras acerca de padrões de correlação entre a estrutura linguística e a organização social, impõem limitações à análise de aspectos ligados, por exemplo, à inserção do falante em redes sociais (MILROY, 1980) e de processos como o de acomodação linguística,

que ganha vulto numa megalópole como o Rio de Janeiro, em que convivem falantes de diferentes origens geográficas. O investimento nessa vertente, através de trabalhos realizados pela pesquisadora Cecília Mollica, tem permitido trazer à luz a forma como esses falantes acomodam características do seu dialeto de origem e o consequente surgimento de novas variantes linguísticas. Além disso, vem fomentar a discussão de questões concernentes a atitudes e crenças linguísticas, resistência e assimilação, orientação para a identidade ou para o prestígio (LABOV, 1972).

Variação e mudança na escrita

Como já foi dito na introdução deste artigo, no período de 2002-2004 os pesquisadores do grupo PEUL se dedicaram à constituição de um acervo de língua escrita em que se contemplaram diversos gêneros do domínio jornalístico. O objetivo era, com base nesse *corpus*, investigar, de forma sistemática, fenômenos variáveis já analisados na fala do Rio de Janeiro, em uma amostra que podemos caracterizar como um padrão de escrita vigente na época. Ao mesmo tempo, a destinação a um público-leitor diferenciado⁸ permitiria avaliar até que ponto haveria correlações de ordem social no uso de uma modalidade de língua, de qualquer forma, mais monitorada do que a fala de nossas entrevistas.

As análises deste material foram orientadas por várias questões, destacando-se dentre elas:

- (a) Estaria, de fato, a variação na escrita correlacionada às mesmas restrições encontradas na fala?
- (b) Até que ponto as diferenças decorrentes do meio de produção poderiam se refletir nos efeitos associados às variáveis estruturais?
- (c) Considerando-se fala-escrita como um *continuum*, qual a interferência dos diferentes gêneros textuais nos processos de variação e mudança? Admitindo que a migração de processos de mudança para a escrita seja gradativa, quais os gêneros de maior receptividade às formas da oralidade? Que princípios mais gerais subjazem a essa migração?
- (d) A constância de alguns efeitos na fala e na escrita, ainda que com um peso diferente, permite estabelecer a existência de princípios de ordem mais geral, e mesmo universal, para a variação e mudança linguísticas?
- (e) Que contribuições essas novas análises poderiam representar para aqueles que se dedicam ao ensino da língua materna e à formação de professores?

⁸ O público-alvo foi determinado em função do preço do jornal nas bancas.

Fenômenos como o preenchimento variável do sujeito (referencial e arbitrário), o uso variável do dativo, o queísmo e o dequeísmo, a expressão variável das relativas, a ordenação de circunstâncias temporais e locativas, a expansão de uso da forma **onde** e a variação na concordância verbal foram alguns dos aspectos examinados em estudos comparativos, cujos resultados estão reunidos em número temático da *Revista Lingüística*, do Programa de Pós-Graduação da UFRJ (2007), e em volume organizado por Paiva e Gomes ([20-]). Em linhas gerais, tais trabalhos confirmaram uma escrita bem mais conservadora do que a fala, apesar de deixarem transparecer alguns aspectos inovadores, com gradativa incorporação de formas linguísticas frequentes na oralidade.

A face mais conservadora da escrita se expressa, por exemplo, na preservação de formas e estruturas praticamente extintas, ou em franco processo de desaparecimento na fala, como é o caso das formas de expressão do sujeito arbitrário, estudadas por Duarte (2007b) e Duarte e Freire ([20-]). Tais trabalhos mostram que, embora os textos jornalísticos em muito se aproximem dos textos orais no que diz respeito, por exemplo, ao controle estrutural do sujeito preenchido, as duas modalidades se distinguem nitidamente quanto às formas preferenciais de expressão do sujeito de referência arbitrária: na escrita, ainda prevalecem o *se* indeterminador e o uso da terceira pessoa do plural, em oposição à ampla disseminação de formas de indeterminação através de pronomes como **você** e **a gente** na fala (SANTOS, 2003).

A natureza mais conservadora da escrita não impede, no entanto, que formas mais próprias da modalidade oral encontrem guarida nos textos escritos, mesmo naqueles de maior formalidade. Mollica et al. (2007) detectam a presença de fenômenos como a ocorrência de pausa entre sujeito e verbo, refletida no uso da vírgula; a omissão da preposição diante do conectivo *que* (o chamado *queísmo*); e a ocorrência da relativa cortadora em textos jornalísticos, principalmente naqueles de feição mais popular. Comparando a escrita jornalística com redações de alunos de um vestibular comunitário, constata-se maior presença das formas menos padrão na produção escrita dos segundos, mostrando que os filtros normativos operam de forma diferenciada em diferentes gêneros textuais.

As pressões de caráter prescritivista não conseguem impedir a migração de variantes não previstas pela norma, quando se trata de variações que envolvem mudança, como é o caso do queísmo e das relativas cortadoras (MOLLICA, 2008). Mesmo em gêneros mais formais e supostamente mais monitorados, como dissertações e teses de áreas distintas como Letras/Linguística e Ciências Exatas, pode ser atestada a presença dessas construções, como evidencia Mollica (2010). Comparando um largo período de tempo, que compreende desde a década de 60 do século passado até os dias atuais, a autora conclui que, a partir dos idos de 80, tais estruturas encontram porta aberta na escrita de textos acadêmicos, sem distinção

de área de conhecimento. Trata-se, sem dúvida, de inovação importada da fala, modalidade em que tais usos, possivelmente, têm avançado nos últimos decênios.

Adotando uma perspectiva funcionalista, associada ao conceito de marcação, a autora atribui a migração dessas estruturas para a escrita ao fato de elas serem menos complexas, com menor custo de produção – e desse ponto de vista menos marcadas, o que permite que a diferença entre elas e a padrão passe despercebida ao falante.

Inovação também presente na escrita é o uso de *onde*, em variação com prep. + art. + QU. Originalmente um relativo de valor locativo, este elemento passa a ser usado em remissões anafóricas a categorias não locativas (BRAGA et al., 2007). Os autores apresentam evidências de ordem diacrônica que permitem interpretar essa evolução como um processo de gramaticalização de *onde*, como conector, que vai se afastando do seu valor espacial e perde suas propriedades anafóricas, no caso a necessidade de antecedente, própria dos relativos. Utilizando para a escrita uma perspectiva diacrônica, os autores comprovam a expansão do uso não padrão em ambas as modalidades, porém com pesos relativos diferentes, e atestam o papel diferenciado da categoria cognitiva do termo retomado (espaço, objeto, instituição, pessoa etc.) particularmente na escrita. Destacam a necessidade de se considerarem os diferentes gêneros e tipos textuais em análises dessa natureza, em ambas as modalidades.

O que se destaca em muitos trabalhos é que a trajetória de incorporação de formas da oralidade envolve o *continuum* de gêneros textuais, comprovando, assim, a fragilidade de uma dicotomização entre fala e escrita. Pode-se afirmar que um dos principais ganhos da amostra jornalística decorreu da sua abrangência, levando-se em conta diferentes gêneros veiculados na imprensa escrita. Essa diversificação permitiu aos pesquisadores examinar mais minuciosamente a correlação entre variantes linguísticas e diferentes “formas estáveis de enunciado” (BAHKTIN, 2003), que poderiam passar despercebidas numa oposição dicotômica fala/escrita e atestar que a porta de entrada para estruturas mais características do oral são os gêneros textuais mais informais ou que se situam num ponto mais intermediário do *continuum* fala-escrita.

O estudo de Gomes sobre a variação no uso do dativo anafórico (2007, [20--]), por exemplo, revela tendências inovadoras na escrita como a substituição da preposição **a**, prototípica dos complementos dativos, pela preposição **para**, de natureza mais informal e em estágio avançado de implementação na fala, como vimos na seção 1. Essa substituição ocorre com mais frequência, no entanto, nos gêneros jornalísticos em que o leitor é diretamente referido, como o *horóscopo*, o que lhes confere uma maior aproximação com a oralidade. A variante *ausência de preposição*, mais estigmatizada socialmente (**dei Ø ele o livro**), é de fato evitada na escrita jornalística, tendo ocorrido apenas num caso isolado.

A importância da variável gênero é notável também na alternância entre SN, pronome e anáfora zero para a expressão do sujeito de terceira pessoa, como mostra Paredes Silva (2007b). Sua análise revela um crescendo na preferência pelo SN nos gêneros jornalísticos analisados, em proporção direta ao aumento no grau de formalidade do texto. Assim, numa comparação entre artigos de opinião, notícias e crônicas, são os dois primeiros que mais utilizam a variante nominal. Para a autora, tal tendência se explica com base em princípios funcionalistas, como a informatividade, ou seja, a necessidade de deixar claro para o leitor o tópico/tema tratado, seja por repetição do termo ou por recategorização do referente. Além disso, aspectos como a distância referencial e o traço semântico de animacidade são variáveis também correlacionadas a essas preferências. Expondo ideias, portanto, num plano inanimado, os articulistas optam pela retomada nominal. Esta tendência vai ao encontro de resultados já observados na análise da escrita informal de cartas pessoais (PAREDES SILVA, 1988), em que se constatou que a escolha do pronome no português é fortemente associada ao traço de animacidade do referente.

No gênero crônica, considerado o mais informal entre aqueles analisados, prevalece a norma de uso da anáfora zero em contextos de continuidade referencial (PAREDES SILVA; COSTA PINTO, 2010), o que vai ao encontro da tendência depreendida para a fala, como foi destacado na seção 1 (PAREDES SILVA, 2003b).

A precedência da anáfora zero sobre outras formas de realização do sujeito em textos de maior formalidade é corroborada no estudo dos denominados *sujeitos de referência estendida* (seção 2), como atestam Paredes Silva e Oliveira ([20-]). Comparando diferentes gêneros de escrita, desde cartas pessoais até gêneros jornalísticos mais formais (editoriais), as autoras constatam maior conformidade dos gêneros *editorial* e *artigo de opinião* à orientação normativa, com maior recorrência da anáfora zero, se comparados, por exemplo, a gêneros mais informais como a carta pessoal.

A ordenação de SPreps circunstanciais temporais e locativos, investigada por Paiva (PAIVA et al., 2007; PAIVA, 2008a, 2008b, 2011, [20-]) à luz do conceito de marcação, é outro exemplo da necessidade de olhar as diferenças entre fala e escrita sob o prisma de um *continuum* de gêneros textuais. Focalizando apenas os SPreps temporais (PAIVA et al., 2007), as autoras depreendem diferenças atribuíveis às condições de produção da fala e da escrita. Em linhas gerais, a produção mais planejada da escrita e a sua forma de recepção propiciam disposições sintagmáticas (como a “intromissão” do circunstancial entre um verbo e seu argumento) evitadas na fala, cujo processamento se faz on-line. Comparando, no entanto, diferentes gêneros (editoriais, notícias/reportagens⁹

⁹ Neste artigo, não estamos fazendo distinção entre *notícias* e *reportagens*, embora reconheçamos que haja, no âmbito jornalístico, necessidade de fazê-lo. Como nosso material se restringiu às notícias *da cidade*, o que

e crônicas) a autora constatou que são os **editoriais** o gênero jornalístico mais propenso a situar os circunstanciais temporais na periferia esquerda da oração. As reportagens, por sua vez, se destacam como o gênero que mais evita esse uso, provavelmente em razão de serem mais centradas nos participantes, que garantem a continuidade referencial através do sujeito/tópico principal. Ora, essa diferença de comportamento observada nos gêneros reforça uma discussão teórica já empreendida por vários autores (GIVÓN, 1990; CROFT, 1990; DRYER, 1995) quanto à relatividade do conceito de marcação, quando se levam em conta diferentes formas de organização textual.

A questão da correlação forma-função, entendida em termos de efeitos discursivos e pragmáticos, e gênero discursivo vai aparecer sobremaneira nos estudos de Roncarati (2003, 2008, 2009, 2010), tanto naquele voltado para o uso de estruturas clivadas na escrita como nos que focalizam a referência demonstrativa ou a construção da dêixis social. Destaca-se para esses diferentes fenômenos linguísticos o papel de variantes linguísticas na construção argumentativa do texto. Analisando textos acadêmicos, Almeida e Roncarati (2007) ressaltam o papel dessas estruturas na construção da proeminência informativa e sua contribuição para a orientação da leitura. Na análise das diferentes formas de tratamento (pronome de tratamento, vocativo, expressões referenciais definidas ou indefinidas), Roncarati (2008) mostra que as formas de tratamento mais frequentes em cada um dos gêneros analisados (reportagem, artigo de opinião, cartas de leitores, editoriais, colunas sociais) são colocadas a serviço de objetivos comunicativos, singularizando uma orientação argumentativa.¹⁰

Duas perguntas instigam os estudos sobre a variação e mudança na escrita: a primeira diz respeito à convergência no efeito de variáveis estruturais; a segunda concerne à generalidade de princípios que autorizam a migração de variantes linguísticas inovadoras para a escrita. A primeira pode encontrar resposta positiva na regularidade do controle estrutural da variação em ambas as modalidades, ainda que com diferenças na magnitude dos efeitos. Os já citados trabalhos de Duarte e o de Paredes Silva sobre a forma de realização do sujeito chegam a conclusões semelhantes: destaca-se também na escrita a importância da relação do sujeito sob análise com sua menção prévia (tratada como condições de referência em Duarte (2003a) e como distância referencial em Paredes Silva (2007b) ou conexão discursiva em Paredes Silva (1988, 2003a). O traço de animacidade do sujeito, um forte promotor do uso do pronome na modalidade falada, como vimos na seção 1, destaca-se igualmente na escrita.

privilegiou textos de natureza mais narrativa e menos opinativa, como seria, por exemplo, o noticiário político, nos permitimos essa variação terminológica.

¹⁰ As reflexões e conclusões advindas dos estudos sobre referenciação estão reunidas em Roncarati (2010).

A ordenação dos SPreps locativos e temporais ilustra igualmente a convergência na ação de princípios mais gerais que operam independentemente das diferentes condições de produção de cada uma das modalidades. Paiva et al. (2007) e Paiva (2008a, 2011, [20--]) discutem em detalhes a similaridade na forma como a ocorrência desses constituintes na margem esquerda da oração interage com a posição do sujeito (VS, SV). Além desse aspecto estrutural, outras similaridades se destacam, como as que decorrem da ação do princípio de peso final, particularmente na sequenciação relativa de circunstanciais de categorias semânticas distintas, como as de tempo, modo e lugar.

A segunda questão, a de princípios norteadores da incorporação, expansão ou retração de variantes inovadoras na escrita, encontra igualmente evidências na análise de alguns fenômenos variáveis, a exemplo dos estudados por Mollica (2003b, 2007, 2010, [20--]b), fenômenos que podem ser considerados como mudança em curso, como o uso das relativas cortadoras, o queísmo ou ainda a alternância entre *muitas das vezes/muitas vezes*. A título de hipótese, a autora postula que a maior chance de ingresso de estruturas queísticas e relativas cortadoras na escrita se explica em termos do princípio segundo o qual estruturas menos marcadas (MOLLICA, 2010, [20--]b) possuem mais acolhida na escrita do que as formas linguísticas mais marcadas. O mesmo princípio pode justificar a forte resistência à introdução de *muitas das vezes* na escrita (MOLLICA, [20--]a), se considerarmos seu status mais marcado. A autora defende, ainda, a hipótese de que pressões relativas à regularização paradigmática e tendências resultantes do processo de hiper correção intervêm igualmente na forma como a mudança opera, tanto na fala como na escrita.

Os estudos de variação na língua escrita têm mostrado, também, que há determinadas estruturas mais frequentes nessa modalidade, embora possam assumir diferentes roupagens, dependendo do gênero textual. Veja-se o caso da concordância verbal com SNs complexos (aqueles que contêm um SPrep encaixado), focalizada por Scherre e Naro (2007, [20--]). Os autores atestam, sobretudo, a importância de traços do sujeito na mudança do controle da concordância verbal¹¹. Estabelecem uma escala que vai dos sujeitos mais fortes aos menos fortes, a partir de traços semânticos (caráter referencial, humano) e morfológicos (presença de expressões com numerais) e demonstram como isso pode influenciar no controle da regra de concordância verbal.¹²

¹¹ No referido artigo, o *corpus* de escrita analisado não se restringe a jornais cariocas, mas abrange cartas circulares, ofícios, memorandos e textos técnicos ou científicos, todos tidos como monitorados. Os autores, porém, não fazem um controle específico dos gêneros utilizados.

¹² Esse caso particular de concordância não pôde ser estudado na fala, uma vez que, nesta modalidade, ao menos no gênero entrevista sociolinguística, é muito rara a presença de SNs complexos.

A forte presença de SNs complexos em gêneros da escrita jornalística e acadêmica tem sido escrutinada por Paredes Silva (2011), com o objetivo de verificar a correlação dos tipos de SNs com suas funções sintáticas e discursivas em diversos gêneros de escrita. Considerando aspectos estruturais como o número de constituintes do SN, o número e o tipo de Spreps encaixados, a presença de nominalizações, é possível identificar uma escala que situa SNs mais complexos em gêneros acadêmicos¹³ decrescendo a complexidade nos gêneros jornalísticos, conforme se trate de artigos de opinião, notícias e crônicas. Tais diferenças demonstram, claramente, que a possibilidade de compactar informações, graças a recursos como nominalizações se distribui de forma diferenciada num *continuum* escrito/oral.

As generalizações destacadas até este ponto emergiram em análises, sobretudo, da Amostra de Textos Midiáticos. O próprio interesse pela investigação do *continuum* fala – escrita, conduz, necessariamente, à exploração de gêneros intermediários, não contemplados na organização inicial do *corpus* jornalístico. A colocação de outras questões exigiu a análise de amostras como entrevistas transcritas em jornais e revistas (PAREDES SILVA; MARTINS, 2008), diálogos de revistas de quadrinhos (SCHERRE, 2008) assim como de redações de alunos. (MOLLICA et al, 2007).

Nessa expansão para novos *locus* de variação e mudança, ganham, ainda, destaque os estudos da linguagem do ciberespaço, utilizada nas redes sociais, moldada para atender aos relacionamentos virtuais assim como a veiculação de informação em plataforma de Ensino a Distância (EAD) (BARBOSA, 2010). Como mostra Mollica (2012), a comunicação via web configura um novo espaço de variação e mudança que pode, inclusive, abrir portas para a entrada de novas formas linguísticas variáveis na língua. Um exemplo é a forma *orkutização*, que tal como *digitalização*, constitui inovação. No entanto, não corresponde ao processo de *orkutizar*, estrutura verbal que nem entrou no português assim como **haqkear*, diferentemente do verbo *twitar*, que teve entrada franca na língua. Abrem-se, assim, novas perspectivas para a verificação das questões centrais sobre contato linguístico, *code-switching*, redes sociais, formação do léxico e sobre mudança.

Retomando as questões levantadas no início desta seção, podemos dizer que:

- a) Existe uma série de restrições aos fenômenos variáveis que se manifesta de forma similar na escrita e na fala. As diferenças no meio de produção podem, todavia, interferir na forma como essas restrições operam.
- b) Mudanças em curso, que se instalaram de forma mais veloz na fala, podem ser incorporadas na escrita. Algumas podem, no entanto, ser inibidas por força do poder conservador da norma e das instituições.

¹³ Foram investigados artigos, resumos de artigos e *abstracts* de dissertações e teses da área de Letras.

- c) À incorporação de mudanças na escrita subjazem princípios que, de certa forma, selecionam candidatos a serem incorporados.
- d) Mesmo bons candidatos a serem incorporados na escrita obedecem a uma trajetória que envolve as distinções de gêneros textuais.

Abrindo outras frentes de trabalho e colocando outras questões

As questões relativas à variação e mudança na variedade carioca, nas suas modalidades falada e escrita, projetam outros questionamentos. Destacamos, nesta seção, as questões ligadas à aquisição da variação e às aplicações teórico-metodológicas da pesquisa sociolinguística.

A aquisição da variação sociolinguística, focalizando o período que antecede a alfabetização formal, com crianças de 2 a 5 anos, é uma vertente que tem sido explorada recentemente no PEUL. Registre-se, no entanto, o trabalho pioneiro de Alzira Tavares de Macedo, em sua tese de doutorado “Uso do Futuro do Subjuntivo em Português: regularização de uma forma verbal”, que além de um estudo de tempo aparente com dados de adultos também utilizou testes com crianças para avaliar os condicionamentos da variação na aquisição.

Estudos sobre aquisição da variação partem do pressuposto de que, se a variação é inerente ao sistema, as estruturas variáveis são adquiridas da mesma forma que as invariáveis (CHAMBERS, 1995). A observação da aquisição linguística sob o enfoque da Sociolinguística Variacionista tem permitido observar de que maneira os condicionamentos que regulam a variação na gramática estabilizada do adulto atuam na aquisição; como são transmitidos às crianças valores sociais associados a variantes; como se dá a aquisição de condicionamentos sociais; de que maneira a variabilidade manifestada pelas crianças reflete o estágio da mudança em curso. Ainda, como qualquer outro estudo dentro deste quadro de referência, os resultados sobre aquisição da variação devem contribuir para a discussão da modelagem teórica da variação na gramática.

Diversos trabalhos já foram realizados e dissertações e teses orientadas utilizando a Amostra AQUIVAR/PEUL/UFRJ, a saber: Gomes (2004), sobre a aquisição de complementos indiretos; Benayon (2006), sobre a aquisição de ditongos; Vieira (2006), sobre a flexão da 3^a pessoa do plural; Gomes (2006), sobre a aquisição da vibrante em coda; Gomes (2010), sobre a aquisição dos grupos consonantais formados por lateral (claro ~ craro) e tepe (precisa ~ picisa); Benayon (2010), sobre a aquisição das fricativas; Gomes et al (2011), sobre a flexão nominal e a flexão verbal na aquisição.

Outros trabalhos discutem questões teóricas e metodológicas centrais no estudo da aquisição da variação, a saber: Gomes, Benayon e Vieira (2006), sobre

a aquisição da variação sob a ótica dos modelos multirepresentacionais; Gomes (2008), sobre questões teóricas e metodológicas da variação na aquisição; Gomes (2011), sobre o papel da variação na modelagem do conhecimento fonológico na aquisição e no adulto. As evidências fornecidas por esses trabalhos têm contribuído para ampliar o conhecimento sobre a aquisição linguística e mostram a importância de se observar a aquisição dos aspectos variáveis da estrutura linguística.

Os estudos sobre aquisição da variação têm atestado a convergência entre princípios mais gerais que regulam a heterogeneidade estruturada na gramática do adulto e a forma como se processa a aquisição dos processos variáveis. No estudo da realização variável da marca de 3^a pessoa do plural, por exemplo, Vieira (2006) observou o efeito relevante de saliência fônica, isto é, uma tendência à ausência de marca flexional naqueles verbos em que a diferença entre a forma de singular e plural é menor; ao contrário, maior presença de marcas flexionais nos casos em que a diferença é maior. A presença da marca de plural é mais evidente, para as crianças, nos três últimos níveis da escala proposta por Naro (1981).

Tomando como base a hipótese de que as representações sonoras se estabelecem em diferentes graus de abstração (PIERREHUMBERT, 2003), os trabalhos sobre aquisição de estruturas sonoras variáveis mostram a relação entre a importância das formas em competição - as variantes fonéticas - entendidas como parte da representação da forma das palavras e os tipos fonológicos que podem ser abstraídos destas representações, e em que níveis de abstração. A relação entre *type* (tipo) e *token* (ocorrência) permite não só compreender o percurso aquisitivo como também autoriza inferências sobre o conhecimento linguístico que está sendo abstraído. O estudo da aquisição dos grupos consonantais revelou, por exemplo, que a diferença na aquisição dos dois padrões silábicos, a saber: *onset* complexo formado por obstruinte e tendo como alvo a lateral, como em [globu] ~ [grobu]; e obstruinte tendo como alvo o tepe, como em [outru] ~ [outu], não pode ser explicada em função de uma hierarquia aquisitiva. Foi observado que os itens que têm a lateral como alvo utilizam ou manifestam a realização da segunda consoante do *onset* antes daqueles que têm o tepe como alvo. Algumas evidências provenientes destes diversos estudos mostram que não há, neste caso, a atuação de uma hierarquia aquisitiva, em que alguns sons são adquiridos antes de outros, com reflexos na aquisição dos itens que possuem essas estruturas. Houve crianças que apresentaram alta incidência de tepe e lateral no *onset* complexo do tipo C(r)V, mas ausência de tepe em *onset*, sendo este substituído por lateral e vice-versa, isto é, apresentaram ausência da segunda consoante no grupo e presença da líquida no *onset*. Esse padrão de aquisição é atribuído à competição entre formas fonéticas e os tipos estruturais a eles subjacentes. No caso da variação entre lateral e tepe, essa alternância não

implica competição de padrões fonológicos diferentes, favorecendo a abstração de um padrão CCV para os itens que se enquadram neste caso (GOMES, 2010).

Do ponto de vista da aplicação, a grande massa de material produzido sobre a variedade carioca garantiu o conhecimento do perfil sociolinguístico dessa comunidade de fala e forneceu as condições ótimas para repensar a forma de traduzir o produto das pesquisas produzidas em subsídios para a educação. Assim, desde o início dos anos 90, a pesquisadora Maria Cecília Mollica tem se dedicado a uma vertente de estudos – Sociolinguística Educacional – cuja preocupação central é a criação e disponibilização de material instrucional (na forma de livros e CDs) que incorpora a variação fono-ortográfica, morfossintática e discursiva. Contribui, ainda, para a formação continuada de docentes em vários níveis de ensino. (MOLLICA, 2000, 2003a, 2006, 2007, 2009; RONCARATI; MOLLICA, 2001).

Estes trabalhos partem do pressuposto de que o conhecimento da realidade sociolinguística das comunidades de fala permite analisar os problemas específicos dos aprendizes e conhecer as dificuldades próprias dos docentes em via de qualificação, a fim de melhor instrumentalizá-los para lidar com a natureza heterogênea da linguagem e seus diferentes registros escritos. Tal pressuposto é primordial também quando se tem a meta precisa de se oferecer soluções para questões importantes e atinentes à alfabetização e ao letramento em todo o seu percurso. A parceria com pesquisadores da área de Matemática tem permitido estender as reflexões para questões relativas à escrita matemática (LEAL; MOLLICA, 2009, 2010, [20-]). A proposta de estratégias de mediação vem se mostrando eficaz na implementação de pedagogia que lida com fenômenos dinâmicos em classes muito heterogêneas sociolinguisticamente, em ambientes institucionalizados ou não institucionalizados. Neste particular, tem sido profícuo o esforço em oferecer subsídios aos alfabetizadores que participam do Programa de Alfabetização da UFRJ cujo objetivo é o de melhor qualificar os docentes para atuar de forma eficaz ao entender e explicar os erros na produção discursivo-textual dos noviços em processo de alfabetização e letramento.

Nesta perspectiva da Sociolinguística Educacional se inserem, ainda, os estudos sobre o contato entre comunidades geograficamente diferentes, rurais ou urbanas, que levam em conta os *continua* de níveis de oralidade e escrita, de monitoramento estilístico e de práticas eminentemente orais comparadas a eventos de letramento (MOLLICA, 2003a; MOLLICA; LEAL, 2008, 2010), bem como suas repercussões para a prática docente e discente.

O mais recente desafio nesta linha de investigação encontra-se em trabalho inédito de Mollica e Roncarati ([20-]) em que as autoras lançam mão de alguns princípios funcionalistas, com vistas a instrumentalizar o professor para melhor trabalhar pedagogicamente com a variação. Para tanto, focalizam conceitos como

arbitrariedade e iconicidade, estatuto informacional do referente, cadeia tópica e parâmetros tais como saliência fônica, marcação, regularização paradigmática, níveis de acessibilidade do referente, com o objetivo de oferecer explicação nova aos docentes em processo de capacitação¹⁴.

A atuação de alguns pesquisadores do PEUL como docentes do curso de Fonoaudiologia da UFRJ vem gerando frutos interessantes para a compreensão, sob a ótica da Sociolinguística, dos quadros de transtornos de aprendizagem das habilidades de leitura e escrita e da escrita matemática (MOLLICA; SILVA, [20--]). Para tanto, a pesquisadora tem estabelecido parceria estreita com professora Marisa Leal (IM/UFRJ) com quem vem pesquisando e orientando trabalhos em literacia nos campos de linguagem e matemática especialmente voltados para a educação de jovens e adultos - EJA (LEAL; MOLLICA, 2009, 2010, [20--]). Nesse particular, tem sido profícuo seu esforço em oferecer subsídios aos alfabetizadores que participam do Programa de Alfabetização da UFRJ, cujo objetivo é o de melhor qualificar os docentes para atuar de forma eficaz ao entender e explicar os erros na produção discursivo-textual dos noviços em processo de alfabetização e letramento.

A exploração das interfaces da Sociolinguística com outras áreas de conhecimento têm se mostrado igualmente instigante na colaboração com profissionais de Ciência da Informação. Um bom exemplo dos resultados dessa parceria é o trabalho de Guedes (2010), que comprova a regularidade da variação de sufixos em nominalizações deverbais em periódicos científicos. A autora utiliza as leis de ZIPF, de 1935, de modo a verificar a zona de maior densidade semântica das formas variáveis. Na mesma direção, a pesquisadora identifica as diferenças estatísticas da variação no *continuum fala/escrita* (GUEDES, 2010, 2012) e analisa a escrita do gênero resumos de cartas científicas, que circulavam entre filósofos naturais desde o século XVII, desempenhando papel crucial na comunicação científica. Os resultados encontrados pela autora contribuem para estabelecer pontos de interação entre a Linguística e a indexação na Ciência da Informação.

Considerações finais

Neste artigo, ressaltamos alguns pontos que constituíram a pauta central de trabalho do grupo PEUL, com ênfase nas questões teóricas e metodológicas ligadas à dinamicidade inerente das línguas naturais. Ressaltaram-se, ainda, os caminhos possíveis que a Sociolinguística pode trilhar na interface com outras áreas da Ciência, abrindo, assim, o espaço para maior compreensão dos múltiplos usos da língua.

¹⁴ A pesquisadora tem envidado esforços tanto no sentido de fornecer subsídios para o processo de letramento de sujeitos típicos como de sujeitos com desenvolvimento atípico, o que se torna possível graças à parceria com profissionais da área de Fonoaudiologia. Essa parceria vem gerando frutos para a compreensão dos quadros de transtornos de habilidades de leitura e de escrita matemática.

■ *ABSTRACT: PEUL (Program for the Study of Language Use) is composed of a group of researchers whose main initial purpose is to investigate patterns of variation and change in Brazilian Portuguese as spoken in Rio de Janeiro (the "carioca" variety). At first, in the 1980's, only spoken language was studied. A new sample of the same community of speakers collected about 20 years later facilitated comparison of the behavior of the community (trend study) and the individuals (panel study). These studies gave rise to the analysis of linguistic change in real time. Recently, investigation of written language, focusing specifically on newspaper genres, has been undertaken. This expansion of both aims and corpora was motivated by comparison of principles correlated to variable phenomena in the modalities under investigation. It also allowed verification of hypotheses related to the migration of change-in-process from spoken to written language. In this article we intend to show how the analysis developed by members of the group has contributed to deepen several relevant theoretical points. The systematicity of principles underlying the variable use of distinct phenomena, the way linguistic change penetrates written genres, or is rejected by them, are some of the issues presented. The constant examination of new questions has expanded research to the interface with other areas of knowledge.*

■ *KEYWORDS: Sociolinguistics. Spoken language. Written language. Text genre. Applications.*

REFERÊNCIAS

- ALMEIDA, M. C.; RONCARATI, C. N. A multifuncionalidade da clivagem na fala e na escrita. *Revista Linguística*, Rio de Janeiro, n.7, v.1, p.21-52, 2007.
- BAKHTIN, M. *Estética da criação verbal*. São Paulo: Martins Fontes, 2003.
- BARBOSA, M. F. S. de O. *(Im)polidez em EAD*. 2010. 136f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.
- BENAYON, A. R. *Aquisição das fricativas sibilantes no português brasileiro: propriedades distribucionais e variação*. 2010. 138f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.
- _____. *A emergência de padrões fonológicos: a aquisição de ditongos orais decrescentes do PB*. 2006. 138f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.
- BRAGA, M. L. E aí se passaram 19 anos. In: DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003. p.159-174.
- BRAGA, M. L. et al. É um problema de visão para perto, inevitável após os 40 anos de idade, onde o indivíduo tem dificuldade para ver imagens. *Revista Linguística*, Rio de Janeiro, v.3, n.1, p.117-132, 2007.

- BRAGA, M. L.; MOLLICA, M. C. M. As estruturas, SN pleno + SN pronominal anafórico no português do Brasil 20 anos depois. In: LIMA-HERNANDES, M. C.; CHULATA, K. A. *Língua portuguesa em foco: ensino-aprendizagem, pesquisa e tradução*. Lecce: Pensa Multimidia, v.1, 2010. p.85-95.
- BRAGA, M. L.; PAIVA, M. C. A. Gramaticalização e sociolinguística variacionista: o tratamento das construções iniciadas por *por*. In: LIMA, M. A. F. et al. *Colóquios linguísticos e literários: enfoques epistemológicos, metodológicos e descritivos*. Teresina: Ed. da UFPI, 2011a. p.125-152.
- _____. Gramaticalização e gramática de construções: estabilidade e instabilidade no uso de orações complexas de causa em tempo real. *Revista Letras & Letras*, Uberlândia, v.27, n.1, p.51-70, 2011b.
- COSTA, A. L. dos P. *O futuro do pretérito e suas variantes no português do Rio de Janeiro: um estudo diacrônico*. 2003. 132f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2003.
- CHAMBERS, J. *Sociolinguistic Theory*. Oxford: Blackwell, 1995.
- CROFT, W. *Typology and universals*. Cambridge: Cambridge University Press, 1990.
- CYRINO, S. *O objeto nulo no português do Brasil: um estudo sintático-diacrônico*. Londrina: Ed. da UEL, 1997.
- CYRINO, S.; DUARTE, M. E.; KATO, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: KATO, M. A.; NEGRAO, E. V. (Ed.). *Brazilian Portuguese and the Null Subject Parameter*. Frankfurt-Madrid: Vervuert-Iberoamericana, 2000. p.55-104.
- DRYER, M. S. Frequency and pragmatically unmarked word order. In: DOWNING, P.; NOONAN, M. (Ed.). *Word Order in Discourse*. Amsterdam: John Benjamins, 1995. p.105-135.
- DUARTE, M. E. L. A representação dos sujeitos de “referência estendida”: um estudo diacrônico. In: LOBO, T. et al. (Org.). *ROSAE: linguística histórica, história das línguas e outras histórias*. Salvador: Ed. da UFBA, 2012. p.123-135.
- _____. Sujeitos de referência definida e arbitrária: aspectos conservadores e inovadores na escrita padrão. *Revista Lingüística*, Rio de Janeiro, v.3, n.1, p.89-115, 2007b.
- _____. A evolução na representação do sujeito em dois tempos. In: PAIVA, M. C.; DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa, 2003a. p.115-128.

_____. O sujeito expletivo e as construções existenciais. In: RONCARATI, C.; ABRAÇADO, J. (Org.). *Português Brasileiro: contato linguístico, heterogeneidade e história*. Rio de Janeiro: 7Letras, 2003b. p.123-131.

_____. The loss of avoid pronom in brazilian Portuguese. In: KATO, M. A; NEGRÃO, E. V. In: *Brasilian Portuguese and the null subject*. Frankfurt am Main: Vervuert: Verlag, 2000, p.17-36.

_____. *A perda do princípio 'Evite pronome' no português brasileiro*. 2005. 140f. Tese (Doutorado) – Universidade de Campinas, Campinas, 1995.

DUARTE, M. E. L.; FREIRE, G. C. Como a escrita padrão recupera formas em extinção e implementa formas inovadoras In: PAIVA, M. A.; GOMES, C. A. (Org.). *A dinâmica da variação e mudança na escrita*. [20--]. No prelo.

GIVON, T. *Syntax: a functional typological introduction*. Amsterdan: John Benjamins Publishing, 1990. v.II.

GOMES, C. A. Uso variável do dativo na escrita jornalística. In: PAIVA, M. C. A; GOMES, C. A. *Dinâmica da variação e da mudança na escrita*. [20--]. No prelo.

_____. Variação sociofonética na aquisição e na modelagem do conhecimento linguístico. *Revista da ABRALIN*, Brasília, v.2, p.209-226, 2011.

_____. Variação linguística e aquisição do onset complexo no português brasileiro. In: MOLLICA, M. C. de M. (Org.). *Usos da linguagem e sua relação com a mente humana*. Rio de Janeiro: Tempo Brasileiro, 2010. p.15-22.

_____. Aquisição da variação estruturada: uma nova perspectiva de pesquisa. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Julius Naro e a linguística no Brasil uma homenagem acadêmica*. Rio de Janeiro: 7 Letras: FAPERJ, 2008. p.107-114.

_____. Uso variável do dativo na fala e na escrita. *Revista Linguística*, Rio de Janeiro, v.3, n.1, p.7-20, 2007.

_____. Aquisição do tipo silábico CV(r) no português brasileiro. *Scripta*, Belo Horizonte, v.10, p.58-77, 2006.

_____. Aquisição linguística em contexto de input variável: a emergência das variantes de dativo. *Revista de Estudos da Linguagem*, Belo Horizonte, v.12, n.1, p.175-190, 2004.

_____. Variação e mudança na expressão do dativo no português brasileiro. In: PAIVA, M. C.; DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003. p.81-96.

GOMES, C. A.; BENAYON, A. R.; VIEIRA, M. C. P. Aquisição linguística sob a ótica dos modelos multirepresentacionais. *Gragoatá*, Niterói, v.21, p.303-318, 2006.

GOMES, C. A. et al. Variação e aquisição de flexão nominal e flexão verbal. *Gragoatá*, Niterói, v. 30, p.39-54, 2011.

_____. Pressões estruturais e discursivas no condicionamento da variação: a ordem dos complementos verbais no português brasileiro. In: RONCARATI, C.; ABRAÇADO, J. (Org.). *Português brasileiro: contato linguístico, heterogeneidade e história*. Rio de Janeiro: 7 Letras, 2003. p.199-205.

GRYNER, H. Equilíbrio e desequilíbrio na evolução das estruturas condicionais. In: DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa : Faperj, 2003. p.175-192.

_____. *A variação de tempo, modo e conexão nas orações subordinadas condicionais do português*. 1990. 550f. Tese (Doutorado em Linguística) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1990.

GUEDES, V. L. S. Breve histórico do artigo de periódico. In: MOLLICA, M. C. M. (Org.). *Linguística e Ciência da Informação*: diálogos possíveis. Rio de Janeiro: APPRIS, 2012.

_____. *Nominalizações deverbais em artigos científicos*: uma contribuição para a análise e a indexação temática da informação. 2010. 107f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

LABOV, W. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

_____. *Principles of linguistic change*: internal factors. Cambridge: Blackwell, 1994.

_____. Building on empirical foundations. In: LEHMANN, W.; MALKIEL, Y. (Ed.). *Perspectives on historical linguistics*. Amsterdam: John Benjamins, 1982. p.17-92.

_____. What can be learned about change in progress from synchronic descriptions. In: SANKOFF, D.; CEDERGREN, H. (Ed.). *Variation Omnibus*. Carbondale; Edmonton: Linguistic Research, 1981. p.177-199.

LEAL, M.; MOLLICA, M. C. Lendo matemática. In: BORTONI-RICARDO, S. M. *Mediação em leitura*. São Paulo: Parábola Editorial, [20--?].

_____. Competências subjacentes à leitura: pilares de acesso saberes formais. In: _____. (Org.). *Construindo o capital formal das linguagens*. Curitiba: CRV, 2010.

_____. *Letramento em EJA*. São Paulo: Parábola Editorial, 2009.

LIGHTFOOT, D. *The development of language acquisition, change and evolution*. Oxford: Blackwell, 1999.

MARTINS, K. C. *A expressão variável de anterioridade a um ponto de referência passado na escrita midiática*. 2010. 131f. Dissertação de mestrado (Mestrado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010.

MILROY, L. *Language and social networks. Language in society* 2. Oxford: Blackwell, 1980.

MOLLICA, M. C. Primeiras notas sobre muitas das vezes. In: PAIVA, M. C.; GOMES, C. A. (Org.). *Dinâmica da variação e mudança na escrita*. [20--]a. No prelo.

_____. Sobre restrições de mudança linguística. In: Congresso de la Delegación Argentina de la Asociación de Filología de América Latina, 1., [20--]b, La Plata. *Atas...* La Plata: Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata, [20--]b. No prelo.

_____. Neologismos na era cibernetica. In: PEREIRA, J. *Neologia e neologismo no Brasil*. Curitiba: Primas, 2012. p.277-288.

_____. Mudança e resistência no contínuo oral/escrito. In: MOLLICA, M. C. (Org.). *Usos da linguagem e sua relação com a mente humana*. Rio de Janeiro: Edições Tempo Brasileiro, 2010. p.91- 97.

_____. Formação em linguagem. In: _____. (Org.). *Linguagem para a formação em letras, educação e fonoaudiologia*. São Paulo: Contexto, 2009. p.27-33.

_____. Aportes para uma teoria da mudança na escrita. In: VOTRE, S.; RONCARATI, C. *Anthony Julius Naro e a Linguística no Brasil: uma homenagem acadêmica*. Rio de Janeiro: FAPERJ : 7LETRAS, 2008. p.242-53.

_____. Da escola para vida: a importância do letramento escolar. *Acolhendo a alfabetização nos países de Língua Portuguesa*. São Paulo: Ed. da USP, 2007. p.19-30.

_____. Português e matemática: parceria indispensável em política educacional. In: SILVA, C. R.; DA HORA, D.; CHRISTIANO, M. E. (Org.). *Linguística e práticas pedagógicas*. Santa Maria: Pallotti, 2006. p.33-54.

_____. *Da linguagem coloquial à escrita padrão*. Rio de Janeiro: 7LETRAS, 2003a.

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

_____. *Influência da fala na escrita*. Rio de Janeiro: Edições Tempo Brasileiro, 2000.

MOLLICA, M. C. et al. Marcas de fala na escrita de jornal e de escola. *Revista Linguística*, Rio de Janeiro, v.3, n.1. p.53-68, 2007b.

MOLLICA, M. C.; GONÇALEZ, M. *Linguística e Ciência da Informação: diálogos possíveis*. Curitiba: APPRIS, 2012.

MOLLICA, M. C.; LEAL, M. Ler e interpretar informação. In: MOLLICA, M. C.; GONZALEZ. *Linguística e Ciências da Informação: diálogos possíveis*. Curitiba: APPRIS, 2012. p.201-209.

_____. Modos de Ancoragem na interpretação da leitura de algumas bulas e listas. *Revista Domínios de Linguagem*, Uberlândia, ano 4, n.1, p.115-130, 2010.

_____. Oralidades em linguagens. In: GONÇALVES, C. A; ALMEIDA, M. L. L. (Org.). *Língua portuguesa: identidade, difusão e variabilidade*. Rio de Janeiro: AILP : UFRJ, 2008. v.2. p.493-502.

_____. Letramentos na escola e na vida. In: SILVEIRA, M. L. S. (Org.). *Educação popular e leituras do mundo*. Rio de Janeiro: UFRJ, 2007. p.35-42.

MOLLICA, M. C.; LISBÔA, S. G. Produtividade lexical em artigos científicos e a indexação temática da informação. In: SOARES, A et al. *Línguas pluricêntricas: variação linguística e dimensões sociocognitivas*. Braga: Universidade Católica Portuguesa, 2011. p.393-408.

MOLLICA, M. C.; RONCARATI, C. Como a escola pode explicar erros gramaticais e inovações? In: BORTONI-RICARDO, S. M. *Por que a escola não explica assim?* São Paulo: Parábola Editorial. [20--]. No prelo.

MOLLICA, M. C.; SILVA, P. *O letramento de sujeitos típicos e atípicos*. São Paulo: Contexto, [20--]. No prelo.

NARO, A. J. Sobre o efeito do princípio da saliência na concordância verbal na fala moderna, na escrita antiga e na escrita moderna. In: MOURA, D.(Org.). *Os múltiplos usos da língua*. Maceió: Ed. da UFAL, 1999. p.26-37.

_____. The social and structural dimensions of a syntactic change. *Language*, Baltimore, n.57, p.63-98, 1981.

NARO, A. J.; SCHERRE, M. *A trend study revisited: remodeling the age variable in number concord in Brazilian Portuguese*. Washington: Georgetown University, 2011. Comunicação apresentada no Congresso NAWAV40.

_____. Fluxos e contrafluxos: movimentos sociolinguísticos da comunidade de fala brasileira. In: MOLLICA, M. C (Org.). *Usos da linguagem e sua relação com a mente humana*. Rio de Janeiro: Tempo Brasileiro, 2010. p.79-90.

_____. Estabilidade e mudança linguística em tempo real: a concordância de número. In: PAIVA, M. C.; DUARTE, M. E. (Org.). *Mudança em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003. p.47-62.

OLIVEIRA SILVA, G. M.; SCHERRE, M. M. P. *Padrões sociolinguísticos*: análise de fenômenos variáveis do português falado no Rio de Janeiro. Rio de Janeiro: Tempo Brasileiro, 1996.

OMENA, N. P. de. A referência à primeira pessoa do plural: variação ou mudança? In: PAIVA, M. C.; DUARTE, M. E. L. (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa : Faperj, 2003. p.63-80.

PAIVA, M. C. Restrições à posição de SPreps temporais na modalidade falada. *Alfa*, São Paulo, v.56, n.1, p.29-53, 2012.

_____. Configurações XSV e XVS no português contemporâneo: complementaridade sintático-semântica e discursiva. *Diadorim*, Rio de Janeiro, n.8, p.245-270, 2011.

_____. Ordem não marcada de circunstanciais locativos e temporais. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Naro e a linguística no Brasil*: uma homenagem acadêmica. Rio de Janeiro: 7 Letras, 2008a. p.254-264.

_____. A ordem de circunstanciais no português brasileiro escrito. *Diacrítica*, Braga, v.22, n.1, p.73-91, 2008b.

_____. O percurso da monotongação de [ey]: observações no tempo real.. In: PAIVA, M. C. A.; DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003. p.31-46.

_____. Gramaticalização de conectores no português do Brasil. *Scripta*, Belo Horizonte, v.5, n.9, p.35-46, 2001.

_____. Ordenação de circunstanciais na escrita: princípios em competição? In: PAIVA, M. C. A; GOMES, C. A (Org.). *Dinâmica da variação e da mudança na escrita*. [20--]. No prelo.

PAIVA, M. C.; BARROS, E. C. M. Construções modais com ter: gramaticalização e variação. *Diacrítica*, Braga, v.25, n.1, p.259-282, 2011.

PAIVA, M. C.; DUARTE, M. E. L. Introdução. In: _____. *Mudança linguística em tempo real*. Rio de Janeiro: ContraCapa: Faperj, 2003. p.13-30.

PAIVA, M. C. A; GOMES, C. A (Org.). *Dinâmica da variação e da mudança na escrita*. [20--]. No prelo.

PAIVA, M. C.; SCHERRE, M. M. P. Retrospectiva sociolinguística: contribuições do PEUL. *D.E.L.T.A*, São Paulo, v.15, n. especial, p.201-222, 1999.

PAIVA, M. C. et al. Extensão e posição de circunstanciais temporais. *Cademos do CNLF*, Rio de Janeiro, v.13, n.4, p.1880-1891, 2009.

_____. Padrão não marcado de ordenação de circunstâncias temporais: regularidades e divergências entre fala e escrita. *Revista Linguística*, Rio de Janeiro, v.1, n.3, p.69-88, 2007.

PAREDES SILVA, V. L. P. É isso aí: verbo ser e demonstrativos em função coesiva no português. In: ENCONTRO NACIONAL DE LINGUÍSTICA, 10., 1985, Rio de Janeiro. *Anais...* Rio de Janeiro: PUC-RJ, 1985. Mimeografado.

_____. *Cartas cariocas*: a variação do sujeito na escrita formal. 1988. 330f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.

_____. O retorno do pronome *tu* à fala carioca. In: RONCARATI, C.; ABRAÇADO, J.. (Org.). *Português brasileiro*: contato linguístico, heterogeneidade e história. Rio de Janeiro: FAPERJ: 7Letras, 2003a. p.160-169.

_____. Motivações funcionais no uso do sujeito pronominal: uma análise em tempo real. In: PAIVA, M. C.; DUARTE, M. E. L. (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003b. p.97-114.

_____. O uso da anáfora zero em textos jornalísticos. *Cadernos de Pesquisa em Linguística*, Porto Alegre, v.3, n.1, p.52-61, 2007a.

_____. Continuidade de referência: Nomes, pronomes e anáfora zero em gêneros da fala e da escrita. *Revista Linguística*, Rio de Janeiro, v.1, n.3, p.159-178, 2007b.

_____. Desfazendo um mito: a repetição na escrita e suas funções. In: RONCARATI, C; ABRAÇADO, J. (Org.). *Português brasileiro*: contato linguístico, heterogeneidade e história. Niterói: Ed. da UFF, 2008a. p.334-344.

_____. Ainda sobre sujeitos pronominais e flexão verbal: revisitando Naro. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Naro e a linguística no Brasil*: uma homenagem acadêmica. Rio de Janeiro: 7 Letras, 2008b. p.381-392

_____. Notícias recentes da presença do pronome *tu* no quadro de pronomes no português falado no Rio de Janeiro. In: REBOLLO COUTO, L.; LOPES, C. R. (Org.). *As formas de tratamento no português e no espanhol*. Variação, mudança e funções conversacionais. Niterói: Ed. da UFF, 2011. p.245-262.

_____. *O uso de sintagmas nominais complexos em gêneros jornalísticos e acadêmicos*. Palestra apresentada no Curso de Pós-Graduação em Educação da Universidade do Minho, Braga, jun./2011. 17p. Mimeografado.

PAREDES SILVA, V. L. P. ; COSTA PINTO, P. I. A crônica: um gênero de vários tipos textuais. In: MOLLICA, M. C. (Org.). *Usos da linguagem e sua relação com a mente humana*. Rio de Janeiro: Tempo Brasileiro, 2010. p.37-52.

PAREDES SILVA, V. L.; MARTINS, A. P. O uso de SNs definidos vs. demonstrativos como rótulos em entrevistas jornalísticas. *Revista Linguística*, Rio de Janeiro, v.4, n.1, p.39-58, 2008.

PAREDES SILVA, V. L.; OLIVEIRA, A. “É isso aí”: a variação na referência estendida em diferentes gêneros de escrita. In: PAIVA, M. C.; GOMES, C. A. (Org.). *Dinâmica da variação e mudança na escrita*. [20--]. No prelo.

PIERREHUMBERT, J. Probabilistic phonology: discrimination and Robustness. In: BOD, R. et al. (Ed.). *Probabilistic Linguistics*. Boston: MIT Press, 2003. p.177–228.

RONCARATI, C. Os mostrativos na fala carioca. In: DUARTE, M. E. L. *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa: Faperj, 2003. p.139-158.

_____. Dêixis social – a designação social referenciada: sabe com quem você está falando? In: VOTRE, S. J.; RONCARATI, C. *Anthony Naro e a linguística no Brasil: uma homenagem acadêmica*. Rio de Janeiro: 7Letras: Faperj, 2008. p.115-147.

_____. A subjetividade do demonstrativo na mídia jornalística. *Investigações*, São Paulo, v.21, p.221-240, 2009.

_____. *As cadeias do texto – construindo sentidos*. São Paulo: Parábola Editorial, 2010.

RONCARATI, C.; MOLLICA, M. C. Questões teórico-descritivas em Sociolinguística e em Sociolinguística aplicada: uma agenda de trabalho. *D.E.L.T.A.*, São Paulo, v. 17, p.45-55, 2001.

SANTOS, G. M. *O uso de pronomes pessoais genéricos no discurso argumentativo oral*. 2003. 79f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003.

SCHERRE, M. M. P. Efeitos da saliência fônica e do tempo/modo na concordância verbal. In: MOLLICA, M. C. (Org.). *Usos da linguagem e sua relação com a mente humana*. Rio de Janeiro: Tempo Brasileiro, 2010. p.71-77.

_____. Paralelismo linguístico. *Revista de Estudos da Linguagem*, Belo Horizonte, v.7, n.2, p.29-59, jul./dez. 1998.

_____. O imperativo gramatical no português brasileiro: reflexo de mudança na escrita de revistas em quadrinhos. In: VOTRE, S.; RONCARATI, C. *Anthony Julius Naro e a linguística no Brasil: uma homenagem acadêmica*. Rio de Janeiro: 7Letras, 2008. p.306-319.

SCHERRE, M. M. P; NARO, A. J. Sobre o deslocamento do controle da concordância verbal. In: PAIVA, M. C. A; GOMES, C. A. *Dinâmica da variação e da mudança na escrita*. [20--]. No prelo.

_____. *A panel study revisited*: interpreting increase in use of number concord in Brazilian Portuguese. Washington: Georgetown University, 2011. Comunicação apresentada no Congresso NNAV40.

_____. O deslocamento do controle da concordância verbal. *Revista Linguística*, Rio de Janeiro, v.3, n.1, p.133-158, jun. 2007.

_____. Marking in discourse. Birds of a feather. *Language Variation and Change*, Cambridge, v.3, n.1, p.23-32, 1992.

SCHERRE, M. M. P.; RONCARATI, C. Programa de estudos sobre o uso da língua (PEUL): origens e trajetórias. In: VOTRE, S.; RONCARATI, C. *Anthony Julius Naro e a linguística no Brasil*: uma homenagem acadêmica. Rio de Janeiro: 7Letras, 2008. p.37-49.

TESCH, L. M. *A expressão do tempo futuro no uso capixaba*: variação e gramaticalização. 2011. 191f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

VIEIRA, M. C. P. *A emergência do padrão flexional variável de 3a pessoa do plural na aquisição do PB como L1*. 2006. 120f. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

WEINREICH, U; LABOV, W; HERZOG, M. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2006.

_____. Empirical foundations for a theory of language change. In: LEHMANN, W.; MALKIEL, Y. (Ed.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p.97-195.

Recebido em abril de 2012.

Aprovado em julho de 2012.

PROJETO PORTVIX: A FALA DE VITÓRIA/ES EM CENA

Lilian Coutinho YACOVENCO*
Maria Marta Pereira SCHERRE**
Leila Maria TESCH*
Marcela Langa L. BRAGANÇA***
Elaine Meireles EVANGELISTA*
Alexandre Kronemberger de MENDONÇA****
Elba Nusa CALMON*
Heitor da Silva CAMPOS JÚNIOR*****
Astrid Franco BARBOSA*****
Jucilene Oliveira Sousa BASÍLIO*****
Carlos Eduardo DEOCLÉCIO*****
Janaína Biancardi da SILVA*
Aline Tomaz Fonseca BERBERT*
Samine de Almeida BENFICA*

- RESUMO: Os objetivos centrais do presente artigo são (1) apresentar à comunidade científica o projeto PortVix, que trata da fala da cidade de Vitória, capital do Espírito Santo; (2) sintetizar os trabalhos desenvolvidos e em desenvolvimento, com base nos dados deste projeto; e (3) tecer considerações sobre o alinhamento da fala de Vitória ao cenário nacional e sobre traços que a caracterizam. A Teoria da Variação e de Mudança Linguística norteia a organização da amostra e, de forma geral, a análise de diversos fenômenos variáveis, a saber, alternância entre futuro do pretérito e pretérito imperfeito; a expressão variável do futuro do presente; usos do

* UFES – Universidade Federal do Espírito Santo - Centro de Ciências Humanas e Naturais. Vitória – Espírito Santo - Brasil. 29075-910 - liliannyacovenco@yahoo.com.br; leilatesch@yahoo.com.br; elainemeirelesevangelista@yahoo.com.br; elbacalmon@yahoo.com.br; janaina_biancard@hotmail.com; aline_bt@hotmai.com; saminebenfica_@hotmail.com

** UFES – Universidade Federal do Espírito Santo – Centro de Ciências Humanas e Naturais. Vitória – Espírito Santo – Brasil. 29075-910 - mscherre@terra.com.br

*** UFFS – Universidade Federal da Fronteira Sul - Colegiado de Engenharia de Aquicultura e de Interdisciplinar em Educação no Campo. Laranjeiras do Sul – Paraná – Brasil. 85301100 - marcelalanga@uffs.edu.br

**** SEDU-ES – Secretaria Estadual de Educação do Espírito Santo – SEDU-PMS – Secretaria de Ensino do Município de Serra. Serra – Espírito Santo – Brasil. 29176-100 - alkomen@yahoo.com.br;jucilenem3@hotmail.com

***** SEDU-ES – Secretaria Estadual de Educação do Espírito Santo. Vitória – Espírito Santo – Brasil. 29052-908. heitorscj@gmail.com;astridfb@terra.com.br

***** IFES – Instituto Federal do Espírito Santo - Coordenadoria de Linguagens. Vitória – Espírito Santo – Brasil. 29040-780 - deoclecio@ifes.edu.br

gerúndio; variação sintática das orações adverbiais finais; alternância *nós/a gente*; a expressão do objeto direto anafórico; concordância nominal; concordância verbal; expressão gramatical do imperativo; alternância indicativo/subjuntivo; alternância *você/cê/ocê*; ausência/presença de artigo diante de antropônimos e de possessivos. Observações do senso comum e resultados das análises permitem evidenciar que, ao lado de aspectos que alinham a fala de Vitória-ES ao cenário nacional, corroborando a generalidade de diversos fenômenos, há outros aspectos que permitem inferir a caracterização da fala capixaba face aos três estados com os quais o Espírito Santo estabelece fronteiras: Rio de Janeiro, Bahia e, especialmente, Minas Gerais.

- PALAVRAS-CHAVE: Projeto PortVix. Fala capixaba. Português brasileiro. Variação linguística. Mudança linguística. Variação regional.

Introdução

A identidade linguística do capixaba não é fenômeno sobre o qual se tenha grande conhecimento. Pode-se afirmar, inclusive, que traços característicos dessa identidade não são imediatamente nítidos para quem entra no Espírito Santo pela Grande Vitória (formada pelas cidades de Vitória, Vila Velha, Serra, Cariacica e Viana), nem mesmo para os próprios capixabas. Muitas são as razões para tal desconhecimento e, para suprir tal lacuna na descrição do português brasileiro (PB), foram constituídos dois grandes projetos de pesquisa, vinculados, inicialmente, ao Departamento de Línguas e Letras (DLL), e, posteriormente, ao Programa de Pós-Graduação em Estudos Linguísticos (PPGEL), do Centro de Ciências Humanas e Naturais (CCHN), da Universidade Federal do Espírito Santo (UFES).

Na área da Sociolinguística, a professora Lilian Coutinho Yacovenco organizou e coordena o Projeto Português Falado na Cidade de Vitória – PortVix. Este projeto, de orientação variacionista, gravou, entre 2001 e 2003, quarenta e seis entrevistas com pessoas nascidas em Vitória, divididas segundo as variáveis relativas ao sexo/gênero, idade e escolaridade, sobre o qual apresentamos detalhes mais à frente, tendo em vista que sua constituição e relatos de pesquisas com base em seus dados são o foco deste texto.

No campo da Geografia Linguística, a professora Catarina Vaz Rodrigues organiza e coordena, desde 2005, o Atlas Linguístico do Espírito Santo (ALES), projeto que tem por objetivo a descrição das principais variantes diatópicas do Espírito Santo, um Estado que apresenta uma complexa rede de migração europeia, à semelhança da Região Sul do Brasil. A partir dos pontos observados (RODRIGUES, 2006), foram possíveis pesquisas sobre alguns deles. Destaca-se o interesse pela área de Santa Maria de Jetibá, em que o bilinguismo da população, falante de pomerano e português, é ponto de interesse de pesquisadores, especialmente pela manutenção do pomerano como primeira língua até, pelo menos, a quinta geração.

Esses dois projetos são um primeiro passo para que se descreva a variedade linguística capixaba e já se podem ver seus primeiros frutos, seja por meio de

comunicações¹ apresentadas em congressos nacionais e internacionais, seja por meio de dissertações concluídas ou iniciadas no PPGEL/UFES.

Desde 1953, quando Antenor Nascentes publica *O linguajar carioca*, se observa uma busca por uma caracterização dos diversos falares que constituiriam o português brasileiro. Segundo a divisão dialetal proposta por Nascentes, baseada na expressão das vogais médias pretônicas, o Espírito Santo comporia, juntamente com o Rio de Janeiro e com parte do leste de Minas Gerais, o subfalar Fluminense, conforme se verifica no mapa abaixo.

Figura 1 - Divisão dialetal do Brasil proposta por Antenor Nascentes. Há seis subfalares: Amazônico, Nordestino, Baiano, Mineiro, Fluminense, Sulista, além de um território indefinido



Fonte: Nascentes (1953, p.19).

O presente artigo objetiva, portanto, situar a fala capixaba no cenário da língua portuguesa falada no Brasil, apresentando uma breve síntese de 13 trabalhos realizados com dados coletados pelo PortVix. Dessa forma, tenciona descrever a variedade capixaba, confrontando os resultados dessas pesquisas com os relativos a outras variedades do PB. Pretende, também, na medida do possível, verificar se a fala capixaba possui marcas próprias ou se, ao contrário, é não marcada, isto é, se nela prevalecem os traços mais gerais do português brasileiro, o que implica

¹ Yacovenco (2009).

dizer que há ausência de marcas dialetais claras. É, então, objetivo desse artigo evidenciar, a partir de observações empíricas e de estudos já realizados sobre a fala de Vitória, alguns traços linguísticos que parecem distinguir a fala capixaba da realizada em outros estados brasileiros, mais especificamente dos vizinhos com os quais o Espírito Santo estabelece fronteira: Minas Gerais (a Oeste e Noroeste), Rio de Janeiro (ao Sul) e Bahia (ao Norte).

A colonização do Espírito Santo e configuração etnográfica de seu povo

O Espírito Santo foi uma das capitarias hereditárias fundada pelos portugueses no início da colonização do Brasil. Em 23 de maio de 1535, Vasco Fernandes Coutinho chega ao Espírito Santo e funda a sede da capitania na hoje Prainha, em Vila Velha. Vários escritos mostram que a colonização não foi tarefa fácil e que em 8 de setembro de 1555, após vitória sobre invasores, a sede da Capitania, que havia sido transferida para a atual ilha de Vitória, é aí estabelecida definitivamente (GONÇALVES, 2000).

A capitania do Espírito Santo não teve um desenvolvimento rápido, nem de sucesso, fato também ocorrido em várias outras capitias. Algumas das principais razões apontadas são a resistência indígena à colonização portuguesa e a parca disponibilidade financeira do donatário, Vasco Fernandes Coutinho. É certo que o Espírito Santo não possuiu um desenvolvimento contínuo e, após a descoberta do ouro em Minas Gerais, teve ainda mais dificuldades para seu crescimento, uma vez que as fronteiras do estado foram fechadas para que o ouro e os diamantes descobertos e retirados das Minas Gerais não fossem contrabandeados pelo Espírito Santo. Foi estabelecida uma espécie de “barreira verde” e o desenvolvimento do território capixaba se dava a passos muito lentos.

Salletto (2000) discute a configuração étnica do Espírito Santo e aponta que a presença dos índios Tupiniquim, Tupinambá e Tremembé, até o meado do século XVII, é marcante no estado. Após essa data, inicia-se fortemente a importação de escravos negros, que iriam trabalhar nas lavouras de açúcar. Entretanto, em 1640, ocorre um período de declínio da capitania espiritossantense, devido ao fim do domínio espanhol e, por conseguinte, ao confisco de terras de vários cristãos novos, acusados de apoiarem o governo espanhol. Este fato abalou a produção açucareira e, logo a seguir, a crise mundial do açúcar encerra esse segundo ciclo de exploração de terras brasileiras.

Salletto (2000) mostra que até 1850 há um enorme incremento do número de escravos negros em terras capixabas em função do plantio de café no sul do estado. A população do Espírito Santo, formada até então, predominantemente, por portugueses e índios, passa por um aumento significativo da população parda.

A partir das duas últimas décadas do século XIX, inicia-se a imigração europeia. Há um grande número de imigrantes alemães, que se deslocam, inicialmente, para Santa Isabel, no atual município de Domingos Martins, zona serrana do Espírito Santo, de clima mais ameno. A população de descendência alemã, em 1960, correspondia a 5% da população capixaba. Apesar da chegada dos alemães, a imigração mais significativa foi a italiana, que, segundo estimativas, fez com que a população capixaba se tornasse uma das maiores colônias italianas do Brasil e, ainda segundo estimativas, 65 a 70% da população capixaba tem origem italiana. Segundo Saletto (2000, p.108), a imigração estrangeira teve forte peso na composição étnica do estado, pois:

[...] diferentemente do que ocorreu [em Minas Gerais e no Rio de Janeiro], no Espírito Santo a imigração estrangeira teve um forte peso, e os imigrantes eram predominantemente de tipo claro e cabelos louros, da Europa Central e do norte da Itália. Áí, sim, estaria uma especificidade capixaba: um povo [antes] com forte herança africana, no qual a presença dos imigrantes louros foi muito mais marcante do que nos outros estados nos quais os descendentes de africanos são numerosos, inclusive Minas e o Rio de Janeiro na sua configuração atual, englobando o antigo Distrito Federal, pois nessa cidade a numerosa imigração estrangeira foi principalmente ibérica, sobretudo portuguesa.

Essa configuração etnográfica da população capixaba revela uma forte miscigenação e, também, um vasto campo de estudos para a Linguística, especialmente para os interessados em pesquisas de cunho dialetológico e sociolinguístico. Dessa forma, podemos dizer que o Espírito Santo é um grande “caldeirão” linguístico, em que inúmeras influências vindas de contatos diversos merecem estudos linguísticos. É, também, um grande campo de estudo para a caracterização de uma variedade não marcada. Podemos perguntar se a miscigenação étnica e linguística ocorrida no estado contribuiu para a ausência de marcas linguísticas na variedade dos capixabas ou se, ao contrário, proporcionou diferenças marcantes a depender da região do Espírito Santo sobre a qual recaia a análise.

Vitória, conforme acima exposto, é uma das capitais mais antigas do Brasil, porém, por a colonização do solo espiritossantense ter-se dado de modo descontínuo, não houve um desenvolvimento semelhante ao das demais cidades fundadas na mesma época. O isolamento do estado e a forte imigração europeia na segunda metade do século XIX mudaram a configuração etnográfica do Espírito Santo e, por conseguinte, de sua capital, Vitória. A partir da imigração europeia e, por consequência, com o desenvolvimento da lavoura de café, inicia-se uma nova fase econômica no ES. No caso específico de Vitória, a partir de 1970, com a implementação dos projetos industriais e com a fixação de duas grandes

companhias, a Vale do Rio Doce e a Companhia Siderúrgica Tubarão (hoje, Acelor Mittal), houve um forte movimento de migração para a capital, especialmente de jovens filhos de imigrantes que partiam para a capital para dar continuidade a seus estudos e para buscar novas oportunidades de trabalho nas grandes companhias e no setor de serviços, alavancado pelo incremento da área portuária da cidade. É essa cidade, de colonização antiga, porém descontínua, que é o objeto do Projeto PortVix.

O projeto PortVix

Conforme visto acima, a configuração etnográfica do ES pode ter contribuído para que em Vitória houvesse uma variedade não marcada. Aliado a esse fator, o isolamento da cidade, durante séculos, também pode ter contribuído para essa característica. Assim, a configuração da variedade capixaba ainda não é bem clara, nem para a comunidade acadêmica nem para os leigos nem mesmo para os próprios habitantes do Espírito Santo. Algumas marcas lexicais são percebidas, destacando-se o verbo “pocar”, com diversos sentidos, como em a) O menino **pocou** o balão de gás; b) Ele se **pocou** de rir da piada contada; c) Ele **pocou** a cabeça no acidente e d) A praia ‘**tá pocando**’. Outra marca seria o termo “gastura”, que significa “agonia”, ou, ainda, o uso de “taruíra” por “lagartixa”. Outra característica é a expressão “ir pru rock”, mesmo quando a dança for qualquer outro ritmo ou mesmo significar “ir pra balada”.

Assim, Vitória, capital do estado, apesar de já ter completado mais de 460 anos de fundação, não possuía um registro sistemático de sua língua oral, fato este que impossibilitava estudos de natureza sociolinguística. Surge, então, o interesse na formação de um banco de dados de cunho sociolinguístico, com o intuito de registrar o vernáculo dos habitantes da capital capixaba. Em 2000, após parceria com Maria da Conceição Paiva e Christina Abreu Gomes, do Programa de Estudos sobre o Uso da Língua (PEUL/RJ), que ministraram curso sobre formação de um banco de dados sociolinguístico, surge o Projeto “O português falado na cidade de Vitória” (PortVix).

O PortVix, projeto de orientação variacionista, gravou quarenta e seis entrevistas com informantes nascidos em Vitória, divididos segundo as variáveis relativas ao sexo/gênero do informante, à sua idade e à sua escolaridade. Após a formação das células, dividiram-se, por meio de sorteio, os perfis dos entrevistados pelas sete regiões administrativas da cidade, estabelecendo-se a seguinte tabela:

Tabela 1 - Distribuição das células sociais

(idade [→])	07-14		15-25		26-49		50-...	
	H	M	H	M	H	M	H	M
Ensino Fundamental	4	4	2	2	2	2	2	2 =20
Ensino Médio			3	3	2	2	2	2 =14
Ensino Universitário			2	2	2	2	2	2 =12
	<i>Número total de entrevistados</i>							=46

Fonte: Yacovenco (2002, p. 108).

O PortVix buscou o vernáculo, isto é, conforme define W. Labov, “[...] o estilo em que se presta o mínimo de atenção ao monitoramento da fala” (LABOV, 2008, p.244). Entretanto, as entrevistas registram não o vernáculo, mas, sim, a fala monitorada (*careful speech*), que é “[...] o tipo de fala que normalmente ocorre quando a pessoa está respondendo perguntas que são formalmente reconhecidas como “parte da entrevista” (LABOV, 2008, p.102). Buscaram-se, então, diversos procedimentos para que o vernáculo emergisse, isto é, para que a pessoa prestasse menos atenção à sua própria fala (ao *como* dizer) e passasse a ficar mais atenta ao próprio enunciado (sobre o *que* dizer). Um desses procedimentos foi a busca do envolvimento emocional com o próprio fato enunciado, sendo um dos procedimentos a pergunta sobre situações de risco de vida pelas quais a pessoa já houvesse passado. Entre os procedimentos metodológicos, destaca-se que as entrevistas foram realizadas, em sua maioria, por dois entrevistadores, num grupo de seis a oito pessoas treinadas para a realização da tarefa.

No atual momento, o projeto está em fase final de revisão das entrevistas, para em breve disponibilizar, por meio eletrônico, toda a amostra aos interessados. As transcrições já foram realizadas, possibilitando o acesso a alguns pesquisadores, especialmente aos alunos do PPGEL. Assim, já há análises concluídas e, outras, em andamento, como as que serão comentadas.

As pesquisas do PortVix

A Teoria da Variação e da Mudança Linguística de base laboviana (WEINREICH; LABOV; HERZOG, 2006; LABOV, 2008) norteia a organização da base de dados e as análises de diversos fenômenos variáveis. A ordem de apresentação dos trabalhos tem como base um gradiente de maior a menor alinhamento com o cenário linguístico nacional, para que se chegue aos objetivos do presente texto, que são a descrição da fala capixaba e a reflexão sobre sua caracterização no cenário nacional. Sendo assim, os 12 temas estudados serão apresentados na seguinte ordem:

- 1) Alternância entre futuro do pretérito e pretérito imperfeito (TESCH, 2007);
- 2) Expressão variável do futuro do presente (BRAGANÇA, 2008; TESCH, 2011);
- 3) Usos do gerúndio (BASÍLIO, 2011);
- 4) Variação sintática das orações adverbiais finais (DEOCLÉCIO, 2011)
- 5) Alternância *nós/a gente* (MENDONÇA, 2010);
- 6) Expressão gramatical do imperativo (EVANGELISTA, 2010);
- 7) Alternância indicativo/subjuntivo (BARBOSA, 2011);
- 8) A expressão do objeto direto anafórico (YACOVENCO; BERBERT, 2012);
- 9) Concordância nominal (SILVA; SCHERRE, 2012);
- 10) Concordância verbal (BENFICA; SCHERRE, 2012);
- 11) Alternância *você/cê/ocê* (CALMON, 2010);
- 12) Ausência/presença de artigo diante de antropônimos e de possessivos (CAMPOS JÚNIOR, 2011).

Os trabalhos que constam deste texto analisam, em maior ou menor grau, dados do PortVix, mas nem todos, em função de seus objetivos, analisam só os dados do PortVix. Para alguns, foi necessária a análise de dados escrita e da mídia televisiva capixaba e outros precisaram de amostras complementares para melhor entendimento da variação em jogo. Nos momentos oportunos, explicitaremos esta questão, bem como apresentaremos exemplos dos fenômenos analisados. A apresentação levará em conta aspectos mais gerais dos trabalhos, tendo em vista os objetivos aqui traçados.

Alternância entre futuro do pretérito e pretérito imperfeito

A primeira dissertação realizada com base na amostra PortVix é de Leila Maria Tesch, que, na Iniciação Científica, participou efetiva e ativamente da composição do banco de dados. Em sua dissertação, desenvolvida no Programa de Pós-graduação em Linguística da UFRJ, sob a orientação de Vera Lúcia Paredes Silva, estuda a variação, no âmbito do *irrealis*, do futuro do pretérito e do pretérito perfeito na amostra capixaba, tendo por fundamentação teórica a Sociolinguística Variacionista e o Funcionalismo, especialmente as ideias de Joan Bybee (apud TESCH, 2011, p.49) quanto ao papel da frequência de uso ou de ocorrência (*token*) e de tipo (*type*) na variação linguística.

Tesch (2007) identifica inicialmente quatro variantes, a seguir exemplificadas:

Futuro do pretérito (FP) “não é muito bom colocar uma criança no mundo
não... acho que a gente **prejudicaria** elas”(TESCH, 2007, p.74)

Pretérito imperfeito não perifrástico (IA) “se você dissesse que apanhou
da professora **apanhava** de novo” (TESCH, 2007, p.74)

Pretérito perfeito perifrástico (IA+V infinitivo) “um excelente pedreiro...
técnico de segundo grau que eu falei... ele deveria dar aula no ensino técnico...
ele **ia pegar e fazer**... é complemente diferente”(TESCH, 2007, p.32)

Futuro do pretérito perifrástico (IRIA+V infinitivo) “**iria gerar** desemprego
pro Estado”

Seguem, abaixo, resultados com todos os dados (Tabela 2), excluindo-se, da presente análise, os resultados relativos a verbos modais, que apresentam comportamento semântico distinto em função das variantes identificadas (TESCH, 2007, p.31-34). Na Tabela 2, os resultados de Tesch serão comparados aos de Costa (apud TESCH, 2007, p.54), que tratou do mesmo tema com dados do PEUL/RJ.

**Tabela 2 - Distribuição das variantes de futuro do
pretérito no PortVix e no PEUL com todos os dados**

Amostra	Futuro do pretérito	Pretérito imperfeito	IA + V infinitivo	IRIA + V infinitivo
PortVix (TESCH, 2007)	443/1080 41%	402/1080 38%	229/1080 21%	06/1080 0%
PEUL (COSTA, 1997)	287 34%	346 41%	194 23%	16 2%

Fonte: Adaptação da Tabela 3 de Tesch (2007, p.67). Dados relativos a Costa (apud TESCH, 2007) apresentados em termos de frequência relativa (Tesch, 2007, p.54).

Para evidenciar a oposição entre formas do futuro do pretérito e de pretérito imperfeito, Tesch (2007) amalgamou o futuro do pretérito à perífrase IRIA + infinitivo e o pretérito imperfeito à perífrase IA + infinitivo, conforme pode ser visto na Tabela 3.

Tabela 3 - Distribuição das variantes no futuro do pretérito e pretérito imperfeito do indicativo com todos os dados do PortVix e do PEUL

Amostra	Futuro do pretérito e IRIA + V infinitivo	Pretérito Imperfeito e IA + V infinitivo
PortVix	449/1080	42 %
PEUL	303/843	36 %

Fonte: Tabela 6 de Tesch (2007, p.70), com adaptações. Dados relativos a Costa (1997) apresentados em termos de frequência relativa (Tesch, 2007, p.54).

Na Tabela 3, podemos verificar que, mesmo que haja diferenças entre os resultados das duas amostras, é nítida a alternância entre as formas de expressão do futuro do pretérito, havendo, na fala carioca, uma variação um pouco mais forte do que na fala capixaba, com predomínio, em ambas as amostras, de formas ligadas ao pretérito imperfeito, forma mais inovadora.

A seguir, na Tabela 4, relativa aos dados do PortVix, observamos o efeito da faixa etária sobre a perífrase formada por IA + infinitivo contraposta ao futuro do pretérito em verbos não modais.

Tabela 4 - Influência do grupo de fatores faixa etária na escolha de IA + Infinitivo vs. futuro do pretérito em verbos não modais

FATORES	APLIC/TOTAL	FREQ. %	PESO RELATIVO
7 a 14 anos	53/130	40 %	0,64
15 a 25 anos	84/288	29 %	0,51
26 a 49 anos	35/140	25 %	0,42
> 49 anos	50/154	32 %	0,42
Total	222/712	31 %	

Fonte: Tabela 25 de Tesch (2007, p.105), com adaptações.

Tesch (2007, p.103-108) afirma, então, com base nos resultados da Tabela 4, que há uma mudança em curso na variedade capixaba, já que os falantes mais jovens são os que mais favorecem a perífrase com verbo IR no pretérito imperfeito. Conforme visto nas Tabelas 2 e 3, a fala capixaba se aproxima da carioca, havendo, nesse caso, um alinhamento com o cenário linguístico nacional.

Expressão variável do futuro do presente

Bragança (2008) estuda a gramaticalização do verbo IR em dois *corpora*: um, composto por 12 falantes de nível universitário do PortVix, e outro, formado por

editoriais do Jornal *A Gazeta*, jornal mais antigo do estado do Espírito Santo e destinado às classes de maior poder aquisitivo. Os resultados globais da pesquisa de Bragança (2008, p.89) encontram-se na Tabela 5:

Tabela 5 - Distribuição geral das variantes de futuro do presente no PortVix e em *A Gazeta*

	PortVix	Editorial de <i>A Gazeta</i>	
Futuro do presente	4	1,12 %	622
Ir (Futuro) + V infinitivo	1	0,38 %	124
Ir (Presente) + V infinitivo	350	98,5 %	81
Total	355	100 %	827
			100 %

Fonte: Tabela 1 de Bragança (2008, p.89), com adaptações.

Bragança aponta no PortVix uma mudança praticamente concluída na expressão do futuro: a forma sintética possui uma frequência bastante baixa e a forma inovadora, composta pelo verbo *Ir + infinitivo*, tem uso quase categórico. Na escrita, em textos de caráter mais formal, como os editoriais do jornal de maior influência de Vitória, a forma inovadora, apesar de ter frequência de, aproximadamente, 25%, começa a aparecer com maior destaque. Dessa forma, a autora conclui que há dois processos em curso, um na oralidade, que se apresenta praticamente concluído, e outro na escrita, que está em curso.

Em sua tese de doutorado², Tesch (2011) também analisa a expressão de futuro do presente. Além da amostra PortVix, a autora estuda alguns gêneros do jornal *A Gazeta* das décadas de 30, 70 e 2000, com o objetivo, agora, de verificar a mudança em tempo real. A seguir, na Tabela 6, demonstram-se os percentuais globais encontrados (TESCH, 2011, p.122, p.129, p.141 e p.164):

Tabela 6 - Distribuição geral das variantes do futuro do presente no PortVix e em *A Gazeta*

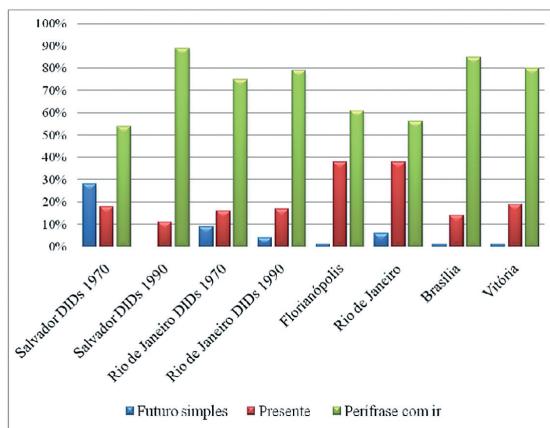
		PortVix	<i>A Gazeta</i> 2010	<i>A Gazeta</i> 1970	<i>A Gazeta</i> 1930		
Futuro	do	4	0,4 %	368	54,3 %	324	81,8 %
presente						178	88,1 %
Ir (Futuro) + V		0	0 %	16	2,4 %	03	0,8 %
Presente		206	19,1 %	158	23,3 %	27	6,8 %
Ir (Presente) + V		867	80,5 %	136	20,1 %	42	10,6 %
Total		1077	100 %	678	100 %	396	100 %
						202	100 %

Fonte: Adaptação das Tabelas 8, 9, 15 e 37 de Tesch (2011, p.121, p.128, p.140 e p.163).

² Assim como sua dissertação de mestrado, Tesch desenvolveu sua tese de doutoramento no Programa de Pós-graduação em Linguística da UFRJ, sob a orientação de Vera Lúcia Paredes da Silva.

Ao confrontar seus resultados com os obtidos por outras pesquisas baseadas em dados de outras cidades brasileiras, Tesch (2011) ratifica que a mudança na expressão do futuro do presente na fala capixaba está alinhada ao que ocorre no cenário nacional, conforme podemos ver no Gráfico 1.

Gráfico 1 - Comparação dos resultados de pesquisas sobre a expressão do futuro do presente em diversas localidades brasileiras



Fonte: Correspondente ao gráfico 10, de Tesch (2011, p.181).

Podemos, portanto, notar que a mudança na expressão do futuro verbal ocorre na fala e na escrita do PB. Entretanto, há diferenças quanto ao estágio da mudança, quando confrontamos os resultados das duas modalidades: enquanto, na fala, a mudança já se encontra concluída, na escrita, com base nos dados de *A Gazeta*, podemos confirmar a mudança em curso em tempo real, ao observarmos que o futuro do presente, em sua forma sintética, começa a ceder lugar à forma perifrástica.

Usos do gerúndio

Em Basílio (2011), foi examinado o uso variável de estruturas formadas por perífrases e construções complexas com *estar+gerúndio*. Buscou-se entender os fatores que influenciam a alternância no campo do presente frequentativo, com leitura habitual, em construções como “[...] como você sempre se esquece de tudo: a gente TÁ FAZENDO questão de ligar pra te lembrar da festa...”, no campo infinitivo não futuro, em enunciados do tipo “*o secretário veio ai pra TÁ ASSINANDO o documento atual...*”, e no campo do infinitivo futuro, em estruturas como “*Primeiro é sobre o seminário que a gente VAI ESTAR FALANDO*”.

Em um estudo piloto, Basílio (2011) analisa minuciosamente uma das entrevistas do PortVix, e codifica 392 dados, a fim de delimitar os contextos de realização das ocorrências do fenômeno a ser investigado. Todavia, nos 392 dados codificados, só foram encontrados nove casos de perifrases com gerúndio, sendo cinco dados do campo do presente frequentativo e dois do campo do infinitivo (um futuro e outro não futuro), ou seja, apenas 2,3% da configuração de *estar+gerúndio*.

Então, Basílio (2011) foi à busca de dados pertinentes em outras amostras de fala, de natureza supostamente mais casual, imaginada como sendo o *locus* deste tipo de variação, embora esta tentativa já tivesse sido também levada a cabo, sem sucesso, por Santos (2008). A segunda tentativa foram duas pequenas gravações, ambas em Vitória, com dados da fala de dois adolescentes de 13 anos de idade. Todavia, Basílio (2011) também não logrou sucesso na obtenção de perifrases com gerúndio, que tanto nos chamam atenção no dia a dia.

Assim, Basílio (2011) adotou o método da coleta de eventos de fala anônima, denominado por Labov (2008, p.246) de “observações assistemáticas”, que se caracteriza pela busca de dados nos usos públicos da língua, para captar o maior número de dados apropriados à pesquisa pretendida. Basílio (2011, p.36-38) organizou a coleta de dados com gravações de falas, também em Vitória, em seu contexto social natural, porém em ambientes que caracterizam contextos de gêneros discursivos como **o relato de procedimento, relato de opinião, relato de experiência vivida, relato de acontecimentos**, em ocasião de planejamento escolar, momentos informativos em ambientes religiosos, reuniões de departamento, atuação institucional, debates, anúncios orais públicos, reuniões escolares e intervenções pedagógicas, e outros contextos argumentativos tais como propagandas, vendas, atendimento a clientes, noticiário de jornal. O *corpus* de Basílio (2011), submetido a tratamento quantitativo foi, então, em grande parte, constituído de gravações em locais públicos, mas também foram realizadas gravações rápidas e anônimas com a participação enunciativa da pesquisadora. A base de dados que Basílio (2011) organizou é formada de 26 gravações, com cerca de 17 horas, totalizando 784 dados pertinentes à sua análise.

Para os objetivos do presente texto, é importante salientar que a pesquisa de Basílio (2011) evidencia que a expansão dos usos de gerúndio na capital capixaba é semelhante ao que ocorre em outras partes do português falado no Brasil. Não se trata de uma variação de natureza geográfica ou social, mas, sim, associada a gêneros discursivos específicos. Basílio (2011) relata que o trabalho de Santos (2008, p.12 e p.60) viveu a mesma situação: a dificuldade de captar a variante inovadora, da ordem de 2% em dados de fala informal (6/279), obtidos em gravações de 15 horas do programa “Pânico e Translouca”, das redes Jovem Pan e Transamérica. Para empreender sua análise, Santos (2008) lançou mão de dados da variante inovadora anotados logo que ouvidos, mas o grande desafio

para Basílio (2011) foi exatamente o de captar a frequência da estrutura inovadora, nas situações discursivas propícias ao seu aparecimento, em gravações públicas anônimas, discursivamente orientadas. A base de dados de Basílio (2011) apresenta assim um percentual bastante significativo das estruturas com gerúndio, como se pode ver na Tabela 7, que exibe os percentuais de três amostras.

**Tabela 7 - Distribuição geral de estruturas
estar+gerúndio em três amostras.³**

Amostras	Santos (2008)	Basílio (2011)	Basílio (2011, p.90-100)
Campos de expansão de usos do gerúndio	Amostra informal de gravações de entrevistas nos Programas Pânico e Trasnluca	Uma entrevista do PortVix	- Gravações públicas e anônimas, discursivamente orientadas
<i>estar+gerúndio</i> no campo do presente frequentativo		5	30/137=22,0%
<i>estar+gerúndio</i> no campo do infinitivo não futuro		1	94/417=22,5%
<i>estar+gerúndio</i> no campo do infinitivo não futuro	6/279=2,15%	1	35/155=20,9%
Total	6/279=2,15%	7/392=2,3%	159/709=20,9%

Fonte: Santos (2008); Basílio (2011).

Na amostra constituída de gravações públicas anônimas discursivamente orientadas, a percentagem geral destas estruturas fica na casa dos 21% (nove vezes mais do que nos resultados obtidos na pesquisa do projeto-piloto fundamentado no PortVix).

Variação sintática das orações adverbiais finais

O trabalho de Deoclécio (2011) analisa no PortVix a variação sintática das orações subordinadas adverbiais finais, ora desenvolvidas, ora reduzidas cujas ocorrências se ilustram nestes exemplos: “eu tenho que gerenciar e organizar... fazer um esquema... *pra que tudo funcione muito bem*” (dado de final desenvolvida) e “tinha que voltar a tomar de novo daqui a seis horas, *pra febre passar novamente*” (dado de final reduzida). Deoclécio (2011) explicita que seu trabalho assume também a orientação da Sociolinguística Variacionista, complementada pelo uso de princípios do Funcionalismo norte-americano, como o da marcação de Givón, o da economia e o da iconicidade de Haiman. Além de analisar um

³ A amostra de Santos (2008) é composta de entrevistas dos programas Pânico e Translouca, das emissoras de rádio Jovem Pan e Transamérica. Os dados de Basílio (2011) são retirados das tabelas apresentadas na p.90 e p.100.

corpus da língua falada formado por 19 entrevistas do PortVix, Deoclécio também pesquisou um *corpus* da língua escrita composto por 35 reportagens da revista Superinteressante, contemporâneas ao das entrevistas, com, respectivamente, 302 e 282 estruturas relevantes à análise proposta. Em termos globais, o percentual de finais desenvolvidas foi de 3% (9/302) na fala e 16% (45/282) na escrita. Estamos diante de um fenômeno de natureza semicategórica na fala.

Para os objetivos gerais do presente texto, destacam-se na Tabela 8, como apontado por Deoclécio (2011, p.68), os resultados da fala da capital capixaba, com base na amostra PortVix, em comparação aos de Finck (apud DEOCLÉCIO, 2011), para as orações adverbiais finais na fala da capital catarinense.

Tabela 8 - Distribuição geral das finais desenvolvidas e reduzidas nos dados da fala: Vitória (PortVix) vs. Florianópolis (VARSUL)

Capitais	Desenvolvidas	Reduzidas
<i>Vitória</i>	9/302=3%	293/302=97%
<i>Florianópolis</i>	5/187=3%	182/187=97%

Fonte: Adaptação da Tabela 2 de Deoclécio (2011, p.68).

Deoclécio (2011, p.68) enfatiza “[...] que as duas capitais apresentaram percentuais de ocorrências para as variantes em análise exatamente iguais – 3% de desenvolvidas e 97% de reduzidas –, embora o total de dados seja diferente, isto é, 302 para Vitória e 187 para Florianópolis”.

Deoclécio (2011) conclui que a variação das adverbiais finais não é um fenômeno regional, por estar alinhada aos resultados encontrados em trabalhos desenvolvidos previamente, como os de Finck (apud DEOCLÉCIO, 2011, p.68), para a fala, e os de Azevedo (apud DEOCLÉCIO, 2011, p.69), para a escrita. Considera que se trata, portanto, de um fenômeno substancialmente interno ao sistema linguístico do PB, embora aspectos externos ligados ao letramento - grau de escolarização dos falantes e modalidade falada ou escrita - influenciem a variação em questão, como se pode ver na distribuição global dos dados da final reduzida ao se levar em conta a fala de pessoas de escolarização fundamental e média (1/183=0.5%), a fala de pessoas de escolarização universitária (8/119=6,7%) e os dados da escrita (42/282=16%) (DEOCLÉCIO, 2011).

Em relação à descrição da fala capixaba, os resultados obtidos por Deoclécio permitem que seja observado o alinhamento dessa variedade às demais do PB, especialmente aos resultados obtidos por Finck (apud DEOCLÉCIO, 2001, p.68).

Alternância *nós/a gente*

O objetivo do trabalho de Mendonça (2010, p.6) consiste na descrição e análise do uso das formas variantes *nós* e *a gente* na fala dos moradores de Vitória. No intuito de alcançar o objetivo traçado, foi investigado, em um *corpus* constituído por 40 das 46 entrevistas do PortVix, o papel de fatores linguísticos e sociais no uso das formas variantes *nós* (... *Nós temos três linhas ...*) e *a gente* (*a gente tá economizando*). A análise das variáveis linguísticas revela a importância de fatores como a referencialidade, o tempo verbal, o paralelismo linguístico e a posição sintática.

Quanto à análise dos fatores sociais, destacamos a atuação da variável faixa etária, que, confirmando os estudos de Oména apud Mendonça (2010, p.41), revela um processo de mudança em curso. Segundo Mendonça (2010, p.6), “[...] os resultados deste trabalho indicam que Vitória se alinha a outras pesquisas realizadas no Rio de Janeiro (LOPES), Florianópolis (SEARA), Jaguarão e Pelotas (BORGES) e Porto Alegre (ZILLES) que concluíram que o sistema pronominal do português brasileiro está em pleno processo de mudança”. Os resultados da faixa etária obtidos pela análise de Mendonça (2010, p.62-63) revelam, conforme observamos na Tabela 9, a implementação da forma *a gente*, variante inovadora, que é nitidamente mais favorecida por pessoas mais jovens.

Tabela 9 - Atuação da variável social faixa etária no uso de *a gente* pelos moradores de Vitória

FAIXA ETÁRIA	APLICAÇÃO/OCORRÊNCIAS	%	PESO RELATIVO
07 a 14 anos	269/315	85,4	0,76
15 a 25 anos	454/541	83,9	0,70
26 a 49 anos	244/416	58,7	0,36
50 ou + anos	269/473	56,9	0,23
TOTAL	1236/1745	70,8	

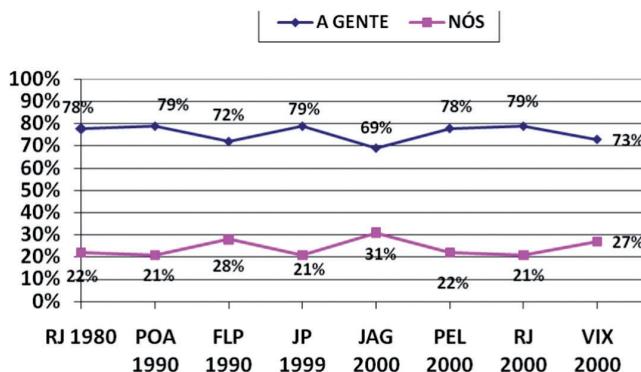
Fonte: Adaptação da Tabela 18 de Mendonça (2010, p.71).

Em sua busca para situar a fala de Vitória no cenário linguístico nacional, Mendonça (2010, p.87) apresenta dois gráficos. Um deles, a seguir, diz respeito à distribuição global das variantes em diversos estudos do PB, os do [...] Rio de Janeiro (OMENA, 1996, 2003); João Pessoa (FERNANDES, 1996); Florianópolis (SEARA, 2000); Jaguarão e Pelotas (BORGES, 2004) e Porto Alegre (ZILLES, 2007)."

Mendonça (2010, p.87) conclui que, em termos percentuais de uso, a introdução da forma *a gente* na função de pronome pessoal vem ocorrendo em todas as

localidades estudadas de modo bastante acentuado, ficando nítida a escolha do falante pela variante *a gente* em detrimento do pronome *nós*. Na comparação entre os efeitos da faixa etária nas diversas localidades, Mendonça enfatiza que, guardadas as diferenças de análise, o padrão de mudança do PB em direção ao uso da forma *a gente* é regular e irreversível e a fala de Vitória/ES segue o padrão nacional.

Gráfico 2 - Percentuais de uso da forma *a gente* nos trabalhos consultados por Mendonça (2010)



Fonte: Gráfico 8 de Mendonça (2010, p.100).

Expressão gramatical do imperativo

A pesquisa de Evangelista (2010) analisa a alternância do imperativo gramatical na cidade de Vitória/ES, no que diz respeito às formas contemporaneamente associadas ao modo indicativo (*fala/olha/deixa/diz*) ou ao modo subjuntivo (*fale/olhe/deixe/diga*) em enunciados afirmativos e negativos da fala e da escrita, no contexto exclusivo do pronome *você*. Foram utilizados quatro amostras, a saber: (1) entrevistas do PortVIX, – diálogos de língua falada; (2) propagandas e títulos de colunas em dois jornais impressos locais, *A Tribuna* e *A Gazeta* – escrita sem formato de diálogo; (3) tirinhas de *Marly, a solteirona*, personagem capixaba criada pelo cartunista e escritor Milson Herinques há mais de 30 anos – escrita com formato de diálogo; e (4) fala da mídia televisiva em dois programas locais, *Balanço Geral* e *Tribuna Notícias*. As amostras utilizadas nos deram uma visão mais ampla das tendências, já observadas em outras pesquisas, acerca da variação e da mudança do imperativo no português brasileiro. O principal objetivo deste estudo foi verificar qual é o alinhamento do uso do imperativo da cidade de Vitória no contexto nacional. Para tanto, vamos nos fixar na apresentação dos

resultados de Evangelista (2010) relativos ao PortVix, após a devida exemplificação do fenômeno estudado a seguir, em que se ilustra a variação já observada por diversos outros pesquisadores com a análise da fala de outras localidades:

[...] “tô vivo aqui, tô conversando com a senhora numa boa, **DÁ** licença”.
“**DÊ** pra alguém criar... sei lá entendeu?”

Na análise dos dados da fala da cidade de Vitória, Evangelista (2010) pesquisou as 34 entrevistas do PortVix, disponíveis à época. Seus resultados confirmaram tendências já reveladas em estudos anteriores sobre a alternância de uso do imperativo gramatical no PB.

Tendo em vista o objetivo de verificar qual a posição da cidade de Vitória no cenário nacional, Evangelista (2010) apresenta análise comparativa entre as capitais Vitória/ES, Rio de Janeiro/RJ e Salvador/BA. Esse interesse se justifica pelo fato de o estado do Espírito Santo estar situado geograficamente, seguindo a linha do litoral, entre estados antagônicos no que diz respeito ao uso do imperativo gramatical: os estados do Rio de Janeiro e o da Bahia, região Sudeste e região Nordeste, respectivamente. A Tabela 10, organizada por Evangelista (2010, p.51), apresenta os resultados para as três capitais.

Tabela 10 - Frequência de uso do imperativo associado à forma indicativa (*fala/vem*) – dados de fala das cidades de Vitória/ES (EVANGELISTA, 2010), Rio de Janeiro/RJ e Salvador/BA (década de 90/séc.XX) (SAMPAIO, 2001)

Localidade	Nº de ocorrências/Total	Porcentagem
Vitória/ES – Sudeste	260/266	97%
Rio de Janeiro/RJ – Sudeste	243/258	94%
Salvador/BA – Nordeste	136/479	28%

Fonte: Adaptação da Tabela 2 de Evangelista (2010, p.51).

Os resultados obtidos na análise de Evangelista (2010) mostram o perfeito alinhamento de Vitória/ES com o Rio de Janeiro/RJ: em ambas as cidades há percentuais elevados de imperativo associado à forma indicativa (*fala/vem*), outro fenômeno também de configuração semicategórica. Apesar de a cidade de Salvador usar o pronome *você*, e não o *tu*, como pronome de segunda pessoa

do singular, o percentual de imperativo associado à forma indicativa é de 28% na capital da Bahia, muito diferente do encontrado na capital do Espírito Santo, da ordem de 97%, que também é de contexto exclusivo do pronome *você* (e sua variante *cê*), como se verá com clareza no trabalho de Calmon (2011), mais adiante. Esse percentual suscita curiosidade a respeito do uso do imperativo nas cidades do Espírito Santo que fazem fronteira com o estado da Bahia, curiosidade esta que pode ser satisfeita com pesquisas futuras.

Alternância indicativo/subjuntivo

O trabalho de Barbosa (2011, p.36) trata da variação de modo indicativo vs subjuntivo em orações completivas e orações absolutas com a presença do modalizador *talvez*, com base em dados extraídos de 43 das 46 entrevistas que compõem o PortVix, exemplificados a seguir:

“...quer que eu **FALE** cada um deles?”

“Quer que eu **EXPLICO** como começou?”

Nos exemplos acima, temos o verbo *querer* na oração matriz ou principal. Esse verbo volitivo possui carga semântica que expressa desejo, criando um ambiente linguístico propício ao uso do subjuntivo na oração subordinada, mas o que se observa é a possibilidade de variação em ambientes de expectativa de uso do subjuntivo. A alternância também pode ser percebida com a presença do modalizador *talvez*, em que há também expectativa de uso da forma subjuntiva. A pesquisa de Barbosa (2011) objetiva entender a variação e o eventual processo de mudança linguística envolvendo este fenômeno.

Também com o objetivo de estabelecer uma comparação com resultados de outras pesquisas e de verificar qual é o comportamento deste fenômeno no cenário nacional, Barbosa (2011), em seu estudo, considera tanto variáveis sociais quanto linguísticas (verbo da oração matriz, assertividade, tempo verbal da oração matriz), as quais são as mesmas adotadas nos trabalhos de Rocha, Carvalho e Oliveira (apud Barbosa, 2011, p.113), que analisam a mesma alternância, respectivamente, no Rio de Janeiro/Brasília, Cariri e João Pessoa.

A variável mais importante foi o verbo da oração matriz, que engloba três grandes campos:

1. Campo de expectativa de uso da forma subjuntiva;
2. Campo de expectativa de alternância da forma subjuntiva e da indicativa;
3. Campo de expectativa de uso da forma indicativa.

Barbosa (2011, p.67) observa que, de forma geral, o campo de expectativa do subjuntivo na oração principal tende a favorecer a forma subjuntiva na oração encaixada e o campo de expectativa do indicativo na oração principal tende a favorecer a forma indicativa na oração subordinada nos dados do PortVix, mas não há categoricidade nem distribuição complementar. Além disso, o verbo *gostar* apresenta comportamento fora das expectativas, tendo em vista que pertence ao grupo de expectativa do subjuntivo, mas o desfavorece. Entretanto, o que chama mesmo a atenção nos resultados obtidos por Barbosa (2011, p.67) é o campo registrado como de alternância, que tem verbos com comportamento diferenciado, desde o que mais favorece a forma subjuntiva na oração subordinada, como o verbo *crer* (0,84), até o que mais a desfavorece, como o verbo *achar* (0,17). Há outras questões em jogo, além da classe semântica do verbo da subordinada, como a frequência de ocorrência do verbo *achar*, que sozinho perfaz 48% dos casos (167/348).

Na Tabela 11, a seguir, podem ser vistos os resultados comparativos de quatro amostras, de forma mais sintetizada. Pelos resultados comparativos, Barbosa (2011, p.113) observa que, no campo de expectativa de ocorrência da forma subjuntiva, a frequência de uso de subjuntivo na subordinada nos dados do Nordeste é quase categórica, da ordem de 97%. Já no Sudeste/Centro-Oeste, embora este também seja o campo que mais favoreça a ocorrência da forma subjuntiva, é evidente a entrada do indicativo no campo de expectativa de uso do subjuntivo: a ocorrência da forma subjuntiva cai para 71%. Como observa Barbosa (2011), os verbos *gostar* e *concordar* têm comportamento bastante diferenciado em Vitória, mas são estes os verbos que mais desfavorecem a forma subjuntiva nos dados do Sudeste/Centro-Oeste, de Rocha (1997), com 59% (17/29).

O fato é que, no campo de expectativa do subjuntivo, o comportamento dos dados de Vitória se alinha ao dos outros da região Sudeste e Centro-Oeste e se distingue, relativamente, do comportamento dos dados da região Nordeste.

Tabela 11 - Comparação da frequência relativa de uso do subjuntivo em função da carga semântica do verbo da matriz nos trabalhos de Rocha (1997), Carvalho (2007), Oliveira (2007) e Barbosa (2011)

	Rocha (1997) Sudoeste e Centro-Oeste (Rio de Janeiro e Brasília)	Carvalho (2007) Nordeste (Cariri-CE)	Oliveira (2007) Nordeste (João Pessoa-PB)	Barbosa (2011) Sudeste (Vitória-ES)
Verbo da oração principal	Janeiro e Brasília)			
Campo 1: expectativa de forma subjuntiva na oração subordinada				
Querer/esperar, desejar/pedir, deixar/temer, ser necessário, gostar/concordar	93/131=71%	31/32=97%	224/227=97%	54/76=71%
Campo 2: expectativa de alternância na oração subordinada				
Imaginar/pensar, acreditar/crer, considerar	44/107=47%	15/33=45%	12/130=9%	21/44=48%
Achar	9/248=4%	21/173=12%	12/533=2%	11/168=7%
Parecer	1/ 17=6%	0/ 35=0%	0/ 1=0%	
Campo 3: expectativa de forma indicativa na oração subordinada				
Saber/falar, dizer/ser certo	3/32=9%	3/38=8%	2/219=0,9%	2/22=9%

Fonte: Adaptação da Tabela 16 de Barbosa (2011, p.113).

Com relação aos campos de expectativa da alternância de indicativo/subjuntivo e de expectativa de uso da forma indicativa, a distinção mais evidente entre as amostras, que não estabelece oposição entre regiões, fica por conta do grupo de verbos *imaginar/pensar/acreditar/crer/considerar*, que tende a desfavorecer a forma subjuntiva nos dados de João Pessoa. O verbo *achar* tem comportamento bastante uniforme – desfavorece a forma subjuntiva na oração subordinada – e as pesquisas ressaltam que um dos fatores que propicia o uso de subjuntivo na subordinada com o verbo *achar* e com os demais verbos de expectativa de uso da forma indicativa é a presença da negação na oração principal.

Em síntese, Barbosa (2011) ressalta que, com relação ao que foge à expectativa do uso do subjuntivo nas orações subordinadas, ou seja, à entrada da forma indicativa no campo de expectativa de uso da forma subjuntiva, Vitória se alinha mais a outras pesquisas com dados das regiões Sudeste e Centro-Oeste, embora, neste caso, use uma forma mais marcada.

A expressão do objeto direto anafórico

Yacovenco e Berbert (2012), tomando por base os estudos de Duarte (apud YACOVENCO; BERBERT, 2012), objetivam observar e sistematizar a mudança

em curso no quadro dos clíticos acusativos no PB, a partir das formas variantes para a expressão do objeto direto anafórico na fala capixaba.

No PortVix, há, assim como relatado por Duarte (apud YACOVENCO; BERBERT, 2012), quatro possibilidades de preenchimento do objeto direto anafórico, a saber:

1) **clíticos acusativos de terceira pessoa**

*(...) é bem melhor você ir comprando peças, pra montar a sua bicicleta... pra você conhecê-**la** (...)*

2) **Pronome lexical (pronome nominativo)**

*"(...) A Fernando Ferrari ... dizem que vão melhorar **ela** (...)"*

3) **Sintagma Nominal Anafórico** (forma plena do SN correferente com outro SN previamente mencionado)

*(...) ninguém respeita mais a polícia... naquela época... o cidadão respeitava a **polícia** (...)*

4) **Sintagma Nominal apagado (Categoria Vazia)**

*(...) você faz um molho de manteiga com alcaparra... de azeite com alcaparra... e joga [**0**] por cima daquela carne branquinha ... fica muito bom!*

A Tabela 12 mostra a distribuição das formas usadas para o preenchimento do objeto direto anafórico. Yacovenco e Berbert (2012) observam que, conforme já afirmado por diversos autores, entre os quais se destacam Amaral, Nascentes, Omena, Duarte e Cyrino (apud YACOVENCO; BERBERT, 2012), o clítico acusativo de 3^a pessoa está em vias de desaparecimento na língua falada. Apenas 0,3% dos 3.238 dados é dessa categoria, caracterizando, inclusive, uma mudança já praticamente concluída.

Tabela 12 - Distribuição das variantes de objeto direto anafórico no PortVix

VARIANTES	FREQUÊNCIA
Clítico	9/3238 = 0,3%
Pronome Lexical	365/3238 = 11,3%
Sintagma Nominal Anafórico	1020/3238 = 31,5%
Categoria Vazia	1844/3238 = 56,9%

Fonte: Berbert; Yacovenco; Scherre (2012).

Confrontamos na Tabela 13 os resultados obtidos no PortVix com os de outras pesquisas, que, evidenciam, de modo irrefutável, uma mudança no preenchimento do objeto direto anafórico no PB: o clítico acusativo de 3^a pessoa cede lugar a outras três variantes. Mesmo com diversidade de análise entre as pesquisas, nota-se que a categoria vazia é a variante mais frequente. Os resultados obtidos com base nos dados de Vitória mostram-se um pouco diversos dos encontrados nas outras duas pesquisas, já que na fala capixaba há uma forte presença do preenchimento do objeto anafórico por sintagmas nominais anafóricos.

Tabela 13 - Comparação dos resultados do preenchimento do objeto direto anafórico em Vitória, São Paulo e Rio de Janeiro

VARIANTES	Vitória (amostra de 2000 - PortVix -) (Yacovenco & Berbert)		São Paulo (amostra de 1980) (Duarte)		Rio de Janeiro (1980) (amostra PEUL, Schwenter & Silva)	
	N	%	N	%	N	%
Clítico	9	0,3	97	4,9	4	0,3
Pronome lexical	365	11,3	304	15,4	151	12,0
SN anafórico	1020	31,5	338	17,1	193	15,4
Categoria vazia	1844	56,9	1235	62,9	909	72,5
Total	3238		1974		1257	

Fonte: Berbert; Yacovenco e Scherre (2012).

Com o objetivo de se observarem os pesos relativos das variantes, foram retirados da análise os clíticos acusativos. Segundo Yacovenco e Berbert (2012), a estratégia de preenchimento primeiramente apontada por Amaral (1981), Nascentes (1953) e Câmara Jr. (1972) é o pronome lexical, mas, por outro turno, essa é a estratégia mais estigmatizada, conforme se vê nas palavras desses mesmos autores. Assim, foram amalgamadas as duas novas estratégias de preenchimento, o sintagma nominal anafórico e a categoria vazia, e contrapostas ao pronome lexical. Foram, então, consideradas estatisticamente significativas as variáveis animacidade do objeto, faixa etária, tempo/modo verbal e gênero. A animacidade do referente também é apontada por Duarte (apud YACOVENCO; BERBERT, 2012) como fator de extrema importância na escolha da variante pronome lexical.

Concordância nominal

O fenômeno linguístico analisado no trabalho de Silva e Scherre (2012) é a presença ou ausência de marca de número nos elementos do sintagma nominal, nos dados do PortVix, sob uma perspectiva laboviana clássica. O fenômeno é exemplificado a seguir: “*eu acho que deve ser uma das drogas*

mais consumidas depois do cigarro" (presença de concordância nominal) e "pro governo ficar com dinheiro mas... tipo assim... os/ **os contraBAndo esses negócio** (...)." (ausência de concordância nominal).

O principal objetivo nesse momento da pesquisa é apresentar resultados de efeitos das variáveis sociais e discutir semelhanças e/ou diferenças com relação, em especial, aos resultados obtidos pelo PEUL/RJ. Até o presente momento, Silva e Scherre (2012) analisaram um total de 9132 elementos nominais, extraídos de 43 das 46 entrevistas que compõem a amostra PortVix. Destes 9.132 elementos nominais analisados, 8.573 apresentaram marcas de plural explícitas, com 89,5% de concordância. Segundo Silva e Scherre (2012), Scherre e Naro registram aumento de concordância no estudo comparativo de duas comunidades aleatórias da cidade do Rio de Janeiro, gravadas pelo PEUL na década de 1980 e na década de 2000. Na década de 1980, havia um percentual global de 71% de concordância nominal, com um aumento de 18 pontos percentuais na década de 2000, que atinge um percentual de 89% de concordância. Na amostra do PortVix, gravada também na década de 2000, foi registrado percentual idêntico de concordância nominal (89,5%). Salienta-se, todavia, que a amostra capixaba abriga falantes universitários e que a amostra do Rio de Janeiro só contempla falantes do ensino fundamental e médio.

Os resultados para a faixa etária, os anos de escolarização e o sexo/gênero obtidos por Silva e Scherre (2012) são apresentados na Tabela 14. Os falantes de 50 anos ou mais e os de 26 a 49, relativamente, são os que menos favorecem a concordância plural, seja em termos percentuais (79,8% e 86,8%) ou em termos de pesos relativos (0,29 e 0,31). Os falantes de 15-25 anos e de 7-14 favorecem sensivelmente o aumento de concordância plural em termos percentuais (93,7% e 96,3%) e em termos de pesos relativos (0,59 e 0,82).

Visto que a concordância plural ocorre com maior frequência na fala dos mais jovens, este fato fornece novos indícios de que, especialmente nas áreas urbanas, está ocorrendo uma mudança no sentido de aquisição da concordância por parte dos membros da comunidade. Trata-se de uma reversão de um processo de perda, relatada em Silva e Scherre (2012), nos termos previstos por Naro (apud SILVA; SCHERRE, 2012), e rediscutidos por Naro e Scherre (apud SILVA; SCHERRE, 2012), cujo modelo para o entendimento deste fenômeno variável é o de fluxos e contrafluxos na grande e complexa comunidade de fala brasileira. A variável faixa etária é bastante robusta, com uma diferença de 53 pontos.

No tocante aos anos de escolarização, os resultados se mantiveram na direção socialmente esperada: os falantes com nível de escolaridade universitária apresentaram maior índice de favorecimento da variante de prestígio, a marca de

plural explícita em elementos do sintagma nominal, com um percentual de 94,1% (e peso relativo de 0,64), seguidos pelos falantes de nível médio (9 a 11 anos de escolarização) com 90,9% (e peso relativo de 0,56) e, por último, observam-se os falantes com nível fundamental de escolaridade, com 85,8% (e peso relativo de 0,38).

No PortVix, os resultados obtidos em relação ao sexo/gênero, como uma macrocategoria, fugiram ao esperado no concernente às tendências que têm sido observadas em outras análises, nas quais se evidencia que as mulheres tendem a favorecer mais concordância do que os homens. Os resultados de Silva e Scherre (2012) evidenciam que, diferentemente, nos dados do PortVix, os homens é que estão na liderança da presença de concordância nominal.

Tabela 14 - Atuação da faixa etária, anos de escolarização e sexo/gênero na concordância nominal em dados da fala de Vitória/ES (PortVix)

Variáveis	Frequência de concordância plural	Peso relativo dos fatores
Faixa etária		
7-14 anos	1379/1432=96,3%	0,82
15-25 anos	3410/3641=93,7%	0,59
26-49 anos	1808/2844=86,8%	0,31
50 anos ou mais	1576/1975=79,8%	0,29
Anos de escolarização		
1-8 anos (Fundamental)	3610/4208=85,8%	0,38
9-11 anos (Médio)	2017/2219=90,9%	0,56
Mais de 11 anos (Universitário)	2546/2705=94,1%	0,64
Sexo/gênero		
Masculino	3987/4288=93,0%	0,58
Feminino	4186/4844=86,4%	0,43
Total	8573/9132=89,5%	

Fonte: Silva e Scherre (2012).

Silva e Scherre (2012) revelam, portanto, que, no que diz respeito aos efeitos das variáveis sociais convencionais faixa etária e anos de escolarização, há alinhamento dos resultados da fala de Vitória/ES com os do Rio de Janeiro, mas, com relação ao efeito da variável sexo/gênero, a tendência é invertida. É necessário um aprofundamento maior da pesquisa e do conhecimento da configuração da sociedade capixaba para que se busquem explicações para tal tendência.

Concordância verbal

O trabalho de Benfica e Scherre (2012) analisa a concordância verbal na amostra PortVix, em dados da seguinte natureza: “**eles ficam fazendo corpo duro** pra não melhorar a situ/o salário... **eles fala que o correio** tá trabalhando no vermelho”. Benfica e Scherre (2012) detectam também maior frequência do uso da concordância verbal na fala dos mais novos (de 7 a 14 anos e de 15 a 25 anos – pesos relativos: 0,73 e 0,64, respectivamente) do que na fala dos falantes mais velhos (de 26 a 49 anos e de 50 a 71 anos – pesos relativos: 0,31 e 0,35), como pode ser visto na Tabela 15. Este fato nos fornece indícios de que está ocorrendo uma mudança, no sentido de aumento da concordância plural verbo/sujeito, na mesma direção de Naro e Scherre para dados do Rio de Janeiro.

Os resultados relativos aos anos de estudos revelam, de forma não surpreendente, que os falantes com escolaridade fundamental apresentam menos concordância (peso relativo de 0,40) do que os falantes com ensino universitário (peso relativo de 0,85).

Tabela 15 - Atuação da faixa etária, anos de escolarização e sexo/gênero na concordância nominal e verbal em dados da fala de Vitória/ES (PortVix)

Variáveis	Concordância nominal (43 falantes)	Concordância verbal (25 falantes)
	Peso relativo dos fatores	Peso relativo dos fatores
Faixa etária		
7-14 anos	0,82	0,73
15-25 anos	0,59	0,64
26-49 anos	0,31	0,31
50 anos ou mais	0,29	0,35
Anos de escolarização		
1-8 anos (Fundamental)	0,38	0,40
9-11 anos (Médio)	0,56	Dados em fase de codificação
Mais de 11 anos (Universitário)	0,64	0,85
Sexo/gênero		
Masculino	0,58	0,56
Feminino	0,43	0,44
Total	8573/9132=89,5%	1555/2003=77,6%

Fonte: Benfica e Scherre (2012).

Os resultados que refletem as diferenças de fala entre os gêneros masculino e feminino nos permitem afirmar que os homens, assim como os resultados obtidos para a concordância nominal por Silva e Scherre (2012), tendem a favorecer um pouco mais a concordância verbal (peso relativo de 0,56) do que as mulheres (peso relativo de 0,44). Esses resultados, embora não sejam polarizados (como normalmente não são as diferenças entre os gêneros), são contrários aos pressupostos teóricos sobre essa variável social em função da noção de prestígio.

Os resultados expressos em termos de pesos relativos para a concordância nominal e verbal são todos paralelos. Então, nos mesmos moldes de Silva e Scherre (2012), observa-se alinhamento dos resultados da fala de Vitória/ES com os do Rio de Janeiro, no que diz respeito aos efeitos das variáveis sociais convencionais faixa etária e anos de escolarização; mas, com relação ao efeito da variável sexo/gênero, a tendência é também invertida.

Alternância *você/cê/ocê*

O trabalho de Calmon (2010) trata do estudo dos pronomes *você*, *ocê*, *cê* e *te* falados na cidade de Vitória/ES. Os dados dessa pesquisa foram extraídos de duas amostras: PortVix e Fala Casual, constituído este por duas gravações, realizadas em Vitória, de conversas sem conhecimento prévio do evento. Calmon (2010) analisou o fenômeno variável em função das variáveis sociais faixa etária, anos de escolarização e sexo/gênero e, também, em função de variáveis linguísticas relativas a aspectos sintáticos (função sintática das variantes) e a aspectos semânticos (referência genérica ou específica). A faixa etária e o sexo/gênero apresentam significância estatística. Com os resultados da faixa etária (Tabela 16), Calmon (2010) infere que está ocorrendo uma mudança da forma *você* para *cê* na faixa etária de 26 a 49 anos. Contudo, observa que, na faixa etária 15 a 25 anos, há indício de reversão dessa mudança, mas de *cê* para *você*, que se acentua na faixa 7-14 anos. São estes os resultados de Calmon (2010) que serão explorados neste texto, tendo em vista que há fortes indícios de que, com o retorno do pronome *você* à fala capixaba, aliado à quase inexistência da forma *ocê* e à sua rejeição em Vitória, os capixabas estejam marcando distinção com relação aos mineiros, que têm forte presença no estado do Espírito Santo, e que utilizam com mais frequência as formas *cê* e *ocê* em suas falas cotidianas. Em Calmon (2010, p.36 e p.68), exemplos da alternância, dois transcritos abaixo, podem ser encontrados com as variantes explícitas, mas os zeros, devidamente controlados, também foram considerados na análise:

Num é que ele andou, se você fazer os cálculos, cê gasta até menos, mas cê tem que contar caminhão, eh, ônibus, é uma parada... Mas se você viajar à noite, uma hora dessa no “reveillon”, você num vai pegar ningum na estrada.

A distribuição global dos dados do PortVix para a alternância *você/cê* revela um percentual de 74,44% de uso de *você* (1.696/2.190), que se eleva para 75,50 (1.477/1.954) se forem retirados os respectivos casos de omissão do pronome (477 casos de *você* e 107 casos de *cê*). Controlados os contextos de efeito categórico da variante *cê* (por causa da questão da cliticização), Calmon (2010) submeteu 1.819 dados a cálculos de pesos relativos, dos quais 1.350 são da forma *você*, ou seja, 74,20%. Como já observado, os resultados da variável faixa etária da análise de Calmon (2010, p.99) para os dados do PortVix revelam que a linha de mudança que se observava de *você* para *cê* na faixa de 26 a 49 anos (0,164) começa a mudar na faixa de 15-25 anos (0,310) e aumenta na faixa de 7-14 anos (0,574), como pode ser visto em detalhes na Tabela 16, a seguir.

Tabela 16 - Efeito da faixa etária do falante no uso do pronome *você* vs. *cê* na amostra do PortVix

Faixa etária	Frequência absoluta de você	Total das formas você e CE	Percentual da forma você	Peso relativo dos fatores
7 a 14 anos	113	130	86,90%	0,547
15 a 25 anos	450	689	65,30%	0,310
26 a 49 anos	140	320	43,80%	0,164
Mais de 49 anos	647	680	95,10%	0,823
Total	1350	1819	74,20%	

Fonte: Tabela 23 de Calmon (2010, p.99), com adaptações.

A hipótese de Calmon (2010) para a reversão da mudança é sedutora (e merece, claro, confirmação futura, minimamente com a análise das 46 entrevistas do PortVix - até agora Calmon analisou dados de 24 entrevistas). Calmon (2010, p.99-102) considera que tudo indica que estamos diante de algo semelhante ao que aconteceu com os falantes da ilha de Martha’s Vineyard, no estudo pioneiro de Labov (2008), em que a reversão dos ditongos centralizados foi interpretada como um mecanismo de resistência, uma marca de identidade de nativo vinhardense, em função do grande afluxo de turistas e de moradores externos que a ilha recebia.

Assim, segundo Calmon (2010, p.99),

[...] a ilha de Vitória é uma cidade que recebe muitos turistas e, com o advento de grandes empresas, como a Companhia Vale do Rio Doce e a Arcelor Mittal Tubarão, houve um aumento de pessoas vindas de outros estados e países, conforme dados demográficos do IBGE referente ao último Censo de Vitória. De acordo com os dados, é o estado de Minas Gerais que apresenta um maior índice de imigração no estado do Espírito Santo, pois há um contingente de 29.095 imigrantes, o que corresponde a 43,7% do total de imigrantes.

Conforme podemos verificar na Tabela 17, a seguir, adaptada de Calmon (2010):

Tabela 17 - Imigrantes externos por local de nascimento – 2000

MG	BA	RJ	SP	Demais estados	Outros países	Total
29.095	10.460	12.383	4.340	8.820	1.477	66.575
43.7%	15.7%	18.6%	6.5%	13.2%	2.2%	

Fonte: Adaptação de Calmon (2010).

Calmon (2010, p.101) acrescenta:

Segundo informações obtidas por meio do site do Governo do Estado do Espírito Santo, a Setur (Secretaria de Estado de Turismo) e o Sebrae (Serviço Brasileiro de apoio a Micro e Pequenas Empresas), por meio da Futura Pesquisa (<http://www.futuranet.ws/>), realizaram um estudo denominado “Pesquisa de Turismo Receptivo na Região Metropolitana de Vitória”, no ano de 2008, com a intenção de avaliar o fluxo turístico de turistas na alta temporada. (...) Os resultados da pesquisa feita pela Empresa *Futura* também confirmam que o Estado de Minas Gerais é aquele que apresenta um maior contingente de turistas na Região Metropolitana de Vitória.

Evidências adicionais para fortalecer a hipótese de Calmon (2010) vêm de alguns fatos bastante relevantes:

- 1) Dos 1.713 dados levantados por Calmon (2010, p.81) de pronomes *você/cê/ocê* explícitos, há exatamente, e apenas, 17 casos do pronome *ocê*. E mais: 15 destes casos encontram-se na faixa etária de 26-49 anos, a faixa que revelava mudança em direção ao pronome *cê*; os outros dois casos estão nos dados da faixa de 15-25 anos e não há nenhum caso da forma *ocê* na faixa dos mais velhos e dos mais novos, os que mais favorecem o uso de *você*; a amostra de Fala Casual não apresenta nenhum caso do pronome *ocê*;
- 2) Calmon (2010, p.81 e p.120-124) relata que um teste de atitude independente, de Ponzo (apud CALMON, 2010), para captar, entre outros aspectos, a naturalidade

da configuração “p’ce”, revela rejeição da forma *ocê* por falantes capixabas. A consciência de não uso da forma *ocê* pelos capixabas é clara: basta perguntar e a resposta “não uso” é imediata, mas a consciência de maior uso de forma *você* não é tão cristalina – em verdade, na amostra de Fala Casual, Calmon (2010, p.81) não registra uso da forma *ocê*, mas observa aumento da forma *cê*,

- 3) Embora Calmon (2010, p.117) relate que as pesquisas sobre a fala de Minas Gerais evidenciem maior ocorrência da forma *ocê* em áreas rurais, mesmo na capital mineira, em Belo Horizonte, segundo seus relatos, há pesquisas que registram 12% da forma *ocê*.

Em síntese, o trabalho de Calmon (2010) evidencia que os capixabas revelam, consciente ou inconscientemente, distinção linguística em relação à fala dos mineiros. Minimamente podemos dizer que o capixaba não usa a forma *ocê*, uma forma social e geograficamente marcada no PB. Assim, a fala capixaba não se alinha à fala mineira com relação a esse traço. Além do mais, a forma mais frequente nos dados analisados por Calmon é a forma *você* no PortVix (73%) e na Fala casual (53%). Comparando pesquisas com entrevistas tipicamente labovianas, percebemos que os resultados da fala capixaba se alinham mais aos resultados de fala de Brasília (coletada em 1991), avaliada também como uma fala não marcada (CALMON, 2011, p.117). A fala capixaba só se alinha à fala mineira com relação à ausência do pronome *tu*, com base nos dados até agora analisados.

Ausência/presença de artigo diante de antropônimos e de possessivos

Campos Júnior (2011, p.6) investigou a variação morfossintática ausência/presença de artigo definido antes de antropônimos e possessivos em 20 entrevistas do PortVix, “a fim de delimitar, em última instância, a tendência capixaba para esse aspecto morfossintático, estabelecendo-o como (um) traço de identidade linguística dentro do cenário nacional”. Foi analisada a alternância entre a presença e a ausência do artigo definido diante de antropônimos, em construções como “totalmente diferente *de* □ *Isabela*”/“só que *o Edmundo* deu show”, e de possessivos, em enunciados do tipo “*mora* □ *minhas tias... os meus tios...*”. Mesmo buscando as semelhanças entre os dois fenômenos, Campos Júnior (2011) relata que a análise dos dados foi feita em etapas separadas: uma etapa para os antropônimos, com um total de 300 dados e um percentual global de uso do artigo definido de 39%, em que quatro grupos de fatores foram apontados como estatisticamente significativos: *gênero do antropônimo, traços supra-segmentais, faixa etária e gênero do falante*; outra etapa para os possessivos, com um total de 1.016 dados e um percentual global de uso do artigo de 33%, com cinco grupos estatisticamente significativos: *ausência/presença e tipo de preposição, tipo do possessivo e outras formas, natureza do possuído, nível de escolaridade e gênero do falante*.

Campos Júnior (2011, p.6) assim conclui:

[...] Em linhas gerais, os resultados apontam que, em termos de percentuais globais de uso do artigo, os capixabas da cidade de Vitória (ES) usam menos artigo do que outras regiões do país. Esse percentual foi de 39%, no contexto dos antropônimos, e 33% no contexto de pronomes possessivos, o que alinha a capital do Espírito Santo, no contexto dos antropônimos, com o Rio de Janeiro (43%). Por outro lado, no contexto do possessivo, não há comparação razoável, haja vista que Vitória (ES) está consideravelmente afastada das demais capitais brasileiras inventariadas até o momento: as cidades de Recife e Salvador, que apresentam os menores índices, estão na casa dos 60% contra apenas 33% da capital capixaba. Destarte, a tendência à ausência do artigo definido parece configurar-se como uma marca identitária inconsciente na capital capixaba.

Os resultados comparativos apresentados por Campos Júnior (2011, p.74) relativos aos percentuais globais de uso de artigo definido diante de antropônimos para Vitória (VIX), Belo Horizonte (BH), Recife (RE), Salvador (SSA), Rio de Janeiro (RJ), São Paulo (SP) e Porto Alegre (POA) podem ser vistos na Tabela 18.

Tabela 18 - Realização do artigo definido diante de antropônimos de acordo com a origem geográfica

	VIX	BH	RE	SSA	RJ	SP	POA
Freq.	118/300	--	12/71	10/24	27/85	20/23	50/63
(%)	(39)	(76)	(17)	(32)	(43)	(87)	(79)

Fonte: Tabela 7 de Campos Júnior (2011, p.74).

Com relação ao contexto dos possessivos, Campos Júnior (2010, p.75-76) enfatiza que “todas as capitais brasileiras inventariadas até o momento apresentam um padrão geral de maior uso do artigo definido. Na contramão dessa tendência, os capixabas [se individualizam], apresentando uma inclinação [acentuada] à ausência de artigo definido diante de possessivos”, o que pode se ver na Tabela 19 organizada pelo autor, com base em diversas pesquisas brasileiras, nos mesmos moldes feitos para o uso de artigo diante de antropônimos. Em síntese, as ponderações de Campos Júnior (2011) não consideram que os índices de uso de artigo diante de antropônimos e possessivos alinhem parcialmente a capital capixaba às capitais da região Nordeste, mas, sim, que estes índices, tão próximos, respectivamente, 39% e 33%, tipificam a fala capixaba, no sentido de usar, sistematicamente, uma variante perceptual e regionalmente mais marcada: a ausência mais sistemática de artigos nos contextos analisados.

Tabela 19 - Realização do artigo definido diante de possessivos de acordo com a origem geográfica

	VIX	BH	RE	SSA	RJ	SP	POA
Freq.	118/300	--	12/71	10/24	27/85	20/23	50/63
(%)	(39)	(76)	(17)	(32)	(43)	(87)	(79)

Fonte: Campos Jr. (2011, p.76).

Cumpre ressaltar que este comportamento é paralelo ao que foi observado por Calmon (2010), no sentido de a fala capixaba se distinguir da fala mineira, especialmente da fala dos mineiros da capital, que apresentam 76% de uso do artigo diante de antropônimos. Infelizmente, parece não haver estudos disponíveis para o uso de artigo diante de possessivos na fala da capital mineira. Dia virá, para que possamos ver se as conjecturas aqui levantadas têm efetivamente uma contraparte com a realidade linguística.

Considerações finais

Retomando o início de nosso texto, a configuração da variedade capixaba ainda não é bem clara nem para a comunidade acadêmica nem para os leigos nem para os próprios habitantes de Vitória. Na busca de “colocar em cena a variedade capixaba”, faz-se mister afirmar que o PortVix muito tem contribuído para revelar não somente a identidade linguística capixaba, mas, também, o alinhamento dessa variedade às demais do PB.

Conforme relatado, a comunidade capixaba ficou isolada do restante da região Sudeste por razões históricas e esse isolamento pode ter contribuído para a constituição de uma variedade menos marcada. A capital capixaba somente passa a ter um fluxo maior de imigração a partir da instalação de duas grandes indústrias na década de 70. A forte imigração do interior do estado do Espírito Santo, marcadamente influenciada por alemães e italianos, também pode ter contribuído para a formação dessa variedade menos marcada. Assim, os falantes capixabas não percebem claramente suas características linguísticas nem mesmo são reconhecidos por outros falantes brasileiros.

As pesquisas desenvolvidas com base no PortVix, todas, até o momento, baseadas em fenômenos sintáticos ou morfossintáticos, contribuíram para a descrição da variedade linguística capixaba. Quase todos os fenômenos analisados apontaram para o alinhamento a outras variedades do PB. A implementação do futuro do presente e do futuro do pretérito perifrástico, a alternância dos usos do

gerúndio, a variação entre orações adverbiais finais desenvolvidas ou reduzidas, a implementação do pronome de primeira pessoa *a gente*, a alternância do imperativo e a variação entre indicativo e subjuntivo mostram que a variedade linguística capixaba se assemelha às demais variedades do PB. A supressão do clítico acusativo de terceira pessoa e sua substituição por sintagmas nominais anafóricos e categorias vazias e a variação das concordâncias nominal e verbal mostram semelhanças e diferenças. O maior uso do pronome *você* e a quase inexistência do pronome *ocê* por falantes de Vitória parecem ser uma característica da variedade capixaba, revelando ser uma marca identitária, ao se contraporem os resultados aos de outras pesquisas baseadas na fala de Minas Gerais. Do mesmo modo, a ausência de artigos definidos antes de nomes de pessoas e de possessivos também caracteriza a fala capixaba.

O PortVix, dessa forma, cumpre seu papel de descrever a variedade linguística da capital do Espírito Santo e de colocar luzes para uma comunidade desconhecida por brasileiros e estrangeiros. Cumpre, também, seu papel de formar um grupo de pesquisas forte, cujo objetivo é desenvolver estudos de cunho variacionista em Vitória e, agora, em Santa Leopoldina, município do interior do Estado do Espírito Santo. Assim, novas pesquisas se iniciam, seja em nível de dissertações de mestrado, como as que abordam a variação *nós/a gente* e de concordância nominal na fala de Santa Leopoldina (região serrana do Espírito Santo); expressão do sujeito e variação das vogais médias pretônicas na fala de Vitória (PortVix); e, também, em nível de Iniciação Científica, pesquisa que focaliza a expressão do objeto direto anafórico de terceira pessoa no jornal *A Gazeta*.

É importante relembrar que o Estado do Espírito Santo possui uma configuração etnográfica bastante variada, sendo interessante o desenvolvimento de pesquisas que possibilitem o estudo de diferentes variedades linguísticas, bem como as pesquisas que enfatizem o contato com diversas línguas, sejam elas de comunidades indígenas, quilombolas ou pomeranas, além de pesquisas que caracterizem o forte contato com comunidades de descendência italiana e alemã. Paralelamente às amostras já constituídas, é urgente a constituição de outras para que seja possível delimitar a identidade linguística dos falantes do Espírito Santo, em todas as suas fronteiras. Esforços como os aqui já realizados são de extrema importância, mas há necessidade de outros estudos para que se integre o Espírito Santo à descrição das variedades brasileiras já pesquisadas, deixando, então, de se pensar que o Espírito Santo não possua marcas próprias ou que seja uma ilha dentro da região Sudeste. Um fato se destaca na paisagem deslumbrante do mar e das montanhas: a beleza ímpar do Estado de Espírito Santo, aos olhos atentos de uma mineira canela verde, a segunda autora deste texto, que chega ao seu final, abrindo perspectivas para o futuro.

YACOVENCO, L. C. et al. Portvix: a Sociolinguistics Project - the Speech of Vitória (State of Espírito Santo) on the Setting. *Alfa*, São Paulo, v.56, n.3, p.761-796, 2012.

- *ABSTRACT: The main objectives of this paper are (1) to show the academic community the PortVix, a project which deals with the speech of the city of Vitória, capital of Espírito Santo; (2) to summarize the works concluded and in development, based on data from this project, and (3) to speak about the alignment of Vitória's speech with the national setting and about characteristics that individualize it. The Theory of Language Variation and Change guides the organization of the sample and, in general, the analysis of several variable phenomena, namely the variation between verb forms of conditional and of imperfect indicative; variable expression of the future tense; use of the gerund; variation of final adverbial clauses; alternation between the expressions nós/a gente (we); expression of anaphoric direct object; nominal agreement; verbal agreement; grammatical expression of imperative form; alternation of indicative/subjunctive form; variation of second pronouns, specially among você/cê/ocê; presence or absence of definitive article in front anthroponyms and possessive pronouns. Common sense observations and test results evidenced that there are aspects that align the speech of Vitória-ES with the national setting, confirming the generality of some phenomena, but there are other characteristics that allow individualization of the capixaba speech (speech from the state of Espírito Santo) against the three states with which Espírito Santo establishes boundaries: Rio de Janeiro, Bahia, and especially Minas Gerais.*
- *KEYWORDS: PortVix Project. Capixaba Speech (Speech from the state of Espírito Santo). Brazilian Portuguese. Linguistic Variation. Linguistic Change. Geographical variation.*

REFERÊNCIAS

- BASÍLIO, J. O. S. *Usos das estruturas com estar+gerúndio no português brasileiro*. 2011. 121f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2011.
- BARBOSA, A. F. *Alternância de formas indicativas e subjuntivas na fala de Vitória (ES)*. 2011. 216f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2011.
- BENFICA, S. de A.; SCHERRE, M. M. P. *A concordância verbal na fala capixaba: a variável escolaridade*. Comunicação oral apresentada no III Seminário de Sociolinguística, Vitória, Universidade Federal do Espírito Santo, 2012.
- BERBERT, A.T.F; YACOVENCO, L.C.; SCHERRE, M.M.P. *Eu encontrei ele em Vitória à variação do objeto direto anafórico*. Comunicação oral apresentada no III Seminário de Sociolinguística, Vitória, Universidade Federal do Espírito Santo, 2012.
- BRAGANÇA, M. L. L. *A gramaticalização do verbo IR e a variação de formas para expressar o futuro do presente: uma fotografia capixaba*. 2008. 146f. Dissertação

(Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2008.

CALMON, E. N. *Ponte da passagem: você e cê transitando na fala de Vitória (ES)*. 2010. 139f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2010.

CAMPOS JUNIOR, H. da S. *A variação morfossintática do artigo definido na capital capixaba*. 2011. 110f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2011.

DEOCLÉCIO, C. E. *Variação sintática das orações adverbiais finais: similaridades e diferenças entre fala e escrita*. 2011. 122f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2011.

EVANGELISTA, E. M. *Fala, Vitória: o imperativo na cidade de Vitória/ES e sua posição no cenário nacional*. 2010. 172f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2010.

GONÇALVES, F. de C. A colonização no Espírito Santo entre 1535 e 1700 e a contribuição dos diversos segmentos populacionais na formação socieconômico-cultural. *Dimensões*, Vitória, n.11, p.53-61, 2000.

LABOV, W. *Padrões Sociolinguísticos*. São Paulo: Parábola, 2008.

MENDONÇA, A. K. de. *Nós e a gente em Vitória: análise sociolinguística da fala capixaba*. 2010. 135f. Dissertação (Mestrado em Estudos Linguísticos) - Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, 2010.

NASCENTES, A. *O linguajar carioca*. Rio de Janeiro: Organização Simões, 1953.

RODRIGUES, C. V. Atlas linguístico do Espírito Santo: uma pesquisa em andamento. In: ZAIDAN, J.; SOARES, L. E. (Org.). *Letras por Dentro III*, Vitória, p. 103-110, 2006.

SALETTTO, N. Sobre a composição étnica da população capixaba. *Dimensões*, Vitória, n.11, p.99-109, 2000.

SANTOS, P. T. de A. *Só um instante, senhora, que eu vou tá verificando se o livro tá disponível na editora*. 2008. 123f. Dissertação (Mestrado em Linguística) - Instituto de Letras, Universidade de Brasília, Brasília, 2008.

SILVA, J. B. da; SCHERRE, M. M. P. A concordância nominal na fala capixaba: fatores sociais. In: CARDOSO, C. R.; SCHERRE, M. M. P.; SALLES, H.; PACHECO, C. (Org.). *Variação, mudança linguística e educação: contribuições do III Encontro*

do Grupo de Estudos Avançados em Sociolinguística da Universidade de Brasília. São Paulo: Pontes, 2012.

TESCH, L. M. *A expressão do tempo futuro no uso capixaba: variação e gramaticalização*. 2011. 192f. Tese (Doutorado em Linguística) - Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

_____. *A variação no âmbito do irrealis entre as formas de futuro do pretérito e pretérito imperfeito do indicativo na fala capixaba*. 2007. 153f. Dissertação (Mestrado em Linguística) - Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2007.

YACOVENCO, L. C. *Em busca da identidade capixaba*. Comunicação apresentada no Congresso ABRALIN - em Cena Espírito Santo. Universidade Federal do Espírito Santo, 2009.

_____. O projeto “O português falado na cidade de Vitória”: coleta de dados. In: LINS, M. da P.; YACOVENCO, L. C. (Org.). *Caminhos em linguística*. Vitória: Nuples: DLL: Ufes, 2002. p.102-111.

_____; BERBERT, A. T. F. Preenchimento do objeto direto anafórico: um olhar sobre a fala de Vitória/ES. CARDOSO, C. R. et al. (Org.). *Variação, mudança linguística e educação: contribuições do III Encontro do Grupo de Estudos Avançados em Sociolinguística da Universidade de Brasília*. São Paulo: Pontes, 2012.

WEINREICH, U., LABOV, W.; HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola, 2006.

Recebido em abril de 2012.

Aprovado em julho de 2012.

PROJETO DE VARIAÇÃO LINGUÍSTICA NO CENTRO-OESTE (VALCO)

Rachel do Valle DETTONI*
Cíntia da Silva PACHECO**
Carolina Queiroz ANDRADE***
Maria Marta Pereira SCHERRE****

- **RESUMO:** O objetivo deste artigo é apresentar o Projeto Variação Linguística no Centro-Oeste (VALCO) e sintetizar resultados de pesquisa sociolinguística em áreas geográficas distintas da região Centro-Oeste e em temáticas também distintas. Quatro pesquisas foram realizadas com dados da grande Brasília ou Distrito Federal e analisam os pronomes de segunda pessoa *tu/você/cê/ocê*. Nestas pesquisas, apresenta-se a ideia central de que a variedade brasiliense em formação está introduzindo a variante *tu* em seu repertório linguístico, com focalização de um traço marcado no português brasileiro. Duas outras pesquisas foram realizadas com dados da Baixada Cuiabana, Estado do Mato Grosso, e analisam a variação de gênero no plano discursivo e sintagmático. Nestas duas pesquisas, diferentemente, apresentam-se traços linguísticos de natureza semicategorica, em provável situação de extinção, mas evidencia-se que a variação de gênero nos dados analisados é linguisticamente encaixada no sistema da língua portuguesa. Sendo assim, o presente artigo, que se orienta pela Teoria da Variação e da Mudança Linguística, apresenta para a comunidade científica um pouco mais da riqueza da língua portuguesa em terras brasileiras, mais especificamente, no Brasil central tradicional e no Brasil central expandido.
- **PALAVRAS-CHAVE:** Variação linguística. Região Centro-Oeste. Variedade brasiliense. Variedade Cuiabana. Pronomes de segunda pessoa. Concordância de gênero. Focalização dialetal.

Breve histórico do projeto VALCO

A preocupação em se pesquisar a fala de Brasília não é recente (BORTONI-RICARDO, 1985; HANNA, 1986; MALVAR, 1992; DIAS, 1993; CORRÊA, 1998), e essa preocupação cresceu à medida que os brasilienses foram sentindo necessidade de uma identidade linguística frente aos falantes de outras regiões do país. Como essa necessidade de “identificação” é algo natural nos brasilienses, foi natural também

* UnB – Universidade de Brasília. Departamento de Linguística, Português e Línguas Clássicas. Brasília – Distrito Federal – Brasil. 71000-000 - rdetttoni@gmail.com

** UnB – Universidade de Brasília. Departamento de Linguística, Português e Línguas Clássicas. Brasília – Distrito Federal – Brasil. 71000-000 - cintialetras@yahoo.com.br

*** UnB – Universidade de Brasília. Departamento de Linguística, Português e Línguas Clássicas. Brasília – Distrito Federal – Brasil. 71000-000 - carollwith@gmail.com

****UFES – Universidade Federal do Espírito Santo. Centro de Ciências Humanas e Naturais – Departamento de Línguas e Letras. Vitória – Espírito Santo – Brasil. 29075-910 - mscherre@terra.com.br

que algumas pesquisas alocadas na Universidade de Brasília se propusessem a entender esse dialeto nascituro.

Pesquisas mais recentes (ANDRADE, 2004; LUCCA, 2005; DIAS, 2007; ANDRADE, 2010) corroboram o interesse dos pesquisadores de analisar cientificamente o que pode ocorrer linguisticamente em uma cidade planejada, com um crescimento diferenciado, como é o caso de Brasília.

Portanto, pensando em um projeto que pudesse identificar, documentar e caracterizar uma variedade linguística própria do Distrito Federal, dentro do cenário linguístico nacional, no segundo semestre de 2008, reuniu-se um grupo de professores e estudantes de pós-graduação da Universidade de Brasília ligados à área de pesquisa da Sociolinguística Variacionista para discutir a implementação de um projeto de registro de fala. Dessa reunião, surgiu a ideia de agregar, aos corpora do Distrito Federal, corpora de outras cidades da região Centro-Oeste. Desse modo, a exemplo do que fizeram pesquisadores paraibanos, com o VALPB, e sulistas, com o VARSUL, esses pesquisadores, sob coordenação da professora Rachel Dettoni (UnB), resolveram organizar um material de fala já coletado por pesquisadores envolvidos em projetos de pesquisa das universidades de Brasília, Goiás, Mato Grosso e Mato Grosso do Sul para compor um banco de dados on-line. Assim, surgiu oficialmente, em julho de 2008, o Projeto Variação Linguística no Centro-Oeste (VALCO).

O primeiro objetivo do VALCO é reunir e socializar o conhecimento e a pesquisa linguística que vem sendo desenvolvida no âmbito da variação linguística, disponibilizando material de fala coletado por meio de entrevistas labovianas típicas para todos os profissionais que atuem na pesquisa e no ensino de língua portuguesa, em nível fundamental, médio ou superior, uma vez que, em geral, esse conhecimento permanece restrito ao ambiente acadêmico. Um banco de dados on-line possibilitará o desenvolvimento de novas pesquisas sobre a diversidade linguística do Centro-Oeste, propiciando a ampliação do conhecimento acerca da identidade linguística e sociocultural dessa região do Brasil.

A formulação desse projeto contou com a colaboração fundamental da professora Marta Scherre (UnB/UFES), como uma figura que semeou boa parte do escopo do Valco (deixemos claro, no entanto, que o Valco é um projeto aberto a novas pesquisas de outras áreas da linguística), atuando como orientadora de todas as pesquisas descritas neste artigo e outras que aqui não entraram, sendo ao todo oito dissertações de mestrado e uma tese de doutorado com dados coletados na região. Scherre continua orientando alguns doutorandos alocados na UnB, sendo dois deles estudos de dados de fala do Centro-Oeste. Além de Scherre, o Valco também conta com a coordenadora e idealizadora Rachel Dettoni (UnB), que entendeu a necessidade de alocar em um ambiente todo esse nicho de pesquisa linguística. Compõem ainda o grupo Valco as

pesquisadoras-colaboradoras, doutorandas da UnB, Carolina Andrade e Cíntia Pacheco. O intuito maior é ainda buscar colaboradores de outras universidades federais e/ou estaduais da região Centro-Oeste.

Os dados coletados estão disponíveis on-line em arquivos de áudio e de transcrição, que poderão ser acessados mediante autorização prévia dos administradores do site apenas para audição e visualização on-line, não sendo permitido o download de quaisquer conteúdos sem prévia autorização. Os resultados de pesquisas realizadas com dados provenientes do projeto VALCO devem dar os devidos créditos ao Projeto e podem ser divulgados no site.

Esse siteé hospedado pelo servidor da Universidade de Brasília, e o Projeto VALCO é financiado com recursos da Fundação de Apoio à Pesquisa do Governo do Distrito Federal (FAPDF).

Resumo do VALCO

O Projeto VALCO pertence à área de Linguagem e Sociedade do Programa de Pós-graduação em Linguística da Universidade de Brasília, e sua linha de pesquisa é língua, interação sociocultural e letramento.

Esse projeto visa à organização de um banco de dados de língua falada da região Centro-Oeste a ser disponibilizado on-line para promover o estudo e a pesquisa das variedades linguísticas da região. A relevância desse projeto está apoiada nos seguintes aspectos:

- a) Necessidade de se identificar, documentar e caracterizar uma variedade linguística própria do Distrito Federal, dentro do cenário linguístico nacional;
- b) Necessidade de investigação científica de uma identidade linguística regional, principalmente no que diz respeito ao Distrito Federal;
- c) Urgência em estabelecer, na região Centro-Oeste, um alinhamento com uma tendência nacional da pesquisa linguística, no sentido de organizar bancos de dados representativos para documentar e estudar as manifestações linguísticas regionais, contribuindo para a construção de um perfil mais completo da diversidade do português falado no Brasil, à luz do que já vem sendo desenvolvido por outras universidades federais em projetos como, só para citar alguns, o VARSUL, na região Sul; VALPB, na Paraíba; PEUL, no Rio de Janeiro; PHPB (Projeto para a História do Português Brasileiro), que envolve várias universidades federais e estaduais (UFMG, UFRJ, UFBA, UNICAMP, USP, UFPB, UFPE, UEFS);
- d) Desenvolvimento de novas tecnologias visando à divulgação da pesquisa científica na área de sociolinguística, seus resultados e contribuições para o ensino em geral, promovendo a articulação entre a universidade e demais profissionais da área

de educação, especialmente os docentes de escola pública do Distrito Federal, considerando a inclusão dos estudos sobre variação linguística no ensino de língua portuguesa para os níveis fundamental e médio, prevista pelos Parâmetros Curriculares Nacionais (PCN) definidos pelo Ministério da Educação.

Para desenvolver o referido projeto, já dispomos de algum material coletado, em pesquisas individuais, por alunos de graduação e pós-graduação do curso de Letras (Linguística e Língua Portuguesa) da Universidade de Brasília, que forma professores de língua portuguesa para atuarem nas escolas do Distrito Federal, e também por alunos e pesquisadores de outras universidades da região Centro-Oeste – UFG, UEG, UFMT e UFMS.

A organização dos dados se dá por meio da criação de um site hospedado pelo servidor da Universidade de Brasília. As entrevistas que compõem a amostra são digitalizadas e devem estar transcritas pelo próprio pesquisador interessado, segundo as normas de transcrição do VALCO (Anexo I).

Todo acesso ao banco de dados só será efetivado mediante autorização prévia da equipe de coordenação do projeto, composta inicialmente por professores e alunos de pós-graduação do Departamento de Linguística, Português e Línguas Clássicas da UnB.

Pretende-se que a manutenção e a constante alimentação do site sejam efetuadas por monitores ligados ao Departamento de Linguística da UnB. Por razões éticas, e de acordo com o que preconiza a ética das pesquisas científicas, a identidade dos informantes das entrevistas será preservada.

Além do banco de dados de fala da região, pretende-se inserir no site um espaço para a divulgação de artigos, teses, monografias e trabalhos advindos das pesquisas realizadas com o referido material.

Principais propostas e objetivos do VALCO

A proposta desse projeto parte do princípio de que há, no Distrito Federal, uma variedade linguística regional em processo de constituição e definição de seus traços típicos em tempo real, caso curioso e de extrema relevância para a pesquisa linguística tanto quanto para a formação dos profissionais de Letras que atuam na educação.

A documentação sobre a linguagem falada no Distrito Federal tem dado seus primeiros passos, restritos ainda ao meio acadêmico, necessitando de mais incentivo para que se possa configurar uma amostra mais representativa que nos permitirá caracterizar, com mais segurança e precisão, os traços linguísticos que estão se consolidando como formas típicas desta região e construindo uma identidade local.

Nesse sentido, partimos da hipótese de que há uma manifestação linguística em processo de definição em Brasília e em todo o Distrito Federal que ainda não foi suficientemente documentada e estudada de modo que possamos definir suas características e contornos locais.

Além disso, diferentemente do que vem sendo realizado em outros estados e regiões do Brasil, a região Centro-Oeste não dispõe ainda de nenhum grupo de pesquisa sociolinguística consolidado, apesar de muitas pesquisas isoladas já terem sido desenvolvidas ou estarem sendo feitas neste momento, nesta área, por doutorandos, mestrandos e alunos de graduação.

O material que tem sido coletado por esses pesquisadores e professores em formação representa uma amostra já bastante significativa do que é possível construir e documentar em termos de pesquisa e construção do conhecimento científico sobre a língua portuguesa falada no Brasil e suas diversas manifestações regionais.

Assim sendo, os objetivos principais do VALCO são:

- a) Ampliar a amostra de língua falada no Distrito Federal e entorno, o que permitirá mapear a diversidade linguística própria dessa região, que congrega, em seu território, um mosaico de manifestações e falares de diversas regiões do Brasil;
- b) Realizar o estudo e a sistematização dos traços linguísticos que caracterizam a(s) variedade(s) linguística(s) do Distrito Federal, trabalho ainda não realizado de modo sistemático, que terá também repercuções em nível nacional na medida em que fornecerá novos elementos para a história e o conhecimento do português do Brasil;
- c) Organizar e implementar um banco de dados linguísticos de língua falada no Distrito Federal e demais estados da região Centro-Oeste, cujo grande impacto consiste em viabilizar, para o maior número possível de profissionais (professores e pesquisadores) da área da educação, o acesso tanto aos dados linguísticos e amostras de fala de toda essa região quanto aos resultados das pesquisas e estudos (teses, dissertações e artigos) que forem sendo desenvolvidos pelos pesquisadores interessados na área;
- d) Promover e incentivar o estudo, o conhecimento e a caracterização, por meio de pesquisa científica, das variedades linguísticas do Distrito Federal e demais estados do Centro-Oeste;
- e) Capitalizar e socializar os resultados das pesquisas sociolinguísticas que vêm sendo desenvolvidas não só pelo Programa de Pós-Graduação em Linguística (PPGL) e pelo Programa de Iniciação Científica de alunos graduandos da UnB mas também pelos Programas de Pós-Graduação das outras universidades federais da região Centro-Oeste;

- f) Promover a utilização de novas tecnologias a serviço da socialização do conhecimento científico e da articulação entre a UnB e escolas de nível fundamental e médio do DF;
- g) Oferecer suporte teórico-pedagógico para professores e pesquisadores que atuam no ensino da língua portuguesa, em nível fundamental (6º a 9º ano) e médio, divulgando e incentivando o pensamento científico e o estudo da diversidade linguística nos outros níveis de ensino, além do superior.

Busca-se, então, legitimar uma identidade cultural por meio do patrimônio linguístico, caracterizando a identidade sociolinguística e cultural do Distrito Federal e demais regiões, por meio do estudo sistemático dos traços linguísticos que definem as variedades regionais locais. Assim, é importante também oportunizar a formação de crianças e jovens e a produção científica de novos pesquisadores, incentivando o estabelecimento de uma cidadania participativa por meio de reconhecimento das diversidades linguísticas e consequente respeito a elas.

Metodologia do projeto e das pesquisas

O projeto comporta, em princípio, duas ações básicas a serem desenvolvidas. A primeira diz respeito à criação e organização do banco de dados, o que se dará a partir da reunião seletiva das amostras de fala regional já coletadas por pesquisadores e professores em formação na UnB. Essa ação prevê que, numa segunda etapa, sejam contempladas também as amostras coletadas e organizadas por pesquisadores das outras universidades do Centro-Oeste, a saber: UFG, UEG, UFMT e UFMS.

De posse desse material já coletado, proceder-se-á à organização do banco de dados on-line, o que permitirá o acesso de pesquisadores e professores dos vários estados.

A segunda ação refere-se à constante alimentação do site e do banco de dados com as novas coletas que forem sendo realizadas a partir dos trabalhos de novos alunos e pesquisadores interessados, além da divulgação dos trabalhos de pesquisa, como dissertações, teses, artigos e trabalhos de iniciação científica de graduandos que forem sendo desenvolvidos. A metodologia específica da pesquisa e coleta de dados será orientada pelos pressupostos teórico-metodológicos da sociolinguística variacionista de base laboviana, que prevê a coleta de dados de língua falada em contextos reais de comunicação e interação.

Em suma, os objetivos são incentivar e alimentar constantemente o espírito e a prática de pesquisa científica sobre a linguagem, ampliar o banco de dados, alimentando-o sempre com novo material coletado, divulgar resultados de

pesquisas e registros científicos sobre a linguagem do Centro-Oeste e Distrito Federal que forem sendo desenvolvidos e tantos outros. Para isso, o VALCO já tem o material sobre a Baixada Cuiabana disponível para a pesquisa. Posteriormente, serão introduzidas as entrevistas do Distrito Federal.

Para acessar o VALCO, os pesquisadores interessados deverão enviar uma mensagem para o e-mail (valco.unb@gmail.com) do projeto e solicitar permissão para o acesso às transcrições e aos áudios. Na parte das transcrições, é possível fazer uma busca pelo fenômeno linguístico desejado ou por qualquer outro aspecto que interesse ao pesquisador.

Resultados das primeiras pesquisas variacionistas que constam (constarão) no projeto VALCO

Os pronomes de segunda pessoa no Distrito Federal

Dentro do escopo do Projeto VALCO, temos as pesquisas de cunho variacionista sobre os pronomes de segunda pessoa do singular na fala de Brasília. Até o momento foram realizadas quatro dissertações sobre o tema: Andrade (2004); Lucca (2005); Dias (2007) e Andrade (2010). Alguns dos principais resultados das pesquisas serão apresentados a seguir.

Andrade (2004) realizou uma pesquisa com as variantes você, ocêecê em Brasília, com uma amostra de fala da faixa etária entre 10 e 14 anos, incluindo também os dados da pesquisadora, com, na época, 25 anos. A amostra foi constituída em 1991, feita a partir de gravações labovianas típicas, para dissertação de mestrado de Malvar (1992), a organizadora da amostra. A localidade específica de Brasília em que a amostra foi composta foi a RA Sobradinho, fundada em 1960. A amostra é composta por falantes residentes em áreas urbanas e rurais da mesma RA.

Os resultados que a autora alcançou da variação dos pronomes estão descritos na tabela a seguir:

Tabela 1 - Distribuição dos pronomes entre os falantes de 10 a 14 anos de Sobradinho em 1991

Área	TU	VOCÊ	CÊ	OCÊ	TOTAL
Área urbana	0%	63%	31%	6%	90
Área rural	0%	44%	50%	6%	32

Fonte: Andrade (2004).

A ocorrência da variante *cê*, de forma geral, foi bastante significativa, embora tenha sido menor que a ocorrência da variante *você*. Há que se considerar que se tratou de uma coleta de dados tradicional, do gênero discursivo entrevista, em que os informantes sabiam previamente da gravação, fato que influencia a fala a ficar mais monitorada.

Outro resultado interessante foi relacionado ao sexo¹, em que as pessoas do sexo masculino favoreceram o uso do *cê*, com o peso de 0,61, enquanto as pessoas do sexo feminino desfavoreceram esse uso, com peso de 0,40 (ANDRADE, 2004).

Andrade (2004) também analisa a função sintática, e um resultado relevante é a verificação da fala da entrevistadora (com 25 anos na época da gravação). Vejamos a tabela a seguir:

Tabela 2 - Função sintática da variação *você* / *cê* / *ocê* em Brasília (dados da entrevistadora)

	VOCÊ	CÊ	OCÊ
Sujeito	602/ 996 = 60%	226/ 996 = 23%	15/ 996 = 2%
Objeto de verbo	8/ 8 = 100%	0/ 8 = 0%	0/ 8 = 0%
Objeto de preposição	38/ 38 = 100%	0/ 38 = 0%	0/ 38 = 0%
Dupla função	5/ 6 = 83%	0/ 6 = 0%	1/ 6 = 17%
Total	674/ 1069 = 63%	226/ 1069 = 21%	16/ 1069 = 1%

Fonte: Andrade (2004, p.50) – Com adaptações e retirando os casos de: zeros, 153 casos, que correspondem a 15%, e demais casos, que somam 21 ocorrências.

Os dados da tabela anteriormente colocada sugerem que, quando o pronome serve à dupla função (sujeito e objeto) ele até pode ser *ocê*, mas categoricamente não poderá ser *cê*, pois, nos dados de Andrade (2004), não há casos dessa forma como objeto. O *ocê* parece não ter restrição sintática, apesar de essa forma ser escassa nas amostras analisadas. Já o *cê* é sintaticamente restrito à função de sujeito. A função de sujeito, por seu turno, é a mais produtiva para todas as variantes em estudo (uma vez que forma a ordem natural do português em SVO), sendo, portanto, essa a função irrestrita para todas as variantes.

Por fim é importante ressaltar que Andrade (2004) não encontra em sua amostra da fala de Brasília a ocorrência da forma *tu*. Lucca (2005) realizou pesquisa sobre a variação dos pronomes *tu* e *você* (em que os casos decê foram amalgamados ao *você*) na fala de jovens brasilienses em três regiões administrativas diferentes do Distrito Federal, a saber: Brasília, Taguatinga e Ceilândia. A autora revela que

¹ Nas pesquisas da sociolinguística variação, tem-se utilizado a codificação relacionada a sexo biológico, e Labov (2001, 2008) considera que deve ser assim senão perdemos a comparabilidade dos resultados, mas a interpretação dos dados tem sido de gênero, uma noção social e/ou cultural.

sua coleta de dados ocorreu em circunstâncias prototípicas para o uso da variante *tu*, uma vez que a maioria dos dados foi coletada em gravações de conversas espontâneas e ocultas, entre jovens rapazes que eram amigos. Essa pesquisa não teve a idade como fator social de investigação, uma vez que a pesquisadora considerou as pequenas diferenças de idade entre os informantes como indicador de que tais informantes pertenciam à mesma faixa etária, entre 15 a 19 anos. A seguir, apresentamos resultados gerais da pesquisa:

Tabela 3 - Distribuição dos pronomes entre os falantes de 15 a 19 anos na Grande Brasília em 2005

Região	TU	VOCÊ	CÊ	TOTAL
Ceilândia	87%	8%	5%	121
Taguatinga	66%	25%	9%	244
Brasília	68%	5%	28%	87
Total	72%	17%	11%	452

Fonte:Lucca (2005).

Os fatores analisados e selecionados pelo programa foram, por um lado, os linguísticos: 1) paralelismo formal (tendência de formas semelhantes se agruparem) e 2) tipo de estrutura (se afirmativa, interrogativa ou exclamativa); por outro lado, os extralingüísticos: 1) gênero do falante, 2) tipo de relação entre os pares, 3) região administrativa e 4) familiaridade com o tema.

O fator gênero do falante, na pesquisa de Lucca (2005), foi o primeiro a ser selecionado pelo programa estatístico, e os resultados apontaram para o peso relativo da variante *tu* na ordem de 0,55 para o gênero masculino, enquanto o gênero feminino ficou na ordem de 0,09. A diferença entre os pesos revelou que o gênero masculino favorece de forma indubitável o uso do *tu*, ao passo que o gênero feminino o desfavorece (é importante salientar que cerca de 90% dos dados de Lucca são de falas masculinas).

Em função da região administrativa, a cidade satélite de Ceilândia foi a que mais favoreceu o uso do *tu*, com o peso de 0,68, enquanto Taguatinga e Brasília ficaram, cada uma, com 0,43 de peso relativo. Segundo a pesquisadora, a variante *tu* teria aparecido no repertório brasiliense em função do grande contingente de imigrantes nordestinos que hoje vive no Distrito Federal. O fato de Ceilândia ter se destacado com mais ocorrências da variante *tu* corrobora essa hipótese, já que, entre as três regiões administrativas estudadas, Ceilândia conta com o maior contingente de imigrantes nordestinos.

A autora conclui que o *tu* é mais recorrente nas seguintes situações conversacionais: “[...] falantes jovens do gênero masculino que interagem com

pares solidários tratando de temas cotidianos" (LUCCA, 2005, p.113) e que essa variação ocorre principalmente segundo fatores sociais.

Dias (2007) estudou a mesma variação no Distrito Federal, com dados apenas da região administrativa de Brasília, porém com dados de ambos os sexos equilibradamente (51,6% dos dados de falantes do sexo feminino e 48,4% dos dados de falantes do sexo masculino) e em três faixas etárias diferentes, a saber: 13 a 19 anos (19% dos dados); 20 a 29 anos (47,1% dos dados) e mais de 30 anos (33,9% dos dados). A seguir, os resultados gerais da pesquisa:

Tabela 4 - Distribuição dos pronomes por faixa etária no Plano Piloto ampliado, sem Vila Planalto, em 2007

Faixa Etária	TU	VOCÊ	CÊ	TOTAL
13-19 anos	30%	15%	55%	171
20-29 anos	13%	22%	65%	424
30-48 anos	4%	36%	60%	305

Fonte: Dias (2007).

Os fatores analisados pela autora (DIAS, 2007, p.64-94) e selecionados pelo programa foram, por um lado, o único linguístico selecionado: 1) tipo de fala (*frames* ou *enquadres*) - (a) conversa casual, com peso de 0,56; (b) conversa profissional ou acadêmica, com peso de 0,17; (c) repreensões, com o peso de 0,34; e (d) observações irônicas, deboches e brincadeiras, com peso de 0,80. Por outro lado, os fatores extralingüísticos (ou sociais): 1) sexo; 2) faixa etária; 3) estilo do falante; 4) tipo de relacionamento com o interlocutor e 5) faixa etária do interlocutor. Os resultados alcançados pela autora em função da faixa etária estão na tabela a seguir:

Tabela 5 - A variação tu/ você no Distrito Federal em função da idade

Faixa Etária	Frequência de TU	Peso relativo
13 a 19 anos	51/ 171 = 29,8%	0,76
20 a 29 anos	53/ 424 = 12,5%	0,56
Mais de 30 anos	11/ 305 = 3,6%	0,28
Totais	115/ 900 = 12,8%	-

Fonte: Dias (2007, p.70) - com adaptações.

Os pesos indicaram o favorecimento do *tu* por parte da faixa etária de 13 a 19 anos. Outra observação da autora é que os informantes com mais de trinta anos

tenderam a usar a variante *tu* em situações mais específicas, enquanto as outras faixas etárias tenderam a usar essa variante em diversos tipos de fala. Nas palavras de Dias (2007, p.74), “[...] as frequências de uso de *tu* nas brincadeiras/ironias cresce à medida que elevamos a faixa etária, e o inverso acontece com outros tipos de fala”; e, ainda, “[...] o *tu* está passando de um uso altamente específico para uso de contextos variados”.

Quanto ao fator sexo, a autora chegou aos seguintes resultados:

Tabela 6 - A variação tu/ você no Distrito Federal em função do sexo

Sexo	Frequência de <i>TU</i>	Peso relativo
Feminino	$50/ 464 = 10,8\%$	0,41
Masculino	$65/ 436 = 14,9\%$	0,60
Totais	$115/ 900 = 12,8\%$	-

Fonte: Dias (2007, p.75) – com adaptações.

Ao analisar ambos os fatores – sexo e faixa etária –, a autora chegou aos seguintes percentuais em função da variante *tu*:

Tabela 7 - A variação tu/ você no Distrito Federal em função do sexo e idade

Sexo	13 – 19 anos	20 – 29 anos	Mais de 30	Total
Feminino	$24/ 106 = 22,6\%$	$26/ 269 = 9,7\%$	$0/ 89 = 0\%$	$50/ 464 = 10,8\%$
Masculino	$27/ 65 = 41,5\%$	$27/ 155 = 17,4\%$	$11/ 216 = 5,1\%$	$65/ 436 = 14,9\%$
Totais	$51/ 171 = 29,8\%$	$53/ 424 = 12,5\%$	$11/ 305 = 3,6\%$	$115/ 900 = 12,8\%$

Fonte: Dias (2007, p.76) – com adaptações.

Dias (2007) analisa as demais variáveis selecionadas e não selecionadas. A autora confirmou suas hipóteses de que, quanto mais novo o falante, mais *tu* usaria e também a tendência de *tu* ocorrer mais na fala de pessoas do sexo masculino. Lucca (2005), quando optou por estudar o grupo prototípico dessa variante, demonstrou que partiu exatamente dessa hipótese. Podemos perceber não os mesmos resultados numéricos, mas, de forma geral, as mesmas tendências nos dados de Dias (2007) e Lucca (2005), em que a maior tendência de ocorrência de *tu* é na fala de falantes do sexo masculino na faixa etária entre 13 e 19 anos, em relações solidárias ou íntimas. Na comparação, porém, os resultados de Dias lançam a semente do possível aumento da ocorrência de *tu* na fala das falantes do sexo feminino.

Andrade (2010) realizou sua pesquisa baseada nas três primeiras pesquisas, sobre a variação pronominal entre *cê/tu/você*, na região administrativa nominada Vila Planalto (com apenas 30% dos dados de fora da Vila, no Plano Piloto).

A coleta de dados de Andrade (2010) se deu por gravações de conversas tomadas como espontâneas. Porém, como os falantes tinham conhecimento da gravação, esse fato pode ter enviesado a naturalidade nos dados. Os falantes gravados foram contatados por meio de uma escola pública da Vila Planalto e compreendem as idades entre 07 a 15 anos. Andrade (2010) encontrou fatos linguísticos interessantes em função das origens de pais e mães dos falantes, pois percebeu que o uso dos pronomes pelas crianças estudadas parecia corresponder, relativamente, ao uso dos pronomes nas localidades de origens de seus pais. Filhos de pais mineiros, por exemplo, apresentavam maior tendência de uso da forma *cê*, enquanto filhos de pais nordestinos, maior tendência de uso da forma *tu*. Apesar de Andrade (2010) ter dividido as idades em duas faixas etárias – de 7 a 11 anos, as crianças; e de 12 a 15 anos, os adolescentes –, os resultados demonstraram que tanto as crianças como os adolescentes usaram os três pronomes com a mesma frequência e esse fato é relevante para o entendimento em relação à focalização dialetal, por um lado, e em relação à aquisição linguística, por outro. A seguir, apresentaremos resultados em função do sexo:

Tabela 8 - Efeito do fator sexo em percentuais e pesos (dados da Vila Planalto)

Sexo	VOCE	CÊ	TU
Feminino	$113/330 = 34\%$ Peso = 0,44	$72/330 = 22\%$ Peso = 0,30	$145/330 = 44\%$ Peso = 0,25
Masculino	$36/244 = 15\%$ Peso = 0,24	$80/244 = 33\%$ Peso = 0,34	$128/244 = 52\%$ Peso = 0,41

Fonte: Andrade (2010, p.100).

Os resultados obtidos em função da variável sexo em rodada eneária com três variantes (peso 0,33 como referência) indicaram que ambos os sexos tendem a usar a variante *cê* de forma neutra, com pesos de 0,30 e 0,34. Já os comportamentos das variantes *vocé* e *tu* ficaram inversamente favorecidos pelos sexos, por um lado, o sexo feminino favorecendo o *vocé*, com o peso de 0,44, e, por outro, o sexo masculino favorecendo o *tu*, com o peso de 0,41.

Os resultados alcançados pela autora também indicaram que, em comparação com os resultados de 2005, vêm diminuindo as diferenças entre os sexos masculino e feminino quanto ao uso da variante *tu*. Mas esse resultado também pode ter ocorrido devido às diferenças quanto à coleta de dados. Embora os resultados de 2010 tenham apresentado os percentuais de 44% de uso de *tu* pelo sexo feminino, ficando bem acima dos percentuais alcançados por Dias em 2007, os pesos ficaram desfavorecidos por esse fator (sexo feminino), de forma semelhante aos resultados

de 2007, com pesos um pouco abaixo do peso de referência. Assim, a diminuição entre as diferenças de uso do *tu* por ambos os sexos não foi revelada na amostra de 2010 pelos resultados em pesos.

Andrade (2010), como apontado anteriormente, analisou diversos fatores que influenciam a presente variação. Entre eles, Andrade (2010) chama a atenção para a importância da origem dos pais em relação à variante mais usada pelos jovens falantes, fato apontado na análise de Lucca (2005) e posteriormente comprovado e aprofundado na análise realizada em 2010.

Apresentaremos, a seguir, um quadro com os resultados gerais da variação *tu/você/cê* no Distrito Federal (DF), região Centro-Oeste:

Tabela 9 - A variação *tu/você/cê* no DF, região Centro-Oeste

	VOCE	CE	TU
Adriana (2004)²	$108/ 189 = 57\%$	$48/ 189 = 25\%$	Não há registro
Lucca (2005)	$125/ 453 = 23\%$		$327/453 = 72\%$
Dias (2007)	$785/ 900 = 87,2\%$		$115/ 900 = 12,8\%$
Andrade (2010)	$318/835 = 38,1\%$	$229/835 = 27,4\%$	$288/835 = 34,5\%$

Fonte: Andrade (2004, p.45); Lucca (2005, p.78); Dias (2007, p.64) e Andrade (2010, p.79) - com adaptações - Retirados os “nulos e ocês”, de Andrade (2004), e amalgamados o “você” e o “cê” nos resultados de Lucca e Dias (2005; 2007).

Com base nas quatro pesquisas, chegamos a diversas conclusões, entre as quais destacamos as seguintes:

- 1) O *tu* é uma variante inovadora no Distrito Federal, configurando, entre outros aspectos, que em Brasília, além de haver um movimento em direção à focalização dialetal, houve também uma mudança no paradigma pronominal. O *cê*, como uma variante instigante e sintaticamente restrita, necessita ser mais analisado em termos sociais e linguísticos, contudo os estudos de 2010 apontaram que este possui um papel de esquiva. O pronome *você* parece ser a variante *default*, o que significa que ela é mais neutra, menos marcada, ou que possui mais *status* de prestígio (porém essa interpretação depende de um referencial interacional).
- 2) Muitos fatores sociais são instigantes quanto ao estudo da presente variação, destacando-se entre eles: sexo, faixa etária, origem dos pais (junto à localização dentro do Distrito Federal), comunidade de prática e estilo do falante. Tais fatores necessitam de mais análises para a descrição mais acurada da presente variação.

² É importante lembrar que a pesquisa foi realizada em 2004, mas os dados foram gravados em 1991. 2

- 3) O estudo dos pronomes nos permite ter uma visão linguística e extralingüística do comportamento das variações em línguas naturais.

O site do projeto Valco, além de dar acesso às dissertações (aqui comentadas) na íntegra, em PDF, também dá acesso aos arquivos em áudio, das gravações realizadas e suas respectivas transcrições.

Vale lembrar que os arquivos em áudio e suas transcrições dão acesso à fala brasiliense em diferentes anos compreendidos entre 1991 e 2009. Os dados analisados foram em função da variação dos pronomes, mas trazem luz aos tipos de metodologias utilizadas para sua coleta. Assim, os próximos parágrafos são uma compilação do que os áudios e as transcrições trazem para o pesquisador que deseje acessá-las.

Andrade (2004) examinou o *corpus* de língua falada recolhido em parte das entrevistas constantes da dissertação de mestrado de Malvar (1992), em Sobradinho. As entrevistas foram realizadas no ano de 1991, em que foram analisados dados de quinze falantes, nas faixas etárias entre 10 e 14 anos, além da fala da própria entrevistadora, que à época das entrevistas estava com 25 anos de idade. Com relação ao sexo, oito falantes de 10 a 14 anos eram meninos e sete, meninas. A entrevistadora é natural do Rio de Janeiro, mas vivia em Brasília há cerca de 19 anos. Ao todo, foram levantados 189 dados.

Em Lucca (2005), os dados coletados compreendiam falantes da mesma faixa etária, entre as idades de 15 a 19 anos, sendo cerca de 90% dos dados de fala de meninos, residentes em três RAs distintas: Plano Piloto, Taguatinga e Ceilândia. Ao final da coleta dos dados, Lucca obteve cerca de 55 minutos de gravação com falantes de Ceilândia, 82 minutos com falantes de Taguatinga e 45 minutos com falantes da região administrativa de Brasília, somando 182 minutos de gravações e resultando em 453 dados.

Em Dias (2007), foram analisadas aproximadamente 23 horas e 30 minutos de conversa, que totalizaram 900 dados, sendo todos os falantes residentes no Plano Piloto. As faixas etárias estudadas foram: de 13 a 19 anos; de 20 a 29 anos; e mais de 30 anos. Foram escolhidos três falantes de cada sexo para cada uma das faixas etárias, resultando num total de 18 falantes.

Andrade (2010) realizou 14 transcrições diferentes (sendo 16 gravações ao total, mas duas delas sendo continuações), o que resultou em um total de 11 horas e 24 minutos de gravação. Os resultados globais contam com 953 dados, sendo cerca de 70% dos dados de pessoas moradoras da Vila Planalto e 30% dos dados de moradores de outras localidades, sejam elas: Jardim Botânico no Lago Sul (3 meninas e 1 menino); Asa Norte (2 meninos); Sudoeste (2 meninas e 1 menino). Excetuando a Vila Planalto, que é socialmente híbrida, as demais

regiões administrativas não são muito diferentes entre si em termos de classe social. Há na amostra estudada 24 meninas e 18 meninos que compõem a faixa etária escolhida, entre 7 e 15 anos.

As pesquisas sobre a concordância de gênero na Baixada Cuiabana

A Baixada Cuiabana é a região do Estado de Mato Grosso compreendida entre o Pantanal mato-grossense (ao Sul) e a “portentosa” Chapada dos Guimarães (ao Nordeste), constituída pelos municípios que margeiam os rios Cuiabá e Paraguai. O principal polo de desenvolvimento da baixada é a cidade de Cuiabá, atual capital de Mato Grosso.

As 30 entrevistas do Mato Grosso, mais especificamente da Baixada Cuiabana, foram realizadas por Rachel do Valle Dettoni, entre julho de 2001 e agosto de 2002, para sua tese de doutorado: “A concordância de gênero na anáfora pronominal: variação e mudança linguística no dialeto da Baixada Cuiabana – Mato Grosso” (2003). Anos depois, Cíntia da Silva Pacheco utiliza parte dessas entrevistas para seu trabalho de mestrado: “Padrões sociolinguísticos da concordância de gênero na Baixada Cuiabana” (2010).

O fenômeno linguístico analisado por Dettoni (2003) é a concordância de gênero em configurações discursivas que se destacam por apresentarem nomes explicitamente femininos retomados por pronomes na forma masculina: **Uma dor segura todinho meu corpo, beleza. Quando ele vai aliviano eu tô podre, criatura.**

Pacheco (2010) analisa a concordância de gênero nas relações internas dos sintagmas nominais (**noite inteiro, aquele argola bonito, cara feio**) e na relação entre sujeito e predicado ou nos sintagmas predicativos (**minha mãe eravivo, esta mão tavabobo**).

Ambas as pesquisas são de linha variacionista, baseadas na Teoria da Variação e da Mudança Linguística (LABOV, 1972, 2008; WEINREICH; LABOV; HERZOG, 1968). A análise quantitativa dos dados foi feita pelo pacote de programas VARBRUL (PINTZUK, 1988) na pesquisa de Dettoni (2003) e pelo GOLDVARB-X (SANKOFF; TAGLIAMONTE; SMITH, 2005) na pesquisa de Pacheco (2010).

Sobre o trabalho de Dettoni (2003), a autora afirma que não há possibilidade de gênero neutro, porque o falar cuiabano não apresenta itens lexicais e funções sintáticas específicas para um terceiro gênero (o neutro). Outra possibilidade descartada é a de que a marcação de gênero seja um fato gramatical ausente no dialeto da baixada cuiabana, porque se trata, na verdade, de uma forma variável de marcar o gênero.

A anáfora, por ser uma relação de referência que o falante faz por meio das expressões linguísticas, deve ser examinada como um fenômeno que se situa na interface entre a sintaxe, a semântica e a pragmática.

As intenções do falante, que estão associadas à representação mental que ele faz do antecedente, determinam muito mais a escolha da forma pronominal presente em uma relação anafórica do que, estritamente, as características gramaticais do antecedente. E isto tem a ver com questões de ordem pragmática.

Das 18 entrevistas codificadas por Dettoni (2003), obteve-se um total de 829 dados, sendo 233 do pronome *ele* retomando antecedentes femininos, ou seja, 28% dos dados. Exemplificando o fenômeno analisado, temos, respectivamente, a presença da concordância e a ausência da concordância na retomada anafórica:

Presença da concordância de gênero na retomada anafórica

Essa banana mariquita é só crua, **ela** num frita.

Ausência da concordância de gênero na retomada anafórica

Um dia tava entrando **uma cobra** lá dentro, a gente tava assistindo, quando vê **ele** vinha entrando, sabe.

Entre as variáveis linguísticas selecionadas estatisticamente pelo programa VARBRUL, a anáfora com a retomada do antecedente feminino pelo pronome *ele* aparece mais fortemente condicionada pela referência genérica (0,84) do que pelo traço [inanimado]: é o aspecto da referência genérica e não o da animacidade que entra em jogo no entendimento desse fenômeno variável. Assim, o *ele* retoma qualquer antecedente, seja feminino ou masculino, prioritariamente quando a referência é genérica (Se **a pessoa** já vem com aquela intenção de não respeitar você, você mostra que respeita **ele**).

A natureza morfológica do antecedente também interfere nesse fenômeno, pois nomes invariáveis ou inanimados favorecem a retomada por *ele* (0,54). Segundo Dettoni (2003), o traço [-humano, -animado] (Tira, putchaa **brasa**, apaga, djoga na água, né, aí, num demora **ele**ta fria.) funciona como um divisor de águas no sistema de atribuição de gênero desse dialeto.

A análise da variável presença x ausência de determinante junto ao antecedente revelou que a ausência de determinante (0,61) é um fator favorecedor da retomada por *ele*, corroborando o paralelismo formal, no sentido de que formas gramaticais semelhantes tendem a aparecer juntas (SCHERRE, 1988).

Com relação às variáveis sociais, o grau de escolaridade confirma a hipótese de que a escolaridade (0,34) mais alta tende a inibir a retomada de itens lexicais femininos por *ele*. Nas faixas etárias, a retomada anafórica com o pronome *ele* é mais generalizada na faixa etária acima de 60 anos e mais restrita na fala dos informantes de 15 a 25 anos.

Tendo em vista que a variedade linguística da Baixada Cuiabana encontra-se em processo de mudança linguística no sentido da perda de traços típicos dialetais, que são estigmatizados, em prol da aquisição e fixação de formas mais prestigiadas do português, Dettoni (2003) aprofunda a discussão sobre os cinco problemas da mudança linguística na Baixada Cuiabana, a saber, os problemas da restrição, da transição, do encaixamento, da implementação e da avaliação.

O problema da restrição refere-se a essa variação que ocorre em contextos cada vez mais restritos e tem baixa frequência na fala dos jovens, sendo mais comum na fala de pessoas menos escolarizadas. A mudança linguística em curso no dialeto da Baixada Cuiabana pode ser descrita, atualmente, como a perda da codificação morfossintática do traço semântico [\pm referência genérica].

O problema da transição envolve a descoberta do caminho pelo qual um estágio da mudança evoluiu a partir de um estágio anterior (LABOV, 1972). Segundo Weinreich, Labov e Herzog (1968, p.184): “[...] a mudança ocorre (1) quando um falante aprende uma forma alternativa; (2) enquanto duas formas coexistirem em sua competência; (3) quando uma das formas se torna obsoleta”. Logo o processo de mudança linguística na Baixada encontra-se, atualmente, na transição do segundo para o terceiro momento apontado pelos autores acima.

O problema do encaixamento diz respeito à mudança linguística na matriz social da comunidade investigada, ou seja, é possível identificar, hoje, na Baixada Cuiabana, os seguintes processos em andamento: (i) a transformação de um dialeto regional em um dialeto urbano (LABOV, 1972, p.286) comum à comunidade brasileira como um todo; (ii) o espalhamento de traços linguísticos de prestígio na fala dos usuários do dialeto, em função das pressões sociais recebidas de cima e de fora.

O problema da implementação da mudança deve-se ao pouco acesso ao nível superior em Mato Grosso até o início dos anos setenta; à assimilação de novas formas linguísticas pelos alunos que retornavam a Cuiabá; à melhoria e ao incremento das oportunidades de escolarização na região da Grande Cuiabá; ao início da re-colonização (anos 70 do século XX); ao contato com outras variedades do português.

O problema da avaliação, por meio das falas dos colaboradores, ajuda a entender o verdadeiro conteúdo subjacente às avaliações negativas de pessoas de fora que são vistas como melhores e superiores. Por isso, há três diferentes atitudes sociais em relação ao falar cuiabano: (i) uma que se identifica com o conteúdo das

pressões externas e ratifica o estigma; (ii) outra que tem consciência do estigma e se sente pressionada a anular sua identidade local; (iii) uma terceira que procura preservar a identidade cultural, apesar das atitudes contrárias.

Pacheco (2010) também analisa a concordância de gênero na Baixada Cuiabana, mas no interior do sintagma nominal (**a noite inteiro**) e do sintagma predicativo (**a situação** dele era muito **sério**).

Além disso, Pacheco (2010) coletou vários dados de observação participante e fez uma análise separada apenas da ausência de concordância de gênero nos sintagmas predicativos (**A situação** da Amazônia está a cada dia mais **complicado**) e no sujeito posposto (**É muito mínimo essa concordância**) do português brasileiro na língua falada formal e informal e na língua escrita formal, incluindo dados da mídia. Diante desses dados, percebe-se que há sistematicidade da ausência de concordância de gênero nos predicativos e sujeitos pospostos no português brasileiro como um todo.

Nos dados do português europeu, coletados por Lima (2007), também há esse tipo de exemplo (**A cabeça** do dedo fica muito **grosso**), o que corrobora a existência de um encaixamento linguístico muito maior e mais amplo do que apenas na Baixada Cuiabana.

Para a análise variacionista, Pacheco (2010) utilizou 21 entrevistas do *corpus* organizado por Dettoni (2003). De um total de 2.928 dados coletados, apenas 173 dados representam a ausência de concordância de gênero, totalizando 5,9%, ou seja, trata-se de um fenômeno raro e próximo à categoricidade, com 94,1% de uso pleno da concordância de gênero.

Exemplificando o fenômeno analisado, temos, respectivamente, um dado de sintagma nominal e outro de sintagma predicativo:

Uso pleno da concordância de gênero

Uma criação severa, de verdade.

A senhora olha na Bíblia, tem a Bíblia pra ler, **a Bíblia está escrita**.

Uso não pleno da concordância de gênero

Dai, **moçadinhamái novo**.

Ela é da chapada, **nascido** lá né.

Segundo Pacheco (2010), a hipótese geral é que, apesar de a concordância de gênero não ser um fenômeno tão produtivo no português brasileiro, apresenta padrões muito semelhantes aos da concordância de número, nos termos de

Scherre (1988, 1996), principalmente com relação à posição dos elementos na sentença, já que ambas as concordâncias são mais produtivas à esquerda dos sintagmas nominais. Posto isso, podemos estar diante de um padrão mais geral do português brasileiro em que a concordância de gênero e a de número seguem os mesmos padrões linguísticos.

As variáveis selecionadas como estatisticamente relevantes pelo programa VARBRUL foram, em ordem de significância, a configuração sintagmática, a natureza morfológica, a faixa etária e o grau de letramento.

Sobre as variáveis linguísticas, Pacheco (2010) fez um controle rígido de todos os tipos de sintagma (configurações sintagmáticas) que apareceram nas entrevistas, e a hipótese era de que a posição à esquerda do nome favoreceria a concordância, principalmente nos artigos e possessivos.

Através da configuração dos sintagmas, Pacheco (2010) analisou também a posição dos elementos. O resultado foi que o artigo definido *a*, adjetivo e possessivo + nome (**a roupa da minha mãe/ boa fé/ mia madrinha**) são os grandes responsáveis pela concordância de gênero, ambos à esquerda do nome, com 0,92 de peso relativo. O artigo indefinido *uma* e *numeral* + nome (**uma sala de aula/ uma hora**) vêm logo adiante, também favorecendo a concordância à esquerda, com 0,56. Os elementos que mesmo à esquerda do nome desfavoreceram relativamente a concordância foram o sintagma com dois elementos à esquerda (**o primeiro injeção**), com 0,40; o demonstrativo e indefinido (**esse menina e vários substância**), com 0,27; e o quantificador (tudo criancada), com 0,05.

Todos os elementos à direita do nome desfavoreceram bruscamente a concordância de gênero, como os sintagmas com elementos à esquerda e à direita do nome (**mia casa todo desarrumado**), com 0,09; o possessivo e o adjetivo (**casa nosso, pessoa cego**), com 0,07; o predicativo (**minha família toda é cuiabano**), com 0,02; e o quantificador (**as vacina tudo**), com 0,01 (SCHERRE, 1988, 1996).

É importante registrar que, como se trata de um fenômeno morfossintático e raro, a pouca quantidade de dados da variante ausência de concordância de gênero pode interferir nos resultados finais. Assim sendo, é necessário polemizar um pouco mais sobre os elementos à esquerda que teoricamente desfavoreceram a concordância de gênero. Sintagmas de três elementos com dois elementos à esquerda (**o primeiro injeção**), os demonstrativos e indefinidos à esquerda do nome (**esse menina/ vários substância**) obtiveram um peso relativo inferior a 0,50. No entanto os dados dos sintagmas de três elementos com elementos à esquerda são muito poucos em comparação com os demais. Essa diferença também deve ser levada em consideração na análise, porque, com um *input* de 0,99, qualquer valor um pouco mais abaixo será desfavorecedor (PACHECO, 2010).

Logo, para se obter resultados mais sintéticos e gerais, os sintagmas foram reunidos em quatro fatores: (i) elementos à esquerda do nome; (ii) elementos à direita do nome; (iii) quantificadores em ambas as posições; e (iv) predicativos em ambas as posições. Assim sendo, pode-se generalizar que apenas os elementos à esquerda favorecem a concordância de gênero (0,69), em detrimento dos elementos à direita (0,16), quantificadores (0,06) e predicativos (0,04). Portanto, a hipótese da posição à esquerda para a concordância de gênero é confirmada, principalmente com o artigo + nome, adjetivo + nome e possessivo + nome.

Para a natureza morfológica, Pacheco (2010) também fez um detalhamento grande, que se polarizou entre nomes variáveis por léxico diferente (**galinha de casa só**, **criado co milho**), com peso relativo de 0,88; itens lexicais morfológicamente variáveis (**esse menina**), com 0,71; e itens lexicais invariáveis (**banana fritado**), com 0,43. Essa variável mostra como a flexão dos nomes e a saliência fônica, no caso dos nomes variáveis por léxico diferente, favorecem a concordância de gênero, além de os nomes variáveis também terem a presença do traço [+animado +humano], com 0,80.

Nas variáveis sociais, de acordo com Pacheco (2010), a faixa etária foi bastante consistente, em termos do favorecimento da concordância apenas para os falantes de 15 a 25 anos (0,75). Tanto a faixa etária de 30 a 45 anos (0,48) quanto a faixa acima de 60 anos (0,38) desfavorecem o uso da concordância de gênero. Esse é um grande indício de mudança linguística, em que os jovens estão à frente no uso da concordância de gênero, corroborando os resultados de Dettoni (2003).

Outra variável social significativa, em Pacheco (2010), foi o grau de letramento, constituído a partir da escolaridade e dos graus de letramento que os informantes possuíam, através de uma análise qualitativa e quantitativa. A identificação individual de cada informante também foi importante para ver a questão do desempenho linguístico diferente que muitos tinham.

O objetivo de criar uma variável social surgiu do resultado do grau de escolaridade, já que os colaboradores analfabetos (0,53), com cinco a oito anos de estudo (0,59) e mais de nove anos de estudo (0,59) favoreceriam a concordância de gênero. Assim, buscou-se interpretar, qualitativamente por meio das entrevistas, o que realmente estava em jogo diante desses resultados.

As práticas sociais dos colaboradores revelaram que os traços não marcados do letramento e da escolaridade juntamente (0,32), ou seja, ausência de letramento e ausência de escolaridade, é que foram desfavorecedores da concordância de gênero, enquanto os traços [-escolaridade +letramento], [+escolaridade +letramento], [+escolaridade +letramento] favoreceram a concordância, respectivamente com 0,59, 0,55 e 0,52.

Essa variável não é uma categoria social fixa como as que normalmente são analisadas pela Sociolinguística Variacionista, pois é uma reanálise da escolaridade com base em outros fatores que vão além da sala de aula, como o nível sociocultural, econômico, acesso à mídia, ocupação, participação em eventos sociais, viagens, moradia na zona rural ou urbana etc. Nesse ponto, trata-se de uma visão ampla e mais complexa de vários fatores sociais inseridos no grau de letramento.

Com base nas pesquisas de Dettoni (2003) e Pacheco (2010), é possível analisar o encaixamento linguístico da concordância de gênero na Baixada Cuiabana, fazendo o seguinte paralelo:

- 1) Segundo Dettoni (2003), do ponto de vista do encaixamento linguístico, a mudança na concordância de gênero na Baixada Cuiabana inicia-se pelo contexto Determinante + Nome, espalhando-se, em seguida, para outras estruturas; as construções Nome + Adjetivo e Sujeito + Predicativo apresentam-se como os contextos mais resistentes à mudança;
- 2) Pacheco (2010) discute a possibilidade de a mudança – presença categórica de concordância – não se concretizar nos sintagmas predicativos ou, pelo menos, demorar mais a chegar, uma vez que a ausência da concordância nesses casos já acontece de forma mais geral e não se limita à área geográfica de Cuiabá, ocorrendo em todo o território brasileiro, tanto na fala como na escrita. Juntamente com os predicativos, os quantitativos também são os grandes obstáculos para a concordância de gênero, uma vez que o peso relativo desses fatores fica abaixo de 0,06.
- 3) De acordo com Pacheco (2010), o peso relativo baixo dos predicativos indica que eles têm menos probabilidade de ter concordância de gênero do que os SNs. Logo os SNs estão na direcionalidade contrária da mudança, já que a probabilidade da presença da concordância de gênero é maior, principalmente nos artigos, possessivos, adjetivos e numerais à esquerda do nome. Sendo assim, percebe-se que os predicativos têm um encaixamento linguístico maior que o encaixamento social dentro do sistema linguístico da língua portuguesa.
- 4) Na pesquisa de Dettoni (2003), o contexto mais favorável à retomada pelo anafórico *ele* é a referência genérica, ausência de determinante, ausência de flexão de gênero, faixa etária mais idosa e escolaridade baixa. Na pesquisa de Pacheco (2010), o contexto mais favorável à ausência de concordância de gênero é a posição à direita do nome, predicativos e quantificadores, ausência de flexão de gênero, faixa etária mais idosa e grau de letramento menor.
- 5) Diante do exposto, a concordância de gênero nos sintagmas nominais pode estar em processo de mudança linguística, no sentido de um uso cada vez mais restrito da ausência de concordância nos SNs, já que o *input* para a realização plena de concordância é de 0,99, um resultado próximo à categoricidade, e um uso um pouco menor da concordância de gênero nos predicativos isolados, com um *input* de 0,74 (PACHECO, 2010).

- 6) Em suma, as duas análises corroboram a idéia de que a variação na concordância de gênero na Baixada Cuiabana segue os padrões maiores de variação de concordância de número no português brasileiro, o que tem sido atestado pelos trabalhos de Scherre para a concordância de número. Então, o padrão linguístico é o mesmo, ou seja, formas à esquerda concordam mais do que formas à direita do núcleo. Além disso, Pacheco (2010) mostra que a ausência de concordância de gênero nos sintagmas predicativos, independentemente de estarem na ordem direta ou indireta, é generalizada no português brasileiro.

Conclusão

O Projeto Valco atingiu, até o momento, os objetivos inicialmente apresentados, uma vez que traz em seus arquivos registros de língua em uso real, com seus respectivos áudios e transcrições, adicionados às reflexões já realizadas sobre os dados. Além disso, veicula alguns artigos e trabalhos científicos sobre o Centro-Oeste.

As pesquisas aqui apresentadas formam o escopo de estudos amplamente analisados e divulgados no site do Projeto Valco. Tal projeto, porém, ainda precisa ser plenamente difundido no meio acadêmico e alimentado com mais entrevistas, a fim de se fortalecer e se estabelecer como importante veículo difusor de conhecimentos sobre a língua em uso no Centro-Oeste brasileiro.

Como afirmado anteriormente, acreditamos que esse projeto tem muito a contribuir para o estudo do português em diversas áreas linguísticas, bem como para historiografia do português em terras brasileiras, entre tantos outros estudos na área de Letras; e, por fim, acreditamos que esse projeto seja um meio de desenvolver e difundir resultados dos estudos sociolinguísticos realizados nessa região do país.

DETTONI, R. do V.; PACHECO, C. da S.; ANDRADE, C. Q.; SCHERRE, M. M. P. Linguistic variation project in the midwest region of Brazil – Valco. *Alfa*, São Paulo, v.56, n.3, p.797-823, 2012.

ABSTRACT: The aim of our article is to present Valco Project (Linguistic Variation of Brazil's Midwest Region) and summarize the results of some researches in sociolinguistics from different geographical areas (in the Midwest) and different themes. Four researches were developed with data from Brasília and theyh analyze the second person pronouns: Tu/Você/Ocê/Cê. In these studies, we present a central idea that the variety from Brasília is currently introducingthe new variant TU in its linguistic repertoire, focusing on a trend in Brazilian Portuguese. Two other researches were developed with data from the BaixadaCuiabana region, in MatoGrosso, and they analyzed the gender variation present on the discourse and syntagmatic plans. Unlike

the previous researches mentioned here, those two studies present linguistic features of semi categorical nature, which are probably in an extinction situation. But it is evident that gender variance in the analyzed data is linguistically embedded in the system of Portuguese language. Thus, the present article, which is based on Language Variation and Change Theory, brings to the scientific community a wider knowledge of Portuguese language in Brazilian lands, more specifically, in traditional Central Brazil and in expanded Central Brazil.

■ KEYWORDS: *Linguistics Variation. Midwest region. Linguistic variety from Brasília. Linguistic variety from Cuiabá. Second person pronouns. Gender agreement. Dialectfocusing.*

REFERÊNCIAS

- ANDRADE, A. L. V. S. *A variação você, cê eocê no português brasileiro falado.* 145f. 2004. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 2004.
- ANDRADE, C. Q. *Tu e mais quantos? A segunda pessoa na fala brasiliense.* 132f. 2010. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 2010.
- BORTONI-RICARDO, S. M. *The urbanization of rural dialects speakers: a sociolinguistic study in Brazil.* Cambridge: CUP, 1985.
- CORRÊA, C. da C. *Focalização dialetal em Brasília: um estudo das vogais pré-tonicas e do /s/ pós-vocálico.* 98f. 1998. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 1998.
- DETTONI, R. do V. *A concordância de gênero na anáfora pronominal: variação e mudança linguística no dialeto da baixada cuiabana – Mato Grosso.* 256f. 2003. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal de Minas Gerais, Minas Gerais, 2003.
- DIAS, M. C. A. *A variação na concordância nominal: um contraste entre o urbano e o rural na fala brasiliense.* 174f. 1993. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 1993.
- DIAS, E. P. *O uso do tu no português brasiliense falado.* 114f. 2007. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 2007.
- HANNA, E. S. *Difusão e focalização dialetal: o caso de Brasília.* 184f. 1986. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 1986.
- LABOV, W. *Padrões sociolinguísticos.* Tradução Marcos Bagno, Maria Marta Pereira Scherre, Caroline Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008.

- _____. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- LIMA, J. L. *A variação na concordância do gênero gramatical no falar cuiabano*. 248f. 2007. Tese (Doutorado em Linguística) – Universidade de Campinas, São Paulo, 2007.
- LUCCA, N. N. G. *A variação tu/você na fala brasiliense*. 139f. 2005. Dissertação (Mestrado em Linguística) - Universidade de Brasília, Brasília, 2005.
- MALVAR, E. da S. *A realização do objeto direto em cadeia anafórica no português do Brasil*. 153f. 1992. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 1992.
- PACHECO, C. da S. *Padrões sociolinguísticos da concordância de gênero da baixada cuiabana* 162f. 2010. Dissertação (Mestrado em Linguística) – Universidade de Brasília, Brasília, 2010.
- PINTZUK, S. *Programas VARBRUL*. Tradução de Ivone Isidoro Pinto. Rio de Janeiro: UFRJ, 1988.
- SANKOFF, David; TAGLIAMONTE, Sali; SMITH, Eric. Goldvarb X: *A variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto. 2005.
- SCHERRE, M. M. P. Sobre a influência de três variáveis relacionadas na concordância nominal em português. In: SILVA, G. M. de O.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos - análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro: Ed. da UFRJ, 1996. p.41-62.
- _____. *Reanálise da concordância nominal em português*. 555f. 1988. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1988.
- WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical foundations for a theory in language change. In: LEHMANN, W. P.; MAKIEL, Y. (Ed.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p.95-195.

Anexo I

Normas de Transcrição do Projeto VALCO

Cada pesquisador, ao se interessar pelo banco de dados do VALCO, deverá adaptar suas entrevistas às normas de transcrição elaboradas para o projeto VALCO, no intuito de padronizar as entrevistas e facilitar a consulta por outros interessados.

Tipos propostos de transcrição:

- a) Fonográfica** – transcrição da fala, respeitando e reproduzindo, na medida do possível, a oralidade;
- b) Convencional** – adaptação da ortografia aos objetivos pretendidos.

Orientações gerais:

- a) A transcrição não pode ser sobre carregada de símbolos e deve ser limpa e legível.
- b) A transcrição deve permitir a compreensão do significado do texto.
- c) A produção oral (tá, tô, tamos, pra, eles foi, os menino, memo) deve ser respeitada, sem a necessidade de uma transcrição fonética.
- d) Os falantes devem ser indicados apenas com a sigla do nome (as três primeiras letras).
- e) Deve-se procurar facilitar ao leitor a criação de uma “imagem” do texto elaborado no plano da oralidade.

Normas para transcrição de dados:

- 1) Letra Times New Roman, 12, espaçamento 1.5.
- 2) Os participantes devem ser numerados, de acordo com as fichas catalográficas de cada interação.

Ex.:

Participante: MAR.

Escolaridade: analfabeta.

Ocupação: lavradora, dona de casa e merendeira da escola.

Lugar de origem: Córrego Puba (Vendinha), em Pombal.

Idade: 50 anos.

Data: 27/07/1997.

Tipo de entrevista: entrevista tipicamente laboviana.

Lugar da entrevista: Cotidiano familiar.

Duração: 40 minutos

Temas abordados: Emancipação do município de Santa Rita do Novo Destino.

- 3) Ao transcrever os dados, coloque em negrito os enunciados do pesquisador e sem destaque os do pesquisado.

Ex.: **E - Como foi isso?** (E, de entrevistador)

P - Foi trapaiado, foi. (P de participante);

- 4) **Ênfase** - usar letras maiúsculas.

Ex.: Mais eu QUERO brincar.

- 5) **Pontuação** – o ponto de interrogação, nas frases interrogativas, e o de exclamação, nas exclamativas, são mantidos.

- 6) **Dúvida quanto à palavra** – a palavra sob dúvida deve ficar entre colchetes angulares.

Ex.: Ele <andava> muito.

- 7) **Fala sobreposta** (Cruzamento de vozes) – os enunciados pronunciados por dois falantes ao mesmo tempo devem estar conectados por um colchete.

Ex.: Que legal!!

Que legal!!

- 8) **Alongamento de vogal** – após a vogal alongada são usados quatro pontos.

Ex.: Ele gostava, e co::mo gostava!

- 9) **Silabação** – colocar hífen no meio da palavra.

Ex.: ca-fé, ca-mi-nha-da.

10) Repetições – reduplicar palavras ou sílabas repetidas.

Ex.: Aí, e ele foi pra casa de de Carlos.

11) Comentários do pesquisador acerca de atitudes não linguísticas do informante devem ser colocados entre parênteses.

Ex.: Ele gosta de mim, (risos)

(baixou o tom de voz)

12) Comentário de outro participante – colocar o símbolo # (3^a pessoa falando).

Ex.: # (3^a pessoa falando)_Ele, ele num cunhece. {inint} é melhó. {inint} má lá embaixo daqui lá.

13) Palavra ou trecho ininteligível – o comentário deve estar entre chaves.

Ex.: Maria queria comprar {inint}, a mãe dela falou que não queria.

14) Pausa pequena – usar vírgula.

15) Pausa média – usar ponto.

16) Reticências: suspensão do pensamento e truncamento de palavras.

Ex.: Eu gostaria de ir para...sei lá, acho melhor morar em cas...apartamento.

Referências das Normas de Transcrição do VALCO

- Projeto Vozes do Amapá.
- Núcleo de estudos da história linguística de Goiás. Normas para coleta e transcrição de dados. Faculdade de Letras, UFG.
- MARCUSCHI, L. A. **Análise da conversação**. Editora Ática, 2000.

Recebido em março de 2012.

Aprovado em julho de 2012.

BANCO DE DADOS VARSUL: A RELEVÂNCIA DE SUAS CARACTERÍSTICAS E A ABRANGÊNCIA DE SEUS RESULTADOS

Gisela COLLISCHONN*
Valéria de Oliveira MONARETTO**

- RESUMO: O Banco de Dados VARSUL (Variação Linguística Urbana do Sul do País) reúne dados de fala de localidades sócio e culturalmente significativas de três Estados do Sul do Brasil. Este artigo apresenta uma reflexão sobre a utilização dos dados de coletâneas de entrevistas levantadas segundo metodologia laboviana, tanto na área de variação linguística quanto em diversas outras áreas de pesquisas, especialmente no que diz respeito a pesquisas fonológicas. A importância e as vantagens obtidas por análises em *corpora* de língua falada espontânea são abordadas. Discutimos também a possibilidade que levantamentos feitos em moldes de entrevistas sociolinguísticas oferecem para a interface com outras áreas de pesquisa. Algumas características da constituição do Banco VARSUL são apresentadas, tais como a metodologia de obtenção de dados, critérios de seleção dos informantes, modelo de transcrição, facilidades de buscas, entre outras. Ao final, apresentamos um breve levantamento de pesquisas realizadas com esses dados e que tiveram, como objeto de análise, fenômenos fonético-fonológicos variáveis a fim de ilustrar a abrangência de fenômenos já estudados e de mostrar a possibilidade de generalização de resultados.
- PALAVRAS-CHAVE: VARSUL. Banco de Dados de Fala. Variação Fonológica.

Introdução

A documentação da língua falada é fundamental para que os estudiosos da linguagem tenham onde buscar comprovação empírica para suas hipóteses e teorias. Nesse sentido, parece-nos que um banco de dados de fala tenha relevância implícita, independentemente da metodologia de coleta e de organização dos dados. Entretanto, dado que as abordagens em termos de coleta de dados e de foco da investigação linguística vão se aperfeiçoando e se modificando continuamente, cremos que cabe a discussão sobre a questão da relevância de um banco, como o VARSUL (Variação Linguística Urbana do Sul do País).

* UFRGS – Universidade Federal do Rio Grande do Sul. Instituto de Letras – Departamento de Línguas Clássicas e Vernáculas. Porto Alegre – Rio Grande do Sul – Brasil. 91350-300 - giselac@via-rs.net

** UFRGS – Universidade Federal do Rio Grande do Sul. Instituto de Letras – Departamento de Lingüística, Filologia e Teoria Literária. Porto Alegre – Rio Grande do Sul – Brasil. 91350-300 - monar@terra.com.br

O Banco VARSUL é fruto de cuidadosa coleta de dados de língua falada que compreende amostras representativas de cidades dos três estados do Sul: Paraná, Santa Catarina e Rio Grande do Sul. Na formação do Projeto VARSUL, definiu-se que quatro cidades de cada estado fariam parte do *corpus*. Cada cidade é representada por um conjunto de 24 entrevistas. Os indivíduos da amostra estão distribuídos por sexo, escolaridade e idade. Segundo a metodologia laboviana, as entrevistas buscam ser representativas da linguagem falada (autêntica e natural, na medida do possível). Por isso, busca-se deixar o informante discorrer sobre diversos temas (em geral, sobre algum aspecto da história de vida do entrevistado), sem muita interferência do entrevistador, justamente para tentar minimizar a artificialidade da situação. Os dados de fala são gravados e posteriormente transcritos.

O VARSUL foi criado em uma época em que eram, no País, extremamente escassos registros de língua falada e estudos empíricos descritivos sobre variáveis linguísticas em geral. O projeto pioneiro nesse sentido foi o Projeto Censo, proposto em 1980 com o objetivo de se estudar a variação e mudança no Rio de Janeiro. Entre 1980 e 1984, foram coletadas entrevistas com informantes distribuídos por diferentes faixas etárias, níveis de escolaridade e sexo. Essa amostra foi ampliada ao longo dos anos, formando o acervo de dados do Programa de Estudos sobre os Usos da Língua (PEUL). O projeto PEUL tornou-se referência para os demais projetos de coleta de língua falada, como o VARSUL, VALPB e outros.

Na região Sul, especificamente, dados que atestassem o uso da língua falada, haviam sido coletados, pioneiramente, pelo Projeto Norma Urbana Culta (NURC), que buscou investigar a fala de indivíduos com curso superior nas cinco capitais mais populosas à época, no início da década de setenta, e que incluiu a cidade de Porto Alegre, no Rio Grande do Sul.

Ao fazer uma retomada da história dos estudos sociolinguísticos, Eckert (2012) identifica três ondas sucessivas. A primeira focava o estabelecimento de padrões de correlação entre variáveis linguísticas e as grandes categorias classe social, sexo, etnia e idade, consideradas significativas em qualquer comunidade linguística. A segunda, com o emprego dos métodos da etnografia, buscava explorar categorias mais locais e identificar as configurações que constituem as categorias mais amplas. Nestas duas primeiras ondas, a perspectiva principal era de que a variação seria reflexo das categorias sociais pré-estabelecidas. Na terceira onda, segundo Eckert (2010), o estudo da variação focaliza o significado social dos empregos das formas linguísticas e procura identificar as categorias que são construídas socialmente em uma determinada comunidade.

A constituição do Banco VARSUL está inserida na primeira onda, que Eckert (2012) chama de “era dos inquéritos”. Levantamentos em grande escala, com base no modelo pioneiro de Labov (1966), foram sendo desenvolvidos para diversas línguas, envolvendo um grande número de informantes. Com isso, foi

possível obter uma visão ampla da variação e dos padrões de distribuição social das variáveis. Entre as principais virtudes desses levantamentos estão a sua cobertura e replicabilidade. No entanto, essas vantagens vêm com o senão de que dependem do emprego de categorias sociais pré-determinadas, empregadas para classificar os informantes. Como resultado, não permitem uma visão mais detalhada dos papéis que os informantes de fato desempenham dentro do grupo.

A segunda onda utilizou métodos etnográficos para se aproximar das dinâmicas locais de variação e, por isso mesmo, concentrou-se em pequenas comunidades, por períodos relativamente longos de tempo. Em vez de pressupor que determinadas categorias sejam relevantes, os métodos buscam descobrir quais são as categorias localmente relevantes. Esta segunda onda tomou forma a partir das conclusões da onda anterior, isto é, a partir da constatação de que nem sempre as macrocategorias apresentavam o comportamento esperado a partir das supostas características sociais do fenômeno analisado.

Na terceira onda, inverte-se a abordagem. Em vez da perspectiva de variação como um reflexo do lugar social, assume-se a perspectiva da variação como um recurso para a construção de sentido social. Segundo Eckert (2012), a terceira onda nos afasta da perspectiva macrossociológica, olhando para a negociação momento a momento das relações de poder entre os falantes.

Neste artigo, em que reapresentamos aos leitores o Banco VARSUL, é importante – além de destacarmos resultados que já foram encontrados – refletirmos a respeito do acesso ao Banco por uma comunidade mais ampla de pesquisadores e sobre a relevância que dados coletados no contexto de uma perspectiva sociolinguística characteristicamente da primeira onda podem ter para outras perspectivas de investigação. Nas próximas duas seções, trazemos algumas reflexões nesse sentido, considerando o valor que pesquisas de *corpora* têm adquirido nos desenvolvimentos teóricos recentes e as vantagens de *corpora* de língua falada espontânea em comparação com a fala captada em condições de laboratório. Em seguida, passamos a apresentar a caracterização do Banco, explicitando os cuidados metodológicos que foram tomados para garantir uma coleta tão fidedigna quanto possível. Na última seção, apresentamos resultados já obtidos, no que se refere às pesquisas na área de variação fonológica.

Por que *corpora* em larga-escala?

Segundo Eckert (2012), nenhuma onda substitui a anterior, embora cada uma represente uma forma bastante diferente de se pensar sobre a variação e também uma prática metodológica e analítica distinta. Como destaca Eckert (2012), os amplos padrões demográficos de variação não deixaram de ser importantes. *Corpora* como o VARSUL, por se constituírem de amostras aleatoriamente

estratificadas segundo padrões sociais pré-estabelecidos, têm valor por possibilitar o (re)conhecimento da distribuição de variáveis linguísticas de acordo com esses padrões. Dada a riqueza e abrangência das entrevistas coletadas, ainda há muito o que pesquisar em termos dessas correlações.

Também deve-se destacar que o estudo da variação ganhou novo impulso, deixando de ficar restrito à comunidade sociolinguística. No que concerne à relevância da variação em fonologia, por exemplo, há uma clara mudança em curso. Coetze e Pater (2011), em recente capítulo sobre a variação em fonologia, na segunda edição do *Handbook of Phonological Theory* (GOLDSMITH; RIGGLE; YU, 2011), notam que, na primeira edição do Handbook, sequer havia um capítulo sobre variação, ao passo que, na atualidade, qualquer texto que tenha como proposta abranger a área de fonologia adequadamente e que pretenda dar uma perspectiva consequente do estado da arte na fonologia teórica não pode deixar de ter uma seção ou um capítulo dedicado à variação.

Ainda em Hinskens, Hout e Wetzel (2000, p.8-9) observavam o seguinte:

[...] nem todos os linguistas orientados mais teoricamente parecem estar suficientemente cônscios de que a variação é uma característica essencial da língua, tanto quanto um pré-requisito para a evolução linguística. [...] Os poucos linguistas teóricos que mostram sensibilidade para o problema da variação raramente têm idéias claras com relação ao lugar da variação linguística em seus modelos.

Essa realidade parece já ter mudado bastante. Há uma maior integração, que se deve, em parte, aos desenvolvimentos das teorias fonológicas, no sentido de incorporar um componente probabilístico, e, em parte, ao fato de que os trabalhos em variação cada vez mais buscam levar em consideração as análises elaboradas pela fonologia teórica na tentativa de entender as relações estruturais entre os fenômenos linguísticos.

Dados quantitativos lançam luz sobre muitas questões que não podem ser respondidas simplesmente pela intuição e permitem comparar estruturas hipotéticas com estruturas atestadas. Ou seja, estudos de *corpora* sociolinguísticos podem contribuir para uma sustentação mais sólida na realidade de análises de caráter mais teórico. Além disso, conforme nos alerta Durand (2009, p.39), baseando-se apenas na intuição, o linguista pode estar correndo o risco de formalizar seus pressupostos normativos sobre a língua. Por mais que a linguística moderna reconheça a prioridade da fala sobre a escrita e rejeite o prescritivismo, é muito fácil para o analista tornar-se refém de conceitos derivados de ortografia e do ensino escolar caso não se debruce com alguma regularidade sobre dados de língua falada.

Finalmente, é de se destacar que *corpora* como o do Banco VARSUL permitem uma interface com outras áreas, tais como a pesquisa de *corpus* e a área de pesquisas geolinguísticas, como demonstram pesquisas como a de Brandão (2009), Mota (2009), Callou (2010), entre outros.

Kendall (2011) observa que é na pesquisa sociolinguística da primeira onda, e talvez também na da segunda onda, que estudos de linguística de *corpus* podem encontrar correspondência. Segundo o autor:

[...] it seems likely that continued technological advancements in the development, annotation, and analysis of corpora will lead to increased opportunities for sociolinguistic engagement with corpora. This is especially true for research investigating aspects of language and social structure (i.e. work in the first and second waves), though I believe it is still the case for work that is less interested in large-scale quantitative study. (KENDALL, 2011, p.371).

A importância desta questão está no fato de que o desenvolvimento e o compartilhamento de *corpora* atendem a um dos requisitos básicos do método científico, a replicabilidade. Segundo Kendall (2011, p.371), “[...] *By working from a shared pool of data, researchers are best able to collectively develop agreed upon knowledge about language. This, I believe, is a major benefit of corpus-based work*”. Embora tenha se tornado mais comum, o compartilhamento de dados ainda não é prática corrente entre os pesquisadores da área da sociolinguística. Kendall identifica duas razões principais. A primeira é que o levantamento de *corpora* é um trabalho árduo, que nem sempre é reconhecido; portanto, faz sentido que os grupos que trabalharam no seu levantamento queiram limitar o acesso aos dados enquanto estão trabalhando em suas análises. A outra razão diz respeito ao direito à privacidade dos informantes. As coletas foram feitas, via de regra, com a autorização do informante de que os dados fossem utilizados em pesquisas com um determinado fim. A publicização dos dados nem sempre resguarda a garantia dada ao informante. Além disso, em entrevistas sociolinguísticas, frequentemente surgem temas privados do informante que este possivelmente não queira tornar públicos. Assim, a publicização dependeria de se poder recontatar todos os informantes e de se obter deles a autorização de compartilhamento. Como esse é um problema comum a entrevistas sociolinguísticas, é importante que se divulguem e se discutam protocolos de anonimização, como o que foi desenvolvido pelo Laboratório MUSL (Memorial University Sociolinguistic Laboratory) (CHILDS; VAN HERK; THORBURN, 2011).

O Banco VARSUL tem se ocupado com essas questões ao desenvolver o Projeto Amostra Digital, no qual entrevistas de informantes das três capitais, Porto Alegre, Florianópolis e Curitiba, são disponibilizadas a um público mais

amplo¹¹. Até o momento, a solução encontrada tem sido a de disponibilizar um conjunto de trechos das entrevistas das capitais Porto Alegre, Florianópolis e Curitiba, de forma a que pesquisadores de diversas áreas tenham uma amostra do material registrado no Banco. Para a escolha dos trechos disponibilizados, buscaram-se aquelas passagens em que o entrevistado discorre livremente, com o mínimo de intervenção do entrevistador. Mas também buscou-se privilegiar a escolha de passagens que não identificassem o informante. O conjunto completo das entrevistas somente está disponível aos pesquisadores associados ao Projeto.

Entrevista sociolinguística versus outros tipos de coleta

Corpora de língua falada espontânea, coletados sistematicamente, têm uma série de vantagens em relação a outros tipos de levantamento. Segundo Ernestus e Baayen (2011, p.377):

[...] speech corpora are substantially broadening the empirical scope of phonological research. Corpora show that many well-established basic facts that constitute a kind of canon feeding both phonological theory and psycholinguistic theories involve substantial simplifications that do not do justice to the variability and the range of phenomena that characterize actual speech.

Em fonologia e fonética, análises de fala de laboratório são muito comuns e, muitas vezes, são a única forma disponível para se obter uma análise segura e confiável. No entanto, não deixam de ter desvantagens. As quantidades de dados coletados tendem a ser pequenas e normalmente não podem ser reutilizadas para fins diferentes. Além disso, estilos de fala eliciados em contexto de experimentos tendem a ser formais e não espontâneos, e materiais são apresentados de forma isolada ou em contextos pequenos, muitas vezes artificiais.

Em comparação com a fala de laboratório, obtida em resposta a instruções específicas, a fala casual apresenta enorme variabilidade na realização, refletindo não somente a filiação do falante a uma determinada comunidade linguística mas também padrões prosódicos que transmitem a estrutura de informação e o contexto pragmático de um enunciado, condições que nem sempre estão disponíveis na fala eliciada. Na visão de Ernestus e Baayen (2011), *corpora* de língua falada são importantes complementos para as análises experimentais.

Os autores citam como característica dos *corpora* de língua falada a revelação de que determinados processos fonético-fonológicos considerados de aplicação

¹¹ Disponível em: <www.varsul.org.br>.

categórica são, na verdade, de aplicação opcional, alguns dos quais com taxas de aplicação menores do que 50%. Um exemplo apresentado refere-se ao processo de assimilação regressiva de vozeamento do holandês, considerado de aplicação categórica em muitas análises fonológicas; no entanto, um levantamento em *corpus* de língua falada mostrou que somente em 43% das ocorrências analisadas houve aplicação do vozeamento, contrariando os 100% esperados. Outro aspecto que os *corpora* de língua falada revelam são os complexos padrões de redução de formas linguísticas. Segundo Ernestus e Baayen (2011), estas reduções não parecem ter caráter puramente fonético, já que os padrões de redução parecem estar sujeitos a uma variedade de condições intrinsecamente fonológicas. Os autores mencionam como exemplo o fato de que um elevado grau de redução é observado apenas para palavras que não portam acento sentencial e que se encontram em posição medial de enunciado, evidenciando que o padrão de redução está sujeito às mesmas condições de estrutura prosódica de processos de ordem fonológica.

A vantagem de *corpora* já coletados é economia de tempo e de recursos e também o acesso a um Banco de dados muito maior do que um único pesquisador individual poderia construir. Além disso, existe a possibilidade de tomar como base outros trabalhos que já tenham sido feitos usando o mesmo *corpus*, ou utilizando os resultados anteriores como referência para testar novos métodos de pesquisa. No entanto, para que uma coleção de arquivos de fala possa transcender os objetivos das pesquisas iniciais, que levaram à sua coleta, é preciso que os dados possam ser localizados através de transcrições, sendo o tipo de transcrição mais comum a transcrição ortográfica simples, que serve à função de fornecer pistas para acessar as formas a serem analisadas nos arquivos de voz.

No caso do Banco VARSUL, o *corpus* principal está totalmente transscrito, e cada transcrição é acompanhada de um arquivo com informações de características sociais dos falantes, incluindo idade, sexo, escolaridade, residência, tipo de trabalho que realiza, entre outros. Além disso, uma transcrição fonológica ampla está sendo desenvolvida com a ajuda de reconhecimento automático de fala para os trechos de fala disponibilizados no site da Amostra Digital. Este tipo de levantamento poderá ser útil, em pesquisas futuras, para a identificação de padrões de redução, medidas de taxa de elocução e de outras características da fala conectada e para a realização de uma mesma palavra ou locução em diferentes contextos prosódicos.

O Banco de Dados VARSUL

Os dados de língua falada do Projeto VARSUL foram coletados em situações espontâneas e, em certa medida, formam um *corpus* em larga-escala, por abrangerem amostras de diferentes regiões de um certo espaço geográfico e por contemplarem indivíduos de diferentes estratificações socioculturais. Desde o

seu início, o Projeto era interdisciplinar, o que, em nosso entender, trouxe uma série de vantagens ao grupo.

O Banco de dados VARSUL reúne um grupo de professores pesquisadores ligados a três universidades da região sul do Brasil: Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Paraná (UFPR) e Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Iniciou-se em 1990, com objetivo geral de oferecer subsídios para a descrição do português falado no País (BISOL; MENON; TASCA, 2008). Os dados estão disponíveis para pesquisas da comunidade acadêmica em geral, de acordo com regras estabelecidas pelo Projeto.

O Banco de dados dispõe de fichas sociais digitalizadas, contendo algumas informações sobre os entrevistados como estado civil, quantidade de filhos, renda familiar, religião, etc. As entrevistas foram realizadas em forma de diálogo entre duas pessoas, em *estilo não controlado*, tendo como tema principal a vida na cidade. Nas entrevistas, buscou-se obter narrativas sobre assuntos que fossem, ao mesmo tempo, de interesse do entrevistado e que pudessem ter pontos em comum com os relatos de outros informantes, como histórias que os pais contavam, um momento de raiva/alegria, etc. Busca-se, assim, captar, tanto quanto possível, a fala vernacular.

O material gravado foi transscrito segundo um sistema de transcrição próprio do Projeto, constituído por três linhas, para se possibilitar o acesso a diferentes informações e pesquisas em diferentes níveis linguísticos. Na primeira linha, é registrada, em forma ortográfica, a sintaxe real da fala dos informantes com interrupções e hesitações. Na segunda linha, registram-se aspectos fonéticos variáveis e as pausas. Na terceira linha, apresentam-se informações sobre a ênfase, mudança de velocidade de fala e uma classificação morfossintática dos itens lexicais. O roteiro de transcrição bem como os códigos específicos de cada linha estão disponíveis no *Manual do Usuário* do Projeto².

Essa forma de transcrição possibilitava, já na época de constituição do Banco, pesquisas de busca de palavras ou de sequência de palavras, assim como a sequência de caracteres da primeira linha ou de códigos usados nas linhas dois e três. Para auxiliar nesse sentido, foi criado um programa especial para o Projeto, denominado de *Interpretador*, que possibilita pesquisas lexicais, morfológicas e morfossintáticas, além das fonológicas nos arquivos de transcrição dos dados. Atualmente, este programa continua sendo usado, mas sua aplicabilidade é restrita à localização das ocorrências pretendidas e não permite contagens de ocorrências ou outros levantamentos mais sofisticados. Além disso, tendo sido concebido ainda na pré-história da linguagem computacional, por diversas razões, não pôde ser atualizado; em função disso, apesar de sua utilidade, acaba tendo um emprego limitado. Entretanto, novas ferramentas de busca desenvolvidas pela

² Knies; Costa (1996). Ver também Menon; Fagundes; Lorean-Penkal (2009).

linguística de *corpus* podem ser adaptadas ao emprego do Banco, o que deve se constituir em um dos desenvolvimentos subsequentes do Projeto VARSUL. Os dados de fala e suas transcrições estão disponíveis nas quatro sedes do Projeto³.

As amostras foram coletadas de 1988 a 1996 e são constituídas por dados de fala de indivíduos pertencentes a cidades étnico e/ou culturalmente representativas de cada Estado. Foram coletadas, seguindo a metodologia laboviana, 24 entrevistas, 96 por Estado, formando um total de 288, cada uma delas com duração aproximada de 60 minutos. Os indivíduos da amostra estão distribuídos por sexo, escolaridade e idade. A escolaridade é controlada por três graus de instrução, cuja nomenclatura antiga, primário, ginásio e 2º grau, assim como sua estratificação, é substituída, nos dias atuais, geralmente, pela classificação por anos de estudo. Assim, os informantes estão divididos em até quatro, até oito e até onze anos de instrução. A idade, originalmente disposta em duas faixas, é, em alguns trabalhos, remodelada para três. Eis as cidades que compõem o Banco de dados básico:

(1) *Cidades da Amostra Básica do Banco de Dados VARSUL (período de coleta- 1988 a 1994)*

- Rio Grande do Sul: Porto Alegre, Flores da Cunha, Panambi e São Borja.
- Santa Catarina: Florianópolis, Blumenau, Chapecó e Lages.
- Paraná: Curitiba, Londrina, Irati e Pato Branco.

Outras cidades e localidades foram incorporadas ao longo dos anos no Banco VARSUL, como contribuições de trabalhos de mestrandos e doutorandos, desenvolvidos na linha laboviana. Há atualmente a incorporação de entrevistas com dados de fala, resultantes de teses de doutorado:

(2) *Amostras Incorporadas ao Banco VARSUL, resultantes de Teses*

Rio Grande do Sul:

- a) amostra Bisol (1981): cidades de Taquara, Monte Bérico (distrito de Monte Bérico), Santana de Livramento, Porto Alegre⁴.
- b) amostra Amaral (2000): cidade de São José do Norte⁵.

³ Informações socioculturais breves de cada cidade, bem como pequenos trechos ilustrativos de fala de cada lugar, podem ser obtidos pelo site <www.varsul.org.br>.

⁴ A amostra Bisol (1981) segue os preceitos da amostra básica do VARSUL com a mesma distribuição de informantes em sexo, escolaridade e idade. Foi coletada em 1978, pela própria autora, para fins de trabalho de doutorado. Abrange vinte e quatro entrevistas de cada cidade. Os dados encontram-se armazenados na PUCRS, disponíveis em fitas cassetes e em transcrições.

⁵ A amostra Amaral (2002) comprehende 40 entrevistas coletadas na zona rural do município, em 1997, para fins de trabalho de tese de doutorado. Os informantes estão distribuídos por sexo, escolaridade (de 0 a 4 anos e até 4 anos).

Santa Catarina:

amostra Brescancini (2002): cidade de Florianópolis e distritos de Ribeirão da Ilha e Barra da Lagoa⁶.

Além de coletas de dados de teses, cedidas para o VARSUL, há o armazenamento na sede VARSUL/UFRGS de alguns dados de Porto Alegre do Projeto Norma Urbana Culta (NURC). Esses dados, coletados entre 1971 e 1978, segundo Menon (2008, p.337), encontram-se em rolos de fitas magnéticas, o que dificulta o acesso de pesquisadores. Também há o acréscimo de novas amostras, contemplando novas regiões, escolaridade, faixas etárias, como a fala de jovens de Porto Alegre (até 25 anos), de universitários e de quinze informantes recontatados do Projeto NURC, em 1999. Atualmente, a maior parte desses dados não se encontra disponível em meio eletrônico e não tem transcrição, mas projeta-se essa tarefa. Há também a incorporação dos dados de Flores da Cunha (RS) do Banco BDSer (Banco de Dados de Fala da Serra Gaúcha), cedidos pela Universidade de Caxias do Sul.

Nesses últimos anos, foram realizadas dezenas de trabalhos com os dados do Banco VARSUL, com divulgação nacional e internacional. Artigos, capítulos de livros, dissertações e teses trataram de expor diferentes estudos sobre variação e mudança linguística nas áreas de fonologia, morfologia, sintaxe, léxico, pragmática e discurso, em relação à fala do Sul do Brasil.

Os objetivos a que se propôs desde seu surgimento (BISOL; MENON; TASCA, 2008) foram e estão sendo, sem dúvida, cumpridos. Retomando-os, aqui, são os que seguem: oferecer subsídios para a descrição do português no País; dar condições para teste e desenvolvimento de teorias linguísticas; fornecer condições para formação de novos pesquisadores e subsídios para programas educacionais, promovendo o conhecimento e o respeito às variedades linguísticas.

No que diz respeito à descrição do português, estudos na área de Fonologia têm se encarregado de abranger diversos processos e contextos linguísticos, como forma de dar visibilidade, em termos gerais, ao panorama de fenômenos variáveis presentes na fala da região. Dezenas de trabalhos expõem, por meio de dados do Projeto VARSUL, resultados valiosos para a compreensão da língua portuguesa falada no Brasil, contemplando a teoria da Fonologia e a proposta da Sociolinguística Quantitativa.

⁶ A amostra Brescancini (2002) compõe-se por 100 informantes, sendo 48 referentes à região urbana do município e 52 às regiões interioranas.

Pesquisas com o Banco VARSUL na Área de Fonologia

A variação fonológica no sul do País tem sido persistentemente pesquisada no Banco VARSUL por um grupo de pesquisadores encabeçado por Leda Bisol. Com o objetivo de traçar um panorama dos trabalhos realizados pelo grupo de pesquisadores, ligados a diferentes universidades do Rio Grande do Sul, listamos a seguir os processos fonológicos examinados.⁷

(3) *Fenômenos Fonológicos Examinados com o Banco VARSUL*

- a) **alçamento de vogais pretônicas**, como as realizações m[i]nino, melh[u]ria, p[i]queno, por exemplo;
- b) **elevação de vogais postônicas final e não final**, como em bol[u], lev[i], núm[i]ro, côm[u]do;
- c) **apagamento da vogal postônica não final**, como em abób[Ø]ra, véspr[Ø]ra;
- d) **redução de ditongos nasais átonos**, como em home[Ø], órg[u];
- e) **epêntese vocálica em grupos consonantais**, como em p[i]neu, ad[e]vogado;
- f) **fenômenos de juntura vocabular**, como degeminação (menin[a]amada), elisão (verd[a]marelo) e ditongação (verd[ja]marelo);
- g) **variação de líquidas laterais**, como em sa[w]gado (salgado), mu[l]e (mulher);
- h) **variação de vibrantes** prevocálicas (rato), intervocálicas (carro) e pós-vocálicas (mar, carta);
- i) **variação de oclusivas dentais /t/ e /d/ diante de [i]**, como em palavras como *tia* e *dia*;
- j) **variação de /s/ posvocálico** em palavras como *paz* e *peste*, por exemplo;
- k) **monotongação de ditongos decrescentes**, como em fe[j]ra (feira), co[j]ro (courro);
- l) **ditongação em contextos seguidos por sibilantes**, como em fa[j]z (faz) e de[j]z (dez);
- m) **haplogolia sintática** em sequência de sílabas com /t/ e /d/ subjacentes: *resto do ano>res do ano; qualidade de vida>qualida de vida.*

Em relação à abrangência dos fenômenos fonológicos estudados, as descrições contemplam amostras de quase todas as cidades do Banco. De fato, a região sul pode, por meio das amostras do Banco, oferecer conhecimento sobre o comportamento variável de diversos processos fonológicos. A questão que se

⁷ Uma descrição mais abrangente desses estudos pode ser encontrada em Bisol e Brescancini (2002), Bisol e Collischonn (2009) e Revista de Letras (2005).

coloca, depois de tantos anos de investigação, é: o que os dados mostram sobre a fala do sul do País? É possível se fazer inferências sobre um levantamento exaustivo de casos em que um elemento variável poderia ter ocorrido, mas que não ocorreu? Labov (2008) defende que os dados numéricos permitem identificar implicações e fazer inferências, atendendo-se ao *Princípio de Explicabilidade*.

A importância dos dados é reconhecida para estudo teórico de língua. Por isso, um dos objetivos básicos da Sociolinguística Quantitativa é produzir resultados que podem ser generalizados no que diz respeito ao comportamento de uma extensa população. Para Bailey e Tillery (2004), generalização implica em *veracidade*, que pode ser obtida quando há resultados semelhantes em observações repetidas do mesmo fenômeno, e em *intersubjetividade*, quando há o mesmo resultado para um mesmo fenômeno observado por diferentes pesquisadores.

No caso do Banco Varsul, há pesquisas que podem ser comparadas por abrangerem mesmo fenômeno linguístico, na tentativa de se generalizar resultados. As regras variáveis examinadas distribuem-se por diferentes pesquisas, como pode se ver no quadro 1, sobre os trabalhos realizados em diferentes coletas, a respeito de determinados fenômenos fonológicos.

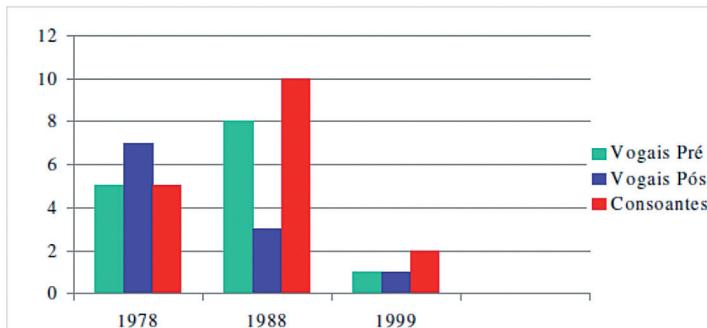
Quadro 1 - Pesquisas comparáveis com Dados do VARSUL por abrangerem o mesmo fenômeno

- **Harmonia Vocálica:** Bisol (1981), Battisti (1993), Schwindt (1995, 2002), Casagrande (2004), Silva (2012).
- **Alçamento de /e/ e de /o/ Postônico:** Schmitt (1987), Vieira (1994, 2002, 2009), Roveda (1998), Carniato (2000).
- **/r/ Posvocálico:** Monareta (1992, 1997, 2000, 2002, 2009), Monguilhott (1997), Gregis (2003), Pimentel (2003), Silveira (2010).
- **/l/ Posvocálico:** Quednau (1993), Tasca (1999, 2002), Espiga (2001), Collischonn e Costa (2005), Costa (2003), Collischonn e Quednau (2009).
- **Palatalização das Oclusivas Dentais /t/ e /d/ diante de [i]:** Bisol (1991), Almeida (2000), Kamianecky (2003), Pires (2007), Zamboni (2006), Battisti e Guzzo (2009).

Fonte: Elaboração própria.

O Banco VARSUL comprehende amostras coletadas, em termos gerais, em três períodos, que deram origem a diferentes pesquisas, como pode ser visto pela ilustração do gráfico 1. O ano de 1978 representa a amostra de Bisol (1981); o de 1998, o de coleta da amostra base do VARSUL; e o de 1999, o de coleta de universitários e de recontato do NURC.

Gráfico 1 - Distribuição de pesquisas por períodos de coletas de amostras do VARSUL



Fonte: Elaboração própria.

Segundo Bailey e Tillery (2004), para certificar a generalização de resultados, é preciso verificar a *origem dos dados* (época de coleta); a *constituição da amostra* (comunidades e estratificação) e os *procedimentos analíticos* utilizados na pesquisa (levantamento de dados, delimitação de variáveis dependentes e independentes, amálgamas, cruzamentos, etc.). Descrições, algumas vezes, podem ser divergentes em termos de resultados por questões metodológicas utilizadas nas pesquisas. As características sociais do entrevistador, por exemplo, ou a presença de outra pessoa na entrevista, podem provocar efeitos nos entrevistados que afetam os dados. A seleção da amostra e os meios de sua estratificação podem também acarretar diferenças de resultados, assim como as estratégias analíticas utilizadas em trabalhos sobre um mesmo fenômeno.

Em relação às amostras utilizadas, apesar de as localidades serem distintas, é possível de se testar a regularidade e intersubjetividade, fundamentais, segundo Bailey e Tillery (2004), para a generalização de dados.⁸ A regra variável de apagamento da vibrante pós-vocálica, por exemplo, ao ser examinada por diferentes pesquisadores e localidades do Banco VARSUL, e por estratégias analíticas, por vezes, também divergentes, pode ser generalizada por ter sido apontada, unanimemente em todos trabalhos, por ocorrer em final de verbos no infinitivo. O papel das variáveis sociais não é consenso entre as pesquisas, mas indica, preliminarmente, os homens como líderes do processo.

Outros fenômenos fonológicos podem ser verificados em termos de generalização de resultados. Há muitas pesquisas já realizadas que podem atestar a veracidade dos dados do Banco VARSUL. Eis, pois, uma preciosidade imensa para a descrição da fala do português brasileiro.

⁸ Propostas de exercícios comparativos entre pesquisas sobre mesmos fenômenos podem ser vistas em trabalhos sobre a discussão de efeitos de metodologia em regras variáveis com dados do VARSUL, em Brescancini (2011) e Monaretto (2008).

Considerações Finais

Neste artigo, mostramos inicialmente que as características do Banco VARSUL o inscrevem na primeira onda dos estudos sociolinguísticos, conforme a divisão proposta em Eckert (2012), e procuramos mostrar que a sua relevância não se esgota em pesquisas de cunho variação. Mostramos que *corpora* de língua falada em larga escala são, potencialmente, fonte útil de observação de dados de língua também para pesquisadores de outras áreas, especialmente das áreas de linguística de *corpus* e da geolinguística. Enfatizamos também a importância de dados de fala espontânea para análises fonológicas na atualidade, destacando algumas de suas vantagens, que podem complementar estudos feitos na base de fala de laboratório. Também abordamos as dificuldades inerentes à constituição e transcrição de Bancos de dados linguísticos de larga escala. O Banco de dados VARSUL é reapresentado, com exposição das amostras disponíveis para pesquisas em diferentes áreas, assim como um retrato dos trabalhos desenvolvidos na área de variação fonológica. Mostramos também que os levantamentos permitem atualmente elaborar pesquisas de cunho metametodológico, no sentido de testar a veracidade dos resultados, na linha do que propõem Bailey e Tillery (2004).

COLLISCHONN, G.; MONARETTO, V. de O. Database VARSUL: relevance of its features and scope of its results. *Alfa*, São Paulo, v.56, n.3, p.XX-XX, 2012.

- *ABSTRACT: The VARSUL Database (Urban Language Variation in the South of the Country) assembles speech data from social and culturally significant locations of the three southern states of Brazil. This article presents some considerations on the use, -both in the field of language variation and in several other fields of research, especially as regards to phonological phenomena - of data collections gathered according to the Labovian approach. The importance and advantages obtained by corpora analyses of spontaneous spoken language are addressed. We also discuss the possibilities that collections of sociolinguistic interviews provide for interface with other areas of research. Some features of the constitution of the VARSUL Database are presented, such as data collection methodology, criteria for informants selection, transcription protocols, as well as search devices developed for this Database. Finally, a survey of researches conducted on these data involving phonetic-phonological variables is presented to illustrate the scope of phenomena that have been studied and to show the possibility of generalization of results.*
- *KEYWORDS: VARSUL. Speech Database. Phonological Variation.*

REFERÊNCIAS

- ALMEIDA, M.A.B. de. *A variação das oclusivas dentais na comunidade bilíngüe de Flores da Cunha: uma análise quantitativa.* 2001. 106f. Dissertação (Mestrado em Lingüística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001.
- AMARAL, M. P. *As proparoxíticas: teoria e variação.* 2000. 222f. Tese (Doutorado em Lingüística Aplicada) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2000.
- BAILEY, G.; TILLERY, J. Some sources of divergent data in sociolinguistics. In: FOUGHT, C. *Sociolinguistic variation: critical reflections.* New York: Oxford University, 2004. p.11–30.
- BATTISTI, E. *Elevação das Vogais Médias Pretônicas em Silaba Inicial de Vocáculo na Fala Gaúcha.* Porto Alegre. 1993. 125f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1993.
- BATTISTI, E. ; GUZZO, N. B. Palatalização das oclusivas alveolares: o caso de Chapecó (SC). In: BISOL, Leda; COLLISCHONN, Gisela. (Org.). *Português do sul do Brasil: variação fonológica.* Porto Alegre: Ed. da PUCRS, 2009. v.1. p.114-149.
- BISOL, L. Palatalization and its variable restriction. *International Journal of Sociology of Language*, New York, n.89, p.107-124, 1991.
- _____. *Harmonia vocálica: uma regra variável.* 1981. 332f. Tese (Doutorado em Lingüística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- BISOL, L.; BRESCANCINI, C. (Org.). *Fonologia e variação: recortes do português brasileiro.* Porto Alegre: Ed. da PUCRS, 2002.
- BISOL, L.; COLLISCHONN, G. (Org.). *Português do sul do Brasil: variação fonológica.* Porto Alegre: Ed. da PUCRS, 2009. Disponível em: <<http://www.pucrs.br/edipucrs/portuguesdosuldobrasil.pdf>>. Acesso em: 25 set. 2012.
- BISOL, L.; MENON, O. P.; TASCA, M. VARSUL: um banco de dados. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Naro e a lingüística no Brasil: uma abordagem acadêmica.* Rio de Janeiro: Letras, 2008.
- BRANDÃO, S. F. Sociolinguística e geolinguística: uma perspectiva histórica. In: BARRETTO, M.; SALGADO, A. C. (Org.). *Sociolinguística no Brasil: uma contribuição de estudos sobre línguas/contato.* Rio de Janeiro: 7 Letras, 2009. p.88-92.
- BRESCANCINI, C. A elevação da vogal postônica final no Rio Grande do Sul: uma leitura à luz de Bailey e Tillary. In: COLLISCHONN, G.; BATTISTI, E. (Org.) *Língua e linguagem: perspectivas de investigação.* Pelotas: EDUCAT, 2011. p.227-251.

BRESCANCINI, C. *A fricativa palato-alveolar e sua complexidade*: uma regra variável. 2002. 362f. Tese (Doutorado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2002.

BRESCANCINI, C.; MONARETTO, V. Os róticos no sul do Brasil: panorama e generalizações. *SIGNUM*, Londrina, v.11, n.2, p.51-66, dez. 2008.

CALLOU, D. Quando dialetologia e a sociolinguística se encontram. *Estudos Linguísticos e Literários*, Salvador, n.41, p.29-48, jan./jun. 2010.

CARNIATO, M. C. *A neutralização das vogais postônicas finais na comunidade de Santa Vitória do Palmar*. 2000. 111p. Dissertação (Mestrado em Letras) - Universidade

Católica de Pelotas, Pelotas, 2000.

CASAGRANDE, G. P. B. *Harmonização vocálica*: análise variacionista em tempo real. 2004. 171f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2004.

CHILDS, B.; HERK, G.; THORBURN, J. Safe harbour: ethics and accessibility in sociolinguistic corpus building. *Corpus Linguistics and Linguistic Theory*, Ypsilanti, v.7, n.1, p.163–180, maio 2011.

COETZEE, A. W.; PATER, J. The place of variation in phonological theory. In: GOLDSMITH, J.; RIGGLE, J.; YU, A. C. L. (Org.). *The handbook of phonological theory*. 2. ed. Malden: Wiley-Blackwell, 2011. p.401-434.

COLLISCHONN, G.; COSTA, C. F. Ressilabação da lateral pós-vocálica final e sua limitação prosódica. *Letras de Hoje*, Porto Alegre, v.40, n.3, p.7-38, 2005.

COLLISCHONN, G.; QUEDNAU, L. As laterais variáveis da Região Sul. In: BISOL, L.; COLLISCHONN, G. (Org.). *Português do sul do Brasil*: variação fonológica. Porto Alegre: Ed. da PUCRS, 2009. p.154-176.

COSTA, C. F. Fonologia lexical e controvérsia neogramática: análise das regras de monotongação de /ow/ e vocalização de /ʌ/ no PB. 2003. 129f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

DURAND, J. On the scope of linguistics: data, intuitions, corpora. In: KAWAGUCHI, Y.; MINEGISHI, M.; DURAND, J. (Ed.). *Corpus analysis and variation in linguistics*. Amsterdam: John Benjamins, 2009. p.25-52.

ECKERT, P. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, Palo Alto, v.41, p.87–100, 2012. Disponível em: <<http://www.stanford.edu/~eckert/PDF/ThreeWaves.pdf>>. Acesso em: 30 mar. 2012.

ERNESTUS, M.; BAAYEN, H. Corpora and exemplars in phonology. In: GOLDSMITH, J.; RIGGLE, J.; YU, A. C. L. (Org.). *The handbook of phonological theory*. 2. ed. Malden: Wiley-Blackwell, 2011. p.374-400.

ESPIGA, J. *O português dos campos neutrais: um estudo sociolíngüístico da lateral posvocálica nos dialetos fronteiriços de Chuí e Santa Vitória do Palmar*. 2001. 154f. Tese (Doutorado em Linguística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2001.

GOLDSMITH, J.; RIGGLE, J.; YU, A. C. L. (Org.). *The handbook of phonological theory*. 2. ed. Malden: Wiley-Blackwell, 2011. p.401-434.

GREGIS, H. *O apagamento da vibrante pós-vocálica em Porto Alegre*. 2002. 111f.

Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

HINSKENS, F.; VAN HOUT, R.; WETZELS, L. Um balanço de dados e teoria no estudo da variação e da mudança fonológica. Tradução de Marisa Porto do Amaral. *Letras de Hoje*, Porto Alegre, v.35, n.1, p.7-46, mar. 2000.

KAMIANECKY, F. *Palatalização das oclusivas dentais nas comunidades de Porto Alegre e Florianópolis: uma análise quantitativa*. 2003. 114f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

KENDALL, T. Corpora from a sociolinguistic perspective. *RBLA*, Belo Horizonte, v.11, n.2, p.361-389, 2011.

KNIES, C.; COSTA, I. *Manual do usuário – Banco de Dados linguísticos VARSUL*. 1996. Disponível em: <<http://www.varsul.org.br/?modulo=secao&id=1>>. Acesso em: 30 mar. 2012.

LABOV, W. Quantitative reasoning in linguistics. *Linguistics* 563, Philadelphia, p.24, jan. 2008.

_____. *The social stratification of English in New York City*. Washington: Center of Applied Linguistics, 1966.

MENON, O. P. Uniformitarismo ou transmissão oral. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Naro e a linguística no Brasil: uma abordagem acadêmica*. Rio de Janeiro: Letras, 2008. p.337-351.

MENON, O. P.; FAGUNDES, E. D.; LOREGIAN-PENKAL, L. The Varsul Database. *Linguistik online*, Unitobler, v.38, n.2, p.2, 2009. Disponível em: <http://www.linguistik-online.de/38_09/menonEtAl.html>. Acesso em: 25 set. 2011.

MONARETTO, V. N. O. Descrição da vibrante no português do sul do Brasil. In: BISOL, L.; COLLISCHONN, G. (Org.). *Português do sul do Brasil: variação*

fonológica. Português do sul do Brasil: variação fonológica. Porto Alegre: Ed. da PUCRS, 2009. v.1. p.141-151.

_____. A vibrante pós-vocálica em Porto Alegre. In: BISOL, L.;

BRESCANCINI, C. *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: Ed. da PUCRS, 2002. p.253-268.

_____. O apagamento da vibrante pós-vocálica nas capitais da fala do sul do Brasil. *Letras de Hoje*, Porto Alegre, v.35, n.1, p.275-284, mar. 2000.

_____. *Um reestudo da vibrante: análise variacionista e fonológica*. 1997. 213f. Tese (Doutorado em Lingüística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1997.

_____. *A vibrante: representação e análise sociolinguística*. 1992. 104f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1992.

MONGUILHOTT, I. O. S. A vibrante em final de palavra na fala de Santa Catarina. In: ENCONTRO DO CÍRCULO DE ESTUDOS LINGUÍSTICOS DO SUL, 2., 1997,

Florianópolis. **Anais...** Florianópolis: UFSC: Celsul, 1997. CD-ROM.

MOTA, J. A. Sociolinguística e sociolinguística no Brasil: caminhos e encontros. In: BARRETTO, M.; SALGADO, A. C. (Org.). *Sociolinguística no Brasil: uma contribuição de estudos sobre línguas/contato*. Rio de Janeiro: 7 Letras, 2009. p.93-98.

PIMENTEL, R. *A variação lingüística do fonema /r/ na posição pós-vocálica em falantes da cidade de Porto Alegre*. 2003. 105f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

PIRES, L. B. A palatalização das oclusivas dentais em São Borja. *Revista Virtual de Estudos da Linguagem – ReVEL*. Edição especial n.1, p.1-23, 2007.

QUEDNAU, L. R. *A lateral pós-vocálica no português gaúcho: análise variacionista e representação não-linear*. 1993. 150f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.

REVISTA DE LETRAS. Porto Alegre: PUCRS, n.141, set. 2005.

ROVEDA, S. D. *Elevação da vogal média átona final em comunidades bilingues: Português e Italiano*. 1998. 87f. Dissertação (Mestrado em Lingüística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1988.

SCHMITT, C. J. *Redução vocálica postônica e estrutura silábica*. 1987. 139f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 1987.

SCHWINDT, L. C. A regra variável de harmonização vocálica no RS. In: BISOL, Leda; BRESCANCINI, Claudia (Org.). *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: Ed. da PUCRS, 2002. p.161-182.

_____. *A harmonia vocálica nos dialetos do sul do país: uma análise variacionista*. 1995. 76f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1995.

SILVA, M. E. da. *O alçamento das vogais médias pretônicas na fala de São José do Norte/RS: harmonia vocálica*. 2012. 144f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

SILVEIRA, G. da. *O Apagamento da vibrante na Fala do Sul do Brasil: sob ótica da palavra*. 2010. 129f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

TASCA, M. Variação e mudança no segmento lateral na coda silábica. In: BISOL, L.; BRESCANCINI, C. *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: Ed. da PUCRS, 2002. p.269-302.

_____. *A lateral em coda silábica no sul do Brasil*. 1999. 147f. Tese (Doutorado em Lingüística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1999.

VIEIRA, M. J. B. As vogais médias átonas nas três capitais do sul do país. In: BISOL, L.; COLLISCHONN, G. (Org.). *Português do sul do Brasil: variação fonológica. Português do Sul do Brasil: variação fonológica*. Porto Alegre: Ed. da PUCRS, 2009. v.1. p.50-72.

_____. As vogais médias postônicas: uma análise variacionista. In: BISOL, L.; BRESCANCINI, C. (Orgs.). *Fonologia e variação: recortes do português Brasileiro*. Porto Alegre: Ed. da PUCRS, 2002. p.127-159.

_____. *Neutralização das vogais médias postônicas*. 1994. 139f. Dissertação (Mestrado em Lingüística e Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1994.

ZAMBONI, A de P. *A palatalização das oclusivas dentais*. 2006. 152f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

Recebido em abril de 2012.

Aprovado em julho de 2012.

PROJETO ATLAS LINGUÍSTICO DO BRASIL: ANTECEDENTES E ESTÁGIO ATUAL

Suzana Alice CARDOSO*
Jacyra Andrade MOTA**

- RESUMO: Neste artigo, trata-se do Projeto Atlas Linguístico do Brasil, destacando-se os antecedentes que, na história da Geolinguística brasileira, explicam o seu surgimento, em 1996, por ocasião do Seminário *Caminhos e Perspectivas para a Geolinguística Brasileira*, realizado na Universidade Federal da Bahia, com a participação de pesquisadores da área, particularmente de autores de atlas publicados e de atlas em andamento. Apresentam-se informações sobre a sua estrutura, seus objetivos, como vem se desenvolvendo a pesquisa de campo, o estágio em que se encontra, atualmente - com 92,8% de localidades pesquisadas -, e resultados que as análises preliminares têm revelado. Destacam-se dois aspectos da sua organização: a realização de reuniões sistemáticas do Comitê Nacional que coordena o Projeto (a última das quais, a de número XXXIV, se realizou em fevereiro de 2012) e *workshops* para discussão, acompanhamento do trabalho e formação de inquiridores e de pesquisadores auxiliares. Do ponto de vista metodológico, o projeto se insere na geolinguística pluridimensional contemporânea, priorizando não só a variação diatópica, mas também a variação diageracional, diagenérica e diastrática.
- PALAVRAS-CHAVE: Atlas Linguístico do Brasil. Geolinguística. Diatopia. Variação linguística.

Introdução

O Projeto Atlas Linguístico do Brasil (Projeto ALiB) constitui-se na primeira tentativa, em nível nacional, de descrição do português brasileiro com base em dados coletados, *in loco*, nas diversas regiões geográficas, a partir da investigação em uma rede de pontos que se estende do Oiapoque (ponto 001) ao Chuí (ponto 250). Trata-se, portanto, de um projeto que se desenvolve no campo da variação linguística, mais especificamente no campo da Dialetologia e com base na Geolinguística, caminho metodológico que se ocupa da cartografia dos fatos de língua, cuja produção de maior relevância se consubstancia nos atlas linguísticos. Dá curso a uma tendência dos estudos da linguagem que, iniciada na Europa com

* UFBA – Universidade Federal da Bahia. Instituto de Letras – Departamento de Letras Vernáculas. Salvador – Bahia – Brasil. 40170-290 - suzalice@ufba.br

** UFBA – Universidade Federal da Bahia. Instituto de Letras – Departamento de Letras Vernáculas. Salvador – Bahia – Brasil. 40170-290 - jacymota@ufba.br

o *Atlas Linguistique de la France* (1902-1910), obra de Jules Gilliéron, se expande e se diversifica no que concerne à abrangência da área considerada — atlas regionais, nacionais, de família de línguas e de um continente.

O Projeto ALiB responde a um desejo expresso por filólogos e linguistas brasileiros e enfaticamente reivindicado por Nascentes (1958, p.7) que, nas suas *Bases para a elaboração do atlas linguístico do Brasil* (1958), assinala a importância dos atlas linguísticos: “[...] Depois de Gilliéron, o fundador da geografia linguística, nenhum verdadeiro estudioso da filologia nega o valor dos atlas linguísticos nem o seu caráter de indispensáveis.”

Com tal entendimento, apresenta-se, neste artigo, uma visão do Projeto ALiB, trazendo informações sobre os antecedentes, os objetivos e a metodologia seguida e fornecendo uma visão do estágio atual do seu desenvolvimento.

Antecedentes do projeto ALiB

A partir da segunda metade do século XX, os pesquisadores brasileiros começam a ressaltar a importância e a necessidade de realizar-se um atlas linguístico nacional, obra que colocaria o país junto a alguns países europeus, que já contavam com trabalhos dessa natureza.

Com relação à língua portuguesa, Silva Neto (1957), em “breve comunicação”, apresentada no 1º. *Colloquium Internacional de Estudos Luso-Brasileiros*, realizado, em 1950, em Washington, refere-se ao fato de ainda não se haver realizado, àquela época, o *Atlas Lingüístico-Etnográfico de Portugal e Ilhas*, o que considera “uma das mais prementes necessidades da Filologia Portuguêsa” e “obra de capitalíssima importância, que virá revolucionar os nossos conhecimentos de Filologia Portuguêsa e facilitar os nossos incipientes estudos acerca dos falares brasileiros”. Ressalta a urgência desse atlas, inspirando-se nas palavras de Meillet a respeito do Atlas Ítalo-Suíço, publicado entre 1928 e 1940, como se lê no trecho:

É preciso ter no espírito, como um aguilhão a apressar-nos, aquelas palavras de Antoine Meillet¹: “Qui a manié un Atlas comme celui-ci (Atlas Ítalo-Suíço) a l'impression qu'une linguistique que non dispose pas d'un pareil instrument est infirme.” (SILVA NETO, 1957, p.9).

Entendendo a justa valia de um atlas linguístico e numa demonstração clara do interesse pela realização do atlas linguístico do Brasil, o Governo Brasileiro, ao promulgar o Decreto nº. 30.643, de 20 de março de 1952, fixa como um dos principais objetivos da Comissão de Filologia da então criada Casa de Rui Barbosa

¹ Quem teve em suas mãos um atlas linguístico como este (Atlas Ítalo-Suíço) fica com a impressão de que uma linguística que não dispõe de um instrumento similar não tem valia.

a elaboração de um atlas linguístico do Brasil — “[...] sendo sua finalidade principal a elaboração do Atlas Lingüístico do Brasil” (Art. 3º) —, disposição que vem a ser regulamentada pela Portaria n.º 536, de 26 de maio do mesmo ano.

No entanto, apesar desse pronunciamento favorável por parte do governo brasileiro e do esforço de alguns pesquisadores (notadamente Serafim da Silva Neto, Celso Cunha, Antenor Nascentes e Nelson Rossi) em prol de obra de tal natureza, inúmeras dificuldades impediram a sua realização, o que levou Silva Neto e Celso Cunha a apresentarem, em 1957, no *III Colloquium de estudos luso-brasileiros*, em Lisboa, a proposta de Atlas Linguístico-Etnográfico do Brasil por regiões, com a divisão do país, de acordo com Silva Neto, em sete áreas culturais, como se lê no parágrafo inicial do texto:

É tarefa das mais urgentes a organização de um *Atlas Linguístico-Etnográfico do Brasil*, obra que irá tornar possível, pela primeira vez, uma visão de conjunto do desenvolvimento e do estado actual da língua portuguesa na América. (CUNHA; SILVA NETO, 1960, p.405).

Nascentes (1958, p.8), nas *Bases para a elaboração do Atlas Lingüístico do Brasil*, retoma essa proposta, especificando:

Temos de dividir o país em regiões, cada uma com o seu superintendente, ficando os superintendentes sob a direção de um técnico, havendo um órgão central supervisor, a Comissão de Filologia.

Começaremos pela divisão em regiões.

[...] o professor Serafim da Silva Neto [...] dividiu o nosso país nas seguintes áreas culturais:

1. Amazônia (Acre, Amazonas, Rio Branco, Pará e Amapá);
2. Nordeste Litoral (Maranhão, Piauí, Ceará, Rio Grande do Norte, Fernando de Noronha, Paraíba, Pernambuco, Alagoas e Sergipe);
3. Nordeste Mediterrâneo (Maranhão, Piauí, Ceará, Paraíba, Pernambuco, Alagoas e Bahia);
4. Planalto Central (Minas Gerais e sudoeste da Bahia);
5. Centro-Oeste (Goiás e Mato Grosso);
6. Centro-Leste (Rio de Janeiro, Distrito Federal, Espírito Santo, Sul de Minas e São Paulo);
7. Sul (Paraná, Santa Catarina e Rio Grande do Sul).

Devemos começar pelos atlas de regiões onde já tenham sido feitas sondagens.

Estas regiões são: a Amazônia, a Bahia e o Rio Grande do Sul.

No seu objetivo de delinear os caminhos da execução de obra como essa, apresenta, a seguir, como sugestão, “[...] um questionário típico, de caráter geral [...] e uma lista de pontos em que devem ser feitos os inquéritos” (NASCENTES, 1958, p.8).

A decisão de se elaborarem atlas regionais, embora não seguida dentro do esquema previsto por Nascentes, inspirou inúmeras pesquisas, dando origem a diversos atlas, o primeiro dos quais, o *Atlas Prévio dos Falares Baianos*² — idealizado por Nelson Rossi e realizado com a colaboração de Dinah Isensee e Carlota Ferreira, publicado em 1963 — inaugura a Geolinguística brasileira.

Em 1996, quando se inicia o Projeto *Atlas Linguístico do Brasil* (ALiB), já estavam publicados mais quatro atlas regionais — o *Esboço de um Atlas Lingüístico de Minas Gerais* (EALMG), em 1977; o *Atlas Lingüístico da Paraíba* (ALPB), em 1984, o *Atlas Lingüístico de Sergipe* (ALS), em 1987, e o *Atlas Lingüístico do Paraná* (ALPR), em 1994, encontrando-se, em andamento, o *Atlas Lingüístico-Etnográfico da Região Sul* (ALERS), que seria publicado em 2002³.

Projeto Atlas Linguístico do Brasil (ALiB)

Em consonância com esse percurso dos estudos dialetais brasileiros e por iniciativa de um grupo de pesquisadores ligados à área da Dialetologia/Sociolinguística, é retomada a ideia de elaboração de um atlas linguístico nacional, no tocante à língua portuguesa, por ocasião do Seminário *Caminhos e Perspectivas para a Geolinguística no Brasil*, ocorrido em novembro de 1996, no Instituto de Letras da Universidade Federal da Bahia, momento em que se constituiu um Comitê Nacional para coordená-lo.

Acatada a iniciativa, compôs-se o Comitê Nacional que, no seu primeiro momento, ficou constituído pelos autores dos atlas regionais até então publicados — Suzana Alice Cardoso (escolhida como Presidente do Comitê), Jacyra Andrade Mota (Diretora Executiva), ambas da UFBA, Maria do Socorro de Aragão (UFPB/UFCE), Vanderci de Andrade Aguilera (UEL) e Mário Roberto Lobuglio Zágari (UFMG) — e por um representante dos atlas em andamento — Walter Koch (UFRGS) — como Diretores Científicos. Em 2002, com a publicação do ALERS, Walter Koch passa a figurar como autor de atlas publicado, abrindo espaço para a inclusão de Aparecida Negri Isquierdo, como representante dos atlas em andamento. Com o falecimento de Walter Koch (em 2008) e de Mário Roberto Lobuglio Zágari (em 2010), passam a integrar também o Comitê Nacional Cleo Altenhofen (UFRGS), Felício Margotti (UFSC) e Ana Paula Rocha (UFOP).

² Confira: ROSSI; CALLOU; FERREIRA (1963).

³ Confira: RIBEIRO, et al. (1977); ARAGÃO; MENEZES (1984); FERREIRA, et al., (1987); AGUILERA (1994); KOCH, et al. (2002).

Objetivos

O Projeto ALiB, que vem ser assumido por esse grupo de dialetólogos, tem ao lado do objetivo geral de fornecer dados empíricos, sistematicamente recolhidos e analisados, para o conhecimento do português do Brasil, os seguintes objetivos específicos:

1. Descrever a realidade lingüística do Brasil, no que tange à língua portuguesa, com enfoque prioritário na identificação das diferenças diatópicas (fônicas, morfossintáticas, léxico-semânticas e prosódicas) consideradas na perspectiva da Geolinguística.
2. Estabelecer isoglossas com vistas a traçar a divisão dialetal do Brasil, tornando evidentes as diferenças regionais através de resultados cartografados em mapas lingüísticos e de estudos interpretativos de fenômenos considerados.
3. Examinar os dados coletados na perspectiva de sua interface com outros ramos do conhecimento — história, sociologia, antropologia, etc. — de modo a poder contribuir para fundamentar e definir posições teóricas sobre a natureza da implantação e desenvolvimento da língua portuguesa no Brasil.
4. Oferecer, aos interessados nos estudos lingüísticos, um considerável volume de dados que permita aos lexicógrafos aprimorarem os dicionários, ampliando o campo de informações; aos gramáticos atualizarem as informações com base na realidade documentada pela pesquisa empírica; aos autores de livros didáticos adequarem a sua produção à realidade cultural de cada região; aos professores aprofundar o conhecimento da realidade linguística, refletindo sobre as variantes de que se reveste a língua portuguesa no Brasil e, consequentemente, encontrando meios de, sem desprestigar os seus dialetos de origem, levar os estudantes ao domínio de uma variante tida como culta.
5. Contribuir para o entendimento da língua portuguesa no Brasil como instrumento social de comunicação diversificado, possuidor de várias normas de uso, mas dotado de uma unidade sistêmica (COMITÉ NACIONAL, 2001, p.16).

Definidos os objetivos, implementou-se a metodologia, que vai encontrar suas bases na Dialetologia Pluridimensional contemporânea, que, além de priorizar a dimensão diatópica, vai se fazer valer dos avanços que a nova realidade dos espaços e de sua ocupação imprimiram à metodologia dos estudos diaetais, considerando também outras dimensões, como a diageracional, diagenérica, diastrática, diafásica.

Metodologia

Assim, ao adotar os parâmetros da Geolinguística pluridimensional contemporânea, o Projeto ALiB filia-se a uma nova concepção nos estudos geolinguísticos, assim identificada por Thun (2000, p.408):

La nouvelle géolinguistique se caractérise par l'élargissement de son champ d'observation et par un travail en profondeur plus poussé. Elle passe de l'analyse de la superficie, constituée par la dimension diatopique, à celle de l'espace linguistique formé par la prise en considération de variables comme la dimension diastratique, diaphasique ou d'autres.

Elle n'emploie plus toute son énergie à la recherche du dialecte pur rural mais elle entre également dans les villes, elle analyse des langues régionales, focalise des situations de contact, questionne aussi des gens démographiquement mobiles.⁴

A “nouvelle géolinguistique”, a que se refere Thun, busca unir ao princípio da diatopia, pelo qual se vem regendo a Dialetologia desde os seus primórdios, princípios sociolinguísticos que, juntando-se ao primeiro, vão ensejar cartas linguísticas⁵ que permitem ao leitor não só saber **onde** se diz tal coisa, mas que **tipo de falante** — homem-mulher, jovem-velho, escolarizado-não escolarizado — é responsável por aquele enunciado.

Para alcançar a abrangência desejada, de acordo os parâmetros metodológicos da Geolinguística Pluridimensional, o Projeto ALiB insere, ao lado da pesquisa da variação diatópica, também a de outras dimensões variacionais, como a diageracional, a diagenérica, a diastrática, a diafásica e a diarreferencial.

Do ponto de vista diatópico, a rede, constituída de 250 pontos, distribui-se, geograficamente, por todo o território nacional, levando em conta a densidade demográfica de cada região e de cada Estado, com os ajustes necessários nos casos de áreas com densidade muito baixa, como a região norte e centro-oeste, ou demasiadamente elevada, como na região sudeste, principalmente em São Paulo. E, na seleção das localidades, além da distribuição espacial, consideram-se a importância de cada uma delas na região e no Estado e dados de sua formação, além de limites interestaduais e internacionais. Incluem-se cidades de grande e

⁴ A nova Geolinguística se caracteriza pela ampliação de seu campo de observação e por um trabalho de maior profundidade. Passa da análise da superfície, constituída pela dimensão diatópica, para a do espaço linguístico voltado para a consideração de outras dimensões como a diastrática e a diafásica.

Ela não emprega mais toda a sua energia na busca do dialeto rural puro, mas entra igualmente nas cidades, analisa os dialetos regionais, focaliza situações de contato, questiona também os grupos demograficamente móveis.

⁵ Mister se faz esclarecer que a Dialetologia monodimensional, nada obstante priorizar na cartografia os dados diatópicos, considera, na seleção de informantes, falantes diferenciados em idade, escolaridade e de sexos distintos.

médio porte, inclusive as capitais de Estado — à exceção do Distrito Federal e de Palmas, no Tocantins, cidades de formação relativamente recente nas quais ainda não se pode dispor de três gerações nascidas e criadas na área e com idade requerida para a seleção de informantes.

Para atender à variação social, os informantes, em número de quatro em cada ponto — exceto nas capitais de Estado, onde são oito —, estratificam-se quanto ao gênero — quinhentos e cinquenta homens e quinhentas e cinquenta mulheres —, à faixa etária — a primeira de 18 a 30 anos e a segunda, de 50 a 65 anos — e, nas capitais, ao nível de escolaridade — fundamental e universitário.

De acordo com a metodologia geolinguística, o informante deve ser natural da localidade, onde deve ter passado 2/3 de sua vida e filho de pessoas naturais da mesma área linguística. Os afastamentos não devem coincidir com os primeiros anos de vida do indivíduo (fase de aquisição da linguagem), nem com os anos imediatamente anteriores àquele em que se realiza o inquérito.

A variação diafásica ou estilística pode ser analisada a partir das respostas aos questionários previamente elaborados para a constituição do *corpus* (COMITÊ NACIONAL, 2001), confrontando-se os registros de fala mais monitorada, nas respostas monovocabulares, obtidas através de perguntas diretas, aos de elocuções mais espontâneas, surgidas por iniciativa do próprio informante, em comentários livres, ou por sugestão do inquiridor, na parte final da entrevista.

Para a análise do que se identifica como variação diarreferencial, incluem-se seis questões de natureza metalinguística, e se apresenta, finalizando a aplicação dos questionários, um pequeno texto para leitura, com vistas a ampliar o espectro das considerações diafásicas.

Os questionários, que passam a ser descritos, indicam um dos caminhos adotados — a diversidade na interrogação — para caminhar-se ao encontro da amplitude de enfoque na coleta de dados.

Questionários

Os questionários utilizados para a constituição do *corpus* do ALiB compreendem questões especificamente dirigidas para a obtenção de fatos fônicos, inclusive de natureza prosódica (questionário fonético-fonológico-QFF), variantes lexicais (questionário semântico-lexical-QSL) ou de estruturas morfossintáticas (questionário morfossintático-QMS), ao lado de questões de pragmática (QP), questões de natureza metalinguística (QM), temas para discurso semidirigido (TDS) e texto para leitura.

A estrutura das questões se distingue em função do objetivo de cada tipo de questionário, procurando-se obter, no QFF, o registro de uma variante fônica em vocábulos selecionados; no QSL, variantes lexicais características da localidade, a partir do fornecimento dos semas pertinentes; no QMS, variantes morfossintáticas, em questões diretamente dirigidas ao fato ou passíveis de serem flagradas na descrição de uma gravura ou no relato feito pelo informante, como nos exemplos a seguir:

QFF 055 – NOITE – Quando fica tudo escuro e as pessoas vão dormir é a __ ?

OFF 056 – DIA – E depois da noite, o que é que vem?

QSL 154 – PRESÉPIO – No Natal, monta-se um grupo de figuras representando o nascimento do Menino Jesus. Como chamam isso?

QSL 158 – PAPAGAIO DE PAPEL / PIPA – Como se chama aqui o brinquedo feito de varetas cobertas de papel que se empina no vento por meio de uma linha?

QMS 027 – COMIGO / COM EU / MAIS EU – Quando alguém não quer tomar café sozinho, diz para convidar outra pessoa: “Quer tomar café __?”

QMS 021 – COMPARATIVO DE GRANDE / PEQUENO – Essas duas casas (*mostrando um desenho de duas casas, uma grande e outra pequena*) têm o mesmo tamanho? A primeiro é como? E a segunda? Poderia comparar as duas casas em termos de tamanho?

QMS 046 – TER/HAVER em sentido existencial – Como era esta cidade antigamente, em termos de festas? Antigamente, esta cidade era mais desenvolvida? Por quê?

As quatro questões de pragmática procuram saber como um rapaz jovem, ao encontrar um objeto na rua, se dirige a outro jovem, do sexo masculino e do sexo feminino, ou a um indivíduo idoso, também dos dois sexos, para dizer-lhe que o objeto caiu de seu bolso.

As seis questões metalinguísticas indagam sobre o modo de falar local, as pessoas que, segundo o informante, “falam de modo diferente” na localidade e em outras áreas brasileiras, o falar de épocas anteriores, solicitando exemplos dessas variantes.

Nos temas para discurso semidirigido, o informante deve discorrer sobre uma experiência pessoal, comentar um programa de televisão, descrever sua atividade profissional ou relatar um fato de seu conhecimento.

Os inquéritos são totalmente gravados e cópias em DVDs são encaminhadas ao Arquivo Nacional do Projeto ALiB, sediado no Instituto de Letras da Universidade Federal da Bahia, onde se procede ao controle do feito e à catalogação dos registros. Resultam, todos eles, do trabalho de campo porque a documentação é integralmente coletada *in loco*, não importando as distâncias a serem vencidas.

A pesquisa de campo

A pesquisa de campo tem de maneira muito forte e incisiva dois veios: o da busca dos dados científicos e o do contato humano que caracteriza o encontro investigador-pessoa investigada. Para um trabalho que segue as pegadas da metodologia geolinguística, essas duas trilhas têm um sentido muito particular. De um lado, permitem registrar informações dos diferenciados usos de uma língua, usos muitas vezes marcados pelo aspecto arcaizante, ou inovador, ou denunciador de certa interferência de grupos imigrantes, ou ainda indicador das mudanças que começam a ser processadas naquela língua. De outro lado, propiciam ouvir de viva voz — e esse é um dos muitos pontos positivos —, *in loco*, no ambiente em que vive o falante, no seu próprio contexto, no calor humano do aparentemente neutro interrogatório linguístico, histórias, relatos, explicações que exibem o peso da sua inserção cultural e da sua inserção no mundo. Se uma entrevista aguça nos pesquisadores todas essas faces, 1.100 entrevistas feitas em todo o país — meta e ideário do Projeto ALiB —, despertarão muito mais.

Neste item, pois, faz-se uma reflexão sobre a pesquisa de campo desse projeto nacional, não para descrevê-la na sua inteireza, mas para pinçar pontos e aspectos que mostram como se tem realizado. Para isso, traz-se uma consideração de caráter geral que diz respeito ao processo de realização das entrevistas — a preparação para ida ao campo, as viagens, o contato com facilitadores do trabalho e a busca dos informantes —, a que se rotula de **ida ao campo**, e outra, de cunho particular, que se refere ao sentido acadêmico, social e humano da pesquisa dialetal, a que se denomina **no campo**.

A ida ao campo

A ida ao campo tem um antes, um durante e, necessariamente, um depois e, como afirmam Cardoso e Ferreira (1994, p.23):

A metodologia do trabalho dialetal — nada obstante reconhecer-se que o **fazer** é que **ensina**, o fazer é que **dita o método** — pode ter alguns passos definidos, no sentido de ordenar e proporcionar melhor rendimento da investigação e disciplinar a pesquisa.

A ida, pois, a um campo com 8.511.000 km², área do Brasil, tem particularidades muito próprias e no que se refere à preparação requer um conhecimento amplo, pelo menos, da realidade brasileira e uma disposição sem limites para observar, ver, sentir e reagir harmonicamente a cada situação com a qual se depara o investigador. Nessa preparação prévia à viagem estão, assim, o conhecimento pleno do instrumento de trabalho, os questionários; o estudo da área/localidades

que serão percorridas; a previsão de material a ser levado, desde aqueles de caráter acadêmico-científico aos de proteção contra insetos da região; o contato com os facilitadores do trabalho, as pessoas às quais se recorre como auxiliares indispensáveis na busca de informantes e que se tornam os grandes “avalizadores” da pertinência e da seriedade do destino do material a ser recolhido.

Nessa primeira parte, trabalha-se junto com dados atualizados do IBGE, com informações das Prefeituras locais, das Secretarias de Educação dos Municípios, das Igrejas que atuam na área e, sobretudo, com aquele filão que se chama a **rede de amizade e conhecimento** que se vem construindo por meio dos alunos, de diferentes procedências, que têm passado pelas mãos dos que compõem a Equipe ALiB por todo o Brasil. A partir daí, enfrenta-se o deslocamento para atingir a grande meta — o ponto da rede ALiB a ser inquirido.

Tem-se conhecimento das distâncias que unem e separam os diferentes rincões do país, tem-se, também, clareza da precariedade, particularmente em algumas regiões, das vias e dos meios de transporte, e tem-se, pelo menos, a ideia de que a alguns lugares se chega por terra, a outros pelo ar, a outros pela água, a outros se conjugam meios, e, ao atingir-se a terra, se faz o deslocamento interno a pé, de taxi e, hoje, nos tão atuais, pelo menos no interior, mototaxis. Nesse caminhar, tem-se que vencer as distâncias que separam o inquiridor dos pontos a serem inquiridos, as dificuldades decorrentes das intempéries, que fogem, obviamente, ao controle dos investigadores, e enfrentar, não poucas vezes, a travessia por áreas de litígio de terras e de situações de questionamentos indígenas. Para exemplificar as distâncias vencidas, uma vez que não cabe aqui e agora a contabilidade final dos quilômetros percorridos pelo projeto ALiB, traz-se o total atingido pelo Regional Mato Grosso do Sul, coordenado por Aparecida Negri Isquierdo, Diretora Científica do Projeto, responsável pela documentação em 26 pontos da rede que se distribuem por oito estados (Mato Grosso do Sul, Mato Grosso, Paraná, Roraima, Acre, Tocantins, Amazonas, Goiás, Espírito Santo): 51.150 km. Esse número significa: uma volta e mais um quarto de volta em torno do nosso planeta, na circunferência da terra na altura do Equador, ou seja, na sua parte mais ampla, cuja extensão é de 40.075 km.

Vencida a distância, a outra batalha diz respeito à localização de informante que preencha fielmente os critérios estabelecidos pela metodologia do Projeto e à obtenção do seu consentimento explícito de que concorda em ser entrevistado e em disponibilizar um pouco de seu tempo, do seu modo de pensar e, às vezes, até de seus sentimentos, para que se conheça mais e melhor o português do Brasil.

Pode parecer fácil, mas quem já foi ao campo sabe que nem sempre se consegue fazer convergirem todos os critérios numa pessoa, a pessoa ideal. Cumprida a missão, a volta do campo torna-se uma viagem prazerosa, com sabor de vitória.

No campo

Ao lado das questões relativas ao processamento dos dados coletados, necessária se faz uma consideração sobre o sentido social e humano da pesquisa dialetal.

Em primeiro lugar, destacar o momento da primeira entrevista do ALiB, feita em Quirinópolis, Goiás, por Vanderci de Andrade Aguilera, Diretora Científica do Projeto, e as lembranças que relata. Eram os idos de junho de 2001, exatamente o dia 1º, quando se faz a primeira entrevista das 1.100 estabelecidas, assim descrita:

Na manhã do dia 1º de junho de 2001, realizei a primeira entrevista do ALiB, junto ao R. (R.G.S.), de 29 anos, numa das salas do próprio Centro Social Urbano, em ambiente tranqüilo, poucas vezes interrompida pela intervenção de um ou outro funcionário desavisado de nossa presença naquele local. Mas, no decorrer da entrevista, eis que um caminhão de som ambulante passa a circular várias vezes pelos arredores, anunciando a grande liquidação da loja tal e... perturbando o meu trabalho. Essa entrevista foi identificada como 126/1, em que o numeral antes da barra significa o ponto linguístico Quirinópolis e o 1 após a barra identifica o informante masculino jovem. (AGUILERA, 2006, p.233).

Começavam, assim, as entrevistas do ALiB, e se pensava naquele então: “Foi feita a primeira, faltam 1.099!”. Do que relata a inquiridora, no hotel, no seu primeiro café da manhã em Quirinópolis, vê-se que manteve conversa com a copeira, com a faxineira, com o gerente do hotel e todos se propõem a ajudá-la, a começar por ensinar-lhe o caminho para chegar à Prefeitura. Mantém contato com a coordenadora do Centro Social que se “[...] empolga com a minha exposição e já começa a fazer uma lista dos possíveis informantes” (AGUILERA, 2006, p.233). Uma primeira lição nos vem daí: a compreensão que pessoas totalmente desvinculadas de um trabalho dessa natureza são capazes de manifestar e, em decorrência, a capacidade de assumir o papel de coadjuvantes.

Na linha da compreensão do projeto, outro inquérito vem para ilustração. Fazia-se a documentação do informante faixa etária II, portanto dos 50 anos para cima, homem de escolaridade fundamental. Foi entrevistado na sala da sua casa, casa simples nos arredores de Recife, ato presenciado, além dos partícipes — o informante, a inquiridora e a auxiliar de pesquisa —, pela sua esposa. A documentação dos dados corria muito bem, acabava-se de fazer a aplicação do QFF, do QSL e do QMS — como se sabe, uma espécie de pingue-pongue em que o inquiridor pergunta e o informante responde — e passava-se às questões semidirigidas, nas quais se pede ao informante o relato de fatos sobre temas orientados. Naquele exato momento, se desgruda da sala a esposa do informante que volta, incontinente, com um sanduíche quente e um suco de laranja, e

sussurra ao ouvido da inquiridora: “Estou vendo que agora é ele que fala mais e aí a senhora pode comer uma coisinha”. Compreensão da pesquisa, respeito ao método, entendimento perfeito da mecânica do interrogatório linguístico, uma solidariedade ilimitada ao trabalho dialetal. É preciso que se diga: acabado o inquérito, ao entrarem no táxi inquiridora e auxiliar, aquela não pode conter as lágrimas da emoção seguradas no final da gravação.

São essas e outras lições que o trabalho de campo permite ao pesquisador aprender ou reafirmar o seu aprendizado.

Estado atual

Passando por todas essas etapas, o Projeto Atlas Linguístico do Brasil (ALiB) encontra-se, atualmente: (i) na fase final de levantamento de dados para constituição do *corpus*, tendo já alcançado a documentação de 93.6% das 250 localidades previstas com 94.2% dos 1.100 informantes programados, e contando com 20 estados onde a pesquisa foi integralmente realizada e se encontra concluída; e (ii) na etapa de análise dos dados, especialmente das capitais de Estado, com vistas à publicação de um primeiro volume de cartas, acompanhado de estudos sobre os fatos linguísticos objeto da cartografia. Ao lado disso, procede-se à constituição de um Banco de Dados informatizado, que deverá permitir acesso aos dados e consulta ao atlas falante.

O Projeto ALiB chega ao estágio em que se encontra, presentemente, graças ao trabalho conjunto e coordenado do Comitê Nacional e das Equipes Regionais, estruturado sob a direção de membros do Comitê Nacional, que constituem o que já se convencionou chamar de *Família ALiB*. Para tanto, dois tipos de ação têm sido de fundamental importância: as reuniões do Comitê Nacional e os *workshops* nacionais.

As reuniões do Comitê Nacional, das quais a XXXV realizou-se em Salvador, em julho de 2012, por ocasião X *Workshop* do Projeto Atlas Linguístico do Brasil, têm um papel fundamental na condução do projeto. Por um lado, constituem-se, inicialmente, em momentos de montagem do seu arcabouço, definição dos princípios que devem ser seguidos, articulação de políticas que lhe deem sustentabilidade, avaliação do andamento, análise de dados e preparação de materiais para publicação.

De não menor importância para a sustentação do projeto, está a realização de *workshops* nacionais, o último dos quais, o décimo, aconteceu em Salvador, no mês de julho de 2012. Os *workshops* se constituem em momentos de fundamental importância para o desenvolvimento do projeto, considerando-se que nesses encontros discutem-se a metodologia do trabalho, o andamento da pesquisa,

os resultados de análises parciais e as atividades relacionadas à sustentação do trabalho; efetiva-se a discussão continuada da metodologia do trabalho, avaliando a sua aplicação e procedendo-se a reajustes, quando necessários e indicados pela experiência do trabalho de campo; são analisados os materiais reunidos e submetidos a estudo, buscando-se definir o quadro geral de fenômenos selecionados para informação sobre a realidade no território nacional.

Destaca-se que os três primeiros eventos destinaram-se à preparação de inquiridores para a execução dos 1.100 inquéritos linguísticos que deveriam recobrir 250 localidades, distribuídas de norte a sul do país. Convém assinalar que, sem extrapolar o número de participantes compatível com a natureza do trabalho a ser desenvolvido, tem-se facultado a participação nesses *workshops* a membros de equipes regionais envolvidas com a elaboração de atlas linguísticos de seus Estados e a pesquisadores interessados na temática.

À guisa de conclusão

O Atlas Linguístico do Brasil, proposto como se vem de apresentar, constitui-se em instrumento que nos revela a realidade do português brasileiro, mostra a diversidade de usos e a unidade a que é capaz de chegar a língua no Brasil. Em relação a resultados, os primeiros volumes do Atlas Linguístico do Brasil, referentes aos dados das capitais, deverão ser publicados em breve, espera-se que até 2012, com a apresentação de cartas fonéticas (inclusive prosódicas), morfossintáticas e lexicais, acompanhadas, como nos atlas ditos de 2^a geração, dos comentários a cada uma delas, juntamente com o volume de Introdução.

Em versão preliminar, as análises que constituirão o 1º volume de cartas foram apresentadas no IX *Workshop* do Projeto ALiB, realizado no Instituto de Letras da UFBA, em julho de 2011, algumas das quais se passa a fazer uma breve referência a seguir.

Entre os dados fonético-fonológicos, cita-se a distribuição do <s> em coda silábica, como em *estrada*, *casca*, *mesmo*, *desvio*, *três*, *arroz*. Quanto a esse fato, os dados do ALiB referentes às capitais mostram, quanto à variação entre alveolares e palatais, áreas não contíguas, assim distribuídas: (i) áreas de predominância de variantes palatais, representadas, principalmente, por Rio de Janeiro, Belém e Florianópolis, mas onde se incluem, com índices menos elevados e, em alguns casos, restritos ao contexto medial, capitais do Norte (Macapá, Manaus, Boa Vista, Rio Branco e Porto Velho), do Nordeste (Recife, Salvador, Aracaju e Natal) e do Centro-Oeste (Cuiabá); (ii) áreas em que se documentam índices mais elevados de alveolares, destacando-se Porto Alegre e Curitiba, no Sul, as três capitais do Sudeste (Belo Horizonte, São Paulo e Vitória) e, no Centro-Oeste, Goiânia e Campo

Grande. Ainda se incluem nesse grupo, embora com índices um pouco mais elevados de variantes palatais, no Nordeste, Teresina, Maceió, Fortaleza, João Pessoa e São Luís (MOTA; JESUS; EVANGELISTA, 2010).

No nível morfossintático, cita-se, para ilustração, o tratamento dado ao interlocutor na conversação em geral e numa mesma situação de discurso. Os dados revelam que há uma preferência acentuada pelo uso de *você*; no entanto observa-se a presença de áreas onde se registra a preferência ou a tendência para o uso de *tu*. Quanto a estas, destacam-se os seguintes aspectos: (i) uma área na qual o uso de *tu* se situa entre 77%-97% — Porto Alegre; (ii) uma outra área cuja presença de *tu* está entre 35%-55% — São Luís; e (iii) áreas nas quais se esboça uma presença de *tu* que se destaca das demais regiões, com índice entre 15%-34%, como se verifica nas capitais Rio Branco, Manaus, Macapá, Belém, Recife e Florianópolis.

No nível lexical, os estudos têm revelado uma significativa diversidade de usos. Por exemplo, a carta experimental *galinha d'angola* apresenta dez variantes para a identificação da ave: *angola*, *angolista*, *capote*, *catraia*, *cocar*, *galinhola*, *guiné*, *picote*, *saquê*, *tô fraco*. O levantamento e o estudo das formas demonstraram que, embora haja um número significativo de variantes, pode-se perceber uma distribuição uniforme para as regiões: (i) Norte: predominância da variante *picote*; (ii) Nordeste: predominância da variante *guiné*, coocorrendo com *galinha d'angola*; (iii) demais regiões: predominância de galinha d'angola.

Ao lado das análises dos dados das capitais com vistas à publicação do 1º volume de cartas do ALiB, contam-se para a divulgação de resultados preliminares com estudos mais restritos, discutidos em reuniões científicas e publicados em Anais e periódicos da área pelas equipes responsáveis pelo ALiB, e com a realização de dissertações de Mestrado e teses de Doutorado, realizados sob a orientação de membros do Comitê Nacional.

CARDOSO, S. A.; MOTA, J. A. The project linguistic atlas of Brazil: antecedents and current stage. *Alfa*, São Paulo, v.56, n.3, p.845-860, 2012.

- *ABSTRACT: This article deals with the Project Linguistic Atlas of Brazil, paying special attention to its antecedents which, in the history of Brazilian Geolinguistics, explain its appearance in 1996, by the time of the Ways and Perspectives for Brazilian Geolinguistics Seminar, held at the Federal University of Bahia, that counted on the participation of researchers of this field of knowledge, mainly authors of published and yet to be published atlases. The article presents information about the Project, its goals, how the field research has been done, the stage it is at, presently with 92,8% of localities already researched, and the results that preliminary analyses have shown. Two aspects of its organization are evident: the regular meetings of the National Committee that coordinates the Project (the last one, numbered XXXIV, took place in February, 2012) and workshops for discussion, work assistance and the*

formation of inquirers and assistant researchers. From the methodological point of view, the Project belongs to the contemporary pluridimensional geolinguistics, and it prioritizes not only the diatopic variation, but also the diagenerational, diageneric and diastratic variation.

■ KEYWORDS: *Linguistic Atlas of Brazil. Geolinguistics. Diatopic variation. Linguistic variation.*

REFERÊNCIAS

- AGUILERA, V. de A. De Quirinópolis-GO: a entrevista nº. 1 do ALiB. In: CARDOSO, S. A.; MOTA, J. A. (Org.). *Documentos 2: projeto Atlas Lingüístico do Brasil*. Salvador: Quarteto, 2006. p.229-238.
- _____. *Atlas Lingüístico do Paraná*. Curitiba: Imprensa Oficial, 1994. v. I.
- ARAGÃO, M. do S.; MENEZES, C. *Atlas Lingüístico da Paraíba I-II*. Brasília: Universidade Federal da Paraíba: CNPq, 1984.
- BRASIL. Decreto n.º 30.643, de 20 de março de 1952. Institui o Centro de Pesquisas da Casa de Rui Barbosa e dispõe sobre seu funcionamento. *Diário Oficial da União*, Brasília, 20 mar. 1952.
- CARDOSO, S.; FERREIRA, C. *A dialectologia no Brasil*. São Paulo: Contexto, 1994.
- COMITÊ NACIONAL. *Atlas Lingüístico do Brasil: questionários 2001*. Londrina: Ed. da UEL, 2001.
- _____. *Projeto Atlas Lingüístico do Brasil*. Salvador: Instituto de Letras, 1998. Não publicado.
- _____. *Projeto Atlas Lingüístico do Brasil*. Salvador: Instituto de Letras, 2003. Não publicado.
- CUNHA, C. F. da; SILVA NETO, S. da. *Atlas Linguístico-Etnográfico do Brasil*. In: COLÓQUIO INTERNACIONAL DE ESTUDOS LUSO-BRASILEIROS, 3., 1957, Lisboa. *Actas...* Lisboa: [S.l.], 1960. v. II. p.405-412.
- FERREIRA, C. et al. *Atlas Lingüístico de Sergipe*. Salvador: Universidade Federal da Bahia: Fundação Estadual de Cultura de Sergipe, 1987.
- GILLIÉRON, J.; EDMONT, E. *Atlas Linguistique de la France*. Paris: Honoré Champion, 1902-1910.
- KOCH, W.; KLASSMANN, M. S.; ALTENHOFEN, C. V. *Atlas Lingüístico Etnográfico da Região Sul do Brasil (ALERS)*. Porto Alegre: UFRGS; Florianópolis: UFSC; Curitiba: UFPR, 2002. 2 v.
- MOTA, J.; JESUS, C. de; EVANGELISTA, G. K. O <S> em coda silábica em capitais brasileiras: dados do projeto atlas linguístico do Brasil (ALiB). *Estudos linguísticos e literários*, Salvador, n.41, p.189-228, jan./jun.2010.

NASCENTES, A. *Bases para a elaboração do Atlas Lingüístico do Brasil*. Rio de Janeiro: MEC: Casa de Rui Barbosa, 1958. v. I.

RIBEIRO, J. et al. *Esboço de um atlas lingüístico de Minas Gerais*. Rio de Janeiro: Fundação Casa de Rui Barbosa: Universidade Federal de Juiz de Fora, 1977. v. 1.

ROSSI, N.; CALLOU, D.; FERREIRA, C. *Atlas prévio dos falares baianos*. Rio de Janeiro: Ministério de Educação e Cultura - Instituto Nacional do Livro, 1963.

SILVA NETO, S. da. *Guia para estudos dialectológicos*. Belém: Conselho Nacional de Pesquisas - Instituto Nacional de Pesquisas da Amazônia, 1957.

THUN, H. Atlas linguistique et variabilité – introduction à la table ronde. In: CONGRÈS INTERNATIONAL DE LINGUISTIQUE ET DE PHILOLOGIE ROMANES, 22., 1998, Bruxelles. *Actes...* Tübingen: Max Niemeyer, 2000. v. III. p.407-409.

Recebido em abril de 2012.

Aprovado em junho de 2012.

PARA UM ATLAS PLURIDIMENSIONAL: PESQUISAS E PESQUISADORES

Vanderci de Andrade AGUILERA *
Fabiane Cristina ALTINO **

- RESUMO: Este artigo propõe apresentar à comunidade científica um balanço das atividades do Projeto Atlas Linguístico do Brasil (ALiB) no que tange à formação de novos pesquisadores e, ao mesmo tempo, demonstrar o que a equipe do Comitê Central produziu na década deste início de século. Para isso, dividiu-se o texto em quatro seções: (i) a constituição da equipe de investigadores que trata da composição do Comitê Nacional e das atribuições de cada um de seus componentes; (ii) a produção discente em projetos de Iniciação Científica, desenvolvidos nas várias regionais do ALiB; (iii) a produção dos discentes da pós-graduação *lato* e *stricto sensu* sob a orientação dos diretores científicos e (iv) a produção dos coordenadores e diretores científicos. Os resultados apontam para a diversidade de enfoques teórico-metodológicos na abordagem de fenômenos de variação no nível da fonética, do léxico e da morfologia, entre outros, e indicam a possibilidade de associar alguns aspectos da variação linguística registrada pelo ALiB à proposta de Nascentes (1953).
- PALAVRAS-CHAVE: Atlas linguístico do Brasil. Composição do Comitê Nacional. Produção científica. Divisão dialetal do Brasil.

Apresentação

A inovação no campo da Geolinguística atual, iniciada há cerca de duas décadas por Thun (2000), diz respeito à importância e à necessidade de associar dados diastráticos, diagenéricos, diageracionais, diarreferenciais e diamésticos aos dados diatópicos da Geolinguística iniciada por Wenker e Gilliéron, no final do século XIX.

Nessa perspectiva, o Atlas Linguístico do Brasil (ALiB), projeto em andamento, nacional e interinstitucional, agrupa preceitos da Sociolinguística Variacionista à sua proposta de descrever a língua portuguesa falada no Brasil a partir de uma visão diatópica. Dessa forma, em se tratando de um atlas dessa envergadura,

* UEL – Universidade Estadual de Londrina. Departamento de Letras Vernáculas e Clássicas. Londrina – Paraná – Brasil. 86051-980 - vanderci@uel.br

** UEL – Universidade Estadual de Londrina. Departamento de Letras Vernáculas e Clássicas. Londrina – Paraná – Brasil. 86051-980 - fabiane_altino@uol.com.br

isto é, que se propõe mapear a diversidade do português falado em um país cujo território abrange mais de 8.500.000 km², o ALiB não poderia ser um projeto nem individual nem de curta duração. Na realidade, acredita-se que o material já coletado e aquele que está para ser coletado servirão de objeto de estudo para várias gerações de pesquisadores.

Sem perder de vista a necessidade de preparar novos pesquisadores para dar continuidade à proposta iniciada em 1996, durante o Seminário Nacional de Geolinguística, o Comitê Nacional que dirige o ALiB, conseguiu, ao longo desses 15 anos, aliar a experiência dos veteranos ao vigor e dinamismo dos mais jovens de tal sorte que contabiliza como resultado parcial do projeto uma gama diversificada de produtos. Sobre eles pretende-se discorrer com o objetivo de apresentar à comunidade científica um balanço das atividades do Projeto ALiB no que tange à formação de novos pesquisadores e, ao mesmo tempo, demonstrar o que a equipe do Comitê Central produziu na década deste início de século. Para isso, estruturou-se o texto em quatro seções: 1. a constituição da equipe de investigadores; 2. a produção no nível da Iniciação Científica; 3. a produção no nível da pós-graduação *lato* e *stricto sensu* e 4. a produção dos coordenadores e diretores científicos.

A constituição da equipe

No lançamento do Projeto ALiB, durante o Seminário Caminhos e Perspectivas para a Geolinguística no Brasil, realizado na cidade de Salvador – BA, em 1996, representantes das Universidades Federais da Bahia, da Paraíba, de Juiz de Fora, do Rio Grande do Sul e da Universidade Estadual de Londrina¹ constituíram um Comitê Nacional para dirigir os trabalhos voltados para verificar as diferenças léxico-semânticas, investigar os aspectos fonético-fonológicos, morfossintáticos e prosódicos no português brasileiro, na tentativa de delimitar os contornos, ainda que tênues, dos dialetos ou falares nas diferentes áreas do Brasil. A primeira preocupação foi incluir no Comitê pesquisadores que tivessem consciência da real dimensão de um trabalho dessa natureza para que o projeto não se perdesse no caminho. Assim, foram designados autores de atlas já publicados: Bahia, Minas Gerais, Paraíba, Sergipe e Paraná e um representante de atlas em andamento, na época, o Atlas Linguístico e Etnográfico da Região Sul (ALERS). Cada representante passou a ser designado Diretor científico com a atribuição de um número de estados para coordenar. Mais tarde com a publicação do ALERS, uma pesquisadora do Atlas Linguístico do Mato Grosso do Sul passou a integrar

¹ UFBA – Profa. Dra. Suzana Alice Marcelino da Silva Cardoso e Profa. Dra. Jacyra Andrade Mota; UFPB – Profa. Dra. Maria do Socorro Silva de Aragão; UFJF – Prof. Dr. Mário Roberto L. Zágari; UFRGS – Prof. Dr. Walter Koch; UEL – Profa. Dra. Vanderci de Andrade Aguilera.

o Comitê Nacional como representante de atlas em andamento. Como é natural, o tempo e as circunstâncias foram demandando alterações na equipe diretora.

Além da Diretora-Presidente, da Diretora Executiva e dos Diretores Científicos, o Projeto conta com a equipe de Entrevistadores, responsável pela aplicação dos Questionários *in loco*. Sobre a preparação dos entrevistadores é importante sinalizar para a exaustão do treinamento para a captação dos dados, que foi uma constante desde o *Workshop sobre técnicas e métodos de trabalho de campo para a preparação de inquiridores para o Atlas Lingüístico do Brasil – ALiB* – realizado em Salvador – Bahia, nos dias 06 a 10 de setembro de 1999 e uma das tópicas das discussões em diversas outras reuniões do Comitê Nacional. A preocupação com a forma adequada na condução do inquérito, os procedimentos metodológicos e os problemas arrolados por Labov (1972) tornaram-se tema constante. A constituição do *corpus* do ALiB, embora ainda em construção, já permite que sejam feitas análises sobre o português falado no Brasil. Até o final do mês de março de 2012 já estava concluída a coleta de dados em 232 localidades do Brasil, correspondendo a 92,8% do total da rede de pontos, o que representava 1.028 informantes entrevistados, ou 93,5% do total de inquéritos.

Outra equipe considerada de suma importância para o avanço dos trabalhos do ALiB é a dos transcritores e revisores das entrevistas gravadas no campo. Trata-se de um grupo de bolsistas de Iniciação Científica que tem a incumbência de executar e manter as cópias de segurança, seja nos bancos de dados regional e nacional, seja sob a forma de arquivo sonoro em ambas as instâncias. Além do registro e arquivamento das entrevistas e cópias, é a equipe que transcreve grafemática e foneticamente os dados coletados, sob a supervisão do Diretor Científico.

De posse do material coletado, transcrito e revisado, duas outras equipes entram em cena: a dos informáticos, na alimentação do Banco de dados central, e a dos cartógrafos, na elaboração das cartas experimentais e definitivas, ambas sediadas na UFBA.

Sobre a disseminação dos resultados parciais, inicialmente direcionou-se para a descrição e análise da metodologia adotada pelo projeto. Na sequência, voltou-se para o significativo volume de trabalhos sobre as entrevistas experimentais realizado por discentes de graduação (voluntários ou bolsistas da Iniciação Científica), de pós-graduação, e por docentes.

Passados 15 anos de instauração do projeto, outros trabalhos, agora com os dados dos inquéritos definitivos nas capitais e no interior, passam a compor o quadro da disseminação dos estudos realizados por esses pesquisadores. Somam-se a estes nove publicações:

i) inicialmente, Aguilera (1998), verificando que não havia uma obra específica que proporcionasse ao público uma visão do conjunto dos trabalhos na área, reúne

17 pesquisadores, nove dos quais participaram da elaboração de Atlas estaduais, e organiza a obra *A geolinguística no Brasil: caminhos e perspectivas*, que faz uma retomada da Dialetologia sob o método geolinguístico no país, além de apresentar dados e discussões sobre a metodologia adotada pelo projeto ALiB. Cardoso, no prefácio da obra, discorre:

A ideia de uma publicação mostrando os caminhos trilhados pela Geolinguística no Brasil (...) vem em boa hora. Primeiramente, porque um final de milênio sempre nos inspira e sugere um *balanço*, uma tomada de posição diante da história, uma visão crítica sobre o que se tem construído com projeção dos caminhos a percorrer. Em segundo lugar, porque a Dialetologia no Brasil vem passando por um processo de expansão, quando se amplia a produção de dissertações de Mestrado e teses de Doutorado na área e surgem novas incursões no campo da Geolinguística com a iniciativa de realização dos atlas linguísticos da Região Sul, de São Paulo, do Rio de Janeiro, de Mato Grosso do Sul, do Acre, do Pará, do Ceará e, mais recentemente, do Atlas Linguístico do Brasil. (CARDOSO, 1998, p.8).

ii) A publicação dos questionários do projeto é de suma importância para a Dialetologia nacional. Dos encontros promovidos pelo Comitê Nacional surgiram contribuições para o aperfeiçoamento dos questionários utilizados no trabalho de campo por meio dos três instrumentos de coleta de dados: o Questionário Fonético-Fonológico (QFF), composto de 159 questões; o Questionário Semântico-Lexical (QLS), com 202 questões, e o Questionário Morfossintático (QMS), com 49 perguntas. Duas versões preliminares (publicadas em 1998 e 2000) foram testadas em diversos inquéritos experimentais e analisadas nos encontros nacionais, chegando à versão definitiva publicada em 2001 (COMITÊ NACIONAL DO ALiB, 2001).

iii) Decorridos sete anos dessa publicação, e já com a edição esgotada, Aguilera retoma a obra de 1998 e recorre aos novos geolinguistas brasileiros para organizar uma nova versão (AGUILERA, 2005a) acreditando que, entre 1991 (ano da publicação de *A geografia linguística no Brasil*, por Brandão) e 2005, devido à ampliação do campo dos referidos estudos, seria preciso voltar os olhos para a atualização dos dados e oferecer à academia uma obra que espelhasse os avanços da geolinguística nos últimos anos. Assim, *A geolinguística no Brasil: trilhas seguidas e caminhos a percorrer* (AGUILERA, 2005) acrescenta àqueles textos mais uma dezena de outros para informar sobre os novos atlas publicados e em andamento, mas principalmente para mostrar as novas diretrizes da Geolinguística no Brasil que, seguindo a tendência mundial e certamente influenciados pela metodologia adotada para o ALiB, associa dados diatópicos a diastráticos na feitura de atlas.

(iv) e (v) A quarta e quinta obra que expõem os primeiros passos do ALiB são os *Documentos I* (AGUILERA; MOTA; MILANI, 2004) e *Documentos II* (MOTA; CARDOSO, 2006) que reúnem textos fundamentais para o entendimento do Projeto e para a compreensão da dimensão de que se reveste a pesquisa que desenvolve.

(vi) A sexta obra, *Dos sons às palavras: nas trilhas da língua portuguesa* (RIBEIRO; COSTA; CARDOSO, 2009), organizada em homenagem à Dr.^a Jacyra de Andrade Mota, Diretora Executiva do Projeto, trata da variação do português em seus diversos aspectos, focalizando, igualmente, resultados parciais do ALiB.

(vii) A sétima obra, *Pelos caminhos da Dialetologia e da Sociolinguística: entrelaçando saberes e vidas* (RAMOS; BEZERRA; ROCHA; OLIVEIRA; RAZKY, 2010), livro em homenagem à Dr.^a Maria do Socorro Aragão, Diretora Científica do ALiB, coordenadora regional em parte do Nordeste brasileiro, traz, além de estudos de renomados linguistas do Brasil, do Uruguai e de Portugal, artigos voltados para resultados parciais do ALiB.

(viii) A oitava publicação, *Geolinguística: tradição e modernidade*, da Diretora Presidente do ALiB (CARDOSO, 2010), retoma criticamente os procedimentos metodológicos adotados pelos vários atlas estaduais brasileiros e faz um balanço da Geolinguística no Brasil e em Portugal.

(ix) A Jornada dos Dicionários, evento tradicional e de alcance internacional, discute, com profundidade, temáticas relacionadas à Lexicografia e aos Dicionários. Em sua última edição, realizada em 2010, nos domínios da Universidade Federal da Bahia, foram três dias de discussões profícias entre pesquisadores brasileiros e franceses sobre as temáticas da construção dos dicionários e da relação das obras lexicográficas com os estudos geolinguísticos. A nona publicação do projeto, a obra *Os dicionários: fontes, métodos e novas tecnologias*, organizada por Cardoso e Mota, diretoras do ALiB, e Salah Mejri (2011), diretor do LDI (Laboratoire Lexiques, Dictionnaires, Informatique), da Universidade Paris 13, é o fruto destas discussões e os artigos aí contidos ilustram a preocupação em concatenar os estudos das diversas áreas em prol do tratamento informatizado do léxico do projeto ALiB.

Além desses, há dezenas de artigos inseridos em obras nacionais e estrangeiras escritos por pesquisadores do ALiB, sobre aspectos léxico-semântico-lexicográficos, fonético-fonológicos, morfossintáticos e metalinguísticos. Alguns deles abordam esses aspectos sob a ótica histórico-crítica da evolução do ALiB, conforme descrevemos nas seções seguintes.

O ALiB e a Iniciação Científica

A formação de novos pesquisadores foi uma das metas priorizadas desde os primeiros momentos de vigência do ALiB, seja pela necessidade de contar com colaboradores jovens para a gigantesca tarefa de transcrição², revisão e armazenamento de dados, seja pela urgência de criar e manter a mentalidade dialetológica da qual se ressentiram os pesquisadores que nos antecederam, como Nascentes e Silva Neto, que atribuíam à ausência dessa mentalidade a principal dificuldade para elaborar um atlas linguístico do país, não obstante outros empecilhos, como a grande extensão territorial e a ausência de apoio governamental.

Até o momento, registram-se 125 trabalhos realizados pelos alunos de IC orientados pelos diversos Diretores Científicos, que utilizaram como *corpus* os dados tanto das capitais, como do interior de um estado ou de uma região, e foram apresentados sob a forma de resumos, resumos ampliados, artigos em anais, relatórios parciais ou finais de pesquisa para os órgãos mantenedores da bolsa de IC, além da apresentação oral ou sob a forma de pôsteres em eventos de várias naturezas.

Desses 125 trabalhos de IC, 49 têm como objeto de estudo a variação lexical em torno das respostas dadas a questões de algumas áreas semânticas, como fenômenos atmosféricos (arco-íris, orvalho, sereno; nevoeiro, cerração, neblina), atividades agropastoris (tangerina, mexerica; penca de bananas; banana dupla, gêmeas; parte terminal da inflorescência da bananeira, umbigo, coração; mandioca, aipim), fauna (urubu, corvo; gambá), corpo humano (pálpebras, capela do olho; pomo-de-adão, gogó), convívio e comportamento social (pessoa pouco inteligente; assassino pago; marido enganado; prostituta; toco de cigarro), religião e crenças (diabo), jogos e diversões infantis (cambalhota, bolinha de gude; papagaio de papel, pipa; amarelinha), habitação (vaso sanitário), alimentação e cozinha.

Na sequência, 46 trabalhos versam sobre fenômenos fonéticos, como o /s/ em coda silábica interna ou externa, como em *festa*, *costas*, *fósforo*, *casca*; a monotongação de ei>e, de ai>a, de ou>o, como em *travesseiro*, *torneira*, *caixa*, *tesoura*, a ditongação de vogais finais seguidas de /s/, como em *luz*, *paz*, *arroz*; alternância de /v/ > /b/ e vice-versa, como em *varrer*>*barrer*, *assobio*>*assovio*, *travesseiro*>*trabissero*; o /r/ em início de sílaba ou palavra, como em *varrer*, *terreno*, *rosa*, *arroz*; a alternância de qüe/que, qüi/qui, qua/ca, entre outros. As regionais do Paraná e do Mato Grosso do Sul, em duas oportunidades, recolheram todo esse material e o publicaram em forma de CD para que não se perdesse o registro dessa importante etapa do Projeto ALiB na formação de novos pesquisadores (ISQUERDO; ALTINO; AGUILERA, 2009, 2012).

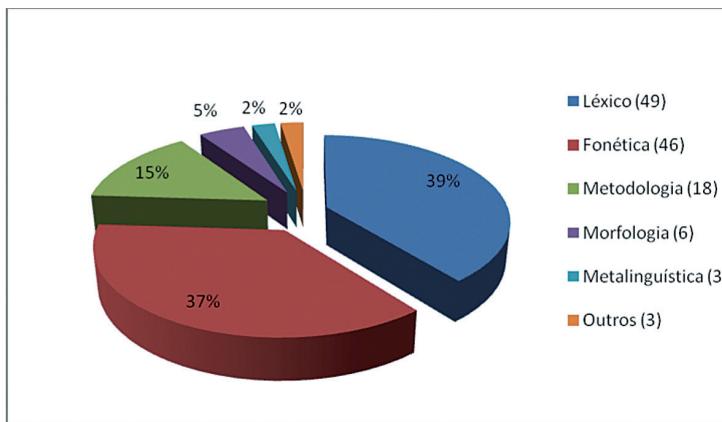
² É bom lembrar que, ao final, serão 1.100 entrevistas gravadas, cada qual com a duração média de 2h30min, transcritas e revisadas.

Os trabalhos sobre o Questionário Morfossintático, em número de seis, dizem respeito, sobretudo, à variação de *nós* e *a gente*; de *ter* e *haver*, no sentido de existir, junto aos falantes das capitais investigadas.

As discussões sobre a Metodologia adotada pelo ALiB resultaram em 18 trabalhos desenvolvidos por graduandos de Letras da UFBA, UFMS, UFPA, UFSC e UEL. Além desses, três deles exploraram as Questões de Metalinguística e os demais se ativeram à prosódia e à análise dos dados fonéticos e lexicais ligados à história social do povoamento das várias regiões do Brasil, com ênfase no papel disseminador da variante provavelmente usada pelos bandeirantes e, posteriormente, pelos tropeiros, que deu origem aos traços caipiras na fala do interior de São Paulo e do Paraná, do sul de Minas Gerais, Mato Grosso do Sul, parte do Mato Grosso, de Goiânia, de Santa Catarina e do Rio Grande do Sul.

O gráfico 1 mostra a distribuição dos trabalhos de IC por área analisada:

Gráfico 1 - Trabalhos de IC



Fonte: Textos construídos com dados do ALiB entre 2003 e 2011.

Muitos dos trabalhos desenvolvidos como projetos de IC estão auxiliando na discussão da divisão dialetal do português brasileiro proposta por Nascentes, em 1953, que divide o país em dois grupos de falares: o do Norte e o do Sul. Neste último estão inseridos os falares: Baiano, Fluminense, Mineiro e Sulista; enquanto naquele, os falares Amazônico e Nordestino. Para Nascentes havia, ainda, uma terceira área denominada de território incaracterístico.

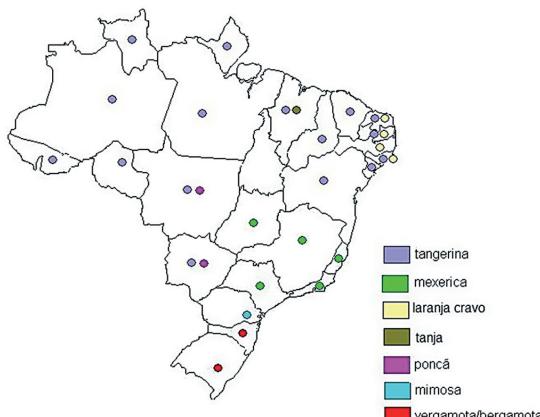
Como ilustração dos inúmeros trabalhos de IC, Romano, em co-autoria com Aguilera (2009), em artigo de natureza geolinguística e lexicológica, chega às seguintes conclusões acerca da distribuição diatópica das variantes, coletadas nas capitais, para a questão 039 sobre a *fruta da família da laranja que se descasca com as unhas*:

A análise das variantes para a questão 039, apresentadas nas 25 capitais, permite tecer as seguintes considerações:

- (i) *tangerina* é a forma mais produtiva e mais frequente que se distribui diatopicamente por quase todas as regiões, embora não apareça como a primeira resposta em algumas delas, sobretudo em capitais onde prevalecem as formas dialetais, como *tanja*, *bergamota*, *laranja-cravo* e *mimosa*;
- (ii) as variantes se distribuem de modo muito próximo ao que preconizou Nascentes (1958) ao propor a divisão dialetal do Brasil em falares do Norte e do Sul, de modo que *tangerina* se realiza como variante mais produtiva no Norte e *mexerica* no Sul/Sudeste/Centro-Oeste;
- (iii) nem todas as variantes regionais constam dos seis dicionários consultados, como ocorre com *laranja-cravo*, *poncã*, *mimosa* e *tanja*;
- (iv) os dicionários elegem a forma *laranja* + a variante como a forma padrão: *laranja-tangerina*, *laranja-mimosa*, *laranja-cravo* que depois, na linguagem oral e informal, se reduz ao segundo termo;
- (v) a maioria dos lexicógrafos faz referência às variantes regionais, embora não as localizem nas respectivas regiões;
- (vi) dicionários mais antigos não trazem as formas regionais, como *mimosa*, que vão aparecer somente nas edições mais recentes como Houaiss (2001) e Ferreira (2004);
- (vii) destaca-se a importância dos estudos geolinguísticos como fonte segura para os lexicógrafos. (ROMANO; AGUILERA, 2009, p.157).

Os autores ilustram com uma carta experimental (figura 1) que traz as formas predominantes em cada uma das 25 capitais.

Figura 1 - Carta experimental (preparada *ad-hoc*) sobre as variantes de *tangerina* registradas nas 25 capitais

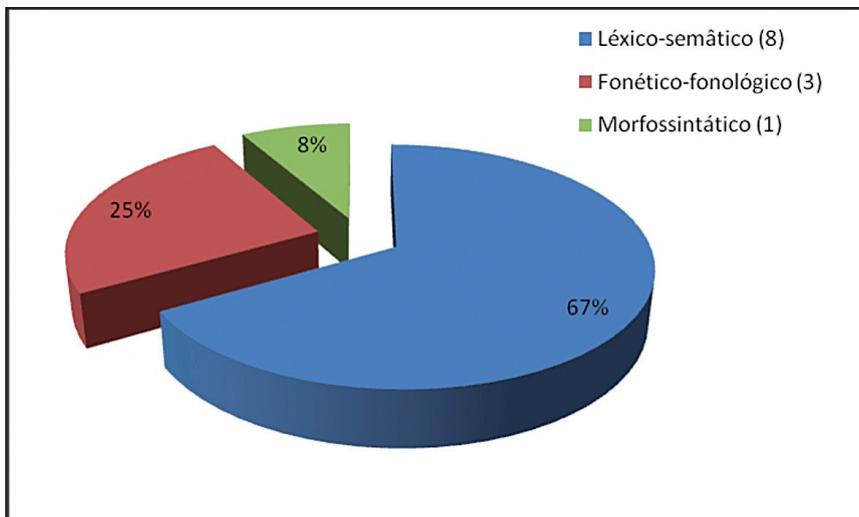


Fonte: Romano e Aguilera (2009).

A produção no nível da Pós-Graduação *Lato e Stricto Sensu*

Mais de uma dezena de trabalhos em nível de pós-graduação *stricto sensu* foram, ou estão sendo, desenvolvidos pelas diversas IES do país. São estudos no âmbito da fonética-fonologia, da morfossintaxe e do léxico, como sintetiza o gráfico 2:

Gráfico 2 - Síntese da produção na pós-graduação lato e stricto sensu com dados do ALiB



Fonte: Banco de dados do ALiB.

O levantamento feito a partir dos dados semântico-lexicais permite organizar a apresentação do universo vocabular para as respostas obtidas às questões, segundo alguns critérios, como: arcaísmos, brasileirismos, criações neológicas, variações no âmbito da diatopia, dialetismos, formação dos vocábulos, africanismos, vestígios das línguas indígenas; coletados nas entrevistas e que poderão ser apresentados nas cartas semântico-lexicais.

Dentre os estudos, em nível de Mestrado que tratam desta área, destacam-se os trabalhos apresentados na tabela 1:

Tabela 1 - Dissertações desenvolvidas com os dados do projeto ALiB

Título	Autor	Instituição	Ano
O campo léxico-semântico corpo humano no português do Brasil	Daiane Cunha dos Santos	UFBA	2008
Variantes lexicais no campo semântico da alimentação e cozinha no Brasil	Vanessa Yida	UEL	2009
Em busca de vestígios das línguas indígenas no português falado nas regiões Norte e Centro-Oeste do Brasil: um estudo geolinguístico	Daniela de Souza Silva Costa	UFMS	Início 2010
Um estudo do vocabulário rural na região Centro-Oeste: perspectiva geolinguística	Luciene Gomes Freitas	UFMS	Início 2010
Tabus linguísticos na fala dos habitantes das capitais brasileiras documentados pelo Projeto ALiB	Vanessa Cristina Martins Bencke	UFMS	Início 2010
Áreas dialetais do Brasil: em busca do falar amazônico	Danyelle de Almeida Saraiva	UFMS	Início 2010

Fonte: Dados extraídos do Currículo Lattes de cada autor.

Os estudos com o léxico, registrado nas entrevistas do projeto, rendem ainda dois trabalhos em nível de doutorado. São as teses *Áreas lexicais do português brasileiro*, recentemente defendida, e *O léxico brasileiro: confronto entre áreas segundo os dados do projeto ALiB*, ambas desenvolvidas na Universidade Federal da Bahia, sede do projeto, respectivamente por Márcia Verônica R. de Macêdo e Genivaldo da Conceição Oliveira. Ambas buscam delimitar as áreas dialetais dos falares brasileiros.

Os trabalhos sobre a fonética e fonologia do português brasileiro demonstram que este é um campo fértil para o desenvolvimento de dissertações e teses. A exemplo disto, temos os estudos em nível de mestrado e doutorado listados neste artigo.

A dissertação de Hélen Cristina da Silva, desenvolvida na Universidade Estadual de Londrina – UEL, sob o título de *O /R/ caipira no triângulo mineiro: um estudo dialetológico e de atitudes linguísticas* (2012), situa-se no âmbito da Dialetologia Pluridimensional e da Sociolinguística, e utilizou três *corpora* distintos: os dados do EALMG, os do ALiB coletados nas cidades de Campina Verde e Uberlândia e os resultantes da pesquisa de campo realizada pela mestrandona, com os objetivos de apresentar e discutir a distribuição espaço-social do /r/ caipira nas cidades de Campina Verde, Frutal, Ituiutaba, Iturama, Prata e Uberlândia, situadas no Triângulo Mineiro, analisada em tempo aparente e tempo real. Além de verificar a provável existência de mudança em curso refletida na diminuição da área do /r/ retroflexo, a mestrandona buscou averiguar aspectos extralingüísticos que possam influenciar nesse processo.

O estudo intitulado *Variação fonética dos ditongos /ey/ e /ow/ nas capitais do Norte Nordeste do Brasil*, desenvolvido por Williane Brasil dos Santos na

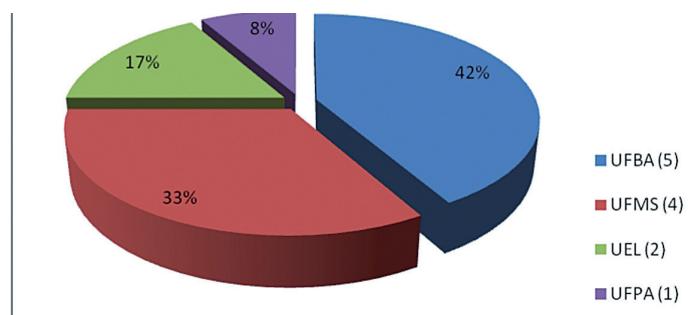
Universidade Federal do Pará – UFPA (2012) propôs verificar a manutenção e/ou apagamento dos ditongos na língua falada em algumas capitais do país, averiguando, por meio da metodologia da Sociolinguística e da Dialetologia Pluridimensional, quais aspectos interferem na realização/apagamento desses ditongos. Para essa pesquisa, foram utilizados os 120 inquéritos dos informantes (8 para cada localidade) das seguintes capitais: Macapá, Boa Vista, Manaus, Belém, Rio Branco e Porto Velho, na região Norte; e Salvador, São Luis, Teresina, Fortaleza, Natal, João Pessoa, Recife, Maceió e Aracaju, na região Nordeste.

A tese de doutoramento de Silvana Ribeiro, *Regiões dialetais do Brasil: em busca do falar baiano, na trilha de Antenor Nascentes*, recentemente defendida na Universidade Federal da Bahia, tem por objetivo examinar a vitalidade da divisão dialetal do Brasil proposta por Antenor Nascentes (1952) a partir da distribuição dos nomes para brinquedos e brincadeiras infantis. A tese de Ribeiro revisita o estudo de Nascentes e discute a pertinência dos limites estabelecidos em face da realidade linguística atual do país. Para a pesquisa foram utilizadas as 72 entrevistas do projeto ALiB nas 9 capitais do Nordeste.

No âmbito da morfossintaxe, o Projeto conta com a dissertação de mestrado *VOCÊ ou TU? Nordeste versus Sul: o tratamento do interlocutor no português do Brasil*, a partir dos dados do Projeto ALiB, desenvolvida na Universidade Federal da Bahia por Viviane Gomes de Deus. Trata-se de um estudo que buscou verificar a manutenção ou a substituição do pronome pessoal reto *tu* pelo pronome de tratamento *você*, além de analisar os aspectos linguísticos e/ou extralinguísticos que interferem na seleção dos pronomes pelos falantes.

Os estudos elencados acima demonstram a importância do projeto ALiB na formação acadêmica de novos pesquisadores nas diferentes áreas do saber linguístico. Fica evidente, ainda, a abrangência do projeto, que mobiliza pesquisadores de diversas regiões do país, como demonstra o gráfico 3.

Gráfico 3 - Produção de trabalhos acadêmicos na pós-graduação, com dados do ALiB, por IES



Fonte: Banco de dados do ALiB.

A atuação dos coordenadores regionais e diretores científicos

Como uma nova fase da Dialetologia brasileira, a retomada de um projeto de atlas linguístico nacional, em 1996, revitaliza a área e favorece a instalação de uma “mentalidade dialetológica” preconizada por Silva Neto (1957), já mencionada aqui, inaugurando uma nova etapa dos estudos geolinguísticos. Esse 4º período já proposto por vários pesquisadores³ confirma-se pelas inovações metodológicas adotadas, pela envergadura do projeto ALiB, pela disseminação da área de pesquisa junto aos órgãos de fomento e às instituições de ensino superior (IES). Diante deste cenário, vê-se a profícua produção acadêmica do Comitê Nacional. São inúmeros os estudos publicados nos eventos e revistas em nível nacional e internacional e capítulos de livros como demonstramos no quadro 1:

Quadro 1 - Produção acadêmica por tipo de publicação nacional e internacional

Tipo de publicação	Número de publicações	%
Nacional: revistas e anais	56	70
Capítulos de livros	12	15
Internacional: anais, revistas e capítulos de livros	12	15
Total	80	100

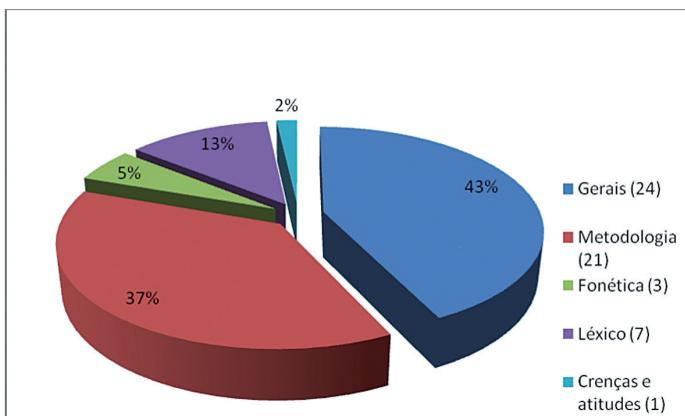
Fonte: Elaboração própria.

Quanto à temática dos trabalhos, inicialmente foi necessário explicitar e divulgar as intenções do grupo do ALiB quanto à retomada do projeto já preconizado em 1952, com a publicação do Decreto 30.643, que previa a elaboração de um atlas linguístico em que os falares brasileiros fossem identificados.

Nos anos iniciais do projeto ALiB, a discussão da metodologia era a constante preocupação e disseminar as bases de tão grande empreitada passou a ser a tônica dos trabalhos. Dos 56 estudos publicados em anais e revistas nacionais, pode-se observar a concentração de artigos que descrevem e explicam o funcionamento do projeto e, ainda, discutem a metodologia aplicada. É o que se observa no gráfico 4:

³ A periodização da Dialetologia brasileira foi apresentada por Nascentes em 1952, oportunidade em que o autor descreve as duas primeiras fases retomadas por Mota e Cardoso (2006).

Gráfico 4 - Distribuição por temas das produções em revistas e anais nacionais



Fonte: Banco de dados do ALiB.

Uma vez implantado o Projeto, a próxima etapa seria a de torná-lo conhecido do grande público acadêmico. Inicia-se, dessa forma, a etapa de divulgação das diretrizes que iriam nortear o trabalho de campo da equipe: como estava sendo planejada e concretizada a formação dos inquiridores, transcritores e revisores, basilares para a composição do banco de dados. Cabe aqui salientar o papel importante do ALiB na formação da mentalidade dialetológica e no estabelecimento e discussão das diretrizes norteadoras do trabalho do geolinguista no Brasil.

Não há como negar que o projeto ALiB é um divisor de águas para a Geolinguística no país. A partir da disseminação da metodologia, muitos trabalhos passaram a utilizar os seus questionários, os critérios de seleção de informantes e a rede de pontos, para citar os mais evidentes, em seus estudos. Esta popularização se dá por meio das publicações que tratam exaustivamente do projeto, exibem as bases e estratégias de ação. Como exemplos, listam-se os artigos de Mota e Cardoso (1998), Aguilera (2001), Isquierdo (2004), Aragão (2004) e Cardoso (2009a).

Uma das inovações do Projeto, e que recebeu atenção especial, foi a elaboração e posterior publicação do instrumento de pesquisa. A prerrogativa da fase de implementação recaiu sobre o instrumento de coleta de dados, elaborando-se questionários que pudessem refletir os objetivos do Projeto, indicando com clareza as noções cujas denominações se pretendem extrair dos informantes, os traços fonéticos e prosódicos recorrentes em cada região estudada. A posterior disseminação desta etapa metodológica foi campo frutífero para a elaboração de artigos. É o que se pode constatar nos artigos de Cardoso (1999), Mota e Cardoso (1999), Aguilera (1999) e Mota (1999), utilizados aqui como amostragem.

Cardoso (1999) discute a importância de ajustar o instrumento de coleta de dados à realidade cultural dos informantes. Devido à dimensão continental e à diversidade cultural bastante saliente no Brasil, haja vista os costumes do norte comparados aos do sul, por exemplo, se fez necessária uma reflexão cuidadosa sobre a pesquisa de campo para que se pudesse abranger todas as regiões. Da mesma forma, os trabalhos de Aguilera (1999) e Mota (1999) demonstram a preocupação da equipe em compor questionários adequados aos fenômenos pesquisados.

A produção científica dos coordenadores regionais e diretores do projeto ALiB abrange, também, a publicação de capítulos de livros, organizados em diversas instituições no país. Esses estudos versam sobre a metodologia adotada (50% dos trabalhos), a fonética (25%), o léxico (17%) e crenças e atitudes linguísticas (8%), e estão presentes nas publicações de oito estados brasileiros (Goiás, Paraná, São Paulo, Rio de Janeiro, Minas Gerais, Mato Grosso, Mato Grosso do Sul e Bahia). A título de exemplos, citamos: (i) sobre a metodologia: Aguilera (2002); (ii) sobre léxico: Cardoso (2003), Aguilera (2005b), Aguilera (2010); (iii) sobre fonética: Mota (2008); (iv) sobre crenças e atitudes: Aguilera (2008).

O projeto ALiB está presente, ainda, nas publicações de nível internacional. A coordenação regional e os diretores científicos, até o momento, publicaram 12 artigos em anais de congressos fora do país e em coletâneas que homenageiam personalidades da área. Esses artigos estão levando o Projeto para discussões em novas esferas do saber e possibilitando alianças para o melhor aproveitamento dos dados coletados. Sendo assim, artigos como “La dialectologie au Brésil. Aperçue historique e bilan actuel” (CARDOSO, 2002); “O Projeto ALiB – das origens ao momento atual” (CARDOSO, 2008); “A rede de pontos: uma reflexão sobre a história social das capitais de estado que integram o Projeto ALiB” (ISQUERDO, 2008) e “Revisitando os conceitos de rural e urbano no português do Brasil: contribuições do Projeto ALiB” (ISQUERDO, 2010) levam para a discussão dos colegas estrangeiros a metodologia empregada e já bastante veiculada nos meios acadêmicos nacionais.

Os capítulos de livros publicados no exterior encerram a exposição das publicações sobre o ALiB. São produções científicas que tratam da questão metodológica e trazem alguns dos primeiros resultados das análises. É o que se verifica nas publicações de Cardoso e Mota (2003), Mota e Souza (2009), Aguilera (2010) e Mota e Cardoso (2011).

Conclusões

Neste artigo procurou-se demonstrar, em linhas gerais, a trajetória percorrida pelo ALiB em seus 15 anos de vigência. Para isso, apresentou-se a constituição da equipe de pesquisadores que atuam nas diversas etapas e funções do Projeto,

seguindo-se os comentários sobre a ampla produção discente no que se refere aos trabalhos de IC, desenvolvidos sob a orientação dos Diretores Científicos que compõem o Comitê Nacional. Na terceira parte, relatam-se os trabalhos de pós-graduação *lato e stricto sensu*, também sob a orientação dos Diretores Científicos em cada Regional, que aprofundam os temas abordados pelos graduandos. Na quarta e última parte, discorre-se sobre os trabalhos dos docentes pesquisadores. Embora acredite-se não estar completa a relação das pesquisas realizadas nos diferentes estratos do Comitê, este é um retrato do que se fez para e pelo ALiB nesses 15 anos. A meta é, dentro de pouco tempo, trazer ao público os primeiros volumes do tão almejado Atlas Linguístico do Brasil.

AGUILERA, V. de A.; ALTINO, F. C. For a Multidimensional Atlas: Research and Researchers. *Alfa*, São Paulo, v.56, n.3, p.861-879, 2012.

■ *ABSTRACT:* This article aims at presenting to the scientific community a report on the activities of the Project *Atlas Linguístico do Brasil* (ALiB- Lingusitic Atlas of Brazil) concerning the constitution of new researchers and, at the same time, to demonstrate what the Central Committee team has produced since the beginning of this century. We divided the text into four sections: (i) the constitution of the team of investigators that deals with the formation of the National Committee and the attributions of each of its members; (ii) students' production in Scientific Initiation Projects developed in the several ALiB branches; (iii) the post-graduation students' production tutored by the scientific directors; and (iv) the production of the coordinators and scientific directors. The results indicate the diversity of theoretical-methodological perspectives in the approach of phenomena related to phonetic, lexical and morphological variation, among others, and also indicate the possibility of associating some aspects of linguistic variation found by ALiB to the proposal by Nascentes (1953).

■ *KEYWORDS:* *Linguistic Atlas of Brazil. Constitution of the National Committee. Scientific Production. Dialectal Division of Brazil.*

REFERÊNCIAS

AGUILERA, V.A. Arcaização, mudança e resistência lexicais em atlas linguísticos brasileiros: o rural e o urbano. In: DÖLL, C. et al. (Org.). *De arte grammatica: Festschrift für Eberhard Gärtner zu seinem 65. Geburtstag*. Frankfurt am Main: Valentia, 2010. p.19-32.

_____. Estudos dialetológicos no Paraná: caminhos e perspectivas. In: GÄRTNER, E.; HUNDT, C.; SCHÖNBERGER, A. (Org.). *Estudos de geolinguística do português americano*. Frankfurt am Main: Valentia, 2009.

_____. Crenças e atitudes lingüísticas: quem fala a língua brasileira? In: RONCARATI, C.; ABRAÇADO, J. (Org.). *Português brasileiro II: contato lingüístico, heterogeneidade e história*. Niterói: Ed. da UFF, 2008. p.311-333.

_____. (Org.). *A Geolinguística no Brasil*: trilhas seguidas, caminhos a percorrer. Londrina: Ed. da UEL, 2005a.

_____. Léxico regional, léxico rural ou vocabulário de curiosidade? Um olhar sobre aspectos lexicais de Cuiabá. In: ALMEIDA, M. M. S.; COX, M. I. P. (Org.). *Vozes cuiabanas: estudos lingüísticos em Mato Grosso*. Cuiabá: Cathedral Publicações, 2005b. p.115-136.

_____. Caminhos da dialetologia: os Atlas lingüísticos do Brasil. In: HENRIQUES, C. C.; PEREIRA, M. T. G. (Org.). *Língua e transdisciplinaridade: rumos, conexões, sentidos*. São Paulo: Contexto, 2002. p.77-92.

_____. Atlas Lingüísticos do Cone Sul: a caminhada do Atlas Lingüístico do Brasil. *Revista de Letras*, Curitiba, v.1, n.56, p.45-64, 2001.

_____. ALiB: considerações sobre o Questionário Léxico-semântico. A Cor das Letras, Feira de Santana, v.1, n.3, p.203-215, 1999.

_____. (Org.). *A geolinguística no Brasil*: caminhos e perspectivas. Londrina: Ed. da UEL, 1998.

AGUILERA, V. A; MOTA, J. A.; MILANI, G. A. L. *Documentos I: Projeto do Atlas Lingüístico do Brasil*. Salvador: Ed. da UFBA, 2004.

ARAGÃO, M. S. S. *Atlas Lingüístico do Brasil - Estado da Paraíba - ALiB-PB. Conceitos*, João Pessoa, v.5, p.42-48, 2004.

BRANDÃO, S. F. *A geografia linguística no Brasil*. São Paulo: Ática, 1991.

CARDOSO, S. A. M. *Geolinguística: tradição e modernidade*. São Paulo: Parábola, 2010.

_____. Projeto Atlas Linguístico do Brasil-Projeto ALiB: descrição e estágio atual. *Revistada ABRALIN*, Brasília, v.8, p.185-198, 2009a.

_____. Projeto Atlas Linguístico do Brasil (Projeto ALiB): uma resposta da geolinguística ao conhecimento do português brasileiro. In: GÄRTNER, E; SCHÖNBERGER, A. (Org.). *Estudos sobre o português brasileiro*. Frankfurt am Main: Valentia, 2009b. v. VIII. p.153-161.

_____. O Projeto ALiB: das origens ao momento atual. In: CONGRESO INTERNACIONAL DE LA ASOCIACIÓN DE LINGÜÍSTICA Y FILOGÍA DE AMÉRICA LATINA, 15., 2008, Montevidéu. **Anais**... Montevidéu: ALFAL, 2008. p.397-422.

_____. Arco-íris, estrela cadente e via láctea: que traçados fazem do português do Brasil? *Revista Portuguesa de Filologia*, Coimbra, v.24, n.1, p.77-100, 2003.

- _____. La dialectologie au Brésil. Aperçue historique et bilan actuel. *Géolinguistique*, Grenoble, v.2, p.197-229, 2002.
- _____. Geolinguística: convergências e divergências na coleta de dados. *Estudos Lingüísticos e Literários*, Salvador, v.23/24, p.143-153, 1999.
- _____. Prefácio. In: AGUILERA, V. A. (Org.). *A Geolinguística no Brasil: caminhos e perspectivas*. Londrina: Ed. da UEL, 1998. p.8.
- CARDOSO, S. A. M.; MOTA, J. A. A geolinguística no Brasil e o Projeto ALiB. *Bollettino dell' Atlante Linguistico Italiano*, Torino, III série, n.27, p.255-267, 2003.
- CARDOSO, S. A. M.; MOTA, J. A.; MEJRI, S. (Org.). *Os dicionários, fontes, métodos e novas tecnologias*. Salvador: Vento Leste, 2011.
- COMITÊ NACIONAL DO PROJETO ALiB. *Questionários 2001*. Londrina, Ed. da UEL, 2001.
- ISQUERDO, A. N. Revisitando os conceitos de rural e urbano no português do Brasil: contribuições do Projeto ALiB. In: CONGRES INTERNATIONAL DE LINGUISTIQUE ET DE PHILOLOGIE ROMANES, 25., 2010, Innsbruck. *Actes...* Berlim: De Gruyter: 2010. p.137-146, v. IV.
- _____. A rede de pontos: uma reflexão sobre a história social das capitais de estado que integram o Projeto ALiB. In: CONGRESO INTERNACIONAL DE LA ASOCIACIÓN DE LINGÜÍSTICA Y FILOLOGÍA DE AMÉRICA LATINA, 15., 2008, Montevideo. *Actas...* Montevideo: Diretoria da ALFAL, 2008, p.1-22, v. I.
- _____. Projeto ALiB: veredas. *Estudos Linguísticos*, Campinas, v.33, p.1-6, 2004.
- ISQUERDO, A. N; ALTINO, F. C.; AGUILERA, V. A. (Org.). *Atlas Linguístico do Brasil: descrevendo a língua, formando jovens pesquisadores*. Londrina: Ed. da UEL, 2012. v. 2.
- _____. *Atlas Linguístico do Brasil: descrevendo a língua, formando jovens pesquisadores*. Londrina: Ed. da UEL, 2009.
- LABOV, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- MOTA, J. A. Como fala o nordestino: a variação fônica nos dados do Projeto Atlas Lingüístico do Brasil. In: LIMA-HERNANDES, M. C. et al. (Org.). *A língua portuguesa no mundo*. São Paulo: FFLCH-USP, 2008.
- _____. Atlas Lingüístico do Brasil: o questionário fonético-fonológico e o questionário morfossintático. *A Cor das Letras*, Feira de Santana, v.3, p.123-137, 1999.

MOTA, J. A.; CARDOSO, S. A. Variantes lexicais e estratificação social no português brasileiro: uma visão com base nos dados do Projeto Atlas Linguístico do Brasil. In: SILVA, A. S. da; TORRES, A.; GONÇALVES, M. (Org.). *Línguas Pluricêntricas, variação linguística e dimensões sociocognitivas*. Braga: Universidade Católica Portuguesa, 2011.

_____. *Documentos II*: projeto do Atlas Lingüístico do Brasil. Salvador: Quarteto, 2006.

_____. O Atlas Linguístico Galego e o Projeto Atlas Lingüístico do Brasil: questões metodológicas. In: REI-DOVAL, G (Org.). *A linguística galega desde além mar*. Santiago de Compostela: Universidade de Santiago de Compostela, 1999.

_____. Um desafio: o Atlas Lingüístico do Brasil. *Estudos Lingüísticos e Literários*, Salvador, n.21-22, p.87-95, 1998.

MOTA, J. A.; SOUZA, L. N. Estudo de fatos fônicos em inquéritos experimentais do ALIB: o apagamento dos róticos em coda silábica. In: GÄRTNER, E.; HUNDT, C.; SCHÖNBERGER, A (Org.). *Estudos sobre o português brasileiro*. Frankfurt am Main: Valentia, 2009.

NASCENTES, A. *O linguajar carioca*. 2. ed. Rio de Janeiro: Organização Simões, 1953.

_____. Études dialectologiques au Brésil. ORBIS – *Bulletin Internacional de Documentation Linguistique*, Louvain, Tome I, n. 1, 1952.

RAMOS, C. M. A. et al. *Pelos caminhos da dialetologia e da sociolinguística: entrelaçando saberes e vidas*. São Luís: Ed. da UFMA, 2010.

RIBEIRO, S. S. C.; COSTA, S. B. B.; CARDOSO, S. A. M. (Org.). *Dos sons às palavras: nas trilhas da língua portuguesa*. Salvador: Ed. da UFBA, 2009.

ROMANO, V. P.; AGUILERA, V. A. A distribuição diatópica das variantes para tangerina: um estudo geo-sociolinguístico. In: ISQUERDO, A. N.; ALTINO, F. C.; AGUILERA, V.A. (Org.). *Atlas Lingüístico do Brasil*: descrevendo a língua, formando jovens pesquisadores. Londrina: Ed. da UEL, 2009. p.148-157.

SANTOS, W. B. dos. *Variação fonética dos ditongos /ey/ e /ow/ nas capitais do norte e nordeste do Brasil*. 280f. 2012. Dissertação (Mestrado em Lingüística) – Universidade Federal do Pará, Belém, 2012.

SILVA, H. C. *O /R/ caipira no triângulo mineiro*: um estudo dialetológico e de atitudes linguísticas. 186f. 2012. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Estadual de Londrina, Londrina, 2012.

SILVA NETO, S. da. *Guia para estudos dialetológicos*. Belém: Conselho Nacional de Pesquisa: Instituto Nacional da Amazônia, 1957.

THUN, H. La géographie linguistique romane à la fin du XX^e siècle. In: CONGRES INTERNATIONAL DE LINGUISTIQUE ET DE PHILOLOGIE ROMANES, 22., 1998, Bruxelles. *ACTES...* Tübingen: Niemeyer, 2000. p.367-388. v. III.

Recebido em março de 2012.

Aprovado em agosto de 2012.

DISCUTINDO A DIMENSÃO SOCIOLINGUÍSTICA DO PROJETO ALIB: UMA REFLEXÃO A PARTIR DO PERFIL DOS INFORMANTES

Aparecida Negri ISQUERDO*
Valter Pereira ROMANO**

- **RESUMO:** Este trabalho centra-se na discussão da dimensão pluridimensional dos estudos dialetológicos e tem como objetivos: i) discutir aspectos da dimensão sociolinguística do projeto ALIB (Atlas Linguístico do Brasil), com base no perfil estabelecido para a seleção dos informantes e ii) apresentar um panorama das características dos 200 informantes do ALIB entrevistados em 25 capitais brasileiras, discutindo tendências evidenciadas nesse universo de falantes da língua portuguesa contemporânea, especialmente em termos de nível de escolaridade, de contatos com os meios de comunicação - televisão e rádio -; de hábitos de leitura de jornal e de hábitos de diversões - cinema e teatro. Para tanto, recupera dados historiográficos relativos à disciplina Dialetologia, destacando aspectos que apontam, desde os seus primórdios, para a necessidade de se considerar, além da horizontalidade, aspectos relacionados à dimensão vertical dos estudos linguísticos, adotando-se também variáveis sociais como parâmetro para a descrição e análise da fala de grupos sociais fixados em zonas geográficas previamente delimitadas.
- **PALAVRAS-CHAVE:** Dialetologia pluridimensional. Atlas Linguístico. Brasil. Informantes.

Introdução

A Dialetologia nasceu sob o signo da monodimensionalidade à medida que tradicionalmente priorizava a diatopia como objeto de investigação, selecionando a fala de áreas rurais de pequeno porte e tomando como informantes para a recolha de dados dialetais, preferencialmente, homens de vida sedentária, idosos, analfabetos e nascidos e residentes na localidade, tendência essa inaugurada na segunda metade do século XIX, quando desponta a Dialetologia como desmembramento e também uma reação ao movimento dos neogramáticos, no que diz respeito à regularidade e à uniformidade das mudanças fonéticas que ocorrem no âmbito de cada língua. Iniciada com o trabalho de Georg Wenker, a partir de 1876, na Alemanha, e consolidada por Jules Gilliéron, com a publicação

* UFMS – Universidade Federal de Mato Grosso do Sul. Centro de Ciências Humanas e Sociais. Campo Grande – Mato Grosso do Sul – Brasil. 79070-900 - aparecida.isquierdo@gmail.com

** UEL – Universidade Estadual de Londrina. Centro de Letras e Ciências Humanas. Londrina – Paraná – Brasil. 86051-980 - valter.romano@hotmail.com.

do Atlas Linguístico da França (1902-1910), essa nova linha de investigação solidifica-se como uma nova orientação para o estudo da variação, no caso, tomando como contexto um espaço geográfico e buscando elucidar a relação entre língua e meio social: “[...] la encuesta francesa de Gilliéron ha influido enormemente, y debido a la eficacia del proyecto desde sus comienzos hasta su publicación y también a la calidad de sus resultados, se ha convertido en piedra de toque para cualquier encuesta posterior” (CHAMBERS; TRUDGILL, 1994, p.41). Os mesmos linguistas salientam a influência da obra de Gilliéron nos projetos de estudos dialetais subsequentes, como a pesquisa sobre os dialetos italianos da Itália e do sul da Suíça (Karll Jaberg e Jakob Jud) e, posteriormente, a contribuição de Jakob Jud e de Paul Scheurmeier na preparação de inquiridores para o *Atlas Linguístico dos Estados Unidos e do Canadá*, iniciado a partir de 1930. Destacam ainda os mesmos autores que “[...] lazos semejantes, bajo la forma de contacto personal y de consultas profesionales, unen Gilliéron directa o indirectamente con las encuestas dialectales nacionales de España, la Romania e Inglaterra, además de con varias encuestas regionales realizadas en Europa” (CHAMBERS; TRUDGILL, 1994, p.41-42). Essa orientação monodimensional expandiu-se para a América e tem caracterizado a grande maioria dos atlas linguísticos produzidos em território europeu e americano.

Todavia o embrião da pluridimensionalidade logo começa a surgir motivado pela própria dinâmica da língua. Assim, mesmo que os procedimentos metodológicos condicionassem descrições puramente diatópicas do material linguístico documentado em uma área espacial, esses mesmos dados davam e continuam dando mostras de que a fala, mesmo circunscrita a um espaço geográfico delimitado, pode evidenciar tanto zonas marcadas pela homogeneidade quanto áreas que apontam para traços de heterogeneidade linguísticos, culturais, étnicos. O *Atlas linguistique et Ethnographique de l'Italie et de la Suisse Méridionale* (AIS), por exemplo, “[...] proporciona para alguns pontos urbanos, os resultados dos famosos *relevés doublés*, feitos em duas camadas citadinas diferentes” (THUN, 2005, p.64), procedimento metodológico que aponta para o emergir de uma dimensão diastrática no âmbito dos estudos dialetais. Segundo o mesmo dialetólogo, também Griera (1923), no prefácio do *Atlas Linguistic de Catalunya*, primeiro atlas publicado na Península Ibérica, “esboça um programa completo de uma dialetologia pluridimensional que lamentavelmente – pensa o autor – nem ele nem algum atlas poderia realizar, porém, só uma série de monografias”. Registra Griera que “[...] l’ideal seria, evidentemente, de tener transcrita, de cada localitat, la llengua dels vells, la llengua de les personnes de mitjana edat i la llengua dels infants” (GRIERA, apud THUN, 2005, p.64). Sentencia ainda o mesmo dialetólogo alemão que, provavelmente, tenha sido:

[...] o afastamento da burguesia do mundo rural que causou nos dialetólogos acadêmicos do século passado [XX] a ilusão de uma uniformidade da vida e cultura dos camponeses à qual se amolda como síntese e peça de conservação de museu, o atlas linguístico estritamente diatópico (THUN, 2005, p.63-64).

A Dialetologia, seja assinalando para uma perspectiva pluridimensional, seja mantendo-se fiel ao preconizado pelos dialetólogos tradicionais foi ora ganhando espaço no âmbito das pesquisas sobre os falares regionais, os dialetos, ora recebendo severas críticas (sobretudo quanto ao número e ao perfil dos informantes selecionados para a coleta de dados) de linguistas estruturalistas e em especial dos sociolinguistas que implantaram uma metodologia calcada em variáveis sociais que interferem no uso da língua. Todavia, a despeito disso:

A Dialectología apresenta-se, no curso da história, como uma disciplina que assume por tarefa identificar, descrever e situar os diferentes usos em que uma língua se diversifica, conforme a sua distribuição espacial, sociocultural e cronológica. Dois aspectos fundamentais estão, pois, na sua gênese: o reconhecimento das diferenças ou das igualdades que a língua reflete e o estabelecimento das relações entre as diversas manifestações lingüísticas documentadas ou entre elas e a ausência de dados registrados, circunscritos a espaços e realidades pré-fixados (CARDOSO, 2002, p.1).

O advento da Sociolinguística a partir da década de 60 do século XX¹, sobretudo a vertente liderada por William Labov, de cunho variacionista, pautada no princípio das regras variáveis e com destaque para a dimensão vertical (social) dos estudos linguísticos, lançou questionamentos sobre as diretrizes da Dialetologia, no que tange à prioridade concedida à dimensão horizontal (geográfica) na análise da fala. Ramirez (2009, p.41), ao situar a Sociolinguística como disciplina que se ocupa do estudo dos sistemas linguísticos em seu contexto social, priorizando os centros urbanos que, por sua vez, concentram maiores índices de diversidades (lingüística e social), menciona quatro enfoques distintos da disciplina: “[...] 1) las relaciones entre lengua y sociedad; 2) los usos de la lengua en la sociedad; 3) la lengua en los contextos socioculturales y 4) la lengua y las redes sociales”.

O pesquisador espanhol estabelece ainda correlações entre as disciplinas Dialetologia e Sociolinguística, argumentando que “[...] la dialectología e la sociolinguística son disciplinas hermanas que estudian la lengua dentro de la sociedad. Las dos tradiciones, aun con distintos propósitos, se complementan con sus hallazgos lingüísticos y orientaciones metodológicas.” Destaca também

¹ Valendo-se do pensamento de Weinreich, Labov e Herzog apud Ramírez (2009, p.43) reitera que “[...] la sociolinguística tiene como uno de sus postulados básicos que la variación lingüística es inherente a la estructura del lenguaje”.

o fato de dados obtidos por meio de estudos orientados pelos parâmetros da Geografia Linguística terem sido utilizados por pesquisadores em trabalhos de sociolinguísticos, dentre outros os de William Labov que “[...]modificó los métodos de la dialectología para sus estudios del inglés en la isla de Martha's Vineyard (1963) y luego em Nueva York (1966)” (RAMÍREZ, 2009, p.46-47).

Na verdade, a partir dos anos 60, os estudos geolinguísticos tomaram uma nova orientação, pois passaram a inserir alguns pressupostos metodológicos da Sociolinguística, agregando ao fator diatópico variáveis sociais. Sobre esse assunto, Brandão (1991, p.26) faz as seguintes ponderações:

Assim, hoje, torna-se imperativo, por exemplo, incluir, entre os critérios de escolha dos indivíduos que servirão de informantes para a formação do *corpus* de um atlas linguístico, variáveis como idade, sexo, nível de instrução, ou mesmo situação socioeconómica, a fim de que se revelem ao máximo as particularidades do sistema dialetal focalizado e se possam melhor conhecer os condicionamentos socioculturais que presidem à distribuição geográfica dos fenômenos linguísticos.

Além das variáveis sociais, fez-se necessário incluir diferentes níveis de interlocução, como discursos livres com temas para que os informantes os desenvolvessem e textos de leitura, buscando, dessa forma, a variação diafásica.

Nota-se, pois, que, à medida que os estudos de Geografia Linguística começaram a se desenvolver, a metodologia da pesquisa se aprimorava, diferenciando-se conforme a época, os objetivos do trabalho e a região pesquisada. Segundo Almeida (2008, p.41), na América, por exemplo, Hans Kurath (1939) com seus colaboradores elaboraram o *Linguistic Atlas New of England* (LANE), principal resultado do *Atlas Linguístico dos Estados Unidos da América e Canadá*, inovando quanto ao critério de escolha dos informantes, somando ao aspecto geográfico, o social, o que evidencia relações entre os dados linguísticos e os de cunho sociocultural. Para tanto, os informantes foram selecionados considerando duas variáveis: o estrato social (escolaridade) e a faixa etária (jovens; meia idade; idosos). Sobre o LANE, Cardoso (2001) sintetiza:

Se com Wenker e no atlas da Alemanha se consagra a relevância da intercomparabilidade de dados e com o atlas de Gilliéron se assenta a importância da inquirição *in loco*, ao passo etnolinguístico dado por Jud e Jaberg no atlas italo-suíço se pode acrescentar outro avanço significativo trazido por Hans Kurath: a escolha de informantes que não apenas representassem os pontos constituintes da rede de localidades definida por região a partir de uma única faixa etária ou de um estrato social também único, mas que trouxessem também um aporte sociocultural e etário.

Um olhar retrospectivo para a história da Dialetologia e da Geolinguística/Geografia Linguística permite, pois, observar que, em diferentes continentes e em épocas distintas, houve o que Thun (2005) denomina de “apelos à pluridimensionalidade”, embora a maioria desses “apelos” tenha se limitado a introduzir determinadas variáveis sociais na metodologia do projeto e a sinalizar para a necessidade de o fenômeno linguístico ser examinado também segundo variáveis sociais, já que a par de fatores espaciais, condicionantes sociais interferem de forma significativa nos processos de variação e mudança linguísticas.

Ramírez (2009, p.47), por exemplo, destaca que o projeto do *Atlas Lingüístico de Hispanoamérica* (Alvar y Quilis, 1984), em desenvolvimento, e o estudo de Lope Blanch (1990) sobre o espanhol falado nos Estados Unidos “[...] tienen mucho que aportar para los estudios con enfoque sociolingüístico.” O *Atlas Linguístico da Hispanoamérica* foi projetado por Manuel Alvareem 1980 e teve como objetivo retratar o comportamento do espanhol em território americano. O principal resultado desse atlas é o que se refere ao espanhol do Sul dos EUA na área fronteiriça com o México (*El español en sur del Estados Unidos*) publicado em 2000.

Thun (2005, p.64-66), por seu turno, ao discutir as dimensões que têm orientado os caminhos percorridos pela Geografia Linguística, faz um inventário de atlas linguísticos produzidos na Europa e na América, destacando características desses atlas que evidenciam três quadros distintos: caminho monodimensional (tendência majoritária); caminho que introduz de forma direta ou indireta “apelos à pluridimensionalidade” e, por fim, “a pluridimensionalidade como programa”. Na primeira categoria de atlas, situa a Geografia Linguística do mundo ibero-peninsular e americano, ao mesmo tempo em que destacam dois fatores como justificativa para a predominância de atlas monodimensionais na Espanha: o “[...] poderoso modelo que forneceu Manuel Alvar com a sua impressionante série de atlas espanhóis regionais e ao florescer da sociolinguística que, com franco desdém para a sua co-disciplina, pretende privar a dialetologia de toda dimensão que não seja a areal²” (THUN, 2005, p.64). Já no que diz respeito aos atlas que, embora produzidos numa perspectiva da monodimensionalidade, antecipam características que apontam para a necessidade de revisões de posições teórico-metodológicas à luz do novo quadro da sociedade em geral, numa era de modernidade, de êxodos rurais e de ênfase na vida urbana que estão desenhando um novo panorama do mundo contemporâneo fortemente marcado pela industrialização, pelo avanço dos meios de comunicação de massa, pelos conflitos de classes, o que se reflete nas formas de uso da língua, intensificando o fenômeno da variação motivado por condicionantes sociais e desencadeando processos singulares de mudanças linguísticas. Enfim, a dicotomia entre o rural

² “A separação é a característica da geolinguística românica até hoje, pelo menos de sua corrente maior (RADTKE; THUN, apud THUN, 2005, p.64).

e o urbano sede lugar para a manifestação de um *continuum* rural/urbano/rural a que Bortoni-Ricardo (1985) denomina de dialeto *rurbano*. Nesse contexto de “[...] apelos à pluridimensionalidade”, Thun (2005) cita dois exemplos ilustrativos dessa tendência: o primeiro diz respeito ao programa delineado por Tomás Navarro Tomás (autor do modelo de estudo sobre o “*Español en Puerto Rico*”, de cunho monodimensional) que retoma as palavras de Griera (1923), com a finalidade de argumentar a favor de novos caminhos para as pesquisas dialetológicas sobre o espanhol:

Está por realizar el plan ideal que abarque el estudio del español de toda la isla en sus centros urbanos y en sus barrios campesinos y señale sus distintos niveles entre ancianos, adultos y jóvenes, entre personas instruidas, semicultas e iletradas, y entre hombres y mujeres [...] (NAVARRO apud THUN, 2005, p.65).

O segundo exemplo citado, o *Atlas Lingüístico y Etnográfico de Andalucía* (ALEA), é considerado por Thun (2005, p.65) como uma obra que “[...] apela indiretamente à pluridimensionalidade quando introduz, nos trinta pontos urbanos, as *encuestas múltiples* e distingue, nas notas marginais de alguns mapas, entre o modo de falar dos homens e das mulheres, entre faixas e níveis socioculturais”.

Por fim, no terceiro quadro delineado por Thun (2005), em que o autor apresenta a pluridimensionalidade como programa, o pesquisador considera o *Atlas Lingüístico de México* (ALM), de Lope Blanch (1991) como “potencialmente pluridimensional”, em que o autor sublinha que “[...] um atlas não pode substituir a detalhada análise sociolinguística, ‘pero sí puede dejar entrever las principales distinciones sociolingüísticas entre unas y otras clases de hablantes’ (1992)” (THUN, 2005, p.65).

Já na Península Ibérica, o dialetólogo alemão considera que apenas três trabalhos geolinguísticos deixam explícita a pluridimensionalidade: 1) a monografia *Untersuchungen zur Sprachvariation des Gaskognischen im Val d' Aran (Zentralpyrenäen)*, Tubinga, de O. Winkelmann (1989); 2) a tese de doutorado de F. Boller (1995), *Die Isoglossenstaffelung in der galicisch-portugiesisch-spanischen Kontaktzone und im LombadaÁliste-Grenzgebiet*, Kiel, e 3) o projeto *Atlas Lingüístico y Etnográfico de Castilla-La Mancha*, de F. Moreno Fernandez e P. García Mouton³ (THUN, 2005, p.66).

Os estudos geolinguísticos que vêm se desenvolvendo fora da Europa, principalmente, na América do Sul, agregam à Dialetologia conhecimentos advindos

³ O *Atlas Lingüístico y etnográfico de Castilla - La Mancha* configura-se como um atlas regional da língua espanhola falada na região central de Península Ibérica. Já está disponível online uma amostra de mapas contendo dados fonéticos, lexicais e gramaticais. O site fornece ainda informações acerca da metodologia adotada para a realização da pesquisa que está subsidiando a elaboração do Atlas que, quando concluído, pretende reunir em torno de 2.500 mapas a serem gerados automaticamente. Esse Atlas já conta com parte dos dados cartografados e está disponível <<http://www.lenguas.net/alecman/>>. Acesso em: 29.mar. 2012.

da Sociolinguística. Esses novos trabalhos passam a controlar sistematicamente, na coleta de dados, as variáveis sociais, tais como sexo, faixa etária, escolaridade. Consideram também na dimensão diatópica, além do informante topoestático (informantes fixos à localidade), o informante topodinâmico (pessoas oriundas de outras regiões). Já a dimensão diafásica é incluída no instrumento de coleta por meio de diferentes níveis de interlocução e passa-se a fazer uma Dialetologia urbana, considerando cidades de pequeno e de grande porte. Outra tendência também dos estudos pluridimensionais é incluir áreas de investigação onde há o contato de línguas a fim de “[...] documentar não somente a coexistência de língua e variedades, mas também a mútua influência que exercem umas sobre outras” (RADTKE; THUN, 1996, p.41).

No âmbito da América Latina⁴, Thun (2005) cita o *Atlas Lingüístico y Etnográfico de Chile*, planejado por G. Carrillo que, a exemplo do *Atlas Linguístico do México* (ALM), apresentou um projeto pluridimensional. Essa nova tendência de verificar o fenômeno da variação linguística veio constituir a Dialetologia pluridimensional, conceito desenvolvido por pesquisadores alemães como Edgar Radtke e Harald Thun, da Universidade de Kiel.

Ainda tratando do programa pluridimensional da Dialetologia, Thun (2005) apresenta os fundamentos teóricos e metodológicos do *Atlas Lingüístico de contatos do Rio da Prata*, um macroprojeto que abriga três projetos parciais de atlas: ADDU – *Atlas Lingüístico diatópico y diastrático del Uruguay*; ALRG - *Atlas Lingüístico Guaraní-Románico*; ALMRP - *Atlas Lingüístico de los Mennonitas Rioplatenses*⁵.

O *Atlas Diatópico y Diastrático del Uruguay* (ADDU), de Harald Thun e Adolfo Elizaincín, que visa a descrever, sob a perspectiva geolinguística, a região do Rio da Prata, teve dois volumes publicados em 2000 referentes à fonética e à morfossintaxe. Esse atlas orienta-se por diferentes dimensões e bem ilustra o modelo pluridimensional: dialingual (espanhol / português); diatópica (topoestático); diatópico-cinética (topostático / topodinâmico); diastrática (classe alta / classe baixa); diageracional (geração II / geração I); diassexual (mulheres / homens); diafásica (conversa livre; questionário; leitura); diarreferencial (fala “objetiva” / fala metalingüística) (THUN, 2005, p.71).

Já o *Atlas Lingüístico Guarani-românico* (ALGR), de Wolf Dietrich, Amídio Aquino e Harald Thun, abrange três países: Brasil (21 pontos), Argentina (35) e Paraguai (49). Dois tomos desse atlas foram publicados em 2002. O ALGR é pioneiro, na área da Geolinguística, no que se refere à descrição de línguas indígenas⁶.

⁴ O primeiro atlas nacional latino-americano que teve êxito foi o de Luiz Florez, o *Atlas linguístico-etnográfico da Colômbia* (ALEC), publicado entre os anos de 1981-1983. Este atlas teve sua coleta de dados iniciada em 1958 e é composto por seis volumes, totalizando 1.500 cartas linguísticas (ALMEIDA, 2008).

⁵ Projeto ainda em fase de desenvolvimento.

⁶ O ALIP - *Atlas Lingüístico de Ponta Porã* (REIS, 2006) inseriu na sua metodologia a dimensão dialingual, avançando, assim, na documentação das línguas em contato na área de fronteira entre o Brasil e o Paraguai

Nota-se, pois, que, se no continente europeu os estudos geolinguísticos ganharam principalmente o enfoque monodimensional, em solo americano, frente à nova realidade, esse método de trabalho tomou uma nova orientação, a pluridimensional, havendo, pois, uma adequação quanto aos instrumentos de coleta de dados, ao perfil dos informantes e à escolha das localidades.

Tomando por base o quesito dimensões da variação, Thun (2005) classifica os atlas linguísticos em *monodimensionais*, *bidimensionais* e *pluridimensionais*. O atlas linguístico de Wenker e o atlas de Gilliéron, por exemplo, têm como característica marcante a arealização dos dados linguísticos, ou seja, contemplam exclusivamente a variação linguística no espaço geográfico, a variação diatópica, logo, esses dois atlas são monodimensionais. Já o Atlas Ítalo-Suíço (AIS) contempla a variação diastrática por meio da arealidade ao incluir informantes de duas classes sociais distintas, podendo, pois, ser classificado como atlas bidimensional, por selecionar duas variáveis ou dimensões: a diatópica e a diastrática. Já os atlas linguísticos que contemplam mais de duas variáveis enquadraram-se na categoria dos atlas pluridimensionais, a tendência contemporânea dos estudos dialetais. Thun (2000, p.375) esclarece que:

Les premiers représentants d'un atlas linguistique systématiquement pluridimensionnel sont l'Atlas Linguistique de Sergipe [...] le microatlas aranéen contenu dans la monographie de Otto Wilkelmann (1989) et l'Atlas Lingüístico do Paraná [...] Les deux atlas brésiliens ajoutent à la dimension diatopique la variable diasexuelle que s'étend, en principe, à tous les points de réseau.

Nesse contexto, não é demais assinalar que, produtos do método da Geografia Linguística, as cartas que compõem os atlas linguísticos possibilitam uma ampla visão areal da variação interna dos dialetos estudados pelos métodos da Dialetologia à medida que visualizam a distribuição espacial de fenômenos linguísticos (sobretudo fonéticos e lexicais), permitindo identificar áreas conservadoras e inovadoras que, por sua vez, evidenciam o caráter dinâmico da língua e podem fornecer pistas para o estudo de determinado fenômeno, também na perspectiva histórica. A noção de isoglossas⁷ vinda dos fundamentos teóricos da Dialetologia muito bem ilustra os diferentes veios da variação dos sistemas linguísticos.

coberta pelo município brasileiro de Ponta Porã, à medida que mapeou variantes lexicais do português, do espanhol e do guarani para designar um mesmo conceito.

⁷ "El término isoglosa lo usó por primera vez J.G.A. Bielenstein, un dialectólogo letón, en 1892. Al parecer, acuñó el neologismo basando-se en el término meteorológico isotermo, la línea que divide dos lugares con la misma temperatura media. Isoglosa significa literalmente «igual lengua» (iso + glosa). Presumiblemente quiere expresar el hecho de que una línea trazada a través de una región mostrará dos áreas en cada una de las cuales coincide algún aspecto del uso lingüístico, pero que difieren una de la otra, los dialectólogos han usado de dos modos ligeramente diferentes la palabra, con el resultado de que la isoglosa se ha representado gráficamente de dos modos distintos; [...] se traza una línea entre los lugares de dos hablantes que muestren rasgos distintos [...] se traza una línea que enlaza los puntos de los hablantes que comparten el rasgo y una segunda línea

Isso porque as cartas linguísticas não são representações opacas da realidade, cujos registros se esgotam em si, mas se oferecem em sua dupla feição como registros da história dos homens. Trata-se, portanto, de uma dupla arealização: (i) *do espaço e das rotas ou trajetos entre os pontos*; (ii) *da sociedade e das projeções que os dados acionam no interior das variáveis extralingüísticas* (BUSSE, 2009, p.125, grifo nosso).

Em se tratando dos atlas pluridimensionais a cartografia dos dados representa uma opção metodológica particularmente desafiadora de maneira a garantir uma “[...] organização inteligente e racional do grande rol de dados coletados” (THUN apud CARDOSO, 2010, p.64). O *Atlas Diatópico y Diastrático del Uruguay - ADDU* (THUN; ELIZAINCÍN, 2000) é um exemplo de atlas que elaborou uma cartografia pluridimensional, permitindo o cruzamento das variáveis (diatópica, diageracional, diagenérica, diastrática e diafásica) e exibindo os resultados que mostram o comportamento de cada uma delas. Margotti (2008, p.09), por seu turno, ao tratar dos desafios metodológicos no âmbito da Geolinguística pluridimensional, conclui: “através dos mapas geolinguísticos pluridimensionais, mais simples ou mais complexos, é possível visualizar o grau e o modo de variação do fenômeno, fornecendo argumentos mais consistentes para as conclusões”. Nesse particular, Cardoso (2002, p.10-11) assim sintetiza a tendência contemporânea da cartografia nos estudos geolinguísticos:

A Dialectologia e especificamente os estudos geolinguísticos deixam de apresentar-se numa visão predominantemente diatópica e passam a exhibir, também cartograficamente, dados de natureza social. Isso vem mostrar, ainda, que a uniformidade diatópica pode sofrer desdobramentos sem, contudo, perder a sua inteireza e a sua integridade. E aí estaria “o novo”, a novidade não encontrada em tempos pretéritos nem amplamente explorada, como se verifica hoje — à diversidade de espaços físicos e geopolíticos junta-se a consideração dos parâmetros diagenérico, diageracional, diastrático, diafásico, dia-referencial ou nas especificações da diatopia, diatópico-topoestático, diatópico-topodinâmico, e de outros mais a que se pode e possa chegar.

A situação brasileira

Os estudos dialetológicos têm início no Brasil em 1920, com a obra *O Dialeto Caipira*, de Amadeu Amaral. Todavia a primeira manifestação que se pode caracterizar, numa visão ampla, de natureza dialetal sobre o português do Brasil, ocorre no ano de 1826, e deve-se a Domingos Borges de Barros, o Visconde de Pedra Branca que, a serviço de Adrien Balbi, anotou algumas peculiaridades

para unir a los que comparten el rasgo , cuando esos hablantes son contiguos al otro grupo, esto es, cuando se encuentran en la intersección entre las dos regiones.” (CHAMBERS; TRUDGILL, 1994, p.139-141).

do português brasileiro para integrar o *Atlas Ethnographique du Globe* (1924-1925). Callou (2010, p.33), ao discorrer sobre as inter-relações entre Dialetologia e Sociolinguística no Brasil, assevera:

Não se pode negar que o conhecimento sobre a realidade linguística brasileira teve início no âmbito da Dialetologia e atingiu o ápice com a Sociolinguística, mais especificamente, com a sociolinguística variacionista laboviana e os sofisticados métodos de análise estatística, um modelo de análise seguido em centenas de estudos na área.

Mota e Cardoso (2006, p.15-26), numa abordagem de caráter histórico, propõem uma divisão para os estudos dialetais no Brasil, formada por quatro fases. Para tanto, tomam como parâmetro a proposta de Nascentes (1952, 1953) que estabelece duas fases para os estudos dialetológicos no Brasil e a apresentada por Ferreira e Cardoso (1994) que, ampliando a periodicidade do dialetólogo brasileiro, dividem os estudos dialetais em três momentos. Essa periodicidade coincide com a estabelecida por Nascentes, no que diz respeito às duas primeiras fases, ou seja, a primeira recobre aproximadamente um século, estendendo-se de 1826-1920, período caracterizado por estudos acerca do léxico, com ênfase na elaboração de glossários e de dicionários, enquanto a segunda fase vai de 1920 a 1952 e caracteriza-se, basicamente, como os primeiros passos da Dialetologia brasileira, em que se destacam os seguintes trabalhos monográficos: *O Dialeto Caipira*, de Amadeu Amaral (1920), *O Linguajar Carioca*, de Antenor Nascentes (1922/53), *A língua do nordeste*, de Mário Marroquim (1934). Ainda nessa fase, desenvolveram-se trabalhos que podem ser classificados em quatro grupos diferenciados conforme o enfoque: glossários regionais; obras de caráter geral que tratam de questões mais abrangentes da língua portuguesa no Brasil; estudos monográficos de caráter regional que abordam, particularmente, fenômenos específicos de dada região e estudos acerca da contribuição do elemento africano à língua⁸. Já a terceira fase estabelecida por Ferreira e Cardoso (1994) tem seu início em 20 de março de 1952, data em que o governo brasileiro publicou o Decreto 30.643 que definia as finalidades da Comissão de Filologia da Casa de Rui Barbosa, dentre as quais, a primordial era a elaboração de Atlas linguístico do Brasil. Nessa fase, segundo Mota e Cardoso (2006, p.19):

[...] não só se sedimenta a mentalidade dialectológica preconizada por Serafim da Silva Neto no seu *Guia para estudos dialectológicos no Brasil* (1957), mas, também, se dá início aos trabalhos de natureza geolinguística, com as pesquisas para o *Atlas Prévio dos Falares*

⁸ Confira: Ferreira e Cardoso (1994).

Baianos e com a sua publicação em 1963. A esse primeiro atlas, seguem-se-lhe outros de caráter regional o que consolida, na opinião das autoras [Ferreira e Cardoso] uma terceira fase, cuja ênfase está no desenvolvimento dos estudos geolinguísticos.

As mesmas autoras, considerando o novo momento vivido pela Dialetologia no Brasil, a partir de 1996, com a retomada do projeto de um Atlas Linguístico no Brasil, propõem uma quarta fase para os estudos dialetológicos no Brasil, apresentando como justificativas: i) o crescimento de pesquisas nessa área, em sua maioria, direta ou indiretamente, relacionadas ao projeto ALiB; ii) as discussões acerca de questões metodológicas nas reuniões nacionais do Comitê Nacional⁹ de coordenação do projeto ALiB e nos *workshops* nacionais organizados pelo projeto; iii) o aumento do número de atlas regionais e início de outros projetos de atlas regionais (MOTA; CARDOSO, 2006, p.20-21).

Nesse contexto, merece ainda destaque o papel do projeto ALiB nas pesquisas linguísticas no Brasil, representando um divisor de águas nas pesquisas na área de Dialetologia e Geolinguística, não só pelo que representa em termos de empreendimento como um projeto nacional de grandes dimensões, mas também por ter estabelecido e consolidado uma diretriz metodológica que tem orientado as pesquisas na área. Há que se considerar ainda a questão da sintonia com as pesquisas dialetológicas e geolinguísticas na contemporaneidade, considerando-se que:

Do ponto de vista metodológico, essa nova fase coincide com a incorporação dos princípios implementados pela Sociolinguística a partir da década de 60 do século passado, abandonando-se a visão monodimensional – monoestrática, monogeracional, monogenérica, monofásica, etc. – que predominou na geolinguística hoje rotulada de “tradicional” (MOTA; CARDOSO, 2006, p.21).

Recentemente, Romano (2012, p.67-68), considerando especificamente a produção dos atlas linguísticos regionais no Brasil e o advento do projeto ALiB, argumenta que os estudos geolinguísticos em território brasileiro podem ser divididos em dois momentos. O primeiro que tem como marco inicial a publicação do primeiro atlas linguístico estadual, o APFB, em 1963, e se estende até 1996 – ano em que o projeto ALiB deu início às suas atividades com o esforço de dialetólogos de todo o País. Os atlas linguísticos desenvolvidos ou projetados neste período de 33 anos seguem pressupostos metodológicos anteriores aos do ALiB, portanto,

⁹ Por ocasião do Seminário Nacional Caminhos e Perspectivas para a Geolinguística no Brasil, realizado em Salvador, Bahia, foi constituído o Comitê Nacional do ALiB, formado por autores de atlas regionais e um representante de atlas em andamento. A estrutura desse Comitê comporta um presidente, uma diretora executiva e os diretores científicos. Atualmente integram o Comitê Nacional do ALiB pesquisadores de 08 universidades públicas brasileiras: UFBA, UFMS, UFPA, UFSC, UFRGS, UFTO, UFPB, UFC, sob a presidência de Suzana Alice Marcelino Cardoso (UFBA).

divergem quanto à metodologia, ora priorizando a diatopia (EALMG, APFB, ALERS), ora incluindo outras dimensões da variação, como a diagenérica (ALS, ALPR). São trabalhos pioneiros no campo da Geolinguística brasileira que revelam o amadurecimento de uma mentalidade dialetológica já preconizada em 1957 por Silva Neto. Alguns desses atlas foram divulgados e publicados após 1996, porém, pertencem ao primeiro momento devido às suas particularidades metodológicas. O segundo momento da Geolinguística brasileira vai de 1996 até os dias atuais. Os trabalhos desse período já apresentam influência dos pressupostos metodológicos do ALiB, caminhando para uma visão pluridimensional da variação¹⁰.

Projeto ALiB: um atlas nacional pluridimensional

A falta de uniformidade metodológica dos atlas estaduais produzidos no Brasil entre 1963 e 1996 não permite, em alguns casos, estudos comparativos e contrastivos que revelem uma visão geral da língua falada no Brasil. Dessa forma, em novembro de 1996, por ocasião do Seminário *Caminhos e Perspectivas para a Geolinguística no Brasil* ocorrido em Salvador/BA, dialetólogos de todo o país, assessorados pelo pesquisador francês Prof. Michel Contini, da Universidade de Grenoble, reuniram-se e decidiram assumir o grande desafio de elaborar um atlas nacional, retomando a ideia dos eminentes linguistas brasileiros do início da segunda metade do século XX. Esse atlas possuiria, pois, uma metodologia única e revelaria a realidade linguística brasileira. Estava lançado o projeto Atlas Linguístico do Brasil (projeto ALiB).

Como já assinalado anteriormente, esse projeto interinstitucional, sediado na Universidade Federal da Bahia, é coordenado por um Comitê Nacional, formado por pesquisadores de 08 universidades públicas brasileiras e tem como objetivo principal escrever a variante brasileira da língua portuguesa nos níveis fonético-fonológico, semântico-lexical, morfossintático e pragmático-discursivo sob a perspectiva da Dialetologia pluridimensional.

O projeto ALiB, de caráter nacional e de grande domínio no que tange à área de abrangência (um país de dimensões continentais), foi concebido em sintonia com as diretrizes contemporâneas dos estudos dialetológicos e geolinguísticos por isso o projeto contempla diferentes dimensões, além da diatópica, que é garantida por uma rede de pontos formada por 250 localidades do interior dos 26 Estados e capitais de 25 deles. Em cada ponto do interior entrevistam-se quatro informantes de nível fundamental e nas capitais somam-se a estes quatro informantes de nível universitário. No total, o *corpus* do projeto constará de um volume de gravações que revelarão características da língua falada no século XXI por 1.100 brasileiros.

¹⁰ Sobre a questão dos atlas linguísticos publicados e em andamento no Brasil, ver Isquierdo (2006).

O Questionário Linguístico utilizado como instrumento norteador para a coleta dos dados, além de perguntas que buscam apurar fenômenos fonético-fonológicos, lexicais e morfossintáticos, reúne perguntas voltadas para questões prosódicas, pragmáticas, metalingüísticas, além de temas para discursos semidirigidos e texto para leitura. O Quadro 1, na sequência, sintetiza as dimensões consideradas na metodologia do projeto ALiB e os parâmetros considerados para cada uma delas. Essas características técnicas configuram o ALiB como um atlas pluridimensional.

Quadro 1 - A pluridimensionalidade do projeto ALiB

DIMENSÕES	PARÂMETROS
Diatópica	Atlas topostático que reúne uma rede de pontos de 250 localidades brasileiras (cidades de pequeno, médio e grande porte), incluído as capitais de Estados, exceto Palmas (capital de Tocantins) e Brasília (capital do Distrito Federal) em virtude de serem cidades de fundação recente. As 250 localidades estão distribuídas pelos 26 Estados da Federação e foram selecionadas a partir dos seguintes critérios: densidade demográfica, áreas de fronteira, áreas dialetais já delineadas por pesquisas geolinguísticas produzidas no Brasil (atlas regionais).
Diageracional	Falantes de duas faixas etárias: Faixa I ⇒ 18 e 30 anos; Faixa II ⇒ 50 e 65 anos.
Diassexual	Falantes do sexo masculino. Falantes do sexo feminino.
Diastrática	Falantes com dois graus de escolaridade: Ensino Fundamental (localidades do interior e nas capitais); Superior (capitais).
Diafásica	Questionário linguístico Texto para leitura Discursos semidirigidos (temas para que o informante discorra livremente)
Diarreferencial	Apuração de fatos metalingüísticos e epilingüísticos

Fonte: Metodologia do projeto ALiB (2012)¹¹

O perfil do informante do projeto ALiB

Como já assinalado, o perfil dos informantes do Projeto ALiB considerou quatro variáveis sociolinguísticas: idade, sexo, escolaridade e naturalidade. Assim, em cada uma das 25 capitais, foram entrevistados oito informantes,

¹¹ O Quadro 1 foi elaborado pelos autores do texto, com base nas dimensões e parâmetros propostos por Thun (2005) e na metodologia do projeto ALiB disponível em <www.alib.ufba.br>. Acesso em: 25.fev.2012.

distribuídos equitativamente segundo essas variáveis, perfazendo um total de 200 informantes. Nota-se que o projeto ALiB avançou na perspectiva sociolinguística, agregando, à variável espaço geográfico, outras dimensões da variação linguística, como a diastrática (escolaridade), a diassetual (gênero), a diageracional (faixa etária), constituindo-se, pois, como um atlas linguístico pluridimensional, segundo o preconizado por Thun (2000).

Além dessas mencionadas, indiretamente, são consideradas outras variáveis sociolinguísticas que, muito embora não se configurem como um critério para a seleção dos informantes, fornecem informações sociolinguísticas que em muito auxiliam na interpretação e análise dos dados linguísticos documentados por meio dos inquéritos realizados *in loco* pelos pesquisadores do projeto ALiB. Essas informações são registradas na ficha do informante¹², que é preenchida parte antes e parte depois de cada inquérito, com a finalidade de coletar dados extralingüísticos que, por ventura, possam influir no uso da língua. Assim, a ficha do informante fornece, além de informações sociais tais como nível na escala social, hábitos sócio-culturais, escolaridade, também dados que caracterizam o momento do inquérito. Essa ficha compõe-se de 50 itens, dos quais os 37 primeiros são preenchidos com a ajuda do informante, antes da entrevista. Os itens de 1 a 19 referem-se aos dados pessoais do informante (nome, endereço, estado civil, escolaridade, naturalidade, profissão etc.). Os itens de 20 a 29 são atinentes ao contato do informante com os meios de comunicação (televisão, rádio, jornal, revista etc.) e registro da religião professada pelo falante. Os itens de 30 a 37 referem-se à participação do informante em diversões, tais como cinema, teatro, shows, esportes, dentre outras. Por fim, os itens 38 a 50, que são preenchidos pelos pesquisadores depois de concluído o inquérito, versam sobre as impressões do inquiridor acerca do informante e do ambiente da pesquisa – características psicológicas do informante (tímido, vivo, perspicaz, sarcástico); espontaneidade da elocução (total, grande, média, fraca), postura do informante durante o inquérito (cooperativa, não cooperativa, agressiva, indiferente); categoria social do informante (A, B, C, D); grau de conhecimento entre inquiridor e informante (grande, médio, pequeno, nenhum). Além disso, a ficha possui campos destinados ao registro de informações sobre interferências de circunstâncias, ambiente da entrevista, a família do informante.

A partir deste ponto do artigo, tecemos considerações sobre o perfil dos 200 informantes das 25 capitais brasileiras que integram a rede de pontos do projeto ALiB com base nas informações constantes das fichas dos informantes já armazenadas no Banco de Dados do Projeto. Para tanto, selecionamos 11 itens da ficha: 11 (escolaridade), 21 (assiste TV), 23 (ouve rádio), 25 (lê jornal), os itens de

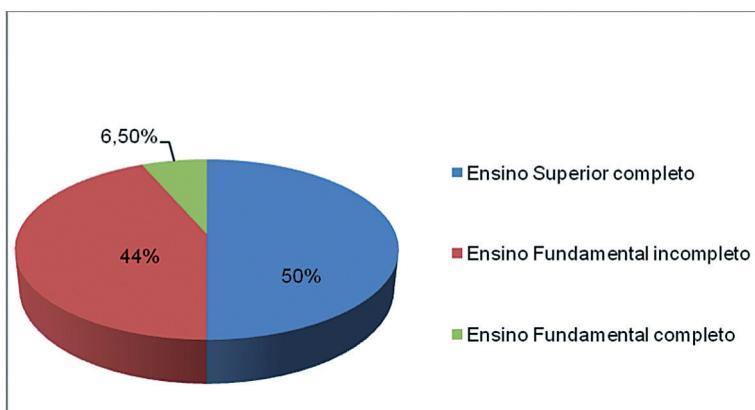
¹² A ficha do informante integra o Questionário Linguístico do projeto ALiB (2001).

30 a 36 (participação em diversões). Priorizamos a abordagem quantitativa, por isso os dados são apresentados por meio de gráficos e tabelas, acompanhadas de descrição sumária dos resultados obtidos.

Escolaridade

A dimensão diastrática, no âmbito do Projeto ALiB, é controlada pelo fator escolaridade. Assim, nas capitais brasileiras foram entrevistados 200 informantes, 100 com formação universitária e 100 com formação básica, falantes alfabetizados que cursaram até a última série do Ensino Fundamental. Dessa feita, dentre os 200 informantes inquiridos nas capitais, 50% possuem nível universitário completo, enquanto 44% têm como nível de escolaridade o curso fundamental incompleto e apenas 6,5% situam-se entre os que cursaram o Ensino Fundamental completo, conforme o visualizado no Gráfico 1:

Gráfico 1 - Informantes do ALiB/capitais – nível de escolaridade

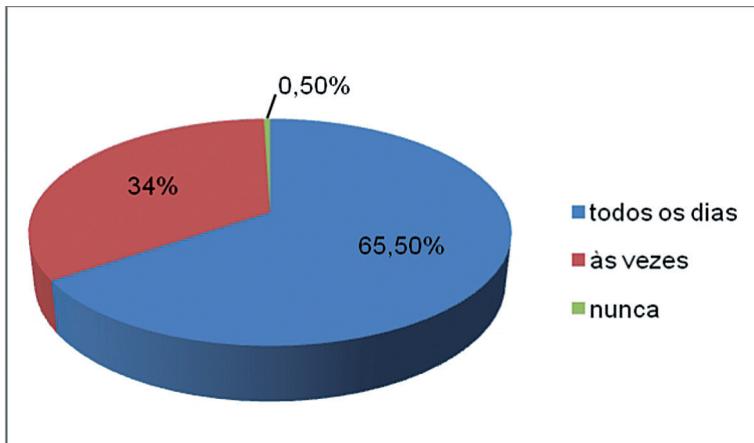


Fonte: Banco de Dados do ALiB (2012).

Contatos com os meios de comunicação - televisão e rádio

Os dados armazenados no Banco de Dados do ALiB revelam que é significativo o contato dos habitantes das capitais brasileiras selecionados como informantes com os meios de comunicação de massa, como a televisão e o rádio, pois 65,5% dos informantes declaram que assistem a programas veiculados pela televisão todos os dias, 34% às vezes e apenas 0,5% dos entrevistados declaram nunca ter contato com programas de televisão, conforme ilustra o Gráfico 2.

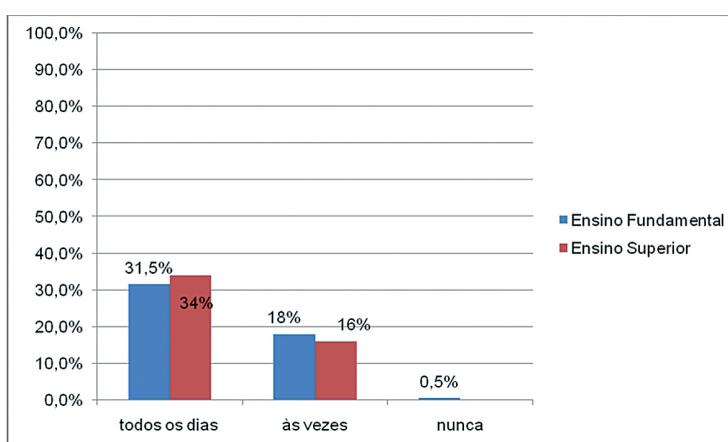
Gráfico 2 - Informantes do ALiB/capitais – contato com os meios de comunicação/televisão



Fonte: Banco de dados do ALiB (2012).

O cruzamento entre os dados relativos à variável escolaridade com os concorrentes ao hábito de assistir à televisão evidencia que, em termos percentuais, não há informações contrastantes, pois dos 65,5% que declararam assistir a programas de televisão todos os dias, 34% são de nível superior e 31,5% de nível fundamental. Já dentre os 34% que declararam às vezes ter como meio de diversão a televisão, 18% cursaram Ensino Fundamental e 16% Curso Superior. Por fim, apenas um informante de nível fundamental afirma não ter como hábito assistir à televisão (Gráfico 3).

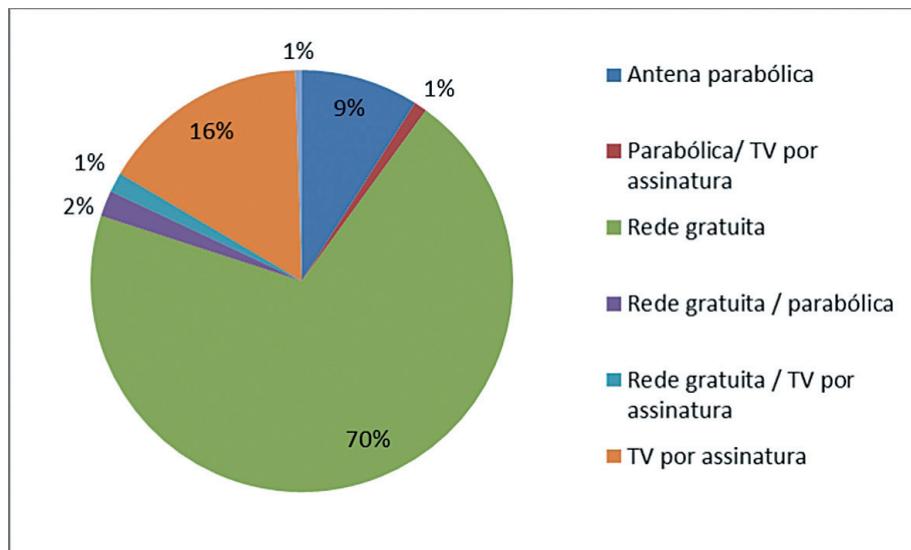
Gráfico 3 - Informantes do ALiB/capitais – contato com os meios de comunicação/televisão, segundo a variável escolaridade



Fonte: Banco de dados do ALiB (2012).

Outro dado a ser pontuado diz respeito à distribuição percentual do tipo de transmissão de canais de televisão. Nota-se, pelo Gráfico 4, que a grande predominância recai no tipo de rede gratuita (70%), seguida da TV por assinatura (16%), da parabólica (9%) e, em pequenos índices, os dados apontam para a convivência com mais de um tipo de transmissão.

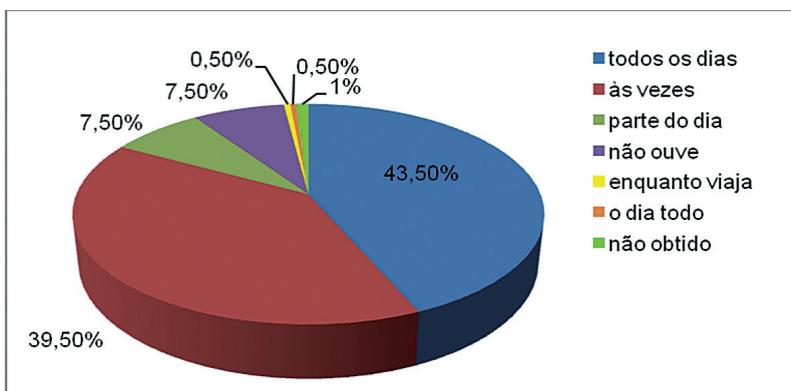
Gráfico 4 - Informantes do ALiB/capitais – contato com os meios de comunicação/televisão, segundo o tipo de transmissão



Fonte: Banco de dados do ALiB (2012).

Tratando agora dos dados relativos ao quesito “audição de rádio”, o Banco de Dados do projeto ALiB regista o seguinte panorama: 43,5% dos informantes declara ouvir rádio todos os dias; 38,5% às vezes; 7,5% dos informantes passa parte do dia ouvindo programas de rádio; 7,5% dos entrevistados afirmam não ouvir rádio e, em menor percentual, 0,5% ouvem rádio enquanto viajam e também 0,5% têm como hábito ouvir o dia todo. 1% dos entrevistados não forneceram esse tipo de dado (Gráfico 5).

Gráfico 5 - Informantes do ALiB/capitais – contato com os meios de comunicação/rádio

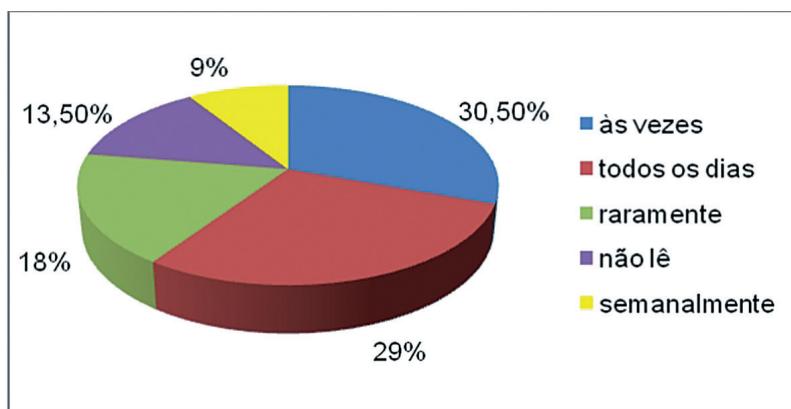


Fonte: Banco de dados do ALiB (2012).

Hábito de leitura de jornal

Ainda sobre o contato com os meios de comunicação, consideraremos o item 25 que vai revelar o contato dos informantes com a leitura de jornal. Do total de todos os entrevistados, 30,5% deles declaram ler jornal às vezes, 29% leem todos os dias, 18% raramente, 13,5% não lê e 9% leem semanalmente (Gráfico 6).

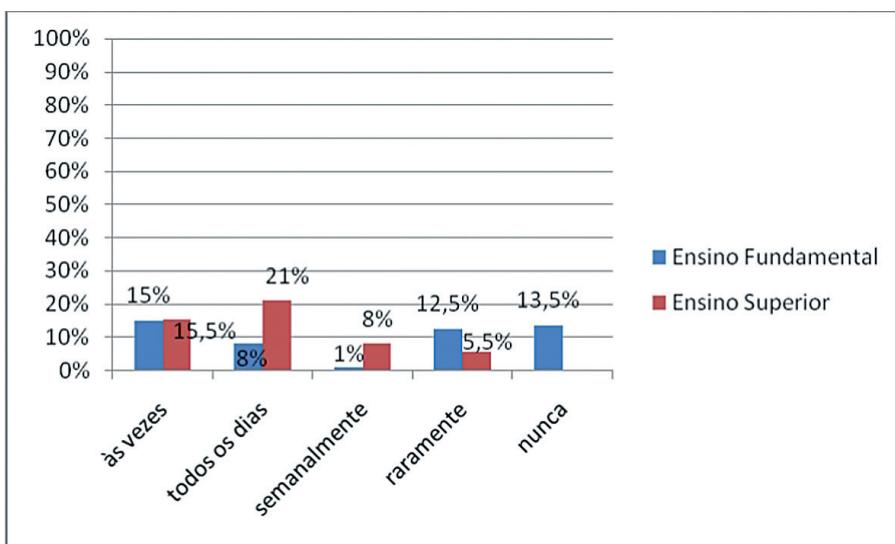
Gráfico 6 - Informantes do ALiB/capitais – contato com os meios de comunicação/leitura de jornal



Fonte: Banco de dados do ALiB (2012).

Considerando agora o cruzamento entre os dados relativos ao hábito de leitura de jornal e a variável escolaridade (Gráfico 7), nota-se que, conforme o previsível, o contato com a leitura é mais frequente entre os informantes de nível superior, uma vez que os dados percentuais revelam que a frequência de leitura é mais comum entre os falantes dessa categoria – 21% deles declararam ler jornal todos os dias e 15,5% às vezes. Já entre os informantes com menor escolaridade, 15,5% declaram ler jornal às vezes; 12,5% raramente têm esse hábito de lazer e 8% dos entrevistados afirmam ler todos os dias o jornal. Também entre os informantes de nível fundamental, verifica-se que 13,5% deles atestam não ler jornal.

Gráfico 7 - Informantes do ALiB/capitais – contato com os meios de comunicação/leitura de jornal, segundo a variável escolaridade



Fonte: Banco de dados do ALiB (2012).

Já o Quadro 2, a seguir, apresenta a frequência da leitura de jornal em dados percentuais considerando a região administrativa do Brasil.

Quadro 2 - Informantes do ALiB/capitais – contato com os meios de comunicação/leitura de jornal, segundo a região administrativa do Brasil

Região \	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Todos os dias	31,25%	20,83%	37,50%	31,20%	37,51%
Às vezes	31,25%	33,34%	20,83%	28,20%	33,33%
Raramente	16,66%	13,89%	20,83%	25%	20,83%
Não lê	14,59%	18,05%	12,50%	12,50%	-----
Semanalmente	6,5%	13,89%	8,33%	3,10%	8,33%
Total por região	100%	100%	100%	100%	100%

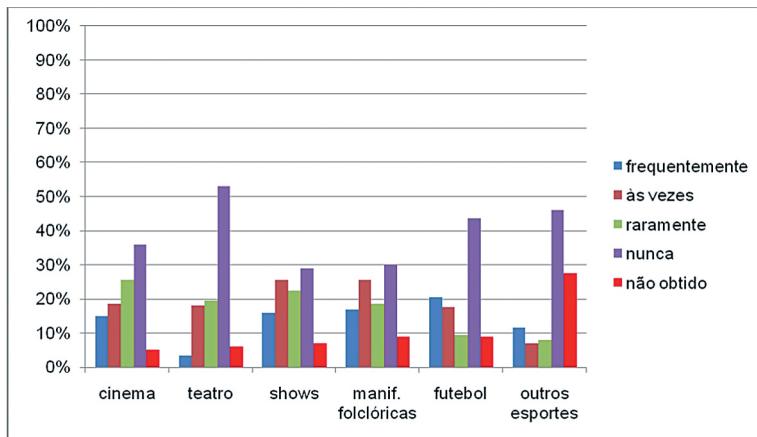
Fonte: Banco de dados do ALiB (2012).

Conforme se observa no Quadro 2, curiosamente as regiões que evidenciam o maior índice de leitura são a Sul e a Centro-Oeste, onde, respectivamente, 37,51% e 37,50% dos informantes declararam ter o hábito de ler todos os dias. Já a leitura esporádica (às vezes) de jornal destaca-se nas regiões Nordeste (33,34%) e Sul (33,33%), enquanto 25% dos informantes do Sudeste leem jornal raramente. E o destaque principal é para a região Nordeste que apresentou o maior índice de não leitura de jornal – 18,05% de informantes atestaram essa realidade. Das cinco regiões brasileiras, apenas na região Sul não houve casos de falantes que não têm como hábito a leitura de jornal.

Hábitos de diversões: cinema e teatro

Na sequência, o Gráfico 8 visualiza os resultados relativos às respostas aos itens 30 a 36 da ficha do informante, acerca de hábitos de diversões como cinema, teatro, shows diversos, manifestações folclóricas, futebol e outros esportes.

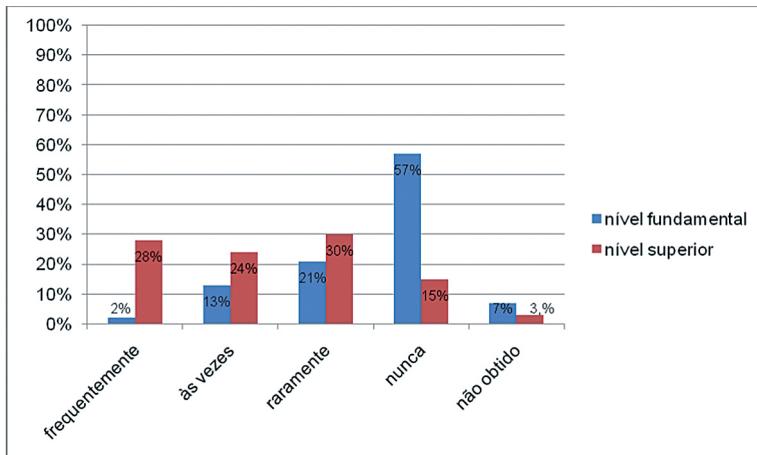
Gráfico 8 - Informantes do ALiB - participação em diversões



Fonte: Banco de dados do ALiB (2012).

Nota-se, pelo Gráfico 8, que os informantes do ALiB em análise não cultivam determinados hábitos de lazer, o que fica evidente pela alta produtividade da opção “nunca” em todas as atividades lúdicas mencionadas, com destaque para o teatro (53%) e para outros esportes (46%). Em contrapartida, a atividade indicada como prática mais frequente foi futebol (20,5%), seguida de manifestações folclóricas, shows e cinema. Consideramos a seguir as atividades de lazer cinema (Gráfico 9) e teatro (Gráfico 10), sob a perspectiva da variável escolaridade.

Gráfico 9 - Informantes do ALiB – participação em diversões/cinema, segundo a variável escolaridade

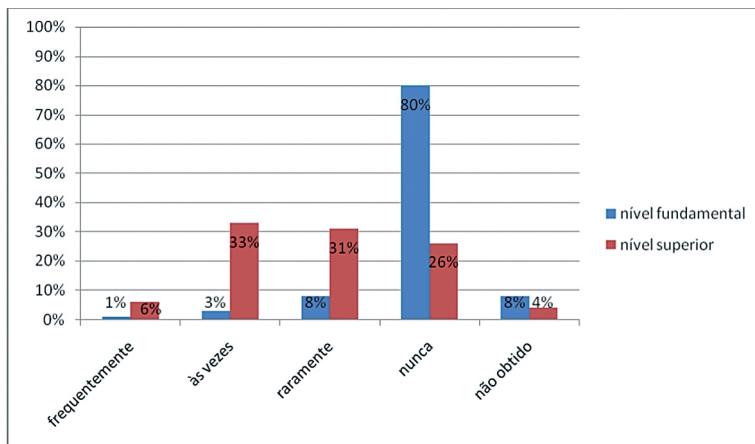


Fonte: Banco de dados do ALiB (2012).

Os dados visualizados no Gráfico 9 demonstram que nas capitais brasileiras o maior índice de não participação em teatro recai no grupo de falantes de baixa escolaridade (57%). Já entre o grupo dos informantes com Curso Superior apenas 28% vão com frequência ao cinema; 24% assinalaram às vezes e 30% raramente vão ao teatro. Esses dados apontam para uma significativa diferença diastrática entre o grupo de informante.

Resultados similares ficam evidentes com relação à participação em teatro (Gráfico 10), à medida que 80% dos entrevistados com ensino fundamental nunca participam dessa modalidade de lazer. Já dentre os falantes com formação em nível superior, 33% assinalam que às vezes frequentam o teatro; 31% raramente vão ao teatro e 26% nunca frequentam essa atividade de lazer.

Gráfico 10 - Informantes do ALiB – participação em diversões/teatro, segundo a variável escolaridade



Fonte: Banco de dados do ALiB (2012).

As características dos informantes do Projeto ALiB destacadas nos gráficos anteriores fornecem um panorama dos informantes das capitais brasileiras que, por sua vez, refletem a realidade dos brasileiros dos grandes centros urbanos em termos de opções de lazer.

Algumas considerações finais

Buscamos, ao longo deste texto, discutir a questão das dimensões consideradas nas pesquisas dialetológicas, permeada por uma perspectiva histórica em que se procurou recuperar diferentes momentos da Dialetologia e respectivas

características que identificaram os atlas linguísticos em diferentes fases dessa disciplina linguística. Focamos particularmente a questão da pludimensionalidade dos atlas linguísticos, fruto das contribuições assimiladas da Sociolinguística, sobretudo a partir da década de 60 do século passado. Nesse contexto situamos o Projeto do Atlas Linguístico do Brasil, no que tange à dimensão sociolinguística evidenciada nos parâmetros teórico-metodológicos que orientam a pesquisa. Centramos nossa reflexão no perfil dos informantes, não só na discussão desse perfil, mas também na apresentação de dados relativos aos 200 informantes entrevistados nas capitais brasileiras com vistas a demonstrar aspectos do viés sociolinguístico manifestado nas características desses informantes.

A amostra de variáveis sociais focalizadas neste estudo demonstra que as características dos informantes, para além de subsidiar a análise dos dados linguísticos coletados pelo projeto, fornecem um panorama de características dos habitantes brasileiros de grandes centros urbanos que também denunciam características da sociedade brasileira, em termos de hábitos culturais, como leitura de jornais e revistas, participação em atividades culturais como cinema, teatro. Os dados demonstram também a relação entre a prática desses hábitos e o nível de escolaridade dos informantes.

Por fim, é preciso assinalar que os estudos em curso sobre fenômenos linguísticos selecionados para integrar o volume I do Atlas Linguístico do Brasil, como também trabalhos pontuais desenvolvidos em nível de pós-graduação com dados do Banco de dados do ALiB têm demonstrado a importância das variáveis tomadas como parâmetro nas dimensões consideradas na metodologia do ALiB na interpretação de dados linguísticos. Nesse particular, citamos, a título de ilustração, o estudo de Cardoso (2012), sobre o uso do *tu/você* nas capitais do Brasil; de Marins (2012), acerca da questão do rural e do urbano no léxico dos habitantes da região Centro-Oeste, e de Benke (2012), que estudou manifestações de tabus linguísticos no léxico de habitantes das capitais brasileiras. E assim o projeto ALiB vai seguindo o seu curso com o propósito de contribuir para o conhecimento da realidade linguística do Brasil, coletando e analisando fatos linguísticos coletados por meio da metodologia da Dialetologia pluridimensional, perspectiva teórica que orienta contemporaneamente os estudos geolinguísticos.

ISQUERDO, A. N.; ROMANO, V. P. Discussing the sociological dimension of the ALiB project: a reflection based on the informants profile. *Alfa*, São Paulo, v.56, n.3, p.881-906, 2012.

- *ABSTRACT: This paper focuses on the discussion of the multidimensional dimension of dialectology studies and it aims to: i) discuss aspects of sociolinguistics dimension of the ALiB project (Linguistic Atlas of Brazil), based on the profile established for the selection of informants and ii) present an overview of the characteristics of 200 informants of ALiB who were interviewed in 25 Brazilian capitals, discussing highlighted trends in this universe of*

speakers of contemporary Portuguese language, especially in terms of education level, contacts with the media – television and radio; newspaper reading habits and entertainment: cinema and theater. In order to do so, it retrieves historiographical data related to the Dialectology subject, highlighting aspects that show, since its inception, the necessity to consider, in addition to horizontality, the vertical dimension aspects of linguistic studies, also taking social variables as parameters for the description and analysis of the speech of social groups fixed in geographical areas that were previously delimited.

- KEYWORDS: Multidimensional dialectology. Linguistic Atlas. Brazil. Informants.

REFERÊNCIAS

- ALMEIDA, F. da S. C. *Micro atlas fonético do Estado do Rio de Janeiro (Micro AFERJ)*: uma contribuição para o conhecimento dos falares fluminenses. 2008. 162f. Tese (Doutorado em Língua Portuguesa) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008. 2v.
- BENKE, V. C. M. *Tabus linguísticos nas capitais do Brasil*: um estudo baseado em dados geolinguísticos. 2012. 310f. Dissertação (Mestrado em Estudos de Linguagens) - Universidade Federal de Mato Grosso do Sul, Campo Grande, 2012.
- BORTONI-RICARDO, S. M. *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil*. Cambridge: Cambridge UP, 1985.
- BRANDÃO, S. F. A geografia lingüística do Brasil. São Paulo: Ática, 1991.
- BUSSE, S. Atlas Linguístico-Etnográfico da Região Oeste do Paraná/ALERO: uma descrição preliminar do movimento diatópico e diastrático da fala. *SIGNUM*, Londrina, v.12, n.1, p.123-144, jul. 2009.
- CALLOU, D. Quando dialetologia e sociolinguística se encontram. *Estudos Linguísticos e Literários*, Salvador, n.41, p.29-48, jan.-jun. 2010.
- CARDOSO, S. A. M. *O Atlas Linguístico do Brasil*: uma visão crítica dos caminhos seguidos e perseguidos. Comunicação apresentada no II Congresso Internacional de Linguística Histórica, Universidade de São Paulo, São Paulo, fev. 2012.
- _____. *Geolinguística: tradição e modernidade*. São Paulo: Parábola, 2010.
- _____. Geolinguística no terceiro milênio: monodimensional ou pluridimensional? *Revista GELNE*, Fortaleza, v.4, n.2, p.1-16, 2002. Disponível em: <http://www.gelne.org.br/Site/RevistaGelne/arquivos/artigos/art_34a01e3a7b2f8deaa71b52a3df2d54c0_12.pdf>. Acesso em: 24. mar. 2012.

_____. Dialetologia: trilhas seguidas, caminhos a perseguir. *D.E.L.T.A*, São Paulo, v.17, 2001. Disponível em: <http://www.scielo.br/scielo.php?pid=S0102-44502001000300003&script=sci_arttext>. Acesso em: 30 de mar. 2012.

COMITÊ NACIONAL DO PROJETO ALIB. *Atlas Lingüístico do Brasil*: questionários 2001. Londrina: UEL, 2001.

_____. *Projeto Atlas Linguístico do Brasil* – Projeto ALIB. Disponível em: <www.alib.ufba.br>. Acesso em: 25 fev. 2012.

CHAMBERS, J. K.; TRUDGILL, P. *La Dialectología*. Tradução de Carmen Morán Gonzalez. Madrid: Visor Libros, 1994.

FERREIRA, C.; CARDOSO, S. *A dialectologia no Brasil*. São Paulo: Contexto, 1994.

ISQUERDO, A. N. Os atlas regionais brasileiros publicados e em curso: percursos metodológicos. In: MOTA, J. A.; CARDOSO, S. A. M. (Org.). *Documentos 2*: projeto atlas lingüístico do Brasil. Salvador: Quarteto, 2006. p.67-94.

MARGOTTI, F. W. Geolingüística pluridimensional: desafios metodológicos. In: ENCONTRO DO CÍRCULO DE ESTUDOS LINGÜÍSTICOS DO SUL, 8., 2008, Porto Alegre. *Anais...* Pelotas, Educat, 2008, p.1-9. Disponível no site: <http://celsul.org.br/Encontros/08/geolinguistica_pluridimensional.pdf>. Acesso: 29 mar. 2012.

MARINS, L. G. F. *O rural e o urbano: novos e velhos olhares no Centro-Oeste do Brasil*. 2012. 325f. Dissertação (Mestrado em Estudos de Linguagens) - Universidade Federal de Mato Grosso do Sul, Campo Grande, 2012.

MOTA, J. A.; CARDOSO, S. A. M. Para uma nova divisão dos estudos dialeitais brasileiros. In: MOTA, J. A.; CARDOSO, S. A. M. (Org.). *Documentos 2*: Projeto Atlas Lingüístico do Brasil. Salvador: Quarteto, 2006. p.15-34.

RADTKE, E.; THUN, H. Nuevos caminos de la geolinguística románica. Un balance. In: RADTKE, E.; THUN, H. *Neue Wege der Romanischen Geolinguistik*. Kiel: Westensee-Verlag, 1996. p.25-49.

RAMÍREZ, A. G. Dialectología y sociolingüística. In: ALVAR, M. (Director). *Manual de dialectología hispánica*: El Español de España. Barcelona: Ariel Lingüística, 2009. p.37-48.

REIS, R. C. P. *Atlas Lingüístico do município de Ponta Porã-MS*: um registro das línguas em contato na fronteira do Brasil com o Paraguai. 2006. 2 v. Dissertação (Mestrado em Letras) - Universidade Federal de Mato Grosso do Sul, Aquidauana, 2012.

ROMANO, V. P. *Atlas geossociolinguístico de Londrina*: um estudo em tempo real e tempo aparente. 2012. 2 v. Dissertação (Mestrado em Estudos da Linguagem) - Universidade Estadual de Londrina, Londrina, 2012.

THUN, H. A dialetologia pluridimensional no Rio da Prata. In: ZILLES, A. M. S. (Org.). *Estudos de variação linguística no Brasil e no Cone Sul*. Porto Alegre: Ed. da UFRGS, 2005. p.63-92.

_____. La géographie linguistique romane à la fin du XXe Siècle. In: CONGRES INTERNATIONAL DE LINGUISTIQUE ET PHILOLOGIE ROMANES, 1998, 22., Bruxelles, 1998. *Actes...* Tubingen: Niemeyer, 2000. p.367-388.

THUN, H; ELIZAINCÍN, A. *Atlas lingüístico Diatópico y Diastrático del Uruguay (ADDU)*. Kiel: Westensee, 2000.

Recebido em março de 2012.

Aprovado em junho de 2012.

BANCOS DE DADOS SOCIOLINGUÍSTICOS DO PORTUGUÊS BRASILEIRO E OS ESTUDOS DE TERCEIRA ONDA: POTENCIALIDADES E LIMITAÇÕES

Raquel Meister Ko. FREITAG*

Marco Antonio MARTINS**

Maria Alice TAVARES***

- RESUMO: Bancos de dados linguísticos de fala – especialmente aqueles elaborados para a pesquisa de orientação sociolinguística variacionista – têm sido fonte privilegiada para a descrição do português brasileiro. Neste texto, discutimos procedimentos metodológicos que deveriam ser adotados para a organização de novos bancos de dados. Fazemos um breve retrospecto dos bancos de dados já constituídos e sugerimos a coleta e expansão de *corpora* de diferentes comunidades de fala – e de diferentes comunidades de prática. De acordo com proposta defendida por Eckert (2012), os estudos sociolinguísticos podem ser distinguidos em três ondas de análise que refletem modos distintos de abordagem à variação linguística. Sugerimos estratégias para padronizar os procedimentos de organização de bancos de dados sociolinguísticos que levem em conta as três diferentes ondas da pesquisa sociolinguística, e destacamos a terceira onda, ainda incipiente no Brasil. A padronização dos bancos de dados sociolinguísticos facilitaria a realização de investigações contrastivas de diferentes dialetos brasileiros, contribuindo, dessa forma, para o estabelecimento e refinamento de generalizações e princípios de variação e mudança universais.
- PALAVRAS-CHAVE: Sociolinguística. Banco de dados. Variação e mudança linguística. Fatores sociais. Estilo.

Introdução

Na literatura sociolinguística variacionista, a referência à metodologia costuma ocupar duas a três linhas, quando muito um parágrafo: “como *corpus* foram selecionados X informantes do banco de dados Y, estratificados em Z células sociais”. A voz passiva da construção e a exiguidade do espaço dedicado à

* UFS – Universidade Federal de Sergipe, Centro de Educação e Ciências Humanas – Departamento de Letras Vernáculas. São Cristóvão – Sergipe – Brasil. 49100-000 - rkofreitag@uol.com.br

** UFRN – Universidade Federal do Rio Grande do Norte, Centro de Ciências Humanas, Letras e Artes. Natal – Rio Grande do Norte – Brasil. 59072-970 - marcoamartins.ufrn@gmail.com

*** UFRN – Universidade Federal do Rio Grande do Norte, Centro de Ciências Humanas, Letras e Artes. Natal – Rio Grande do Norte – Brasil. 59072-970 - aliceplp@hotmail.com

metodologia de organização do *corpus* não condizem com o real esforço e tempo dispendidos no processo de constituição de um banco de dados, desde a prospecção e seleção de informantes até a transcrição, armazenamento e disponibilização.

Bancos de dados linguísticos de fala (especialmente os que seguem a orientação da Sociolinguística Variacionista) têm sido fonte privilegiada para a descrição do português brasileiro. Porém, a constituição de *corpus* que procure considerar as variedades do português brasileiro é tarefa dispendiosa não só quanto a recursos financeiros, mas também quanto ao tempo. A cada projeto que constitui seu banco de dados em uma comunidade de fala, o mapeamento das variedades do português no Brasil vai se efetivando, mas só a padronização dos procedimentos metodológicos permitirá a realização de estudos contrastivos entre as variedades, para, então, possibilitar uma descrição mais acurada do português brasileiro.

Iniciativas para viabilizar estudos linguísticos variacionistas otimizando os recursos têm se tornado prática no cenário nacional, afinal, a comparação entre resultados obtidos para fenômenos variáveis é um recurso analítico que permite grandes avanços teóricos para a pesquisa linguística, uma vez que transcender os limites de uma única variedade linguística possibilita o estabelecimento, refinamento e fortalecimento de generalizações e princípios de variação e mudança universais (TAVARES, 2002).

Como um dos objetivos da Sociolinguística variacionista é obter resultados que possam ser generalizados, sua metodologia deve ser pautada em confiabilidade (os mesmos resultados devem ser repetidos na análise do mesmo fenômeno) e intersubjetividade (dois pesquisadores diferentes devem obter os mesmos resultados seguindo a mesma metodologia) (BAILEY; TILLERY, 2004). Ao tecerem avaliações de ordem metodológica – campo ainda pouco explorado nos estudos variacionistas – Bailey e Tillery discutem três razões possíveis para explicar a divergência de resultados em abordagens sociolinguísticas quantitativas a partir da premissa de que diferenças metodológicas resultam em divergências de resultados, especialmente quanto às diferenças de entrevistador e às diferenças na amostra da população.

Bailey e Tillery (2004) retomam o estudo de Rickford e McNair-Knox (1994), em que o mesmo informante afroamericano foi entrevistado por duas entrevistadoras: uma também afroamericana e outra branca. A frequência com que traços característicos do AAVE apareciam na entrevista realizada pela afroamericana era sensivelmente superior à da entrevista realizada pela entrevistada branca. Outro aspecto que Bailey e Tillery destacam é que a experiência do pesquisador de campo que realiza a entrevista (e, em menor efeito, o tópico) também mostra efeitos na frequência de fenômenos. Quanto aos efeitos da amostra, Bailey e Tillery (2004) ressaltam que os estudos sociolinguísticos costumam

trazer poucas informações acerca da seleção dos seus informantes ou de sua representatividade na comunidade. Os autores sugerem, a título de recomendação, que os pesquisadores precisam especificar exatamente qual é a amostra da população em estudo, assim como especificar quais os procedimentos para definir a amostra, de modo que possam garantir a confiabilidade e a intersubjetividade da análise. A discussão de Bailey e Tillery (2004) sugere que os resultados de uma investigação sociolinguística são não raro muito mais a consequência de escolhas metodológicas do que o comportamento dos informantes. O que torna a situação problemática, segundo os autores, é que a sociolinguística quantitativa não tem um corpo de pesquisadores que se dediquem ao método, nem literatura que explore os efeitos de diferentes entrevistadores, diferentes estratégias de elicitização dos dados, procedimentos de amostragem ou estratégias analíticas. À esteira da constatação de Bailey e Tillery (2004), neste texto, tecemos reflexões acerca da metodologia das abordagens sociolinguísticas de orientação variacionista desenvolvidas no Brasil, especificamente no que tange à constituição de bancos de dados. Fazemos um breve retrospecto dos bancos de dados já constituídos e prospectamos ações futuras neste campo, com a expansão e ampliação de amostras de variedades linguísticas (novos bancos de dados e coletas piloto). A orientação da discussão segue a proposta de Eckert (2012) a respeito das três ondas da Sociolinguística e as ponderações de Bailey e Tillery (2004), já apresentadas, sobre dados divergentes em sociolinguística (e que procedimentos metodológicos são pertinentes para minimizá-los).

As três ondas da Sociolinguística e os bancos de dados brasileiros

Propondo uma discussão sobre os rumos do significado social no estudo da variação, Eckert (2012)¹ faz uma abordagem programática dos estudos sociolinguísticos com o propósito de relevar o estudo da variação com ênfase no significado social: como o sistema de significado social é estruturado? Que tipos de significados sociais são expressos na variação? Em seu retrospecto, Eckert destaca que os estudos sociolinguísticos podem ser agrupados em três ondas de estudos, não substitutivas nem sucessivas, mas que se configuram como modos distintos de pensar a variação, com práticas analíticas e metodológicas peculiares. A proposta das três ondas dos estudos sociolinguísticos de Eckert vem recebendo sinalizações de que merece atenção no cenário sociolinguístico brasileiro (BENTES, 2009; CAMACHO, 2010; SCHERRE, 2011; HORA; WETZELS,

¹ Esse texto é a versão revisada e modificada do trabalho intitulado Variation, convention, and social meaning, que foi apresentado por Penelope Eckert na Annual Meeting of the Linguistic Society of America, em Oakland, no ano de 2005. Nessa versão mais recente de sua proposta, a autora sugere a integração entre os níveis social e cognitivo na dinâmica da variação que potencialmente os estudos de terceira onda poderiam abranger. A versão mais antiga do texto está disponível em <<http://www.stanford.edu/~eckert/EckertLSA2005.pdf>>.

2011, entre outros), motivando a discussão acerca de seu impacto para o campo de estudos. Apresentamos, a seguir, a proposta de Eckert (2012) para a abordagem de cada uma das ondas de estudos sociolinguísticos, com ênfase nos aspectos metodológicos, especialmente no que diz respeito à constituição de bancos de dados sociolinguísticos brasileiros. Ao final da seção, avaliamos as potencialidades e as limitações de cada uma das abordagens quanto à constituição de bancos de dados, delimitando os aspectos que os novos bancos de dados potencialmente precisam contemplar para contribuir de modo efetivo aos estudos que enfocam o significado social da variação.

A primeira onda de estudos sociolinguísticos inicia com os estudos de Labov sobre a estratificação do inglês na cidade de Nova Iorque, cujos resultados foram replicados em uma série de estudos em comunidades urbanas que corroboraram um padrão regular de estratificação socioeconômica das variáveis, em que o uso das variantes não padrão está inversamente relacionado ao *status* socioeconômico do falante (ECKERT, 2012). A primeira onda estabeleceu uma base sólida para o estudo da variação, evidenciando as correlações entre variáveis linguísticas e categorias sociais primárias, como classe socioeconômica, sexo, idade, escolaridade etc. Os padrões regulares e sistemáticos de covariação social e linguística levantaram questões sobre relações sociais subjacentes às categorias sociais primárias, o que conduziu ao surgimento da segunda onda, caracterizado por estudos etnográficos de populações mais localmente definidas.

A premissa dos estudos de primeira onda é, pois, que as variedades linguísticas carregam o status social de seus falantes. A metodologia dos estudos de primeira onda é calcada na correlação entre as variáveis linguísticas e as categorias socioeconômicas em sentido amplo (cuja classificação se dá de forma estável, homogênea e padronizada de modo a permitir a replicação, como faixa etária, sexo, etnicidade, escolaridade), com a estratificação dos falantes em células sociais, a constituição de bancos de dados linguísticos e resultados quantitativos refinados (especialmente com o uso de técnicas estatísticas aprimoradas para o modelo da variação linguística, como a regressão logística com o cálculo de desvio da média ponderada (SANKOFF; TAGLIAMONTE; SMITH, 2005)).

No Brasil, os estudos quantitativos com bancos de dados estratificados de acordo com características sociodemográficas amplas têm se consolidado como modelo hegemônico, com os bancos de dados do Programa de Estudos sobre o Uso da Língua (PEUL), da Universidade Federal do Rio de Janeiro, que foi o pioneiro no Brasil a implementar esse modelo de constituição de amostra². A partir deste, foram replicados projetos em diferentes regiões do Brasil, com adaptações em sua metodologia (Projeto Variação Linguística Urbana na Região

² Para mais detalhes sobre o banco de dados PEUL, ver Scherre e Roncarati (2008).

Sul do Brasil – VARSUL, da equipe formada pela Universidade Federal do Paraná, Universidade Federal de Santa Catarina; Universidade Federal do Rio Grande do Sul e Pontifícia Universidade Católica do Rio Grande do Sul³; Projeto Variação Linguística na Paraíba – VALPB, da Universidade Federal da Paraíba; Banco de Dados Sociolinguístico da Fronteira e da Campanha Sul-Rio-Grandense – BDS-Pampa, da equipe da Universidade Federal de Pelotas e da Pontifícia Universidade de Pelotas; Banco de Dados por Classe Social – VarX, da Universidade Federal de Pelotas, entre outros).⁴ Esse tipo de banco de dados possibilita captar tendências amplas de variação e mudança em uma comunidade de fala. Implica, entretanto, a homogeneização da amostra, como discutimos mais à frente.

É importante destacar que a elaboração desses *corpora* permitiu a descrição do português brasileiro em diferentes aspectos linguísticos e considerando distintas variedades. De algum modo, têm-se uma descrição da variação na(s) gramática(s) do português do Brasil envolvendo diferentes fenômenos e a correlação destes com variáveis sociais.

Os estudos de segunda onda são também de natureza quantitativa, mas de abordagem etnográfica, abarcando categorias sociodemográficas mais abstratas, a fim de evidenciar como o vernáculo assume valor local. Os estudos etnográficos enfocam comunidades menores por períodos de tempo relativamente longos com o objetivo de descobrir as categorias sociais localmente mais salientes. Essas categorias podem ser instanciações locais das categorias primárias que guiam os estudos quantitativos, mas o traço distintivo crucial desse tipo de estudo é a descoberta do lugar dessas categorias na prática social local. Nesse tipo de abordagem, o foco recai nos conceitos de comunidades de fala e de identidade de grupo. Eckert (2012) traz, em referência à segunda onda, três exemplos: (1) o estudo de Labov sobre o inglês afroamericano (AAVE), cujos resultados apontam para o uso de traços vernaculares por adolescentes como indexadores do status entre o grupo de comunidade de prática; (2) o estudo de Milroy (1980), que enfoca comunidades de classe operária e examina a relação entre engajamento local e uso do vernáculo, correlacionando o uso de variáveis vernaculares locais com a densidade e a multiplicidade da rede de relações sociais do falante; e (3) o estudo da própria Eckert sobre o papel das categorias *jokers* e *burnouts* na indexação de classe socioeconômica em grupos adolescentes (ECKERT, 2000).

No cenário brasileiro, esse tipo de abordagem não recebeu a mesma ênfase que os estudos quantitativos baseados em categorias sociais amplas. Dentre os poucos estudos que se encaixam na segunda onda, escolhemos o de Ferrari (1994) para ilustração. Ferrari (1994) selecionou doze traços fonológicos, escalonados entre

³ Ver Bisol, Menon e Tasca (2008).

⁴ Nesta relação, não incluímos o projeto Norma Urbana Culta – NURC porque, apesar de subsidiar também descrições de cunho sociolinguístico, seu banco de dados não foi constituído para essa finalidade.

discretos – em que a variável indica uma delimitação nítida entre grupos sociais contíguos – ou gradientes – em que a variável não se apresenta com frequência significativamente maior de um grupo social para outro –, e um traço sintático-semântico (variação de preposição locativa “em” vs. “ni”), a fim de verificar as relações entre variação e redes sociais na comunidade do Morro dos Caboclos, no Rio de Janeiro. A rede de relações sociais do indivíduo estabelecida na comunidade não se configura como um indicador sociodemográfico amplo, como sexo, idade, escolarização etc.; trata-se de um indicador que só é captado com um estudo investigativo individualizado, aos moldes etnográficos. Os resultados da investigação de Ferrari (1994) apontam que redes sociais relativamente fechadas possibilitam a focalização de traços linguísticos (conservação dos traços linguísticos característicos da comunidade do Morro dos Caboclos), enquanto redes sociais pouco coesas associam-se à difusão linguística (abandono dos traços linguísticos da comunidade em troca de traços que os aproximam dos moradores de bairros da zona oeste carioca, nas proximidades do Morro dos Caboclos): os falantes que trabalhavam na cidade faziam uso de traços linguísticos diferentes daqueles que nunca desciam o morro. Estudos dessa natureza permitem uma avaliação mais acurada do fenômeno da variação, com ênfase no valor social das variáveis. São, entretanto, estudos dispendiosos e demorados, cujo *corpus* de análise não segue o alinhamento dos bancos de dados constituídos de acordo com a estratificação social baseada em indicadores sociodemográficos amplos.

Os estudos de primeira e segunda ondas, segundo Eckert (2012), têm como foco a descrição da estrutura – um retrato estático. Os estudos de terceira onda incorporam a dinamicidade da estrutura, ou seja, como a estrutura se molda no cotidiano, com os condicionamentos sociais impostos e as relações de poder estabelecidas atuando sobre ela. Eckert (2012) salienta que não está negando a estrutura, mas sim enfatizando o papel da estrutura no condicionamento da prática paralelamente ao papel da prática na produção e reprodução da estrutura, a fim de captar com mais detalhes a dinâmica do valor social das variáveis.

Os estudos de terceira onda combinam os postulados dos estudos de primeira e de segunda onda, com uma mudança no foco: da comunidade de fala para a comunidade de prática. Enquanto, na definição laboviana, comunidades de fala são agrupamentos de indivíduos que compartilham não necessariamente dos mesmos traços linguísticos, mas sim do mesmo juízo de valor acerca desses traços, e os reconhecem como legítimos para a identificação do grupo, a comunidade de prática (WENGER, 1998; ECKERT; MCCONNELL-GINET, 2010; ECKERT; MCCONNELL-GINET, 1997) é um agrupamento de indivíduos (comunidade) que partilham perspectivas em comum, valores e conhecimento (domínio), e que interagem entre si para se aperfeiçoarem e replicarem esses valores e conhecimentos (prática). Trata-se de uma construção social, e, como tal, está sujeita às práticas diárias dos indivíduos, que interagem entre si e com outras comunidades.

Em lugar de conceber o indivíduo como uma entidade à parte, pairando sobre o espaço social, ou como um ponto em uma rede, ou como membro de um conjunto específico ou de um conjunto de grupos, ou como um amontoado de características sociais, precisamos enfocar as comunidades de prática. Tal foco possibilita-nos ver o indivíduo como agente articulador de uma variedade de formas de participação em múltiplas comunidades de prática. (ECKERT; MCCONNEL-GINET, 2010, p.103).

A terceira onda, que se desenvolveu mais recentemente, centra o foco na variação vista não como o reflexo do lugar social num ponto da escala, mas como um recurso para a construção de significado social. Eckert (2012) se volta à necessidade de conectar essas categorias sociais mais abstratas, arraigadas na experiência do falante, com as comunidades imaginárias mais amplas, centrando foco na construção do conceito de comunidade de prática. Uma comunidade de prática é um agregado de pessoas que se juntam para engajar-se em algum empreendimento comum. Na esteira desse engajamento, a comunidade de prática desenvolve meios para fazer coisas que se traduzem em práticas e essas práticas envolvem a construção de uma orientação compartilhada em relação ao mundo em volta – uma definição tácita que os indivíduos assumem um em relação ao outro e em relação a outras comunidades de prática.

Os estudos de terceira onda combinam a metodologia quantitativa, presente nas ondas anteriores, o *corpora* constituídos de modo a contemplar a dimensão mais cotidiana (o que não é necessariamente captado pela entrevista sociolinguística), com observações participantes, por exemplo.

O conceito-chave para o processo de construção é o de prática estilística. Até aqui, nos estudos variacionistas, o estilo tem sido tratado como ajustes à (in) formalidade da situação mediante o uso de variáveis individuais. A face renovada de estilo o identifica com o modo como os falantes combinam variáveis para criar modos distintivos de fala, que fornecem a chave para a construção da identidade. A identidade consiste, por sua vez, em tipos particulares explicitamente localizados na ordem social. Continuamente, os falantes atribuem significado social à variação de um modo consequente, situação que implica certo grau de agentividade.

Eckert (2012) postula que toda variação tem potencial para receber significado social, ainda que nem toda variação seja conscientemente controlada ou mesmo socialmente significativa. A indexação de variáveis fonológicas não é tão transparente quanto, por exemplo, o uso de partículas honoríficas, mas é justamente a fluidez das primeiras que as torna acessíveis a uma grande variedade de propósitos sociais. É necessário haver apenas tempo e continuidade suficientes para convencionar a relação entre uma variável e um significado social. É por essa razão que variáveis estáveis, como a redução de (-ing) no inglês americano, têm

significados tão extremamente claros que podem ser referidos como estereótipos, ao passo que variantes representando mudanças em progresso são recursos mais instáveis, mais transitórios e, por isso, mais disponíveis para assumirem significado social. Um contínuo da convencionalização acompanha um contínuo de intencionalidade, num processo que torna o sujeito agente dos processos sociais que constroem sua própria identidade.

Na linha dos estudos de terceira onda, Moore (2010) analisou a variação entre *were/was* em uma comunidade de prática em Bolton, Inglaterra. Seus resultados globais seguem o padrão da variação *were/was* obtidos por estudos baseados em bancos de dados sociolinguísticos (TAGLIAMONTE, 1998; CHESHIRE; FOX, 2009). Porém, sua metodologia de coleta etnográfica possibilitou captar a correlação entre o uso não padrão e a estrutura social da comunidade de prática, configurando a variável como um índice de prática social.

A investigação de Moore (2010) intenta mostrar como os fatores sociais interagem e avaliar como cada fator condiciona o uso de *were* em contextos de primeira e terceira pessoa do singular. A constituição da amostra se deu em um período de dois anos de observação etnográfica de adolescentes da *Midlan High School*, em Bolton, Inglaterra. Especificamente, foram consideradas as gravações de fala de 39 garotas. Para coletar esses dados, primeiramente a pesquisadora foi à escola no horário do almoço e se envolveu em atividades diversas (como almoçar na cantina, assistir a um ensaio de peça de teatro, sair com os fumantes para a área externa). Nessa etapa foram tomadas notas de campo; interações com as adolescentes só foram gravadas após seis meses de observação etnográfica, e nunca foram realizadas sob a forma de uma entrevista sociolinguística clássica, mas sob a forma de um grupo de discussão e relato de atividades, envolvendo entre duas e quatro adolescentes, o que resultou em 50 horas de gravação. Além disso, questionários circularam entre as participantes do estudo, para coletar informações sobre as práticas sociais das informantes (a fim de validar as observações etnográficas), além de informações sobre classe social, identificação de si mesmo, de família e da sua localidade de nascimento.

A partir dessa observação etnográfica, Moore (2010) identificou quatro comunidades de prática, e cada adolescente membro foi avaliada quanto a com que ocupava seu tempo na escola, com quais atividades ela se engajava, sua orientação em relação aos pares e entorno, seu estilo pessoal e sua aparência e sua avaliação (por si mesma e pelos membros do grupo).

As “populares” exibem uma atitude antiescola, têm um estilo esportivo e feminino de se vestir e se engajam em atividades moderadamente rebeldes, como beber e fumar. Moore (2010) destaca que, no meio do trabalho de campo, as “*townies*” fundiram-se com as “populares”, quando estas começaram a se engajar em atividades de certo risco, como envolvimento com drogas e atividade sexual.

Os amigos das “townies” são rapazes mais velhos, com quem elas ocupam seu tempo na escola. As “geeks” exibem uma atitude positiva em relação à escola, engajando-se em atividades como a banda da escola, esportes etc. Por fim, as “Eden Village”, assim nomeadas em função de residirem em bairro de status, também são orientadas para valores institucionais da escola, e se vestem de acordo com o estilo *teen* da moda, ocupando seu tempo com dança, compras e festas de pijama.

Após analisar os dados coletados considerando fatores linguísticos e sociais, controlando subamostras e especialmente como cada adolescente se identifica e é identificada pelo grupo, Moore (2010) traz evidências quantitativas (baseada na frequência de uso individual) de que a comunidade de prática “townie” tende a fazer uso da forma não padrão (*were*), a comunidade de prática das “populares” se mostra neutra, enquanto a comunidade “geeks” assume uma postura desfavorável à forma não padrão e a comunidade “Eden Village” tem um padrão de uso da forma padrão (*was*) muito próximo do categórico.

Moore (2010) conclui que, para analisar o fenômeno de modo mais eficiente, é preciso observar o que se passa nas relações além do contexto institucional. No caso da variação entre *was* e *were*, os resultados globais se aproximam ao que outros estudos constataram; a distribuição considerando as comunidades de prática permite observar que a identidade não é apenas uma entidade social que é correlacionada a aspectos linguísticos, mas um fenômeno sociolinguístico, que é construído com o valor simbólico de características sociais e linguísticas.

Estudos de terceira onda têm tomado como objeto comunidades de prática variadas, como *yuppies* em Beijin (ZHANG, 2008), *gays* (PODESVA, 2002); PODESVA; ROBERTS; CAMPBELL-KIBLER, 2002), adolescentes em escolas (ECKERT, 2000; MOORE, 2010), que não podem ser aprioristicamente definidas, como vimos na descrição detalhada do estudo de Moore (2010).

O cenário sociolinguístico brasileiro atual vem acenando com entusiasmo para os estudos de terceira onda. Camacho (2010, p.160) diz que “O entendimento que temos da teoria sociolinguística permite assumir que o terceiro ciclo, na visão de Eckert (2005), é o ponto de vista mais consistente com o postulado de que a linguagem é um sistema adaptativo.” É possível que o entendimento de Camacho (2010) seja acertado – apesar de não ser baseado em resultados de estudos empíricos no português –; porém, é preciso reconhecer que os estudos baseados em categorias sociodemográficas amplas (os de primeira onda) são particularmente importantes para respaldar os estudos de terceira onda. Apenas para exemplificação dessa importância, vejamos o estudo de Bentes (2009, p.118-119):

Esta análise, que considera necessária a articulação entre diferentes recursos e níveis de linguagem para a explicação de elaboração de

registros e de estilos linguísticos (sejam eles cultos ou populares), insere-se na agenda de estudos sociolinguísticos da chamada “terceira onda” (ECKERT, 2005), que pretende dar visibilidade aos complexos processos de elaboração de identidades, registros e estilos a partir da manipulação dos recursos das diferenças linguísticas no interior dos grupos sociais. (COUPLAND, 2001; BELL, 2001).

Ao analisar a fala de Mano Brown, Bentes (2009) foca traços linguísticos que associa ao português não padrão, como ausência de concordância explícita de número, e conclui que o comportamento do sujeito quanto ao traço considerado “apenas corrobora a tendência já afirmada em estudos sociolíngüísticos” (NARO; SCHERRE, 2007; SCHERRE; NARO, 2007 apud BENTES, 2009, p.126). Os estudos referidos pela autora tomam por *corpus* amostras de bancos de dados sociolinguísticos – especificamente o PEUL –, o que os caracteriza como estudos de primeira onda. Hora e Wetzels (2011) também destacam a importância dos estudos de primeira onda para os estudos de terceira: ao analisarem os efeitos estilísticos da variação – em uma abordagem que procura se alinhar aos estudos de terceira onda – entre o uso de oclusivas dentais e africadas na fala de João Pessoa, os autores dizem que “Neste estudo, os dados coletados na Paraíba (VALPB) refletem esse momento que Eckert denomina de primeira onda.” (HORA; WETZELS, 2011, p.162). Assim, apesar da visibilidade e da ênfase aos estudos alinhados à tendência da terceira onda dos estudos sociolinguísticos, os estudos de primeira onda – e, particularmente, as abordagens baseadas em bancos de dados sociolinguísticos – continuam a ter um significativo papel e importância na sociolinguística brasileira.

De um modo geral, bancos de dados constituídos de acordo com a metodologia da Sociolinguística são, ainda, importantes fontes para os estudos sociolinguísticos. Faz-se necessário, no entanto, aprimorá-los para contemplar a dimensão da comunidade de prática, do estilo e da *personae*, ou da terceira onda, nos termos de Eckert. Reflexões nessa direção são tecidas na seção a seguir.

Bancos de dados sociolinguísticos no Brasil: potencialidades e limitações

Como vimos, apesar de a dinâmica dos estudos de terceira onda focar relações entre estrutura e prática, a pesquisa sociolinguística baseada em bancos de dados segue mantendo seu espaço no cenário brasileiro, na medida em que possibilita captar tendências amplas em uma comunidade de fala. Essa relação implica, entretanto, a homogeneização da amostra e suas consequências supergeneralizantes.

A discussão sobre a homogeneização da amostra de comunidades de fala não é recente. Severo (2009, p.16) faz um retrospecto, apontando que atualmente

“[...] as pesquisas sociolinguísticas de variação/mudança (incluindo as labovianas) têm valorizado as dimensões micro de estudo, sendo que as unidades de análise deixam de se centrar na comunidade de fala, para integrar as ideias de redes sociais e de comunidades de prática.”

Bancos de dados baseados em comunidades de fala caracterizam-se pela seleção aleatória de seus informantes, que sejam nascidos na comunidade e onde tenham vivido pelo menos 2/3 da vida; que sejam filhos de pais com as mesmas características, além de serem reconhecidos pelos pares como membro da comunidade de fala. A estratificação dos informantes se dá em função de características sociodemográficas (sexo, idade, escolaridade etc.), gerando células sociais (confluência de fatores estratificadores), que devem tender à ortogonalidade (GUY, 2007). Uma amostra estatisticamente representativa da comunidade de fala precisa contar com 0,5% do total da população, margem de erro assumida nas ciências humanas.

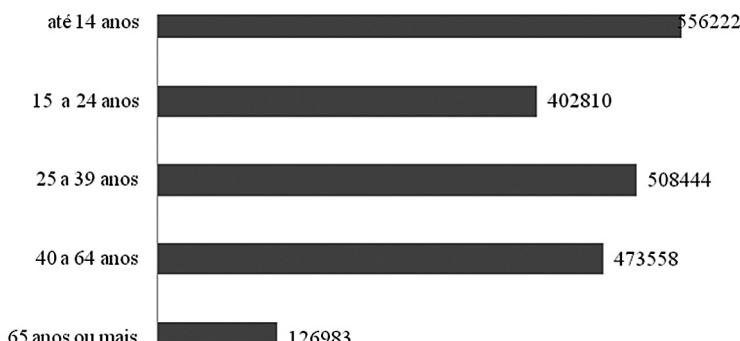
Freitag (2011b) também discute aspectos acerca da metodologia da constituição de amostras para bancos de dados sociolinguísticos:

Se os bancos de dados têm como objetivo subsidiar a descrição de uma dada variedade de língua, e esta descrição, por conta da orientação teórico-metodológica, contempla a dimensão social, será que a estratificação das amostras homogeneizadas, como nos bancos de dados do PEUL e do VARSUL, reflete a estrutura social do Brasil? (FREITAG, 2011b, p.44).

Apenas a título de ilustração, vejamos a aplicação dos critérios de amostragem e representatividade assumidos pelos bancos de dados na aplicação aos bancos de dados *Falantes Universitários de Itabaiana/SE* e *Falares Sergipanos*, que serão apresentados com mais detalhamento na seção a seguir. Suponhamos que um banco de dados seja constituído a partir dos seguintes indicadores sociodemográficos: sexo, idade, escolaridade e zoneamento. Esse perfil social (dois sexos, três escolaridades, cinco faixas etárias e dois zoneamentos) geraria 60 células sociais. Idealmente, são necessários cinco falantes por célula social para garantir a confiabilidade e a representatividade da amostra (MOLLICA; BRAGA, 2004); a condição mínima de constituição de células sociais prevê dois falantes por célula (este é o padrão adotado no banco de dados do projeto VARSUL; o banco de dados do PEUL apresenta amostras com três falantes). Assim, na condição metodológica ideal, a amostragem seria constituída por 300 falantes; na condição mínima, 120 falantes. Agora vejamos o quanto fidedigna é a amostragem constituída: tomando por base os dados do Censo 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística – IBGE (2010), publicados no Diário Oficial da União do dia 04/11/2010 e disponibilizados na internet para consulta interativa – , a população

do município de Itabaiana/SE é de 86.019 habitantes; aplicando-se o corte de 0,5% (o mesmo que Labov usou em Martha's Vineyard), a amostra representativa da comunidade de fala deveria ser constituída por 430 falantes. A população do município de Aracaju/SE é de 552.365 habitantes; aplicando-se o corte de 0,5%, a amostra representativa da comunidade de fala deveria ser constituída por 2.762 falantes. A amostra ideal para Itabaiana/SE está relativamente próxima da amostra estatisticamente significativa; já para Aracaju/SE, a diferença é sensivelmente alta. O quanto fidedigna é uma modelagem que homogeneiza duas amostras de populações sensivelmente diferentes? As distorções da homogeneização vão além: vejamos a distribuição da população do estado de Sergipe quanto à estratificação etária, considerando os dados do Censo 2010, apresentados no gráfico 1.

Gráfico 1 - Distribuição da população do estado de Sergipe por faixas etárias.



Fonte: Elaboração própria com dados do IBGE (2010).

A distribuição da população por faixas etárias apresentada no gráfico 1 aponta para uma redução sensível da faixa etária mais velha (com 65 anos ou mais), com aumento de população na faixa etária de até 14 anos. Se a amostragem do banco de dados respeitasse a proporção da estratificação da população, em um cenário de quatro indivíduos por célula social, a faixa etária de 65 anos ou mais teria apenas um representante, e a faixa etária até 14 anos teria cinco representantes. A opção metodológica por homogeneizar as amostras tem por implicação a restrição, ou ressalva, da generalização dos resultados, embora reconheçamos trabalhos que assumam a supergeneralização. As ponderações acima não significam a condenação dos bancos de dados constituídos nessa perspectiva; ao contrário, são registros sistemáticos e altamente produtivos para a identificação de tendências na comunidade, motivo que reforça a necessidade de continuidade desse trabalho de armazenagem, mas não sem uma readequação metodológica de

modo a contemplar as premissas da terceira onda. Na seção a seguir, discutimos aspectos metodológicos que devem ser considerados na constituição de novos bancos de dados sociolinguísticos (e também na ampliação dos já existentes), de modo que seja possível a comparabilidade entre as amostras.

Como vimos destacando, bancos de dados linguísticos têm sido fonte privilegiada para a descrição do português brasileiro e a tendência recente dos estudos de terceira onda ratifica sua importância para apontarem tendências linguísticas na comunidade. Seguindo as premissas da confiabilidade e da intersubjetividade (BAILEY; TILLERY, 2004), para dar continuidade a essa prática produtiva, novos bancos de dados precisam conservar minimamente as estratificações dos bancos de dados já existentes, pois a comparação de dados em tempo real permite análises mais acuradas com estudos de painel e de tendência (LABOV, 2001). Assim, as variáveis demográficas amplas – sexo, idade, escolarização etc. – precisam continuar a ser controladas nas novas coletas e nas novas amostras constituídas; é desejável, entretanto, que as novas coletas aprimorem o controle do falante, suas características individuais e de práticas, de modo a permitir que se construa um perfil social que contemple indicadores sociodemográficos mais amplos e abstratos. Um exemplo desse tipo de controle é o banco VarX: “[...] a construção do VarX surgiu da necessidade de se estudar com mais profundidade aspectos referentes a classes sociais (ocupação/profissão, renda/patrimônio e escolaridade) e suas implicações linguísticas.” (AMARAL, 2003, p.63). Nesse banco de dados, informações mais acuradas sobre a região onde o falante mora, se o tipo de sua casa corresponde ao padrão do zoneamento, que tipos de bem de consumo possui, se sua profissão é manual, técnica ou intelectual, correlacionadas ao nível de escolarização, possibilitam um enquadramento mais próximo da realidade social do informante, desfazendo a homogeneização da amostra. O banco de dados VarX estratifica esses indicadores, mas a simples inclusão de mais questões relacionadas a fatores subjetivos nos questionários/roteiros de entrevistas sociolinguísticas já torna possível que esse tipo de informação seja extraída das amostras. É nessa linha também que vem seguindo o banco de dados Português Paulistano (MENDES, 2011), cujo objetivo é a constituição de um *corpus* contemporâneo do português paulistano que permita a sua descrição e análise nos moldes da sociolinguística variacionista.

Os novos bancos de dados sociolinguísticos têm também investido em coletas que privilegiam a diversidade de tipos/sequências textuais, de modo a captar estilos linguísticos mais próximos do dia a dia; é o caso do projeto Amostra Linguística do Interior Paulista – ALIP, cujo banco de dados Iboruna é constituído por amostras do português falado na região de São José do Rio Preto, e cidades circunvizinhas, na região noroeste do Estado de São Paulo.

Em relação à tendência de homogeneização da amostra, é importante destacar que o Iboruna é um banco de dados em que a distribuição dos informantes por célula social se dá proporcionalmente à densidade populacional das cidades da região, como podemos ver no quadro 1.

Quadro 1 - Distribuição dos informantes do bando de dados IBORUNA.

Cidades da Região de São José do Rio Preto	População	Número de informantes
1. Bady Bassit (12 km ao sul de SJRP)	11.475	04
2. Cedral (14 km, ao sul de SJRP)	6.690	02
3. Guapiaçu (16 km, ao leste de SJRP)	14.049	05
4. Ipiguá (18 km, ao norte de SJRP)	3.461	01
5. Mirassol (14 km, a oeste de SJRP)	48.233	16
6. Onda Verde (25 km, ao norte de SJRP)	5.407	02
7. São José do Rio Preto	357.705	122
Total da população representada	447.020	152

Fonte: Gonçalves (2008, p.2729).

O cuidado em dimensionar a amostra de modo a garantir uma representação proporcional da população, assim como as informações do perfil dos falantes quanto a fatores sociodemográficos mais amplos colaboram para a confiabilidade e a intersubjetividade da análise, de acordo com as recomendações de Bailey e Tillary (2004), apresentadas na introdução deste texto.

Para respaldar estudos de terceira onda, novos bancos de dados precisam também promover a mudança do foco da comunidade de fala para a comunidade de prática. Eckert e McConnel-Ginet (2010) explicitam o que entendem por comunidades de prática:

Uma comunidade de prática pode ser constituída por pessoas trabalhando juntos em uma fábrica, *habitués* de um bar, companheiros de brincadeira em uma vizinhança, a família nuclear, parceiros policiais e seu etnógrafo, a Suprema Corte etc. Comunidades de prática podem ser grandes ou pequenas, intensas ou difusas; elas nascem e morrem, podem sobreviver a muitas mudanças de membros e podem estar intimamente articuladas a outras comunidades. As pessoas participam de múltiplas comunidades de prática, e a identidade individual é baseada nesta participação. Em lugar de conceber o indivíduo como uma entidade à parte, pairando sobre o espaço social, ou como um ponto em uma rede, ou como membro de um conjunto específico ou de um conjunto de grupos, ou como um amontoado de características sociais, precisamos enfocar as comunidades de prática. Tal foco possibilita-nos ver o indivíduo como agente articulador de uma variedade de formas de participação em múltiplas comunidades de prática. (ECKERT; MCCONNEL-GINET, 2010, p.102-103).

A constituição de *corpus* de comunidade de prática permite, por exemplo, a depender do tamanho da comunidade, que sejam considerados todos os indivíduos. Permite, também, que se proceda ao mapeamento acurado das redes de relacionamento, observando os graus de integração dos indivíduos dentro da comunidade de fala.⁵

Apesar dos pontos favoráveis apresentados à abordagem de comunidades de prática, a sua implementação única e exclusiva como fonte para os estudos sociolinguísticos não é benéfica; além de quebrar a série histórica da comparabilidade de amostras de comunidades de fala (com coletas iniciadas na década de 1980, como vimos na seção anterior), a abordagem de comunidades de prática, sem um estudo anterior baseado em comunidades de fala para levantar a(s) tendência(s) ampla(s) a ser(em) analisada(s), é um tiro no escuro que pode ou não resultar em uma boa investigação sociolinguística. Cada uma das abordagens apresenta suas particularidades e suas especificidades, como sistematizamos no quadro 2.

Quadro 2 - Comparaçāo entre abordagens sociolinguísticas de comunidades de fala e de comunidades de práticas.

Abordagem de comunidade de fala	Abordagem de comunidade de práticas
- estratificação baseada em fatores sociodemográficos amplos	- estratificação baseada em valores localmente estabelecidos
- distribuição homogênea, tanto quanto ao tamanho quanto às categorias controladas	- distribuição variável, definida caso a caso
- categorias definidas a priori	- categorias definidas a posteriori
- permissão para captar tendências amplas da comunidade	- permissão para captar valores sociais localmente estabelecidos nas relações
- coleta padronizada (entrevista sociolinguística)	- coleta etnográfica (observação participante, interações entre grupos)
- constituição da amostra em curto prazo	- constituição da amostra em longo prazo

Fonte: Elaboração própria.

Como podemos facilmente constatar, cada uma das abordagens apresenta peculiaridades que não permitem a implementação simultânea, por conta de incompatibilidades de natureza teórico-metodológica. Entendemos, no entanto,

⁵ Milroy e Gordon (2003) diferenciam as abordagens de investigação de redes sociais das de comunidades de prática: enquanto, em redes sociais, o objetivo é identificar o nó social que é importante para o indivíduo; em comunidades de prática, o objetivo é identificar o agrupamento que forma o local da prática linguística e da prática social.

que é viável a articulação entre as abordagens, e que essa articulação deve constar como proposta programática nas novas coletas de dados. O ponto de partida é, no entanto, a abordagem de comunidades de fala, pois é a partir dessa coleta que é possível delinear as tendências amplas da comunidade (e garantir a comparabilidade entre amostras, na medida que é adotada uma metodologia já consolidada) e captar pistas para definir abordagens de comunidades de prática.

Nas seções a seguir, nos dedicamos à apresentação de três novos bancos de dados no nordeste brasileiro – Bancos de fala culta de Itabaiana/SE, Falares Sergipanos, e FALA-Natal –, contemplando aspectos metodológicos da terceira onda de estudos sociolinguísticos.

Banco de fala culta de Itabaiana/SE

Localizada na região do agreste central sergipano, a cidade de Itabaiana é a cidade mais importante do Estado de Sergipe fora da região da Grande Aracaju. Abrigando uma central de abastecimento e uma feira de porte significativo para a região, a cidade é conhecida por sua fama de “comércio forte”, atuando como entreposto comercial na circunvizinhança. O itabaianense tem uma atitude muito positiva quanto a si e quanto à sua cidade (FREITAG; SANTOS; SANTOS, 2009). Apesar de já haver um polo universitário particular, a implantação do *campus* de Itabaiana da Universidade Federal de Sergipe, em 2006, decorrente do programa do governo federal de expansão e interiorização da educação superior no Brasil, provocou grandes alterações na cidade de Itabaiana e circunvizinhança. Com oferta anual em dez cursos de graduação, escolhidos de acordo com as peculiaridades e necessidades da região (nas áreas de gestão e educação), a estrutura do *campus* de Itabaiana recebe diariamente cerca de 2.500 alunos, em três turnos de funcionamento, com concentração no período noturno. Ser universitário é uma conquista familiar da maioria: pesa a responsabilidade de ser o primeiro universitário em uma família de pais que não tiveram oportunidade de acesso à escolarização. Os primeiros exames vestibulares de Itabaiana tiveram uma concorrência superior à concorrência dos mesmos cursos no campus sede da Universidade Federal de Sergipe, trazendo de volta aos estudos alunos que estavam no mercado de trabalho. Ser universitário da Universidade Federal de Sergipe em Itabaiana é um diferencial para esses indivíduos em seus nichos familiares; traz a responsabilidade e o compromisso com o estudo. Durante pelo menos quatro horas por dia, esses estudantes travam contatos próximos, compartilhando valores e conhecimentos (não só dos seus cursos específicos, mas do saber universitário e sua função social). A estrutura da unidade acadêmica – multicursos – facilita o contato e a interação entre todos os alunos; além disso, por muitos virem de cidades circunvizinhas (algumas distantes mais de 50 km de Itabaiana, como é

o caso do município de Carira, na divisa entre Sergipe e Bahia), além do contato na universidade, há o contato durante o trajeto, em ônibus de transporte escolar (muitos subsidiados pelas prefeituras). Esse vínculo entre os universitários fica explícito materialmente por meio da identificação no vestuário: embora não seja obrigatório o uso de uniforme, os graduandos de cada curso se unem e elaboram a camiseta do seu grupo (que pode ser do curso todo, de uma turma do curso ou de parte de uma turma), que é utilizada diariamente como um uniforme, mas com a finalidade de marcar a identidade e o pertencimento ao grupo não entre os pares universitários, mas nas suas redes de relacionamento de origem. O acesso à universidade propicia, também, oportunidades de inserção nos programas institucionais remunerados (iniciação científica, iniciação à docência, extensão, monitoria etc.), de caráter meritocrático. A participação nesses programas é vista como positiva e gera expectativas de continuidade e ascensão nos estudos, alcançando os universitários rumo à pós-graduação.

O engajamento social verificado entre os universitários do *campus* de Itabaiana da Universidade Federal de Sergipe nos permite defini-los como constituintes de uma comunidade de prática, nos termos do que propõem Eckert e McConnell-Ginet (2010). Não há um limite geográfico específico para definir essa comunidade, mas um limite de comportamento, preservação e compartilhamento de valores associados ao ser universitário de uma instituição pública no interior. Em termos linguísticos, ser universitário pressupõe a passagem por 11 anos de escolarização formal (ou, em casos de exames supletivos, demonstrar o domínio de conteúdos equivalente a esse tempo de escolarização), em que há um contato direto com a cultura letrada e, por hipótese, o domínio da norma culta da língua. Na sociedade brasileira, o ser universitário está associado ao saber (havia um programa de televisão no formato pergunta-resposta que facultava aos participantes dispor da “ajuda universitária”). Pelo significado social que possui, pelo tipo de exigência do próprio contexto (que prioriza seminários, exposições orais, debates etc. como formas de avaliação), o ambiente universitário é, pois, um dos espaços onde a norma culta falada se manifesta. Ainda que em outros momentos, no seu núcleo familiar, por exemplo, o universitário faça uso de uma variedade linguística marcada por traços estigmatizados (o rotacismo, por exemplo), e ainda que o seu colega também faça uso desse traço, enquanto estiverem desempenhando seus papéis nessa comunidade de práticas universitária, ambos tendem ao monitoramento para a não ocorrência do traço em questão. O banco de dados Falantes Cultos de Itabaiana/SE (CAAE - 0301.0.107.000-11) foi constituído tomando por base o constructo de comunidade de prática. É composto por 20 entrevistas, estratificadas por gênero do entrevistador e do entrevistado e abrangendo a faixa etária dos 19 aos 32 anos, realizadas de acordo com protocolo da entrevista sociolinguística (envolvendo temas como risco de vida, narrativas da infância etc.), mas com o diferencial de que o entrevistador e o entrevistado são membros da comunidade de

prática sob análise, o que faz com que a assimetria seja sensivelmente diminuída. Outro fator importante a ser considerado é que todos os entrevistados e todos os entrevistadores se conhecem, o que não costuma acontecer, por exemplo, em coletas em comunidades de fala.⁶ Por haver esse envolvimento, a condução da interação em alguns momentos sai do molde da entrevista sociolinguística, especialmente naqueles em que o tópico discorrido está relacionado com o que constitui a comunidade de prática: o que é ser o primeiro universitário da família, a oportunidade de ser universitário em Itabaiana, prospecções para o futuro. Esses temas são comuns a entrevistado e a entrevistador e, em alguns momentos, a interação assume a forma de desabafo, confidênciia, compartilhamento, o que é mais uma evidência em favor da constituição efetiva da comunidade de prática.

É importante destacar que o banco de dados não foi constituído para subsidiar, em princípio, estudos ditos de terceira onda. Seu propósito inicial foi dar suporte ao estudo do fenômeno de variação e mudança em categorias verbais do português, em uma perspectiva baseada em frequências de uso.⁷ A modelagem da amostra, porém, permite não só estudos de cunho quantitativo, mas também estudos de caráter mais etnográfico, na medida em que as entrevistas são ricas em informações sobre o ser universitário. O banco de dados encontra-se em fase de revisão e em breve será disponibilizado à comunidade científica como mais uma fonte para estudos descritivos de variedades do português falado.

Banco de falares sergipanos

A constituição e/ou ampliação de bancos de dados sociolinguísticos, contemplando uma variedade do português brasileiro ainda não mapeada (ou pouco mapeada), como é o caso de Sergipe, é altamente desejável, motivo que levou à proposição do banco de dados *Falares Sergipanos* (CAAE - 0386.0.107.000-11). Por entendermos as ondas propostas por Eckert (2012) não como suplementares, mas complementares, a constituição de novos bancos de dados não pode abrir mão da comparabilidade com os bancos de dados já constituídos. Nessa linha de raciocínio, o banco de dados Falares Sergipanos

⁶ Deve-se destacar que o corpus do projeto SP2010 é constituído pelo critério “bola de neve” de seleção de informantes (MILROY; GORDON, 2003) – um indica o outro –, procedimento que não se verifica em outros bancos de dados, a exemplo do VARSUL.

⁷ Projeto Variação na expressão do tempo verbal passado na fala e escrita de Itabaiana/SE: funções e formas concorrentes (FREITAG, 2009), financiado pela FAPITEC (Edital FAPITEC/FUNTEC-SE Universal 06/2009 Processo n. 019.203.00910/2009-0) e CNPq (Edital MCT/CNPq/MEC/CAPES 02/2010 - Ciências Humanas, Sociais e Sociais Aplicadas Processo n. 401564/2010-0).

(FREITAG, 2011a),⁸ em processo de implementação, segue duas linhas de coleta – a de estratificação homogeneizada e a de comunidades de prática.

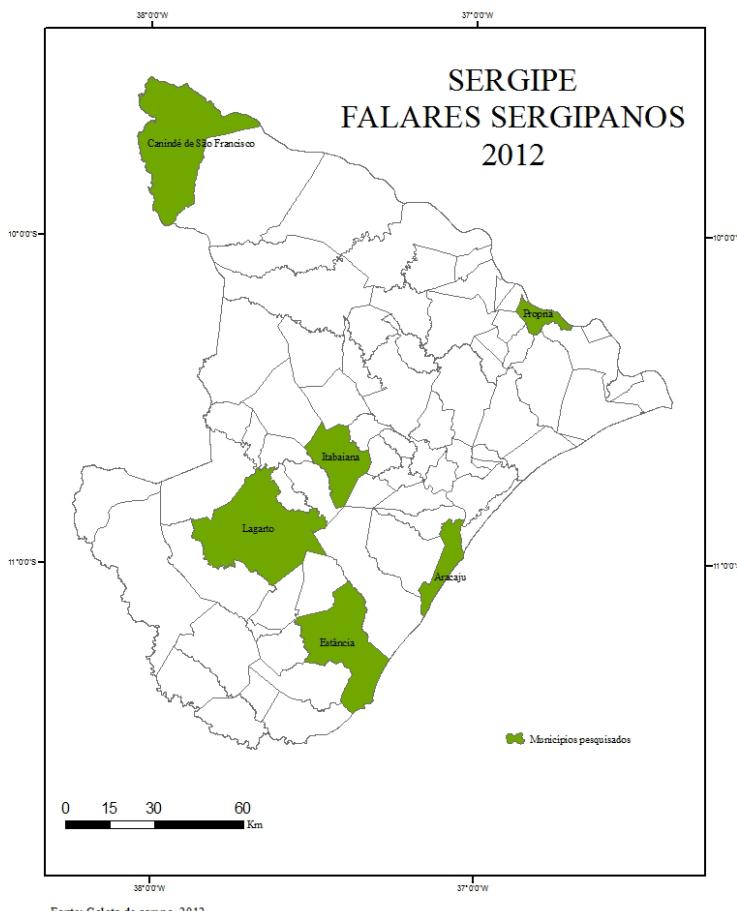
A estratificação homogeneizada é predominante nos bancos de dados já constituídos, como apresentamos em seção anterior. Para o dimensionamento da amostra, foram selecionadas seis cidades representativas do estado de Sergipe, por territórios – Canindé de São Francisco, Itabaiana, Lagarto, Estância, Propriá e Aracaju (figura 1). A estratificação etária dos informantes segue a padronização do IBGE (2010), computando cinco faixas (até 14 anos; 15-24; 25-39; 40-64; mais de 65 anos). A seleção dos informantes (inicialmente dois para cada célula social) seguirá a abordagem “bola de neve”, a partir do contato inicial de pesquisador de campo da comunidade, o que será viabilizado pelo fato de a Universidade Federal de Sergipe contar com o curso de Letras na modalidade a distância e a disciplina obrigatória Sociolinguística. Dadas as dimensões do estado e o fato de haver 14 polos universitários, alguns dos quais nas cidades escolhidas para a constituição da amostra, não está prevista, inicialmente, a estratificação por nível de escolarização, o que será feito depois, ao ritmo da coleta, a partir de mapeamento qualitativo. Desse modo, a coleta inicial fornecerá 40 entrevistas por cidade, totalizando 240 entrevistas.

Paralelamente à coleta de estratificação homogeneizada, serão realizadas coletas voltadas para: i) identificação de comunidades de prática, aos moldes do banco de dados Falantes Cultos de Itabaiana/SE, apresentado anteriormente; e ii) observação de efeitos de sexo/gênero, com coletas de dados (entrevistas sociolinguísticas) realizadas por dois entrevistadores: um homem e uma mulher, considerando que Bailey e Tillery (2004) relatam que há estudos que apontam indícios de que o sexo/gênero do entrevistador influencia nos traços linguísticos do entrevistado, propiciando assim análise mais acurada acerca da variação estilística.

Ademais, a partir da identificação de tendências amplas na amostra de estratificação homogeneizada, podem ser realizadas coletas de dados mais particularizadas, que não podem ser definidas aprioristicamente, mas sim a partir da observação empírica de uma realidade.

⁸ O projeto Falares Sergipanos integra um projeto maior, intitulado Da expressividade da língua ao mal na literatura: base de pesquisas interinstitucionais do PPGL/UFS, financiado pelo edital CAPES/FAPITEC/SE 06/2012, que, em parceria com o Programa de Pós-Graduação em Linguística da Universidade Federal de Santa Catarina (PPGLg/UFSC) e o Programa de Pós-Graduação em Literatura da Universidade Federal de Minas Gerais (Pós-Lit/UFMG), tem como um dos seus objetivos constituir o banco de dados Falares Sergipanos para respaldar a pesquisa da área de concentração Estudos Linguísticos do PPGL/UFS para descrição e estudos comparativos do português e aplicações para o ensino de língua estrangeira e materna.

Figura 1 - Distribuição das cidades constituintes do banco de dados Falares Sergipanos.



Fonte: Freitag (2011a, p.04).

Banco de Dados FALA-Natal

Como já dito, estudos feitos sob a égide da sociolinguística variacionista vêm fomentando a ampliação do conhecimento sobre o português brasileiro desde a década de 1970, através da descrição e da análise de fenômenos variáveis nos âmbitos fonológico, morfológico, sintático, semântico e discursivo. Contudo, há

estados da federação em que tais pesquisas são ainda incipientes ou mesmo inexistentes. É o caso do Rio Grande do Norte, que não conta com um banco de dados de fala com as características necessárias para a pesquisa sociolinguística. Para suprir essa lacuna, Tavares e Martins (2012) propuseram-se a organizar um *corpus* de fala, que será denominado Banco de Dados da Fala do Rio Grande do Norte (FALA-RN) e contará com amostras representativas de diferentes comunidades de fala norte-rio-grandenses. O marco inicial da organização do FALA-RN será a constituição do Banco de Dados FALA-Natal, que congregará entrevistas sociolinguísticas a serem feitas com membros da comunidade de fala do município de Natal, que é a capital e maior centro urbano do estado potiguar. Posteriormente, serão coletadas entrevistas sociolinguísticas em comunidades de fala do interior.

Os informantes do Banco de Dados FALA-Natal serão socialmente estratificados de modo similar a informantes de bancos de dados já existentes no país, a exemplo do PEUL, do VARSUL e do VALPB. Inicialmente, o Banco de Dados FALA-Natal será composto por 48 entrevistas sociolinguísticas com cerca de 60 minutos de duração. Essas entrevistas serão distribuídas, em termos de estratificação social, quanto ao *sexo* (24 informantes de sexo feminino e 24 informantes de sexo masculino), *idade* (12 informantes de 8 a 12 anos, 12 informantes de 15 a 21 anos, 12 informantes de 25 a 50 anos e 12 informantes de mais de 50 anos) e *nível de escolaridade* (12 informantes com ensino fundamental I completo, 12 informantes com ensino fundamental II completo e 12 informantes com ensino médio completo, além de 12 informantes cursando o ensino fundamental I – os indivíduos de 8 a 12 anos). Serão gravados informantes de diferentes bairros das quatro zonas de Natal.

De acordo com o Censo de 2010, a capital norte-riograndense tem, atualmente, 803.739 habitantes (IBGE, 2010). Se considerarmos uma amostragem na condição metodológica ideal, aplicando o corte de 0,5% da população, o banco de dados FALA-Natal deveria contar com 4.019 entrevistas. Em condições reais, o desenvolvimento de um banco de dados com esse número de entrevistas demandaria anos de realização. Com o significativo crescimento da população, se adotássemos tal condição para o desenvolvimento do banco, quando a última entrevista fosse realizada, a comunidade, com certeza, já não seria a mesma. Além disso, a quantidade de informantes também depende de financiamento e de quanto tempo se dispõe para a organização do banco de dados, fatores que, em geral, impedem a coleta de um grande número de entrevistas.

De qualquer forma, um número menor de entrevistas pode ser representativo de tendências gerais da comunidade. Segundo Sankoff (1988, apud TAGLIAMONTE, 2006, p.23), é necessário “[...] não que a amostra seja uma versão em miniatura da população, mas apenas que tenhamos a possibilidade de fazer inferências sobre a população com base na amostra.” Cada banco de dados deve ter um

mínimo de representatividade com base em idade, sexo, classe social e/ou nível de educação, o que assegura que a diversidade linguística da comunidade de fala esteja representada na amostra.

Lembramos que a maior coleta de entrevistas sociolinguísticas já feita foi dirigida por Shuy et al., tendo sido gravadas 702 entrevistas em Detroit, nos Estados Unidos. No entanto, as análises mais detalhadas desse *corpus* utilizaram apenas 48 dessas entrevistas, com os informantes distribuídos simetricamente em quatro classes sociais, em um total de 12 informantes por classe (TAGLIAMONTE, 2006).

No caso do Brasil, os bancos de dados costumam ter de dois a três informantes por célula social, o que tende a ser suficiente para a obtenção dos padrões gerais de variação de uma comunidade de fala no que diz respeito a diversos fenômenos variáveis. Quanto ao Banco de Dados FALA-Natal, caso algumas características de uso linguístico variável chamem, por alguma razão, a atenção no conjunto das 48 entrevistas sociolinguísticas iniciais, outras entrevistas poderão ser realizadas – com os mesmos ou outros informantes –, no sentido de possibilitar uma análise mais refinada desses usos.

Uma vez coletadas e armazenadas, as entrevistas integrantes do Banco de Dados FALA-Natal poderão servir de *corpus* para pesquisas que objetivem: i) a descrição e a análise da fala de Natal; ii) a comparação com outros dialetos brasileiros, com o intuito de descrever o português brasileiro de modo mais abrangente e detalhado, e de observar diferenças e semelhanças interdialetais; iii) a comparação com outras vertentes do português, como a europeia; iv) a testagem de teorias linguísticas; v) investigações de natureza social, histórica, antropológica, psicológica, entre outras.

Na constituição do banco de dados FALA-Natal, que está em desenvolvimento,⁹ temos nos defrontado com uma série de questões para as quais temos buscado soluções. Entre essas questões, apontamos: i) representatividade da amostra; ii) dificuldade de localização de informantes com certos traços socioeconômicos; iii) necessidade de maior diferenciação de faixas etárias para testar hipóteses relativas à aquisição e à mudança linguística; iv) estratégias para tornar acessíveis à comunidade acadêmica os bancos de dados sociolinguísticos (e as questões éticas aí implicadas); v) validade da comparação de análises realizadas com base em dados extraídos de entrevistas sociolinguísticas feitas recentemente com análises realizadas com base em dados extraídos de entrevistas sociolinguísticas feitas há dez ou vinte anos; vi) como considerar aspectos relacionados à questão da análise estilística pelo viés da terceira onda, nos termos de Eckert.

⁹ Estamos em fase de elaboração da estrutura do banco de dados e de seleção de informantes, bem como de treinamento da equipe que realizará as entrevistas. A previsão é que sejam gravadas até março de 2013 as 48 entrevistas que comporão o banco de dados em sua fase inicial.

Em relação a esse último tópico, inicialmente serão coletadas entrevistas sociolinguísticas em uma comunidade de fala ampla – a de Natal – para que seja possível a realização de mapeamentos de tendências gerais de variação e de mudança em relação a essa comunidade, ou seja, em sua primeira fase, o Banco de Dados Fala Natal será composto por entrevistas que permitirão a realização de estudos alinhados à primeira onda da sociolinguística.

Todavia, nossa comunidade de fala alvo abriga, naturalmente, inúmeras comunidades de prática. Com a intenção de aprofundarmos nosso conhecimento acerca das comunidades de prática em que se engajam cada um dos informantes a serem selecionados para o banco de dados, elaboramos uma ficha social a ser preenchida previamente à entrevista na qual constam, entre outras, questões que permitem a obtenção de informações a respeito das diferentes comunidades de prática em que se engaja o informante em sua vida cotidiana. Nessa ficha social, solicitamos, por exemplo, para os informantes de 15 a 21 anos, que respondam às seguintes questões: (i) **Como ocupa seu tempo livre?** e (ii) **Participa de algum grupo (igreja/ jovens/ esporte/ clube)? Se sim, com que frequência?**

Também poderão ser propostos, nas entrevistas, tópicos que estimulem o informante a discorrer sobre as diferentes comunidades de prática das quais faz parte. Com esse fim, elaboramos um roteiro para as entrevistas com sugestões de perguntas que o entrevistador pode fazer ao entrevistado. Entre essas perguntas, estão questões do tipo: (i) **Com quem você passa o tempo, além das pessoas da sua família? O que vocês fazem juntos? Que tipo de lazer vocês têm?** (ii) **Você participa de algum trabalho voluntário? Como é?** (iii) **Você participa de algum grupo de jovens? O que vocês fazem juntos?** (iv) **Você participa de algum grupo da igreja? Como é?** (v) **Você frequenta algum clube? Qual? Como é?** (vi) **Algo interessante já aconteceu no clube/grupo de jovens/grupo da igreja quando você estava lá? O que foi?** (vii) **Descreva o que você faz em um dia, desde que acorda até ir dormir.**

Esse maior conhecimento sobre as comunidades de prática em que se integra cada informante que será obtido com base nas fichas sociais e nas próprias entrevistas poderá ser levado em conta na análise dos fenômenos variáveis. Todavia, as informações presentes nas fichas sociais e nas entrevistas não apenas fornecerão subsídios para uma caracterização mais aprofundada de cada informante no que tange a traços sociais e de prática, como também trarão indícios a respeito de quais comunidades de prática – entre as inúmeras de que participa cada indivíduo – são mais importantes para a realização de estudos nos moldes das segunda e terceira ondas da sociolinguística. Esses indícios fundamentarão as etapas posteriores de construção do Banco de Dados FALA-Natal, em que serão coletadas entrevistas adequadas para contemplar as duas últimas ondas.

Ou seja, para a organização do Banco de Dados FALA-Natal, estamos conscientes da necessidade de organizar não somente um conjunto de entrevistas sociolinguísticas que possibilitem a realização de pesquisas afiliadas à abordagem variacionista alinhada a Labov (um retrato amplo de comunidades de fala definidas geograficamente), mas também à abordagem etnográfica alinhada a Milroy (um retrato local, etnográfico, de comunidades de fala definidas geograficamente) e à abordagem da identidade social alinhada à Eckert (um retrato do(s) indivíduo(s) integrante(s) de comunidades de prática, pelo viés do estilo como elemento central de constituição da *persona*). Pretendemos, pois, num futuro próximo, tornar disponíveis fontes de dados viáveis para pesquisas encaixadas em qualquer uma das três ondas da sociolinguística.

Considerações finais

Como destacamos na introdução, a constituição de um banco de dados sociolinguístico é tarefa dispendiosa e ao mesmo tempo altamente produtiva, por subsidiar estudos de fenômenos variáveis em diferentes níveis linguísticos, com diferentes interfaces teóricas, para vários pesquisadores. Considerando o estado da arte da Sociolinguística no Brasil, e observando a tendência ao direcionamento para estudos de terceira onda (ECKERT, 2012), manifestada já por pesquisadores brasileiros, mas ainda não implementada de forma plena, defendemos que os novos bancos de dados devem, sim, contemplar aspectos relacionados a esta abordagem, mas sem abandonar a tradição consolidada de bancos de dados de estratificação homogeneizada baseados em indicadores sociodemográficos amplos, ditos de primeira onda.

Dado que o interesse em metodologia é uma característica da Sociolinguística e que, como dizem Bailey e Tillary (2004), nem sempre a confiabilidade e a intersubjetividade têm prevalecido nos estudos sociolinguísticos de cunho variacionista, defendemos a necessidade de continuidade de uma metodologia de constituição de *corpus* já consolidada, a fim de permitir a comparação entre amostras, inclusive em tempo real.

FREITAG, R. M.; MARTINS, M. A.; TAVARES, M. A. Brazilian Portuguese sociolinguistic databases and third wave studies: potentialities and limitations. *Alfa*, São Paulo, v.56, n.3 p.907-934, 2012.

- *ABSTRACT: Spoken linguistic databases – especially those designed for research with a Variationist Sociolinguistics approach – has been a privileged source for the description of Brazilian Portuguese. In this paper, we discuss methodological procedures that should be adopted in the development of new databases. We trace a short retrospect on already established databases and propose the collection and expansion of corpora from different speech communities – and from different communities of practice. Eckert (2012) describes*

the three analytical practices most commonly embraced by sociolinguistics studies. Each one of these “three waves” (in Eckert’s terms) reflects distinct ways of approaching linguistic variation. We suggest strategies to standardize procedures in the development of databases taking into account the three waves of sociolinguistic inquiry, and we shed additional light on the third wave, still incipient in Brazil. The standardization of sociolinguistics databases would make contrastive investigations of different Brazilian dialects easier, contributing, in this way, to the proposition and refinement of generalizations and universal principles of variation and change.

- KEYWORDS: *Sociolinguistics. Databases. Linguistic variation and change. Social factors. Style.*

REFERÊNCIAS

- AMARAL, L. C. *A concordância verbal de segunda pessoa do singular em Pelotas e suas implicações linguísticas e sociais*. 2003. 203f. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, 2003.
- BAILEY, G.; TILLERY, J. Some sources of divergent data in sociolinguistics. In: FOUGHT, C. *Sociolinguistic variation: critical reflections*. New York: Oxford University, 2004. p.11–30.
- BENTES, A. C. Tudo que é sólido desmancha no ar: sobre o problema do popular na linguagem. *Gragoatá*, Niterói, v.27, p.12-47, 2009.
- BISOL, L.; MENON, O. P. S.; TASCA, M. VARSUL, um banco de dados. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Julius Naro e a linguística no Brasil: uma homenagem acadêmica*. Rio de Janeiro: 7 Letras, 2008. p.50-58.
- CAMACHO, R. G. Uma reflexão crítica sobre a teoria sociolinguística. *DELTA*, São Paulo, v.26, n.1, p.141-162, 2010.
- CHESHIRE, J.; FOX, S. Was/were variation: a perspective from London. *Language, Variation and Change*, Cambrigde, v.21, p.1-38, 2009.
- ECKERT, P. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, Palo Alto, n.41, p.87-100, 2012.
- _____. *Linguistic variation as social practice*. Oxford: Blackwell, 2000.
- ECKERT, P.; MCCONNELL-GINET, S. Comunidades de práticas: lugar onde co-habitam linguagem, gênero e poder (1992). In: OSTERMANN, A. C.; FONTANA, B. (Org.). *Linguagem, gênero, sexualidade: clássicos traduzidos*. São Paulo: Parábola, 2010. p.93-108.

_____. Communities of practice: where language, gender and power all live. In: COATES, J. (Ed.). *Language and gender*: a reader. Oxford: Blackwell, 1997. p.484-494.

FERRARI, L. V. *Variação linguística e redes sociais no Morro dos Caboclos*. 1994. 204f. Tese (Doutorado em Linguística) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1994.

FREITAG, R. M. K. *Banco de dados de falares sergipanos*. 2011. Projeto de Pesquisa, Universidade Federal de Sergipe, Sergipe, 2011a.

_____. O social da sociolinguística: o controle de fatores sociais. *Diadorim*, Rio de Janeiro, v.8, p.43-58, 2011b.

_____. *Variação na expressão do tempo verbal passado na fala e escrita de Itabaiana/SE*: funções e formas concorrentes. 2009. Projeto de Pesquisa, Universidade Federal de Sergipe, Sergipe, 2009.

FREITAG, R. M. K.; SANTOS, J. C.; SANTOS, S. “Fio do canço”: marca linguística identitária do itabaianense. *InterSciencePlace*, [S.l.], v.5, p.1-13, 2009. Disponível em: <<http://www.interscienceplace.org/interscienceplace/article/viewArticle/55>>. Acesso em: 23 mar. 2012.

GONÇALVES, S. C. L. Projeto ALIP (Amostra Linguística do Interior Paulista). In: MAGALHÃES, J. S.; TRAVAGLIA, L. C. (Org.). *Múltiplas perspectivas em linguística*. Uberlândia: Ed. da UFU, 2008. p.2726-2739. Disponível em: <http://www.filologia.org.br/ileel/artigos/artigo_478.pdf>. Acesso em: 22 mar. 2012.

GUY, G. R. Introdução à análise quantitativa da variação linguística. In: GUY, G. R.; ZILLES, A. M. *Sociolinguística quantitativa*: instrumental de análise. São Paulo: Parábola, 2007. p.19-46.

HORA, D.; WETZELS, L. A variação linguística e as restrições estilísticas. *Revista da ABRALIN*, Brasília, n. esp., p.147-188, 2011. Disponível em: <<http://www.abralin.org/revista/RVE1/v4.pdf>>. Acesso em: 22 mar. 2012.

IBGE. *Censo demográfico 2010*. Brasília, 2010. Disponível em: <<http://www.censo2010.ibge.gov.br>>. Acesso em: 22 mar. 2012.

LABOV, W. *Principles of linguistic change*: social factors. Oxford: Blackwell, 2001.

MENDES, R. B. *SP-2010*: construção de uma amostra da fala paulistana. 2011. Projeto de Pesquisa, Universidade de São Paulo, São Paulo, 2011.

MILROY, L.; GORDON, M. *Sociolinguistics*: method and interpretation. Oxford: Blackwell, 2003.

MOLLICA, M. C.; BRAGA, M. L (Org.). *Introdução à sociolinguística*: o tratamento da variação. São Paulo: Contexto, 2004.

MOORE, E. Interaction between social category and social practice: explaining was/were variation. *Language Variation and Change*, Cambrigde, v.22, p.347-371, 2010.

PODESVÁ, R. J. Phonation type as a stylistic variable: the use of falsetto in constructing a persona. *Journal of Sociolinguistic*, Hoboken, v.11, p.478-504, 2002.

PODESVÁ, R. J.; ROBERTS, S. J.; CAMPBELL-KIBLER, K. Sharing resources and indexing meanings in the production of gay styles. In: PODESVÁ, R. J. et al. (Ed.). *Language and sexuality*: contesting meaning in theory and practice. Stanford: CSLI Press, 2002. p.175-90.

SANKOFF, D.; TAGLIAMONTE, S.; SMITH, E. *Goldvarb X*: a variable rule application for macintosh and windows. Department of Linguistics of University of Toronto, Toronto, 2005.

SCHERRE, M. M. P. *Análise e mapeamento de três fenômenos variáveis no português brasileiro*. Projeto de pesquisa, Universidade Federal do Espírito Santo, Vitória, 2011. Disponível em: <<http://www.linguistica.ufes.br/sites/www.linguistica.ufes.br/files/Projeto%20de%20Pesquisa%20PPGEL%20-%20UFES%20-%20Marta%20Scherre.pdf>>. Acesso em: 22 mar. 2012.

SCHERRE, M. M. P.; RONCARATI, C. Programa de Estudos sobre o Uso da Língua (PEUL): origens e trajetórias. In: VOTRE, S.; RONCARATI, C. (Org.). *Anthony Julius Naro e a linguística no Brasil*: uma homenagem acadêmica. Rio de Janeiro: 7 Letras, 2008. p.37-49.

SEVERO, C. G. A comunidade de fala na sociolinguística laboviana: algumas reflexões. *Voz das Letras*, Concórdia, n.9, p.01-17, 2008. Disponível em: <<http://www.nead.uncnet.br/2009/revistas/letras/9/92.pdf>>. Acesso em: 22 mar. 2012.

TAGLIAMONTE, S. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press, 2006.

_____. Was/were variation across the generations: view from the city of York. *Language Variation and Change*, Cambrigde, v.10, p.153-191, 1998.

TAVARES, M. A. *Sequenciação de informações na fala de Natal (RN) e de Florianópolis (SC)*: um estudo sociofuncionalista comparativo. 2002. Projeto de Pesquisa, Universidade Federal do Rio Grande do Norte, Natal, 2002.

TAVARES, M. A.; MARTINS, M. A. *Banco de Dados FALA-Natal*: primeiras considerações. Manuscrito. 2012.

WENGER, E. *Communities of practice*: learning, meaning, and identity. Cambridge: Cambridge University Press, 1998.

ZHANG, Q. Rhotacization and the Beijing Smooth Operator: the social meaning of a linguistic variable. *Journal of Sociolinguistics*, Hoboken, v.12, p.201–222, 2008.

Recebido em abril de 2012

Aprovado em julho de 2012

VOGAIS NA AMAZÔNIA PARAENSE

Regina CRUZ*

- **RESUMO:** Este artigo trata essencialmente de uma apresentação das ações do projeto institucional Norte Vogais vinculado ao Diretório Nacional PROBRAVO. O artigo fornece uma ideia precisa de como a equipe da UFPA vinculada ao PROBRAVO está conduzindo suas investigações sobre a variedade do português falada no Pará. O projeto Norte Vogais conta com amostras de fala de trezentos e dezoito informantes nativos do Pará no seu banco de dados. As descrições sociolinguísticas empreendidas pela equipe da UFPA priorizaram a investigação de três aspectos fonéticos em particular: a) a variação das vogais médias pretônicas; b) a variação das vogais médias postônicas mediais e; c) a nasalidade alofônica. Os resultados obtidos reforçam a hipótese de Silva Neto (1957) de que o Pará compreenderia uma ilha dialetal na classificação de Antenor Nascente entre os dialetos do Norte do Brasil. Com o objetivo de refinar as descrições sociolinguísticas, duas novas ações se impuseram: a) o mapeamento da situação sociolinguística das áreas de contato interdialetal no Pará e; b) a análise acústica do sistema vocálico do português falado na Amazônia Paraense.
- **PALAVRAS-CHAVE:** Variação linguística. Vogais átonas. Português brasileiro. Região Norte.

Introdução

Em 2005, um grupo de investigadores decidiu criar um diretório de pesquisa nacional denominado *Descrição Sócio-Histórica das Vogais do Português (do Brasil)*¹ – PROBRAVO² – com o objetivo de realizar uma investigação multidisciplinar – sociohistórica e linguística – para descrever as realizações fonéticas das vogais nos dialetos do Sul ao Norte do Brasil. O PROBRAVO pretende com os seus estudos investigar: a) como são realizadas foneticamente as vogais no Português do Brasil (PB); b) a base da diversidade de realizações fonéticas das vogais átonas do PB e; c) como os falantes do PB se entendem apesar das diversidades da qualidade vocálica. O PROBRAVO pretende igualmente verificar se é possível explicar essa diversidade gramaticalmente.

* UFPA – Universidade Federal do Pará. Faculdade de Letras. Belém – Pará – Brasil. 66045-580 - regina@ufpa.br

¹ Disponível em: <<http://relin.letras.ufmg.br/probravo>>.

² O PROBRAVO é coordenado por Dr. Marco Antônio de Oliveira (PUCMG) e Dr. Seung-Hwa Lee (UFMG), fazem parte quinze outras instituições nacionais além da UFPA e da própria UFMG que sediam o projeto: PUC-MG, UFU, UNESP, UFPB, USP, UFMT, UFBA, UESB, UNICAMP, UFRJ, UFRGS, UCPEL, UFPEL, UNIR e PUC-RS.

Considerando a importância do tema – vocalismo átono -, são desenvolvidos projetos em todas as regiões do Brasil. No Pará, as investigações sobre o vocalismo átono do português brasileiro estão sendo conduzidas pela equipe da UFPA³ coordenada pela autora do presente artigo.

No caso específico das vogais médias pretônicas, a maior fonte de variação no português do Brasil em termos de vocalismo átono, ainda vigora a divisão dialetal de Antenor Nascente, na qual o Nordeste e Norte apresentariam determinadas variantes, diferentes daquelas do resto do Brasil. Mais especificamente, os dialetos do Norte e Nordeste apresentariam as variantes abertas das vogais médias pretônicas totalmente ausentes nos dialetos do sul. Entretanto, as investigações conduzidas no seio do PROBRAVO tem demonstrado que a questão não é tão simples como parece. De um lado, existem os aspectos estruturais correlacionados ao comportamento dessas vogais, de outro, estão as implicações sociais que se correlacionam não apenas às características dos falantes, mas também à atitude do ouvinte, que, muitas vezes, manifesta um comportamento estigmatizante frente ao falar diferente.

Por essa razão, o conhecimento da realidade de uso das vogais tem implicações positivas, tanto para a perspectiva do conhecimento de falares específicos como para a consolidação da teoria linguística em geral.

Na região Norte, os estudos sobre as vogais átonas foram impulsionados com a entrada da autora do presente artigo e de sua equipe de pesquisa, todos integrantes do projeto Norte Vogais, no Grupo PROBRAVO. Antes do trabalho da equipe do projeto Norte Vogais, tem-se o registro de raros estudos sobre o tema na variedade do português falado na Amazônia Paraense (NINA, 1991; FREITAS, 2001).

Para se ter uma ideia precisa de como a equipe da UFPA vinculada ao PROBRAVO está conduzindo suas investigações sobre a variedade do português falada no Pará, na seção seguinte, trazemos uma apresentação geral do Projeto Norte Vogais. Em seguida, detalhamos os procedimentos metodológicos adotados, enfatizando principalmente como se está procedendo na formação do banco de dados do projeto e no tratamento dos mesmos. Os fenômenos analisados, como variação das átonas, sejam pretônicas ou postônicas mediais, e nasalidade alofônica são abordados e discutidos com relação aos resultados obtidos a partir das análises sociolinguísticas da variedade do português regional paraense. Na quarta seção, apresentamos as tendências do português da Amazônia Paraense a partir das descrições sociolinguísticas empreendidas. A última seção descreve as duas novas frentes de atuação do Projeto Norte Vogais oriundas dos resultados sociolinguísticos obtidos. A conclusão retoma de forma sintética os temas aqui abordados.

³ Portaria ILC/UFPA N. 047/2009.

Projeto Norte Vogais

O grupo de pesquisadores da UFPA integrantes do projeto Norte Vogais também tem seus projetos em consonância com os objetivos do PROBRAVO. Até o presente momento sete municípios estão sendo investigados no estado do Pará: Aurora do Pará, Belém, Breves, Bragança, Cametá, Mocajuba e Breu Branco, tanto nas suas zonas rurais quanto urbanas.

De maneira geral, a equipe da UFPA, vinculada ao PROBRAVO, pretende ao mesmo tempo caracterizar o sistema vocálico átono e suas variantes, com base em amostra estratificada e em termos variacionistas, assim como analisar qualitativamente e explicar o processo de variação das vogais médias pretônicas e postônicas não finais no português falado no Norte do Brasil condicionado por fatores internos.

O projeto Norte Vogais, inicialmente, deu prioridade por investigar e formar *corpora* com amostras de fala das variedades linguísticas do português da Amazônia paraense situadas na zona do português regional paraense na classificação dialetal de Cassique (2006 apud CRUZ, 2012).

Cassique (2006), tomando como base as considerações de Silva Neto (1957), apresenta uma nova divisão dialetal do Pará (figura 1) que está sendo considerada pelos pesquisadores da UFPA ligados ao projeto PROBRAVO e, consequentemente, está na base da escolha das localidades-alvo do presente projeto.

A divisão dialetal do Pará estabelecida por Cassique (2006) considera as várias fases migratórias ocorridas no estado e que foram responsáveis pelas alterações da configuração original da situação sociolinguística da região. A primeira fase corresponde à chegada dos açorianos ao Pará ainda no século XVII, quando o português não conseguia sua implantação definitiva, uma vez que se atesta a situação de língua Franca com o domínio de uma variante do tupinambá, a Língua Geral Amazônica – LGA – (CALDAS et al., 2005).

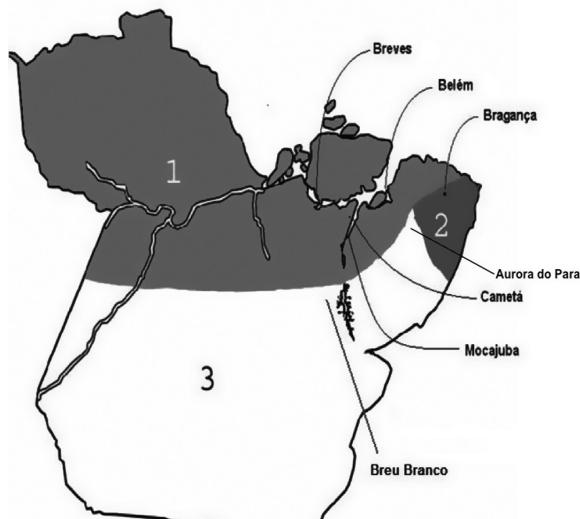
Segundo Rodrigues (1996), a primeira importante leva de falantes nativos de português chegou à Amazônia apenas no primeiro ciclo da Borracha no século XVIII, são os nordestinos que migraram para a Amazônia seduzidos pelo poder econômico da borracha. Até o final do século XIX, o português falado em todo o Pará foi o resultado do contato direto com a LGA, este português resiste ainda hoje na zona 1, da figura 1.

As cidades fundadas na época do Brasil Colonial, como Belém, Cametá, Óbidos e Soure a título de exemplo, mantêm uma variedade linguística com forte influência da LGA. Exceção feita à região Bragantina (zona 2 da figura 1) que, apesar de englobar a cidade de Bragança (também uma cidade do Pará com fundação histórica), possui uma variedade linguística que sofre forte influência

do falar nordestino por conta de sua proximidade geográfica com o Maranhão, e portanto registra um contato intenso com a variedade linguística do oeste maranhense.

Nos anos 1970, o português regional paraense sofreu mais uma alteração, quando governos militares incentivaram a migração do sul para a Amazônia, com a abertura da Transamazônica. A migração ocorrida se concentrou no sul e sudeste do Pará, alterando radicalmente o português falado nessas regiões e criando uma outra norma distinta do português regional paraense, cujo reduto é, de acordo com Cassique (2006 apud CRUZ, 2012, p.207), (a) o Médio Amazonas Paraense; (b) a região do Salgado; (c) a Ilha de Marajó; (d) o Nordeste Paraense e (e) a capital paraense – Belém. Portanto, na figura 1 abaixo, visualiza-se a região do português regional paraense (1), do dialeto bragantino (2) e de contato interdialetal (3).

Figura 1 - Mapa das localidades-alvo do Projeto Norte Vogais.



Fonte: Adaptado de Cruz (2012, p.207).

O português regional paraense é também aquele considerado por Silva Neto (1957) como sendo o de *canua cheia de cucus de pupa a prua*, por conta de sua principal marca dialetal, o alteamento das vogais posteriores em posição de sílaba tônica (RODRIGUES, 2005).

As localidades investigadas e selecionadas para constituirem a área geográfica de atuação do projeto Norte Vogais pertencem ao português regional paraense (Zona 1 da figura 1), Bragança é a única localidade pertencente ao dialeto bragantino (Zona 2 da figura 1), assim como Breu Branco e Aurora do Pará que

são as únicas localidades investigadas até o momento pertencentes à zona de contato interdialetal (Zona 3 da figura 1).

Desde 2007, quando passou a integrar o grupo PROBRAVO, o projeto Norte Vogais está estudando o processo de variação das vogais átonas de cinco localidades do estado do Pará, a saber: i) do português falado em Cametá (RODRIGUES; ARAÚJO, 2007; RODRIGUES; REIS, 2012; COSTA, 2010); ii) em Mocajuba (CAMPOS, 2008); iii) em Breves (CASSIQUE et al., 2009; DIAS; CASSIQUE; CRUZ, 2007); iv) em Belém (SOUSA, 2010; CRUZ et al., 2008) e v) em Breu Branco (MARQUES, 2008; COELHO, 2008; CAMPELO, 2008). Todas são descrições sociolinguísticas de cunho variacionista, que, portanto, apresentam um tratamento quantitativo dos dados. Essas análises seguiram orientações metodológicas comuns, desde a formação de seu *corpus* até o estabelecimento dos fatores a nelas serem considerados, com a finalidade de possibilitar uma comparação de seus resultados quanto ao fenômeno estudado, no caso, as vogais átonas. São justamente estes procedimentos que passaremos a detalhar na seção subsequente.

Procedimentos metodológicos adotados

As amostras de fala, que compõem o banco de dados do Projeto Norte Vogais, foram coletadas em trabalho de campo, com gravações em áudio. Para a coleta de dados, priorizaram-se as narrativas de experiência pessoal nos moldes da teoria da variação (TARALLO, 1988). Utilizou-se para cada variedade investigada uma amostra estratificada em sexo, faixa etária (15 a 25 anos; 26 a 45 anos e acima de 46 anos) e escolaridade (Analfabeto, Fundamental, Médio e Superior).

Uma vez as gravações concluídas, os dados obtidos foram transcritos grafematicamente observando os parâmetros da Análise da Conversação (CASTILHO, 2003).

Um arquivo contendo a triagem dos dados, tomando como unidade de análise o grupo de força como estabelecido por Câmara Jr. (1969), foi criado, por informante. Uma cópia do mesmo foi feita, para nela se proceder à transcrição fonética do vocabulário contendo o fenômeno estudado. Utilizou-se para a transcrição fonética o alfabeto SAMPA⁴.

Uma vez a transcrição fonética concluída, procedeu-se à codificação dos dados. Foram examinados os mesmos grupos de fatores para todos os estudos das variedades do estado do Pará do PROBRAVO⁵.

⁴ Disponível em: <<http://www.phon.ucl.ac.uk/home/sampa/index.html>>

⁵ Este arquivo de Especificação é de autoria de Orlando Cassique e Doriedson Rodrigues.

O arquivo de especificação utilizado contém grupos de fatores de diversas naturezas: a) fonéticos (natureza da vogal tônica, tipo de ataque da sílaba contendo a vogal-alvo, por exemplo); b) morfológicos (classe gramatical do vocábulo contendo a vogal-alvo); c) sintático (posição do vocábulo no grupo de força), entre outros, além dos fatores sociais também considerados na formação da amostra do *corpus*. Por último, realizou-se o tratamento estatístico dos dados pelo programa VARBRUL.

Caracterização dos *corpora* formados

Os *corpora* do projeto Norte Vogais possuem um número total de informantes variando de 24 (vinte e quatro) a 72 (setenta e dois), como podemos constatar no quadro 1. Ao todo são 318 (trezentos e dezito) informantes nativos da Amazônia Paraense, originários de cinco variedades locais: Belém, Cametá, Breves, Breu Branco e Mocajuba, em suas zonas rural e urbana. Como já descrito acima, os informantes que compõem o *corpus* foram estratificados socialmente em sexo, escolaridade, faixa etária e procedência e a situação de fala predominante é a de narrativas de experiência pessoal. Todo o *corpus* encontra-se transscrito grafematicamente e com os dados que atestam ocorrência do fenômeno-alvo – vogais médias pretônicas – transcritos foneticamente. Além das transcrições, o *corpus* contém o áudio das gravações realizadas em trabalho de campo.

Os *corpora* formados já subsidiaram a produção de 5 (cinco) Dissertações de Mestrado concluídas (CASSIQUE, 2002; RODRIGUES, 2005; CAMPOS, 2008; COSTA, 2010; SOUSA, 2010) e 6 (seis) planos de Iniciação Científica relatados (REIS, 2011; PAIVA, 2010; OLIVEIRA, 2008; DIAS, 2008; OLIVEIRA, 2007), além de 7 (sete) trabalhos publicados (DIAS; CASSIQUE; CRUZ, 2007; RODRIGUES; ARAÚJO, 2007; CRUZ et al., 2008; CASSIQUE et al., 2009; CRUZ, 2010; CRUZ, 2012; RODRIGUES; REIS, 2012).

Quadro 1 - Número total de informantes de cada amostra utilizada para a investigação do português falado nas localidades estudadas pelo Projeto Norte Vogais com a indicação da fonte de cada estudo realizado.

Localidade	Total de informantes	Fonte
Breves (urbano)	42 (quarenta e dois)	Oliveira (2007)
Breves (rural)	36 (trinta e seis)	Dias; Cassique; Cruz (2007)
Belém (urbano) ⁶	48 (setenta e dois)	Sousa (2010)
Belém (rural)	24 (vinte e quatro)	Cruz et al. (2008)
Cametá ⁷	48 (quarenta e oito)	Costa (2010)
Mocajuba ⁸	48 (quarenta e oito)	Campos (2008)
Breu Branco	24 (vinte e quatro)	Marques (2008), Campelo (2008) e Coelho (2008)

Fonte: Adaptado de Cruz (2010, p.250). ^{6 7 8}

O fato de o projeto Norte Vogais ter adotado procedimentos metodológicos comuns na formação de seus *corpora* e no tratamento dos dados possibilitou o avanço nas descrições sociolinguísticas das variedades investigadas do português falado na Amazônia Paraense, como veremos a seguir.

Fenômenos investigados

As descrições sociolinguísticas empreendidas pela equipe da UFPA priorizaram a investigação de três aspectos fonéticos em particular: a) a variação das vogais médias pretônicas; b) a variação das vogais médias postônicas mediais e c) a nasalidade alofônica, cujos detalhes serão fornecidos em seguida.

Vogais médias pretônicas

Nos estudos variacionistas do vocalismo átono do português brasileiro, encontram-se diversos trabalhos que investigam um dado dialeto, tendo como objeto de estudo o comportamento das vogais médias pretônicas. A maioria das pesquisas feitas a esse respeito segue orientações labovianas e utiliza dados quantitativos de um *corpus* representativo do dialeto escolhido, a título de ilustração, podemos citar Bisol (1981), Callou e Leite (1986), Nina (1991), Bortoni-Ricardo et al. (1991), Viegas (1995), Freitas (2001), Schwindt (2002) e Rodrigues (2005).

⁶ O *corpus* formado da variedade de Belém possui 15 horas, 30 minutos e 57 segundos de gravação (SOUZA, 2010).

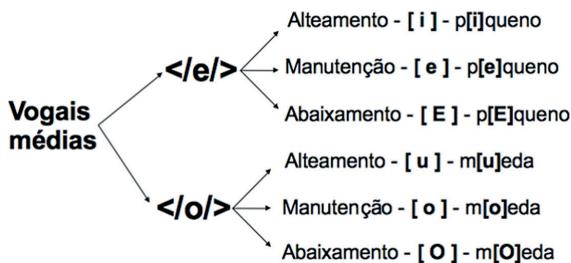
⁷ O *corpus* formado da variedade de Cametá possui 21 horas, 26 minutos e 6 segundos de gravação (COSTA, 2010).

⁸ O *corpus* formado da variedade de Mocajuba possui 25 horas, 49 minutos e 27 segundos de gravação (CAMPOS, 2008).

No presente estudo, buscamos sintetizar apenas os trabalhos referentes à variedade do Norte do Brasil, mais especificamente da Amazônia Paraense de modo que eles nos possam auxiliar em uma visão mais apurada de como esse fenômeno se manifesta na variedade em questão e, assim, verificar a contribuição do projeto Norte Vogais, em particular, para os estudos sociohistóricos das vogais do PB.

O projeto Norte Vogais toma como variável dependente, em suas investigações, as vogais médias pretônicas – </e/> e </o/> – e suas variantes – alteamento, manutenção e abaixamento –, como melhor ilustrado na figura 2 abaixo.

Figura 2 - Variável das vogais médias pretônicas e suas variantes, objeto de estudo do projeto Norte Vogais.



Fonte: Adaptado de Cruz et al. (2012).

O primeiro estudo sobre o assunto de caráter variacionista, envolvendo variedades faladas na Amazônia Paraense, é o de Nina (1991), o qual descreve e analisa o comportamento das vogais médias em posição pretônica na fala de moradores da Região Metropolitana de Belém (PA). Depois de Nina (1991), outros estudos de caráter variacionista se sucederam também, em particular destacaremos os resultados obtidos pela equipe do PROBRAVO na UFPA, a saber: Rodrigues e Araújo (2007), Dias, Cassique e Cruz, (2007), Oliveira (2007), Campos (2008), Marques (2008), Cruz et al. (2008), Cassique et al. (2009) e Sousa (2010).

Rodrigues e Araújo (2007), ao estudarem a variedade do português falada em Cametá (PA), constataram que a presença de vogais nasais propicia o alteamento, seguida da presença de pausa em contexto seguinte, de fricativas glotais, além da presença de vogal alta em posição contígua. Os informantes mais velhos e de menor escolaridade são os que mais realizam o alteamento na variedade observada.

Dias, Cassique e Cruz (2007), em estudo realizado na área rural do município de Breves (PA), tratam igualmente do alteamento das médias pretônicas. Os autores verificaram os condicionamentos dessa variação, muito mais na perspectiva de harmonização vocálica. Os resultados mostraram uma tendência ao não

alteamento (57%) em relação ao alteamento (43%) na variedade investigada. Foi constatado igualmente que a presença da vogal /i/ ou /u/ contígua à sílaba pretônica favorece a aplicação da regra de alteamento, assim como a distância, pois quanto menor a distância maior a possibilidade de ocorrência do fenômeno. No que diz respeito aos fatores sociais, a escolaridade também se mostrou favorecedora da aplicação da regra de alteamento, quanto maior o nível de escolaridade menos probabilidade de ocorrer o alteamento das médias pretônicas, por outro lado a fala dos informantes mais velhos registra uma tendência maior ao alteamento quando comparada a dos mais jovens. A pesquisa também mostrou a presença superior das vogais médias-altas em detrimento das médias-baixas.

Oliveira (2007) analisou dados provenientes de 42 informantes nascidos e residentes do município de Breves (PA). Os dados obtidos a respeito da variável dependente mostraram que a ocorrência de alteamento da vogal média pretônica na área urbana de Breves vem diminuindo consideravelmente. Isso se confirma pelo percentual de alteamento que é de apenas 19%, enquanto que o de não alteamento é de 81%, portanto em bem mais da metade dos dados, os informantes evitaram o uso de alteamento seja de /o/ para [u], seja de /e/ para [i]. Os resultados mostram, ao contrário do que previa Câmara Júnior (1969) no que diz respeito ao ambiente favorecedor do alteamento da vogal média pretônica, que são as vogais tônicas médias (fechadas e abertas) e não altas as que mais favorecem a elevação das vogais médias pretônicas, principalmente quando esta é posterior fechada [o]. Nenhum fator social foi selecionado pelo programa, uma vez que, de um modo geral, todos os falantes evitam o uso das variantes altas.

No estudo de Campos (2008), realizado a partir de dados coletados de 48 informantes no município de Mocajuba (PA), foi constatado que: a) a vogal alta /i/ na tônica como favorecedora do alteamento; b) vogais altas em posição contígua à sílaba tônica aumentam a possibilidade de ocorrência do alteamento; c) a presença de *onset* vazio favorece o alteamento, além de mostrar que o fenômeno ocorre em maior proporção na fala daqueles que possuem menor escolaridade. Campos (2008) constatou igualmente que, no caso das médias pretônicas da variedade de Mocajuba, atesta-se uma situação de variação estável.

Marques (2008) trata do alteamento das vogais médias em posição pretônica na variedade linguística do português falado no município de Breu Branco (PA). No total, identificaram-se 824 ocorrências do fenômeno estudado. Os resultados mostraram que, no caso das vogais médias anteriores, a presença de nasalidade, de vogal contígua alta, de *onset* vazio tanto na sílaba-alvo quanto na sílaba seguinte, assim como as sílabas leves favorecem o alteamento. As vogais médias pretônicas posteriores têm maior probabilidade de altear, quando: (a) há a presença da nasalidade, (b) a vogal contígua for alta e c) o *onset* for vazio. Dos grupos de fatores sociais, foi constatado que os mais velhos e menos escolarizados alteiam

mais em relação aos mais jovens e escolarizados. Marques (2008) mostrou que o alteamento das médias em posição pretônica é um fenômeno de pouca probabilidade e está em via de extinção no falar breuense por conta do intenso fluxo migratório na região sudeste do Pará.

Cruz et al. (2008) tratam da harmonização vocálica das vogais médias pretônicas no português falado nas ilhas de Belém (PA), com um *corpus* de 1.592 ocorrências do fenômeno estudado, coletado de 24 informantes estratificados socialmente. Os resultados obtidos mostraram que a elevação da vogal média pretônica é favorecida: a) por vogais altas na tônica, seja esta oral ou nasal; b) por vogais altas imediatas; c) pelas sílabas com *onset* vazio seja da sílaba contendo a vogal-objeto seja da sílaba seguinte a esta; d) pelo baixo grau de escolaridade e e) pela maior faixa etária. A análise dos dados mostrou que o alteamento das médias em posição pretônica é um fenômeno estável no falar analisado.

Cassique et al. (2009) tratam do alteamento das vogais médias em posição pretônica na variedade linguística do português falada no município de Breves (PA). O *corpus* utilizado conta com relatos de experiência de 78 informantes nascidos e residentes na cidade de Breves, dos quais 42 vivem na área urbana e 36 na área rural; esses informantes foram estratificados socialmente. Identificaram-se 7.320 dados do fenômeno estudado. Os resultados obtidos mostraram que a elevação da vogal média pretônica é favorecida: (i) pela proximidade da vogal pretônica em relação à sílaba tônica, (ii) pela ausência de sufixos, (iii) pelas sílabas pretônicas com *onset* vazio, (iv) pelas sílabas tônicas com *onset* vazio, (v) pelas sílabas pretônicas leves, (vi) pela proximidade da vogal pretônica em relação a uma vogal tônica alta e (vii) pelo baixo grau de escolaridade, pois os informantes que apresentaram pouca ou nenhuma escolaridade foram os que mais realizaram o alteamento.

Verificou-se também a presença superior de vogais médias fechadas em detrimento das médias abertas, assemelhando o dialeto aos falares do extremo Sul do Brasil. A análise dos dados mostrou que o alteamento das médias em posição pretônica é um fenômeno que está sendo extinto do falar rural e urbano de Breves, por se tratar de uma variável estigmatizada, repercutindo no apagamento e perda desta marca do dialeto local.

Sousa (2010) investigou o fenômeno do alteamento das vogais médias pretônicas /e/ e /o/ no português falado na área urbana da cidade de Belém (PA). Para a análise final, foram submetidos ao programa *Varbrul*, 1.434 dados: 776 das variantes de /e/ e 658 das variantes de /o/. Os resultados mostraram que, no dialeto em questão, predomina o não alteamento (64%) das vogais médias pretônicas em detrimento do alteamento (36%). A análise tomou como base 10 (dez) grupos de fatores linguísticos e 3 (três) grupos de fatores sociais que possivelmente pudessem explicar o fenômeno do alteamento. Destes, o programa só não selecionou 1 (um) grupo dos fatores sociais – sexo – como favorecedor do alteamento.

Vogais postônicas não finais

O único trabalho sobre postônicas mediais realizados no seio do PROBRAVO pela equipe da UFPA é o de Costa (2010). A autora verifica o comportamento das vogais médias /e/ e /o/ em posição postônica não final de itens lexicais na área urbana e rural do município de Cametá, Nordeste do estado do Pará, Brasil. O estudo desse fenômeno colaborará para a caracterização do dialeto do município em questão. Como reflexo da posição silábica aqui focalizada, os itens tratados neste trabalho foram somente os vocábulos proparoxítonos.

O *corpus* foi constituído com amostras de fala de 96 informantes estratificados em sexo, faixa etária, nível de escolaridade e procedência. A coleta dos dados foi realizada através de dois tipos de entrevista: a *livre* (48 informantes), em que os informantes discorreram livremente sobre suas experiências de vida; e o *teste* ou nomeação de figuras (48 informantes), que consistiu na identificação de desenhos pelos informantes (interpretados pelo falante como teste de memória). Ao todo foram analisados 2.177 dados no programa *Varbrul*.

Costa (2010) constatou que o fenômeno de alteamento, com um peso relativo de 0,46, apresenta probabilidade menor de ocorrência do que sua ausência com peso relativo de 0,54.

Examinou-se, também, que as vogais médias em posição tônica desempenham um papel bastante favorecedor do alteamento da média postônica /o/ em estudo. Dos fatores sociais estudados, a escolaridade e a procedência foram consideradas relevantes para explicar a regra variável de alteamento em Cametá.

Este trabalho apresenta igualmente uma análise qualitativa do comportamento das vogais médias – /e/ e /o/ – postônicas não finais, na variedade do português falada no município de Cametá (PA), as quais apresentam quatro variantes possíveis: manutenção [e]/[o], alteamento [i]/[u], apagamento [ø] e abaixamento [E]/[O].

Costa (2010) procede igualmente a uma descrição fonológica das vogais médias postônicas – /e/ e /o/ – não finais no português falado na cidade de Cametá (PA). Mais especificamente, analisa qualitativamente a variação das vogais médias postônicas mediais, de forma a verificar como o ambiente fonético é determinante no comportamento das quatro variantes identificadas, a saber: manutenção (abób[o]ra / velocíp[e]de), alteamento (abób[u]ra / velocíp[i]de), abaixamento (abób[O]ra / cér[E]bro) e apagamento (abób[ø]ra / velocíp[ø]i).

Os contextos fonéticos favorecedores de cada uma das variantes identificadas são explicados à luz da geometria dos traços (CLEMENTS, 2004). Aplica-se particularmente a proposta de Wetzels (1991) para explicar a neutralização ocorrida

com as átonas mediais no PB. De forma geral, os dados confirmam o quadro vocálico proposto por Câmara Jr. (1969) para as vogais átonas mediais do PB.

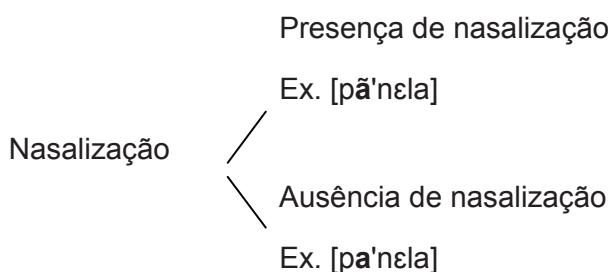
Nasalidade alofônica

Outro estudo sobre vogais átonas no escopo do projeto PROBRAVO foi o de Rodrigues e Reis (2012) sobre a nasalidade alofônica na variedade do português falada em Cametá (PA). Os autores compararam seus resultados com os de Cassique (2002), realizados sobre o mesmo fenômeno na variedade do português falada em Breves (PA) e com os de Abaurre & Pagotto (2002), que estudaram também a nasalidade alofônica em cinco variedades do PB (Recife, Salvador, São Paulo, Rio de Janeiro e Porto Alegre).

O *corpus* final compreendeu a análise de 2.575 ocorrências do fenômeno identificadas na fala de trinta e seis informantes, estratificados socialmente por sexo, faixa etária, escolaridade e procedência emprestados do *corpus* de Rodrigues (2005).

A variável dependente investigada compreendeu a ausência vs presença de nasalização na vogal átona seguida de consoante nasal na posição de *onset* da sílaba seguinte, como ilustrado na figura 3 abaixo.

Figura 3 - Variável dependente da nasalização alofônica e suas variantes.



Fonte: Adaptado de Cruz (2010, p.252).

Segundo Câmara Jr. (1969), existem dois tipos de nasalidade: a) uma de natureza fonológica, distintiva, ou seja, a vogal é nasal por união com um arquifonema nasal, são as chamadas vogais nasais verdadeiras (ex.: 'genro' [geNhu] em oposição a 'gero' [g ru]) e; b) outra de natureza fonética (vogais nasalizadas), isto é, a vogal vai ser nasal por efeito de uma assimilação do traço nasal da consoante da sílaba seguinte (ex.: *canela* > [kãñ la]) – como afirma Cagliari (1981), isso diz respeito ao fato de que, seguido de uma consoante nasal, qualquer segmento vocalico tende a ser nasalizado ou não, portanto a nasalidade neste caso é de caráter alofônico. É sobre este último tipo de nasalidade que trata o trabalho de Rodrigues e Reis (2012).

De acordo com os resultados obtidos em Cametá (PA), há maior probabilidade de ocorrer a variante Presença de nasalização vocálica pretônica, decorrente da assimilação do traço nasal da consoante da sílaba seguinte, em detrimento da variante Ausência de nasalização vocálica pretônica, relacionando o português falado nesse município, em termos de identidade linguística, ao falado em Recife, Salvador e Rio de Janeiro, onde também a presença de nasalização é alta.

O outro trabalho tomado como base de comparação para o fenômeno da nasalidade alofônica é o de Cassique (2002) que estudou o português falado na zona urbana de Breves, nordeste do estado do Pará. Esta pesquisa detectou 2.013 ocorrências de nasalidade alofônica na variedade do português falada em Breves, sendo 1.070 manifestações para a variante nasalizada, consubstanciando um percentual de 53%, e 943 dados atestando a variante não nasalizada, consubstanciando 47%. Segundo o pesquisador, os dados evidenciaram que, em Breves, a nasalização de vogais pretônicas antes de consoante nasal na sílaba seguinte era 6% maior que a não nasalização, o que equivalia, em termos de peso relativo, a 0,54 para a variante presença de nasalização pretônica.

No que diz respeito aos fenômenos investigados relativos ao vocalismo átono, dois achados nos são relevantes, um com relação às vogais pretônicas e outro com relação à nasalidade alofônica, como veremos na seção seguinte.

Tendências do português da Amazônia paraense

De forma geral, as descrições sociolinguísticas realizadas sobre o português falado na Amazônia Paraense tem demonstrado uma tendência à manutenção das vogais médias em posição pretônica, como expresso no quadro 2.

Quadro 2 - Percentual de alteamento nas variedades linguísticas investigadas pelo Projeto Norte Vogais.

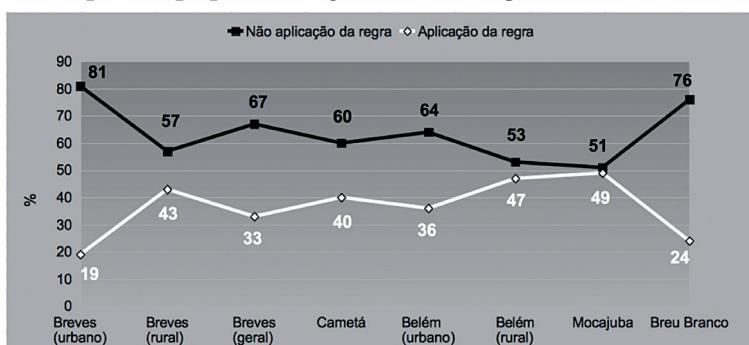
Dialeto	Não aplicação da regra	Aplicação da regra	Fonte
Breves (urbano)	81	19	Oliveira (2007)
Breves (rural)	57	43	Dias et al. (2007)
Breves (geral)	67	33	Cassique et al. (2009)
Cametá	60	40	Rodrigues; Araújo (2007)
Belém (urbano)	64	36	Sousa (2010)
Belém (rural)	53	47	Cruz et al. (2008)
Mocajuba	51	49	Campos (2008)
Breu Branco	76	24	Marques (2008)

Fonte: Adaptado de Cruz (2012, p.202).

Outro resultado relevante compreende a inexpressiva ocorrência de vogais médias baixas nas posições átonas. Tais resultados contrariam de um lado a divisão dialetal de Antenor Nascente que caracteriza os dialetos do Norte do Brasil como apresentando uma tendência à realização das vogais médias abertas nas posições átonas, em oposição aos dialetos do Sul do Brasil que prefeririam as vogais médias fechadas. Por outro lado, os resultados reforçam a hipótese de Silva Neto (1957) de que o Pará compreenderia uma ilha dialetal na classificação de Antenor Nascente entre os dialetos do Norte do Brasil (SILVA, 1989), que menciona nos seus resultados uma predominância das vogais baixas no seu *corpus* formado não somente com amostras de fala do dialeto-alvo – o de Salvador –, que também fôra confrontado com amostras de fala de 50 pontos do território baiano e de uma localidade do estado de Sergipe emprestadas, respectivamente, do Atlas Prévio dos Falares Baiano e de Mota (1979).

Os resultados dos estudos empreendidos pela equipe do Projeto Norte Vogais têm buscado prioritariamente caracterizar o português regional paraense. Nesse sentido, os resultados sobre as vogais médias pretônicas têm demonstrado uma tendência no uso de suas variantes – alteamento, manutenção e abaixamento – com probabilidade de maior ocorrência de manutenção das médias pretônicas em detrimento do alteamento das mesmas, como constatado no gráfico 1, inclusive com índices percentuais muito próximos de ocorrência da manutenção das médias pretônicas entre as variedades investigadas (Breves (rural), Belém, Cametá e Mocajuba). Duas das variedades investigadas (Breves (urbano) e Breu Branco) confirmam a tendência de manutenção, mas apresentam percentuais muito destoantes das quatro outras variedades comparadas.

Gráfico 1 - Tendência ao não alteamento das vogais médias pretônicas no Português da Amazônia Paraense, de acordo com os resultados dos trabalhos realizados pela Equipe do Projeto Norte Vogais da UFPA.



Fonte: Adaptado de Cruz (2012, p.203)

Os resultados do estudo da variação das médias pretônicas no português da Amazônia Paraense mostram que os percentuais de alteamento são muito baixos de modo geral nas zonas dialetais do Pará.

Os índices mais destoantes de Breves (19%) e de Breu Branco (24%), por indicarem a necessidade de uma investigação mais aprofundada sobre a situação sociolinguística destes dois municípios em particular, levaram a equipe da UFPA vinculada ao PROBRAVO a lançar uma nova edição de um outro projeto institucional, o Vozes da Amazônia⁹, destinado a investigar o português falado nas zonas de migração do Pará, como será descrito mais adiante. Breves e Breu Branco apresentam, em comum, o fato de terem sido justamente regiões que receberam um fluxo migratório considerável em decorrência de projetos econômicos da região.

O município de Breves, sozinho, apresenta um terço da população¹⁰ de todo o arquipélago marajoara. O inchaço populacional sofrido pelo município se deu no segundo ciclo da borracha, durante a Segunda Guerra Mundial, quando o governo apostando em um crescimento econômico oriundo da borracha, fez vir nordestinos para trabalharem na exploração da borracha na Amazônia, os ditos Soldados da Borracha. Uma vez terminada a guerra e o declínio do segundo ciclo da borracha, os imigrantes nordestinos não tiveram como voltar para a sua terra de origem e fixaram residência obrigatoriamente na Amazônia, uma boa parte deles ficou justamente na cidade de Breves.

Breu Branco é um dos municípios de criação recente no Pará, seus moradores, em sua maioria, são brasileiros originários de diferentes regiões do Brasil – mineiros, paulistas, gaúchos, paranaenses, maranhenses, cearenses, piauienses, tocantinenses – que migraram para o Pará para trabalhar na construção da hidrelétrica de Tucuruí na década de 1980. Com a conclusão da primeira etapa dos trabalhos de implantação da Hidrelétrica de Tucuruí, a maioria desses trabalhadores fixou residência nos municípios da região. Desta forma, a população atual de Breu Branco se assemelha à de Brasília (DF). Breu Branco, portanto, apresenta a mesma situação linguística atestada no DF e no sul do Pará, onde por questões econômicas – no caso, tal situação foi ocasionada pela construção da hidrelétrica – vários dialetos do PB convivem em uma mesma localidade, ocasionando de tal contato dialetal uma nova norma linguística.

Os resultados dos estudos sobre as vogais médias das variedades da Amazônia Paraense demonstraram que estas duas variedades investigadas fogem completamente a uma característica comum das variedades da Amazônia paraense que é a quase neutralização dos pesos relativos referentes à aplicação

⁹ Portaria N. 075/2009 ILC/UFPA.

¹⁰ A população total do município de Breves é de 80.158 habitantes, destes 40.074 habitantes vivem na área urbana, de acordo com o censo 2000 do IBGE.

e não aplicação da regra de alteamento das médias pretônicas. As variedades de Breu Branco (próximo a Tucuruí) e da zona urbana de Breves (no Marajó) têm como pontos em comum o fato de serem localidades que receberam uma forte migração de falantes do português de outras regiões do Brasil por conta de projetos econômicos. E, neste sentido, elas se assemelham muito às demais variedades da zona 3 da figura 1 de Cassique (2006 apud CRUZ, 2012), pois elas não possuem marcas de identidades (e aí em todos os sentidos) com a Amazônia paraense, e tudo indica inclusive na variedade linguística.

Nossa hipótese é a de que os fatores externos são relevantes no condicionamento da realização das variantes das médias pretônicas e fazem com que tais variedades sejam muito diferentes das demais da Amazônia Paraense. Para comprovar tal hipótese estamos procedendo a uma nova coleta de dados, controlando como principal fator a origem ou ascendência do falante, como fez Bortoni-Ricardo (1985). Acreditamos ser talvez o fator que esteja controlando a realização dessas variantes. Estamos verificando, também, além da variável Origem do Falante, o fator Faixa Etária, em especial, a fala dos mais jovens, a fim de se verificar se é uma variação estável ou mudança em progresso.

Como última hipótese, acreditamos que nas regiões em questão – Breu Branco e Breves – ainda não se cristalizou uma nova norma resultado do contato intervariedades, como ocorrido em Brasília, e o fato desta nova norma ainda não ter sido estabelecida resulta em contraste muito acentuado da realização das variantes atestadas.

Os resultados sobre a nasalidade vêm justamente fortalecer nossa hipótese de sustentação de uma investigação diferenciada para o português falado nas zonas de migração, uma vez que os dados de Breves (CASSIQUE, 2002) contrariam a tendência da nasalidade do português falado no Norte que seria de ocorrência de alto índice de nasalidade.

Comparando-se os resultados de Cametá e Breves com os das cinco capitais brasileiras, presentes em Abaurre & Pagotto (2002), obteve-se o seguinte quadro de tendência de nasalidade do português brasileiro, como visualizado no gráfico 2 abaixo.

Gráfico 2 - Tendência da nasalidade alofônica do Norte ao Sul do Brasil (ABAURRE; PAGOTTO, 2002), exceção da região Centro-Oeste, considerando os dados obtidos para Cametá (RODRIGUES; REIS, 2012) e Breves (CASSIQUE, 2002).



Fonte: Cruz (2010, p.253).

Constata-se, portanto, que há um declínio da nasalidade do Norte ao Sul do Brasil. O índice baixo da variedade de Breves parece não contrariar tal tendência, uma vez que Breves tem indícios de apresentar uma situação sociolinguística particular que levou à elaboração de um projeto de pesquisa específico para investigar o português falado nas regiões formadas em decorrência de forte fluxo migratório, o que será comentado na seção subsequente.

Desdobramentos do Projeto Norte Vogais

No sentido de precisar ainda mais os resultados alcançados, a equipe da UFPA ligada ao PROBRAVO definiu como ações futuras (CRUZ, 2012):

a) investigação mais aprofundada dos fatores externos como hipótese de serem eles os determinantes nos índices destoantes de alteamento *vs* manutenção das médias pretônicas no português falado nas localidades de forte fluxo migratório;

b) análise qualitativa e acústica dos *corpora* de Belém, Mocajuba e Cametá, de modo a precisar se o alteamento das médias pretônicas nessas variedades, quando se dá, é de fato motivado por harmonização vocalica.

São justamente essas ações que passaremos a descrever nesta seção.

Levantamento do português falado nas áreas de migração do Pará

O objetivo central deste projeto é o de mapear a situação sociolinguística diagnosticada por Cruz (2010) identificada na Amazônia paraense, onde se

atesta contato interdialetal decorrente de fluxo migratório intenso motivado por projetos econômicos na região Amazônica.

Até o presente momento três regiões foram selecionadas para uma nova fase de investigação do Vozes no estado do Pará: Marabá (MENDES, Projeto de Pesquisa em andamento), Aurora do Pará¹¹ (FERREIRA, Dissertação em andamento) e Breves (FAGUNDES, Dissertação em andamento); e duas outras localidades estão previstas: Breu Branco e Parauapebas. Mais precisamente, o Vozes busca identificar a influência de fatores extralingüísticos na configuração dos dialetos da Amazônia paraense, em localidades cujo fluxo migratório é considerável em decorrência de projetos econômicos desenvolvidos na região Amazônica. Na figura 1, estão indicadas as localidades investigadas pelo Vozes da Amazônia.

Para tal, toma como base para a formação do *corpus*, o conceito de rede social presente em Bortoni-Ricardo (1985), por ser o melhor instrumento para lidar simultaneamente com as diferenças individuais e com a identificação da variação sutil dos padrões sistemáticos e o da análise das redes sociais dos migrantes, já usado anteriormente em sociolinguística correlacional (LABOV, 1972; MILROY, 1980).

O projeto encontra-se vinculado a dois *campi* da UFPA – o de Belém e o de Marabá – e conta com a infraestrutura destes para a execução de suas atividades. A equipe atual do Vozes da Amazônia, responsável pela condução de suas investigações, é composta por 2 (dois) alunos de Mestrado, 2 (dois) bolsistas de Iniciação Científica e 3 (dois) pesquisadores titulados, todos com vínculo direto com a UFPA, além da coordenadora geral do projeto.

Como o objetivo do Vozes é compor um panorama histórico, antropológico e social do Pará, assim como identificar fatores sociais favorecedores da variação dialetal do português da Amazônia paraense falado nas regiões de forte migração interna, faz-se necessário relacionar aspectos de variação inter e intradialetal. Por essa razão, à medida que se caracteriza sociolinguisticamente o português falado em Marabá, Aurora do Pará e Breves, obtém-se o panorama geral das zonas de migração do Pará.

Assim como Bortoni-Ricardo (1985), utilizaremos o conceito de rede social para as investigações sociolinguísticas envolvendo comunidades linguísticas de migrantes, uma vez que o interesse da investigação não está nos atributos dos indivíduos, mas na caracterização das relações de um com outro, a qual

¹¹ Aurora do Pará fica a 200 km da capital paraense e localiza-se a nordeste do Pará. Atualmente, possui 26.427 habitantes (IBGE, 2010). Parte de seus habitantes é oriunda de outras cidades, inclusive cidades nordestinas, sobretudo do Ceará, os quais vieram para a região em busca de melhorias e fugindo da seca que assolava a região à época. Isso se deu em meados dos anos 1950-1960. A área total da sede é de 1.812 km², localizada no Bioma Amazônia. No que se refere ao aspecto histórico, o município recebeu bastantes migrantes nordestinos, sobretudo, do Ceará, e por isso possui características peculiares em sua cultura e língua que diferem em alguns aspectos daquela conhecida por paraense. Aurora do Pará tem sua criação por meio da Lei Estadual n. 5698, de 13 de dezembro de 1991.

pode predizer e explicar o comportamento destes indivíduos, inclusive o comportamento linguístico.

Por esta razão, para a formação dos *corpora* consideraremos a noção de grupo de referência dos falantes, como fez Bortoni-Ricardo (1985). Segundo esta autora, o grupo de referência é o grupo que serve de alavanca à construção da identidade do indivíduo, ou seja, o falante modela seu discurso de acordo com o grupo com o qual ele busca identificar-se, com o grupo que atende as suas expectativas psicossociais.

Portanto, para a constituição do *corpus*, dois grupos de informantes são formados. Um grupo de ancoragem com 24 informantes (12 de cada sexo), distribuídos em duas faixas etárias de 26 a 46 anos e acima de 50 anos. E um grupo de controle de 12 informantes (6 de cada sexo), no qual todos devem ter algum grau de parentesco com os membros do grupo de ancoragem, ou seja, devem ser filhos, netos ou sobrinhos dos informantes do grupo de ancoragem.

Uma vez o trabalho de campo concluído, o tratamento dos dados seguirá todas as etapas previstas em um estudo sociolinguístico, a saber: (i) transcrição dos dados nos moldes da análise da conversação (CASTILHO, 2003); (ii) triagem dos grupos de força (CÂMARA JR., 1969); (iii) transcrição fonética dos vocábulos contendo marcas dialetais alvo, utilizando-se o alfabeto SAMPA; (iv) codificação dos dados e (v) tratamento quantitativo VARBRUL.

Caracterização acústica do sistema vocálico da Amazônia paraense

Como já mencionado anteriormente, no caso das vogais médias pretônicas, todos os resultados sobre as variedades do português da Amazônia paraense apontam para uma tendência dos dialetos paraenses de preferência pela não aplicação da regra de alteamento das médias pretônicas em detrimento do alteamento, como se pode verificar no gráfico 4.

Os estudos também confirmam que quando ocorre o alteamento das vogais médias pretônicas nos dialetos paraenses, o mesmo se dá motivado por uma clara harmonização vocálica, principalmente condicionada pela vogal da sílaba tônica e pela sílaba imediata. Este processo de harmonia vocálica atestado nos dialetos paraenses contempla as hipóteses de Câmara Jr. (1969) e de Silva Neto (1957) ao mesmo tempo. De acordo com Câmara Jr. (1969), o alcantamento da pretônica é determinado pela altura da vogal da tônica, um dos fatores mais favorecedores do alteamento das vogais médias pretônicas nos dialetos paraenses. Silva Neto (1957), por sua vez, afirmava ser o alcantamento ainda mais favorecido por sílaba com vogal alta contígua e imediata à sílaba da vogal pretônica, o outro fator altamente favorecedor do alteamento das médias pretônicas nos dialetos paraenses.

Logo, os fatores internos são relevantes para explicar o comportamento variacional das vogais médias nos dialetos da Amazônia Paraense e, em particular, os fonéticos como têm demonstrado os estudos realizados.

Outro resultado relevante compreende a inexpressiva ocorrência de vogais médias baixas nas posições átonas, o que contraria, como já mencionado anteriormente, a divisão dialetal de Antenor Nascente de que os dialetos do Norte do Brasil se caracterizariam por apresentarem uma tendência à realização das vogais médias abertas nas posições átonas.

Diante da particularidade das vogais da Amazônia paraense, neste panorama, optou-se por proceder a um refinamento nas análises empreendidas e, desta vez, por observar mais de perto seus aspectos acústicos.

Por essa razão, atualmente o projeto Norte Vogais está investigando as características acústicas das vogais orais átonas das variedades estudadas da Amazônia Paraense pela equipe da UFPA vinculada ao PROBRAVO¹².

Inicialmente quatro variedades da Amazônia Paraense – Belém, Cametá, Mocajuba¹³ e Bragança¹⁴ – estão sendo investigadas com o objetivo principal de analisar acusticamente o seu sistema vocálico tônico e átono. Para tal, formou-se um *corpus* com a mesma estratificação social adotada para as descrições sociolinguísticas: idade, nível de escolaridade¹⁵, sexo e dialeto. Até o presente momento, o Projeto Norte de Vogais dispõe de um material para análise acústica gravado com 108 informantes¹⁶.

Cada informante participou de dois protocolos diferentes de coleta de dados: a) um teste de projeção de imagens¹⁷ e b) a leitura em voz alta de um texto sobre futebol.

Para a formação do *corpus* final, foram utilizados na coleta de dados 74 vocábulos selecionados contendo os mesmos contextos controlados pelos estudos variacionistas, como ‘bebidas’, ‘cerveja’, ‘futebol’, ‘costurar’, ‘namorados’, ‘toalha’, a título de exemplo. Utilizamos prioritariamente os vocábulos indicados por

¹² A primeira iniciativa de análise acústica para dar conta das necessidades do grupo PROBRAVO foi empreendida por Cruz (2011), em seu estágio pós-doutoral no Departamento de Linguística, da *New York University*, no período de setembro de 2010 a março de 2011, quando a coordenadora do presente projeto obteve uma bolsa Fulbright/Capes pelo Edital DRI/CGCI N. 027/2009 (processo BEX 1754/10-6) e lá esteve na condição de pesquisadora visitante.

¹³ Sob responsabilidade de Socorro Campos (UFPA/UFC).

¹⁴ Sob responsabilidade de Carlos Nedson (UFPA).

¹⁵ Como o protocolo experimental prevê etapa de leitura de texto, os sujeitos analfabetos foram excluídos naturalmente.

¹⁶ Sendo 18 nativos da capital paraense, 36 da cidade de Cametá (18 da zona urbana e 18 da zona rural), 36 da cidade de Mocajuba (18 da zona urbana e 18 da zona rural) e 18 da cidade de Bragança.

¹⁷ Inspirou-se para a montagem desse teste em Ribeiro (2007) e Costa (2010), que empregaram este mesmo tipo de instrumento de coleta de dados ao estudar as médias postônicas não finais.

Bisol (1981) contendo vogais pretônicas em contexto de alta variabilidade como ‘formiga’, ‘fogueira’, ‘bonita’, ‘menino’.

Aos 108 (cento e oito) sujeitos foi solicitado primeiro que dissessem o que viam logo depois que cada imagem era projetada. A sequência das 70 imagens foi projetada duas vezes a cada participante da pesquisa. A nenhum deles foi revelado, de imediato, os verdadeiros objetivos da pesquisa. Eles tomaram conhecimento dos reais objetivos da pesquisa, apenas quando lhes foi solicitada a assinatura no Termo de Consentimento esclarecido do projeto. Uma vez aplicado o teste de imagens, foi solicitada a leitura de um texto. Foi dado a cada informante um tempo de 10 minutos para familiarização com o tema do texto – futebol – de modo que a leitura fosse a mais natural possível.

Portanto, de cada informante foram obtidos dois tipos de dados, aqueles produzidos por indução com a projeção das imagens e os dados de fala lida. Apenas os dados da variedade de Belém (PA) estão com suas análises mais avançadas sejam os dados obtidos com o teste de imagens (CRUZ, 2011) sejam os dados de fala lida (REIS, 2011; COSTA; SILVA; CRUZ, 2012; CRUZ et al., 2012).

Concluídas todas as etapas anteriores de coleta de dados, procede-se ao tratamento dos dados, que compreende: a) segmentação do sinal de áudio de cada informante no programa PRAAT em cinco níveis; b) extração dos vocábulos-alvo das gravações originais com o programa PRAAT, seguido de codificação deles¹⁸; c) levantamento das ocorrências das palavras-alvo por informante e por tipo de coleta de dados; d) identificação das variantes ocorridas por vocábulo, considerando os dados de cada informante; e) organização dos dados obtidos em uma planilha Excel para o registro das medidas acústicas tomadas de cada vogal-alvo.

Estão sendo tomadas medidas de F1, F2, F0 e da duração das vogais médias alvo na posição pretônica. As medidas acústicas estão sendo feitas com o auxílio do programa PRAAT. Antes de realizarmos um tratamento estatístico mais robusto, realizar-se-ão tomadas de médias, desvio padrão e teste T com os valores obtidos para verificação de sua relevância.

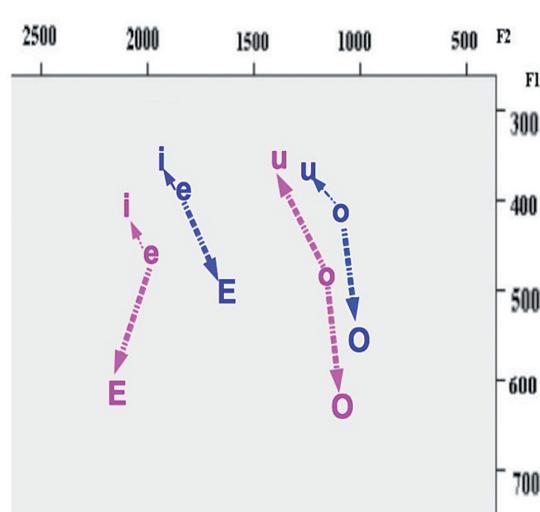
As análises dos dados da variedade de Belém (PA) de caráter ainda preliminar têm demonstrado que a variedade do português falada na Amazônia Paraense possui no nível acústico 4 (quatro) variantes para as vogais médias pretônicas: a) alteamento como em *fut[i]bol* e *c[u]madre*; b) manutenção como em *fut[e]bol* e *c[o]madre*; c) abaixamento como em *fut[E]bol* e *c[O]madre*; d) mas há também casos de enfraquecimento das vogais pretônicas ou mesmo

¹⁸ Os informantes recebem um código que identifica dialeto, sexo, escolaridade e faixa etária de cada um. Adapta-se para o presente projeto um código muito semelhante ao código fornecido pelo projeto AMPER-POR: <<http://pfonetica.web.ua.pt/AMPER-POR.htm>>. Uma vez esse código montado, escreve-se a palavra-alvo seguida de seu número de ocorrência no *corpus* gravado.

de total desvozeamento destas como em *fut[Ø]bol* e *c[Ø]madre*. As análises acústicas têm confirmado os resultados das descrições sociolinguísticas de que a manutenção das médias é a variante de maior ocorrência no português falado na Amazônia Paraense, seguida do alteamento das médias pretônicas, e a variante de abaixamento é a de menor ocorrência nos *corpora*. Com relação ao nível de escolaridade, os falantes de baixa escolaridade possuem comportamento linguístico no que diz respeito à realização das vogais médias pretônicas distinto dos falantes de alta escolaridade (níveis médio e superior) independentemente do sexo.

Uma análise acústica preliminar já foi realizada com os dados de fala lida, apenas uma tomada de médias dos valores em Hz de F1 e F2 fora realizada e apesar de ainda estarmos procedendo a um tratamento estatístico descritivo e nenhum tipo de normalização ter sido realizado com os dados, os resultados são relevantes, pois mostram que, no caso das vogais anteriores, tanto na fala masculina quanto na fala feminina a variante alta ocupa quase o mesmo espaço acústico que a variante média fechada e ambas apresentam-se bem distantes da variante baixa. Por outro lado, as vogais posteriores apresentam-se bem discriminadas acusticamente tanto na fala masculina quanto feminina, como podemos constatar na figura 3 abaixo.

Figura 3 - Médias de valores em Hz de F1 e F2 das três variantes das vogais médias pretônicas dos dados de fala feminina (em rosa) e fala masculina (em azul) do corpus de fala lida da variedade de Belém (BE0).



Fonte: Adaptado de Cruz et al. (2012).

Com o objetivo de tornar mais robustas as análises, futuramente, utilizar-se-á o *software* R para o tratamento estatístico dos dados.

Com relação ao PROBRAVO em si, a presente proposta pretende contribuir diretamente na implementação de uma metodologia adequada para a análise acústica das vogais médias pretônicas de todas as quinze variedades estudadas pela equipe nacional. Acredita-se, portanto, que as ações aqui propostas poderão constituir modelo para ações nacionais mais contundentes e amplas, no que diz respeito ao tema em questão.

A principal meta é a implementação de uma metodologia adequada para a análise acústica das vogais médias pretônicas de todas as quinze variedades estudadas pela equipe nacional do PROBRAVO.

O outro trabalho de que se tem conhecimento sobre o assunto, apesar de mais voltado para as características acústicas da vogais tônicas, utilizou dados artificiais e não existentes no sistema português como um todo (ESCUDERO et al., 2009). Portanto, a presente proposta prioriza uma análise acústica de dados naturais do português.

Conclusão

Apresentamos neste artigo as ações do Projeto Norte Vogais, sediado na UFPA e vinculado ao Diretório Nacional PROBRAVO.

A equipe responsável pelo Projeto Norte Vogais atua desde 2007 e já realizou as descrições sociolinguísticas de cinco variedades do português falado na Amazônia Paraense (Cametá, Mocajuba, Breves, Belém e Breu Branco). O emprego de uma metodologia comum permitiu uma comparação dos resultados obtidos que apontou uma tendência do português falado no Pará à manutenção das vogais médias pretônicas. A comparação apontou também para uma possível dominância de harmonia vocalica quando ocorre o alteamento das médias pretônicas determinada por fatores internos. Por último, os resultados demonstraram a necessidade da aplicação de uma metodologia diferenciada na investigação do português falado nas zonas de contato interdialetal no Pará.

Por esta razão, duas novas ações do Projeto Norte Vogais se instalaram: uma responsável por identificar a influência de fatores extralingüísticos na configuração dos dialetos da Amazônia paraense, em localidades cujo fluxo migratório é considerável em decorrência de projetos econômicos desenvolvidos na região amazônica (Projeto Vozes da Amazônia); uma outra ação responsável por proceder a um refinamento nas análises empreendidas e observar mais de perto aspectos acústicos do sistema vocalico do português falado no Pará. Esta última, mais

avançada, tem confirmado os resultados das descrições sociolinguísticas de que a manutenção das médias é a variante de maior ocorrência no português falado na Amazônia Paraense, seguida do alteamento das médias pretônicas, e a variante de abaixamento é a de menor ocorrência nos *corpora*.

O artigo também tratou de outros aspectos ligados ao vocalismo átono do PB: a nasalidade alofônica e as vogais médias postônicas não finais.

CRUZ, R. Vowels in the north of Brazil. *Alfa*, São Paulo, v.56, n.3, p.935-962, 2012.

■ *ABSTRACT:* This work deals with a presentation of Norte Vogais project's actions linked to the PROBRAVO team. This paper gives a clear idea of how the UFPA team has been conducting its research about Brazilian Portuguese spoken in Amazon. The Norte Vogais project database has samples of 318 native speakers from Pará. The sociolinguistic descriptions prioritized the research of three phonetic aspects: a) unstressed mid vowels variation in pretonic position; b) unstressed mid vowels variation in post-tonic position and; c) allophonic nasalization. The results reinforce Silva Neto's hypothesis (1957) that Pará is a dialectal island in Antenor Nascente's classification in relation to North dialects. In order to precise the sociolinguistic descriptions, two new actions were necessary: a) sociolinguistic situation map of inter dialectal contact areas in Pará and; b) acoustic analysis of the vowel system of Brazilian Portuguese spoken in the Amazon region of Pará.

■ *KEYWORDS:* Linguistic variation. Unstressed vowels. Brazilian Portuguese. North of Brazil.

REFERÊNCIAS

- ABAURRE, B.; PAGOTTO, E. Nasalização no português falado no Brasil. In: KOCH, I. (Org.). *Gramática do português falado*. 2.ed. São Paulo: Ed. da Unicamp, 2002. p.491-515.
- BISOL, L. *Harmonia vocálica: uma regra variável*. 1981. 333f. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- BORTONI-RICARDO, S. M. *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil*. Cambridge: Cambridge University Press, 1985.
- BORTONI-RICARDO, S. M.; GOMES, C. A.; MALVAR, E. S. Um estudo preliminar do /e/ pretônico. *Cadernos de Estudos Linguísticos*, Campinas, v.20, p.75-90, 1991.
- CAGLIARI, L. C. *Elementos de fonética do português brasileiro*. 1981. 192f. Tese (Livre de Docência em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1981.
- CALDAS, R. et al. *Possibilidade de interferência da língua geral amazônica na combinação de orações em Urubú-Ka'apór*. 2005. Trabalho oral apresentado ao Workshop sobre Linguística Histórica e Línguas em Contato: Línguas Indígenas Brasileiras e de Áreas Adjacentes, Brasília, 2005.

CALLOU, D.; LEITE, Y. As vogais pretônicas no falar carioca. *Estudos UFBA*, Salvador, n.5, p.1-253, dez. 1986.

CÂMARA JR, J. M. *Estrutura da língua portuguesa*. Petrópolis: Vozes, 1969.

CAMPELO, M. A. *Variação das vogais médias anteriores pretônicas no português falado no município de Breu Branco (PA)*: uma abordagem variacionista. 2008. 50f. Trabalho de Conclusão de Curso (Licenciatura em Letras) – Faculdade de Letras, Universidade Federal do Pará, Belém, 2008.

CAMPOS, S. *Alteamento vocálico em posição pretônica no português falado no município de Mocajuba-Pará*. 2008. 202f. Dissertação (Mestrado em Letras) –Universidade Federal do Pará, Belém, 2008.

CASSIQUE, O. *Minina bunita... olhos esverdeados (um estudo variacionista da nasalização vocálica pretônica no português falado na cidade de Breves/PA)*. 2002. 90f. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém, 2002.

CASSIQUE, O. et al. Variação das vogais médias pré-tônicas no português falado em Breves (PA). In: HORA, D. (Org.). *Vogais no ponto mais oriental das Américas*. João Pessoa: Ideia, 2009. p.163-184.

CASTILHO, A. *A língua falada no ensino do português*. 5.ed. São Paulo: Contexto, 2003.

CLEMENTS, G. N. Feature organization. In: BROWN, K. (Ed.). *The encyclopedia of language and linguistics*. 2.ed. Oxford: Elsevier, 2004. v.4. p.433-441.

COELHO, M. L. *A variação das vogais médias posteriores pretônicas no português falado no município de Breu Branco (PA)*: uma abordagem variacionista. 2008. 47f. Trabalho de Conclusão de Curso (Licenciatura em Letras) – Faculdade de Letras, Universidade Federal do Pará, Belém, 2008.

COSTA, M.; SILVA, A. C.; CRUZ, R. Análise qualitativa e acústica das vogais médias pretônicas no português falado na Amazônia paraense. In: CONGRESSO INTERNACIONAL DA ASSOCIAÇÃO BRASILEIRA DE LINGÜÍSTICA, 2012, Natal. *Resumos...* Natal: ABRALIN, 2012. p.59-60.

COSTA, R. *Descrição sociolinguística das vogais médias postônicas não-finais /o/ e /e/ no português falado no município de Cametá-PA*. 2010. 100f. Dissertação (Mestrado em Letras) – Curso de Mestrado em Letras, Universidade Federal do Pará, Belém, 2010.

CRUZ, R. Alteamento vocálico das médias pretônicas no português falado na Amazônia paraense. In: LEE, S. H. (Org.). *Vogais além de Belo Horizonte*. Belo Horizonte: Ed. da UFMG, 2012. p.194-220.

_____. Apagamento de vogais médias pretônicas no português da Amazônia paraense: a quarta variante. In: SIMPÓSIO SOBRE VOGAIS: SIS-VOGAIS, 3., 2011, Porto Alegre. *Resumos...* Porto Alegre: UFRGS, 2011. p.7-11.

_____. Panorama sociolinguístico do português falado na Amazônia paraense. In: SIMÕES, S. (Org.). *Navegando entre o rio e a floresta por vias do Marajó: com vista a ensino, pesquisa e extensão*. Belém: EDUFPA, 2010. p.243-261.

CRUZ, R. et al. Vogais médias pretônicas no português falado em Belém (PA): análise qualitativa e acústica. In: SEMINÁRIO INTERNACIONAL DE FONOLOGIA. 4., Porto Alegre. *Anais...* Porto Alegre: Ed. da UFRGS, 2012.

CRUZ, R. et al. As vogais médias pretônicas no português falado nas Ilhas de Belém (PA). In: ARAGÃO, S. (Org.). *Estudos em fonética e fonologia no Brasil*. João Pessoa: GT-Fonética e Fonologia/ANPOLL, 2008. 1 CD-ROM.

DIAS, M. *O alteamento das vogais pré-tônicas /o/ e /e/ no português falado na área rural do município de Breves (PA)*: uma abordagem variação. 2008. Relatório Técnico-Científico de Iniciação Científica, Universidade Federal do Pará, Belém, 2008.

DIAS, M.; CASSIQUE, O.; CRUZ, R. O alteamento das vogais pré-tônicas no português falado na área rural do município de Breves (PA): uma abordagem variação. *Revista Virtual de Estudos da Linguagem: ReVEL*, Porto Alegre, n.9,v.5,p.01-18, jul. 2007. Disponível em: <http://www.revel.inf.br/site2007/_pdf/9/artigos/>. Acesso em: 27 set. 2012.

ESCUDERO, P. et al. A cross-dialect acoustic description of vowels: brazilian and european portuguese. *Journal of the Acoustical Society of America*, New York, v.126, n.3, p.1378-1393, set. 2009.

FAGUNDES, G. *Alteamento das vogais médias pretônicas no português da Amazônia paraense*: a influência do dialeto dos migrantes no português falado em Breves (PA). Belém: Universidade Federal do Pará. (Dissertação em andamento).

FERREIRA, J. E. *Mininu, meninu, mEninu*: variação das médias pretônicas <e> e <o> no falar de Aurora do Pará – PA: um estudo variação. Belém: Universidade Federal do Pará. (Dissertação em andamento).

FREITAS, S. *As vogais médias pretônicas no falar da cidade de Bragança*. 2001. 96f. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém, 2001.

LABOV, W. *Language in the inner city*. Philadelphia: University Press, 1972.

MARQUES, L. C. *Alteamento das vogais médias pré-tônicas no português falado no município de Breu Branco (PA)*: uma abordagem variação. 2008. 52f. Trabalho de Conclusão de Curso (Licenciatura em Letras) – Faculdade de Letras, Universidade Federal do Pará, Belém, 2008.

MENDES, S. *Vozes da Amazônia: a realização das vogais médias pretônicas na comunidade linguística de Marabá*. Marabá: Universidade Federal do Pará. (Projeto de Pesquisa em andamento).

MILROY, L. *Language and social networks*. Oxford: Basil Blackwell, 1980.

MOTA, J. *Vogais antes do acento em Ribeirópolis (SE)*. 1979. 287f. Dissertação (Mestrado em Língua Portuguesa) - Universidade Federal da Bahia, Salvador, 1979.

NINA, T. *Aspectos da variação fonético-fonológica na fala de Belém*. 1991. 216f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.

OLIVEIRA, D. *A neutralização das vogais médias pós-tônicas não finais no português falado na área urbana do município de Cametá (PA): uma abordagem variacionista*. 2008. Relatório Técnico-Científico de Iniciação Científica, Universidade Federal do Pará, Belém, 2008.

_____. *Harmonização vocálica no português falado na área urbana do município de Breves/PA: uma abordagem variacionista*. Relatório Técnico-Científico de Iniciação Científica, Universidade Federal do Pará, Belém, 2007.

PAIVA, E. *Variação das vogais médias pretônicas no português falado no Norte do Brasil: uma abordagem formal*. 2010. Relatório Técnico-Científico de Iniciação Científica, Universidade Federal do Pará, Belém, 2010.

REIS, E. *O Sistema vocálico do português falado em Belém (PA): análise experimental*. 2011. Relatório Técnico-Científico de Iniciação Científica, Universidade Federal do Pará, Belém, 2011.

RIBEIRO, D. *Alçamento de vogais postônicas não finais no português de Belo Horizonte – Minas Gerais: uma abordagem diffusionista*. 2007. 274f. Dissertação (Mestrado em Língua Portuguesa e Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

RODRIGUES, A. As línguas gerais sulamericanas. *Papia*, Brasília, n.4, v.2, p.6-18, 1996.

RODRIGUES, D. *Da zona urbana à rural/entre a tônica e a pretônica: alteamento /o/ > [u] no português falado no município de Cametá/Ne paraense: uma abordagem variacionista*. 2005. 387f. Dissertação (Mestrado em Letras) – Universidade Federal do Pará, Belém, 2005.

RODRIGUES, D.; ARAÚJO, M. As vogais médias pretônicas / e / e / o / no português falado no município de Cametá/PA – a harmonização vocálica numa abordagem variacionista. *Cadernos de Pesquisa em Linguística, Variação no Português Brasileiro*, Porto Alegre, v.3, p.104-126, nov. 2007.

RODRIGUES, D.; REIS, G. Variação da nasalização vocálica pretônica seguida de consoante nasal na sílaba seguinte no português falado no município de Cametá – Pará. In: LEE, S. H (Org.). *Vogais além de Belo Horizonte*. Belo Horizonte: Ed. da UFMG, 2012. p.322-348.

SCHWINDT, L. C. A regra variável de harmonização vocálica no RS. In: BISOL, L.; BRESCANCINI, C. (Org.). *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, 2002. p.161-182.

SILVA, M. B. *As pretônicas no falar baiano*: a variedade culta de Salvador. 1989. 377f Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1989.

SILVA NETO, S. *Introdução ao estudo da língua portuguesa no Brasil*. 4.ed. Rio de Janeiro: Presença, 1957.

SOUZA, J. *A variação das vogais médias pretônicas no português falado na área urbana do município de Belém/PA*. 2010. 371f. Dissertação (Mestrado em Letras) – Curso de Mestrado em Letras, Universidade Federal do Pará, Belém, 2010.

TARALLO, F. *A pesquisa sociolinguística*. São Paulo: Ática, 1988. (Série Princípios).

VIEGAS, M. C. Alçamento de vogais pretônicas e os itens lexicais. *Revista de Estudos da Linguagem*, Belo Horizonte, v.2, n.3, p.101-123, 1995.

WETZELS, W. L. Harmonização vocálica, truncamento, abaixamento e neutralização no sistema verbal do português: uma análise autossegmental. *Caderno de Estudos Linguísticos*, Campinas, n.21, p.25-58, 1991.

Recebido em abril de 2012

Aprovado em julho de 2012

O PAULISTANO NO MAPA SOCIOLINGUÍSTICO BRASILEIRO

Ronald Beline MENDES*
Livia OUSHIRO**

- RESUMO: Este artigo traz notícia acerca dos trabalhos que vêm sendo desenvolvidos pelo Grupo de Estudos e Pesquisa em Sociolinguística da USP (GESOL-USP). Com o intuito de tomar a cidade de São Paulo e o paulistano mais presentes no mapa da sociolinguística brasileira, esses trabalhos organizam-se, sobretudo, em torno da construção de uma nova amostra da fala paulistana, que permita responder às seguintes perguntas centrais: “o que significa falar como paulistano?” e “quais são as variáveis e variantes linguísticas que o identificam?”. Essa nova amostra (SP2010) será disponibilizada (áudio e transcrições) para acesso público e para fins de pesquisa em página própria na Internet. Além de discutir os critérios que definem a construção de tal amostra, bem como os desafios nela envolvidos, esse artigo também faz um panorama dos trabalhos que se têm desenvolvido a partir dos dados que foram coletados durante a fase de preparação e de treinamento de jovens pesquisadores que nela trabalhariam (Amostra SP-Piloto). Finalmente, também delineia caminhos para trabalhos futuros dentro da agenda de pesquisa do grupo.
- PALAVRAS-CHAVE: Variação linguística. Paulistanos. Identidade. *Corpus*.

Introdução

De uma perspectiva internacional, pode-se dizer que a Sociolinguística é um campo de estudos bastante desenvolvido no Brasil. Na reedição de um de seus trabalhos seminais e mais conhecidos, Labov (2006) coloca ao lado de Nova Iorque algumas cidades ao redor do mundo nas quais foram desenvolvidos projetos de pesquisa que considera de impacto na área de estudos de variação linguística, sobretudo em termos de comunidades urbanas. Entre elas aparecem Rio de Janeiro, Belo Horizonte e São Paulo – esta última, o centro urbano que o presente artigo focaliza.

Nas palavras do próprio autor, sua lista é “inevitavelmente parcial” (LABOV, 2006); no caso brasileiro, além da importância dos trabalhos do PEUL, vale lembrar

* USP – Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas. São Paulo – SP – Brasil.
05508-010 - rbeline@usp.br

** USP – Universidade de São Paulo. Faculdade de Filosofia, Letras e Ciências Humanas. São Paulo – SP – Brasil.
05508-010 - livia.oushiro@usp.br

aqueles desenvolvidos no interior de projetos como o VARSUL, o VALPB e o ALIP,¹ entre vários outros. Em cada um deles, variedades regionais e sociais do Brasil vêm sendo extensivamente descritas através de várias publicações, na forma de artigos, dissertações e livros, entre os quais podem ser citados Callou, Moraes (1996), Monaretto (1997) e Hora (2004).

Pode-se dizer que São Paulo, contudo, é um centro urbano ainda pouco estudado sociolinguisticamente. Embora o português paulistano já tenha sido documentado e analisado através de importantes projetos de pesquisa, tais como o Projeto NURC-SP (CASTILHO; PRETI, 1986, 1987; PRETI; URBANO, 1998, 1990), o Projeto Para a História do Português Paulista (CASTILHO, 2007) e o Projeto Para a História do Português Brasileiro (CASTILHO, 2010), foram poucos os estudos, até recentemente, que tenham conceitualizado o falante paulistano sociolinguisticamente, enquanto membro de uma comunidade com características específicas, e que tenham se preocupado com a relação entre usos linguísticos e diferentes identidades sociais.

Devem ser lembrados aqui os trabalhos de Rodrigues (1987) e Coelho (2006). A primeira estudou a variação na concordância verbal em uma amostra de fala de 40 informantes analfabetos ou semialfabetizados de uma favela paulistana – amostra esta que incluiu não somente indivíduos paulistanos, mas também indivíduos provenientes de outras localidades do estado e do país. Neste sentido, além de vários outros interesses sociolinguísticos, pode-se dizer que Rodrigues (1987) é um dos poucos trabalhos que organiza uma amostra do português falado na cidade de São Paulo (ainda que circunscrito localmente numa favela) com objetivos variacionistas. O segundo investigou a variação no emprego dos pronomes de primeira pessoal do plural, bem como a concordância verbal com “nós”, em uma amostra de 24 entrevistas que o pesquisador pôde coletar, também numa favela paulistana, graças ao seu trabalho de observação etnográfica: visitou com frequência a comunidade e deu-se conta de que, ali, os indivíduos se organizavam em grupos cuja identidade era localmente definida: costureiras da cooperativa, trabalhadoras da creche, membros da associação do bairro e suas esposas, frequentadores do projeto para jovens, filhos dos membros da associação do bairro, “manos” dos times de futebol locais. Dessa forma, trata-se de um trabalho cuja amostra foi coletada de acordo com categorias que emergem das práticas sociais dos indivíduos em seus grupos, observados na comunidade, na linha dos estudos da terceira onda preconizados por Eckert (1996).

¹ Respectivamente, “Programa de Estudos sobre o uso da Língua” (<http://www.letras.ufrj.br/peul/>), no Rio de Janeiro; “Variação Linguística Urbana no Sul do País” (<http://www.pucrs.br/fale/pos/varsul>), que reúne pesquisadores de universidades dos três estados do sul do Brasil; “Variação Linguística no Estado da Paraíba” (HORA, 2004); e “Amostra Linguística do Interior Paulista”, no estado de São Paulo. (GONÇALVES, 2003).

Ainda que se lembrem os projetos de pesquisa e trabalhos acima, a cidade de São Paulo ainda carece de descrição sociolinguística. Com tal preocupação em foco, o Grupo de Estudos e Pesquisa em Sociolinguística da USP (GESOL-USP) vem executando a coleta de dados recentes da fala paulistana, no projeto “SP2010”.² Para propor tal projeto, foram previamente gravadas e transcritas, de 2008 a 2010, mais de cem entrevistas (que compõem o que aqui se chama de “Amostra SP-Piloto”), num exercício de observação sociolinguística da cidade, cujos objetivos principais foram promover o treinamento de jovens pesquisadores e definir rumos e parâmetros para o desenvolvimento do então futuro projeto.

Dessa forma, o presente artigo traz notícias acerca da execução de tal projeto de pesquisa (suas bases teórico-metodológicas, seus objetivos principais, desafios e encaminhamentos futuros), bem como reporta pesquisas que foram e vêm sendo desenvolvidas já com os dados da amostra SP-Piloto. As perguntas que congregam esses trabalhos (tanto de construção de *corpora* quanto de sua análise propriamente) são basicamente as seguintes: (i) Como construir uma amostra representativa da fala de uma cidade tão heterogênea e multicultural? (ii) O que significa falar como um paulistano e quais são as variáveis e variantes que o identificam? (iii) Que diferenças sociolinguísticas convivem na capital?

Embora essas perguntas remetam à definição do *paulistano* de maneira geral, como ser social e linguístico, a terceira das questões também tem sido abordada em termos mais específicos, da perspectiva de certas identidades urbanas – sobretudo no que concerne a questões de sexo/gênero e de classe social. Nesse sentido, esse artigo é organizado em três grandes partes: inicialmente, trata do Projeto SP2010, a partir do relato de experiências na construção da “Amostra SP-Piloto”; em seguida, traz notícia de trabalhos que colocam a comunidade paulistana no centro das atenções; finalmente, dá lugar a críticas ao empreendimento variacionista e ao seu possível endereçamento no trabalho sociolinguístico com a cidade de São Paulo.

A construção de uma (nova) amostra da fala paulistana

De 2008 a 2010, o GESOL-USP dedicou-se ao acúmulo de experiência nos procedimentos de coleta e organização de um *corpus* da fala paulistana. Nesse período, foram coletadas e transcritas mais de cem entrevistas, com informantes paulistanos e não paulistanos, de ambos os性os e de diferentes orientações sexuais, de 15 a 89 anos, com escolaridade equivalente ao Ensino Fundamental incompleto até o Ensino Superior completo, de estratos socioeconômicos diversos, residentes de 27 subdistritos e 59 bairros distintos, distribuídos por cinco zonas da capital paulista (Central, Norte, Sul, Leste e Oeste). Tendo em vista a reconhecida

² “SP2010: construção de uma amostra da fala paulistana” (Processo FAPESP nº 2011/09278-6).

complexidade sociodemográfica da cidade, a gravação dessas entrevistas teve um caráter exploratório e objetivou (i) elaborar e aprimorar um roteiro de entrevistas com paulistanos; (ii) elaborar e aprimorar métodos de abordagens a possíveis informantes; (iii) identificar problemas técnicos que podem ocorrer durante a gravação (p.ex., presença de ruídos externos) e metodológicos (p.ex., como conduzir a entrevista de modo que o falante se sinta à vontade), e elaborar soluções para tais problemas; (iv) elaborar critérios para a transcrição das entrevistas; e (v) identificar as variáveis extralingüísticas mais relevantes para a caracterização da fala paulistana.

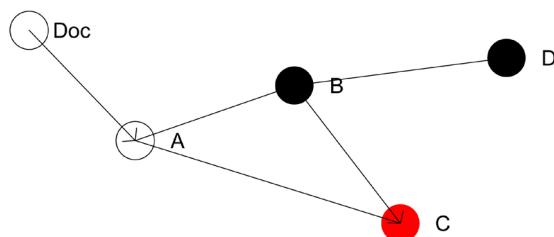
Do modo como foi elaborado, o roteiro de entrevistas objetiva, além de obter amostras de fala propriamente, reunir informações sobre o modo/condições de vida e avaliações sociolinguísticas dos informantes a respeito de certas variantes. Ele é dividido em duas partes: a primeira é mais pessoal e busca o envolvimento do falante nos tópicos apresentados (bairro, infância, família, trabalho, lazer), de modo que ele possa produzir tanto narrativas quanto descrições nos tempos passado (p.ex., através da pergunta “Como foi sua infância no bairro?”), presente (p.ex., “Nas horas de lazer, o que você e sua família gostam de fazer?”), e futuro (p.ex.: “Se você ganhasse na megassena, o que você faria?”). Embora se volte primariamente para questões que envolvam o informante, essa parte também inclui perguntas a respeito de temas mais gerais, que possibilitam a narrativa em terceira pessoa (p.ex., “Como é a vida em família hoje em São Paulo?”). A segunda parte trata de temas mais “objetivos” sobre a cidade (problemas e possíveis soluções para eles, caracterização da cidade e de seus habitantes) e busca uma fala cujo *estilo* – no sentido de Labov (2001b) – seja potencialmente mais distanciado e dissertativo/argumentativo. Essa parte também contém a leitura de uma lista de palavras, de uma notícia de jornal e de um texto com características da oralidade, na qual se espera maior grau de monitoramento da fala, em diferentes graus (LABOV, 2001b, 2006); ela se encerra com questões sobre as avaliações dos falantes a respeito de três variantes de variáveis sociolinguísticas específicas: a pronúncia ditongada de /e/ nasal (em palavras como “entendendo”), a ausência de concordância nominal (p.ex. “dois pastel”) e a pronúncia retroflexa de /r/ em coda silábica (em palavras como “porta”). Embora o roteiro se divida em dois momentos e diferentes subtópicos, ele permite uma transição bastante natural entre assuntos, com duração total média de 60 a 70 minutos, equilibradamente distribuídos entre as duas partes.

Também de acordo com a experiência prévia, constatou-se que a projeção de uma amostra verdadeiramente aleatória pode inviabilizar a realização da coleta. Os habitantes da cidade, de modo geral, mostraram bastante resistência a participar das gravações quando contatados por meios institucionais (através da USP e por associações de bairro), sem a intermediação de um conhecido em comum. Ainda

que algumas pessoas assim abordadas tenham aceitado participar das gravações, essas acabaram muitas vezes sendo mais curtas (de 20 a 40 minutos), uma vez que era visível o desconforto dos falantes em fornecer certas informações ou externar suas opiniões acerca dos temas arrolados. Por outro lado, a intermediação de um conhecido em comum se mostrou bastante produtiva; não só a maioria dos falantes que foram contatados desse modo se sentiu mais à vontade para receber os documentadores em suas residências (o que permite maiores observações que podem ser registradas no diário de campo), como também se revelaram bastante solícitos durante as entrevistas e na indicação de novos informantes.

Desse modo, a amostra é semialeatória, já que os informantes são selecionados e contatados a partir de uma adaptação do método “amigo do amigo” (MILROY, 2004) (ver Figura 1): os membros do GESOL-USP buscam, a partir de suas redes sociais (A), a indicação de paulistanos (B) que potencialmente se enquadram num dos 12 perfis elencados no Quadro 1, observando o critério de que a relação entre documentador e informante não deve ser próxima. A partir desses novos contatos, os documentadores pedem a indicação de novos informantes (pontos C e D) pelas pessoas já gravadas, a fim de ter acesso a um maior número de paulistanos, de perfis variados. Para garantir que um grupo de informantes não se concentre em uma mesma rede social, a nova pessoa indicada (C ou D) não deve ser conhecida da pessoa que indicou o informante atual (A). No exemplo da Figura 1, como C conhece A, C não poderá ser gravado; por outro lado, como D não conhece A, D pode ser um novo informante.

Figura 1 - Seleção de informantes por método semialeatório.



Fonte: Elaboração própria.

Embora se trate de um método semialeatório de amostragem, a partir dele é possível atingir um grande número de falantes de diferentes perfis (como se vê pela abrangência geográfica e sociodemográfica da amostra piloto), bem como obter amostras caracterizadas por maior naturalidade de fala. Os critérios de transcrição dessas gravações também passaram por diversas versões. A fim de permitir fácil manipulação em programas como o R (HORNIK, 2011) (para a extração automática de dados, por exemplo), estabeleceu-se que as transcrições seriam feitas em formato .txt (UTF-8), sem formatação especial como negritos, itálicos, TAB, colunas e tabelas, tampouco marcas gráficas para indicar alongamento, apagamento de segmentos ou hesitações.³ De acordo com a experiência piloto, quanto maior o número de critérios, mais demorada se torna a transcrição e menos padronizados os produtos finais.

A partir da análise qualitativa das gravações da amostra piloto e com base na literatura sociolinguística (CHAMBERS, 1995; CHAMBERS; TRUDGILL; SCHILLING-ESTES, 2004; LABOV, 1972, 1994, 2001a, 2006; RODRIGUES, 1987; TAGLIAMONTE, 2006), considera-se que são seis os parâmetros mais relevantes para a constituição sociolinguística da comunidade paulistana: Sexo/Gênero, Faixa Etária, Escolaridade, Classe Social, Região da Cidade e Geração da Família na cidade. A construção de uma amostra que contemplasse todas essas variáveis seria, contudo, inexequível a curto prazo; por isso, considerou-se, numa primeira fase, que a amostragem da fala paulistana poderia ser satisfatoriamente executada com base nas três primeiras, do modo como resume o Quadro 1 abaixo. Quanto às demais, elas não definem o tamanho da amostra, mas fazem parte dela qualitativamente.

Quadro 1 - Grupos de fatores definidores da amostra SP2010 e perfis sociolinguísticos dos informantes.

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

Fonte: Elaboração própria.

³ Aplicam-se, contudo, reticências para marcação de pausas e os sinais de exclamação e interrogação para indicar a entonação de certas sentenças.

Sexo/gênero

A menos que os propósitos de uma certa pesquisa variacionista fossem especificamente direcionados a uma comunidade de práticas constituída por indivíduos de apenas um dos dois sexos, não faria sentido que uma amostra não incluisse sujeitos-informantes de ambos. Neste sentido, tal grupo de fatores sociais não carece de maiores justificativas, enquanto categoria organizadora de uma amostra de fala. Entretanto, duas esferas de discussão merecem não apenas ser lembradas, como também merecem atenção no desenvolvimento deste projeto: uma delas diz respeito às expectativas de resultados de análises variacionistas, no que concerne ao sexo/gênero dos informantes; a outra se refere à necessidade de descrever as diferenças dos papéis sociais de homens e mulheres no caso específico da comunidade cuja fala se quer amostrar. Muitos são os trabalhos que se dedicam a fazer um quadro sinótico dos resultados das pesquisas sociolinguísticas que incluíram a variável sexo/gênero em suas análises – desde aquele que se considera o primeiro manual desenvolvido para a área (CHAMBERS, 1995), até manuais mais recentes (CHAMBERS; TRUDGILL; SCHILLING-ESTES, 2004; HOLMES; MEYERHOFF, 2005). Em todos eles, invariavelmente se mencionam, quando não se discutem extensivamente: (i) o fato de que, nas sociedades ocidentais especialmente, a frequência de variantes prestigiadas ou positivamente avaliadas numa dada comunidade é maior na fala das mulheres, relativamente à dos homens; (ii) em mudanças linguísticas “de baixo”, as mulheres podem se revelar mais conservadoras; (iii) em mudanças “de cima”, as mulheres tendem a se revelar mais inovadoras (GUY, 1990; LABOV, 2001a).

Ora, num projeto que se propõe a construir uma amostra recente da fala de uma grande comunidade urbana tal como São Paulo, é mister oferecer uma descrição acerca dos papéis sociais (gênero) correlacionados às diferenças entre os dois sexos. Por um lado, é necessário admitir a validade de uma crítica insistente, por parte de sociolinguistas que preconizam, em suas pesquisas, uma observação etnográfica e maior atenção às práticas sociais: aquela que aponta para o fato de que, embora se fale em termos de diferenças gênero, na verdade se trabalha – por vezes, de maneira essencialista – com a diferença de sexo (ECKERT; MCCONNELL-GINET, 1992). Por outro lado, papéis sociais estão em constante processo de construção e reconstrução e, no Brasil, São Paulo talvez seja uma das cidades em que estilos de vida, hábitos urbanos e práticas sociais em geral mudem mais rapidamente. Dessa forma, é importante verificar, no caso de uma comunidade de fala ainda não suficientemente estudada como a cidade de São Paulo, se as tendências sociolinguísticas descritas em inúmeros estudos observam-se também aqui. Além disso, o desenvolvimento deste projeto também traz subsídios para outras pesquisas que se interessam mais especificamente pela variável sexo/gênero, no formato de um elenco de informações sobre as

práticas sociais dos indivíduos, homens e mulheres, que são selecionados para a construção desta amostra – ainda que de maneira relativamente distanciada, como é o caso da entrevista sociolinguística.

Faixa etária

Assim como sexo/gênero, o grupo de fatores faixa etária está entre aqueles mais analisados em estudos sociolinguísticos. Sua ampla utilização na estratificação de amostras de fala se baseia no conceito de *mudança em tempo aparente* e na hipótese de que é possível detectar mudanças linguísticas em progresso através da comparação da fala de pessoas de diferentes idades pertencentes à mesma comunidade (LABOV, 2001a). Além disso, assim como sexo/gênero, a faixa etária frequentemente se revela correlacionada a variáveis linguísticas cujas variantes se diferenciam em graus de prestígio numa comunidade. Desse modo, a inclusão desse grupo de fatores permite investigar, nos trabalhos que se desenvolvem com base nesta amostra, se certos fenômenos linguísticos encontram-se em variação estável ou em mudança no português paulistano, além de avaliar diferentes graus de prestígio ou de estigma de certas variantes.

Cabem aqui, entretanto, alguns esclarecimentos quanto à segmentação etária definida. Esta amostra inclui falantes (i) de 20 a 34 anos, (ii) de 35 a 59 anos, e (iii) com 60 anos ou mais. Trata-se, portanto, apenas de falantes em idade adulta, uma vez que a inclusão de crianças e de adolescentes poderia aumentar significativamente o número mínimo de entrevistas a serem gravadas, inviabilizando a obtenção de 5 informantes por perfil. Por outro lado, a divisão em pelo menos três faixas faz-se necessária a fim de diferenciar variantes em retração ou difusão (em que normalmente se verifica um padrão de curva ascendente ou descendente) e variantes de variáveis estáveis (em que normalmente se observa um padrão de curva em "U").

A presente classificação toma por base, principalmente, a posição relativa das faixas etárias no mercado de trabalho, mas também leva em conta modos/estilos de vida de cada grupo, caracterizados aqui *in abstracto*. O grupo de falantes mais novos, aqueles entre 20 e 34 anos de idade, engloba jovens adultos que, em geral, encontram-se relativamente menos estabilizados do que pessoas das outras duas faixas etárias: na cidade de São Paulo, não é raro encontrar pessoas com até 34 anos de idade que não tenham se casado, que não têm casa própria, que estão fazendo faculdade ou cursos profissionalizantes ou que, de modo mais amplo, levam um estilo de vida mais semelhante ao de pessoas com vinte e poucos anos. O grupo de falantes entre 35 e 59 anos, por sua vez, pretende abranger pessoas mais plenamente inseridas no mercado de trabalho e relativamente mais estáveis (ou que buscam mais seriamente certa estabilidade).

Por fim, o grupo de falantes com mais de 60 anos se refere de modo amplo a pessoas aposentadas ou próximas da aposentadoria.

Essa caracterização, é claro, não é uma regra para todos os falantes – certamente existem pessoas com 23 anos que são casadas e que têm emprego estável, assim como há pessoas com 47 anos que são solteiras e que não têm estabilidade profissional –, mas parece adequada às fases da vida na metrópole. Ao mesmo tempo, não se pretende dar a entender que falantes contatados que não se encaixam na descrição acima não são considerados na amostra; a proposta de 5 informantes por perfil objetiva justamente permitir relativizar tais variações entre indivíduos.

Como se trata de um contínuo, o agrupamento em determinadas faixas etárias não deixa de ser arbitrário até certo ponto, e pode, em princípio, ser reorganizado de outras maneiras. Na construção da amostra SP2010, a ficha de cada informante indica a sua idade (e não apenas a sua faixa etária), de modo que diferentes pesquisas terão a possibilidade de reagrupar os falantes, se assim considerar necessário, ou tratar esse grupo de fatores como variável contínua.

Escolaridade

A escolaridade dos falantes (dividida em dois níveis: até Ensino Médio; Curso Superior completo ou incompleto) compõe a terceira variável extralingüística que estratifica a amostra da fala paulistana em construção no Projeto SP2010. Assim como sexo/gênero e faixa etária, trata-se de uma variável bastante analisada nos estudos sociolinguísticos brasileiros e serve, em princípio, como parâmetro para analisar o impacto da instrução formal no emprego de formas linguísticas. A expectativa é que falantes com graus mais elevados de instrução apresentem maior tendência de emprego de formas consideradas padrão ou de prestígio na comunidade e evitem as formas consideradas não padrão ou estigmatizadas.

Ao mesmo tempo, tal variável tem sido tradicionalmente empregada como modo indireto de avaliar a classe social do falante (RODRIGUES, 2009), uma vez que o grau de instrução se correlaciona com atividades profissionais possíveis e respectivos salários médios dos indivíduos. No entanto, se a realidade socioeconômica brasileira na década de 1970, em que se começaram a coletar amostras de língua falada no país, permitia uma equação relativamente pacífica entre classe social e escolaridade do falante, por outro lado, mudanças nas políticas públicas de educação nas últimas décadas, como a Progressão Continuada (no estado de São Paulo) e a ampliação do acesso aos cursos universitários através do ProUni (no âmbito federal), permitem um sério questionamento dessa equivalência na década de 2010, sobretudo na cidade de São Paulo.

A divisão em apenas dois níveis de escolaridade é decorrência dessas mudanças. A partir da experiência piloto, constatou-se que é cada vez mais raro encontrar paulistanos nativos que não tenham pelo menos ingressado no Ensino Médio (embora tal realidade ainda seja relativamente mais frequente entre migrantes),⁴ uma estratificação mais complexa, além de aumentar o número mínimo de informantes, poderia conduzir à inviabilidade de preencher certas células (por exemplo, falantes da primeira faixa etária que não tenham concluído o Ensino Fundamental). Sugere-se aqui que escolaridade deva ser tratada como um fator de composição da classe social do falante, mas não como sua substituta. Desse modo, a inclusão dessa variável no presente projeto tem o objetivo de instrumentalizá-la posteriormente no desenvolvimento de critérios para o estudo da variável classe social.

Classe social, região da cidade e geração da família na cidade

Essas variáveis, conforme já se apontou, não definem o tamanho da amostra. Entretanto, parece evidente que as três variáveis anteriormente discutidas não abarcam, por si sós, características que potencialmente distinguem São Paulo de outras comunidades linguísticas, dadas as suas especificidades históricas, culturais e sociodemográficas. Nesse sentido, vale também discutir a importância e os modos como essas três variáveis adicionais são observadas na execução do projeto, mantendo-se em perspectiva a dupla finalidade de (i) aumentar a representatividade da amostra; e (ii) aprofundar reflexões sobre a possibilidade de operacionalizar essas variáveis em amostras futuras.

Classe social parece apresentar-se como um fator relevante para a diferenciação sociolinguística na cidade de São Paulo. Nas gravações da experiência piloto, em resposta à pergunta “Dentro da cidade de São Paulo, você consegue identificar se a pessoa é de alguma região ou bairro específico da cidade?”, frequentemente se obtiveram respostas tangentes, que mencionam a classe socioeconômica dos falantes, a despeito de que a pergunta houvesse sido formulada com base em um parâmetro geográfico. Adicionalmente, em um experimento controlado de percepções (MENDES; OUSHIRO, 2011a, 2011b) com aplicação da técnica *matched-guise* (CAMPBELL-KIBLER, 2009), constata-se que classe social é uma das variáveis extralingüísticas mais significativamente correlacionadas à pronúncia variável de /-r/ em coda silábica na cidade de São Paulo: os ouvintes paulistanos atribuíram, em média, níveis de classe social mais baixos quando ouviram estímulos em que falantes empregavam a variante retroflexa do que quando ouviram a variante tepe.

⁴ De fato, de acordo com os Resultados Gerais do Censo Demográfico (IBGE, 2010, p.84 ss.), o percentual de pessoas em idade escolar (7-14 anos e 15-17 anos) que não frequentam escola diminuiu de modo geral em todo o país nos últimos 10 anos, de 22,6% para 16,7%, sendo a Região Sudeste aquela que apresenta menor taxa de evasão escolar (de 20,1% para 15,0%). Tal fato se percebe facilmente no trabalho de campo.

No entanto, classe social é uma variável difícil de abordar. Em primeiro lugar, porque seria necessário estabelecer critérios *sociolinguísticos* – e não estritamente econômicos – que permitissem o estudo da comunidade paulistana. Tais critérios possivelmente devem abranger, além da renda familiar e do já mencionado nível de escolaridade, o tipo e o local de residência, a ocupação, o acesso a bens culturais etc. (ASH, 2004).⁵ Paralelamente à construção da amostra SP2010, Faria (2012) vem trabalhando na definição de critérios suficientemente objetivos para a proposta de uma classificação socioeconômica dos falantes paulistanos.⁶ A observação de diferenciações socioeconômicas na construção da Amostra SP2010 se baseia em perguntas formuladas no roteiro da entrevista sociolinguística com todos os informantes (p.ex., sobre o nível de escolaridade, ocupação, área de residência do informante, meios de locomoção utilizados, formas de lazer, lugares a que já viajou etc.), e no preenchimento de um questionário específico (questões no formato de múltipla-escolha sobre, p.ex., a renda individual e familiar, tipo e posse de habitação etc.). Assim, no Projeto SP2010, a classe social dos falantes é determinada a *posteriori*, e um determinado estudo poderá se valer de tais diferenciações, de acordo com seus objetivos.

Figura 2 - Divisão da cidade de São Paulo em Centro Velho, Centro Expandido e Periferia.



Fonte: SÃO PAULO (2011)⁷.

⁵ Os critérios do IBGE para a classificação socioeconômica dos falantes baseiam-se exclusivamente na renda dos habitantes, o que não parece dar conta dos diferentes níveis de status social na cidade de São Paulo e de suas correlações com usos linguísticos.

⁶ Esse trabalho é descrito com mais detalhes na terceira parte deste artigo.

⁷ A divisão da cidade em três áreas foi feita pelo GESOL-USP, de acordo com os critérios descritos.

De modo semelhante, o local de habitação do falante em diferentes regiões da cidade tem potencial para revelar diferenciações sociolinguísticas. Embora a cidade seja dividida administrativamente em 31 subprefeituras, distribuídas em nove zonas, dois tipos de divisão geográfica surgiram mais frequentemente no discurso de paulistanos nas entrevistas piloto: (i) a divisão entre as zonas Central, Norte, Sul, Leste e Oeste; e (ii) uma divisão entre Centro Velho, Centro Expandido e Periferia (ver Figura 2, acima). A divisão em cinco zonas não apresenta maiores dificuldades, visto que os seus limites são bastante claros. Já a proposta de limites para a divisão em três regiões – Centro Velho, Centro Expandido, Periferia Próxima –, tem base na história de ocupação dos bairros e no grau de desenvolvimento urbano em termos de verticalização e serviços (OUSHIRO, 2011b). Na primeira região, incluem-se bairros como Bela Vista, Consolação, Bom Retiro, Brás, Liberdade, que representam a São Paulo “mais antiga”, em que se encontram mais facilmente informantes cujas famílias estão enraizadas na cidade há mais tempo (há duas ou mais gerações); esses falantes podem apresentar uma tendência maior de empregar formas linguísticas consideradas prototípicamente paulistanas. No Centro Expandido, incluem-se bairros como Pinheiros, Santana, Vila Mariana, Moema, Tatuapé. Assim como o Centro Velho, tais bairros já são amplamente verticalizados e possuem uma infraestrutura relativamente bem desenvolvida de serviços (ônibus, metrô, hospitais, todos os tipos de comércio). No entanto, diferentemente do Centro Velho, aí se localizam os imóveis mais caros da cidade e que relativamente mais se valorizam. Na região mais periférica, incluem-se bairros como Sapopemba, Pirituba, Jabaquara, Freguesia do Ó, Vila Matilde, que são historicamente mais recentes, assim como seu processo de verticalização; neles, ainda é possível ver crianças brincando na rua e vizinhos conversando ao portão de suas residências, fatos de sociabilidade que potencialmente têm consequências para o repertório sociolinguístico desses habitantes. A Figura 2 representa essa divisão da cidade em três regiões aproximadamente concêntricas em relação ao seu centro político e geográfico. Na coleta de entrevistas da presente amostra, o número de informantes por perfil é equilibrado a partir dos dois tipos de categorização geográfica: um informante por zona e um ou dois informantes por região. Esse critério visa a cumprir a tarefa de obter uma amostra representativa e minimamente abrangente da cidade.

Finalmente, quanto à geração da família na cidade, a proposta do Projeto SP2010 amostra é incluir apenas informantes paulistanos (ainda que seus pais possam não sê-lo), e que não se considerem migrantes e imigrantes. Embora estes últimos certamente tenham um papel importante na constituição sociolinguística da cidade, a sua inclusão pressuporia responder certas questões que parecem colocar-se em outro momento da pesquisa, posterior àquele em que correntemente se encontra, tais como: (i) no caso dos migrantes, quantas e quais regiões de origem deveriam ser consideradas? (ii) no casos dos imigrantes, incluir-se-iam “apenas”

aqueles mais tradicionalmente considerados (os japoneses, os italianos, os árabes), ou também grupos de imigração mais recente (os bolivianos, os coreanos)? (iii) no caso de ambos, quais gerações seria importante incluir na amostra? Os mesmos critérios poderiam ser aplicados aos diferentes grupos?

Na amostra em construção, os documentadores devem obter informações acerca de que geração de paulistanos o falante é representante – se os pais (um deles ou ambos) são paulistanos ou não, se um ou mais avós são paulistanos – e da(s) origem(ns) da primeira geração da família que veio para São Paulo (de que outra cidade, estado ou país). Tais informações fazem parte da Ficha do Informante e da Gravação, de modo que, ainda que não constituam um dos parâmetros de estratificação desta amostra, poderão ser objeto de análises linguísticas (para investigar, por exemplo, se a fala de paulistanos exibe padrões quanto ao grau de enraizamento da família na cidade).

O Projeto SP2010 se constitui, então, como um novo movimento na amostragem da fala paulistana, com objetivos variacionistas bem delimitados e definidos sobretudo com base na noção laboviana de comunidade de fala. A partir das experiências acumuladas na construção da Amostra SP-Piloto, seu desenvolvimento poderá levar ulteriormente a propostas de construção de amostras que se definam diferentemente e que complementem a que se desenvolve no presente, aproximando-se progressivamente de um quadro sociolinguístico mais rico da capital paulista; o trabalho de observação da comunidade que tem continuidade ao longo da execução do projeto fornecerá subsídios para o amadurecimento das questões levantadas. Ao final de 2013, a amostra construída pelo Projeto SP2010 deverá ser disponibilizada em página própria na Internet, de modo que outros centros de pesquisa possam utilizá-la.

Estudando o português falado na cidade de São Paulo

Conforme já se indicou anteriormente, alguns trabalhos foram e vêm sendo desenvolvidos já com os dados da Amostra SP-Piloto. Entre os trabalhos já finalizados, encontram-se as dissertações de mestrado de Oushiro (2011a) e Nascimento (2011).

A partir de 53 entrevistas (além de outras amostras de língua escrita), Oushiro (2011a) analisa a variação entre quatro estruturas de Interrogativas-Q no português paulistano contemporâneo: (i) interrogativas-qu (“Onde você mora?”); (ii) interrogativas qu-que (“Onde que você mora?”); (iii) interrogativas é-que (“Onde é que você mora?”); e (iv) interrogativas qu-in-situ (“Você mora onde?”). Um dos aspectos mais interessantes desse trabalho é a discussão da equivalência semântica entre as formas interrogativas, estabelecida com base no conceito de

pressuposição do falante (STALNAKER, 2002) – num exercício de superação do debate Lavandera-Labov (LAVANDERA, 1978; LABOV, 1978). Além disso, Oushiro (2011a) utiliza o conceito de competência comunicativa (HYMES, 1991) como critério para definir os envelopes de variação com que trabalha: diferentes estruturas são consideradas variantes se forem possíveis, factíveis, adequadas e efetivamente empregadas nos mesmos contextos. Com base nesses conceitos, definem-se dois envelopes de variação – e, portanto, duas variáveis: uma envolve a alternância na posição do constituinte interrogativo (*in situ* ou não), e outra que encerra as três estruturas com constituinte interrogativo pré-verbal (-qu, qu-que e é-que).

Em resumo, os resultados das análises quantitativas mostram que interrogativas qu-*in situ* são favorecidas principalmente por fatores morfossintáticos e discursivo-pragmáticos.⁸ Entretanto, da perspectiva social, um dos resultados mais interessantes dessa pesquisa com dados paulistanos é aquele que revela um diferenciação entre homens e mulheres: eles tendem a empregar diferentes estruturas sintáticas em relação a uma mesma estratégia discursiva – para as mulheres, a estrutura de interrogativas qu-*in-situ* parece estar mais associada a perguntas retóricas do que para os homens, que a desfavorecem fortemente no mesmo contexto. Ao cruzar fatores discursivo-pragmáticos com a variável sexo/gênero, Oushiro (2011a) revela evidências muito favoráveis à hipótese de Cheshire (2005) a respeito do encaixamento social indireto da variação sintática (diferentemente do que em geral se observa quando se trata de variáveis fonético-fonológicas). Já no que toca ao emprego de interrogativas qu-que, a autora aponta para evidências de mudança linguística em progresso, uma vez que sua análise em tempo aparente revela o favorecimento da estrutura pelos falantes mais jovens. Neste caso, demonstra que a variação é influenciada principalmente por fatores sintáticos e prosódicos.

Nascimento (2011), por sua vez, trata da variação no uso do pronome de segunda pessoa, cujas formas alternativas empregadas pelos paulistanos são “você(s)” e “cê(s)”. A partir de 36 entrevistas com paulistanos, gravadas pelo GESOL-USP, e de 18 entrevistas do NURC-SP, a autora analisou essa variável sincrônica e diacronicamente. Seus resultados mostram um equilíbrio na distribuição das duas variantes nos seus usos por paulistanos. Embora se evidencie mudança em progresso pelos dados do NURC-SP (coletados na década de 1970), a análise com os dados dos anos 2000 revela um caso de variação estável na comunidade de fala.

Com a amostra recente, Nascimento (2011) mostra que a variante inovadora “cê” tende a ocorrer mais na fala dos mais jovens e das mulheres acima de 50 anos de idade, bem como nas interações entre pessoas que se conhecem mais (comparativamente

⁸ Para uma discussão pormenorizada, mas mais breve e objetiva do que numa dissertação de mestrado, veja-se também Oushiro (2011c).

a indivíduos que são completos desconhecidos entre si). Por outro lado, a forma “cê” tende a ser evitada entre os mais escolarizados. Linguisticamente, o emprego das variantes revela-se correlacionado a fatores semânticos e discursivos, sobretudo à referência do pronome. No que concerne a questões de gramaticalização, os resultados a que chegou a pesquisadora corroboram a hipótese de cliticização da variante reduzida, mas fornecem um contra-argumento à correlação direta entre “erosão fonética” e “abstratização semântica”.

Além dessas pesquisas já finalizadas, as entrevistas coletadas pelo GESOL-USP têm nutrido outras várias, em andamento. Rocha (2012a, 2012b) vem estudando, no nível de mestrado, a estruturação variável de sentenças negativas, com o emprego do advérbio de negação apenas antes do predicado verbal ou com seu emprego antes e depois do verbo, como em “Não vou mais fazer faculdade” (NEG1) e “Não vou mais fazer faculdade não” (NEG2). A partir de 48 entrevistas, Rocha (2012a, 2012b) constata três fatos interessantes: (i) a estrutura NEG2 parece bastante infrequente entre paulistanos (menos de 10% do total de dados); (ii) a variação não parece possível quando se trata de negação de proposição que veicula informação completamente nova no discurso (parece variável quando a informação da proposição que está sendo negada é ativada, direta ou indiretamente, no discurso precedente); (iii) a estrutura NEG2 parece mais frequente entre paulistanos cujos pais são oriundos do nordeste brasileiro. Some-se a esses o fato observado de que uma terceira estruturação possível, em que o advérbio de negação aparece apenas após o verbo – como em “Vou mais fazer faculdade não” (NEG3) – parece muitíssimo infrequente na fala paulistana.

Esses são resultados preliminares que ainda carecem de análises ulteriores, mas sua discussão apresenta-se promissora, não apenas no que diz respeito à descrição do fenômeno na fala paulistana, mas também no que concerne à definição do envelope de variação – tarefa sempre complexa, mas das mais instigantes para o campo da sociolinguística em que se dedica a variáveis que estão “além” do nível fonético-fonológico.

Em outra pesquisa em nível de mestrado, Silva (2011) focaliza a variável concordância nominal (tanto no interior de sintagmas nominais, como em casos de predicativos), a partir de duas amostras; uma delas é constituída de 36 informantes paulistanos *de facto* (nascidos e criados em São Paulo, dentre os já gravados pelo GESOL-USP) e a outra é constituída de 36 alagoanos que se mudaram para a capital paulista (há mais de dez, menos de dez e menos de dois anos). Sua proposta, além de verificar os fatores sociais e linguísticos que regulam esse caso de variação nessa comunidade, é também verificar até que ponto esses subgrupos se assemelham e se dessemelham. Uma das principais contribuições potenciais desse trabalho, parece claro, é a inclusão da figura do migrante num estudo variacionista em São Paulo, de maneira controlada e rigorosa. Esse trabalho

tem preocupações semelhantes às da análise de Mendes (2011a, 2011b) que, com base em uma primeira amostra de entrevistas com falantes paraibanos enraizados em São Paulo, discute as tendências de acomodação de fala de um dos muitos grupos de migrantes residentes na cidade.

Um outro trabalho em que se compararam amostras é a pesquisa de doutorado de Vieira (2012). Neste caso, trata-se do estudo de duas variáveis (a estrutura de orações relativas – padrão, copiadora e cortadora – e o preenchimento variável de preposição em sintagmas preposicionados) com dados da fala paulistana e campo-grandense. Os dados da capital do Mato Grosso do Sul estão sendo coletados de acordo com os perfis definidos pelo Projeto SP2010 (e a partir da experiência com a constituição da Amostra SP-Piloto), com as devidas adaptações que se julgaram necessárias (no roteiro de entrevista, por exemplo) no sentido de coletar informações etnográficas específicas sobre os informantes naturais de Campo Grande. Para as estratégias de relativização, a hipótese de partida de Vieira (2012) é a de que, na norma culta, a variação entre as orações relativas está centrada nas variantes padrão e cortadora, enquanto que o emprego da variante copiadora, de caráter estigmatizado, estaria correlacionado a fatores discursivo-pragmáticos. Para o apagamento de preposições, por outro lado, a pesquisadora questiona, sobretudo, sua produtividade entre falantes cultos.

A correlação entre variáveis na definição de uma comunidade de fala, de maneira geral, não é frequentemente observada, na divulgação da pesquisa sociolinguística brasileira. A partir dessa constatação, e interessada em ir mais longe no estudo da comunidade de fala paulistana, a pesquisa de doutorado de Oushiro (2011b) propõe investigar a produção e a percepção de variantes linguísticas na cidade de São Paulo que parecem estar correlacionadas a diferentes identidades urbanas: a realização de /r/ em coda silábica como tepe ou retroflexo; a pronúncia de /e/ nasal como monotongo ou ditongo; a presença ou ausência de morfema de plural em sintagmas nominais; e a presença ou ausência de morfema de plural em sintagmas verbais. Seu objetivo central é examinar a distribuição linguística e social das variantes, a fim de verificar (i) que identidades urbanas estão associadas ao seu emprego; (ii) como funcionam os mecanismos de adesão ou resistência a certas normas linguísticas em relação a diferentes grupos sociais; e (iii) se a cidade de São Paulo pode ser considerada uma única comunidade de fala. Para a investigação de produção linguística, Oushiro analisa as mais de cem entrevistas sociolinguísticas coletadas em 2008-2010 pelo GESOL-USP; para a análise de percepção linguística, aplica a técnica *matched-guise* em testes com controle de variantes (MENDES; OUSHIRO, 2011a 2011b). Além de contribuir para a descrição e a análise de fatos linguísticos no português paulistano, o exame dessas quatro variáveis sociolinguísticas vai permitir a comparação de padrões de covariância entre variáveis dependentes, bem como deverá levar a uma melhor

compreensão dos processos que conduzem à criação de identidades sociais através de usos linguísticos.

Finalmente, mas não menos importante, deve ser mencionado o trabalho de iniciação científica de Ciancio (2012), que se propõe a construir uma subamostra da fala paulistana especificamente dos moradores do bairro do Itaim Bibi, com dois objetivos centrais: (i) estudar possibilidades de ampliação e continuação do Projeto SP2010; e (ii) encaminhar o estudo da fala de classes mais altas em São Paulo. O interesse nesse bairro se justifica na medida em que, na última década, vem se valorizando extremamente, o que pode ser exemplificado pela alta nos preços dos imóveis, assim como pela multiplicação de empreendimentos prestigiosos, tais como restaurantes, bares, hotéis, cinemas etc. Considere-se ainda que o surgimento do Itaim Bibi como um bairro popular,⁹ por volta dos anos 1920-1930, contrasta com essa valorização dos últimos dez anos, de modo que é possível encontrar e dividir seus moradores da seguinte maneira: (i) paulistanos que moram no Itaim Bibi há um longo tempo (mais de quinze anos) e que, portanto, não escolheram o bairro em virtude de seu desenvolvimento; (ii) pessoas que moram no Itaim Bibi há menos de quinze anos, que acompanharam sua expansão e contribuíram com ela; e (iii) recém-chegados (residentes no bairro há menos de dois anos) que possivelmente buscam os benefícios dessa expansão. Essa variável, de fato, define-se como “grau de enraizamento dos habitantes no bairro”, e seu estudo deverá trazer contribuições interessantes para futuros estudos sociolinguísticos na cidade de São Paulo.

Identidades urbanas

A construção de amostras de fala em comunidades urbanas, e o subsequente estudo de variáveis sociolinguísticas a partir dos dados coletados, em geral têm, na sua base epistemológica, a premissa de que o uso de formas alternativas é linguisticamente indiferente, mas socialmente significativo (CHAMBERS, 2004). Apesar do inegável esforço que caracteriza o empreendimento variacionista – afinal de contas, não é possível descrever e explicar os usos linguísticos numa comunidade de fala a partir de poucos dados ou de dados intuídos – é válida a crítica, muitas vezes advinda de vertentes de estudos sociolinguísticos que se pautam por questões de natureza mais etnográfica e antropológica, de que os estudos variacionistas pecam por lidar com categorias pré-estabelecidas. Em outras palavras, muitas vezes os estudos variacionistas são criticados em virtude do fato de que a descrição e explicação das estruturas linguísticas em uso se deem em detrimento de real observação da realidade social e de categorias “locais”, em vez de pretensamente “universais”.

⁹ Ver: SÃO PAULO ([2011]); ITAIM BIBI ([2011]).

Duranti (2001) esclarece e resume essas críticas em três pontos: (i) não é universalmente aceito, nas ciências sociais, o tratamento de categorias tais como classe social, sexo/gênero, raça e geração enquanto variáveis independentes entre si; (ii) é comum a ausência, nos estudos sociolinguísticos quantitativos, de uma definição de “contexto” como algo que está sempre mudando e que tem na referência a fala um de seus elementos constitutivos; (iii) se a fala é uma realização interacional, não se deveria ter a entrevista sociolinguística como o método mais confiável para acessar falas espontâneas. Todos esses pontos, embora não invalidem as décadas de estudos labovianos ao redor do globo, representam desafios para o futuro do campo.

Com essas questões em mente, o GESOL-USP vem se dedicando também ao trabalho de observar e definir categorias sociais, do modo como elas se constroem localmente no espaço urbano paulistano. Evidentemente, isso vem sendo feito paralelamente a – e não em detrimento de – todas as atividades acima descritas. Colocar-se nessa via intermediária talvez implique incessável atenção à premissa de que a performance linguística e a expressão/construção de uma identidade social estão ligadas por uma via de mão dupla: de um lado, a linguagem porta significados simbólicos e sociais; de outro, os falantes dão-se conta dessa propriedade da linguagem e se valem disso para expressar e para veicular significados sociais.

É o caso, portanto, de observar a variação no estilo de linguagem (no sentido de Eckert (2001)), a fim de se ter acesso às variantes linguísticas que um determinado falante está empregando em determinado momento e situação, bem como às funções sociais específicas que tais variantes desempenham. Antes de tratar especificamente de trabalhos que têm sido desenvolvidos nessa linha, vem ao caso organizar seus pressupostos mais básicos. Primeiramente, é preciso rever como se dá a associação entre uso linguístico e significado social, para depois passar à questão da variação no estilo.

Labov (2006), no seu estudo sobre a estratificação social do inglês falado em Nova Iorque, é considerado um dos pioneiros no demonstrar que variantes linguísticas discretas podem veicular informações sociais que caracterizam o sujeito-falante. Essas variantes podem ser mapeadas numa estratificação social mais ampla, caracterizando o grupo de indivíduos que as empregam. No esforço por desenvolver uma teoria social da linguagem, começou-se a explorar a possibilidade de que as variantes linguísticas não apenas caracterizam a linguagem de um certo grupo social, mas também podem ser usadas por indivíduos que queiram indicar sua afiliação a uma determinada comunidade. O exame desse papel de “índice” que os usos da linguagem podem desempenhar já chegou a ser feito por Trudgill (1974) e Milroy (1980) – em estudos da fala da classe operária de

Norwich e Belfast, respectivamente –, e por Bell (1999), que estudou a linguagem dos noticiários de rádio em Auckland, Nova Zelândia.

Trudgill (1974), em seu exame da pronúncia variável do -*ing* em final de palavra, observou que os homens alegavam usar a variante não padrão [In], enquanto que as mulheres preferiam empregar a variante padrão [Ih]. Essa descoberta ilustra o fato de que os falantes de uma língua fazem avaliações acerca dos seus diferentes estilos de fala e que têm consciência delas. Em seu estudo, Trudgill (1974) demonstra que o emprego de uma variante bem como a avaliação de seu emprego, sobretudo no caso dos homens, serviam ao propósito de marcar seu pertencimento a um determinado grupo social, atestando a associação simbólica entre linguagem e comunidade e oferecendo uma das primeiras contribuições para os estudos da linguagem como meio de construção de identidade.

Milroy (1980) introduziu nos estudos sociolinguísticos a noção de redes sociais. Em seu estudo de três comunidades operárias de Belfast, a autora argumenta que as propriedades que caracterizam a rede social de um indivíduo influenciam sua performance linguística. Ela também afirma que as redes sociais mais densas – aquelas em que todos os membros se inter-relacionam intensamente e interagem uns com os outros numa diversidade de situações – constituem um mecanismo poderoso de normativização de sua fala. Tal mecanismo acaba desempenhando papel decisivo na homogeneização de crenças e práticas sociais dentro do grupo, o que obviamente inclui as práticas linguísticas. Está implícita nessa noção a consciência, tanto da parte dos indivíduos como da parte da comunidade, das práticas sociais que os definem ou que os caracterizam. A conformidade com o padrão linguístico que define o grupo requer a percepção do que constitui um padrão, além da noção de que a linguagem funciona como “índice” de afiliação social dos indivíduos a seu grupo.

Bell (1999), em seu exame da linguagem dos noticiários de rádio em Auckland, analisou a variação no nível individual como um processo que ele chama “*Design da Audiência*”, a partir do que se reconhece como Teoria da Acomodação (GILES; POWESLAND, 1997). O princípio básico dessa teoria é o de que um falante pode optar entre acomodar-se à fala de seu interlocutor – assemelhando sua performance linguística à dele – ou distanciar-se dela. Essa escolha tem significação social no contexto discursivo e indica solidariedade ou diferenciação social, respectivamente. A acomodação linguística é analisada como uma resposta sob a forma de alternância no estilo de falar, na qual o interlocutor adota práticas linguísticas variáveis com base na linguagem do outro. Bell (1999) generalizou tal noção numa teoria de “alternância de estilo”, de modo que a variação observada no nível individual pode ser quantitativamente correlacionada ao interlocutor ou ao tópico da conversa.

A teoria da acomodação está fundada no “Axioma do Estilo”, de acordo com o qual a variação observada na fala do indivíduo mimetiza a variação observada entre dois grupos sociais. Em outras palavras, o estilo de linguagem deriva seu significado social da avaliação que o falante faz das relações entre traços linguísticos e grupos sociais. Apesar de manter a generalização de que as comunidades de fala são caracterizadas por traços linguísticos específicos, e a de que ambos são socialmente avaliados pelo falante, essa teoria é inovadora na medida em que traz a noção de que os próprios falantes podem, individualmente, variar seu estilo de fala a fim de atingir diferentes objetivos sociais. Se os falantes podem adotar traços linguísticos socialmente marcados no intuito de evidenciar sua associação ou dissociação diante de seu interlocutor (unitário ou coletivo), tem-se aí o papel do indivíduo na construção da identidade linguística.

Se, por um lado, essas teorias fornecem um bom modelo para a implementação de uma linguagem socialmente marcada, por outro lado elas deixam a desejar no que diz respeito ao desenvolvimento das diversas avaliações que os fatos linguísticos podem permitir. Entretanto, essa lacuna vem sendo preenchida por trabalhos sobre a interação entre linguagem e ideologias sociais. A teoria da indexação de Ochs (1991) oferece um modelo de análise das relações entre linguagem e identidade, no qual traços linguísticos funcionam como “índices” de ideologias sociais. De acordo com essa autora, a relação entre variantes e identidades é indireta, e, por isso, são raras as variantes linguísticas que funcionam como “índices” de uma única e exclusiva identidade social. Para Ochs, elas funcionam antes como “índices” de atitudes, posicionamentos, atividades, que por sua vez estão associadas a categorias sociais.

Com isso, observa-se o desenvolvimento de uma teoria rigorosa sobre os modos como o significado social e sua avaliação se mapeiam sobre a estrutura linguística. Os indivíduos têm consciência dessa propriedade da linguagem, e manipulam-na para a construção de sua identidade social. Bell (1999) considera que o indivíduo pode alternar seu estilo de fala não apenas como uma “resposta” ao seu interlocutor real, mas também por acomodação a um tópico e/ou interlocutor não presentes no contexto de fala. Ou seja, mais uma vez, o indivíduo num dado contexto de fala faz uso de seu conhecimento da relação entre traços linguísticos e identidade social, adotando marcas para “indexar” sua afiliação a um determinado grupo.

Define-se, dessa forma que pode ser identificada como mais “eckertiana”, uma linha de estudos que se ocupa mais detidamente das variantes linguísticas como índices de posicionamentos e afiliações sociais. Evidentemente, essa linha não implica a negação daquela mais laboviana – até porque se constitui a partir dela. No GESOL-USP, os estudos e discussões nessa linha têm-se debruçado principalmente sobre identidades de sexo/gênero e de classe social.

Sexo/gênero

Mendes (2007, 2011a) desenvolveu um experimento de percepção e avaliação linguística em que perguntou, a mais de 100 paulistanos (homens e mulheres de diversas idades, diferentes orientações sexuais, graus de escolaridade e classes sociais): “[...] o que chama a sua atenção, na fala de um homem, no sentido de fazer você pensar que ele possa ser ‘gay’?” (Observe-se que a mesma pergunta poderia ser feita no sentido de obter pistas acerca do que define uma fala avaliada/percebida como paulistana, como interiorana etc).

Preparado para a dificuldade em identificar marcas ou fatos linguísticos que seus interlocutores pudessem perceber como índices de uma orientação gay, o pesquisador havia previamente gravado cinco leituras de um texto, por cinco homens diferentes. Após alguns minutos de conversa que sucederam aquela pergunta inicial, reproduziu-se a gravação das cinco leituras do tal texto para cada um dos entrevistados e pediu-se, então, a cada um deles, que atribuissem uma “nota” de 1 a 5 para cada uma delas: a nota 1 deveria ser atribuída à leitura que julgassem soar “menos gay”, enquanto que a nota 5 deveria ser atribuída àquela que julgassem soar “mais gay”. A cada leitura deveria ser atribuída uma nota diferente.

Houve grande convergência a respeito de quais leituras soaram “mais gay”. Ilustra-se, desse modo, que a percepção que os sujeitos de uma comunidade de fala têm acerca da produção linguística pode ser bastante homogênea (CAMPBELL-KIBLER, 2008, 2009), apesar da heterogeneidade da produção. Além disso, após ouvir e avaliar as leituras, os entrevistados puderam justificar-se: diferentemente daquilo que se observou ao início das conversas com esses mais de cem paulistanos, vários deles, após a avaliação das leituras, mostraram-se capazes de levantar hipóteses sobre quais fatos linguísticos haviam chamado sua atenção, no sentido de quais notas atribuiriam a quais leituras.

As apreciações de natureza linguística mais mencionadas pelos entrevistados foram: (i) “certas vogais parecem mais longas”; (ii) “a entonação ‘sobe e desce’ muito”; (iii) “as palavras são mais cuidadosamente pronunciadas, gays falam mais ‘direitinho’, cometem menos ‘erros’”; (iv) “homens gays e mulheres parecem usar diminutivos de maneira exagerada”. Apesar da metalinguagem – essas observações linguísticas estão aqui transcritas do modo como mais frequentemente foram enunciadas –, é inegável que essas descrições são bastante sofisticadas, especialmente considerando-se que, em princípio, os entrevistados não se haviam mostrado capazes de apontar fatos linguísticos que os levassem à interpretação de que o sujeito-falante pudesse ser gay.

Todos esses supostos índices – depreendidos experimentalmente – são convites para análises sociolinguísticas. As duas últimas parecem especialmente instigantes: em nenhuma das leituras (todas de um mesmo texto) havia “erros” que

permissem a avaliação de que uma performance teria sido “mais certinha” que outra; em nenhuma das leituras ocorreu emprego de diminutivos. Isso significa que as justificativas em (iii) e (iv) acima, para as notas atribuídas pelos entrevistados às cinco leituras, não foram formuladas de fato com base na audição delas, mas sim com base em percepções e avaliações de natureza sociolinguística que os entrevistados trouxeram das suas vidas. Verificar se essas avaliações encontram correlato nas produções linguísticas é o desafio a se assumir – um desafio, sobretudo, porque pressupõe, idealmente, o trabalho com categorias sociais em constante construção (em vez de categorias pré-estabelecidas – homem, mulher; pobre, rico; gay, heterossexual; paulistano, interiorano etc.).

Numa tentativa de propiciar estudos sobre a correlação entre performance linguística e categorias de sexo/gênero, levando-se em conta o axioma do estilo, Soriano (2012) está trabalhando na coleta de dados da fala de informantes que se voluntariam a gravar suas próprias falas, durante longos períodos (que ultrapassam os limites das entrevistas sociolinguísticas conforme comumente conhecidas), em diferentes momentos e situações, com diferentes interlocutores. Basicamente, a pesquisadora oferece um gravador de fácil portabilidade a informantes selecionados, que o deixam indefinidamente ligado, ao longo de um dia. É evidente que uma coleta de dados feita dessa forma pressupõe um árduo trabalho de edição daquilo que é gravado, antes de ser descrito: partes cuja qualidade acústica é baixa precisam ser descartadas e, mais importante que isso, um termo de consentimento entre a pesquisadora e seus sujeitos-informantes prevê que não se tornem públicas informações pessoais que evidenciem sua identidade. Apesar das dificuldades, contudo, trata-se de um trabalho que poderá permitir estudos inovadores na área de identidades sociais, na sua relação com usos linguísticos.

Classe social

Atenta ao alargamento da classe média brasileira e à reconfiguração dos grupos sociais que compõem cada estrato socioeconômico (SOUZA; LAMOUNIER, 2010), Faria (2012) propõe-se, em sua iniciação científica, a discutir e estabelecer critérios que possibilitem incluir a variável classe social em futuros estudos sociolinguísticos na cidade de São Paulo.

Partindo da consideração de que os estratos socioeconômicos definidos com base numa comunidade nem sempre são aplicáveis a outras (ASH, 2004), Faria (2012) lança-se ao trabalho de verificar as percepções que os habitantes da capital paulista têm a respeito da estratificação socioeconômica dentro da cidade em que moram. Para tanto, vem-se realizando a coleta de entrevistas voltadas especificamente a investigar quais são as camadas sociais e de que modo os paulistanos estão distribuídos nelas. Além disso, seu projeto prevê a aplicação

de um questionário a outro grupo de paulistanos, cujas perguntas deverão incidir sobre os critérios que tiverem aparecido na pesquisa sobre avaliação social (possivelmente nível de escolaridade, ocupação profissional, renda, ocupação e escolaridade dos pais/cônjuge, quantidade de dependentes, posse da habitação e bairro). A partir dessa discussão, deverá ser possível formular critérios, ancorados nas percepções de paulistanos sobre as diferenciações socioeconômicas que compõem a sua comunidade, que poderão ser aplicados em novos estudos sociolinguísticos sobre a cidade de São Paulo.

Considerações finais e trabalhos futuros

Este artigo traçou um panorama dos trabalhos que vêm sendo desenvolvidos pelo GESOL-USP no sentido de consolidar a cidade de São Paulo no mapa de estudos sociolinguísticos brasileiros. A construção de uma nova amostra – as 60 entrevistas do Projeto SP2010 financiado pela FAPESP – está no centro das atividades do grupo, que paralelamente vem desenvolvendo trabalhos de análise de variáveis linguísticas (a partir da Amostra SP-Piloto) e vem delineando caminhos para estudos futuros.

Entre esses caminhos, já começam a ser trilhados aqueles que preveem a inclusão da participação de imigrantes na constituição sociolinguística da cidade. Há projetos em formulação que intencionam focalizar os descendentes de imigrantes árabes, italianos e japoneses, cuja descrição sociolinguística deverá possibilitar comparações interessantes com pesquisas etnolinguísticas desenvolvidas especificamente em outras cidades, como é o caso de Toronto (HOFFMAN; WALKER, 2010). Além disso, uma vez finalizado o Projeto SP2010, pretende-se, como próximo passo, propor a ampliação da amostra, com a inclusão da fala de migrantes, sobretudo nordestinos, cuja presença é tão marcante na cidade, de modo a fornecer uma descrição mais abrangente dessa complexa comunidade.

Agradecimentos

A todos os membros do GESOL-USP que trabalharam na Amostra SP-Piloto, sem os quais nenhum dos trabalhos aqui reportados teriam se concretizado: Alexandre Sobreiro, Andressa Serena, Camila Barbosa Faria, Dayane Celestino de Almeida, Larissa Soriano, Ivanete Nascimento, Mariane Esteves Bieler, Melissa Cruz e Rafael Stoppa Rocha.

O Grupo todo agradece à FAPESP, com cujos auspícios a construção da Amostra SP2010 se tornou possível.

■ *ABSTRACT:* This paper reports on the activities developed by the Study and Research Group on Sociolinguistics at the University of São Paulo. Aiming at positioning the city of São Paulo and its inhabitants, the Paulistanos (born and raised in this city), in the map of Brazilian Sociolinguistics, the group's works revolve mainly around the ongoing construction of a representative sample of Paulistano speech, which main goal is to answer the questions: "what does it mean to sound like a Paulistano?" and "which variables and variants characterize the Paulistano speech?". This new sample (SP2010) is to be made available (audio and transcriptions) on the Internet to the linguists' community. In addition to discussing the criteria that define such sample, this paper also addresses the challenges involved in its constitution. Moreover, it offers an overview of the researches that have been developed by utilizing the data (SP-Pilot) that was gathered in the preparation and training of the young sociolinguists that would work in the group. Finally, this paper also outlines paths for future work.

■ *KEYWORDS:* Language variation. Paulistanos. Identity. Corpus.

REFERÊNCIAS

- ASH, S. Social class. In: CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (Ed.). *The handbook of language variation and change*. Oxford: Blackwell, 2004. p.402-422.
- BELL, A. Styling the other to define the self. *Journal of Sociolinguistics*, Hoboken, v.4, p.523-541, 1999.
- _____. Language style as audience design. *Language in Society*, Cambridge, v.13, p.145-204, 1984.
- CALLOU, D. M. I.; MORAES, J. A. A norma de pronúncia do /s/ e /r/ pós-vocálico: distribuição por áreas regionais. In: CARDOSO, S. (Org.). *Diversidade linguística e ensino*. Salvador: Ed. Da UFBA, 1996. p.133-147.
- CAMPBELL-KIBLER, K. The nature of sociolinguistic perception. *Language Variation and Change*, Cambridge, v.21, p.135–156, 2009.
- _____. I'll be the judge of that: diversity in social perceptions of (ING). *Language in Society*, Cambridge, v.37, n.5, p.637–659, 2008.
- CASTILHO, A. T. . Português Brasileiro: descrição, história, teorização.. Lingüística (Madrid), v. 24, p. 77-100, 2010
- _____. *Projeto para a história do português paulista*. 2007. Projeto temático de equipe 06/5594-0 apresentado a FAPESP, Universidade de São Paulo, São Paulo, 2007.
- CASTILHO, A. T. de; PRETI, D. (Ed.). *A linguagem falada culta na cidade de São Paulo*: materiais para seu estudo. São Paulo: T.A. Queiroz, 1987. (Diálogos entre dois informantes, v.2).

_____. *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo*. São Paulo: T.A. Queiroz, 1986. (Elocuções formais, v.1).

CHAMBERS, J. K. Studying language variation: an informal epistemology. In: CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (Ed.). *The handbook of language variation and change*. Oxford: Blackwell, 2004. p.03-14.

_____. *Sociolinguistic theory*. Oxford: Blackwell, 1995.

CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (Ed.). *The handbook of language variation and change*. Oxford: Blackwell, 2004.

CHESHIRE, J. Syntactic variation and beyond: gender and social class variation in the use of discourse-new markers. *Journal of Sociolinguistics*, Hoboken, v.9, n.4, p.479-508, 2005.

CIANCIO, R. *Estudo sociolinguístico da fala paulistana por falantes do Itaim Bibi*. 2012. Projeto de Iniciação Científica apresentado ao CNPq, Universidade de São Paulo, São Paulo, 2012.

COELHO, R. F. *É nós na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana: o pronome de primeira pessoa do plural e a marcação de plural no verbo*. 2006. 178f. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006.

DURANTI, A. Linguistic anthropology: history, ideas, issues. In: DURANTI, A. (Ed.). *Linguistic anthropology: a reader*. Oxford: Blackwell, 2001. p.01-60.

ECKERT, P. Style and social meaning. In: ECKERT, P.; RICKFORD, J. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001. p. 119-126.

_____. (ay) goes to the city – Exploring the expressive use of variation. In: GUY, G. R. et al. *Towards a social science of language: papers in honor of William Labov*. Amsterdam: John Benjamins, 1996. p. 47-68. (Variation and change in Language and Society, v.1).

ECKERT, P.; MCCONNEL-GINET, S. Think practically and look locally. *Annual Review of Anthropology*, Palo Alto, v.21, n.21, p.461-90, 1992.

FARIA, C. B. de. *Para a inclusão de “classe social” nos estudos sociolinguísticos em São Paulo*. 2012. Projeto de Iniciação Científica, Universidade de São Paulo, São Paulo, 2012.

GILES, H.; POWESLAND, P.F. Accommodation theory. In: COUPLAND, N.; JAWORSKI, A. (Ed.). *A sociolinguistics reader*. Macmillan: Basingstoke, England, 1997. p.232-239.

GONÇALVES, S. C. L. *O português falado na região de São José do Rio Preto: constituição de um banco de dados anotado para o seu estudo*. 2003. Projeto de

Pesquisa apresentado à FAPESP, Universidade Estadual Paulista, São José do Rio Preto, 2003. Disponível em: <www.iboruna.ibilce.unesp.br/histórico/Projeto>. Acesso em: 15 out. 2012.

GUY, G. R. The sociolinguistic types of language change. *Diachronica*, Amsterdã, v.7, n.1, p.47–67, 1990.

HOFFMAN, M.; WALKER, J. Ethnolects and the city: ethnic orientation and linguistic variation in toronto english. *Language Variation and Change*, Cambridge, v.22, p.37-67, 2010.

HOLMES, J.; MEYERHOFF, M. (Ed.). *The handbook of language and gender*. Oxford: Blackwell, 2005.

HORA, D. da (Org.). *Estudos sociolinguísticos*: perfil de uma comunidade. Santa Maria: Palotti, 2004.

HORNIK, K. R. *FAQ*. Disponível em: <<http://cran.r-project.org/doc/FAQ/R-FAQ.html>>. Acesso em: 10 abr. 2011.

HYMES, D. On communicative competence. In: BRUMFIT, C. J.; JOHNSON, K. (Ed.). *The communicative approach to language teaching*. Oxford: Oxford University Press, 1991. p.5-26.

IBGE. *Censo demográfico 2010*: resultados gerais da amostra. Brasília, 2010. Disponível em <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/resultados_gerais_amostra.pdf>. Acesso em: 15 out. 2012.

ITAIM BIBI. *Encontra Itaim*. São Paulo, [2011]. Disponível em: <<http://www.encontraitaim.com.br/>>. Acesso em: 20 jan. 2011.

LABOV, W. *The social stratification of english in New York city*. 2.ed. New York: Cambridge University Press, 2006.

_____. *Principles of linguistic change*: social factors. Oxford: Blackwell, 2001a.

_____. The anatomy of style-shifting. In: ECKERT, P.; RICKFORD, J. R. (Ed.). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001b.

_____. *Principles of linguistic change*: internal factors. Oxford: Blackwell, 1994.

_____. Where does the sociolinguistic variable stop? a response to Beatriz Lavandera. *Working papers in sociolinguistics*, Austin, v.44, p.01-22, 1978.

_____. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

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

MENDES, R. B. Gênero/sexo, Variação linguística e intolerância. In: BARROS, D. L. P. de. (Org.). *Preconceito e intolerância: reflexões linguístico-discursivas*. São Paulo: Mackenzie, 2011a. p.01-30.

_____. A pronúncia retroflexa do /-R/ na fala paulistana. In: HORA, D. da; NEGRÃO, E. V. (Ed.). *Estudos da linguagem: casamento entre temas e perspectivas*. João Pessoa: Ideia, 2011b. p.283-300.

_____. What is 'gay speech' in São Paulo, Brazil. In: SANTAEMILIA, J. et al. (Org.). *International perspectives on gender and language*. València: Universitat de València, 2007. p.296-305.

MENDES, R. B.; OUSHIRO, L. *Production and perception of retroflex /r/ in São Paulo Portuguese*. 2011. Trabalho apresentado ao Variation and Language Processing, United Kingdom, 2011a.

_____. Percepções sociolinguísticas sobre as variantes tepe e retroflexa na cidade de São Paulo. In: HORA, D.; NEGRÃO, E. V. (Org.). *Estudos da linguagem: casamento entre temas e perspectivas*. João Pessoa: Ideia, 2011b. p.262–281.

MILROY, L. Social networks. In: CHAMBERS, J. K.; TRUDGILL, P.; SCHILLING-ESTES, N. (Ed.). *The handbook of language variation and change*. Oxford: Blackwell, 2004. p.573-600.

_____. *Language and social networks*. Oxford: Basil Blackwell, 1980.

MONARETTO, V. N. de O. *Um estudo da vibrante: análise variacionista e fonológica*. 1997. 145f. Tese (Doutorado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1997.

NASCIMENTO, I. B. do. *O uso variável do pronome de segunda pessoa você(s)/cê(s) na cidade de São Paulo*. 2011. 217f. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011.

OCHS, E. Indexing gender. In: DURANTI, A.; GOODWIN, C. *Rethinking context*. Cambridge: Cambridge University Press, 1991. p.335-358.

OUCHIRO, L. *Uma análise variacionista para as interrogrativas-Q*. 2011. 160f. Dissertação (Mestrado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2011a.

_____. *Identidade na pluralidade: produção e percepção linguística na cidade de São Paulo*. Projeto de doutorado apresentado a FAPESP/Processo 2011/09122-6, Universidade de São Paulo, São Paulo, 2011b.

_____. Wh-interrogatives in Brazilian Portuguese: the influence of common ground. *University of Pennsylvania Working Papers in Linguistics*, Pennsylvania, v.17, n.2, p.145–154, 2011c.

- PRETI, D.; URBANO, H. (Ed.). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo*. São Paulo: T.A. Queiroz, 1988. (Entrevistas, v.3).
- PRETI, D.; URBANO, H. (Ed.) *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo*. São Paulo: T.A. Queiroz, 1990. (Estudos, v.4).
- ROCHA, R. S. Negação pós-verbal no português paulistano: restrições discursivo-pragmáticas. In: CONGRESSO INTERNACIONAL DE LINGUÍSTICA HISTÓRICA, 2., 2012, Resumos... São Paulo: Universidade de São Paulo, São Paulo, 2012a.
- _____. *Verbal negative structures in paulistano portuguese*. 2012. Trabalho apresentado no International GSCP Conference: Speech and Corpora, Belo Horizonte, 2012b.
- RODRIGUES, A. C. S. Fotografia sociolinguística do português do Brasil: o português popular em São Paulo. In: CASTILHO, A. T. (Ed.). *História do português paulista*. Campinas: Instituto de Estudos da Linguagem/UNICAMP, 2009. p.151-158.
- _____. *A concordância verbal no português popular em São Paulo*. 1987. 266f. Tese (Doutorado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1987.
- SÃO PAULO (Cidade). Prefeitura de São Paulo. *Município em mapas*: político administrativo. São Paulo, 2011. Disponível em: <<http://sempla.prefeitura.sp.gov.br/mm/panorama/>>. Acesso em: 20 mar. 2011.
- _____. *Histórico*. São Paulo, [2011]. Disponível em: <<http://www.prefeitura.sp.gov.br/cidade/secretarias/subprefeituras/pinheiros/historico/index.php?p=472>>. Acesso em: 20 jan. 2011.
- SILVA, F. G. *Concordância nominal*: um contraste dentro da cidade de São Paulo. Projeto de Mestrado apresentado ao Departamento Linguística, Universidade de São Paulo, São Paulo, 2011.
- SOUZA, A.; LAMOUNIER, B. *A classe média brasileira: ambições, valores e projetos de sociedade*. São Paulo: Elsevier, 2010.
- SORIANO, L. Estudo sociolinguístico de gays paulistanos em diferentes situações de fala. 2012. Projeto de Iniciação Científica, Universidade de São Paulo, São Paulo, 2012.
- STALNAKER, R. Common ground. *Linguistics and Philosophy*, Dordrecht, v.25, p.701-721, 2002.
- TAGLIAMONTE, S. A. *Analysing sociolinguistic variation*. São Paulo: Cambridge University Press, 2006.

TRUDGILL, P. *The social stratification of English in Norwich*. Cambridge: Cambridge University Press, 1974.

VIEIRA, M. *Entre São Paulo (SP) e Campo Grande (MS): fatos discursivo-pragmáticos no uso de variantes morfossintáticas*. 2012. Projeto de Doutorado apresentado ao departamento Linguística, Universidade de São Paulo, São Paulo, 2012.

Recebido em maio de 2012

Aprovado em agosto de 2012

A FALA DO INTERIOR PAULISTA NO CENÁRIO DA SOCIOLINGUÍSTICA BRASILEIRA: PANORAMA DA CONCORDÂNCIA VERBAL E DA ALTERNÂNCIA PRONOMINAL

Cássio Florêncio RUBIO*

Sebastião Carlos Leite GONÇALVES**

- **RESUMO:** Como parte das pesquisas sociolinguísticas desenvolvidas no interior do projeto ALIP (Amostra Linguística do Interior), privilegiamos, neste artigo, a apresentação de resultados gerais decorrentes da investigação de três fenômenos variáveis na fala do interior paulista: (i) a concordância verbal de primeira pessoa do plural; (ii) a concordância verbal de terceira pessoa do plural; (iii) a alternância pronominal entre as formas de codificação da primeira pessoa do discurso do plural *nós* e *a gente*. Assumindo os preceitos teórico-metodológicos da Sociolinguística laboviana, mostramos, por meio de comparações interdialetais, como tais fenômenos variáveis se inserem no cenário mais amplo da pesquisa sociolinguística sobre o português brasileiro. Os resultados permitem constatar aproximações e distanciamentos entre a fala do interior paulista e de outras variedades do português brasileiro, o que põe à mostra a importância da descrição de mais essa variedade, ainda pouco conhecida no cenário sociolinguístico brasileiro.
- **PALAVRAS-CHAVE:** Português paulista. Concordância verbal. Alternância pronominal.

Contextualizando o Projeto ALIP (Amostra Linguística do Interior Paulista)

Considerando os avanços da pesquisa sociolinguística no Brasil nas últimas décadas, pode-se considerar que os estudos variacionistas sobre as variedades paulistas ainda se encontram na sua infância e, portanto, em dúvida com a composição de um retrato sociolinguístico do Português Brasileiro (PB), no que toca ao oferecimento de um espectro mais amplo da fala paulista. Em outras palavras, comparados a trabalhos variacionistas envolvendo outras variedades do PB (citem-se aqui as variedades cariocas, catarinenses, paranaenses, gaúchas, mineiras, pessoenses, dentre outras), podem-se considerar escassos trabalhos nessa mesma linha envolvendo a fala paulista, seja a da capital, seja

* UNILAB – Universidade da Integração Internacional da Lusofonia Afro-Brasileira. Redenção – Ceará – Brasil. 62790-000 - cassiorubio@unilab.edu.br

** UNESP – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. São José do Rio Preto – São Paulo – Brasil. 15054-000 - scarlos@ibilce.unesp.br

a do interior. Seria injusto, entretanto, deixar de lembrar aqui que as variedades paulistas já foram objetos de variados projetos de descrição do PB, mas não sob enfoque sociolinguístico. Citem-se, nessa direção, os estudos descritivos desenvolvidos no âmbito do “Projeto Gramática do Português Falado Culto” (CASTILHO, 1990) e os que vêm se desenvolvendo no interior do projeto “Para a história do português paulista” ou “Projeto Caipira”, os quais conferem lugar de destaque à investigação das variedades paulistas do PB, assumindo como objetivo principal “[...] investigar emparelhadamente a formação da sociedade paulista e de suas variedades linguísticas, tais como testemunhadas no Estado de São Paulo e em sua capital.” (CASTILHO, 2009, p.15).

Em balanço crítico da pesquisa variacionista produzida no estado de São Paulo empregando-se amostras de fala de variedades paulistas, Gonçalves (2012) mostra que só muito recentemente a Sociolinguística vem se firmando no estado, em torno de três projetos de maiores dimensões: o projeto “Amostra Linguística do Interior Paulista” (GONÇALVES, 2007), o projeto “É nós na fita: sobre a formação de registros e a elaboração de estilos no campo da cultura popular urbana paulista.” (BENTES, 2009) e o projeto “SP2010 - Construção de uma Amostra da Fala Paulistana” (MENDES, 2011), sediados, respectivamente, na Universidade Estadual Paulista (UNESP/São José do Rio Preto), na Universidade Estadual de Campinas (UNICAMP) e na Universidade de São Paulo (USP), todos subsidiados pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). O autor registra ainda o pioneirismo dos trabalhos etnolinguísticos sobre a variedade paulista advindo das contribuições de Amadeu Amaral (1976), em *O dialeto caipira*, de Ada Rodrigues (1974), sobre o dialeto caipira de Piracicaba, de Mary Careno (1997), sobre a fala de comunidades negras do Vale do Ribeira, dos trabalhos variacionistas de Duarte (1986), sobre a realização de objeto direto anafórico na fala de paulistanos nativos e na linguagem da televisão, e de Ângela Rodrigues (1987), sobre concordância verbal na fala de favelados da capital paulista.

Cientes dessa carência de estudos sociolinguísticos sobre variedades paulistas, sobretudo do interior do estado, instituímos, no ano de 2003, o projeto ALIP, motivados pelos interesses de pesquisas que tinham como principal diretriz o enfoque da língua inserida em seu contexto social. O projeto teve por objetivo primeiro a composição de um banco de dados com amostras do português falado na região de São José de Rio Preto, recurso fundamental para a consolidação de grupos de pesquisa e para a proposição de projetos coletivos de pesquisa.

Embora inspirados em outros projetos já consolidados (VARSUL, PEUL, VALPB, Gramática & Discurso), o projeto ALIP não assumiu como prerrogativa para sua proposição interesse por qualquer fenômeno linguístico em particular, mas, sob uma concepção mais ampla e mais aberta, manteve a preocupação de captar o máximo possível do dinamismo linguístico da região noroeste do estado de São Paulo, representada por São José do Rio Preto e seis outras cidades que lhe fazem fronteira: Bady Bassit, Cedral, Guapiaçu, Ipiruá, Mirassol e Onda Verde. Assim, entre os anos de 2004 e 2007, constituímos o *banco de dados Iboruna* (= rio preto, em Tupi), composto por dois tipos de amostras de fala: a primeira, coletada de acordo com os critérios da Sociolinguística laboviana (LABOV, 1972), constitui a *Amostra Censo* ou *Amostra Comunidade* (AC, daqui em diante), e a segunda, coletada secretamente em contextos interacionais livres, sem controle de qualquer variável social (RONCARATI, 1996), constitui a *Amostra de Interação* (AI, daqui em diante).¹

Na composição da AC, controlamos as variáveis sociais comprovadamente relevantes nos estudos sociolinguísticos, a saber: sexo/gênero; faixa etária (7 a 15 anos, 16 a 25 anos, 26 a 35 anos, 36 a 55 anos, mais de 55 anos), escolaridade (1º ciclo do ensino fundamental; 2º ciclo do ensino fundamental, ensino médio e ensino superior) e renda familiar (até 5 salários mínimos, 6 a 10 salários mínimos, 11 a 25 salários mínimos e mais de 25 salários mínimos).² Do cruzamento das variantes dessas variáveis resultou um total de 152 perfis sociais, dos quais apenas um ainda não foi encontrado na comunidade. A inclusão de *área geográfica* como mais uma variável de estratificação faria crescer consideravelmente o número de informantes. Assim, para a execução do censo linguístico, os 152 perfis sociais foram distribuídos proporcionalmente por entre as sete cidades da região, considerando a densidade populacional mostrada no quadro 1 a seguir.

¹ O nome *IBORUNA* (= Rio Preto) tem motivação histórica; é um topônimo de origem tupi-guarani que se pretendeu atribuir à cidade de São José do Rio Preto por ocasião da comemoração do seu cinquentenário. A contundente intervenção do episcopado riopretense não só impedi a mudança como conquistou de maneira definitiva a denominação primitiva, São José do Rio Preto, reduzida a Rio Preto de 1906 a 1944.

² Cabe aqui uma rápida justificativa acerca da proposição de algumas variantes: no recorte contínuo de **faixas etárias**, a primeira representa a fase em que os padrões linguísticos estão ainda em fixação; da segunda em diante, a motivação é buscada na pressão social sobre a linguagem do indivíduo, variável de acordo com a inserção no mercado de trabalho (SILVA, 1996); a segmentação das variantes de **escolaridade**, embora pedagogicamente extinta nos dois primeiros ciclos, preserva diferenças salientes em termos de currículo e metodologia de ensino e representa a divisão escolar em vigor no tempo em que a maioria dos informantes se enquadrava; sobre a variável **renda familiar**, reconhecendo a dificuldade de divisão de classe social por indicadores diversos, optamos apenas por esse indicador.

Quadro 1 - Distribuição dos perfis sociais proporcionalmente à densidade populacional das cidades da região.

Cidades da Região de São José do Rio Preto	População	Número de informantes
1. Bady Bassit (12 km ao sul de SJRP)	11.475	04
2. Cedral (14 km, ao sul de SJRP)	6.690	02
3. Guapiáçu (16 km, ao leste de SJRP)	14.049	05
4. Ipuiguá (18 km, ao norte de SJRP)	3.461	01
5. Mirassol (14 km, a oeste de SJRP)	48.233	16
6. Onda Verde (25 km, ao norte de SJRP)	5.407	02
7. São José do Rio Preto ³	357.705	122
Total da população representada	447.020	152

Fonte: (IBGE, 2000).3

A definição do perfil social desses informantes por cidade seguiu o método da distribuição aleatória (SILVA; SCHERRE, 1996) por área geográfica, que consiste nos seguintes procedimentos: (i) em uma primeira urna, colocamos todos os perfis sociais, identificados pelos números de 1 a 152; (ii) em uma segunda urna, depositamos os nomes de todas as cidades da região; (iii) por meio de escolha ao acaso, retirávamos de cada uma das urnas um perfil social e um nome de cidade, definindo-se assim a origem geográfica do perfil social; (iv) em seguida, devolvíamos apenas para segunda urna o nome da cidade escolhida, até que se atingisse o seu número de informantes; (v) nova escolha combinada era feita, até ser definida a origem geográfica de todos os informantes. Esse procedimento permitiu-nos assegurar a probabilidade de quaisquer dos perfis sociais pertencerem a uma dada cidade.

Definidos os perfis sociais, a constituição de AC, incluindo a localização dos perfis na comunidade, a gravação e a validação da entrevista e sua posterior transcrição ficou sob a responsabilidade de uma equipe técnica composta, na primeira fase do projeto, por 30 integrantes.⁴ Segundo roteiro previamente elaborado, as entrevistas de AC foram direcionadas para obtenção de cinco tipos de textos de cada informante, com base na metodologia exposta em Votre e Oliveira (1995), a saber: **narrativa de experiência pessoal, narrativa recontada, relato de descrição, relato de opinião e relato de procedimento.**

Para a constituição da AI, pautada por orientações metodológicas e éticas (RONCARATI, 1996), integrantes da mesma equipe coletaram 11 amostras de interação social livre, cujas descrições são dadas no quadro 2.

³ Inclui os Distritos de Talhados e Engenheiro Schmidt

⁴ Compuseram a equipe um pesquisador coordenador geral, seis pesquisadores colaboradores e 23 auxiliares técnicos.

Quadro 2 - Informações sobre as interações dialógicas de AI.

AMOSTRA	CONTEXTO DE INTERAÇÃO – INFORMANTES ENVOLVIDOS
AI-001-CAS	Conversa entre dois homens e cinco mulheres de diferentes idades e escolaridades, em ambiente familiar.
AI-002-GIL	Diálogo entre duas amigas vizinhas, com perfis sociais semelhantes, no portão da casa de uma delas.
AI-003-ILHA	Diálogo entre tia e sobrinha, de diferentes idades e escolaridades, em ambiente familiar.
AI-004-OND	Diálogo entre duas irmãs, de mesma faixa de idade e diferentes escolaridades, em ambiente familiar.
AI-005-CAS	Diálogo entre duas estudantes de mesmo perfil social, em ambiente universitário.
AI-006-MAR	Conversa entre quatro mulheres de perfis sociais diferentes, em ambiente familiar.
AI-007-FER	Diálogo entre marido e esposa de mesma faixa etária e de níveis de escolaridade diferentes, em ambiente familiar.
AI-008-CAM	Conversa entre três estudantes, do sexo masculino e mesmos perfis sociais, em ambiente universitário.
AI-009-CAS	Diálogo entre um advogado e sua cliente, de perfis sociais diferentes, em escritório de advocacia.
AI-010-CAS	Discussão entre dois advogados, de mesmo perfil social, sobre uma peça jurídica, em escritório de advocacia.
AI-011-CAS	Diálogo entre casal de namorados, estudantes universitários de mesma faixa etária, em ambiente familiar.

Fonte: GONÇALVES (2005).

Na fase final do projeto, a preparação definitiva de AC e de AI ficou sob responsabilidade de quatro integrantes da equipe técnica, aos quais coube a tarefa de revisão minuciosa da documentação linguística de cada uma das gravações armazenadas no banco de dados: as transcrições e os respectivos áudios, as fichas sociais, as fichas de validação e os diários de campo. Além desse material, encontram-se disponíveis também no site do projeto⁵ os relatórios de cada fase do projeto, o manual de transcrição e o roteiro de entrevistas.⁶

Feita essa contextualização do projeto ALIP, a continuidade deste artigo estrutura-se em oito seções: na primeira, apresentamos brevemente pesquisas concluídas que se valeram das amostras do banco de dados Iboruna, para, nas seções seguintes, anteriores à conclusão e centrais deste artigo, ofereceremos um panorama da concordância verbal (CV, daqui em diante) e da alternância pronominal (AP, daqui em diante) na fala do interior paulista, estabelecendo, a partir dos resultados alcançados para os fenômenos variáveis aí envolvidos, comparações interdialetais.

⁵ Disponível em: <www.iboruna.ibilce.unesp.br>.

⁶ Para maiores informações sobre questões teóricas e metodológicas do projeto ALIP, remetemos o leitor a Gonçalves (2008).

Projeto ALIP: esboço de um retrato sociolinguístico da região noroeste do estado de SP

Próximo de completar sua primeira década de existência, o projeto ALIP e o banco de dados Iboruna já propiciaram importantes trabalhos descritivos da fala do interior paulista, tanto da perspectiva sociolinguística quanto de uma perspectiva funcionalista, não variacionista, de modo que um retrato sociolinguístico da região já começa a se esboçar.

Da perspectiva variacionista, os fenômenos já investigados contemplam os níveis fonético-fonológico e morfossintático. Integrados ao “Projeto de descrição sócio-histórica das vogais do português brasileiro” (PROBRAVO), que tem por objetivo descrever as realizações fonéticas de vogais nos dialetos de Sul a Norte do Brasil, trabalhos variacionistas envolvendo o dialeto paulista vêm se consolidando no estado de São Paulo, sob a coordenação da Profa. L. E. Tenani, da Unesp de São José do Rio Preto, a exemplo da descrição de outros dialetos que têm lugar no interior do projeto nacional, a saber: dialetos mineiros, dialetos do Sul, dialeto da Paraíba, dialeto mato-grossense, dialeto do Pará, dialeto de Rondônia, dialeto capixaba, dialeto baiano e dialeto carioca.⁷ Mais especificamente, com base nas amostras de fala do banco de dados Iboruna, as contribuições com PROBRAVO incluem resultados para os seguintes fenômenos variáveis: (i) alcamento vocálico em contextos de postônica medial de nomes, como em *c[o]zinha* ~ *c[u]zinha* e *t[e]soura* ~ *t[i]soura* (SILVEIRA, 2008), e de verbos, como em *d[e]via* ~ *d[i]via* e *p[o]dia* ~ *p[u]dia* (CARMO, 2009); (ii) alcamento e síncope de postônicas mediais, como em *pê.s.e.go* ~ *pê.s.i.go* ~ *pê.s.go* e *a.bó.[b]ra* ~ *a.bó.[bu].ra* ~ *a.bó.[bra]* (RAMOS, 2009).

Não diretamente ligados ao PROBRAVO, outros fenômenos fonéticos também já dispõem de resultados interessantes. Um exemplo é o trabalho de Ferreira (2010) sobre a redução de gerúndio, como em *canta[ndo]* ~ *canta[no]*, traço marcante da fala paulista interiorana, que alcança percentuais elevadíssimos de aplicação da alternante reduzida (72%), frente a outras variedades.

Resultados para fenômenos variáveis de ordem morfossintática incluem a expressão de futuro sintético x futuro analítico (FONSECA, 2010); a expressão de aspecto cursivo por meio de perifrases verbais, como em *andar* x *continuar* x *ficar* x *viver* + gerúndio (FERNANDES, 2010); a realização de preposições com *x* sem contração, como em *com a* ~ *cu'a* ~ *c'a*, *para* ~ *pra* ~ *pa* etc. (MARCATO,

⁷ O PROBRAVO, composto por pesquisadores de 17 universidades brasileiras, foi criado em 2005 e, sediado na Universidade Federal de Minas Gerais, é atualmente coordenado pelos Profs. S. Lee e M. A. Oliveira. De modo articulado, vem procurando responder às seguintes questões: (i) como são realizadas foneticamente as vogais no PB? (ii) como se explica ou o que motiva a diversidade de realizações fonéticas? (iii) como os falantes do PB se entendem, apesar das diversidades da qualidade vocalica? (iv) é possível explicar essa diversidade gramaticalmente? (Informações disponíveis em: <<http://relin.letras.ufmg.br/probravo>>. Acesso em: 22 fev. 2012).

2011); a expressão de cópula em orações matrizes predicativas, como em *é claro que ~ claro que* (FORTILLI, 2011); a alternância indicativo/subjuntivo em orações complexas (SANTOS, 2005); a marcação de plural em sintagmas nominais (FIAMENGUI, 2011) e em contextos de predicativo (SALOMÃO, 2010); a concordância verbal de 3PP (RUBIO, 2008) e padrões de concordância verbal e de alternância pronominal (RUBIO, 2012).

O banco de dados Iboruna tem servido ainda ao desenvolvimento de inúmeras pesquisas de descrição do português falado sob a perspectiva funcionalista, não variacionista. Dentre as várias pesquisas já concluídas, citemos: o uso de marcadores discursivos (GUERRA, 2007; PENHADEL, 2010), a gramaticalização de juntivos (FELÍCIO, 2008; GALBIATI, 2010; SOUZA, 2009; RENCK, 2011), a gramaticalização e a combinação de orações (GONÇALVES, 2011; SANTANA, 2010), a expressão de evidencialidade (VENDRAME, 2010) e o uso de predicações não verbais (FORTILLI, 2007).

Rotulado, de modo mais amplo, de sociofuncionalista, outro conjunto de trabalhos vem sendo desenvolvido, combinando-se as perspectivas da variação e da gramaticalização. Incluem-se nessa vertente a investigação de preposições com verbos de movimento (WIEDEMER, 2011), a alternância *nós x a gente*, o alcamento de constituintes (GONÇALVES, 2012) e as predicações reduzidas encaixadas em predicados avaliativos (LIMA, 2012).

Dentre os inúmeros fenômenos variáveis comprovadamente presentes no PB do interior paulista, nossa motivação para o enfoque, neste artigo, apenas de fenômenos relacionados à CV e à AP, deve-se ao fato de esses constituírem fenômenos amplamente difundidos em grande número de regiões do território brasileiro, o que se comprova pelos inúmeros estudos já realizados para as também inúmeras variedades do PB, o que permite a realização de uma análise sociolinguística comparativa e a comprovação de possíveis padrões de variação nessas variedades de língua portuguesa existentes no Brasil. É do que passamos a tratar nas seções seguintes.

Panorama da concordância verbal e da alternância pronominal na fala do interior paulista

Segundo Weinreich, Labov e Herzog (1968), a mudança linguística necessariamente deve ser concebida como encaixada no sistema linguístico e na matriz social, sem que isso implique, no entanto, concebê-la como um movimento de um sistema inteiro para outro completamente diferente. Em outras palavras, o que ocorre, num processo de mudança, é a alteração gradual de um conjunto limitado de variáveis num sistema. O controle dessa variação pode ser apreendido a partir da competência linguística dos membros da

comunidade de fala. Da mesma forma, no desenvolvimento da mudança linguística, a estrutura social pode pesar de forma diferente sobre o sistema linguístico abstrato. A questão do encaixamento, portanto, lidará com o entrelaçamento das mudanças com outras que poderão afetar tanto a estrutura linguística quanto a estrutura social.

Especificamente, no fenômeno variável da concordância verbal (CV, daqui em diante), a redução no paradigma da conjugação verbal acarreta alterações em outro subsistema da língua, como, por exemplo, na estrutura oracional, levando a um maior preenchimento da posição de sujeito, nos casos em que há o emprego de verbo no singular para sujeito plural, semelhantemente ao que ocorre em outras línguas. Contribui também para a redução do paradigma verbal a implementação de outras formas pronominais, como as formas *você/vocês*, para a segunda pessoa do singular/plural (2PS/2PP, daqui em diante), e *a gente*, para a primeira pessoa do plural (1PP, daqui em diante), que, com maior frequência, se vinculam às formas verbais de terceira pessoa.

A alternância pronominal (AP, daqui em diante) de 1PP talvez seja, dos fenômenos variáveis no PB, o que mais apresente sinais de encaixamento na matriz social, visto que se mostra presente em todos os segmentos sociais de inúmeras variedades, o que significa um grande passo no processo de sua implementação ou aceitação.

Com vistas ao oferecimento de um panorama dos fenômenos de CV e de AP, apresentamos, a seguir, um quadro resumo de realizações pronominais e respectivos padrões de CV verificados em diferentes variedades do PB, com destaque para a variedade falada no interior paulista, os quais podem ser contrastados com os padrões normativos da língua portuguesa.⁸

⁸ Para fenômenos variáveis envolvendo 1PP do discurso e 3PP, procedemos à análise quantitativa e qualitativa e, para a primeira pessoa do singular (1PS, daqui em diante), 2PS e 2PP do discurso e terceira pessoa do singular (3PS, daqui em diante), procedemos à análise apenas qualitativa.

Quadro 3 - Realizações pronominais e formas correlatas e padrões de conjugação verbal em variedades do PB.

Variedades Pessoa	PADRÃO		ESTUDOS DO PORTUGUÊS BRASILEIRO ⁸		PORTUGUÊS BRASILEIRO DO INTERIOR PAULISTA ⁹	
	Pronome e correlatos	Conjugação/ Exemplo	Pronome e Correlatos	Conjugação/ Exemplo	Pronome e Correlatos	Conjugação/ Exemplo
1PS ¹⁰	Eu	1PS / <i>canto</i>	Eu	1PS x (3PS) / <i>canto x (canta)</i>	Eu	1PS / <i>canto</i>
2PS	Tu	2PS / <i>cantas</i>	Tu	2PS x 3PS / <i>cantas x canta</i>	Você	3PS / <i>canta</i>
			Você	3PS / <i>canta</i>		
3PS ¹¹	Ele/a e correlatos	3PS, 3PP / <i>canta x (cantam)</i>	Ele/a e correlatos	3PS x (3PP) / <i>canta x (cantam)</i>	Ele/a e correlatos	3PS x (3PP) / <i>canta x (cantam)</i>
	Coletivo (povo)	3PS x 3PP / <i>canta x cantam</i>	Coletivo (povo)	3PS x 3PP / <i>canta x cantam</i>	Coletivo (povo)	3PS x 3PP / <i>canta x cantam</i>
1PP	Nós e correlatos	1PP / <i>cantamos</i>	Nós e correlatos	1PP x 3PS / <i>cantamos x canta</i>	Nós e correlatos	1PP x 3PS / <i>cantamos x canta</i>
			A gente	3PS x 1PP x (3PP) / <i>canta x cantamos x (cantam)</i>	A gente	3PS x 1PP / <i>canta x (cantamos)</i>
2PP	Vós	2PP / <i>cantais</i>	Vocês	3PP x 3PS / <i>cantam x canta</i>	Vocês	3PP x 3PS / <i>cantam x canta</i>
3PP	Ele/as e correlatos	3PP / <i>cantam</i>	Ele/as e correlatos	3PP x 3PS / <i>cantam x canta</i>	Ele/as e correlatos	3PP x 3PS / <i>cantam x canta</i>

Fonte: RUBIO (2012).^{9 10 11 12}

⁹ Dados relativos ao PB tomam por base os resultados dos trabalhos de Lucchesi, Baxter e Silva (2009) e Lopes e Naro (2011), para a 1PS; Loregian (1996), Paredes Silva (1996), Hausen (2000), Orlando (2004), Amaral (2003) e Modesto (2006), para a 2PS; Scherre e Naro (1998a), Mattos (2003) e Rubio (2012), para a 3PS; Bortoni-Ricardo (1985), Rodrigues (1987), Omena (1986, 2003), Lopes (1999), Naro, Görski e Fernandes (1999), Zilles, Maya e Silva (2000), Zilles (2005, 2007), Vianna (2006), Coelho (2006), Lucchesi Baxter e Silva (2009), para a 1PP; Bechara (2002), Neves (2000) e Perini (2010), para a 2PP; Lemle e Naro (1977), Nina (1980), Nicolau (1984), Rodrigues (1987), Graciosa (1991), Naro e Scherre (1999, 2000a, 2000b, 2003 e 2007), Scherre e Naro (1993, 1997, 1998a, 1998b, 1999, 2001, 2006), Monguilhott e Coelho (2002), Silva e Lucchesi (2006), Gameiro (2005), Monte (2007), para a 3PP.

¹⁰ Dados relativos ao PB do interior paulista tomam por base os trabalhos de Rubio (2008, 2012).

¹¹ As formas entre parênteses constituem fenômenos pouco recorrentes, mas já atestados.

¹² A tradição gramatical admite variação na CV somente nos casos específicos de sujeito partitivo (núcleo singular com elementos plurais adjacentes) ou de sentido coletivo.

Como se observa no quadro acima, no tocante à expressão pronominal, o PB, de forma geral, dispõe de oito formas, duas das quais com alternância: 2PS (*tu x você*) e 1PP (*nós x a gente*). Essa situação se reduz, se consideradas variedades específicas como a do interior paulista, que apresenta sete formas pronominais, com alternância em apenas uma delas (*nós x a gente*).

Quando se consideram os diferentes padrões de CV, à exceção da 2PS, expressa pelo pronome *você*, que apresenta padrão único de CV com 3PS (regra categórica, portanto), a regra de CV é variável para as formas pronominais, em maior ou menor grau, prevalecendo, nos diferentes padrões, para cada uma delas, a 3PS. Na variedade do interior paulista, também escapa a essa constatação a CV com 1PS.

A investigação dos fenômenos variáveis envolvendo tanto o quadro pronominal quanto os padrões de CV no PB demandaria, certamente, trabalho de maior fôlego, motivo pelo qual nos restringimos, neste artigo, ao enfoque dos fenômenos variáveis comuns à amostra do interior paulista (em destaque no quadro apresentado). Fundamentados na literatura pertinente, para a CV variável de 3PS e 2PP, apresentamos apenas ocorrências exemplificativas que comprovam tratar-se de fenômenos variáveis também na fala do interior paulista, e, para AP e CV de 1PP do discurso e CV de 3PP, apresentamos análises qualitativas e quantitativas, focalizando os resultados gerais e contrastando-os com os de outras variedades do PB.

Variação na concordância verbal de terceira pessoa do singular (3PS)

Considerando a 3PS, é possível verificar a ocorrência de variação entre o uso de formas verbais de 3PS e de 3PP nos chamados casos de concordância semântica, nos quais o núcleo do sujeito tem significado coletivo (**povo, multidão, pessoal, grupo, turma**, por exemplo).

As gramáticas normativas aludem ao fenômeno variável como uma “concordância ideológica.” Cunha e Cintra (1984, p.602) afirma que a *silepse*, figura de sintaxe de efeito expressivo, apresenta coesão significativa, ocasionada pelo contexto geral e pela situação. A concordância, nesse caso, faz-se com o sentido e não com a forma gramatical das palavras. No caso da *silepse de número*, um substantivo que se encontra no singular pode ser semanticamente considerado como plural, por ser coletivo (**o povo votaram...**).

Segundo Scherre e Naro (1993), a variação ocorre também nas estruturas complexas que se apresentam normalmente em construções cujo núcleo do sujeito é singular (seja ele de natureza quantitativa ou não) seguido de um *S_{prep}* de núcleo plural, que ocasiona uma leitura quantitativa, coletiva ou partitiva.

No PB do interior paulista, o fenômeno variável também é observado, conforme apresentamos a seguir, nas ocorrências exemplificativas recolhidas do banco de dados Iboruna.¹³

- (1.a) é quase como na saúde né? **Existe meia dúzia de privilegiados né?** que **GAnham** MUIto...[Doc.: hum:] e:: e também existe *a maioria* que **ganhá** muito po(u)co

[BDI-AC-138-305]

- (1.b) *uma parte dos impostos deviam*... ficá(r) pra assim... tê(r) um/ a gente ter o atendimento

[BDI-AC-138-277]

- (1.c) *a maioria das famílias* num **aceita** e eu acabo tendo PEna dessas pessoas porque eles são muito simples...

[BDI-AC-152-540]

Nas ocorrências, verifica-se o uso de formas verbais em 3PS e 3PP, junto de expressões complexas de núcleo singular, invariavelmente, na 3PS, que possuem um SPrep de núcleo plural ((1.a) a (1.c)). Embora, em termos frequenciais, essas ocorrências não sejam suficientes para investigação sociolinguística, elas apontam para existência de variação de CV de 3PS também na variedade brasileira do interior paulista.

Variação na concordância verbal com segunda pessoa do plural (2PP)

A forma pronominal *vocês* (e suas variantes, *ocês*, *cês*, *cêis*), no PB, é utilizada categoricamente para a representação da 2PP do discurso, em lugar da forma padrão, *vós*. Segundo Perini (2010), alguns itens tradicionalmente analisados como pronomes pessoais não ocorrem no PB, como a forma *vós*, que é somente usada na língua escrita em determinados contextos religiosos. No PB, a CV com 2PP do discurso é variável, com emprego de verbos em 3PS e em 3PP junto do pronome *vocês*, fenômeno também presente na variedade do interior paulista, como se observa nas ocorrências extraídas das amostras de interação do Iboruna.¹⁴

¹³ Ao final de cada ocorrência, apresentamos, entre parênteses, respectivamente: a origem das amostras, BDI (banco de dados Iboruna), o tipo de amostra, AC ou AI (amostra censo ou amostra de interação), o número de identificação do perfil social do informante e a linha da ocorrência.

¹⁴ Os tipos textuais que constituem a amostra censo restringem a ocorrência de formas de 2PP, encontradas apenas nas amostras de interação. Embora, nesse tipo de amostra, ocorram 2PS e 2PP, as ocorrências não são suficientes para estudo quantitativo. Nota-se também fenômeno de ordem fonético-fonológica no uso do pronome, com a alternância entre as formas *vocês* e *cê(i)s*.

(2.a) *cês assiste* também esse programa?... nós assiste TÔdo dia... bom eu quando tô em casa

[BDI-AI-002-30]

(2.b) –“mas *cês* num **tá** ensinan(d)o direi::to”– ... ela ainda falava –“eu quero aprende mais”–

[BDI-AI-004-65]

(2.c) a gente vê que *cês dão* muito valor no estudo... *cês* é.... *cês a::cha* importan.:te estudá.:r(r)

[BDI-AI-004-45]

(2.d) então mas vocês nunca/... *vocês* nunca **foram** no fórum pra resolvê(r) isso?

[BDI-AI-009-78]

(2.e) cê sa/ lembra onde tem um xerox?... ali pela rodoviária tal?... *cês* num **entraram** por ali?

[BDI-AI-008-220]

Não obstante a referência, nesses casos, seja à 2PP, o comportamento em relação à CV é semelhante ao verificado junto dos pronomes de 3PP, *eles* e *elas*. As observações sobre o fenômeno de 2PP aqui apresentadas, embora superficiais, apontam, ao menos, para a necessidade de uma investigação mais acurada desses contextos passíveis de variação.

Alternância pronominal no uso da primeira pessoa do plural (1PP) do discurso

Em relação à 1PP do discurso, observam-se, no PB, fenômenos variáveis de AP e de CV. Para o estudo da AP de 1PP no PB do interior paulista, foi analisado um total de 2.173 ocorrências das formas pronominais *nós* e *a gente* explícitas (plenas) e não explícitas (desinenciais ou nulas), com 1.603 casos da forma pronominal *a gente* (1.413 ocorrências do pronome explícito e 190 do pronome não explícito) e 570 da forma pronominal *nós* (477 casos de *nós* explícito e 93 casos de *nós* não explícito). Seguem, na tabela 1, os percentuais de distribuição das ocorrências.

**Tabela 1 - Alternância pronominal de primeira pessoa do plural
(nós x a gente) no português brasileiro do interior paulista.**

VARIEDADE	NÓS		A GENTE		TOTAL
PB – IBORUNA	26,2% (570)		73,8% (1.603)		
SUJEITO PRONOMINAL	explícito 83,7% (477)	não explícito 16,3% (93)	explícito 88,1% (1.413)	não explícito 11,9% (190)	100% (2.173)

Fonte: RUBIO (2012).

Para a fala do interior paulista, os resultados apontam que a forma inovadora *a gente* (como em (3.a)) predomina sobre a forma pronominal conservadora *nós* (como em (3.b)), com percentual de uso de 73,8%. Para ambas as formas de codificação do 1PP, sujeitos explícitos (identificados por (i) nas ocorrências (3.a) e (3.b)) prevalecem acentuadamente sobre os não explícitos (identificados por (ii) nas ocorrências).

- (3.a) lá em casa (i) **a gente...** num *tinha* dINHE(i)ro... (ii) num *tinha* RO(u)pa
[BDI-074-505]
- (3.b) eu trabalhava com trabalho de roça... lá (i) **nós** *mexíamos* com MUda
(ii) *preparávamos* mudas de café...
[BDI-114-60]

Apresentamos o quadro de variáveis consideradas na investigação da AP de 1PP e respectiva ordem de seleção apontada pelo programa estatístico *GOLDVARB*.

Quadro 4 - Ordem de seleção das variáveis consideradas para o fenômeno de alternância pronominal de primeira pessoa do plural no português brasileiro do interior paulista.

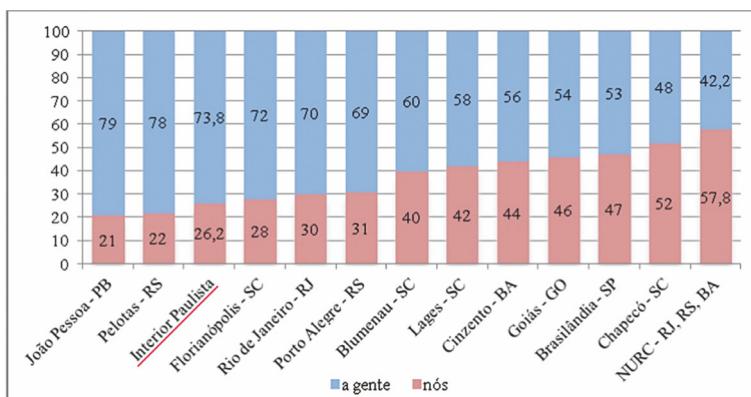
Variáveis	Fenômeno	Nós x A gente PB do interior paulista - IBORUNA
Linguísticas	Paralelismo linguístico discursivo	1º
	Saliência fônica verbal	2º
	Grau de determinação do sujeito	5º
	Tempo e modo verbal	6º
	Preenchimento do sujeito	não selecionada
Sociais	Escolaridade	4º
	Faixa etária	3º
	Gênero	não selecionada

Fonte: RUBIO (2012).

É possível observar, no quadro, que, dentre as variáveis linguísticas, **paralelismo linguístico discursivo** foi selecionado o fator mais relevante no fenômeno. **Saliência fônica verbal, grau de determinação do sujeito e tempo e modo verbal** foram selecionados em 2º, em 5º e em 6º lugares, respectivamente. O grupo de fatores **preenchimento do sujeito**, que controla os contextos de sujeito explícito e de sujeito não explícito, foi o único não selecionado pelo programa *GOLDVARB*, o que revela que a variável possui pouco peso no processo de AP na amostra considerada. Relativamente às variáveis sociais, sobressaem-se **faixa etária e escolaridade**, que foram selecionadas como relevantes em 3º e 4º lugares. **Gênero** foi considerado estatisticamente não relevante. A partir desses resultados, um primeiro julgamento é o de que, no PB do interior paulista, é forte a atuação de variáveis tanto linguísticas quanto sociais na AP de 1PP.

Efetuamos a comparação dos resultados gerais obtidos em nosso estudo com resultados evidenciados em outras variedades do Estado de São Paulo e em variedades de outros estados brasileiros, a fim de verificar as possíveis semelhanças e discrepâncias de percentuais da AP em diferentes variedades da língua portuguesa.¹⁵

Gráfico 1 - Alternância pronominal de primeira pessoa do plural *nós* e a *gente* em diferentes variedades do português brasileiro.



Fonte: Elaboração própria.¹⁶

¹⁵ Lembremos que não é nosso intuito contemplar todos os estudos já realizados sobre AP de 1PP, mas oferecer somente um panorama do fenômeno, na consideração de fatores de ordem diatópica e diastrática.

¹⁶ Características sociais das amostras: João Pessoa-PB (FERNANDES, 1999): escolaridade de nula a superior; faixa etária de 15 a 25, de 26 a 49 e de mais de 50 anos; e gêneros masculino e feminino; Pelotas-RS-Projeto VARX (BORGES, 2004): escolaridade fundamental, média, superior; faixa etária de 16 a 25, de 26 a 37, de 38 a 49, de 50 a 64 e de mais 65; e gêneros masculino e feminino; Interior Paulista-Iboruna (RUBIO, 2012): escolaridade fundamental, média, superior; faixa etária de 16 a 25, de 26 a 35, de 36 a 55 e de mais de 55; e gêneros feminino e masculino; Florianópolis-SC-Projeto VARSUL (SEARA, 2000): escolaridade primária, ginásial e secundária; faixa etária de 15 a 24, de 25 a 50 e de mais de 50; e gêneros feminino e masculino; Rio de Janeiro-RJ (OMENA; BRAGA, 1996): escolaridade fundamental e média; faixa etária de 7 a 14, de 15 a 25, de 26 a 49 e de 50 a 71

Sem perder de vista a diferença na estratificação social dos *corpora* em que cada investigação se baseia, os resultados do gráfico acima permitem evidenciar que as variedades do PB apresentam diferenças no uso das formas pronominais *nós* e *a gente*, com uma discrepância, se comparadas as variedades dos extremos do gráfico, de quase 37 pontos percentuais.

Em variedades de diferentes regiões e estados brasileiros, como a de João Pessoa (PB), a de Pelotas (RS), a do interior paulista, a de Florianópolis (SC), a do Rio de Janeiro (RJ) e a de Porto Alegre (RS), nota-se, pelos percentuais de emprego da forma *a gente* (79%, 78%, 69% e 73,8%, 72%, 70% e 69%, respectivamente), o predomínio acentuado da forma inovadora sobre a forma conservadora, o que, contudo, não se estende a todas as variedades do território brasileiro. É possível verificar certo equilíbrio entre as formas alternantes, em determinadas comunidades, como a de Brasilândia, na periferia de São Paulo, e a do estado de Goiás, que apresentaram, respectivamente, 53% e 54% de emprego da forma inovadora, a despeito de apresentarem estratificação social diferentes (confira nota 15).

O ligeiro predomínio do pronome *nós* foi atestado recentemente em Chapecó (SC), com 52% de emprego dessa forma, em oposição aos 48% de emprego da forma *a gente*, o que aproxima essa variedade popular à do PB culto falado nas cidades do Rio de Janeiro, Salvador e Porto Alegre, consideradas conjuntamente por Lopes (1999), que constatou uma frequência de 57,8% de uso de *nós* contra 42,2% de uso de *a gente*.

Essa constatação, aliada à observação de outros estudos do PB, confirma que a escolaridade não é fator que exerce forte influência no fenômeno variável da AP, já que, apesar de a maioria dos trabalhos considerar informantes de todas as faixas escolares (desde escolaridade nula até superior), constatam-se aproximações e distanciamentos entre as diferentes variedades do PB. Observam-se, por exemplo, aproximações entre a variedade de Cinzento-BA (44% de emprego de *nós*), com informantes de escolarização nula ou baixa; a de Lages-SC (42% de uso de *nós*), com informantes de escolaridade primária, ginásial e

anos; e gêneros masculino e feminino; Porto Alegre-RS (ZILLES; MAYA; SILVA, 2000): escolaridade de baixa até superior; faixa etária de 25 a 49 e de mais de 50 anos; e gêneros feminino e masculino; Blumenau-SC (TAMANINE, 2002): escolaridade primária, ginásial e secundária; faixa etária de até 45 anos e de mais de 50; e gêneros masculino e feminino; Lages-SC (TAMANINE, 2002): escolaridade primária, ginásial e secundária; faixa etária de até 45 anos e de mais de 50 anos; e gêneros masculino e feminino; Cinzento-BA (ANTONINO; BANDEIRA, 2011): escolaridade baixa ou nula; faixa etária de 20 a 40, de 41 a 60, de 61 a 80 e de mais de 80 anos; e gêneros feminino e masculino; Goiás-GO (MATTOS, 2010): escolaridade média e superior; faixa etária de menos de 21, de 21 a 40 e acima de 41; e gêneros masculino e feminino; Brasilândia-SP (COELHO, 2006): escolaridade de nula até 8 anos; faixa etária de menos de 25, de 25 a 50 e de mais de 50 anos; e gêneros masculino e feminino; Chapecó-SC (TAMANINE, 2002): escolaridade primária, ginásial e secundária; faixa etária de até 45 anos e de mais de 50 anos; e gêneros masculino e feminino; Rio de Janeiro, Porto Alegre e Salvador-Projeto NURC (LOPES, 1999): escolaridade superior; faixa etária de 25 a 35, de 36 a 55 e de mais de 56 anos; e gêneros masculino e feminino.

secundária; e a de Goiás-GO (46% de emprego de *nós*), com informantes de escolaridade média e superior. Por outro lado, há discrepância entre variedades geograficamente próximas, como a de Florianópolis-SC (28% de emprego de *nós*) e de Chapecó-SC (52% de uso do *nós*), mesmo ambas contando com informantes de perfis sociais muito semelhantes.

Os resultados de outros trabalhos também demonstram que a variação diatópica não é relevante para a AP de 1PP, pois há variedades geograficamente distantes, como a de João Pessoa-PB, a de Pelotas-RS e a do interior paulista, com comportamentos semelhantes em relação ao uso de *nós* e *a gente* (79%, 78% e 73,8% de emprego do pronome *a gente*, respectivamente).

Variação na concordância verbal de primeira pessoa do plural (1PP)

Após a observação e análise do fenômeno da AP de 1PP, passamos a tratar de dois outros fenômenos relacionados à 1PP: a CV variável junto do pronome *nós* e do pronome *a gente*. A seguir, apresentamos os resultados gerais para esses dois fenômenos.

Tabela 2 - Concordância verbal de primeira pessoa do plural no português brasileiro do interior paulista.

PRONOME	NÓS		A GENTE	
	1PP	3PS	1PP	3PS
VARIEDADE				
PB do interior paulista	85,5% (488/570)	14,5% (82/570)	6% (98/1603)	94% (1505/1603)

Fonte: RUBIO (2012).

No PB do interior paulista, evidencia-se, junto do pronome *nós*, uma frequência de uso de 85,5% de formas em 1PP (como em (4.b)) e de 14,5%, de formas em 3PS (como em (4.b)). Já a regra de CV com *a gente* é praticamente semicategórica, em favor de formas verbais em 3PS (como em (4.c)), com percentual de 94%, contra apenas 6% de emprego de 1PP (como em (4.d)).

- (4.a) *Nós* nunca **desconfiô(u)** de nada... porque viVIA dentro da sua casa...
 a gente conviVIA ali

[BDI-090-500]

- (4.b) aí *nós tivemo(s)* a oportunidade de:... conhecê(r) Fernando de Noronha...

[BDI-051-215]

- (4.c) mais um po(u)co pra frente... já tem uma entradazinha... né? que *a gente fala* que é a Lagoa Seca né?

[BDI-132-195]

- (4.d) meu irmão/ *a gente dividimo(s)* o lo::te... meu irmão tá construindo no fun::do uma casa

[BDI -038-170]

Considerando as diferentes características dos fenômenos variáveis aqui considerados, apresentamos, no quadro que se segue, a ordem de seleção das variáveis neles atuantes.

Quadro 5 - Ordem de seleção das variáveis consideradas para os fenômenos de concordância verbal de primeira pessoa do plural no português brasileiro do interior paulista.

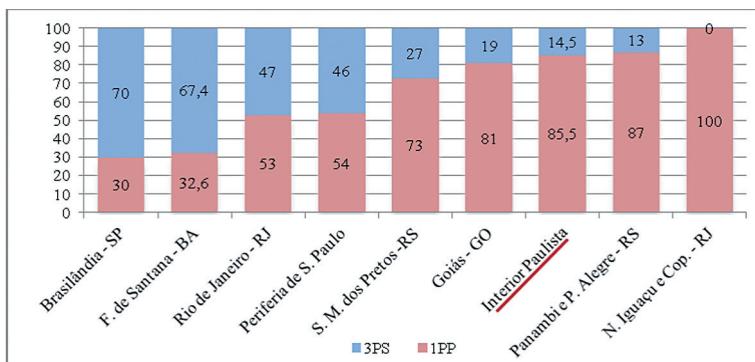
Variáveis	Fenômeno	CV com <i>nós</i> PB do interior paulista	CV com <i>a gente</i> PB do interior paulista
Linguísticas	Explicitude do sujeito	5º	4º
	Paralelismo discursivo	4º	1º
	Saliência fônica verbal	2º	2º
	Tempo e modo verbal	não selecionada	não selecionada
	Grau de determinação do sujeito	não selecionada	3º
Sociais	Gênero	não selecionada	não selecionada
	Faixa etária	3º	5º
	Escolaridade	1º	não selecionada

Fonte: RUBIO (2012).

A seleção das variáveis exibida acima justifica a consideração dos casos de possíveis variações em relação à CV de 1PP de forma individual, uma vez que cada um dos fenômenos variáveis apresenta diferente ordem de seleção dos fatores e diferentes fatores relevantes no processo de variação. Para a CV com *nós*, por exemplo, destaca-se a relevância de dois, dos três fatores sociais considerados, inclusive com a seleção de *escolaridade* como variável mais relevante. Da observação das variáveis consideradas para a CV com *a gente*, entretanto, verifica-se menor influência de variáveis sociais, já que apenas *faixa etária* foi selecionada e, diga-se, como última na ordem de relevância, e maior influência de variáveis estruturais, com a seleção de quatro, das cinco consideradas.

Nos dois próximos gráficos, efetuamos a comparação dos resultados obtidos em nosso estudo para a CV de 1PP com *nós e a gente* com os resultados de outras variedades do PB.

Gráfico 2 - Concordância verbal com o pronome *nós* em variedades do português brasileiro.



Fonte: Elaboração própria.¹⁷

Em relação à forma de 1PP concordante com o pronome *nós*, com base nos dados do gráfico 2, verifica-se que o fenômeno se atesta como variável para a maioria das variedades pesquisadas, em maior ou menor grau de emprego das formas verbais de 1PP, com variação de até 57 pontos percentuais (variedade de Brasilândia vs. variedade de Panambi e de Porto Alegre). Se consideradas as variedades do PB de Copacabana e Nova Iguaçu, no Rio de Janeiro, contudo, constata-se emprego categórico de formas verbais de 1PP junto do pronome *nós*.

A frequência de emprego da desinência de 1PP com o pronome *nós* apresentada nas amostras de fala do interior paulista (85,5%), a princípio, surpreende, por ser mais elevada que a apresentada em inúmeras outras variedades. Entretanto, a

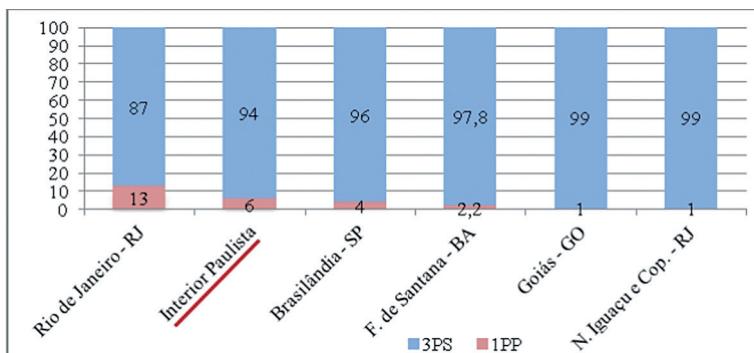
¹⁷ Características sociais das amostras: Brasilândia-Periferia de SP (COELHO, 2006): escolaridade de nula até 8 anos, faixa etária de menos de 25, de 25 a 50 e de mais de 50 anos e gêneros masculino e feminino; Feira de Santana-BA (CARMÓ; ARAÚJO, 2010): escolaridades nula e fundamental (português popular), faixas etárias diversas, informadas no trabalho, e gêneros masculino e feminino; Rio de Janeiro-RJ (NARO; GORSKI; FERNANDES, 1999): escolaridade de até 3 anos e de 4 a 8 anos, faixa etária de 6 a 12, 13 a 20, 21 a 40 e mais de 40 anos e gêneros masculino e feminino; Periferia de São Paulo (RODRIGUES, 1987): escolaridade nula e de até 4 anos, faixa etária de 20 a 35, 36 a 50 e mais de 51 anos e gêneros feminino e masculino; São Miguel dos Pretos-RS (ALMEIDA, 2006): escolaridade não explicitada, faixa etária de 15 a 24, 40 a 64 e 65 a 90 anos e gêneros masculino e feminino; Goiás-GO (MATTOS, 2010): escolaridade ensino médio e superior, faixa etária de menos de 21, de 21 a 40 e acima de 41 anos e gêneros masculino e feminino; Interior Paulista-Iboruna (RUBIO, 2012): escolaridade fundamental, média, superior, faixa etária de 16 a 25, 26 a 35, 36 a 55 e mais de 55 anos e gêneros feminino e masculino; Panambi e Porto Alegre-RS (ZILLES; MAYA; SILVA, 2000): escolaridade de baixa até superior, faixa etária de 25 a 49 e mais de 50 anos e gêneros feminino e masculino; Nova Iguaçu e Copacabana-RJ (VIANNA, 2011): escolaridade fundamental, média e superior, faixa etária de 18 a 35, de 36 a 55 e mais de 55 anos e gêneros masculino e feminino.

observação das características sociais de cada *corpus*, principalmente o nível de escolaridade dos informantes, fornece explicações para a frequência mais elevada da amostra do interior de São Paulo: as amostras que apresentam menores frequências de verbos em 1PP com o pronome *nós* possuem, quase em sua totalidade (exceção feita para a amostra de Goiás, que possui frequência inferior de CV (81%), mas bastante próxima da frequência de CV do interior paulista (85,5%)), informantes com níveis de escolaridade menores do que os da amostra do banco de dados Iboruna e também das amostras de CV de Panambi e Porto Alegre e de Nova Iguaçu e Copacabana, as quais apresentam todos os informantes com escolaridade superior.

A consideração apenas das frequências gerais de emprego das formas de 1PP e de 3PS junto do pronome *nós* e das características sociais dos informantes, obviamente, não permite que se determine com clareza todos os fatores responsáveis pela amplitude de variação na CV entre as diversas variedades do PB, porém é possível confirmar, de antemão, que, diferentemente do fenômeno variável de AP *nós* e *a gente*, o fenômeno variável de CV de 1PP com *nós* é influenciado diretamente pelo nível de escolarização.

Vejamos, agora, o gráfico 3, comparativo da CV com o pronome *a gente*.

Gráfico 3 - Concordância verbal com a forma pronominal *a gente* em variedades do português brasileiro.



Fonte: Elaboração própria.¹⁸

¹⁸ Características sociais das amostras: Rio de Janeiro-RJ (NARO; GORSKI; FERNANDES, 1999): escolaridade de até 3 anos e de 4 a 8 anos, faixa etária de 6 a 12, 13 a 20, 21 a 40 e mais de 40 anos e gêneros masculino e feminino; Interior Paulista- Iboruna (RUBIO, 2012): escolaridade fundamental, média, superior, faixa etária de 16 a 25, 26 a 35, 36 a 55 e mais de 55 anos e gêneros feminino e masculino; Brasilândia-Periferia de São Paulo (COELHO, 2006): escolaridade: de nula até 8 anos, faixa etária de menos de 25, de 25 a 50 e de mais de 50 anos e gêneros masculino e feminino; Feira de Santana-BA (CARMO; ARAÚJO, 2010): escolaridade nula e fundamental (português popular), faixas etárias diversas, não informadas e gêneros masculino e feminino; Goiás-GO (MATTOS, 2010): escolaridade média e superior, faixa etária de menos de 21, de 21 a 40 e acima de 41 anos e gêneros masculino e feminino; Nova Iguaçu e Copacabana-RJ (VIANNA, 2011): escolaridade

A CV com pronome *a gente* não pode ser caracterizada como fenômeno amplamente variável no PB, pois, na maioria das variedades consideradas, o emprego de formas verbais de 3PS supera 95% (Brasilândia-periferia de São Paulo, Feira de Santana-BA, municípios do estado de Goiás, Nova Iguaçu e Copacabana-RJ), sugerindo aplicação semicategórica dessas formas junto do pronome. A amostra do interior paulista apresentou uma variação pouco superior a 5%, o que faz com que o fenômeno de CV junto de *a gente* na comunidade também se classifique como semicategórico, com grande predominância de formas verbais de 3PS junto do pronome *a gente*. Das variedades investigadas, a única para a qual o fenômeno pode ser considerado variável é a do Rio de Janeiro-RJ, que apresentou 13% de uso de 1PP para a forma *a gente*.

Variação na concordância verbal de terceira pessoa do plural (3PP)

São inúmeros os estudos sociolinguísticos que tratam da CV de 3PP nas variedades do PB e muitas são as variáveis linguísticas e sociais que demonstram exercer influência sobre o fenômeno nessas variedades. Apresentamos, a seguir, os resultados relativos à CV de 3PP, para a fala do interior paulista.

Tabela 3 - Concordância verbal de terceira pessoa do plural no português brasileiro do interior paulista

VARIEDADE	DESINÊNCIA DE 3PP	DESINÊNCIA DE 3PS	TOTAL
PB – interior paulista	73% (1.971)	27% (728)	100% (2.699)

Fonte: RUBIO (2012).

Os resultados gerais apontam uma frequência de 73% de emprego de formas em 3PP junto de sujeitos de 3PP (ocorrência (5.a)), e de 27% em 3PS com sujeitos em 3PP (ocorrência (5.b)). As variáveis relevantes para os fenômenos seguem apresentadas no quadro 6, a seguir.

- (5.a) até os noivos **foram** hospitalizados... na Santa Casa de Rio Preto... não **puderam** nem seguir para a lua-de-mel... **ficaram** ali hospitalizados

[BDI-113-85]

fundamental, média e superior, faixa etária de 18 a 35, de 36 a 55 e mais de 55 anos e gêneros masculino e feminino.

(5.b) nessa fazenda... meus avôs **fez** um cercado... **fez** um pomar de:: jabuticaba... então... quando as jabuticabas nasciam... que estavam na época de colher eles...ela tocava né... um... beRRANte...

[BDI-102-174]

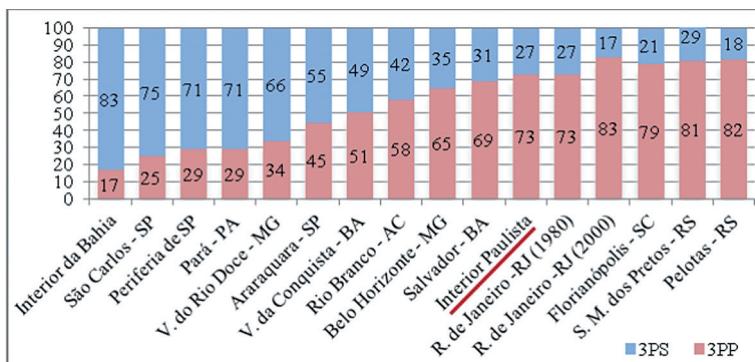
Quadro 6 - Ordem de seleção das variáveis para o fenômeno de concordância verbal de terceira pessoa do plural no português brasileiro do interior paulista.

Variáveis	Fenômeno	CV DE 3PP no PB do interior paulista
Linguísticas	Posição do sujeito	6º
	Traço semântico do referente do sujeito	5º
	Paralelismo formal discursivo	2º
	Saliência fônica verbal	3º
	Paralelismo formal oracional	4º
	Tipo morfológico do sujeito	não selecionada
Sociais	Escolaridade	1º
	Faixa etária	7º
	Gênero	8º

Fonte: RUBIO (2012).

De acordo com os resultados do quadro 6, observa-se que a CV de 3PP na fala do interior paulista sofre a influência direta de variáveis de natureza tanto linguística quanto social, o que se confirma pela seleção de oito delas, das nove consideradas, com destaque para a variável social **escolaridade**, primeira em ordem de relevância, e para as variáveis linguísticas **paralelismo linguístico discursivo** e **saliência fônica**, selecionadas em 2º e 3º lugares, respectivamente. Variáveis relacionadas ao sujeito não se mostraram de grande relevância, já que duas delas foram selecionadas em quinto e sexto lugares (**traço semântico** e **posição**, respectivamente) e a terceira não foi selecionada (**tipo morfológico**). Acrescente-se ainda que a seleção da variável **escolaridade** como a mais relevante aponta que o fenômeno é fortemente estigmatizado na fala do interior paulista, a exemplo do que ocorre no PB, em geral. No gráfico 4, apresentamos a comparação dos resultados gerais obtidos em nosso estudo com os de outras variedades do PB.

Gráfico 4 - Concordância verbal de terceira pessoa do plural em diferentes variedades do português brasileiro.



Fonte: Elaboração própria.¹⁹

Como se pode observar, a CV de 3PP atinge diferentes frequências nas variedades do PB, o que se comprova, inclusive, pela observação dos extremos do gráfico, que revelam discrepância de 65 pontos percentuais entre a amostra do interior da Bahia (17% de verbos em 3PP) e a amostra de Pelotas-RS (82% de CV de 3PP). Da mesma forma, a diferença de frequências entre variedades de regiões próximas, como a da região Noroeste de SP (73% de CV de 3PP) e a da região Central do estado de SP (45% de CV de 3PP), pode chegar a 28 pontos percentuais, e entre cidades vizinhas do interior de SP, como a de Araraquara (45% de uso de 3PP) e de São Carlos (25% de 3PP), a 20 pontos percentuais.

¹⁹ Características sociais das amostras: Interior da Bahia-Português popular (SILVA, 2005): escolaridade nula ou fundamental; faixa etária variável, de 25 a 107 anos; e gêneros masculino e feminino; São Carlos-SP (MONTE, 2007): analfabetos e escolarizados (EJA); faixa etária variável, de 22 a 38 anos; e gêneros masculino e feminino; Periferia de São Paulo (RODRIGUES, 1987): escolaridade nula e de até quatro anos; faixa etária de 20 a 35, de 36 a 50 e de mais de 51 anos; e gêneros masculino e feminino; Pará-PA (NINA, 1980): escolaridade nula (somente analfabetos); faixa etária variável, de 25 a 75 anos; e gêneros feminino e masculino; Vale do Rio Doce-MG (GONÇALVES, 2007): escolaridade fundamental, média e superior; faixa etária de 15 a 18, de 35 a 45 e de mais de 63 anos; e gêneros masculino e feminino; Araraquara-SP (GAMEIRO, 2005): escolaridade variável, de nula até mais de 12 anos, faixa etária de 7 a 15, de 16 a 25, de 26 a 55 e de mais de 55 anos; e gêneros masculino e feminino; Vitória da Conquista-BA (OLIVEIRA, 2005): escolaridade fundamental, média e superior; faixa etária de 15 a 25, de 26 a 49 e de mais de 50 anos; e gêneros masculino e feminino; Rio Branco-AC (RODRIGUES, 1997): analfabetos, 1 a 4 anos e 5 a 8 anos de escolaridade; faixa etária de 20 a 35 anos; e gêneros masculino e feminino; Belo Horizonte-MG (FARIA, 2008): escolaridade fundamental, média e superior; faixa etária de 17 a 29 anos; e gêneros masculino e feminino; Salvador-BA (SOUZA, 2011): escolaridade fundamental, média e superior, faixa etária de 15 a 24, de 25 a 35, de 45 a 55 e de mais de 65 anos; e gêneros masculino e feminino; Rio de Janeiro-RJ (SCHERRE; NARO, 2010): escolaridade de 1 a 4 anos, de 5 a 9 anos e de 9 a 11 anos; faixa etária de 7 a 14, de 15 a 25, de 26 a 49 e de mais de 49 anos; e gêneros masculino e feminino; Interior Paulista-Iboruna (RUBIO, 2012): escolaridade de 1 a 4, de 5 a 8, de 9 a 11 e de mais de 12 anos; faixa etária de 16 a 25, de 26 a 35, de 36 a 55 e de mais de 55 anos; e gêneros masculino e feminino; Florianópolis-SC (MONGUILHOTT; COELHO, 2002): escolaridade de 4 anos e 11 anos, faixa etária de 15 a 25, de 26 a 49 e de mais de 49 anos; e gêneros masculino e feminino; São Miguel do Pretos-RS (ALMEIDA, 2006): escolaridade não explicitada; faixa etária de 16 a 24, de 40 a 64 e de 65 a 90 anos; e gêneros masculino e feminino; Pelotas-RS (WELCHEN, 2009): escolaridade fundamental, média, superior e pós-graduação; faixa etária de 15 a 25, de 26 a 49 e de mais de 49 anos; e gêneros masculino e feminino.

Os resultados alcançados para cada variedade dependem, obviamente, dos fatores sociais considerados na composição das amostras de fala das comunidades investigadas, principalmente o fator *escolaridade*. As variedades que apresentaram menores frequências de pluralização verbal são também as que possuem informantes com os menores níveis de escolaridade (amostra do interior da Bahia, de São Carlos-SP, da periferia da cidade de São Paulo e do Pará). Por outro lado, as amostras que apresentaram maiores frequências de emprego da variante padrão, verbos em 3PP, possuem informantes com escolarização superior.

Se compararmos as frequências para a variedade carioca de 1980, reapresentadas em Scherre e Naro (2010), e as da variedade do interior paulista, é possível notar os mesmos percentuais de CV de 3PP, 73%, ainda que haja distância cronológica, geográfica e social entre as amostras. O mesmo se pode observar também nas pesquisas da periferia de São Paulo e na amostra do Pará, que exibiram, ambas, 29% de emprego de verbos em 3PP para sujeitos de 3PP.

Conclusões

Na exposição dos resultados gerais para os fenômenos variáveis relacionados à CV e à AP no PB do interior paulista, procuramos situá-los no cenário mais amplo da sociolinguística brasileira, mostrando aproximações e distanciamentos desta variedade em relação a outras do PB, sem deixar de nos atentarmos para as características consideradas na estratificação social das comunidades de fala comparadas. Propomos, abaixo, um quadro que sumariza os resultados alcançados para os fenômenos investigados na fala do interior paulista e as respectivas variáveis linguísticas e sociais atuantes em cada um deles.

Quadro 7 - Características dos fenômenos variáveis relacionados à concordância verbal e à alternância pronominal no português brasileiro do interior paulista.

VARIEDADE FENÔMENOS	PORUGUÊS BRASILEIRO DO INTERIOR PAULISTA BANCO DE DADOS IBORUNA
	COMPORTAMENTO / VARIÁVEIS ATUANTES
ALTERNÂNCIA PRONOMINAL DE 1ª PESSOA DO PLURAL	variável paralelismo discursivo > saliência fônica > grau de determinação do sujeito > tempo e modo verbal > escolaridade > faixa etária.
CONCORDÂNCIA VERBAL COM <i>NÓS</i>	variável saliência fônica > paralelismo discursivo > explicitude do sujeito > escolaridade > faixa etária.
CONCORDÂNCIA VERBAL COM <i>A GENTE</i>	semicategórica paralelismo discursivo > saliência fônica > grau de determinação do sujeito > explicitude do sujeito > faixa etária.
CONCORDÂNCIA VERBAL DE 3ª PESSOA DO PLURAL	variável paralelismo discursivo > saliência fônica > paralelismo oracional > traço semântico do sujeito > posição do sujeito > escolaridade > faixa etária > gênero.

Fonte: RUBIO (2012).

Ainda que, por questão de limitação de espaço, não tenhamos discutido em pormenores os resultados para cada uma das variáveis atuantes nos diferentes fenômenos variáveis, a visão panorâmica que oferecemos permitiu constatar que, na fala do interior paulista, observadas as mesmas características sociais de outras comunidades de fala do PB: (i) prevalece, com acentuada frequência, o uso de *a gente* em detrimento de *nós*, com influência apenas da variável faixa etária, como na maior parte das variedades do PB; (ii) a concordância verbal com *nós* em 1PP apresenta índices bastante elevados, como a de outras variedades do sul e sudeste do Brasil; (iii) a concordância verbal com *a gente* constitui regra semicategórica, como se observa também na maior parte dos dialetos do PB; (iv) a concordância verbal de 3PP apresenta índices acima dos 70%, como ocorre em outras variedades do sul e sudeste do país.

Continuam, no interior do Projeto ALIP, os trabalhos de descrição da fala do interior paulista, a partir dos quais esperamos, na próxima década, seja possível revelar, por completo, uma fotografia sociolinguística da região.

RUBIO, C. F.; GONÇALVES, S. C. L. The Speech from the Countryside Region of São Paulo State in the Brazilian Sociolinguistics Panorama: an Overview of Verbal Agreement and Pronominal Alternation. *Alfa*, São Paulo, v.56, n.3, p.993-1024, 2012.

■ *ABSTRACT:* As part of sociolinguistic researches developed by ALIP (speech samples from the countryside region of São Paulo) Project, in this paper we present general results for three variable phenomena of speech from countryside region of São Paulo State: (i) verbal agreement of first person plural; (ii) verbal agreement of third person plural; (iii) pronominal alternation between the forms of first person plural of discourse “nós” (we) and “a gente”. Based on principles of Labovian Sociolinguistics, whereby interdialectal comparison, we show how these variable phenomena are inserted in the Brazilian Sociolinguistics panorama. The results allow us to find similarities and differences among the speech from countryside region of São Paulo State and other varieties from Brazilian Portuguese, highlighting the relevance of the description of this variety still little known in the Brazilian Sociolinguistics panorama.

■ *KEYWORDS:* Paulista Portuguese. Verbal agreement. Pronominal alternation.

REFERÊNCIAS

- ALMEIDA, A. P. *A concordância verbal na comunidade de São Miguel dos Pretos, Restinga Seca, RS*. 2006. 159 f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.
- AMARAL, A. *O dialeto caipira*. 3.ed. São Paulo: Hucitec, 1976.
- AMARAL, L. I. C. *A concordância verbal de segunda pessoa do singular em Pelotas e suas implicações linguísticas e sociais*. 2003. 181 f. Dissertação (Mestrado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

ANTONINO, V.; BANDEIRA, M. Nós, a gente e a concordância em uma comunidade afro-brasileira isolada. *Papia*, São Paulo, n.21, v.1, p.159-176, 2011.

BECHARA, E. *Moderna gramática portuguesa*. 37.ed. Rio de Janeiro: Lucerna, 2002.

BENTES, A. C. *É nós na fita: sobre a formação de registros e a elaboração de estilos no campo da cultura popular urbana paulista*. 2009. Projeto de Pesquisa financiado pela FAPESP, Universidade Estadual de Campinas, Campinas, 2009.

BORGES, P. R. S. *A grammaticalização de “a gente” no português brasileiro*. 2004. 182 f. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

BORTONI-RICARDO, S. M. *The urbanization of rural dialect speakers: a sociolinguistic study in Brazil*. Cambridge: University Press, 1985.

CARENO, M. F. *Vale do Ribeira*: a voz a vez das comunidades negras. São Paulo: Arte & Ciência, 1997.

CARMO, M. C. *As vogais médias pretônicas dos verbos na fala culta do interior paulista*. 2009. 122 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2009.

CARMO, S. D. S.; ARAÚJO, S. S. F. A concordância verbal com a primeira pessoa do plural no português popular falado em feira de Santana-BA. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA, 14., 2010, Santana. *Anais...* Santana: UEFS, 2010. p.575-580.

CASTILHO, A. T. (Org.). *História do português paulista*. Campinas: Ed. da Unicamp, 2009. (Série Estudos).

_____. Português culto falado no Brasil: história do Projeto NURC/BR. In: PRETI, D., URBANO, H. (Org.). *A linguagem falada culta na cidade de São Paulo*. São Paulo: T. A. Queiroz, 1990. (Estudos; v.4).

COELHO, R. *É nós na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana*. 2006. 175 f. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo, 2006.

CUNHA, C.; CINTRA, L. *Gramática do Português contemporâneo*. Rio de Janeiro: Nova Fronteira, 1984.

DUARTE, M. E. L. *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*. 1986. 150f. Dissertação (Mestrado em Linguística) – Pontifícia Universidade Católica, São Paulo, 1986.

FARIA, N. V. M. *A concordância verbal no português de Belo Horizonte*. 2008. 142f. Dissertação (Mestrado em Língua Portuguesa e Linguística) – Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2008.

FELÍCIO, C. P. *Gramaticalização da conjunção concessiva ‘embora’*. 2008. 140f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008.

FERNANDES, E. A. Nós x a gente: variação estável ou mudança em progresso? In: SOARES, M. E.; ARAGÃO, M. S. S. (Ed.). In: JORNADA DE ESTUDOS LINGÜÍSTICOS, 16., 1999, Fortaleza. *Anais...* Fortaleza: UFC, 1999. p.331-334.

FERNANDES, F. O. *Construções com os verbos andar, continuar, ficar e viver seguidos de gerúndio: um estudo na interface sociolinguística/gramaticalização*. 2010. 110f. Trabalho de Iniciação Científica (Licenciatura em Letras) – Universidade Estadual Paulista, São José do Rio Preto, São Paulo, 2010.

FERREIRA, J. S. *O apagamento do /d/ em morfema de gerúndio no dialeto de São José do Rio Preto*. 2010. 145f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto.

FIAMENGUI, A. H. R. *A marcação de pluralidade no SN na fala e na escrita de adolescentes da região de São José do Rio Preto*. 2011. 144f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2011.

FONSECA, A. M. H. *A perífrase verbal ir + infinitivo e o futuro do dialeto riopretano: um estudo na interface sociolinguística/gramaticalização*. 2010. 176f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010.

FORTILLI, S. C. *Orações completivas em posição argumental de sujeito: gramaticalização e dessentencialização de orações matrizes*. 2011. 98f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2011.

_____. *As construções não-verbais no português falado no interior do Estado de São Paulo*. 2007. 145f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2007.

GAMEIRO, M. B. *A concordância verbal na língua falada da região central do estado de São Paulo*. 2005. 198f. Dissertação (Mestrado em Língua Portuguesa) – Universidade Estadual Paulista, Araraquara, 2005.

GONÇALVES, S. C. L. *Orações completivas em posição argumental de sujeito II: sincronia e diacronia*. 2012. Projeto de Pesquisa financiado pelo CNPq, Universidade Estadual Paulista, São José do Rio Preto, 2012.

_____. Orações subjetivas: variância e invariância de padrões na fala e na escrita. *Revista da ABRALIN*, Brasília, v.10, p.87-111, 2011.

_____. Projeto ALIP (amostra linguística do interior paulista): questões teóricas e metodológicas sobre a constituição de um banco de dados de língua falada. In: TAGNIN, E.; VALE, O. A. (Org.). *Avanços da lingüística de corpus no Brasil*. São Paulo: Humanitas, 2008. p.217-245.

_____. *O português falado na região de São José do Rio Preto*: constituição de um banco de dados anotado para o seu estudo. Relatório científico parcial I apresentado à FAPESP, Universidade Estadual Paulista, São José do Rio Preto, 2005. Disponível em <<http://www.iboruna.ibilce.unesp.br/histórico/relatorio1>>. Acesso em: 22 fev. 2011.

GRACIOSA, D. M. D. *Concordância verbal na fala culta carioca*. 1991. 176f. Dissertação (Mestrado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.

GUERRA, A. R. *Funções textual-interativas dos marcadores discursivos*. 2007. 233f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2007.

HAUSEN, T. A. P. *Concordância verbal do pronome tu no interior do estado de Santa Catarina*. 2000. 143f. Dissertação (Mestrado em Letras) – Universidade Federal do Paraná, Curitiba, 2000.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. *Censo demográfico 2000*. Brasília, 2000. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/default_censo_2000.shtml>. Acesso em: 20 out. 2002.

LABOV, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LEMLE, M.; NARO, A. J. *Competências básicas do Português*. Rio de Janeiro: Mobra/Fundação Ford, 1977.

LIMA, A. C. *Predicações não-verbais avaliativas encaixadas em predicados avaliativos*. 2012. 20f. Projeto de Pesquisa para o Mestrado, Universidade Estadual Paulista, São José do Rio Preto, 2012.

LOPES, C. R. S. *A inserção de a gente no quadro pronominal do português: percurso histórico*. 1999. 181f. Tese (Doutorado em Língua Portuguesa) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999.

LOPES, O.; NARO, A. J. Concordância variável de primeira pessoa do singular no PB – amostra variável. In: CONGRESSO NACIONAL DE LINGÜÍSTICA E FILOLOGIA, DO CÍRCULO DE ESTUDOS LINGÜÍSTICOS E FIOLÓGICOS DO RIO

DE JANEIRO, 15., Rio de Janeiro. *Anais...* Rio de Janeiro: UERJ, 2011. Disponível em: <<http://www.filologia.org.br>>. Acesso em: 23 out. 2011.

LOREGIAN, L. *Concordância verbal com o pronome tu na fala do sul do Brasil*. 1996. 156f. Dissertação (Mestrado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 1996.

LUCCHESI, D.; BAXTER, A.; SILVA, J. A. A. A concordância verbal. In: LUCCHESI, D.; BAXTER, A.; RIBEIRO, I. (Org.). *O português afro-brasileiro*. Salvador: Ed. da UFBA, 2009. p.331-371.

MARCATO, F. *Análise prosódica de clíticos preposicionais na variedade rio-pretense*. 2011. 20f. Projeto de Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2011.

MATTOS, S. E. R. A primeira pessoa do plural em Goiás. In: MARÇALO, M. J. et al. (Ed.). *Língua portuguesa: ultrapassar fronteiras, juntar culturas*. Évora: Universidade de Évora, 2010. p.11-22.

_____. *Sujeito coletivo singular em português: concordância e referencialidade*. 2003. 105f. Dissertação (Mestrado em Linguística) – Universidade Federal de Brasília, Brasília, 2003.

MENDES, R. B. *SP2010: construção de uma amostra da fala paulistana*. 2011. Projeto de pesquisa financiado pela FAPESP, Universidade de São Paulo, São Paulo, 2011.

MODESTO, A. T. T. *Formas de tratamento no português brasileiro: a alternância tu/você na cidade de Santos-SP*. 2006. 128f. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo, 2006.

MONGUILHOTT, I. O. S.; COELHO, I. L. Um estudo da concordância verbal de terceira pessoa em Florianópolis. In: VANDRESEN, P. (Org.). *Variação e mudança no português falado na região sul*. Pelotas: Educat, 2002. p.189-216.

MONTE, A. *Concordância verbal e variação: uma fotografia sociolinguística da cidade de São Carlos*. 2007. 114 f. Dissertação (Mestrado) – Universidade Estadual Paulista, Araraquara, 2007.

NARO, A. J.; GÖRSKI, E.; FERNANDES, E. Change without change. *Language variation and change*, Cambridge, v.11, n.2, p.197-211, 1999.

NARO, A. J.; SCHERRE, M. M. P. Sobre o efeito do princípio da saliência na concordância verbal na fala moderna, na escrita antiga e na escrita moderna. In: MOURA, D. (org.) *Os múltiplos usos da língua*. Maceió: EDUFAL, 1999. p.26-37.

_____. A hierarquização do controle da concordância no português moderno e medieval: o caso de estruturas de sujeito composto. In: GROBE, S.; ZIMMERMANN,

K. (Org.). *O português brasileiro: pesquisas e projetos*. Frankfurt am Main: TFM, 2000a. p.167-188.

_____. Variable concord in Portuguese: the situation in Brazil and Portugal. In: MCWHORTHER, J. (Org.). *Language change and language contact in pidgins and creoles*. Amsterdam: John Benjamins, 2000b. p.235-255.

_____. A relação verbo/sujeito: o efeito máscara do que relativo. In: HORA, D.; COLLISCHONN, G. *Teoria linguística: fonologia e outros temas*. João Pessoa: Universitária, 2003. p.383-401.

_____. *Origens do português brasileiro*. São Paulo: Parábola Editorial, 2007.

NEVES, M. H. M. *Gramática de usos do português*. São Paulo: Ed. da UNESP, 2000.

NICOLAU, E. M. D. *A ausência de concordância verbal em português: uma abordagem sociolinguística*. 1984. 179f. Dissertação (Mestrado em Linguística e Língua Portuguesa) – Universidade Federal de Minas Gerais, Belo Horizonte, 1984.

NINA, T. J. C. *Concordância nominal/verbal do analfabeto na microrregião de Bragantina*. 1980. 130f. Dissertação (Mestrado em Língua Portuguesa) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1980.

OLIVEIRA, M. S. *Concordância verbal de terceira pessoa do plural em Vitória da Conquista: um caso de variação estável*. 2005. 177f. Dissertação (Mestrado em Letras) – Universidade Federal da Bahia, Salvador, 2005.

OMENA, N. P. A referência à primeira pessoa do plural: variação ou mudança? In: PAIVA, M. de C.; DUARTE, M. E. L. (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra Capa Livraria, 2003.

_____. A referência variável da primeira pessoa do discurso no Plural. In: NARO, A. J. et al. *Relatório final de pesquisa: projeto subsídios do projeto censo à educação*. Rio de Janeiro: Ed. da UFRJ, 1986. p.286-319.

OMENA, N. P.; BRAGA, M. L. A gente está se gramaticalizando? In: MACEDO, A. T.; RONCARATI, C.; MOLLICA, M. C. (Org.). *Variação e discurso*. Rio de Janeiro: Tempo Brasileiro, 1996. p.75-84.

ORLANDI, P. S. *Usos e (des)usos da flexão verbal de 2ª pessoa do singular em texto orais de informantes de Tubarão (SC): um estudo de caso*. 2004. 86f. Dissertação (Mestrado em Ciências da Linguagem) – Universidade do Sul de Santa Catarina, Tubarão, 2004.

PENHAVEL, E. *Marcadores discursivos e articulação tópica*. 2010. 136f. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 2010.

PAREDES SILVA, V. L. A variação você/tu na fala carioca. In: ENCONTRO DE VARIAÇÃO LINGÜÍSTICA DO CONE SUL, 1, 1996, Porto Alegre. *Anais...* Porto Alegre: UFRGS, 1996. Não paginado.

PERINI, M. A. *Gramática do português brasileiro*. São Paulo: Parábola, 2010.

RAMOS, A. P. *Descrição das vogais postônicas não-finais na variedade do noroeste paulista*. 2009. 177f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista.

RENCK, N. R. *Renovação e gramaticalização: o caso da locução conjuntiva*. 2011. 54f. Trabalho de Iniciação Científica (Licenciatura em Letras) – Universidade Estadual Paulista, São José do Rio Preto, 2011.

RODRIGUES, A. C. S. *A concordância verbal no português popular em São Paulo*. 1987. 189f. Tese (Doutorado em Língua Portuguesa) – Universidade de São Paulo, São Paulo, 1987.

RODRIGUES, A. N. *O dialeto caipira na região de Piracicaba*. São Paulo: Ática, 1974.

RODRIGUES, D. A. *A concordância verbal na fala urbana de Rio Branco*. 1997. 198f. Dissertação (Mestrado em Linguística) – Universidade Estadual de Campinas, Campinas, 1997.

RONCARATI, C. *Banco de dados interacionais*. Rio de Janeiro: Universidade Federal do Rio Janeiro, 1996.

RUBIO, C. F. *Padrões de concordância verbal e de alternância pronominal no português brasileiro e europeu: estudo sociolinguístico comparativo*. 2012. 392f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2012.

_____. *A concordância verbal na região noroeste do Estado de São Paulo*. 2008. 152f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008.

SALOMÃO, M. H. *A variação de pluralidade nas estruturas predicativas da variedade falada na região de São José do Rio Preto*. 2010. 162f. Dissertação (Mestrado em Estudos Linguísticos). Universidade Estadual Paulista, São José do Rio Preto, 2010.

SANTANA, L. *Relações de complementação no português brasileiro: uma perspectiva discursivo-funcional*. São Paulo: Cultura Acadêmica, 2010.

SANTOS, R. M. A. *A alternância indicativo/subjuntivo em estruturas complexas*. 2005. 152f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2005.

SCHERRE, M. M. P.; NARO, A. J. Perceptual vs. grammatical constraints and social factors in subject-verb agreement in Brazilian Portuguese. *U. Penn Working Papers in Linguistics*, Honolulu, v.16, n.2, p.165-171, 2010.

_____. Mudança sem mudança: a concordância de número no português brasileiro. *Scripta*, Belo Horizonte, v. 1, n. 18, 2006. p. 162-185.

_____. Sobre as origens estruturais do português brasileiro: crioulização ou mudança natural? *Papia - Revista de crioulos de base Ibérica*, Brasília: Thesaurus, 2001.

_____. Sobre a concordância de número no português falado do Brasil. In: RUFFINO, G. (Org.). *Dialettologia, geolinguistica, sociolinguistica*. Tübingen: Max Niemeyer Verlag, 1998a. p.153-188.

_____. Restrições sintáticas e semânticas no controle da concordância verbal em português. *Fórum linguístico*, Florianópolis, n. 1, p. 45-71, 1998b.

_____. A concordância de número no português do Brasil: um caso típico de variação inerente. In: HORA, D. (org.) *Diversidade linguística no Brasil*. João Pessoa: Ideia, 1997. p.93-114.

_____. Duas dimensões do paralelismo formal na concordância de número no português popular do Brasil. *DELTA*, São Paulo, v.9, n.1, p.1-14, 1993.

SEARA, I. C. A variação do sujeito *nós* e *a gente* na fala florianopolitana. *Organon*, Porto Alegre, v.14, n.28-29, p.179-194, 2000.

SILVA, J. A. A. *A concordância verbal de terceira pessoa do plural no português popular do Brasil*: um panorama sociolinguístico de três comunidades do interior da Bahia. 2005. 340f. Tese (Doutorado em Letras) – Universidade Federal da Bahia, Salvador, 2005.

SILVA, G. M. O. Variáveis sociais e perfil do corpus CENSO. In: SILVA, G. M. O.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos*. Rio de Janeiro: Tempo Brasileiro, 1996. p.51-81.

SILVA, G. M. O.; SCHERRE, M. M. P. *Padrões sociolinguísticos*: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro. Rio de Janeiro: Tempo Brasileiro, 1996.

SILVEIRA, A. A. M. *As vogais pretônicas na fala culta do noroeste paulista*. 2008. 143f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008.

SOUZA, C. B. A concordância verbal no português falado em Salvador: uma realidade linguística bipolarizada. *Papia*, Brasília, n.21, v.2, p.183-193, 2011.

SOUZA, E. R. F. *Gramaticalização dos itens linguísticos ‘assim’, ‘já’ e ‘ai’ no português brasileiro*: um estudo sob a perspectiva da gramática discursivo-

funcional. 2009. 290f. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 2009.

TAMANINE, A. M. B. *A alternância nós/a gente no interior de Santa Catarina*. 2002. 120f. Dissertação (Mestrado em Letras) – Universidade Federal do Paraná, Curitiba, 2002.

VENDRAME, V. *Os verbos ver, ouvir e sentir e a expressão de evidencialidade em língua portuguesa*. 2010. 140f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010.

VIANNA, J. B. S. *Semelhanças e diferenças na implementação de a gente em variedades do português*. 2011. 235f. Tese (Doutorado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

_____. *A concordância de nós e a gente em estruturas predicativas na fala e na escrita carioca*. 2006. 109f. Dissertação (Mestrado em Língua Portuguesa) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

VOTRE, S.; OLIVEIRA, M. R. *A língua falada e escrita na cidade do Rio de Janeiro: materiais para seu estudo*. Rio de Janeiro: UFRJ, 1995.

WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical foundations for a theory of language change. In: LHEMAN, W.; MALKIEL, Y. (Ed.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p.95-195.

WIEDEMER, M. L. *Variação e gramaticalização de preposições de complementos locativos de verbos de movimento*. 2011. 88f. Relatório Científico, Universidade Estadual Paulista, São José do Rio Preto.

ZILLES, A. M. S. O que a fala e a escrita nos dizem sobre a avaliação social do uso de a gente? *Letras de Hoje*, Porto Alegre, v.42, n.2, p.27-44, 2007.

_____. The development of a new pronoun: the linguistic and social embedding of *a gente* in Brazilian Portuguese. *Language Variation and Change*, Cambridge, v.17, n.1, p.19-53, 2005.

ZILLES, A. M. S.; MAYA, L.; SILVA, K. A concordância verbal com a primeira pessoa do plural em Panambi e Porto Alegre, RS. *Organon*, Porto Alegre, v.14, n.28/29, p.195-219, 2000.

Recebido em março de 2012

Aprovado em maio de 2012

CONCORDÂNCIA NOMINAL E VERBAL: CONTRIBUIÇÕES PARA O DEBATE SOBRE O ESTATUTO DA VARIAÇÃO EM TRÊS VARIEDADES URBANAS DO PORTUGUÊS

Silvia Figueiredo BRANDÃO*
Silvia Rodrigues VIEIRA**

- **RESUMO:** Focalizam-se a concordância nominal de número e a concordância verbal de terceira pessoa do plural em variedades urbanas do Português Europeu, do Português do Brasil e do Português de São Tomé, com base na fala de indivíduos de níveis fundamental, médio e superior de instrução, distribuídos, ainda, por três faixas etárias e sexo. Com o objetivo de descrever as referidas variedades, o trabalho avalia as motivações de natureza estrutural e social segundo os pressupostos da Teoria da Variação e Mudança (WEINREICH; LABOV; HERZOG, 1968). Os resultados das diferentes análises, desenvolvidas no âmbito do Projeto *Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português*, sugerem que, embora as três variedades manifestem tendência à adoção da regra de concordância tanto no Sintagma Nominal quanto no Sintagma Verbal, a concordância apresenta estatuto diferenciado. No Português Europeu, os dados revelam um comportamento que não pode ser considerado efetivamente variável – trata-se, nos termos de Labov (2003), de uma regra categórica no âmbito da concordância nominal e semicategórica no âmbito da verbal; no Português do Brasil e no Português de São Tomé, ambas as regras são variáveis.
- **PALAVRAS-CHAVE:** Concordância nominal. Concordância verbal. Sociolinguística. Variedades do Português.

Introdução

Um dos temas mais focalizados no âmbito do Português, sobretudo no que se refere à variedade brasileira, a concordância nominal e verbal suscita, ainda, grande interesse em virtude não só de fatores linguísticos que se vêm mostrando recorrentes em diversos estudos, mas também de implicações sócio-histórico-culturais para a caracterização das variedades do Português. Em contraste com o que ocorre no Português Europeu (PE), em que a regra canônica de concordância parece ter *status categórico ou semicategórico*,

* Pesquisadora do CNPq, UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras. Rio de Janeiro – Rio de Janeiro – Brasil. 21941-590 - silvia.brandao@terra.com.br

** Pesquisadora da FAPERJ, UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras. Rio de Janeiro – Rio de Janeiro – Brasil. 21941-590 - silviavieira@hotmail.com

observam-se, em variedades não europeias, padrões de concordância variáveis, que, embora com maior ou menor índice de produtividade em certas áreas (rurais, urbanas), obedecem a determinadas restrições comuns do ponto de vista tanto estrutural quanto extralingüístico. Desde 2008, vem-se desenvolvendo o Projeto *Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português*, inicialmente como projeto de cooperação internacional (Brasil e Portugal) apoiado pela CAPES¹ e, a partir de 2011, como um dos projetos da ALFAL. Ao idealizá-lo, pretendia-se, entre outros objetivos, testar hipóteses explicativas do conjunto dos fenômenos observados à luz dos princípios subjacentes a todas as gramáticas e dos parâmetros específicos a subconjuntos de gramáticas, bem como testar hipóteses prévias sobre a origem das diferenças/semelhanças observadas, tendo em conta a história do português.

As pesquisas empreendidas centraram-se, inicialmente, em variedades linguísticas urbanas. As cidades – onde hoje se concentra a grande maioria da população em virtude da crença de que nelas se encontram maiores oportunidades de trabalho e melhores condições de vida – constituem o espaço ideal para observar a interação entre indivíduos de diferentes estratos sociais. Sendo assim, organizou-se um *corpus*, disponível na internet², com entrevistas realizadas em diferentes comunidades das regiões metropolitanas do Rio de Janeiro e de Lisboa. No que se refere à variedade urbana de São Tomé e Príncipe, conta-se com entrevistas realizadas, em 2009, por Tjerk Hagemeijer e que compõem o *Corpus VARPOR*, do Centro de Linguística da Universidade de Lisboa.

Neste artigo, focalizam-se as variedades urbanas europeia (PE), brasileira (PB) e santomense (PST), no intuito de tecer considerações acerca das motivações de natureza estrutural e social à luz das análises empreendidas. Acredita-se que as análises contrastivas permitam observar até que ponto se pode avaliar como semelhante ou diferente a expressão gramatical de um fenômeno em termos quantitativos e qualitativos.

Perfil das comunidades

O confronto aqui realizado tem, necessariamente, de ser relativizado em termos não só das dimensões das áreas metropolitanas consideradas e do contingente populacional que as integra (a Região Metropolitana do Rio de Janeiro, por exemplo, tem população maior que a de Portugal como um todo), mas também de suas especificidades quanto a aspectos históricos, culturais e sócio-econômicos.

¹ Em sua versão inicial, o Projeto contou com a coordenação de Maria Antonia C. da Mota, da Universidade de Lisboa, e de Silvia Rodrigues Vieira, da Universidade Federal do Rio de Janeiro. Como Projeto ALFAL, assumiram a coordenação das atividades as autoras do presente artigo.

² Disponível em: <www.letras.ufrj.br/concordancia>

A Área Metropolitana de Lisboa constitui-se de 18 concelhos que recobrem 2.870km² e onde se concentram 2.819.433 habitantes (2008), “cerca de ¼ da população portuguesa”, bem como 30% das empresas nacionais.³ No concelho de Sintra, encontra-se Cacém, uma de suas 20 freguesias, com 21.289 habitantes.

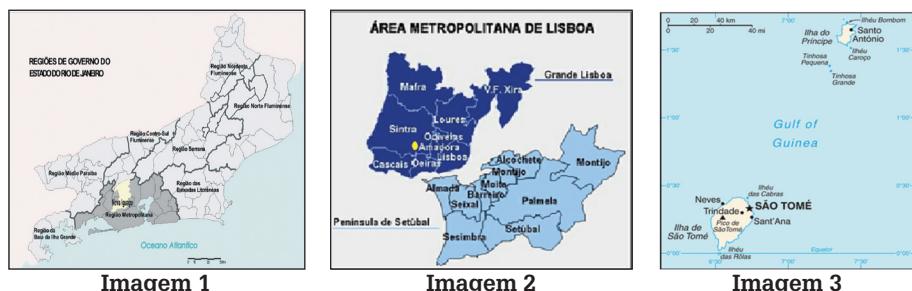
A Região Metropolitana do Rio de Janeiro compõe-se de 19 municípios, distribuídos por 5.645 km², e conta com 12.434.611 habitantes, o que corresponde a 76% da população do Estado. Dentre eles, destaca-se Nova Iguaçu, com 797.637 habitantes⁴, grande parte oriunda de diferentes áreas do Nordeste e de Minas Gerais.

Essas duas áreas metropolitanas têm em comum algumas características, tais como o crescimento desordenado da zona urbana, o significativo número de indústrias e a mobilidade populacional constante. Seus habitantes estão em permanente interação com os moradores das respectivas cidades-sede em função de a maioria deles nelas desenvolverem atividades profissionais.

Na área urbana de São Tomé, vive cerca de um terço da população de São Tomé e Príncipe (206.178 habitantes), arquipélago situado no Golfo da Guiné e onde, em virtude de uma série de contingências históricas, coexistem diversas línguas, entre as quais se destacam o Português e o Forro, que, de acordo com dados do censo realizado em 2001, são falados, respectivamente, por 98,9% e 72,4% dos indivíduos com mais de cinco anos (HAGEMEIJER 2009, p.18). A cidade, onde se desenrolam as principais atividades comerciais, é o principal porto do país e dispõe de estrutura precária.

As três áreas podem ser visualizadas nas imagens a seguir.

Figura 1 – Áreas metropolitanas (Rio de Janeiro, Lisboa e São Tomé)



Áreas Metropolitanas (1) do Rio de Janeiro, com a localização de Nova Iguaçu (mapa com alterações); (2) de Lisboa, com a localização de Cacém (cf. ponto amarelo); (3) de São Tomé (cf. estrela).

Fonte: METRODATA (2012); PLANO REGIONAL DE ORDENAMENTO DO TERRITÓRIO DA ÁREA METROPOLITANA DE LISBOA (2008); ILHA DE SÃO TOMÉ (2012).

³ Disponível em: <<http://www.aml.pt/index.php?&iLevel1=gaml&iLevel2=territorio&iContent=index.html>>. Acesso em: 10 jan. 2012.

⁴ INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. (2010).

Aspectos teórico-metodológicos e características dos corpora

A análise dos dados, tanto da concordância nominal quanto da verbal, parte dos pressupostos assumidos pela Sociolinguística Laboviana (WEINREICH; LABOV; HERZOG, 1968) no que diz respeito à concepção geral de regra variável com base no princípio da heterogeneidade ordenada, à postulação de restrições linguísticas e extralingüísticas que podem acelerar ou refrear processos de variação e mudança, bem como à tipologia proposta em Labov (2003), que diz haver três classes de regras linguísticas, conforme se demonstra a seguir, numa adaptação simplificada do quadro por ele delineado.

Quadro 1 - Tipos de regras linguísticas.

Tipo de regra	Frequência com que opera	Violações ¹
I - Categórica	100%	Nenhuma, na fala natural
II - Semicategórica	95-99%	Rara e relatável
III - Variável	5-95%	Nenhuma por definição e não relatável

Fonte: Com base em Labov (2003, p.243).⁵

Além dos conceitos sociolinguísticos ora assumidos, as investigações da concordância nominal e da verbal valem-se igualmente dos mesmos bancos de dados para a observação do fenômeno – o *corpus Concordância* e o *Corpus VARPOR* já referidos – e dos mesmos procedimentos analíticos, da coleta de dados ao tratamento quantitativo viabilizado pelo pacote de programas computacionais Goldvarb-X.

As análises foram desenvolvidas com base em dados coletados de entrevistas do tipo DID, realizadas com 53 indivíduos, 18 por variedade,⁶ distribuídos por sexo, três faixas etárias (18-35 anos, 36-55 anos e 56 anos em diante) e três níveis de escolaridade (fundamental: 5 a 8 anos; médio: 9 a 11 anos; superior). Os informantes brasileiros são naturais de Nova Iguaçu; os portugueses, de Cacém, e os de São Tomé vivem na capital do país. Além das três referidas variáveis extralingüísticas, controlaram-se (a) na análise da concordância nominal,

⁵ Labov assim descreve as violações, respectivamente, I, II e III: “none in natural speech; rare and reportable; none by definition and unreportable.”

⁶ Tendo em vista as variáveis extralingüísticas consideradas (sexo, três faixas etárias e três níveis de escolaridade), com um informante por célula, deveria haver um total de 54 informantes (18 por variedade). No entanto, no *Corpus VARPOR*, referente a São Tomé, não se encontrou uma mulher com mais de 56 anos com nível superior de instrução, o que se deve às próprias condições sócio-histórico-econômicas do país.

oito variáveis estruturais: (1) saliência fônica; (2) tonicidade e (3) número de sílabas do item singular; (4) classe do vocábulo; (5) posição linear e relativa do constituinte no SN; (6) marcas precedentes; (7) contexto fonológico subsequente; (8) animacidade do núcleo do SN; (b) na análise da concordância verbal, as variáveis (1) posição do sujeito em relação ao verbo; (2) distância entre o núcleo do SN e o SV; (3) presença de elementos intervenientes entre o SN sujeito e o verbo; (4) configuração morfossintática do SN sujeito; (5) paralelismo no nível oracional (entre sujeito e verbo); (6) animacidade do sujeito; (7) saliência fônica; (8) tempo verbal; (9) transitividade verbal.

Em relação a São Tomé, realizaram-se análises complementares sobre a concordância no SN, uma, com a ampliação do número de informantes de níveis fundamental e médio; outra, com base em amostra da fala de estudantes, em que foi possível controlar outras variáveis sociais. Quanto à concordância verbal, também foi realizada uma análise adicional com maior número de indivíduos de níveis fundamental e médio.

Sínteses das análises já realizadas

Aspectos gerais

Considerando os trabalhos já desenvolvidos com base no *corpus Concordância*,⁷ pode-se afirmar, tendo em vista exclusivamente os índices de concretização da marca de número,⁸ que, de um lado, os dados urbanos do PB e do PST em análise apresentam comportamento diferente do verificado no Português Europeu. Embora com expressões diferenciadas, brasileiros e são-tomenses apresentam variação entre concordância e não concordância, com preferência pela realização da marca de número, enquanto portugueses registram quase exclusivamente as marcas verbais e nominais de plural. A tabela e os gráficos a seguir permitem observar as tendências ora delineadas nas amostras constituídas com base na fala dos 53 informantes acima referidos.

⁷ Colaboraram na tarefa de coleta e tratamento dos dados de concordância verbal os então alunos de iniciação científica Fernanda Villares Barreto, Karen Cristina da Silva, Rodrigo Cunha da Silva e Daiane Rangel da Silva. Quanto à concordância nominal, colaboraram os alunos de graduação Aimée de Oliveira Silva e Vinicius Amado de Oliveira Ribeiro.

⁸ No âmbito da concordância verbal, foram coletadas todas as ocorrências de SN na 3ª pessoa do plural que inequivocamente desempenhasse o papel de sujeito do verbo da oração. Assim sendo, foram desprezadas as ocorrências em que tal SN viesse a funcionar como mero tópico discursivo, tendo sido abandonado em favor de uma concepção singular. Também foram desconsiderados os dados em que, por razões fonético-fonológicas, não foi possível determinar se houve ou não realização da marca de número.

Tabela 1 - Distribuição dos dados de concordância (marcação) e não concordância (não marcação) nominal de número e verbal de 3^a pessoa plural nas três variedades estudadas.

Variedade	Concordância nominal		Concordância verbal	
	Marcação	Não marcação	Marcação	Não marcação
PB <i>(Nova Iguaçu)</i>	3439/3777 (91,1%)	338/3777 (8,9%)	1036/1297 (78,1%)	261/1297 (21,9%)
PST <i>(São Tomé)</i>	2439/2612 (93,4%)	173/2612 (6,6%)	640/687 (93,1%)	47/687 (6,9%)
PE <i>(Cacém)</i>	2448/2449 (99,9%)	1/2449 (0,04%)	1499/1515 (98,9%)	17/1515 (1,1%)

Fonte: Elaboração própria.

Gráfico 1: Marcação/não marcação de pluralidade no âmbito do SN.

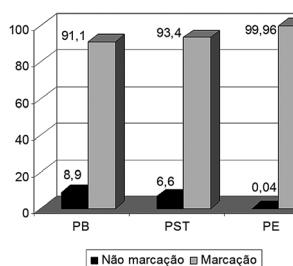
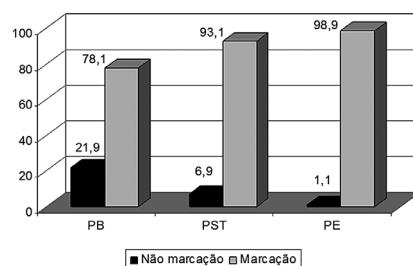


Gráfico 2: Marcação/não marcação de pluralidade no âmbito do SV.



Fonte: Elaboração própria.

Observando-se, apenas, o quantitativo empregado, chama à atenção, primeiramente, o que é comum às três variedades urbanas: a larga preferência pela concretização das marcas de número. Apesar dessa tendência geral, fica evidente, pelo que se verifica no gráfico, que do PB ao PE ocorre um crescimento dos índices da marca de plural, que, na última variedade, beira o valor categórico.

No PST, em posição intermediária, registra-se, em relação ao PE, maior índice de não marcação de pluralidade (6,6% e 6,9%, respectivamente, no âmbito do

SN e do SV). O menor índice de falta de concordância, nos dados são tomenses em relação aos do PB, talvez possa ser atribuído ao fato de a norma adotada oficialmente no PST ser a europeia e alguns membros da comunidade terem um contato mais estreito com os portugueses, tanto por conta de viagens a Portugal, quanto pela presença de professores portugueses nas poucas instituições de ensino, como é o caso do quadro docente do Liceu Português, escola privada que atende, em geral, aos que têm melhor situação sócio-económica. Os altos índices gerais de concordância, todavia, mascaram, de certa forma, uma situação que se supõe semelhante à descrita por Gonçalves (2000, p. 3), que, ao referir-se à variedade moçambicana, diz haver “um continuum polilectal, ao longo do qual se distribuem com maior ou menor frequência” diversos traços gramaticais, *continuum* esse “composto por diversas subvariedades que oscilam entre um extremo muito próximo do padrão europeu, e um extremo ‘basilectal’.”

O presente trabalho confirma as generalizações apresentadas em Lucchesi, Baxter e Ribeiro (2009, p.331), ao propor que a “variação da concordância no PE é na melhor das hipóteses residual e periférica.” Segundo o autor, situação distinta se verifica no PB, em que os índices variáveis de concordância poderiam ser dispostos num *continuum* de polarização sociolinguística, que iria da norma urbana culta (com 94% de concordância, como registrado em Graciosa (1991)) até as comunidades rurais afro-brasileiras (com apenas 16% de concordância, índice verificado por Lucchesi, Baxter e Ribeiro (2009)), *continuum* que evidenciaria o efeito do contato entre línguas. Nos dados brasileiros da presente pesquisa, adotando a proposta ora comentada, pode-se atribuir a alta concretização do plural (91,1% para a concordância nominal e 78,1% para a verbal) ao fato de se tratar de amostra da fala metropolitana do RJ, com acentuado prestígio dos padrões idealizados para os meios escolarizados.

Levando-se em conta exclusivamente os totais de realização de concordância *versus* não concordância no presente estudo, pode-se avaliar o estatuto de cada variedade analisada quanto ao fenômeno. Consoante Labov (2003), acredita-se que a marca de apenas 1% de uma das supostas variantes põe em risco a concepção de regra variável. Segundo o autor, o intervalo de 95 a 99% de concretização de dada forma alternante constituiria parâmetro para o estabelecimento de uma regra linguística semicategórica, e não efetivamente variável, regra que implicaria alternância de formas em 5 a 95% das ocorrências.

A distinção dos tipos de regras propostos pelo autor, longe de constituir mera diferenciação de natureza terminológica, permite observar se as realizações de determinado fenômeno linguístico se configuram mais como uma espécie de parâmetro gramatical,⁹ naturalmente caracterizador de determinada variedade

⁹ Foge ao escopo deste artigo qualquer compromisso em interpretar os fenômenos variáveis à luz do referencial da Teoria de princípios e parâmetros, do quadro gerativista de análise linguística; entretanto, entende-se

linguística / língua, do que como uma forma alternante, que figura, por sua própria natureza, tão disponível quanto outra que lhe é equivalente.

No caso da concordância, tanto a nominal quanto a verbal, embora as três variedades em estudo exibam forte preferência pela realização da marca de número, são os dados do PE (Cacém) aqueles que assumem comportamento que se poderia denominar de semicategórico no que se refere à concordância verbal (98,9%) e de categórico no que toca à concordância nominal (99,6%), tendo em vista que o único registro de ausência de marca em um SN pode ser caracterizado, como se indicará, adiante, como resultante de uma falha de planejamento discursivo. Diferentemente dessa realidade, encontram-se os dados coletados no PST e no PB, cujos índices são inferiores à marca de 95% de realização do plural.

Não obstante o baixo número de dados de não marcação, entende-se ser de fundamental importância, para a determinação do estatuto de cada variedade, a observação detalhada das ocorrências, para que se verifique se algumas variáveis apontadas como relevantes em diversos estudos sobre a concordância de número nominal e verbal de 3^a pessoa operam também no material estudado. Em outras palavras, além dos índices quantitativos, faz-se necessária uma cuidadosa análise qualitativa dos dados. No caso das variedades brasileira e africana do Português, foi possível fazer essa verificação com base no tratamento sociolinguístico provido pelo Goldvarb-X. Quanto à variedade europeia, o procedimento mais produtivo foi, sem dúvida, a observação cuidadosa das poucas ocorrências de constituintes sem a expressa marca de número.

Aspectos específicos

Da concordância nominal

No âmbito da concordância nominal, foram realizadas diferentes análises: (a) confrontaram-se o PB e o PST, levando em conta, por variedade, a fala de 18 indivíduos de três níveis de escolaridade distribuídos, ainda, por sexo e três faixas etárias; (b) focalizou-se apenas o PST (i) considerando, com a ampliação de 12 para 22 do número de informantes, apenas os indivíduos com níveis fundamental e médio de escolaridade e controlando uma nova variável: **frequência de uso de um crioulo**; (ii) avaliando a fala de nove estudantes entre 10 e 19 anos,¹⁰ considerando a série que estavam cursando e, ainda, a

que a tipologia de regras propostas em Labov (2003) seja correlacionável e possa oferecer contribuições aos propósitos daqueles que se ocupam da investigação paramétrica de origem formalista.

¹⁰ O *corpus* Tjerk é constituído de 111 entrevistas. Destas, apenas as nove utilizadas no referido estudo contemplam a fala de indivíduos entre 10 e 19 anos ainda em fase de escolarização.

variável **desempenho individual**; (c) observou-se o PE, com base em amostra eliciada da fala de indivíduos distribuídos segundo os parâmetros explicitados em (a). Os estudos descritos em (a) e (b) foram desenvolvidos com o auxílio do pacote de programas Goldvarb-X, tendo como unidade de análise cada constituinte flexionável do SN; o mencionado em (c) consistiu, com apoio em índices percentuais, na análise qualitativa das ocorrências de SNs tomados como unidade de análise.

Na tabela 2, a seguir, apresentam-se as variáveis selecionadas nas análises empreendidas com amostras do PB e do PST, chamando-se a atenção para alguns aspectos: (a) o baixo *input* da regra em todas as análises empreendidas (.004 a .039); (b) a atuação das variáveis **posição linear e relativa do constituinte no SN e nível de escolaridade**, que aparecem como as mais salientes para a compreensão do fenômeno (em primeiro ou segundo lugar) e **animacidade do núcleo e saliência fônica**, que, embora de forma não tão sistemática, se mostram também relevantes.

Tabela 2 - Variáveis selecionadas em diferentes análises sobre a ausência de marca de pluralidade no SN no PB e no PST.

(a) Confronto entre PB e PST	
PB	PST
Posição linear e relativa do constituinte no SN	Nível de escolaridade
Nível de escolaridade	Posição linear e relativa do constituinte no SN
Faixa etária	Animacidade do núcleo
Saliência fônica	Gênero
Animacidade do núcleo	Contexto fonológico subsequente
Ausência de marca: 338/3777 (8,9%)	Ausência de marca: 173/2612 (6,6%)
Significância: .001	Significância: .023
Input: .039	Input: .016

(b) Análises sobre o PST	
(i) Indivíduos de níveis médio e fundamental	(ii) Estudantes de 10 a 19 anos
Posição linear e relativa do constituinte no SN	Desempenho individual
Nível de escolaridade	Posição linear e relativa do constituinte no SN
Frequência de uso de um crioulo	
Saliência fônica	
Animacidade do núcleo	
Ausência de marca: 305/2375 (12,8%)	Ausência de marca: 31/633 (4,9%)
Significância: .000	Significância: .000
Input: .004	Input: .005

Fonte: Elaboração própria.

No que toca à variável **Posição linear e relativa do constituinte no SN**, como se registra na tabela 3, não obstante os baixos percentuais de não marcação da pluralidade, pode-se depreender, também na fala urbana do PB e do PST, uma restrição indicada como muito atuante em variedades não urbanas: a propensão à ausência da marca de número a partir do núcleo em segunda posição. Pode-se observar que, em quase todas as análises feitas, é a partir dessa posição que se registram pesos relativos acima de .50, indicando o favorecimento da não marcação.

Tabela 3 - Atuação da variável Posição linear e relativa no SN para ausência de marca de número, em diferentes análises com amostras do PB e do PST.

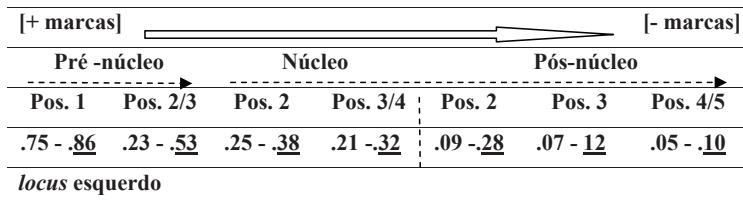
Análise/ Amostra	Pré-nuclear			Nuclear			Pós-nuclear	
	1	2/3	1	2	3/4	2	3	4/5
PB	19/1450	6/155	6/155	238/1628	18/194	9/95	34/109	10/35
	1,3%	3,6%	3,6%	14%	5,1%	9,5%	31,2%	28,6
	.17	.47	.40	.74	.68	.72	.90	.89
(a)	7/1022	3/122	3/67	121/1120	14/144	10/58	10/60	5/19
PST	1%	2%	4%	10%	9,7%	17,2%	10/60%	26,3
	.15	.50	.57	.74	.76	.91	.88	.94
(b)	9/909	6/14	9/909	228/1092	25/122	11/44	16/44	8/15
(i)	1%	5,8%	4,4%	20,9%	20,5%	25%	36,4%	53,3%
	.14	.55	.36	.75	.76	.88	.93	.95
S	4/262	2/26	0/12	16/268	3/35		6/30	
T (ii)	1,5%	7,7%	0%	6%	8,6%		20%	
	.25	.77	---	.62	.79		.90	

Fonte: Elaboração própria.

Invertendo-se os valores dos pesos relativos – de ausência para presença de marcas – registrados na tabela acima, com o objetivo de melhor caracterizar as áreas do SN em que se concentra a tendência à marcação da pluralidade, chega-se ao *continuum* proposto a seguir, cuja validade poderia ser testada em outras variedades urbanas do Português. Leva-se em conta o que aqui se denomina de SN prototípico, isto é, aquele composto por constituinte(s) pré-nuclear(es), núcleo e constituinte(s) pós-nuclear(es). Em cada posição, indicam-se os PRs mínimos e máximos de presença da marca de plural. Assim, a primeira posição linear, no pré-núcleo, tende a ser mais marcada, o que sugere seja esse o *locus* por exceléncia da marca. A partir do núcleo em segunda posição, vai decrescendo gradativamente a presença da marca, quer se considerem os menores ou os maiores índices obtidos (em destaque por meio de sublinhado). No esquema proposto, a linha vertical contínua que separa o pré-núcleo do núcleo representaria o limite do espaço mais saliente

para a marcação (o *locus* esquerdo), enquanto a linha vertical pontilhada entre o núcleo e o pós-núcleo indicaria que essas áreas não sofreriam, quanto à tendência à não marcação, solução de continuidade.

Tabela 4 - Continuum de marcação de pluralidade em SNs prototípicos no PB e no PST, com base em PRs obtidos nas diferentes análises realizadas.



Fonte: Elaboração própria.

SNs em que um constituinte que normalmente ocupa posição à esquerda ou à direita muda de lugar podem comportar-se, quanto à concordância, segundo o *continuum* acima proposto: mais marcas à esquerda (*locus* esquerdo), menos à direita, como se observa nos exemplos em (a) contrapostos aos transcritos em (b):

- 1- (a) todos os seus aluno
(b) aquelas coisa toda
- 2- (a) ter condições financeira boa
(b) ter boas ideia

A estrutura em 2(a) exemplifica, ainda, o que está patente na Tabela X: núcleos em primeira posição tendem a receber a marca (PRs. .64 - 1.0, com a inversão dos valores), diferentemente do que ocorre quando em segunda (PRs. .25 - .38) ou terceira/quarta posição (PRs. .21 - .32).

Ainda do ponto de vista estrutural, mencionem-se as variáveis **animacidade do núcleo** e **saliência fônica**, também selecionadas para o PB e o PST (em algumas das análises) e consideradas em conjunto na tabela 5.

Constata-se que núcleos com traço semântico [+animado] (PRs variando de .39 a .43), embora com índices próximos do obtido para os que apresentam traço [-animado] (PR. .55 em todas as análises), inibem a não marcação. Quanto à saliência fônica são os constituintes com menor diferenciação entre a forma singular e plural (PRs. .52 e .54) os que mais desfavorecem a marca.

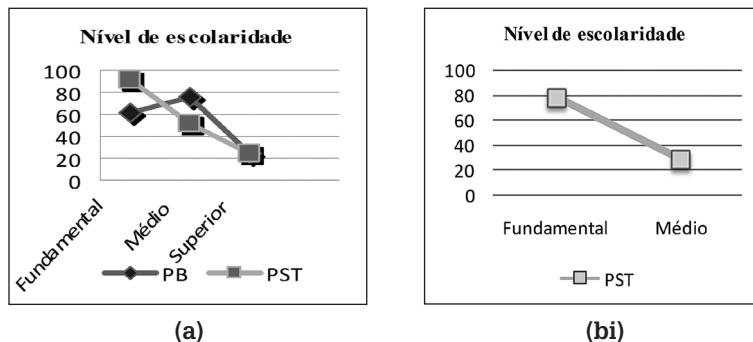
Tabela 5 - Atuação das variáveis animacidade do núcleo e saliência fônica para a não marcação de número no SN em diferentes análises com amostras do PB e do PST.

Análise/Amostra	Animacidade do núcleo		Saliência fônica	
	[+ animado]	[-animado]	[+saliente]	[-saliente]
(a) PB	95/838 11,4% .43	166/1095 15,2% .55	25/374 6,7% .30	313/3403 9,2% .52
	36/489 7% .40	102/842 12% .55	— ¹⁰	—
	PST			
	(i) 59/433 13,6% .39	195/826 23,6% .55	21/218 9,6% .25	268/1.182 22,7% .54
(b) PST	(ii) —	—	—	—

Fonte: Elaboração própria.¹¹

Cabe, agora, discutir a variável **nível de escolaridade**, que parece desempenhar papel preponderante para a implementação e a produtividade dos padrões variáveis de concordância comuns ao PB e ao PST.

Gráfico 3 - Atuação da variável nível de escolaridade para a não marcação de número no SN em diferentes análises sobre o PB e o PST (pesos relativos).



Fonte: Elaboração própria.

Como se salientou inicialmente, os altos índices globais de concordância obtidos nas análises das duas variedades não europeias não deixam transparecer,

¹¹ O traço apresentado nas tabelas indica que a variável não foi selecionada na análise em questão.

de imediato, a existência de um *continuum* em cujos pólos se encontram os padrões usados pelos indivíduos menos escolarizados e os utilizados pelos de nível superior. Em estudo de natureza geolinguística, em que comentava os três *continua* – de urbanização, de oralidade-letramento, de monitoração estilística – segundo os quais Bortoni-Ricardo (2004, p.51-60) sugere que se estude a variação no PB, Brandão (2011) diz ser necessário ter em conta “[...] un quatrième continuum - le niveau d'instruction - non seulement parce qu'il s'agit d'un paramètre fondamental pour représenter les distances sociales au Brésil, mais aussi pour se rattacher étroitement et parfois se superposer au continuum d'urbanisation [...]”, que se apresenta mais adiante, adaptado ao que aqui se discute.

O gráfico 3(a), que retrata os resultados referentes ao confronto entre o PB e o PST realizado com base na fala de indivíduos de níveis fundamental, médio e superior de instrução, mostra que, tanto no PB quanto no PST, há, em relação à concordância nominal, uma situação sociolinguisticamente polarizada, reforçada pelo fato de o nível de escolaridade, em geral, se associar a outras variáveis, como nível socioeconômico, maior ou menor acesso a bens culturais, maior ou menor contato com indivíduos de diferentes estratos sociais.

Quadro 2 - Continuum de nível de instrução, segundo Brandão (2011), adaptado à ausência da marca de concordância nominal no PB e no PST.

Continuum de nível de instrução			
[- marcas] [+ marcas]	Fundamental	Médio	Superior
Não escolarizados/ Escolaridade: 1 a 4 anos	PB: (.62-.76)	PB: .24	
	PST: .90 / .80	PST: .53 / .29	PST: .15 / --

Fonte: Elaboração própria.

Observe-se, no entanto, que, no âmbito da variedade brasileira, que já dispõe de uma norma consolidada, os indivíduos de níveis fundamental e médio se opõem aos de nível superior, enquanto na variedade santomense o caráter francamente escalar de adoção da norma canônica de concordância fica patente.

A análise do PST em que se confrontaram apenas indivíduos de nível médio e fundamental (bi) e aquela em que se focalizou a fala de estudantes de 10 a 19 anos (bii) dão conta da complexidade de fatores que concorrem para a variação.

Na primeira delas, a variável **frequência de uso de um crioulo** foi a terceira mais relevante para a não aplicação da marca, suplantada apenas pelas variáveis referentes à posição dos constituintes e à escolaridade. Embora todos os indivíduos que concorreram para a formação da amostra tenham se declarado falantes de

português como L1, tem-se de levar em consideração o caráter multilíngue da sociedade santomense, em que coexistem três crioulos, além de outras línguas minoritárias: há indivíduos que (a) só usam o português e (i) não dominam ou (ii) dominam um crioulo; (b) usam o português e um ou mais crioulos (i) com menor ou (ii) maior frequência (no mercado, em momentos de lazer com amigos), e assim por diante. O uso do crioulo tem caráter estigmatizante, haja vista que muitos dos informantes mencionaram o fato de seus pais obrigar-los a falar, mesmo em casa, apenas o português, para que, dominando as normas de prestígio, se tomassem aptos a aproveitar as poucas oportunidades de ascensão socioeconômica.

Dante desse quadro, buscou-se verificar se os indivíduos que fazem uso mais frequente de um crioulo (em geral, o forro) que do português teriam maior probabilidade de não usarem a marca de plural. A variável ficou assim definida, quanto à frequência de uso de um crioulo:

- **zero/baixa**, os indivíduos que se expressam fundamentalmente em português;
- **média**, os indivíduos que se expressam em português, mas dominam um crioulo e dele fazem uso eventualmente;
- **alta**, os indivíduos que, embora falem o português, se expressam, regularmente, num crioulo.

Na tabela 6, expõem-se os resultados referentes à atuação da variável.

Tabela 6 - Atuação da variável *frequência de uso de um crioulo* para o cancelamento da marca de número no SN.

Frequência de uso de um crioulo	Oco	Perc.	P.R.
Zero /Baixa (grau zero - G0)	75/1039	7.2%	.41
Média (grau 1 - G1)	140/1117	12.5%	.52
Alta (grau 2 - G2)	90/219	41.1%	.75

Fonte: Elaboração própria.

A classificação dos informantes quanto aos fatores da variável teve por base informações prestadas por eles durante as entrevistas. No primeiro fator, inscrevem-se nove informantes; no segundo, dez e, no terceiro, três. Independentemente dessa distribuição assimétrica, constata-se que são os indivíduos que se comunicam preferencialmente num crioulo os que mais tendem a não implementar a marca de número (41.1%; P.R. .75). Com esses indivíduos, contrastam os que nunca ou pouco se expressam em crioulo (P.R. .41) e os que o utilizam apenas eventualmente (P.R. .52).

A análise variaçãoista sobre a fala de estudantes santomenses, embora não se tenha realizado nos moldes clássicos – uma vez que se baseou na fala de apenas

nove informantes e não houve o preenchimento de todas as células sociais –, indicou a variável **desempenho individual** como a mais relevante. Na tabela a seguir, sintetizam-se os resultados.

Tabela 7 - Atuação da variável *desempenho individual* para a ausência de marca de plural em constituintes do SN, com base em amostra da fala de estudantes de São Tomé.

Informante	<i>Variável desempenho individual</i>			Informante	Oco	%	P.R.
	Oco	%	P.R.				
ST-E1-E6m	8/19	42.1%	.91	ST-E6-FDh	8/88	9.1%	.48
ST-E2E6m	0/28	0%	---	ST-E7-FDh	3/38	7.9%	.38
ST-E3-F8h	0/62	0%	---	ST-E8-FDh	3/83	3.6%	.26
ST-E4-F8m	0/261	0%	---	ST-E9-FDm	0/91	0%	---
ST-E5-F8m	9/26	34.6%	.91				
Input: .05				Significância: .000			

Fonte: Elaboração própria.

Os resultados sugerem que a concordância nominal tem, na sociedade santomense, um forte componente sócio-econômico-cultural. Independentemente do nível escolar em que se encontrem, enquanto, na fala de quatro estudantes, a regra tem caráter categórico, na de cinco outros, tem caráter variável, em maior ou menor grau. Tal quadro, certamente, vincula-se a aspectos que dizem respeito ao ambiente familiar, à maior ou menor exposição aos bens culturais, às línguas faladas na região, ao tipo de escola que frequentam. Vale referir as observações de dois dos estudantes que aplicam categoricamente a regra: um afirmou que o pai lhe propicia todos os meios para seu desenvolvimento intelectual; outro, que os pais preferem que estude na Escola Portuguesa por acharem que aquela escola, diferentemente das públicas, funciona de forma mais regular e, a princípio, propiciaria melhor qualidade de ensino.

Na fala dos estudantes que aplicam variavelmente a regra de concordância, as restrições que regulam a marcação ou não da pluralidade dizem respeito à **posição linear e relativa dos constituintes**, nos moldes aqui já discutidos.

Contrasta fortemente com o quadro descrito para o PB e o PST a variedade europeia, em que a regra de concordância nominal parece ter caráter categórico. A análise de 2.449 SNs, a grande maioria (79,40%) de dois constituintes, presentes na fala de 18 indivíduos de ambos os性os e diferentes faixas etárias e níveis de escolaridade, revelou uma taxa de 99,96% de aplicação da marca de número. O índice de 0,04% de ausência de marca corresponde a um único SN, constituído

por um numeral e um substantivo, que foi produzido por um homem da faixa etária mais alta e com nível fundamental de escolaridade.

Como se poderá aquilatar pelo contexto que se transcreve abaixo, o cancelamento da marca parece ter ocorrido por uma falha de planejamento discursivo, um lapso, como indica Labov (2003) quando explicita o que seria uma regra categórica em oposição a regras semicategóricas e variáveis. O indivíduo organiza seu discurso no sentido de mencionar apenas uma praça, “**uma** pracetazinha”, como ele a caracteriza, continua e reitera que é **uma** e, ao se dar conta de que são **duas**, parece não ter organizado integralmente o enunciado, deixando, assim, de produzir a marca no substantivo.

[...] nada...ali nada tem **uma** pracetazinha...há **uma** ou **duas** **praceta** onde os miúdos se entretêm onde há menos carros pra jogar um bocadinho à bola ou andarem de bicicleta de resto não tem lá nada... aquela zona é mesmo mesmo mesmo sem nada. (Informante CACC1H).

Da concordância verbal

De forma semelhante ao que foi descrito no âmbito da concordância nominal, foram realizadas diferentes análises para o estudo da concordância verbal: (a) confrontaram-se o PB e o PST, levando em conta, em cada localidade, a fala de 18 indivíduos¹² de três níveis de escolaridade distribuídos, ainda, por sexo e três faixas etárias; (b) focalizou-se apenas o PST¹³ considerando a ampliação do número de informantes, de modo que houve maior número de indivíduos com níveis fundamental e médio de escolaridade (22 informantes), e controlando uma nova variável: **frequência de uso de um crioulo**; (c) observou-se o PE, também com base em amostra eliciada da fala de indivíduos distribuídos segundo os parâmetros explicitados em (a). As análises descritas em (a) e (b) também foram desenvolvidas com o auxílio do pacote de programas Goldvarb-X; por impossibilidade decorrente da falta de contextos variáveis, a mencionada em (c) baseou-se fundamentalmente na análise qualitativa dos dados encontrados.

Na tabela 8, a seguir, apresentam-se as variáveis selecionadas nas análises empreendidas com amostras do PB e do PST:

¹² Conforme já se esclareceu, apenas para o PST, o estudo só pôde contar com 17 informantes, o que resulta da dificuldade de localizar falantes de Português com curso superior.

¹³ Diferentemente da análise da concordância nominal, o estudo da concordância verbal também considerou nessa etapa da investigação (b) os informantes de curso superior; além disso, até o momento da pesquisa, não foi realizada a análise apenas com informantes de 10 a 19 anos.

Tabela 8 - Variáveis selecionadas em diferentes análises sobre a ausência de marca de pluralidade no verbo no PB e no PST.

(a) Confronto entre PB e PST	
PB	PST
Paralelismo no nível oracional	Nível de escolaridade
Nível de escolaridade	Sexo
Saliência fônica	Saliência fônica
Animacidade do sujeito	Animacidade do sujeito
Posição do sujeito em relação ao verbo	Faixa etária
Faixa etária	Posição do sujeito em relação ao verbo
Ausência de marca: 338/3777 (8,9%)	Ausência de marca: 47/687 (6,8%)
Significância: .004	Significância: 013
Input: .014	Input: .016

(b) Análise adicional sobre o PST	
Ampliação do número de indivíduos de níveis médio e fundamental + superior	
Nível de escolaridade	
Animacidade do sujeito	
Posição do sujeito em relação ao verbo	
Sexo	
Paralelismo no nível oracional	

Ausência de marca: 100/1053 (9,5%)
 Significância: .036
 Input: .049

Fonte: Elaboração própria.

No âmbito da concordância verbal, também chama a atenção o baixo *input* da regra em todas as análises empreendidas (.004 a .049), confirmando ser a concordância a estratégia preferencial nos dados coletados no Brasil e em São Tomé. Quanto às variáveis relevantes à compreensão do funcionamento dos dados, destaca-se a atuação da variável extralingüística **nível de escolaridade**, cuja descrição permitirá a comparação com os resultados relativos à concordância nominal. As variáveis estruturais **animacidade do núcleo** e **posição do sujeito em relação ao verbo**, em primeiro plano, e **saliência fônica**, em segundo, também se mostraram relevantes em relação ao fenômeno.

A tabela 9, a seguir, permite observar o comportamento das variáveis de cunho estrutural, considerando as análises desenvolvidas.

Tabela 9 - Atuação das variáveis *animacidade do núcleo* e *posição do sujeito em relação ao verbo* para a não marcação de número no verbo em diferentes análises com amostras do PB e do PST.

Análise/ Amostra	Animacidade do núcleo		Posição do sujeito em relação ao verbo ¹³	
	[+ animado]	[-animado]	[anteposto]	[posposto]
(a)	PB	206/1133	55/164	143/947
		18,2%	33,5%	15,1%
		.46	.70	.41
	PST	33/585	14/102	27/467
		5,6%	13,7%	5,8%
		.45	.71	.31
(b)	PST	63/879	37/174	55/688
		7,1%	21,2%	7,9%
		.43	.78	.46
				.86

Fonte: Elaboração própria.¹⁴

No que se refere ao traço semântico de animacidade, fica evidente que, nas três análises, a presença de referente [+animado] no SN sujeito (PRs variando de .43 a .45) desfavorece a não marcação, ao contrário do que ocorre com o traço [-animado], que nitidamente a favorece (PRs variando de .70 a .78).

Quanto à posição do sujeito em relação ao verbo, também fica comprovada a conhecida tendência: sujeitos pospostos, ao contrário dos antepostos, favorecem a ausência de marcação plural (PRs variando de .77 a .86).

A variável saliência fônica – que não se mostrou estatisticamente relevante apenas no caso da análise adicional realizada para o PST¹⁵ – foi controlada, para os dados de concordância verbal, consoante quatro graus de diferenciação das formas singular e plural dos verbos, conforme se expõe na tabela a seguir.

¹⁴ No controle desta variável, foram desconsideradas as ocorrências com sujeito não expresso.

¹⁵ Embora o programa Goldvarb-X não tenha selecionado a variável saliência fônica na amostra em questão, ficou claro, pela observação dos dados e pelo cruzamento entre grupos de fatores, que a atuação do referido grupo estava associado ao tipo de verbo em pauta quanto à transitividade verbal, o que está sendo apurado com mais detalhamento na presente etapa da investigação.

Tabela 10 - Atuação da variável *saliência fônica* para o cancelamento da marca de número no verbo na análise das amostras do PB e do PST.

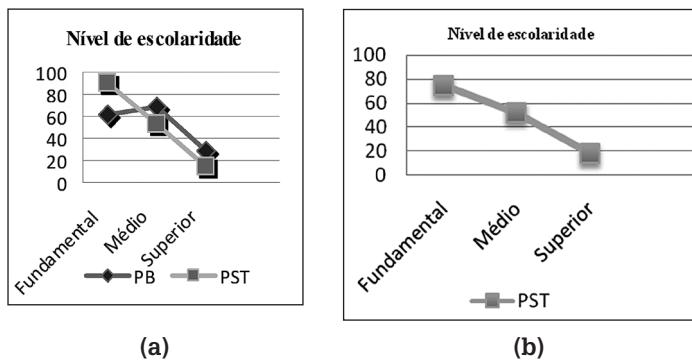
Análise/ Amostra	Saliência fônica				
	Grau 1 <i>come-comem; fala- falam</i>	Grau 2 <i>dá-dão; vai- vão</i>	Grau 3 <i>comeu-comeram; quis- quiseram</i>	Grau 4 <i>veio-vieram; é- são</i>	
(a)	PB	187/783 23,9% .60	46/226 20,4% .46	15/161 9,3% .25	13/127 10,2% .26
	PST	38/450 8,4% .63	2/79 2,5% .17	4/93 4,3% .37	3/65 4,6% .26
	(b) PST	—	—	—	—

Fonte: Elaboração própria.

Nas duas variedades do Português, constata-se maior tendência ao cancelamento da marca de número em verbos que apresentam pouca diferenciação entre as formas singular e plural (.60 no PB e .63 no PST) em relação aos demais (PRs variando entre .25 e .46 no PB e entre .17 e .37 no PST). Na amostra do PB, registram-se índices decrescentes de cancelamento da marca de número consoante o aumento da diferenciação fônica entre as formas verbais. Na amostra do PST, embora não se verifiquem tais índices progressivamente decrescentes, o efeito da **saliência fônica** pode ser aferido do contraste entre os índices obtidos para o grau 1 – que favorecem o cancelamento da marca de número – e os obtidos para os demais graus – que a desfavorecem.

Como se pode observar, fica patente a similaridade na atuação das variáveis estruturais. Resta, então, descrever o comportamento dos dados quanto ao grupo de fatores de cunho extralingüístico, que, como já se observou, desempenha papel preponderante para a implementação e a produtividade dos padrões variáveis de concordância comuns ao PB e ao PST, a variável **nível de escolaridade**. As figuras a seguir permitem observar, agora, o comportamento dos dados relativos à concordância verbal.

Gráfico 4 - Atuação da variável nível de escolaridade para a não marcação de número no verbo em diferentes análises sobre o PB e o PST (pesos relativos).



Fonte: Elaboração própria.

Os gráficos 4(a) e (b) demonstram que, também no âmbito da concordância verbal, tanto no PB quanto no PST, a concretização das marcas de número se associa fortemente ao nível de escolaridade do indivíduo, o que, observados os extremos, configura uma situação sociolinguisticamente polarizada. Na variedade brasileira, os indivíduos de níveis fundamental e médio (com índices acima de .50, indicativos de favorecimento da ausência de marca) opõem-se aos de nível superior (com nítido desfavorecimento da ausência de marca), enquanto na variedade santomense fica nítido que a adoção da norma canônica de concordância é progressiva. Assim, como demonstra o quadro 3, a seguir, também no eixo da concordância verbal, podem-se representar os *continua* de nível de instrução, tais como se configuraram no PB e no PST.

Quadro 3 - Continuum de nível de instrução, segundo Brandão (2011), adaptado à ausência da marca de concordância verbal no PB e no PST.

Continuum de nível de instrução			
[- marcas]		[+ marcas]	
Não escolarizados/ Escolaridade: 1 a 4 anos	Fundamental	Médio	Superior
	PB (.61-.69)	PB: .29	
	PST: .90 / .75	PST: .53 / .52	PST: .15 / .18

Fonte: Elaboração própria.

A análise adicional do PST desenvolvida com os dados referentes à concordância verbal em que se utilizou maior número de indivíduos com nível médio e fundamental, além daqueles que cursaram o nível superior, também permitiu observar a atuação da variável **frequência de uso de um crioulo**.

Embora esse grupo de fatores não tenha sido selecionado pelo Programa na análise estatística dos dados, o cruzamento entre ele e a variável **nível de escolaridade** foi esclarecedor.

Tabela 11 - Cruzamento entre as variáveis nível de escolaridade e frequência de uso de um crioulo para o cancelamento da marca de número no verbo.

Frequência de uso de um crioulo	Ensino Fundamental	Ensino Médio	Ensino superior
Zero /Baixa (grau zero - G0)	14/105 13%	17/195 9%	4/142 3%
Média (grau 1 - G1)	26/154 17%	18/211 9%	3/157 2%
Alta (grau 2 - G2)	18/89 20%	-----	-----

Fonte: Elaboração própria.

Embora seja, de fato, a escolaridade o fator que faz mudar as tendências relativas ao fenômeno, a tabela 11 permite visualizar que, em paralelo ao acesso à escolarização, está a maior ou menor utilização da língua crioula. Só declararam que usam frequentemente o crioulo informantes que cursaram até o ensino fundamental, nível de escolaridade em que se registra maior cancelamento da marca de número. De fato, o prestígio que recebe o Português em terras santomenses e o maior ou menor propósito de dominar essa língua, o que pode se tornar possível graças ao acesso à escolaridade, trazem por consequência o maior ou menor domínio das regras canônicas de concordância.

Também no caso da concordância verbal, é absolutamente distinto o comportamento dos dados da variedade europeia, representada aqui pela comunidade de Cacém,¹⁶ em comparação ao descrito para o PB e o PST: no PE, a regra de concordância verbal assume um perfil de caráter semicategórico.

Das 1.515 ocorrências de construções de 3^a pessoa do plural, produzidas pelos dezoito informantes de ambos os性os e diferentes faixas etárias e níveis de escolaridade, apenas dezessete não apresentaram marca de concordância verbal, o que corresponde a apenas 1,1%. Em termos extralingüísticos, cabe ainda informar que esses dados foram produzidos por apenas dez dos dezoito informantes considerados; em outras palavras, na fala de oito informantes, as marcas verbais plurais ocorreram em todos os dados, de forma categórica.

¹⁶ Já se encontram tratados os dados coletados em Oeiras/Lisboa, os quais não apresentaram comportamento substancialmente diferente do ora descrito, nem quantitativa nem qualitativamente. Para maiores detalhes acerca dos dados dessa localidade, consultar Vieira (2011). A fim de manter a comparabilidade dos resultados com os exibidos para a concordância nominal, consideram-se aqui exclusivamente as ocorrências coletadas na localidade de Cacém.

Tendo em vista que só se registraram dezessete casos de suposta não concordância na amostra do PE, passa-se a apresentar os tipos de estruturas que representam esses dados. A observação das dezessete ocorrências, uma a uma, permitiu detectar os padrões estruturais encontrados. Sistematizando os resultados dessa observação, procedeu-se à distribuição das ocorrências pelos seguintes contextos:

(i) sete dados de estruturas com o verbo “ser” em construções predicativas (exemplos 1 e 2):

(1) D: e: os seus amigos? Que interesse é que têm? Têm os mesmos interesses que o senhor o? L: a/a: os meus amigos meus são de engenharia digamos assim muitos deles **é** de engenharia há muitos de economia [...] (CAC1H)

(2) as melhorias **seria** eu em vez de ganhar quinhentos é assim não vamos ganhar quinhentos vamos ganhar só quatrocentos (CACC1H)

Cabe lembrar que construções do tipo SN sujeito + verbo *ser* + predicativo do sujeito costumam ser caracterizadas de forma particular, especialmente no que tange à concordância. Por vezes, admite-se, desde a tradição gramatical, que a relação de concordância se dê entre o verbo e o constituinte considerado predicativo. Além disso, essas construções podem ser consideradas estruturas “espelhadas”, de modo que as posições de sujeito e predicativo se invertem facilmente (como em “a mudança é a solução” ou “a solução é a mudança”).

(ii) cinco dados de posposição do sujeito (exemplos 3 e 4):

(3) D: há o comboio ali da linha de Sintra
L: sim sim sim passa mesmo no centro do Cacém e depois a partir da estação **existe**:: ... os autocarros para os diversos sítios (CAC3M)

(4) L: tal e qual eles passam a vida a ver aquelas anúncios...então nas épocas de Natal aquelas tretas todas aquilo é horrível mas é assim eles não têm :m não sei **escapa**-lhes as coisas não é? porque eu acho que com a idade deles apesar deles serem crianças eu já tinha outra maturidade pra ver que aquilo não era possível...(CACB3M)

As construções com sujeito posposto constituem, na realidade, um dos poucos contextos estruturais, amplamente debatido na literatura, em que as três variedades igualmente registram casos de cancelamento da marca de número plural. Há que se questionar o estatuto dessas construções quanto à identidade do sujeito. Sabe-se que, além da questão do processamento da informação, as construções com sujeitos pospostos são, em sua maioria, integrantes de categorias especiais de verbos, como os intransitivos e os inacusativos.

(iii) três dados em que o suposto referente sujeito não se encontra adjacente ao verbo (exemplos 5 e 6).

(5) D: não há o mínimo investimento?

L: nada prali não há nada nada e podiam fazer () que aquilo é uma zona onde esta zona aqui é onde eu moro pertence ao Cacém e esta que (é a rua) esta... os prédios daquele lado na mesma rua... já **pertence** a Rio de mouro...

(6) – oh minha menina eu onde eu trabalho já fomos assaltados duas vezes... duas vezes foram esses imigrantes que por aí entram aí não têm nada que fazer vêm para cá.. só **estraga**... e prontos foi uma experiência que não quero repetir espero que não volte a acontecer (CACB1H).

Nesses exemplos, cabe observar que a própria identificação do sujeito pode ser questionada: ser interpretado como nulo com um referente singular. No exemplo 6, é possível propor que “estragar” se relaciona a toda a situação descrita anteriormente (“isso tudo”, “essa situação”) e não especificamente aos “imigrantes” de que se falava.

(iv) dois dados em que o referente-sujeito foi retomado por pronome-cópia:

(7) L: muito alta é muito diferente de todas as outras brasileiras que eu conheço... é um estereótipo completamente diferente de todas as outras brasileiras que eu conheço... já as minhas vizinhas são todas brasileiras já são dentro do estereótipo de brasileiras que eu conheço... assim mais com um cabelo encaracolado uma pele mais morena tem umas feições que **identifica** logo que elas são brasileiras...agora se ela fala a gente vê logo que ela não é de cá assim (CACAC1M)

(8) é por quem lá tá dentro...não vão fazer leis a favorecer o mais pequeno... todas as leis que **é** feito a nível de isto ou de aquilo e de acoloutro é tudo leis (CACC1H)

A presença do pronome relativo *que* nesses exemplos constitui, sem dúvida, fator importante para a realização da forma verbal singular.

(v) um dado de infinitivo pessoal (exemplo 9):

(9) L: talvez pelo tal... por não haver tanta proximidade com os pais...no contacto...e pronto e se calhar o :o...a pô/a presença dos pais não ser tão efectiva e acaba por eles se **tornar** tão autónomos que quando alguém pede/ diz alguma coisa eles começam logo a disparatar (CACB3H)

A forma infinitiva constitui por si só um contexto absolutamente particular, motivo pelo qual muitos estudos variacionistas nem sequer a consideram na contagem dos dados.

Levando em conta o número de dados e os contextos em que se pôde comprovar efetivamente a ausência da marca plural no verbo, ao menos duas conclusões podem ser tiradas:

- (i) em termos quantitativos, a suposta falta de concordância no Português Europeu lisboeta configura-se um caso de regra semicategórica (LABOV, 2003);
- (ii) em termos qualitativos, a suposta falta de concordância no PE lisboeta está concentrada em contextos linguísticos específicos e não ocorre na diversidade de contextos já verificados no PB e no PST.

Reflexões sobre os resultados relativos à concordância nominal e à verbal: o estatuto das variedades do Português

Desenvolvidas as análises variacionistas sobre concordância nominal e verbal com dados das variedades brasileira, são-tomense e europeia, foi possível delinear o quadro a seguir, em que se observam aspectos divergentes e convergentes:

- (i) parece haver, em termos percentuais gerais, tendência à adoção da regra de concordância tanto no SN quanto no SV, embora com estatuto diferenciado, nas diferentes áreas urbanas aqui consideradas:
 - (a) na do Português Europeu, os dados revelam um comportamento que não pode ser considerado efetivamente variável – trata-se, nos termos de Labov (2003), de uma regra categórica no âmbito da concordância nominal e semicategórica no âmbito da verbal;
 - (b) nas do Português do Brasil e do Português de São Tomé, embora se dê preferência à concordância – preferência compatível com meios em que o acesso à escola, em maior ou menor grau, é facultado aos indivíduos –, ambas as regras são variáveis, salientando-se que o PST, que apresenta maiores índices gerais de marcação de pluralidade do que o PB, é a variedade em que o *continuum* de nível de instrução tem caráter nitidamente escalar.
- (ii) variáveis de cunho estrutural nos níveis sintático (**posição linear e relativa dos constituintes do SN; posição do sujeito em relação ao verbo**), semântico (**animacidade**) e fonético (**saliência fônica**) atuam como restrições na concretização das marcas de pluralidade, comprovadamente no PB e no PST, de acordo com as seguintes tendências:

- (a) nas estruturas sintáticas em que cabe indicar pluralidade por meio de flexão, a sua área esquerda é a que mais predispõe à presença de marca:
 - (i) nos constituintes pré-nucleares, quando se trata do SN considerado isoladamente, e no constituinte nuclear, quando ele se encontra em primeira posição;
 - (ii) no verbo, quando o SN sujeito está à sua esquerda.
 - (b) o traço [+ animado] ou [-animado] do núcleo do SN repercute, respectivamente, na presença ou ausência da marca de número tanto no âmbito do SN, quanto no do SV;
 - (c) (c) menor ou maior grau de saliência fônica condiciona, inversamente, maior ou menor índice de marcação da pluralidade nas formas nominais e verbais.
- (iii) o **nível de escolaridade** dos indivíduos atua como variável relevante no PB e no PST, motivação que, por ora, não pode ser atestada no PE.

Essas conclusões advindas das análises sugerem que, em dados contemporâneos, o estatuto da concordância no PE é de natureza absolutamente diferente do assumido nas variedades não europeias do Português ora estudadas. No que se refere à quantidade dos dados, não se detectou variação no âmbito da concordância nominal e registraram-se apenas 17 ocorrências de formas verbais no singular em construções em que se poderia propor a existência de SN sujeito plural.

Quanto à qualidade das construções analisadas, nas poucas estruturas relacionadas à concordância verbal, aventa-se a hipótese de que construções específicas em que não se dá a marca de plural também ocorrem no PE. Considerando apenas os resultados da presente investigação, pode-se supor que a posição do sujeito, a distância do constituinte sujeito ou, ao menos, do enunciado que evoca o referente sujeito, especialmente quando retomado pelo pronome relativo *que*, constituiriam evidências de que alguns parâmetros atuantes no PB e/ou no PST agiriam no PE.

Estudos científicos (NARO; SCHERRE, 2007; VAREJÃO, 2006; MONGUILHOT, 2009) declaram ter comprovado em dados europeus a atuação de outras variáveis, como, por exemplo, a saliência fônica e a animacidade do sujeito. Com base na observação de alguns exemplos e na atuação dessas variáveis, Naro e Scherre (2007, p.114-115) chegam a propor que “[...] o português brasileiro e o português europeu partilham estruturas semelhantes e que as estruturas variáveis de uso comum no Brasil também podem ser encontradas em Portugal, com menos intensidade, tanto hoje quanto antes da colonização.”

Embora não constitua objetivo do presente artigo, é fundamental que se comparem os dados da presente pesquisa com os de outras investigações sociolinguísticas com amostras do PE, sobretudo no âmbito do sintagma verbal.

Trabalhos como os de Varejão (2006) e Monguilhot (2010) registram percentuais gerais de não marcação mais altos do que os verificados nesta pesquisa (cerca de 10% em amostra de indivíduos não escolarizados do *corpus* Cordial-Sin, no primeiro caso, e cerca de 8%, no segundo, considerando amostra com indivíduos com pelo menos o nível fundamental de escolaridade), percentuais que permitiriam postular a existência de uma regra variável em determinadas variedades do PE. Por ora, uma observação preliminar desses trabalhos sugere que tenha havido diferença tanto na coleta de dados considerados variáveis, quanto na interpretação de determinadas estruturas como reveladoras de falta de concordância.

Nesse sentido, considera-se apropriada a hipótese de Naro e Scherre (2007, p.85) relativa à possibilidade de haver, mesmo no Português Europeu contemporâneo, padrões em diversas regiões de Portugal: “[...] são desconhecidas a verdadeira extensão e a intensidade da variação na concordância em terras lusitanas.” É preciso ter claro, por exemplo, que não se dispõe de trabalhos variacionistas baseados em amostras do PE falado por informantes não escolarizados.¹⁷

Considerando no debate ora apresentado apenas a realidade das comunidades urbanas representadas nos *corpora* estudados, pode-se afirmar que os dados europeus de não marcação de plural não ocorrem apenas com menos intensidade; antes, são periféricos quantitativa – como já se comprovou – e qualitativamente – visto que, no âmbito da concordância verbal, só foram encontrados dados de não marcação com verbo copulativo, com sujeitos pospostos ou com sujeitos antepostos cujos referentes são questionáveis ou expressos na forma de pronome cópia. Ainda assim, entende-se ser necessária uma reflexão adicional acerca da natureza das possíveis motivações estruturais para a não marcação de plural, mesmo que sugeridas por um pequeno número de ocorrências.

Em termos qualitativos, supõe-se que as restrições estruturais que atuam inequivocamente nos dados do PB e do PST, por princípio, poderiam atuar no âmbito de qualquer língua, visto que se referem, tanto no âmbito do SN quanto do SV, a componentes de ordem cognitivo-processual que se revelam nos níveis sonoro (saliência fônica), sintático (saliência posicional) e semântico (saliência do traço + animado). Fundamenta essa hipótese o que demonstra, por exemplo, Corbett (2000), ao atestar que, em línguas de perfis bem diferentes, a presença de sujeito posposto faz diminuir a ocorrência da marca de número.

¹⁷ A tese de Doutorado de Alexandre Monte, recentemente defendida na UNESP, contempla entrevistas sociolinguísticas de portugueses em Évora, escolarizados e não escolarizados.

Tabela 12 - Agreement with conjoined noun phrases.

	Animate		Inanimate	
	N	%PL	N	%PL
Subject-predicate				
Medieval Spanish	288	96	243	31
German	1.095	96	1.702	67
Russian	115	100	67	85
Serbo-Croat	21	100	35	91
Predicate-subject				
Medieval Spanish	318	69	239	6
German	379	93	925	40
Russian	89	84	114	28
Serbo-Croat	23	70	62	26

Fonte: Corbett (2000, p.201).

A tabela acima deixa claro que os percentuais de marcas, tanto com sujeito animado quanto com sujeito inanimado, os deste sempre mais baixos do que os daquele, são maiores na ordem sujeito-predicado do que na ordem predicado-sujeito, em línguas românicas ou não. Evidências como essas dão suporte à proposta de que a natureza dessas variáveis – que lidam com algum nível de saliência – está relacionada à atividade linguística geral e não se presta propriamente à indicação dos parâmetros gramaticais que caracterizam cada variedade ou cada língua.

Voltando aos dados da presente pesquisa, há indícios de que é a atuação de determinados fatores externos que acelera ou detém a implementação das restrições linguísticas. Nas realidades brasileira e são-tomense, ao que tudo indica, o valor social que se atribui à não concordância seria compatível com um estereótipo (nos termos de Labov (1972)) ou, na melhor das hipóteses, com um marcador. Por essa razão, o fenômeno variável funciona como nítido identificador do perfil social do usuário da língua. No PST – que configura um caso de variedade do português em formação, com normas objetivas ainda não totalmente definidas e assumidas pelos diferentes grupos de falantes –, sabe-se que dominar a concordância segundo a norma culta que lhe serve de modelo, a do Português Europeu, seria um dos índices de identificação de pleno domínio da Língua Portuguesa. No PB, ao contrário, o uso ou não da concordância canônica é uma opção que está claramente no nível de consciência dos indivíduos, de modo que é interpretado e funciona como forte traço caracterizador de classes sociais. Na realidade europeia, os dados desta pesquisa não permitem inferir diferença de comportamento vinculada a qualquer variável social, o que faz crer que a concordância pode constituir um parâmetro gramatical naturalmente assumido pelos falantes.

A confirmação dessas hipóteses depende, obviamente, de extensa pesquisa com pelo menos as seguintes diretrizes: (i) utilizar outras amostras de maior abrangência em termos diatópicos e diastráticos, (ii) levar em conta dados da aquisição, (iii) empregar metodologia apropriada à investigação de atitudes e, ainda, (iv) desenvolver comparação interlíngüística. São necessários, sem dúvida, não só estudos de cunho sociolíngüístico de outras variedades urbanas do português, mas também a observação dos mecanismos de concordância em outros sistemas língüísticos, de modo a aferir a universalidade das interpretações aqui propostas.

Por ora, ficam a reflexão e o convite à pesquisa, que, por instigante tanto do ponto de vista língüístico quanto social, segue seu rumo, sem limites no interesse de procurar respostas para as questões que inquietam os que buscam não só caracterizar cada variedade língüística, mas também compreender a gênese dessas características.

BRANDÃO, S. F.; VIEIRA, S. R. Nominal and verbal agreement: contributions to the debate on the status of variation in three urban varieties of Portuguese. *Alfa*, São Paulo, v.56, n.3, p.1025-1054, 2012.

- *ABSTRACT: We focus on nominal agreement and on verbal agreement in the third person plural in urban varieties of European Portuguese, Brazilian Portuguese and Portuguese of São Tome, based on the speech of individuals with primary, secondary and university education level, distributed, yet, in three age groups and sex. In order to describe these varieties, structural and social motivations have been evaluated in the light of the assumptions of the Theory of Variation and Change (WEINREICH; LABOV; HERZOG, 1968). Results from different analyzes, developed under the Project Comparative Study of Patterns of Agreement in African, Brazilian and European varieties of Portuguese, suggest that, although the three varieties tend to adopt plural marks in both the Noun Phrase and the Verbal Phrase, agreement presents distinct status in each case. In European Portuguese, linguistic data cannot be described as effectively variable – it is, according to Labov (2003), a categorical rule within the nominal agreement and semi-categorical rule within the verbal one. In Brazilian Portuguese and in Portuguese of São Tome, both rules are variable.*
- *KEYWORDS: Noun agreement. Verb agreement. Sociolinguistics. Varieties of Portuguese.*

REFERÊNCIAS

- BORTONI-RICARDO, S. M. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2004.
- BRANDÃO, S. F. Réalité sociolinguistique brésilienne et géolinguistique pluridimensionnelle. In: LE SYMPOSIUM INTERNATIONAL SUR LES LIMITES ET LES AIRES EN DIALECTOLOGIE, 2011, Lisboa. *Anais...* Lisboa: Centro de Linguística da Universidade de Lisboa, 2011. No prelo.
- CORBETT, G. G. *Number*. Cambridge: Cambridge University Press, 2000.
- GONÇALVES, P. (*Dados para a) história da língua portuguesa em Moçambique*. Maputo, 2000. Disponível em: <<http://cvc.instituto-camoes.pt/hlp/geografia/portuguesmocambique.pdf>>. Acesso em: 05 maio 2009.
- GRACIOSA, D. *Concordância verbal na fala culta carioca*. 1991. 125f. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.
- HAGEMEIJER, T. As línguas de São Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, Macau, n.1, p.1-27, 2009.
- ILHA DE SÃO TOMÉ. In: WIKIPÉDIA a encyclopédia livre. Disponível em: <<http://pt.wikipedia.org/wiki/Ficheiro:Tp-map.png>>. Acesso em: 20 out. 2011.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. *Censo populacional 2010*. Brasília, 29 nov. 2010. Acesso em: 11 dez. 2010.
- LABOV, W. Some sociolinguistic principles. In: PAULSTON, C. B.; TUCKER, G. R. (Org.). *Sociolinguistics: the essential readings*. Oxford: Blackwell, 2003. p.235-250.
- _____. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- LUCCHESI, D.; BAXTER, A.; RIBEIRO, I. (Org.). *O português afro-brasileiro*. Salvador: Ed. da UFBA, 2009.
- METRODATA. Região metropolitana de Rio de Janeiro (RJ): taxa de crescimento anual por municípios. *Observatório das Metrópoles*, Rio de Janeiro. Disponível em: http://www.observatoriodasmetropoles.ufrj.br/metrodata/ibrm/ibrm_rj_tca.htm. Acesso em: 20 out. 2011.
- MONGUILHOT, I. O. S. *Estudo sincrônico e diacrônico da concordância verbal de terceira pessoa do plural no PB e no PE*. 2009. 228f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Santa Catarina, Florianópolis, 2009.
- NARO, A.; SCHERRE, M. M. P. *Origens do português brasileiro*. São Paulo: Parábola, 2007.

PLANO REGIONAL DE ORDENAMENTO DO TERRITÓRIO DA ÁREA METROPOLITANA DE LISBOA. *Área Metropolitana de Lisboa*. 2008. Disponível em: <<http://protaml.inescporto.pt/>>. Acesso em: 20 out. 2012.

VAREJÃO, F. O. A. *Variação em estruturas de concordância verbal e em estratégias de relativização no português europeu popular*. 2006. 187f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

VIEIRA, S. R. O estatuto da regra variável e o fenômeno da concordância verbal em variedades do português. In: CONGRESO INTERNACIONAL DE LA ALFAL, 16., 2011, Alcalá de Henares. *Actas...* Alcalá de Henares: ALFAL, 2011.

WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical foundations for theory of linguistic change. In: LEHMANN, W.; MALKIEL, Y. (Org.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p.97-195.

Recebido em março de 2012

Aprovado em maio de 2012

MOTIVAÇÕES EM COMPETIÇÃO NA VARIAÇÃO SOCIOLINGUÍSTICA: O PLURAL DOS PREDICATIVOS NA VARIEDADE DE SÃO JOSÉ DO RIO PRETO

Roberto Gomes CAMACHO*
Mircia Hermenegildo SALOMÃO**

- **RESUMO:** O objetivo deste trabalho é o de submeter a um tratamento variacionista, de base quantitativa, dados de marcação variável de plural no SN e no SA em contexto de predicativo, obtidos em córpus coletado na região de São José do Rio Preto. O trabalho procura examinar se a marcação de pluralidade nos predicativos pode ser explicada com base em motivações exclusivamente formais, ou exclusivamente funcionais, ou ainda, com base na interação entre ambas, que consistiriam, assim, em motivações em competição. Os resultados indicam que nem as motivações funcionais, nem as motivações formais regem solitariamente o fenômeno, que aparece fortemente condicionado por uma restrição externa específica, o grau de escolaridade. Por essa razão, a explicação mais plausível para a marcação de pluralidade nos predicativos é a de que há motivações em competição, nos termos de Du Bois (1985), e é a marcação de pluralidade o “bem limitado”, pelo qual forças múltiplas, as motivações formais ou internas e funcionais ou externas, competem entre si.
- **PALAVRAS-CHAVE:** Funcionalismo. Formalismo. Pluralidade. Concordância nominal. Predicativo.

Palavras iniciais

A partir da década de 80, a sociolinguística variacionista passou a suspeitar teórico-metodologicamente do caráter explanatório do funcionalismo, entendido, porém, por Labov (1987, 1994) como restrito ao que Du Bois (1985) rotula “funcionalismo transparente”. O resultado mais flagrante desse gesto foi adotar uma versão mais ou menos forte do estruturalismo autônomo, com a predominância de explicações formais e uma concepção autônoma de sintaxe.

Um aspecto que parecia necessário investigar é se não haveria um meio-termo entre essas duas tendências que pudesse amenizar posições tão antagônicas e

* UNESP – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. São José do Rio Preto – São Paulo – Brasil. 15054-090 - camacho@ibilce.unesp.br

** UNESP – Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. São José do Rio Preto – São Paulo – Brasil. 15054-090 - mirciah@yahoo.com.br

a resposta parecia advir do conceito de motivações em competição de Du Bois (1985). O objeto de estudos de uma teoria como a sociolinguística variacionista consiste em fenômenos de grande complexidade, pois o que motiva a variável são, por um lado, condições internas e externas e, por outro, condições internas formais e internas funcionais. É em função disso que se parecia adequar-se bem a ideia de que essa complexidade pudesse ser investigada como um jogo de motivações em competição.

Foram essas as razões que conduziram à proposição do projeto *Motivações formais e funcionais da marcação de pluralidade no português falado na região de São José do Rio Preto*, patrocinado pela FAPESP¹ e desenvolvido entre 2006 e 2008. Os participantes impuseram a si mesmos investigar a marcação variável de número, entendida, na gramática normativa, como regras de concordância nominal e verbal, e numa variedade muito específica – a Amostra Iboruna (GONÇALVES, 2007; GONÇALVES; TENANI, 2008), que havia sido recentemente criada para suprir uma carência de dados sobre a realidade dialetal do interior paulista.

Com efeito, a sociolinguística brasileira, liderada pelo pioneirismo do Projeto Censo do Rio de Janeiro, transformado depois em Projeto PEUL, patrocinou um avanço considerável da pesquisa sociolinguística e, em consequência, do conhecimento da realidade sócio-dialetal do Brasil. No entanto, o conhecimento das variedades praticadas em São Paulo ainda era, há uma década, praticamente derivado de alguns trabalhos isolados e, por isso, longe de fornecer um quadro representativo da fala paulista.

Um passo decisivo para suprir essa lacuna foi a instituição, no ano de 2003, do projeto ALIP - Amostra Linguística do Interior Paulista (GONÇALVES, 2007), motivada pelos interesses de pesquisas que tinham como principal diretriz o enfoque da língua inserida em seu contexto social. O projeto ALIP não teve por objetivo o estudo de nenhum fenômeno particular; na realidade, assumindo para si uma concepção ampla e aberta dos possíveis fenômenos variáveis, procurou captar em grau máximo a diversidade linguística da região noroeste do estado de São Paulo, representada por São José do Rio Preto e seis outras cidades que lhe fazem fronteira: Bady Bassit, Cedral, Guapiaçu, Ipiguá, Mirassol e Onda Verde.

A constituição do Banco de Dados Iboruna (= rio preto, em Tupi), desenvolvida entre 2004 e 2007, representa dois tipos de amostras de fala: a primeira, coletada de acordo com os critérios da Sociolinguística laboviana (LABOV, 1972), constitui a *Amostra Censo*, e a segunda, coletada secretamente em contextos interacionais livres, sem controle de qualquer variável social (RONCARATI, 1996), constitui a *Amostra de Intereração*.

¹ Além do coordenador, constituiu-se uma equipe composta pelos seguintes bolsistas: Francielli Honorato Alves, William da Silva Santana, Ana Helena Rufo Fiamengui e Mircia Hermenegildo Salomão. Os desdobramentos práticos mais significativos dessa pesquisa foram as dissertações de mestrado dos dois últimos integrantes, apoiados, inclusive, com bolsas FAPESP, e a tese de doutorado em andamento de Mircia H. Salomão, também bolsista da mesma agência.

Da uma perspectiva variacionista, que interessa aqui, os fenômenos de concordância já investigados contemplam os seguintes fenômenos: a marcação de plural em sintagmas nominais (FIAMENGUI, 2011) e em contextos de predicativo (SALOMÃO, 2010); a concordância verbal de 3PP (RUBIO, 2008) e padrões de concordância verbal e de alternância pronominal (RUBIO, 2012).

Fechado esse parêntese, retornemos ao projeto a que este trabalho se vincula. Como objetivo geral, o projeto *Motivações formais e funcionais da marcação de pluralidade no português falado na região de São José do Rio Preto* mirava a busca de razões teóricas e metodológicas que permitissem examinar que hipótese é a mais plausível para explicar a ausência e a presença de marcas de pluralidade nessa variedade. Por razões óbvias de espaço, o fenômeno enfocado neste trabalho se restringe aos SNs e SAs em estruturas predicativas.

A pergunta que se faz é se, na variedade estudada, a maior presença ou ausência de marcas tem motivação funcional em função de redundância, ou se tem motivação formal em função de repetição de estruturas em paralelismo formal; e ainda, alternativamente, se condições funcionais e formais competem entre si na motivação desse processo variável. Trata-se, como se vê, de reacender o debate teórico que se iniciou nos anos 80, momento em que a teoria sociolinguística teve que encarar críticas em função da expansão da análise variacionista ao estudo de variáveis do nível sintático (CAMACHO, 2003).

Tratando agora mais especificamente do objeto de estudo, é possível afirmar que os trabalhos realizados, na esteira da pesquisa pioneira de Poplack (1980) para o espanhol americano, tornaram possível formular princípios restritivos de natureza linguística e extralingüística na aplicação de regras variáveis como a de marcação de plural nos nomes. Esse trabalho propiciou um desenvolvimento acentuado do estudo da regra variável de concordância no português do Brasil a partir da década de 80. Os fenômenos que envolvem concordância são, muitas vezes, motivados por mecanismos de erosão fonológica, como processos de redução e/ou supressão consonantal, entre os quais se enquadra apagamento da marca de plural realizada por uma das pronúncias do arquifonema fricativo /S/.

É possível que os processos fonológicos de redução de segmentos fonológicos, como o de marcação de número no SN, possam ser acelerados por conta da redundância desse mecanismo flexional, com a presença de marca formal de plural em todas as palavras constituintes de um sintagma; haja vista a intuição clara de que o /S/ de palavras como *lápis* é frequentemente muito menos apagado do que núcleo de SNs como **os meninos**. Apesar dessa característica sintática, a variante padrão concorre, na prática quotidiana, com sentenças alternativas em que a indicação de pluralidade pode ocorrer num único constituinte, geralmente o que se localiza mais à esquerda do SN sujeito e, portanto, mais à esquerda da estrutura sentencial como um todo. Essa

possibilidade pode reduzir a indicação de número a um único constituinte da sentença, comprometendo a expressão normativa tanto da regra de concordância nominal quanto da regra de concordância verbal.

Em predicativos com a mesma estrutura de SNs não predicativos, as restrições são as mesmas, mas outros fatores, como a marcação de plural no SN sujeito e no SV, interferem igualmente nesse processo variável, tratado nas gramáticas normativas como uma regra de concordância nominal e, como tal, de aplicação compulsória em todos os contextos de uso. Pode-se afirmar que o estudo dos predicativos em si não teve o mesmo sucesso que a pesquisa de outros tipos de sintagmas em termos de concordância nominal.

Além de ser mais reduzido o número de trabalhos, na dialetologia brasileira, sobre a marcação de plural no predicativo, torna-se praticamente uma necessidade incontornável a investigação dessa regra variável numa variedade dialetal ainda pouco conhecida, como a que se fala em uma parte da região noroeste do estado de São Paulo, nucleado em torno de São José do Rio Preto. Por isso, este trabalho se debruça sobre uma amostra de casos extraída do Córpus Iboruna, que justamente representa a fala dessa comunidade.

Este trabalho adota o modelo da Sociolinguística Quantitativa, aplicando-o a alguns princípios funcionalistas como a Condição de Distintividade (KIPARSKY, 1972) e o princípio de economia (HAIMAN, 1983), a fim de examinar as motivações (lingüísticas e extralingüísticas) em competição (DU BOIS, 1985) na indicação de pluralidade nas estruturas predicativas da variedade selecionada. O artigo pretende discutir a natureza explanatória de motivações funcionais e de motivações formais para o enfoque de processos variáveis, restritos aqui à marcação variável de número em SNs e SAs de estruturas predicativas.

A sociolinguística variacionista e a natureza formal e funcional da linguagem

Se pensarmos no que o formalismo e o funcionalismo diferem mais nitidamente entre si, talvez seja no significado que cada qual atribui ao grau maior ou menor de autonomia da sintaxe. A concepção formalista estuda a linguagem como um sistema autônomo centralizado na sintaxe. O formalismo aposta suas fichas na autonomia da sintaxe em relação à semântica e à pragmática; para essa corrente, a prioridade vai da sintaxe à pragmática, via semântica. Já o funcionalismo vê os componentes mais visíveis da organização linguística – o pragmático, o semântico e o sintático como uma relação de maior ou menor grau de abrangência, mas em que é o primeiro o mais abrangente, dentro do qual se devem examinar os outros dois, com a exclusão de qualquer espaço para uma sintaxe autônoma (DIK, 1997).

A relação entre a estrutura e a função linguística é estabelecida pelo enfoque funcionalista da variação em termos de motivações (CAMACHO, 2008). Dessa maneira, um linguista funcionalmente orientado assume que há motivações sistemáticas, que estão sujeitas a princípios regulares subjacentes à relação entre funções discursivas e estruturas morfossintáticas relevantes. As motivações são relacionadas ao modo como as estruturas se projetam nos conceitos correspondentes (iconicidade), à frequência e ao grau de acessibilidade dessas estruturas na mente (economia), à frequência e ao grau de acessibilidade de combinações individuais de estruturas e conceitos.

O princípio funcional de economia (HAIMAN, 1983) representa uma pressão para a simplificação máxima da expressão. A estrutura linguística torna-se, dessa forma, o mais opaca possível, contrapondo-se ao princípio da motivação icônica, que tende à transparência máxima entre forma e função. Pode-se entender o princípio de economia sintagmática como uma tendência para a redução do comprimento ou da complexidade do enunciado, de maneira que as expressões mais frequentes no uso tendem a reduzir-se fonologicamente e a informação redundante ou recuperável no contexto comunicativo tende a ser suprimida.

Situar a sociolinguística nesse quadro panorâmico deixa claro que a abordagem variacionista se inclui mais apropriadamente no espectro das tendências não formalistas, pois “[...] em sua fase inicial, a sociolinguística variacionista sustenta posições teórico-metodológicas que a distanciam da tendência formalista, inclusive como uma abordagem alternativa.” (CAMACHO, 2008, p.19).

Na fase inicial, a sociolinguística variacionista reconhece a possibilidade teórica de que fatores sistemáticos do contexto social interfiram na estrutura linguística, mediante a implementação, a avaliação e a transição de variáveis linguísticas. Mais tarde, na fase em que o estudo da variação passou a tratar também de unidades morfológicas, a sociolinguística mantinha um forte comprometimento com compensações sistêmicas para perda de distinções funcionais relevantes, a chamada regra de mudança estrutural compensatória (*Rule of Compensatory Structural Change*) (LABOV, 1972).

Nos dialetos ingleses da Escócia, a simplificação de grupos consonantais com *t,d* após oclusivas não sonoras tornou-se categórica conduzindo a um uso uniforme *ac'* por *act* e *ap'* por *apt*. Seria perfeitamente natural que a mesma regra eliminasse também a marcação de pretérito em *liked* e *stopped*, mas uma regra generalizada de epêntese eliminou o encontro consonantal, preservando os morfemas de pretérito dessas formas. Com esse exemplo, Labov (1972) pretendia argumentar que o sistema gramatical pode aproveitar-se de outras mudanças sonoras para prover o usuário de todas as necessidades comunicativas.

Apesar dessa tendência francamente funcionalista na formulação inicial da sociolinguística variacionista e das substanciais descobertas de regularidades sob a égide da noção de regra variável, esse conceito entrou rapidamente no fogo cerrado da crítica, quando, na década de 70, o conceito foi estendido aos fenômenos sintáticos com os mesmos métodos aplicados aos fenômenos fonológicos.

O compromisso da análise sociolinguística com explicações funcionalistas e formalistas

Quando a variável se circunscreve a um fenômeno fonológico, o problema do significado não tem pertinência. No entanto, se a variável amplia seu escopo abarcando fenômenos sintáticos, é indiscutível a relevância do significado. Assim, quando os métodos e técnicas da investigação variacionista passaram a ser aplicados a casos de variação em sintaxe, irrompeu a primeira crise na metodologia do paradigma variacionista, em razão dos forte ataques desferidos por Lavandera (1978), Romaine (1981) e García (1985).

As críticas de Lavandera e García visaram o estatuto teórico da regra variável, quando aplicado à sintaxe. Já Romaine mirou o aspecto indutivista da teoria. Lavandera (1978) sustenta que é inadequado alargar a outros níveis de análise da variação o conceito de variável sociolinguística, que foi originariamente desenvolvida com base em dados fonológicos. Segundo a autora os estudos quantitativos da variação que se ocupam de alternâncias morfológicas, sintáticas e léxicas sentem falta de uma teoria bem organizada do significado.

Lavandera (1978) afirma que a análise da variação em fonologia pode ser admitida como uma contribuição para uma melhor compreensão dos tipos de significados, que podem comunicar diferenças formais; no entanto, a extensão da noção de variável para a variação não fonológica deixa de ser reveladora, pois para a autora há diferenças cruciais entre variáveis fonológicas e variáveis não fonológicas. As primeiras dispõem de significado social e estilístico e não precisam ter significado referencial. No entanto, as variáveis não fonológicas veiculam significado social e estilístico, mas também têm significado referencial. Lavandera (1978) destaca a dificuldade metodológica de se lidar com variáveis não fonológicas, principalmente por envolverem significação referencial.

Dessa maneira, a autora propõe relaxar a condição de que o significado referencial deve ser o mesmo para todas as alternativas, introduzindo, assim, uma condição de comparabilidade funcional, já que a equivalência semântica entre duas variáveis sintáticas implicaria numa redução drástica da noção de significado referencial.

Labov (1978) não concorda com essa condição e nega a dificuldade da expansão das técnicas e métodos do modelo às variantes não fonológicas. Segundo o autor, muitas variáveis sintáticas não fornecem qualquer problema para o estabelecimento do mesmo significado representacional, bastando, para tal, reduzir o escopo do significado ao valor de verdade.

É o que fazem Weiner e Labov (1983): a escolha entre passiva não agentiva e ativa com agente genérico no discurso espontâneo deixa-os perfeitamente convencidos de que se trata aqui de uma variável sintática legítima em que as duas variantes têm o mesmo significado em termos de condições de verdade.

Mais recentemente, ao discutir o caráter funcional da mudança, Labov (1994) retoma e mantém essa mesma linha de argumentação. Em face da variação fonológica e morfológica, a hipótese funcional prevê a tendência de o falante selecionar uma variante ou outra de modo a preservar informação. A maioria dos resultados que discute mostra, ao contrário, que, no discurso corrente, a escolha entre variantes descarta a maximização de informação em favor de efeitos mecânicos, como condicionamento fonético e mera repetição de estruturas precedentes. Embora a mudança ameace a capacidade comunicativa, a maioria dos sistemas linguísticos se readjusta ao longo do tempo para preservar os meios de veicular informação.

Labov (1994) sustenta que a caracterização neogramática da estrutura linguística está de fato correta: como as leis fonéticas, a estrutura é um sistema predominantemente mecânico, fora do alcance do reconhecimento consciente ou do ajustamento de seus usuários. Como a linguagem é indubitavelmente programada para veicular informação proposicional, parece estranho que o falante não tenha liberdade de ajustar o sistema para maximizar a eficiência comunicativa.

Para o autor, uma explicação possível se assenta no fato de que a eficiência da linguagem depende de sua natureza automática e, por conseguinte, os esforços por mudá-la conscientemente devem confinhar-se apenas a opções estilísticas de nível superior: a seleção de palavra e a construção de sintagmas e sentenças dentro de um conjunto estreitamente limitado de possibilidades.

Fatores condicionadores como motivações em competição

O breve tratamento das implicações de motivações de natureza funcional ou formal para a teoria variacionista envolve necessariamente a discussão da possibilidade de não se excluírem mutuamente, mas de competirem entre si. Quem primeiramente discutiu a noção de motivações em competição foi Du Bois (1985), que a vinculou à relação entre iconicidade e arbitrariedade.

No arcabouço teórico do estruturalismo clássico, que Du Bois (1985) chama de “estruturalismo autônomo”, a iconicidade é claramente considerada uma anomalia por operar contra a visão de que a linguagem é um sistema completamente autônomo com o conceito de arbitrariedade no centro. O principal arcabouço teórico a herdar de Saussure (1977) esse forte comprometimento com a autonomia da gramática é a teoria gerativa em suas diversas versões.

Em contraste com essa perspectiva, o “funcionalismo transparente” assume que fatores sintáticos aparentemente autônomos constituem, na realidade, os resultados transparentes dos objetivos funcionais do falante, de modo que não é necessário postular nenhum mecanismo sintático arbitrário para a língua (DU BOIS, 1985).

Na medida em que sejam reconhecidas como teorias, o estruturalismo autônomo e o funcionalismo transparente devem ser entendidos como inerentemente opostos e aparentemente irreconciliáveis; entretanto, estudos recentes em gramaticalização têm posto ênfase nas relações entre discurso e gramática, e tentado integrar parcialmente esses dois enfoques.

O estruturalismo autônomo estabelece uma forte distinção entre a linguística interna e a linguística externa. Entende Du Bois (1985) que essa distinção é necessária e é nela que o autor se apoia para mostrar que certos fenômenos podem ser simultaneamente imotivados a partir do ponto de vista sincrônico e motivados a partir de um ponto de vista metagramatical, que trata as gramáticas como sistemas adaptativos.

As gramáticas são, por um lado, parcialmente autônomas, e daí **sistemas**, e, por outro, parcialmente suscetíveis a pressões externas ao sistema, daí **adaptativas**. Segundo Du Bois (1985), essa distinção é metodologicamente útil apenas se dermos pleno reconhecimento à existência de motivações em competição e desenvolvermos um arcabouço teórico que, inicialmente, descreva e analise a interação das motivações com contextos específicos e, posteriormente, seja capaz de prever a resolução da competição entre elas.

Em primeiro lugar, para duas motivações competirem no mesmo âmbito funcional deve haver o que Du Bois (1985) chama de “bem limitado”, pelo qual elas competem. O “bem limitado” é o domínio da forma linguística, sobre as quais forças motivadoras múltiplas não podem prevalecer completamente.

Duas motivações que competem para o controle de um único paradigma linguístico, categoria ou morfema podem ser tanto ambas internas quanto ambas externas ou mesmo pode ocorrer que uma das motivações seja interna enquanto a outra seja externa. Segundo Du Bois, a avaliação do significado da teoria linguística se dá a partir de uma perspectiva que vê as gramáticas como sistemas adaptativos e que, portanto, reconhece, em primeiro lugar, a interação entre forças internas e

externas. Quando as forças que se originam de fenômenos externos penetram no domínio da língua, elas encontram e interagem com as forças internas.

Dizer que a língua é adaptativa por responder a pressões do contexto externo, implica de pronto rejeitar a teoria do estruturalismo autônomo, que é limitada na tentativa de explicar a estruturação de uma língua, por reconhecer apenas forças motivadoras internas. Dizer, por outro lado, que a língua é um sistema, e que, como tal, tem certo grau de continuidade de existência, implica rejeitar o arcabouço teórico do funcionalismo transparente, que falha em explicar o fato de que categorias gramaticais podem ser preservadas, por assim dizer, numa forma mais ou menos congelada. Mais importante ainda que isso talvez, ele falha em construir apropriadamente o aspecto interno mais fundamental da gramática, o processo de gramaticalização em si mesmo.

A compreensão dessas diferenças tem uma importância fundamental para os objetivos desta pesquisa, pois, em grande medida, foram essas razões teóricas que organizaram os procedimentos metodológicos para o processamento dos dados, ou seja, com base em uma metodologia que favorecesse o enfoque da marcação de pluralidade nas estruturas predicativas a partir de duas diferentes perspectivas, a formal e a funcional. A dimensão formal explica as motivações internas, enquanto a dimensão funcional, acrescida dos fatores extralingüísticos, explica as motivações externas. Nesse caso, o “bem limitado”, na visão de Du Bois (1985), é o domínio da variável em busca de resolução seja em direção da categoricidade, mediante predomínio de uma das variantes, seja em direção de estabilidade, mediante distribuição por diferentes comunidades sociais como indício de identidade cultural.

A literatura sobre marcação de pluralidade

De acordo com a norma padrão do português brasileiro, o fenômeno de concordância é considerado regra obrigatória. Como a concordância reitera marcas contendo a mesma informação em pontos diversos da cadeia sintagmática, a variante padrão é necessariamente redundante.

Por outro lado, é também verdadeiro ser a concordância de número uma variação marcada pela identidade social do falante e por motivações emanadas do contexto social. Uma motivação potencialmente relevante para não se usar a regra de concordância nominal é o baixo nível de escolaridade, que impede as pessoas de familiarizar-se suficientemente com a pressão normativa exercida pelo sistema escolar; outra motivação potencialmente relevante é a situação discursiva: mesmo falantes com alto grau de escolaridade não aplicam a regra de concordância em circunstâncias mais informais de uso. Assim, o modo categórico

como a tradição gramatical concebe a variação de pluralidade não se aplica de fato ao uso da língua no contexto social.

Como se sabe, à medida que a língua varia, é possível expressar um mesmo enunciado de modo diferente sem mudar o seu significado, como nas sentenças contidas em 1(a-b)

- (1) a os orelhão (\emptyset) de lá não é igual (\emptyset) aqui (AC029/NE/L.069)
b os orelhões de lá não são iguais aos daqui

Na alternativa contida em (1a), a marca formal de pluralidade se encontra apenas no determinante, enquanto na contida em (1b) as marcas formais de pluralidade estão presentes no determinante e no núcleo do SN sujeito e do predicativo, além da marcação na cópula verbal. A ausência de marcas não é obstáculo para que, ainda assim, seja assegurada a noção de pluralidade sem provocar nenhuma ambiguidade tendo em vista a marcação de número no determinante.

É possível afirmar, então, que, de um ponto de vista funcional, a explicitação de pluralidade no determinante poderia tornar desnecessárias as marcas nos outros constituintes, por razão de economia, em função da redundância inerente da regra de concordância nominal na variedade padrão. Outra explicação possível seria uma motivação por reiteração de formas em paralelismo, tanto na marcação quanto na ausência de marcação (SCHERRE, 1988).

Esse modo variável de marcar número leva a refletir sobre que motivação estaria em jogo no processo de variação da concordância nominal: (i) o paralelismo formal, na medida em que a indicação de pluralidade explícita, ou não explícita, levaria em conta a marcação do mesmo tipo na posição anterior; (ii) a condição funcional de distintividade (KIPARSKY 1972), na medida em que a indicação redundante a partir do constituinte marcado acionaria a desnecessidade de outras marcas subsequentes sem risco de ambiguidade referencial.

Seria possível também supor que poderia haver motivações em competição (DU BOIS, 1985), de natureza interna, ou formal, e externa, ou social. Nesse caso, seria possível supor que os casos alternativos que não apresentam todas as marcas de plural podem representar a escolha de diferentes grupos sociais. Desse modo, haveria grupos sociais em que predominaria a escolha de SNs nominais com marcas formais de plural e, em contraste, outros grupos sociais para os quais a presença de todas as marcas nem sempre é necessária.

As investigações no espanhol cubano realizadas por Terrel (1975) sobre as condições funcionais da supressão do marcador de plural permitiram constatar que os falantes cubanos evitam suprimir todos os traços de indicação morfológica

de número, preservando, dessa maneira, a primeira forma da estrutura sentencial. Outros trabalhos, como o de Braga (1978), com círculo coletado na região do Triângulo Mineiro, confirmam que o determinante, que geralmente ocupa a primeira posição no SN, é a categoria em que mais é observada a retenção da marca de número.

No entanto, essa posição foi amplamente reformulada por Scherre (1988), mediante a reanálise do mesmo fenômeno. A autora rejeita a ideia de que posição do constituinte seja o fator determinante da marcação de pluralidade no SN, mostrando que posição e categoria gramatical são condições fortemente entrelaçadas para a determinação de marca de número no sintagma nominal.

A autora postula um princípio mais geral, baseado na constatação de que, no SN, marca leva à marca e zero leva a zero, princípio esse que denominou de “paralelismo formal”. Ao postular esse princípio, Scherre (1988) recusa também a atuação da Condição de Distintividade de Kiparsky (1972), considerada por ela como uma hipótese funcionalista no arcabouço da teoria gerativa.

Em seus estudos, Kiparsky (1972) se concentra num fenômeno fonológico encontrado em algumas variedades do inglês norte-americano mediante o qual se apaga variavelmente o fonema /t/ em final de palavra. Quando esse fenômeno fonológico atua como marca de pretérito, o fonema é cancelado somente nos casos de verbos irregulares, cujas formas de presente e de pretérito contêm uma vogal diferente na raiz, como no caso de *keep* vs. *kep(t)*. Pode-se afirmar que a regra de apagamento é bloqueada ou aplicada em menor grau quando sua aplicação neutraliza a distinção entre a forma do presente e a do pretérito, como em *passed*, e é liberada quando a distinção se mantém devido à alternância vocálica na raiz (KIPARSKY, 1972). Na tentativa de explicar esse fenômeno variável, o autor postula esse princípio funcional, a Condição de Distintividade, que se caracteriza como uma tendência pela retenção de uma informação semanticamente relevante na estrutura superficial.

Poplack (1980) foi uma das primeiras pesquisadoras a incorporar esses princípios aos estudos sociolinguísticos. A autora investiga a influência de condições formais e funcionais competindo pela marcação variável de plural no espanhol porto-riquenho. Poplack (1980) reivindica a superioridade das condições estruturais para a explicação desse fenômeno variável. Sua pesquisa revela evidências de uma tendência local, isto é, de uma tendência para a concordância no nível da sequência: a ausência de um marcador no segmento precedente à ocorrência favorece a supressão neste elemento, enquanto a presença de uma marca imediatamente precedente favorece a retenção de uma marca na ocorrência. Isso mostra que, se a cadeia manifestar a supressão de marcas nos dois primeiros constituintes, é significativa a contribuição do fator “posição” para a probabilidade de supressão no terceiro elemento. Se a sequência for realizada com

marcas na primeira e na segunda posição, é bem reduzido o efeito da supressão no constituinte da terceira posição. O resultado é, enfim, exatamente oposto do que se descobriu em outros estudos, e é caracterizado por Labov (1994) como um argumento contrafuncional.

Hochberg (1986) retoma a hipótese de Kiparsky (1972) no estudo da variedade porto-riquenha do espanhol, cujos resultados seguem uma direção oposta aos de Poplack (1980), confirmando, assim, a atuação da Condição de Distintividade para a explicação do mesmo fenômeno. Entretanto, em estudo parecido ao de Poplack (1980), Guy (1981) não encontra, nos seus dados sobre o português brasileiro, respaldo para a confirmação da hipótese funcionalista, e acaba postulando que as restrições funcionais atuam apenas quando forma e função são perfeitamente paralelas (GUY, 1981).

Lefebvre (1981), estudando o Quéchua de Cuzco, apresenta também evidências de restrições funcionais e não funcionais em competição, mostrando que algumas situações são previstas pela hipótese funcionalista de Kiparsky e outras não (LEFEBVRE, 1981).

O paralelismo formal foi absorvido como princípio teoricamente explanatório por Labov (1994) como forma de rejeição simultânea de princípios funcionalistas, um movimento que repercutiu largamente na teoria variacionista. Resta, portanto, saber se procedimentos mecânicos sempre sobrepujam os funcionais que requerem marca de pluralidade quando o processo de comunicação exigir eficiência informativa ou se representam motivações competindo entre si, como Lefebvre (1981) sugere para o Quéchua.

Composição da amostra e definição dos condicionamentos da variável

O universo de pesquisa deste trabalho é composto por um recorte do Banco de Dados Iboruna², contendo 128 de 152 gravações das Amostras Censo (GONÇALVES, 2007; GONÇALVES; TENANI, 2008). O recorte aqui realizado se justifica pela ausência, na pesquisa original (SALOMÃO, 2010) da faixa etária de 7 a 15 anos e, portanto, pela preferência por trabalhar com uma apreensão mais consolidada da norma padrão. Cada Amostra Censo é composta por cinco gêneros textuais: **narrativas de experiência pessoal, relato de narrativa recontada, relato descritivo, relato de procedimento e relato de opinião**.

A constituição da amostra provém da gravação de 152 informantes, cujo perfil é determinado pela combinação de fatores sociais: gênero/sexo, escolaridade, faixa

² Apesar de o Banco de dados Iboruna possuir Amostra de Intereração (AI) e Amostra Censo (AC), incorporou-se o segundo tipo de amostra, pois os fatores sociais nas Amostras de Intereração não foram controlados, uma vez que as gravações foram feitas anonimamente e sem o conhecimento prévio dos informantes.

etária e classe social (renda familiar). Desse modo, os grupos de fatores sociais têm a seguinte constituição:

Tabela 1 - Grupos de fatores sociais.

Grupo de fatores	Fatores
1 - Sexo/ gênero	masculino; feminino
2 - Faixa etária	16 a 25 anos; 26 a 35 anos; 36 a 55 anos; mais de 55
3 - Escolaridade	1º ciclo do Ensino Fundamental; 2º ciclo do Ensino Fundamental; Ensino Médio; Ensino Superior

Fonte: Elaboração própria.

A investigação da marcação de pluralidade na variedade falada na região de São José do Rio Preto se restringe aqui à análise de um envelope de variação constituído pela variável binária “presença ou ausência de marcação de pluralidade em SNs e SAs predicativos”, formalizada, basicamente, como <s>, representando o arquifonema /S/. Essa categoria representa, por sua vez, as variantes presença ou ausência de qualquer realização fonética, com alterações morfológicas da marca de pluralidade.

No português brasileiro, presença e ausência de plural se alternam na língua falada em posição final de palavra, contexto que apresenta grande potencial de variação em função de um processo de erosão fonológica por distensão dos órgãos fonadores em posição átona. Essa regra é aplicada tanto em palavras monomorfêmicas (como *lápis, menos*) quanto em palavras bimorfêmicas (como *casas, patos, meninas*). Mesmo representando um morfema de plural, a redundância de marcação no SN leva a uma supressão maior da fricativa em contextos de palavras bimorfêmicas do que em palavras monomorfêmicas.

O contexto de análise dessa variável é representado pela posição sintática de predicativo, que pode ser representada por SNs ou SAs de um único elemento (substantivos, adjetivos, tomados como elementos predicadores propriamente ditos, como mostram (3a-b) e participios em estruturas passivas, como mostra (2). Nos exemplos contidos em (3a-b), o predicativo é também acompanhado por palavras invariáveis, que aparecem grifadas.

- (2) a sala de aula de terce(i)ra e quarta séries são salas onde não existe alfabeto... porque os alunos já estão **alfabetizados**... (AC 88/ L 335-336/ DE)

- (3) a (...) mas as casas ali são bem **antigas**... né?... (AC 120/ L 266/ DE)
- b (...) que nós falávamos que a S. num era boa pesso::a... que num era
boa companhi:a... aí diz que o marido dela respondia assim – “imagina
eles são muito sistemático eles tão com inVEJA dela” (AC 050/ L
134-136/ NR)

Selecionadas as ocorrências, cada uma delas foi analisada segundo os critérios definidos nos grupos de fatores relacionados a seguir, os quais constituem os contextos variáveis definidos para a discussão neste trabalho. Os parâmetros de análise, cuja base teórica se assenta nos trabalhos de Scherre (1988, 1991, 1994) e no de Dias (1996), vão ser apresentados paralelamente à apresentação dos resultados. Para o tratamento quantitativo dos dados, usou-se o Pacote Estatístico Goldvarb (SANKOFF; TAGLIAMONTE; SMITH, 2005).

Motivações internas e externas da marcação de número no SN/SA predicativo

A amostra constitui-se de um total de 526 ocorrências com estruturas predicativas, que revelam a aplicação da regra de CN, dentre os quais 58,0% (310/526) apresentam marcas de plural, enquanto 42,0% (216/526) não apresentam aplicação da regra de concordância nominal (SALOMÃO, 2010).

Os resultados serão apresentados em tabelas vistos da perspectiva do valor da aplicação da regra, ou seja, da presença de marcas formais nas estruturas predicativas. Conforme se verá, o exame do Princípio de Paralelismo Formal pode ser visto em duas dimensões: na dimensão intraoracional e na dimensão interoracional ou discursiva.

Na dimensão intraoracional, as estruturas predicativas são vistas em sua relação com outras categorias sintáticas da oração. Como deveriam ser supostamente marcadas com base na regra de concordância nominal da variedade padrão, essas categorias deveriam ativar a marcação na forma do predicativo. Essas categorias são o sujeito e o verbo, que em geral se confina com a cópula *ser*.

Na dimensão interoracional ou discursiva, o princípio do paralelismo formal é visto como a repetição de estruturas similares, cujo escopo, nesse caso, é o de uma oração inteira contendo o fenômeno em análise com formas similares que podem recorrer ou não no discurso corrente.

A trajetória de análise se inicia com a dimensão intraoracional, mas, ainda mais especificamente, com as características formais do sujeito, vistas em sua relação

com a estrutura predicativa, isto é, examina-se a marcação de plural na estrutura predicativa em correlação com uma configuração flexional similar no sujeito.

O que se busca investigar é a confirmação ou não da hipótese formal de que a presença ou ausência de marcas conduz a uma configuração recorrente no predicativo; na situação inversa, em que essa hipótese seria rejeitada, é necessário verificar se há outro princípio funcional implícito, como a Condição de Distintividade (KIPARSKY, 1972).

Nesse parâmetro, considera-se não explícitos todos os tipos de sujeito considerados por Scherre (1988, 1991, 1994) como nulo e desinencial, conforme mostra (4a). Os casos de manifestação explícita do SN sujeito incluem situações como a de (4b) em que todos os elementos nominais flexionáveis ou somente os últimos estão marcados com plural e em que a última marca está neutralizada em função do contexto fonológico seguinte. Em (4c) aparecem situações em que ou o sujeito explícito é privado da(s) última(s) marca(s) formal(is) explícitas de plural ou o sujeito explícito é coordenado por constituintes no singular. Em (4d) aparece uma situação em que o sujeito explícito tem marca semântica de plural (*nós* e numerais isolados). Finalmente, em (4e), o sujeito explícito aparece com marca formal de plural totalmente neutralizada.

- (4) a então nós casamos e:: a gente trabalhava muito né?... mas **éramos** felizes passeávamos... e viajávamos muito (AC 138/NE/L.15-16)
- meus pais** são bastante são evangélicos né?... (AC035/NR/L.202)
- b as avenidas **as avenidas são** largas sabe bem espaçosa é um lugar limpo... (AC045/DE/L.251-252)
- minha vó num queria o namoro porque na época éh...**as coisa** era muito rígida... (AC034/NR/ L.42)
- c num tinha muita informação dentro de casa... porque **meu pai e minha mãe** sempre foi muito crente... (AC 036/ L 229/ NR)
- d aí surgiu a briga... mas hoje **nós** é colega normal (AC065/NE/L.29)
- e então é bem gostoso... e lá também... éh:: **eles são** bem unidos então... tem sempre... um pessoal da nossa idade... (AC46/DE/L.276-277)

A Tabela 2 mostra os resultados quantitativos da relação entre estrutura predicativa e as características formais do sujeito acima descritas.

Tabela 2 - Características formais do sujeito e pluralidade na estrutura predicativa.

Fatores	Freq.	%	PR
Sujeito não explícito (desinencial ou nulo)	39/56	69	.64
Sujeito explícito com marca formal de plural totalmente neutralizada	30/45	66	.55
Sujeito explícito: (i) com todos os elementos nominais flexionáveis marcados; (ii) com os últimos elementos com marca formal de plural; (iii) ou com a última marca neutralizada por contexto fonológico seguinte.			
Sujeito explícito com marca semântica de plural (nós e numerais isolados)	219/330	66	.52
Sujeito explícito sem as últimas marcas formais explícitas de plural ou sujeito coordenado singular	10/62	36	.39
Total	310/526	59	.595 (input)

Fonte: Elaboração própria.

Os resultados da Tabela 2 são reveladores. Supondo que o sujeito contivesse todas as marcas formais de plural, seria de esperar que o predicativo também dispusesse de alto grau de marcação de pluralidade. No entanto, essa expectativa não se confirma, já que na situação em que o sujeito contém todas as marcas formais explícitas, é provável haver marcação formal na estrutura predicativa, mas com uma frequência (PR .52) que beira o limite da neutralidade.

No entanto, uma tendência oposta se verifica nos casos em que a estrutura predicativa é precedida por um sujeito explícito desprovido das últimas marcas formais, ou por um sujeito coordenado. A reduzida incidência de marcas formais no predicativo, (PR .27), indica que o princípio atua mais no sentido de ausência de marca ativar ausência do que presença de marca ativar presença.

Esse resultado é recorrente, já que, nas situações em que o sujeito explícito contém marca semântica de plural, como pronomes de primeira pessoa do plural e numerais, a probabilidade de a estrutura predicativa ser marcada formalmente é também reduzida, como indica o PR de .39.

Já nas situações em que a estrutura predicativa for precedida por um sujeito explícito com marca formal de plural neutralizada, o comportamento quantitativamente evocado é para uma tendência discreta de marcação de pluralidade, já que próximo da neutralidade (PR .55). Como as marcas estão neutralizadas, principalmente por causa da proximidade com o verbo *ser*, não é

possível descrever, sem um aparato acústico adequado, se o informante emite ou não a marca de pluralidade. Desse modo, é como se o sintagma nominal apresentasse todas as marcas, numa análise simples, de oitiva.

É, por outro lado, curioso observar que a maior probabilidade estatística (.64) de marcação se aplica aos casos em que o sujeito é não explícito, isto é, desinencial ou nulo. Curioso, porque aponta para uma tendência oposta, como uma motivação funcional, já que, se o Princípio de Paralelismo Formal tivesse uma atuação consistente, aplicando-se de modo constante e regular, seria de esperar que sujeitos zeros, no caso dos nulos, acionassem ausência de marcas no predicativo e, inclusive, até menos marcas que os casos submetidos a outros fatores. Todavia, os dados parecem preservar a intuição de que, na ausência de marca formal no sujeito, alguma pluralidade tem que ser marcada em algum outro lugar da estrutura da oração e esse lugar acaba sendo a estrutura predicativa. Essa é uma evidência de que o comportamento dos informantes tende mais favoravelmente para a preservação de informação.

Esses resultados mostram que, na variedade riopretense, o Princípio de Paralelismo Formal, aplicável aos dados de Scherre (1988) para a variedade carioca, parece não ser a única motivação para a marcação de pluralidade na estrutura predicativa, e, portanto, parece não ter poder explanatório suficiente para explicar o processo variável em questão no Córpus Iboruna. Ou, pelo menos, o máximo que pode se inferir desses resultados é que pouco é possível deduzir deles a respeito de que princípio atua mais vigorosamente para marcar ou não marcar o constituinte em função de estrutura predicativa nos dados do Iboruna: se a Condição de Distintividade ou se o Princípio de Paralelismo Formal; mais provável é que constituam motivações em competição atuando como duas forças em sentido contrário.

Passemos, agora, a examinar outro aspecto parcial que pode indicar a atuação do Princípio de Paralelismo Formal: a relação entre pluralidade explícita na estrutura predicativa e as características formais do verbo.

A expectativa em torno dessa variável seria testar a possível atuação do Princípio de Paralelismo Formal, que se deduz da comparação entre os fatores formalmente diferentes, mas inter-relacionados. Em outros termos, a marcação de plural na estrutura predicativa em correlação com a mesma configuração flexional no verbo tem a finalidade de testar a hipótese formal de que a presença ou ausência de marca de plural no verbo implica configuração formalmente recorrente na estrutura predicativa. Uma relação inversamente proporcional nos dois casos não permitiria admitir a hipótese formal.

Nessa fotografia parcial do paralelismo formal, o fenômeno em análise é visto em sua relação com o verbo, que pode, por conseguinte, não manifestar-se

explicitamente, como (5-a), ou, alternativamente, manifestar-se explicitamente com marca formal de plural como (5-b) ou sem marca formal de plural com (5-c).

- (5) a esse pessoal eles não são dono... são case(i)ro lá... esse pessoal é primo
 do meu marido... são um amor de pessoas **[Ø] super hospitaleiro**
 (AC028/NE/L.54-56)
- b porque:: é aquela coisa né? a maioria das pessoas **são** preconceituosas
 quan/... quanto a pessoas crentes né?... (AC45/NE/L.19-20)
- c minha vó num queria o namoro porque na época éh...as coisa **era**
 muito rígida... (AC034/NR/ L.42)

Os resultados relativos a esse grupo de fatores estão expostos na Tabela 3.

**Tabela 3 - Relação entre as características formais
 do verbo e a marcação de pluralidade.**

Fatores	Freq.	(%)	PR
Verbo sem marcas	7/78	9	.14
Verbo com marcas	288/426	67	.57
Zero Verbal	15/22	68	.65
Total	310/526	59	.595 (input)

Fonte: Elaboração própria.

À primeira vista, os resultados mostram influência positiva do paralelismo formal. Se, por um lado, a presença de marcas formais explícitas no verbo ativa positivamente a formulação de estruturas predicativas com marcas formais explícitas (.57), por outro, o inverso também é verdadeiro: a ausência de marca formal no verbo é diretamente proporcional à ausência de marca no predicativo (.14).

É necessário observar, no entanto, que a incidência de marcas no predicativo com verbo flexionado tem uma frequência discreta (.57) e que, além disso, os casos de zero verbal ativam um grau muito elevado de marcação de plural na estrutura predicativa com o índice de aplicação de regra atingindo seu ápice (.65). O contexto de justaposição entre sujeito e estrutura predicativa parece reforçar a marcação de natureza estilisticamente formal da construção, fato que deve favorecer a marcação de pluralidade.

No entanto, é necessário ver esses dados de uma perspectiva mais abrangente que envolva variáveis externas principalmente em função da importância que

a escolaridade formal tem no Brasil. Dessa maneira, afirmar que a presença de marcas no predicativo, quando não houver verbo na sentença, seja apenas o resultado da atuação do Princípio de Paralelismo Formal pode não ser ainda uma conclusão definitiva.

Vale lembrar a esse respeito que, majoritariamente, as ocorrências com verbo sem marca explícita de plural dispõem de uma ou duas marcas formais no SN sujeito. É possível deduzir, portanto, nesse caso, que informação já fornecida não precisa ser repetida nem no verbo nem no predicativo, sem incorrer em qualquer risco de ambiguidade referencial, como mostram os exemplos (6a-b).

- (6) a (...) minha vó num queria o namoro porque na época é as coisa era muito **rígida**... (AC 034/NR/L.42)
- b (...) porque:: naquela época num era que nem hoje que os bastardo era **reconhecido** cê tinha que brigar muito (AC 100/NR/L.119)

Vale a pena salientar que, mesmo nos casos em que é relativamente majoritária a aplicação do Princípio de Paralelismo Formal, o efeito não tem um escopo categórico, isto é, não é reconhecido por toda a comunidade linguística, mas representa apenas uma tendência estatisticamente significativa em favor da aplicação positiva de concordância no predicativo.

O que se deve reter aqui é que a maior parte dos casos revela que a presença de marcas no sujeito aciona a pluralidade no verbo; similarmente, quando o sujeito não dispuser de todas as marcas formais, é frequente o uso de verbos também não flexionados. Essa correlação entre as marcas do sujeito e do verbo acaba por provocar efeito paralelo no predicativo.

Uma terceira dimensão do paralelismo formal, que, como grupo de fatores, foi estatisticamente selecionado pelo programa Goldvarb, tem a ver com o número de marcas precedentes no sujeito e no verbo. Não havendo marca formal de plural em todos os constituintes do SN e no verbo, é provável que a estrutura predicativa seja marcada, caso o comportamento dos informantes seja guiado pelo princípio da economia (HAIMAN, 1983) e/ou da Condição de Distintividade de Kiparsky (1972). Os resultados dessa análise estão contidos na Tabela 4.

Tabela 4 - Número de marcas precedentes no sujeito e no verbo.

Fatores	Freq.	(%)	PR
Três ou mais marcas	149/188	79	.65
Duas marcas	115/214	53	.44
Uma marca	46/124	37	.36
Total	310/526	59	.595 (input)

Fonte: Elaboração própria.

Os resultados apontam muito mais para a confirmação da atuação positiva do Princípio de Paralelismo Formal do que para o de economia ou da Condição de Distintividade, já que a incidência de marcação formal na estrutura predicativa é diretamente proporcional ao número de marcas precedentes em outros constituintes da sentença.

Há, no entanto, uma ressalva importante a fazer: o peso relativo tem significação estatística apenas para a aplicação do fator três ou mais marcas (.65). Nesse caso, constituintes precedentes com duas marcas ou com uma única marca desfavorecem a aplicação da regra de concordância na estrutura predicativa. Vale a pena indagar, portanto, por que razão uma ou duas marcas precedentes não favoreceriam a aplicação da regra nos mesmos termos em que favorecem três ou mais marcas precedentes. Haveria aí motivações em competição?

Passemos, agora, à análise de outra dimensão do paralelismo formal, que é a que se opera no discurso, uma forma de estender a atuação desse princípio do nível da sentença para o nível do discurso. Por paralelismo formal no discurso, entende-se a recursividade ou não da construção em análise. Ou o fenômeno aparece isolado, como em (7a), ou aparece em série, como o primeiro, como em (7b), ou como o segundo ou terceiro da série, podendo, nesse caso, ser precedido de outra construção cuja estrutura predicativa tem plural explícito, como (7c), ou plural zero, como (7d).

- (7) a (...) é bonito as praias são **bonitas** mas a cidade assim o centro assim... é muito... é muito poluído assim... (AC 044/DE/L 164-165)
- b (...) sessenta por cento dos brasileiros são **católicos**...Desses sessenta só cinco por cento são **praticantes**...né? (AC023//L.544-545)
- c (...) eu acho que as crianças tão muito assim... éh: **desorientadas** éh éh só pensam nisso malemá estão fican(d)o **mocinhos** já ...já tão atrás disso... (AC38/RO/L. 253-255)
- d (...)aí ela falo(u) pra mim – “não nós tá **brigado** tal cê a::cha todo mundo sabe que nós tá **brigado**... (AC019/NE/L.10-11)

Diferentemente, porém, do paralelismo entre constituintes no nível interno da sentença, a ideia subjacente a esse grupo de fatores é a de que uma estrutura predicativa formalmente marcada na construção precedente favoreceria a marcação formal na estrutura predicativa das construções seguintes; o inverso também seria verdadeiro: estruturas predicativas sem marcas formais de plural na construção precedente refreariam a aplicação da CN na estrutura predicativa das construções seguintes. Os resultados dessa análise estão contidos na Tabela 5.

Tabela 5 - Relação entre paralelismo formal no discurso e marcação de pluralidade.

Fatores	Freq.	%	PR
Estrutura predicativa precedida de estrutura predicativa com plural explícito	62/67	92	.85
Estrutura predicativa primeira de uma série	55/87	63	.54
Estrutura predicativa isolada	187/331	56	.46
Estrutura predicativa precedida de estrutura predicativa com plural zero	6/41	14	.10
Total	310/526	59	.595 (input)

Fonte: Elaboração própria.

Os resultados mostram que, na série, uma estrutura predicativa precedida de outra similar com marcas de plural recebe um alto índice de marcação formal (.85) e, ao contrário, uma estrutura predicativa precedida por estrutura congênere destituída de marcação formal terá maior probabilidade ainda de não receber marca formal (.10). Portanto, na variedade analisada, a aplicação da regra de CN parece ser severamente condicionada por esse fator.

Neutraliza-se muito esse efeito nas situações em que a estrutura predicativa for a primeira de uma série de outras construções com predicativo/particípio passivo: os dados apontam para uma tendência de marcação de pluralidade, mas, com um PR próximo de .50, tendência quantitativa que beira a neutralidade. Nas situações de predicativos isolados no discurso, fica estatisticamente mais evidente a probabilidade de não aplicação da regra de concordância nominal (.46).

Somados, os resultados desses dois grupos de fatores apontam para uma forte influência do Princípio de Paralelismo Formal, uma vez que marca formal em construção similar precedente implica outra marca formal de pluralidade na construção seguinte; o inverso também é verdadeiro: ausência de marca na construção precedente implica ausência de marca na construção seguinte, em confirmação, portanto, às descobertas de Scherre (1991) para a variedade carioca. Vejamos, todavia, o que pode nos revelar o cruzamento estatístico entre paralelismo formal no discurso e número de marcas precedentes no sujeito e no verbo, conforme disposto na Tabela 6.

Tabela 6 - Cruzamento entre paralelismo formal no discurso e número de marcas precedentes.

Marcas precedentes Paralelismo Formal	Uma marca			Duas marcas			Três ou mais marcas		
	Ap.	%	PR	Ap.	%	PR	Ap.	%	PR
Predicativo isolado	32/84	38	.41	61/121	50	.37	94/126	74	.53
Predicativo 1º de uma série	5/16	31	.26	21/39	53	.41	29/32	90	.80
Predicativo precedido de predicativo c/ plural explícito	9/10	90	.76	29/32	90	.81	24/25	96	.89
Predicativo precedido de predicativo c/ plural zero	0/14	0	n	4/22	18	.10	2/5	40	.19

Fonte: Elaboração própria.

Esse cruzamento entre dois grupos de fatores internos ou linguísticos é muito revelador no que diz respeito às motivações em jogo para o processo variável de concordância nominal, principalmente nas situações discursivas em que a estrutura predicativa aparece isolada. Era de esperar que houvesse uma relação diretamente proporcional entre o acréscimo no número de marcas precedentes e incidência positiva de marcas formais de plural nos predicativos/particípios passivos.

No entanto, essa expectativa não se confirma. É indiferente que um predicativo isolado apareça numa construção com uma marca ou duas marcas precedentes de plural, já que a probabilidade mais recorrente é a de ausência de marcas formais na estrutura predicativa. Apesar disso, nota-se que a frequência de predicativo isolado com duas marcas precedentes é inferior, em termos de peso relativo, à frequência de predicativo isolado com uma marca precedente. Esse resultado se choca frontalmente com o Princípio de Paralelismo Formal, pois quanto mais marcas precedentes ao predicativo maior deveria ser a chance de ele ser marcado também.

Outra questão instigante diz respeito à situação em que uma construção de predicativo discursivamente isolado contenha três ou mais marcas precedentes. A expectativa mais provável, em termos da aplicação do Princípio de Paralelismo Formal, é a de uma incidência muito mais significativa que .53, peso relativo mais a beirar a neutralidade que a favorecer a marcação explícita. Esse resultado só confirma a hipótese aqui defendida de que há outras motivações regendo o processo de concordância nominal em estruturas predicativas.

Já na situação em que a construção em foco aparece no discurso como a primeira de uma série, a frequência de marcação positiva aumenta progressivamente conforme o número de marcas precedentes. Fica, todavia, sem uma explicação plausível, por que as duas situações parecidas em que uma construção predicativa se identifica como isolada ou como a primeira de uma

série no discurso dispõem, em construções com três marcas precedentes, de incidências de uso tão desiguais, .53 e .80 respectivamente.

O Princípio de Paralelismo Formal aparece confirmado em duas situações opostas: a que a estrutura predicativa aparece precedida por outra construção formalmente marcada no discurso e por outra construção não formalmente marcada. É totalmente indiferente o número de marcas nessas duas situações, já que a probabilidade de manifestação de plural é alta na primeira e extremamente baixa na segunda. Não seria, no entanto, intuitivamente reveladora, a ideia de que há quem siga a regra normativa de concordância de número e há quem não a siga?

Sobre esses dados pouco confiáveis para uma generalização definitiva sobre a atuação do Princípio de Paralelismo Formal, valeria a pena cruzar os resultados dos fatores internos com um único fator externo, o grau de escolaridade. Esse procedimento se justifica não apenas porque tem peso decisivo sobre a institucionalização da norma culta no Brasil, mas também porque é outra variável independente selecionada pelo programa estatístico juntamente com essas variáveis internas.

Como uma consequência natural de que a variação e a mudança constituem propriedades constitutivas da linguagem (WEINREICH; LABOV; HERZOG, 2006), as forças centrípetas que produzem uniformidade podem, inclusive, levar alguns países a sentir a necessidade de estabelecer princípios de padronização, para determinar que variedades devam ser consideradas mais apropriadas para publicações governamentais, ensino público, uso na mídia e demais instituições públicas. Alguns países atribuem às academias a tarefa de estabelecer os critérios de padronização. Papel importante nesse âmbito tem sido atribuído à *L'Académie Française*, para determinar o padrão do francês, e *La Real Academia Española*, para o do espanhol.

A discussão desse assunto enveredou para um campo aparentemente neutro ao se entender a padronização como um procedimento de seleção entre variantes com a finalidade de obter uniformização em alguns usos mais formais da modalidade oral e da modalidade escrita. Entretanto, a noção de padronização tem aspectos simbólicos que ultrapassam muito seus aspectos puramente técnicos e pragmáticos. Mesmo nos casos em que se atribui a padronização ao trabalho das academias, as pessoas passam a acreditar que a língua descrita nas gramáticas e nos dicionários é a única variedade correta para todas as circunstâncias de interação.

A noção de correção, atribuída por algum critério padronizador, é, portanto, entendida como correta num sentido tão absoluto que produz no imaginário popular a ideia de que deixar de usar a variedade padrão equivale a deixar de usar a língua real. O conceito de variedade padrão embutido nesse imaginário é o de que há uma escala que se deve adotar como medida mínima de aceitabilidade.

A concordância nominal na estrutura predicativa é uma regra no sentido de padrão mínimo baseado em pressões fortemente normativas e a língua representada nas gramáticas mais tradicionais é considerada padrão para uma parcela da população que constitui uma elite sócio-econômica e cultural.

No entanto, a regra representada nas gramáticas de cunho normativo não se baseia na realidade da língua falada em uso na comunidade, modo de ser que instaura um forte sentimento de preconceito social em relação aos que usam uma regra da variedade não padrão, como a de concordância nominal. Essa regra é visivelmente variável, mas não para os que adotam a variedade padrão, que veem a variante não padrão como incorreta e, consequentemente, seus usuários como indivíduos socialmente estigmatizados.

Esse imaginário fornece, portanto, uma forte razão para examinar também o efeito de escolaridade como variável externa, social, em termos da relação mais ou menos proporcional entre grau de adesão aos princípios normativos e grau de escolaridade. A ideia aqui é examinar, primeiramente, o efeito da variável isoladamente para, em seguida, analisar o cruzamento dela com as variáveis internas já discutidas. Iniciemos o percurso pela análise da relação entre marcação de número na estrutura predicativa e grau de escolaridade, conforme mostra a Tabela 7.

Tabela 7 - Relação entre escolaridade e a marcação de pluralidade.

Fatores	Freq.	%	PR
1º Ciclo do Ensino Fundamental	25/77	32	.28
2º Ciclo do Ensino Fundamental	53/111	47	.37
Ensino Médio	118/186	63	.54
Ensino Superior	114/152	75	.65
Total	310/526	59	.595 (input)

Fonte: Elaboração própria.

Os resultados da Tabela 7 permitem concluir que a incidência de marcas no predicativo é diretamente proporcional ao acréscimo de grau de escolaridade, mas com uma nítida linha divisória, em termos estatísticos, entre estar no primeiro e no segundo ciclo do ensino fundamental e estar no ensino médio e no superior. A primeira situação praticamente desfavorece o uso de marcas de plural e a segunda situação favorece uma probabilidade maior de aplicação da regra de concordância de número e em escala ascendente.

A razão da maior probabilidade de adoção da regra padrão de concordância pelos informantes com nível médio e superior de ensino está certamente arraigada no contato maior que esses indivíduos mantêm com a atividade

normativa e que se aplica como uma injunção prescritiva rigorosa no sistema escolar. Uma ressalva necessária: embora seja elevada a incidência de aplicação positiva, a frequência estatística de .65 é uma demonstração clara de que, nos predicativos, a regra não é apenas variável para os indivíduos de ensino superior, mas também que está muito distante do estatuto de regra categórica.

O Princípio de Paralelismo Formal é claramente definível como uma motivação propriamente interna e, como tal, a aplicação da regra de concordância nominal passaria, então, a ser um procedimento meramente mecânico, determinado por mecanismos puramente estruturais. Como, todavia, acabamos de ver que, para a variedade riopretense, a escolaridade de ensino médio representa uma fronteira significativa a separar os que tendem à marcação positiva dos que tendem à marcação negativa, parece muito instrutivo verificar a interação entre uma motivação linguística como o Princípio de Paralelismo Formal e uma motivação extralingüística, aqui restrita à escolaridade.

Uma hipótese que logo se coloca é se o paralelismo formal – marca levando a marca e ausência levando a ausência – não poderia representar, ao menos para a variedade investigada, uma implicação social cuja distribuição refletiria, na realidade, uma relação proporcional do seguinte tipo: o comportamento linguístico dos indivíduos de 1º e 2º ciclos do ensino fundamental favoreceria ausência de marcas, enquanto o dos indivíduos com ensino médio e superior, a presença de marcas. Continuando, então, o percurso, é necessário, refletir agora sobre a relação entre motivações de natureza diversa competindo entre si. Além disso, o cruzamento de fatores internos com externos tem a finalidade de verificar a possível influência de escolaridade sobre o Princípio de Paralelismo Formal.

O cruzamento com escolaridade limita-se aqui a dois parâmetros internos do paralelismo formal, da perspectiva intraoracional, a saber, características formais do verbo e número de marcas precedentes, e a um parâmetro interno da dimensão interoracional, que é o paralelismo formal no discurso.

E o primeiro passo dessa trajetória está na distribuição entre uma dimensão do paralelismo formal – características formais do verbo e escolaridade, exposta na Tabela 8.

Tabela 8 - Características formais do verbo e escolaridade.

Características formais do verbo	Zero verbal			Verbo marcado			Verbo não marcado			
	Escolaridade	Ap.	%	PR	Ap.	%	PR	Ap.	%	PR
1º ciclo EF		2/2	100	n	23/54	42	.32	0/21	0	n
2º ciclo EF		3/6	50	.39	48/81	59	.48	2/24	08	.05
Ensino Médio		7/10	70	.60	109/152	71	.62	2/24	08	.05
Ensino Superior		3/4	75	.65	108/139	77	.69	3/9	33	.24

Fonte: Elaboração própria.

Os resultados confirmam haver uma relação diretamente proporcional entre escolaridade e presença de marcas, mas apenas nas situações em que o verbo aparece marcado positivamente. Embora o resultado seja o mesmo nas situações em que o fenômeno em análise é precedido por um zero verbal, a incidência da aplicação é muito pouco significativa para ser possível tirar conclusões definitivas.

Como tendência, pode-se notar que os dados, vistos sob uma orientação vertical de leitura, mostram que as situações de zero verbal e de verbo com marca de plural representam uma nítida linha divisória a separar o comportamento dos informantes que dispõem de ensino fundamental dos que dispõem de ensino médio e superior. Aqueles apresentam um uso reduzido de marcas formais em formas de estrutura predicativa precedidas por verbos marcados, conforme atestam os exemplos (8a-b), enquanto estes apresentam comportamento inverso, com uma incidência elevada de marcação de plural nas estruturas predicativas, conforme atestam os exemplos (9a-b).

- (8) a eu acho que hoje as meninas ficam **mocinha** muito depressa né?
(AC 096/RO/L 403)\)
- b (...) Eles que num num são **realista** num são **verdade(i)ro**... mas
são **muito falso** (AC 029/RO/L 185)
- (9) a (...) a sala de aula de terce(i)ra e quarta séries são salas onde não
existe alfabeto... porque os alunos já estão **alfabetizados**... (AC 88/
DE/L 335-336)
- b (...) que:: na época é a:: vaga era pra uma pessoa... mas das três
pessoas... duas foram consideradas **aptas** (AC 114/NE L 55-56)

É, todavia, um tanto diferente o comportamento dos informantes nas situações em que o fenômeno em análise é precedido por um verbo não marcado

em número. Poucos informantes da amostra, até mesmo os dotados de ensino médio e superior, mostram tendência pela marcação positiva. Essa situação específica, na qual a diferença de grau de escolaridade não parece exercer influência, parece apontar para a independência da gramática em relação aos fatores externos, mais especificamente, para confirmação do Princípio de Paralelismo Formal.

No entanto, também esses resultados precisam ser relativizados, já que o número de ocorrências, extremamente reduzido, pode mascarar uma interpretação mais qualitativa, exatamente como deve ser relativizada a aparente influência de escolaridade sobre a situação de predicativos precedidos por zero verbal.

Passemos agora para o outro cruzamento entre número de marcas precedentes e escolaridade, que aparece na Tabela 9.

Tabela 9 - Cruzamento entre marcas precedentes e escolaridade.

Marcas precedentes Escolaridade	Uma marca			Duas marcas			Três ou mais marcas		
	Ap.	%	PR	Ap.	%	PR	Ap.	%	PR
1º ciclo EF	2/23	08	.14	16/39	41	.29	7/15	46	.34
2º ciclo EF	6/33	18	.20	20/44	45	.37	27/34	79	.56
Ensino Médio	22/45	48	.38	42/74	56	.52	54/67	80	.66
Ensino Superior	16/23	69	.68	37/57	64	.57	61/72	84	.76

Fonte: Elaboração própria.

Essa disposição de dados fornece algumas deduções relevantes. Em primeiro lugar, os resultados revelam um aumento progressivo de marcas formais, que contempla uma relação diretamente proporcional ao acréscimo de grau de escolaridade. Em segundo lugar, quanto maior o número de marcas precedentes, tanto maior a probabilidade de a estrutura predicativa receber marca formal de pluralidade. Em terceiro lugar, é certo que essas correlações são motivadas por escolaridade.

Prosseguindo na trajetória expositiva, passemos agora a ver a relação entre a dimensão discursiva do paralelismo formal e o grau de escolaridade, cujos resultados aparecem distribuídos na Tabela 10.

Tabela 10 - Paralelismo formal no discurso e escolaridade.

Escolaridade Paralelismo formal no discurso	1º ciclo EF			2º ciclo EF			Ensino Médio			Ensino Superior		
	Ap.	%	PR	Ap.	%	PR	Ap.	%	PR	Ap.	%	PR
Predicativo isolado	11/42	26	.18	28/41	40	.30	76/123	61	.50	72/93	77	.68
Predicativo primeiro de uma série	7/16	43	.33	12/19	63	.52	18/27	66	.55	18/25	72	.61
Pred. precedido de pred. c/ plural explícito	6/7	85	.79	13/13	100	n	22/23	95	.93	21/24	87	.81
Predicativo precedido de pred.c/ plural zero	1/12	08	.05	0/8	0	n	2/11	18	.12	3/10	30	.21

Fonte: Elaboração própria.

Na situação em que a construção predicativa está isolada no discurso ou aparece pela primeira vez em uma série, o comportamento linguístico dos informantes dos dois ciclos do Ensino Fundamental tende para um uso reduzido de adesão às normas da variedade padrão, enquanto o dos informantes de Ensino Médio e Superior tende para a adesão às normas. Enfim, estar no Ensino Superior é condição favorável para o emprego da flexão de plural nesse tipo de estrutura.

Um fato interessante a observar aqui é que, contrariamente aos informantes de outros níveis de escolaridade, os do Ensino Superior tendem para uma frequência maior de uso nas situações de estrutura predicativa isolada do que nas de estrutura predicativa primeira de uma série. O que há de interessante nessa distribuição está justamente no fato de compartilharem a mesma estrutura frasal, cuja configuração é SN sujeito, verbo e predicativo, como se vê nos exemplos (10a-b), com predicativos isolados, e (11a-b), com predicativos primeiro de uma série, respectivamente.

- (10) a a sala de aula de terce(i)ra e quarta séries são salas onde não existe alfabeto... porque os alunos já estão **alfabetizados**... (AC 88/DE/L 335-336)
- b (...) e ela abandonô(u) ela foi embora não aguentô(u)... veja bem o quanto que... que esses filhos e esse marido foi **ruim** pra ela... (AC 110/NR/L 162-163)

- (11) a minhas tia são muito católica em José Bonifácio mas são assim beata mesmo sabe?... se soubé(r) de um negócio desse ela me mata... (AC 100/RO/L 380)
- b (...) ele tem:... os bancos dele são diferentes também... foram trocados... não são originais... pra ficá dá um aspecto mais esporte... (AC 049/DE/L 132-133)

Vistos de uma perspectiva horizontal, os dados da tabela 10 apontam para uma correlação diretamente proporcional entre acréscimo de escolaridade e acréscimo no uso de marcas formais.

Outro aspecto notável está nos resultados referentes às situações impermeáveis à diferença de escolaridade, que são as de estrutura predicativa precedida de estrutura predicativa com plural explícito e com marcação zero. Na primeira situação, o comportamento é, de certa forma, linear, cuja incidência recai sobre pesos relativos acima de .79, com diferenças reduzidas entre os agrupamentos sociais. Na segunda situação, também os resultados se mostram indiferentes a diferenças de escolaridade, mas a incidência de adesão às formas normativamente reguladas é extremamente baixa por agrupamento social, não ultrapassando a marca de .21 para os informantes no topo do sistema de ensino.

De qualquer modo, os índices de aplicação são extremamente baixos para que seja confiável uma generalização do tipo marca leva a marca e ausência leva a ausência, que confirmaria o princípio do paralelismo formal.

Considerações finais

Ao longo deste trabalho, estabelecemos como objetivo desvendar se a marcação de número nas estruturas de SNs e SAs predicativas poderia ser explicada por motivações tipicamente formais, ou tipicamente funcionais, ou ainda, pela devida interação entre esses dois tipos de motivações, competindo entre si.

A análise dos resultados mostra, por um lado, que os princípios funcionais, como o princípio de economia (HAIMAN, 1983) e a Condição de Distintividade (KIPARSKY, 1972), atuam na marcação de pluralidade, mas não têm força suficiente para governar todo o processo na variedade estudada. Mostram, por outro, que as motivações formais, gerenciadas pelo Princípio de Paralelismo Formal, também exercem influência significativa na marcação explícita de pluralidade nos dois fenômenos. Similarmente com o que ocorre com as motivações funcionais,

também as motivações formais não atuam categoricamente de modo a regerem solitariamente o fenômeno como têm mostrado os estudos de outras variedades.

A atuação do Princípio de Paralelismo Formal, na variedade estudada, sofre severa restrição da condição social dos usuários da variedade, especialmente, a formação escolar; desse modo, ao menos no que tange à variedade investigada, não é possível dar a esse princípio o poder explanatório que lhe atribui Labov (1994), isto é, o de que a pluralidade tem uma motivação essencialmente interna, derivada da própria configuração formal do sistema gramatical, que governa, neste caso específico, a variedade riopretense. Por essa razão, a explicação mais plausível para a marcação de pluralidade nos predicativos é a de que há motivações em competição, nos termos de Du Bois (1985) e é a marcação de pluralidade o “bem limitado”, pelo qual forças múltiplas, as motivações formais ou internas e funcionais ou externas, competem entre si.

CAMACHO, R. G.; SALOMÃO, M. H. Competing motivations in sociolinguistic variation: the plural in NP and AP predicates in the São José do Rio Preto Spoken Variety. *Alfa*, São Paulo, v.56, n.3, p.1055-1087, 2012.

■ *ABSTRACT. THE MAIN PURPOSE OF THIS PAPER IS TO SUBJECT TO A SOCIOLINGUISTIC QUANTITATIVE-BASED TREATMENT SOME DATA OF VARIABLE PLURAL MARKING IN PREDICATE NPS AND APS EXTRACTED FROM A SAMPLE OF SPOKEN LANGUAGE COLLECTED IN THE REGION OF SÃO JOSÉ DO RIO PRETO. A MORE SPECIFIC AIM IS TO EXAMINE WHETHER THE BEST EXPLANATION FOR VARIABLE MARKING OF PLURAL CAN BE BASED ON EITHER EXCLUSIVELY FORMAL OR EXCLUSIVELY FUNCTIONAL CONSTRAINTS, OR PERHAPS BASED ON THE INTERACTION BETWEEN BOTH OF THEM, WHICH WOULD CONSIST IN TRULY COMPETING MOTIVATIONS (DU BOIS, 1985). THE RESULTS SHOW THAT NEITHER FUNCTIONAL NOR FORMAL MOTIVATIONS ALONE RULE THE PHENOMENON, WHICH HAPPENS TO BE STRONGLY CONSTRAINED BY A SPECIFIC EXTERNAL FACTOR: education level. That is why the most likely explanation for plural marking in predicates is that there are competing motivations, which is a term used by Du Bois (1985), and plural markings are the “limited means”, for which multiple forces, the formal or internal and the functional or external motivations, compete among themselves.*

■ *KEYWORDS:* Functionalism. Formalism. Plurality. Nominal agreement. Predicate.

REFERÊNCIAS

BRAGA, M. L. *A concordância de número no sintagma nominal no Triângulo Mineiro*. 1977. 88f. Dissertação (Mestrado em Linguística) – Departamento de Letras, Pontifícia Universidade Católica, Rio de Janeiro, 1977.

CAMACHO, R. G. *Motivações formais e funcionais da marcação de pluralidade no português falado na região de São José do Rio Preto*. 2008. 110f. Relatório Científico FAPESP, São José do Rio Preto, 2008.

_____. O formal e o funcional na teoria variacionista. In: RONCARATI, C.; ABRAÇADO, J. (Org.). *Português brasileiro: contato linguístico, heterogeneidade e história*. Rio de Janeiro: Viveiros de Castro Editora, 2003. p.55-65.

DIAS, J. F. V. *A concordância de número nos predicativos e nos participios passivos na fala da região sul*: um estudo variacionista. 1996. 124f. Dissertação (Mestrado em Comunicação) - Universidade Federal de Santa Catarina, Florianópolis, 1996.

DIK, S. C. *The theory of functional grammar*: part I: the structure of the clause. 2nd rev. ed. Berlin: Mouton de Gruyter, 1997. (Série Functional grammar series ; 9).

DU BOIS, J. W. Competing motivations. In: HAIMAN, J. *Iconicity in syntax*. Amsterdam: John Benjamins, 1985. p.342-365.

FIAMENGUI, A. H. R. *A marcação de pluralidade no SN na fala e na escrita de adolescentes da região de São José do Rio Preto*. 2011. 144f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2011.

GARCÍA, E. C. Shifting variation. *Lingua*, Amsterdam, v.67, p.189-224, 1985.

GONÇALVES, S. C. L. *O português falado na região de São José do Rio Preto*: constituição de um banco de dados anotado para o seu estudo. Relatório científico parcial III apresentado à FAPESP, Universidade Estadual Paulista, São José do Rio Preto, 2007. Disponível em: <<http://www.iboruna.ibilce.unesp.br/histórico/relatorio3>>. Acesso em: 16 out. 2011.

GONÇALVES, S. C. L.; TENANI, L. E. Problemas teórico-metodológicos na elaboração de um sistema de transcrição de dados interacionais: o caso do projeto ALIP Amostra Linguística do Interior Paulista. *Gragoatá*, Niterói, v.25, p.165-184, 2008.

GUY, G. *Linguistic variation in Brazilian Portuguese*: aspects of the phonology, syntax and language history. 1981. 391f. Tese (Doutorado em Linguística) – Universidade da Pennsylvania, Philadelphia, 1981.

HAIMAN, J. Iconic and economic motivation. *Language*, Baltimore, v.59, p.781-819, 1983.

HOCHBERG, J. G. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language*, Baltimore, v.62, n.3, p.609-621, set. 1986.

KIPARSKY, P. Explanation in phonology. In: PETERS, S. (Ed.) *Goals of linguistic theory*. Englewood Cliffs: Prentice-Hall, 1972. p.189-227.

LABOV, W. *Principles of linguistic change*. Oxford: Basil Blackwell, 1994. (Internal factors; v.1).

_____. The overestimation of functionalism. In: DIRVEN, R.; FRIED, V. (Ed.). *Functionalism in linguistics*. Amsterdam: John Benjamins, 1987. p.311-32.

_____. Where does the linguistic variable stop? a response to Beatriz Lavandera. *Working Papers in Sociolinguistics*, Austin, v.44, 1978.

_____. *Sociolinguistic patterns*. Pennsylvania: University of Pennsylvania Press, 1972.

LAVANDERA, B. R. Where does the sociolinguistic variable stop? *Language in Society*, Cambridge, v.7, p.171-82, 1978.

LEFEBVRE, C. Variation in plural marking: the case of cuzco quechua. In: SANKOFF, D.; CEDERGREN, H. (Ed.). *Variation Omnibus*. Canadá: Linguistic Research, 1981. p.73-84.

POPLACK, S. Delection and disambiguation in Puerto Rican Spanish. *Language*, Baltimore, v. 56, n. 2, p.371-385, 1980.

ROMAINE, S. The status of variable rules in sociolinguistic theory. *Journal of Linguistics*, Cambridge, v.17, p.93-119, 1981.

RONCARATI, C. *Banco de dados interacionais*. Rio de Janeiro: Universidade Federal do Rio Janeiro, 1996.

RUBIO, C. F. *Padrões de concordância verbal e de alternância pronominal no português brasileiro e europeu: estudo sociolinguístico comparativo*. 2012. 392f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2012

_____. *A concordância verbal na região noroeste do Estado de São Paulo*. 2008. 152f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008.

SALOMÃO, M. H. *A variação de pluralidade nas estruturas predicativas da variedade falada na região de São José do Rio Preto*. 2010. 162f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010.

SANKOFF, D.; TAGLIAMONTE, S.; SMITH, E. *Goldvarb X*: a variable rule application for Macintosh and Windows. University of Toronto, Department of Linguistics, 2005. Software. Disponível em: <<http://individual.utoronto.ca/tagliamonte/goldvarb.htm>>. Acesso em: 16 out. 2011.

SAUSSURE, F. *Curso de linguística geral*. Tradução de Antônio Chelini, José Paulo Paes e Izidoro Blkstein. São Paulo: Cultrix, 1977.

SCHERRE, M. M. P. Aspectos da concordância de número no português do Brasil. *Revista Internacional de Língua Portuguesa*, Pelotas, v.12, p.37-49. 1994.

_____. A concordância de número nos predicativos e nos participios passivos. *Organon*, Porto Alegre, v.5, n.17, 1991.

_____. *Reanálise da concordância nominal em português*. 1988. 555f. Tese (Doutorado em Linguística) - Faculdade de Letras, Universidade Federal Rio de Janeiro, Rio de Janeiro, 1988.

TERREL, T. Functional constraints on deletion of word-final /s/ in Cuban Spanish. *Berkeley Linguistics Society*, Berkeley, v.1, p.431-7, 1975.

WEINER, E. J.; LABOV, W. Constraints on the agentless passive. *Journal of Linguistics*, Cambridge ,v.19, p.29-58, 1983.

WEINREICH, U.; LABOV, W.; HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. Tradução de Marcos Bagno e Carlos Alberto Faraco. São Paulo: Parábola, 2006.

Recebido em março de 2012

Aprovado em agosto de 2012

DA FALA PARA A LEITURA: ANÁLISE VARIACIONISTA

Dermerval da HORA *

Maria de Fátima S. AQUINO**

- RESUMO: Os estudos variacionistas no Brasil têm focalizado, em sua maioria, os níveis linguísticos fonológico e sintático. Os vários projetos que se pautam em dados coletados em diferentes comunidades (VARSUL, PEUL etc.) procuram abordar seus resultados, utilizando todo o aparato teórico-metodológico que a Sociolinguística Variacionista tem colocado à disposição. Com base em resultados obtidos ao longo da implementação do Projeto Variação Linguística no Estado da Paraíba – VALPB (HORA, 1993), que tem seu *corpus* estratificado segundo o sexo, a faixa etária e os anos de escolarização, e buscando lançar um olhar para um novo conjunto de dados, dessa vez voltados para a leitura, este artigo contempla a variação linguística na leitura oral de alunos da primeira fase do ensino fundamental, tendo como processos de análise: ditongação, monotongação e apagamento do “d” no grupo –ndo. Os dados analisados apontam uma forte influência da variante dialetal do aluno na leitura oral. É essa relação entre o que o aluno fala e o que o aluno lê que motiva a busca por propostas de análise que vão além daquilo que a escola preconiza.

- PALAVRAS-CHAVE: Sociolinguística. Variação. Leitura e ensino.

Introdução

Em 1993, foi iniciado o Projeto Variação Linguística no Estado da Paraíba – VALPB, apoiado pelo CNPq, tendo como principal objetivo traçar o perfil linguístico do falante paraibano, considerando os aspectos fonológicos, gramaticais e textuais. A estratificação social do projeto contemplou as seguintes restrições: sexo, anos de escolarização e faixa etária. Ao todo, foram entrevistados 60 informantes, por cerca de 45 a 60 minutos de gravação com cada um deles.

Os dados, depois de transcritos e armazenados eletronicamente e também editados sob forma impressa, serviram de base para análises as mais variadas possíveis. Foram inúmeros os trabalhos que neles se pautaram. Teses, dissertações, artigos, apresentações em eventos, todas essas modalidades se valeram do *corpus* para promover discussões e contribuir para que se

* UFPB – Universidade Federal da Paraíba. João Pessoa – Paraíba – Brasil. 58059-970 - ho_ra@hotmail.com

** UFPB – Universidade Federal da Paraíba. João Pessoa – Paraíba – Brasil. 58059-970 - fatimaaquinouepb@yahoo.com.br

compreenda como dados de uma comunidade específica podem auxiliar na compreensão da teoria linguística como um todo.

Com o avanço dos estudos sociolinguísticos, novas perspectivas de aplicação desses estudos foram surgindo, em alguns casos de forma interdisciplinar com outras correntes teóricas da linguística contemporânea. Desta forma, os estudos que, inicialmente, eram de caráter apenas descritivo, isto é, tinham como objetivo descrever o perfil da comunidade de fala, passaram a ter uma nova aplicação teórica ou prática, para atender às novas demandas das pesquisas científicas advindas dos novos conhecimentos adquiridos com as “descobertas” sociolinguísticas.

Neste artigo, serão analisados três processos de formas variantes na oralização do texto escrito e sua interferência no processo de leitura. Todos os três trabalhos resultaram de pesquisa com dados do VALPB, tendo como produto final a dissertação de suas autoras. A análise toma por base processos da fala espontânea que foram estudados por Aquino (2004), Silva (2004) e Martins (2004). Aquino (2004), pesquisando sobre o processo de ditongação em formas como “fal[j]s”, “tre[j]s”, constata que os falantes menos escolarizados são os que mais aplicam a regra. No estudo de Silva (2004), a escolaridade exerce forte influência no processo de monotongação, como em “f[e]ra” para “feira”, “[o]tro” para “outro”. A autora constata que quanto mais alto o nível de escolaridade menor é o uso da forma monotongada. Martins (2004), observando o comportamento da variável anos de escolarização, constata que o apagamento da oclusiva ‘d’ no grupo –ndo, como em “falano”, está em uma relação inversamente proporcional ao nível de escolaridade, isto é, à medida que aumenta o nível de escolaridade, diminui o uso da regra do apagamento.

Para dar conta da proposta, assim está estruturado o artigo: na seção 1, será apresentada uma visão sobre a leitura no cenário da variação linguística; na seção 2, será apresentada a metodologia utilizada; na seção 3, uma visão geral dos fenômenos controlados durante a leitura e, em seguida, uma análise de cada um deles.

A leitura no cenário da variação linguística

As primeiras contribuições do estudo da variação para a educação foram trazidas pelas pesquisas de Labov sobre o inglês dos negros nos Estados Unidos. Os resultados de suas pesquisas refutaram a visão corrente, à época, de que as crianças falantes dessa variedade do inglês apresentavam deficiências de habilidades linguísticas. Tais estudos contribuíram para a valorização dos dialetos falados pelos grupos minoritários.

No campo da leitura, ressalta-se a importante contribuição das pesquisas de Labov e Baker (2001, 2003). Os autores analisam o processo de leitura oral de

crianças das classes minoritárias (afro-americanos, latinos e brancos) e mostram como a língua da comunidade dessas crianças influencia a leitura oral. Os autores diferenciam os erros de leitura (decodificação errada da palavra) dos supostos erros decorrentes da influência do dialeto do aluno. Eles defendem que os perfis de erro de leitura de grupos étnicos são, de certo modo, previsíveis da estrutura de suas línguas maternas e que métodos de leitura que levam em conta as diferenças étnicas, sociais e econômicas, terão sucesso, reduzindo a diferença no aprendizado de leitura.

No Brasil, destacam-se as pesquisas de Bortoni-Ricardo (2004) em que enfatiza o monitoramento da fala para a análise do português brasileiro no ambiente escolar e em outros ambientes sociais. Bortoni-Ricardo (2005) denomina de Sociolinguística Educacional todas as propostas e pesquisas que tenham por objetivo contribuir para o aperfeiçoamento do processo educacional.

Destacam-se, também, as importantes contribuições das pesquisas de Mollica (1998, 2003) com análise de formas variantes da fala e sua possível influência na escrita de alunos da educação básica.

Neste artigo, será investigada a influência de formas variantes da fala na leitura de alunos da primeira fase do ensino fundamental.

A variação em pauta: metodologia

Os dados aqui apresentados são de uma pesquisa realizada em uma escola pública no município de Guarabira-PB. O *corpus* da pesquisa é composto por leitura de um texto realizada por 30 (trinta) alunos do 3º ao 5º ano do ensino fundamental pertencentes à classe socioeconômica baixa e idade variando entre 08 e 11 anos. Crianças em idade escolar compatível, em média, com o ano cursado. Aos alunos, foi apresentado um texto e solicitada a leitura individual em voz alta; antes de proceder à leitura para gravação, o aluno tinha um primeiro contato com o texto. Assim que ele sinalizasse que estava pronto, tinha início a gravação da leitura em voz alta.

Como o procedimento metodológico guiou-se por uma concepção sociolinguística, mesmo estando em uma situação de linguagem que exigia certo monitoramento da fala, os alunos foram estimulados a fazerem uma leitura o mais natural possível.

A leitura gravada foi transcrita ortograficamente e, em seguida, foi feita a codificação e quantificação dos dados para a análise. Foram estabelecidos, previamente, alguns fenômenos variáveis para serem analisados: monotongação, ditongação, apagamento do 'd' no grupo -ndo, apagamento do rótico, rotacismo, simplificação do grupo consonantal etc. Apenas os três primeiros serão aqui

analisados. Em seguida, os dados codificados e quantificados foram submetidos ao tratamento estatístico por meio do pacote de programa estatístico R (R DEVELOPMENT CORE TEAM, 2007), para que pudesse ser obtida a frequência de ocorrência de cada forma variante em cada ano pesquisado.

Em uma análise inicial da leitura dos alunos, considerou-se que todos se comportaram de forma espontânea na realização do evento de leitura. No início, alguns demonstraram certo desconforto para com o momento, mas logo em seguida, ao término do primeiro parágrafo, por exemplo, já se comportavam de forma mais espontânea, sem prestar muita atenção à sua linguagem. A atenção, nesse momento, já estava mais voltada para o conteúdo do texto. O envolvimento com o conteúdo assegura a continuidade do fluxo de fala (LABOV, 2008).

Análise dos dados: ditongação, monotongação e apagamento do 'd' no grupo –ndo

Na quantificação dos fenômenos variáveis, foi observado que, em todos os anos pesquisados, ocorreram todos os tipos de fenômenos estabelecidos previamente, muitos dos quais não serão aqui tratados. Os resultados quantitativos apresentados nos revelaram aspectos interessantes dos fenômenos variáveis pesquisados. Na tabela 1 tem-se a porcentagem de ocorrência de cada forma variante por ano de escolaridade.

Tabela 1 - Porcentagem de ocorrência de formas variantes por ano de escolaridade

<i>Fenômeno variável</i>	Ano de escolaridade		
	3º	4º	5º
Ditongação	28,89%	32,63%	34,21%
Monotongação	23,08%	31,54%	68,46%
Apagamento /d/ -ndo	3,33%	8,89%	2,22%

Fonte: Elaboração própria.

Ao serem comparadas as formas variantes controladas, percebe-se que o percentual de ocorrência mais alto está relacionado à monotongação e no quinto ano. O processo de ditongação tem comportamento muito próximo para os três anos, e é o apagamento do 'd' no grupo –ndo que tem percentuais mais baixos.

Comparando a porcentagem de ocorrência de alguns fenômenos entre os anos de escolaridade, percebe-se que o 3º ano é o que apresenta menor porcentagem de

ocorrência nos seguintes fenômenos: Ditongação – 3º ano (28,89%), 4º ano (32,63%) e 5º ano (34,21%); Monotongação – 3º (23,08%), 4º ano (31,54%) e 5º ano (68,46%).

Esses valores apresentados para os ditongos e monotongos vão de encontro às asserções apresentadas por Aquino (2004) e Silva (2004) sobre a relação entre nível de escolaridade e o uso da forma com ditongação e monotongação, respectivamente, com análise de língua falada. Em seus estudos, as autoras constataram que quanto maior o nível de escolaridade menor o uso da forma monotongada e da forma com ditongo. Então, por que na leitura oral ocorre o processo inverso e o menor ano de escolaridade é o que apresenta a menor frequência de uso?

Relacionando os resultados quantitativos às observações do comportamento de leitura dos alunos, infere-se que a explicação para a menor ocorrência nos alunos de 3º ano reside no fato de grande parte deles, por ainda não terem um domínio eficiente de leitura, fazer uma leitura silabada, em que é feita a decodificação da palavra por meio da relação letra/som. Nesse processo, o aluno pronuncia exatamente todos os fonemas que compõem a palavra, o que inibe o apagamento ou acréscimo de algum elemento.

Comparando, ainda, as porcentagens de ocorrência das formas variantes por ano de escolaridade, observa-se que no 5º ano houve a menor ocorrência do apagamento do 'd' no grupo -ndo. Os resultados encontrados para o apagamento do /d/ no grupo -ndo mostram que os alunos, na leitura oral, usam pouco a forma com apagamento, o que confirma, em parte, os achados de Martins (2004) com análise de dados de língua falada, uma vez que, para a autora, a regra de apagamento diminui à medida que a escolaridade avança. Nos dados aqui analisados não há essa relação inversamente proporcional, pois a maior ocorrência da forma com apagamento se verifica no 4º ano.

Realizada essa análise geral dos resultados, será feito um detalhamento a seguir.

Monotongação

O sistema ortográfico do português do Brasil, representado nas gramáticas normativas e nos manuais de ensino da língua, reconhece como ditongo o encontro de uma vogal e uma semivogal na mesma sílaba. Funcionam como semivogais o **i** e o **u**, que são representados fonologicamente por /y/ e /w/. No entanto, nem todo vocábulo que comporta um ditongo na escrita tem o mesmo comportamento na fala espontânea. Assim, em vocábulos como **faixa**, **feira**, **vários**, que o sistema ortográfico considera portadores de ditongo, na língua falada, esses mesmos vocábulos podem comportar uma vogal simples, o monotongo, <faxa>, <fera>, <varos>; já outros, como **leite**, **jeito**, **pai**, por exemplo, não permitem o

apagamento da semivogal. Contudo, essa particularidade do ditongo quase não é discutida nas gramáticas normativas.

Cunha e Cintra (1985, p.49, grifo do autor), em sua *Nova gramática do português contemporâneo*, fazem referência ao apagamento da semivogal, na seguinte observação:

Nem na pronúncia normal de Portugal nem na do Brasil se conserva o antigo ditongo [ow], [...]. Na pronúncia normal reduziu-se a [o], desaparecendo assim a distinção de forma como *poupa / popa, boubá / boba*.

Câmara Junior (1986, p.170, grifo do autor) define a monotongação enfatizando seu caráter fonético:

Mudança fonética que consiste na passagem de um ditongo a uma vogal simples [...]. Para pôr em relevo o fenômeno da monotongação chama-se, muitas vezes, MONOTONGO à vogal simples resultante, principalmente quando a grafia continua a indicar o ditongo e ele ainda se realiza numa linguagem mais cuidadosa. Entre nós há, nesse sentido o monotongo ou /ô/, em qualquer caso, e ai /a/, ei /é/ diante de uma consoante chiente; exs.: (p)ouca, como (b)ooca, (c)caixa, como acha, (d)deixa), como fecha.

O processo de monotongação, aqui referido, é resultante, portanto, do apagamento das semivogais [y] e [w], o que ocasiona a redução do ditongo a vogal simples, o monotongo. O autor ressalta a ocorrência da monotongação do ditongo /ow/ em qualquer contexto e de /ay/ e /ey/ diante de chiente. A essas ocorrências foi acrescentada a deste último, diante de outra consoante como em *dinhero* para **dinheiro**. Como, também, a ocorrência da monotongação com outros ditongos: / y/ *munto* para **muito**, /yo/ *varos* para **vários**. Essa tendência a reduzir os ditongos é atestada no português ao longo de sua história, conforme mostram os manuais e gramáticas históricas (CÂMARA JUNIOR, 1976; COUTINHO, 1978).

No português do Brasil a monotongação tem sido abordada por vários estudiosos: Bisol (1989, 1994), Paiva (1996a, 1996b), Silva (1997, 2004), Mollica (1998), tanto na linguagem oral quanto na escrita. No que se refere à linguagem falada, o uso variável da forma monotongada é observada na fala de diversas comunidades brasileiras, fato verificado através dos vários estudos de cunho variacionista em diversas regiões do País (PAIVA, 1996a, 1996b; SILVA, 1997, 2004). Dentre esses interessa de forma particular o trabalho de Silva (1997, 2004), uma vez que a comunidade analisada pela autora se situa no mesmo Estado do presente estudo. Portanto, os falantes das duas comunidades analisadas possuem características linguísticas muito próximas, o que favorece a comparação entre os resultados obtidos na análise da linguagem falada de forma espontânea (SILVA, 1997, 2004) e neste estudo, com a leitura oral.

O trabalho de Mollica (1998), assim como a presente pesquisa, está voltado para as questões de variação no contexto de sala de aula, isto é, para o processo de ensino e aprendizagem da língua. Em seu estudo, a autora analisou o processo de monotongação relacionado à aquisição da língua escrita, buscando contribuir para uma melhor compreensão da influência da linguagem oral sobre a escrita.

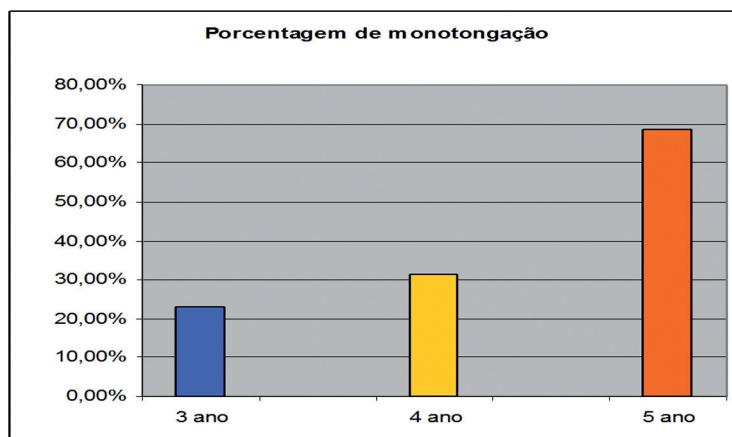
A autora dividiu os alunos pesquisados em dois grupos: um que recebeu instruções sobre as regras de monotongação na fala e outro que não as recebeu. Com base nos resultados, ela apresenta alguns pontos conclusivos, entre os quais destacam-se: (1) os grupos que receberam instrução sobre a possível interferência da fala na escrita apresentaram um índice menor de monotongação, “[...] o que leva a supor que a consciência explícita por parte do falante acerca da relação fala/escrita concorre para melhorar o desempenho no processo de alfabetização.” (MOLLICA, 1998, p.79); (2) turmas de alfabetização e de 1^a série ainda não são sensíveis a uma orientação explícita sobre a influência da fala na escrita, uma vez que nessas turmas “[...] o alfabetizando está assimilando ainda a relação fonema/grafema, processo suficientemente complexo para introduzir-se mais informação sobre fala/escrita.” (MOLLICA, 1998, p.79). (3) A 2^a série é considerada o momento adequado para se introduzir essa orientação sobre a relação fala/escrita, pois é a partir dessa série que os alunos começam a entender as diferenças entre uma e outra forma de linguagem. (4) Observando os ditongos /ey/ e /ow/ constatou-se que a correção para o ditongo /ow/ é mais lenta que a do ditongo /ey/, fato que pode ser explicado em virtude de “[...] quanto mais afetada na fala a mudança, maior resistência à instrução, aplicada como estratégia pedagógica em sala de aula no processo de letramento.” (MOLLICA, 1998, p.80).

Estudos com dados de língua falada concluem que o apagamento da semivogal do ditongo /ow/ configura uma mudança totalmente implementada no sistema (PAIVA, 1996a), ou um estado de mudança praticamente consumado (SILVA, 2004).

Nos dados analisados para este artigo, a supressão da semivogal apresenta comportamento diferente dos apresentados por Paiva (1996a, 1996b), Mollica (1998) e Silva (2004) em relação à escolaridade. Nos resultados apresentados por Paiva (1996b), com relação às crianças, a supressão da semivogal /y/ independe do grau de escolarização. Assim, as crianças do primário (atualmente 1^a fase do fundamental - 1º ao 5º ano) e as do ginásio (2^a fase do fundamental – 6º ao 9º ano) apresentaram probabilidades semelhantes na aplicação da regra de supressão: .51 e .50, respectivamente. Os resultados obtidos por Mollica (1998) levam-na a concluir que, com a evolução do ano de escolaridade, o apagamento da semivogal tende a reduzir-se mais rapidamente quando há orientação específica sobre a relação fala/escrita. Em sua análise, Silva (2004) observa que, com o ditongo /ow/, a escolarização não interfere no apagamento da semivogal; já com os ditongos /ey/ e /ay/, os falantes com mais anos de escolaridade são menos favoráveis ao apagamento da semivogal.

Nesta pesquisa com dados de leitura oral, os resultados evidenciam uma relação oposta entre ano de escolaridade e uso do ditongo, ou seja, na medida em que avança na escolaridade há diminuição do uso da forma com ditongo, conforme pode-se observar no Gráfico 1.

Gráfico 1 - Porcentagem de monotongação



Fonte: Elaboração própria.

Como pode ser observado no Gráfico 1, a manutenção do ditongo é inversamente proporcional ao avanço da escolaridade, enquanto os alunos dos 3º e 4º anos apagam em (21,08%) e (31,54%) respectivamente, o 5º ano aplicou a monotongação em 68,46% das ocorrências, o que mostra uma diferença acentuada entre o 3º e o 5º ano. Para melhor compreensão dos resultados obtidos na análise quantitativa, procedeu-se a uma análise qualitativa da leitura dos alunos, o que revelou aspectos importantes sobre o uso da monotongação. Depreende-se que a elevada porcentagem da forma monotongada para os alunos do 5º ano pode estar relacionada à aplicação, quase que categórica, da regra de apagamento da semivogal do ditongo /ow/, o que pode ser justificado pelo fato de, na linguagem falada, a forma monotongada desse ditongo se encontrar em estado de mudança praticamente implementada no sistema (PAIVA, 1996a; SILVA, 2004). Embora o contexto de uso exigisse uma monitoração maior da linguagem, os alunos do 5º ano, por terem maior proficiência na leitura, leram de forma mais espontânea que os do 3º e 4º anos, o que contribuiu para o uso da linguagem mais próxima da fala. A justificativa para a menor ocorrência da forma monotongada no 3º ano pode ser a mesma aplicada para o apagamento do rótico, isto é, o uso intenso da mediação fonológica no processo de leitura, favorecendo a pronúncia de todos os fonemas da palavra.

O que também se observa é que enquanto o apagamento da semivogal do ditongo /ow/ é muito elevado, a manutenção dos ditongos /ey/ de **jeito** e /yo/ de **vários** é quase categórica. Com dados de língua falada, Silva (2004) constatou que o ditongo /ey/ seguido da oclusiva /t/ conserva a semivogal quase que de forma categórica. Desse modo, pode-se concluir que, com os alunos pesquisados, principalmente os do 5º ano, há tendência à aplicação, na leitura, das mesmas regras de monotongação aplicadas à fala espontânea.

Com base na análise das ocorrências dos dois tipos de ditongo na leitura dos alunos, depreende-se que parece haver diferença no grau de estigmatização dos ditongos crescentes e decrescentes, em que os primeiros tendem a ser mais sensíveis à estigmatização, o que os caracteriza como pertencentes a uma fala menos cuidada, própria de uma classe sociocultural menos prestigiada; já os ditongos decrescentes, em particular o ditongo /ow/, não sofrem estigmatização sendo, portanto, característicos de qualquer classe social. Desta forma, a monotongação dos ditongos crescentes é mais passível de correção pelo professor, que a dos ditongos decrescentes, fato que pode ter colaborado, na leitura dos alunos, para a restrição de uso da monotongação para os primeiros e sua ampliação para os últimos.

Ditongação

A ditongação é um processo que ocorre em direção contrária à monotongação. Enquanto nesta existe a supressão da semivogal, na ditongação há a inserção. Assim, a ditongação pode ser entendida como um fenômeno essencialmente fonético, cuja realização acontece na fala; estando, portanto, sujeita a variações sociolinguísticas (linguísticas e extralinguísticas), a exemplo do tamanho do vocábulo, do contexto fonético, da escolarização, do nível de registro de fala etc.

Coutinho (1978, p.110) aponta, na história do português, algumas causas que contribuíram para a manutenção e aparição de ditongos: a síncope ou queda de fonema medial: *malu* > *mau*, *lege* > *lei*, *palu* > *pau*; a vocalização ou transformação de consoantes em vogal, em certos grupos consonantais: *alt(e)ru* > **autro* > *outro*; *factu* > **faito* > *feito*; *regnu* > *reino*; a metátese ou transposição de fonemas: *primariu* > **primaíro* > *primeiro*, *librariu* > **livraíro* > *livreiro*; a epêntese de vogal para desfazer hiatos: *creo* (<*credo*) > *creio*, *tea* (<*tela*) > *teia*. No português do Brasil, o autor chama a atenção para a ditongação na fala, com vogal em posição tônica final de palavras seguida de /s/ ou /z/: *gás* ~ *gais*, *rapaz* ~ *rapais*.

Câmara Junior (1986, p.100) define a ditongação como “[...] mudança fonética que consiste na formação de um ditongo sistemático a partir de uma vogal simples.” Ao tratar do fenômeno na língua portuguesa moderna, o autor afirma que a ditongação se deu nos seguintes casos:

- 1) Vogal tônica em hiato, quando a) média anterior, com o desenvolvimento de um ditongo /éy/ ou /éy/, indicado na grafia moderna - *ideia, veia*; b) média posterior fechada com o desenvolvimento de um ditongo /ôw/, não indicado na grafia e inexistente nas zonas dialetais em que houve a monotongação do ditongo /ôw/ - *boa ~ bôwa*.
- 2) Dialetalmente, para vogal tônica final travada por /s/ pós-vocálico, com o desenvolvimento dos ditongos de pospositiva /y/ - *pás, és, fez, sós, flux*, pronunciadas *pays, feys, sóys, fluys*. Dá-se, então, a neutralização da oposição entre ditongo e vogal simples, desaparecendo a distinção em casos como: *pás* e *pais* (CÂMARA JUNIOR, 1986, p.101).

Em estudos mais recentes sobre o português do Brasil, Bisol (1989, 1994), em uma análise fonológica do ditongo, afirma que em muitas variedades do português há tendência para o surgimento de um *glide* em palavras como *três, mas, rapaz*, e que isto acontece somente em sílaba acentuada e em posição final de palavra. A autora classifica os ditongos em verdadeiros e falsos; os primeiros estão representados por duas vogais no nível subjacente, os últimos possuem um *glide* epentético, sem representação na estrutura subjacente, formado na estrutura de superfície por expansão de um traço vocálico da consoante vizinha. Seguindo essa proposta, a ditongação presente na fala, em contextos que na escrita se realiza com vogal simples – *faxina ~ fayxina, faz ~ fays*; se insere na classificação dos falsos ditongos.

Como pode ser observado nos estudos referidos, a ditongação é um fenômeno antigo na história do português e ainda recorrente, conforme atestam, também, pesquisas com dados de língua falada no português brasileiro (LEIRIA, 1995; AQUINO, 1998, 2004; LEITE; CALLOU; MORAES, 2003), inclusive com a inserção do *glide* em outros contextos - *nayscimento ~ nascimento* – além da posição final da palavra.

O uso variável da ditongação em diversas comunidades brasileiras é atestado em estudos de base variacionista com dados de várias regiões do País. Leiria (1995), analisando dados de fala da região sul (VARSUL) e considerando apenas sílabas tónicas finais, destacou alguns fatores linguísticos relevantes para a formação do ditongo, como sândi externo, a qualidade da vogal e o ponto de articulação da consoante. Aquino (1998, 2004), investigando dados do Nordeste (VALPB), constatou que os fatores linguísticos exercem maior influência no processo de ditongação que os fatores sociais; entre os fatores linguísticos favorecedores, destacam-se: a tonicidade da sílaba, a qualidade da vogal e o status morfêmico da sibilante; em relação aos fatores sociais, os dados revelaram que os falantes mais escolarizados inibem a aplicação da forma ditongada.

Leite, Callou e Moraes (2003), analisando dados do Rio de Janeiro (NURC), observaram que a inserção do *glide* pode dar-se com qualquer uma das realizações

do arquifonema /S/, com preferência para a palatal e para vocábulos monossilábicos – *paz* ~*pays*, *faz* ~*fays*. Os autores constataram, também, a ocorrência do fenômeno na posição medial de palavra, embora restrita a poucos vocábulos.

Dentre os estudos referidos, interessa, de forma particular, o trabalho de Aquino (1998, 2004), uma vez que a comunidade analisada pela autora se situa no mesmo Estado do presente estudo, o que favorece, portanto, a comparação entre os resultados obtidos nas duas análises.

Comparando os dados de leitura com os de fala na análise de Aquino (1998, 2004), constata-se que a ditongação apresenta comportamento divergente nas duas pesquisas em relação à influência da escolaridade. Os resultados obtidos na análise de fala espontânea mostram que os falantes mais escolarizados desfavorecem a aplicação da regra de ditongação; já nos resultados com os dados de leitura, são os alunos mais escolarizados aqueles que mais usam a forma ditongada, conforme o Gráfico 2.

Gráfico 2 - Porcentagem de ditongação



Fonte: Elaboração própria.

Os valores obtidos, conforme o Gráfico 2, mostram que o processo de ditongação se comporta de forma inversamente proporcional ao avanço da escolaridade, embora não haja uma diferença tão acentuada entre os três anos: 3º ano (28,89%), 4º ano (32,63%) e 5º ano (34,21%). No entanto, considerando-se resultados obtidos por Aquino (1998, 2004) e a tendência nos estudos de língua falada em associar a predominância de formas padrão a falantes mais escolarizados, era de se esperar que os alunos do 5º ano usassem menos ditongação que os do 3º e 4º anos.

De posse dos valores estatísticos por cada ano de escolaridade, é necessário voltar aos dados com o objetivo de observar quais aspectos linguísticos das palavras analisadas favorecem o uso da ditongação. Constata-se, então, que as palavras monossilábicas (vez, fez, voz) foram lidas com ditongo (/veys/, /feys/, /voys/) por todos os alunos, independentemente do ano de escolaridade.

Possivelmente isto tenha contribuído para a pequena diferença percentual entre os anos. Essa constatação do amplo uso do ditongo nas palavras monossilábicas corrobora resultados obtidos em pesquisas sobre língua falada. Aquino (1998, 2004) constata que, com os monossílabos de uso tônico, a aplicação da ditongação é de (97%). Leite, Callou e Moraes (2003) encontram resultados que vão nesta mesma direção, uma vez que os resultados apontam a preferência da inserção do *glide* em vocábulos monossilábicos.

De certo modo, alguns resultados dos dados de fala foram confirmados na análise dos dados de leitura, conforme pode ser visualizado na Tabela 2, em que se apresenta a porcentagem de ocorrência da ditongação com dados de língua falada e a ocorrência na leitura.

Tabela 2 - Comparaçao entre dados de fala e dados de leitura

Fatores	% de ocorrência na fala (AQUINO, 1998)	Ocorrência na leitura
Monossílabo de uso tônico (três)	97 %	Uso ditongado em todas as ocorrências (100 %)
Final postônica de palavra (aqueelas)	3 %	As duas palavras que apresentaram esse contexto foram lidas com a vogal simples (0 %)
Inicial pretônica de palavra (nascer)	6 %	Houve 16 usos de ditongação com a palavra nascido
Inicial tônica de palavra (festa)	0 %	Houve 4 usos da forma ditongada com a palavra festa

Fonte: Elaboração própria.

Um fato que chamou a atenção foi o amplo uso do ditongo na palavra **nascido**. Dos 30 (trinta) alunos que leram o texto, 16 (dezesseis) usaram a forma ditongada. Pode-se ver nesse fato a presença da interferência da norma da comunidade na fala dos alunos e sua transposição para a leitura ou, mais especificamente, a rede social do indivíduo (pessoas com quem ele interage nos diversos domínios sociais, determinando as características de seu repertório linguístico (BORTONI-RICARDO, 2004). Tal fato leva a inferir, portanto, que, na comunidade de fala desses alunos, o uso vigente da palavra **nascido** é a forma ditongada *nayscido*.

Com base na análise quantitativa e qualitativa e, também, nos resultados encontrados nas referidas pesquisas com língua falada, pode-se afirmar que a ditongação diante de /S/ em monossílabos tônicos e na posição tônica final já tem seu uso bem consolidado no português brasileiro. Constatou-se também que, por se tratar de uma forma que não sofre estigma social, passa despercebida pelo crivo corretivo da escola, o que a faz ser usada tanto em situações informais como em contextos sociais em que há maior monitoramento da linguagem, como no contexto de leitura.

Apagamento da oclusiva dental 'd' no grupo –ndo

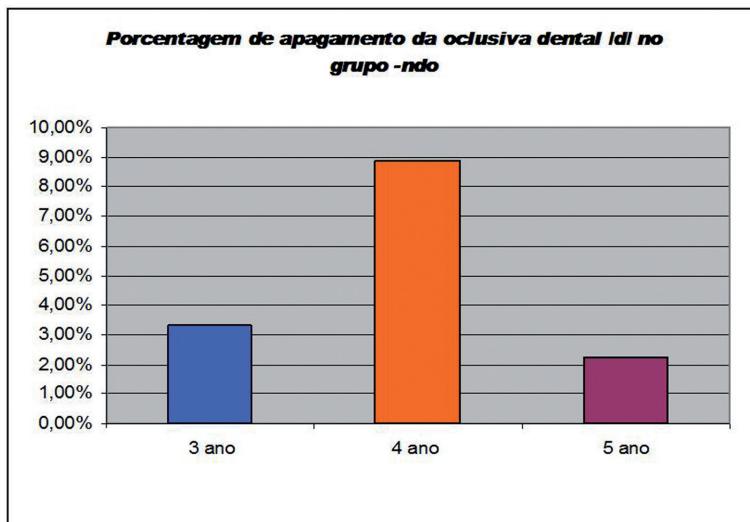
O fenômeno em análise consiste no resultado da assimilação do fonema /d/ pelo fonema /n/ em contextos como: imaginando ~ imaginano, falando ~ falano, quando ~ quano, ou seja, há uma assimilação do /d/ pelo /n/ para, em seguida, haver o apagamento do fonema (-nd- > -nn- > -n-). Este processo de apagamento da oclusiva /d/ não é exclusividade do português brasileiro; ele está presente também em outras línguas (MARTINS, 2004).

Na literatura pertinente, pode-se constatar a produtividade do fenômeno na língua falada em diversas regiões do País. No Atlas Linguístico da Paraíba (ARAGÃO, 1984), há registros de apagamento do 'd' no grupo –ndo em várias localidades como Mamanguape, Campina Grande, Guarabira. Martins (2004), utilizando dados do VALPB, constatou que existe um maior apagamento da oclusiva dental 'd' (58%) do que de sua preservação (42%) na fala dos pessoenses.

Com o objetivo de observar a influência da fala na escrita, Mollica (1998) analisou a assimilação da dental em **-ndo** na escrita de alunos da 5^a à 8^a série. Os resultados mostraram que fatores linguísticos e sociais, como a extensão do item lexical, escolaridade, sexo relevantes para a assimilação do fenômeno na fala foram, também, relevantes na escrita. A autora destaca que os ambientes favorecedores da assimilação da dental são os de maior dificuldade para a aprendizagem da escrita.

Neste estudo, o apagamento do 'd' no grupo –ndo foi o que apresentou a menor frequência (4,81%). Analisando a frequência do fenômeno por ano de escolaridade, verifica-se o menor índice de ocorrência no 5º ano, conforme pode ser visualizado no Gráfico 3.

Gráfico 3 - Porcentagem de apagamento da oclusiva dental /d/ no grupo -ndo



Fonte: Elaboração própria.

Os resultados mostram porcentagem maior de apagamento da oclusiva no 4º ano, o que foi surpreendente, pois se esperava um decréscimo de apagamento com o avanço da escolaridade, fato já constatado em pesquisa com língua falada (MARTINS, 2004).

Voltando aos dados, verifica-se que todas as ocorrências de apagamento do 'd', em todos os anos, aconteceram com a forma verbal no gerúndio, cujo resultado corrobora os achados de Martins (2004), com dados de língua falada, em que o índice mais alto de ocorrência se deu com o gerúndio (68%), e também Mollica (1998), com dados de escrita, em que o maior índice de ocorrência se deu com o gerúndio, com destaque para o gerúndio fático. Embora não tenha sido objetivo deste trabalho controlar as variáveis condicionadoras do fenômeno em análise, os resultados dão indicativos de que o gerúndio é o contexto favorecedor da ocorrência do apagamento da dental em **-ndo**, também na leitura.

Com base nos resultados, percebe-se que o apagamento da oclusiva dental como fenômeno variável não se apresentou de forma produtiva na leitura dos alunos pesquisados, o que leva a concluir que este fenômeno não pode ser concebido como um problema para o desenvolvimento da leitura oral na primeira fase do ensino fundamental (3º ao 5º ano).

Considerações finais

Os resultados apontaram o caráter variável, na leitura, de alguns fenômenos linguísticos característicos da oralidade. Comparando os dados de leitura com os de fala espontânea, constata-se que alguns resultados corroboram os encontrados em estudos sobre a língua falada na comunidade pessoense (SILVA, 1997, 2004; AQUINO, 1998, 2004; MARTINS, 2004).

Nos fenômenos variáveis analisados, os que apresentaram maior índice de ocorrência são, em geral, formas não estigmatizadas, já bastante consolidadas na fala espontânea. No geral, a realização dessas formas variáveis na leitura dos alunos não traz influência negativa no desempenho da leitura, pois não comprometem sua fluência nem a compreensão do que é lido. As formas estigmatizadas, como são mais propensas à correção pelo professor, ocorrem com menos frequência e tendem a declinar de uso com o avanço da escolaridade.

Os resultados deste estudo mostram uma realidade de sala de aula ainda pouco discutida no cenário acadêmico e educacional. O conhecimento, por parte do professor, das formas variantes na leitura dos alunos e sua interferência no processo de ensino/aprendizagem são muito importantes para o desenvolvimento da competência comunicativa do aluno no ambiente escolar e fora dele. Nesse sentido, os estudos sociolinguísticos de base variacionista têm contribuído de forma significativa ao trazer à tona variedades linguísticas presentes em qualquer ambiente social, dentre eles, o escolar. O trabalho de sala de aula que leva em consideração esses aspectos possibilita, ao aluno, a compreensão das variadas formas de uso da linguagem apropriadas aos diversos contextos sociais.

HORA, D.; AQUINO, M. F. From speech to reading: variation analysis. *Alfa*, São Paulo, v.56, n.3, p.1089-1105, 2012.

- *ABSTRACT: Variation studies in Brazil have focused mostly on the phonological and syntactic linguistic levels. The various projects which are based on data collected in different communities (VARSUL, PEUL, IBORUNA, VALPB etc.) seek to address their findings by using all the theoretical apparatus that Variation Sociolinguistics has made available. Based on the results of the implementation of Projeto Variação Linguística no Estado da Paraíba (Linguistic Variation in the State of Paraíba Project) - VALPB (HORA, 1993), which corpus is stratified by gender, age and education level, and attempting to cast an eye on a new corpus, this paper considers the linguistic variation in oral reading done by students of the first phase of elementary school. The analyzed data indicate a strong influence of the dialectal variant of the student in oral reading. Of the various processes involved in the original project, three were selected, namely: monophthongization, diphthongization and the deletion of 'd' in the group -ndo. This relation between what the student speaks and what he/she reads motivates the search for analysis proposals that go beyond what the school recommends.*

- *KEYWORDS: Sociolinguistics. Variation. Reading and teaching.*

REFERÊNCIAS

- AQUINO, M. F. S. Uso variável do ditongo em contexto de sibilante. In: HORA, D. (Org.). *Estudos sociolinguísticos: perfil de uma comunidade*. Santa Maria: Pallotti, 2004. p.45-54.
- _____. *A ditongação na comunidade de João Pessoa: uma análise variacionista*. 1988. 84f. Dissertação (Mestrado em Letras)-Universidade Federal da Paraíba, Centro de Ciências Humanas, Letras e Artes, João Pessoa, 1998.
- ARAGÃO, M. S. S. *Atlas linguístico da Paraíba: análise das formas e estruturas linguísticas encontradas*. Brasília: Ed. da UFPB, 1984.
- BISOL, L. Ditongos derivados. *DELTA*, São Paulo, v.10, n.esp., p.123-140, 1994.
- _____. O ditongo na perspectiva da fonologia atual. *DELTA*, São Paulo, v.5, n.2, p.185-224, 1989.
- BORTONI-RICARDO, S. M. *Nós chegoumu na escola, e agora? sociolinguística e educação*. São Paulo: Parábola, 2005.
- _____. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2004.
- CÂMARA JUNIOR, J. M. *Dicionário de linguística e gramática: referente à língua portuguesa*. 13.ed. Petrópolis: Vozes, 1986.
- _____. *História e estrutura da língua portuguesa*. Rio de Janeiro: Padrão, 1976.
- COUTINHO, I. L. *Pontos de gramática histórica*. 7.ed. Rio de Janeiro: Ao Livro Técnico, 1978.
- CUNHA, C.; CINTRA, L. F. L. *Nova gramática do português contemporâneo*. 2.ed. Rio de Janeiro: Nova Fronteira, 1985.
- HORA, D. *Projeto variação linguística no Estado da Paraíba*. 1993.
- LABOV,W.; BAKER, B. *What is a reading error?* University of Pennsylvania, 22 sep. 2003. Disponível em: <<http://www.ling.upenn.edu/~wlabov/Papers/WRE.html>>. Acesso em: 13 out. 2006.
- _____. *Testing the effectiveness of an individualized reading program for African-American, Euro-American and Latino Inner city children*. Pennsylvania: University of Pennsylvania, 2001. Disponível em: <<http://www.ling.upenn.edu/~wlabov/UMRP/UMRP.html>>. Acesso em: 13 out. 2006.
- LEIRIA, L. L. *A ditongação variável em sílabas tônicas finais travadas por /s/*. 1995. 74f. Dissertação (Mestrado em Letras) - Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1995.

LEITE, Y.; CALLOU, D.; MORAES, J. Processos em curso no português do Brasil: a ditongação. In: HORA, D.; COLLISCHONN, G. (Org.). *Teoria linguística: fonologia e outros temas*. João Pessoa: Universitária, 2003. p.232-250.

MARTINS, I. F. M. Apagamento da oclusiva dental /d/: perspectivas variacionistas e fonológicas. In: HORA, D. (Org.). *Estudos sociolinguísticos: perfil de uma comunidade*. Santa Maria: Pallotti, 2004. p.55-82.

MOLLICA, M. C. *Da linguagem coloquial à escrita padrão*. Rio de Janeiro: 7Letras, 2003.

_____. *Influência da fala na alfabetização*. Rio de Janeiro: Tempo Brasileiro, 1998.

PAIVA, M. C. A. Supressão das semivogais nos ditongos decrescentes. In: SILVA, G. M. O.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996a. p.217-236.

_____. Atuação das variáveis sociais na supressão da semivogal anterior nos ditongos decrescentes. In: SILVA, G. M. O.; SCHERRE, M. M. P. (Org.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996b. p.325-333.

R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Vienna, 2007. Disponível em: <<http://www.R-project.org>>. Acesso em: 24 abr. de 2010.

SILVA, F. S. O processo de monotongação em João Pessoa. In: HORA, D. (Org.). *Estudos sociolinguísticos: perfil de uma comunidade*. Santa Maria: Pallotti, 2004. p.29-43.

_____. *O processo de monotongação em João Pessoa*. 1997. 120f. Dissertação (Mestrado em Letras) - Universidade Federal da Paraíba, João Pessoa, 1997.

Recebido em abril de 2012

Aprovado em junho de 2012

PALATALIZAÇÃO DAS PLOSIVAS ALVEOLARES EM FLORES DA CUNHA (RS): VARIAÇÃO LINGUÍSTICA E PRÁTICAS SOCIAIS

Elisa BATTISTI*

Adalberto Ayjara DORNELLES FILHO**

- RESUMO: A palatalização variável das plosivas alveolares no português brasileiro falado em Flores da Cunha é moderada, o que contribui para caracterizar o falar local. A análise de regra variável (LABOV, 1972, 1994, 2001) de dados de 48 entrevistas sociolinguísticas de Flores da Cunha do BDSer revela uma proporção total de aplicação de 29%. Os condicionamentos são tanto linguísticos quanto sociais e a palatalização progride na comunidade: jovens, vogal alta fonológica, habitantes de zona urbana e consoante-alvo desvozeada favorecem a palatalização. Na análise da variação como prática social (ECKERT, 2000), o estudo da rede social dos informantes mostra que a alta densidade da rede, nucleada por informantes de grupos etários mais velhos, refreia a palatalização. O estudo etnográfico revela que os habitantes de Flores da Cunha, em especial os jovens, realizam práticas sociais inovadoras ao lado de práticas tradicionais, introduzindo a palatalização na comunidade. O emprego das formas palatalizadas é valorado como não local e é relativamente prestigiado.
- PALAVRAS-CHAVE: Variação linguística. Palatalização. Plosivas alveolares. Práticas sociais.

Introdução

O português falado no Brasil tem características fonético-fonológicas resultantes dos contatos linguísticos ocorridos ao longo da história do país, alguns ainda em andamento. Na antiga região colonial italiana do Rio Grande do Sul (RCI-RS), Estado situado no extremo Sul do Brasil, o português guarda, do contato com o italiano, alternâncias como o emprego de vibrante simples em lugar de múltipla em *onset silábico* (*churrasco~churasco*, *carro~caro*); de *-on* em lugar de *-ão* (*pão~pon*, *caminhão~caminhon*). Essas características constroem o sotaque local. Outros traços variáveis contribuem para essa construção, não porque se originem do contato italiano-português, mas porque, bastante frequentes em outras variedades de português brasileiro, têm baixa ocorrência no português

* UFRGS – Universidade Federal do Rio Grande do Sul. Instituto de Letras. Porto Alegre – Rio Grande do Sul – Brasil. 91540-000 - elisa.battisti@pq.cnpq.br

** UCS – Universidade de Caxias do Sul. Centro de Ciências Exatas e Tecnologia. Caxias do Sul – Rio Grande do Sul – Brasil. 95070-560 - aadomef@ucs.br

falado na região. É o que acontece com a palatalização das plosivas alveolares /t/ e /d/ (*tia-[tS]ia, dia-[dZ]ia, parte-par[tS]i, onde-on[dZ]i*).

No presente estudo, analisa-se a palatalização variável de /t/ e /d/ no português falado em Flores da Cunha, pequeno município de base étnica italiana localizado na RCI-RS. Além dos condicionamentos linguísticos e sociais do processo, objetiva-se verificar a proporção total de aplicação da regra, sua tendência a progredir, regredir ou estabilizar-se no sistema da comunidade e o valor social das formas palatalizadas.

A questão a que se procura responder relaciona-se ao padrão de palatalização de outros municípios da região, Antônio Prado e Caxias do Sul. A proporção de aplicação não ultrapassa os 30%. O processo tende a se estabilizar na primeira comunidade e a progredir na segunda. Já na capital do Estado, Porto Alegre, a palatalização supera os 90%. Qual é o padrão de Flores da Cunha? Que elementos da cultura local, da estrutura linguística e social sustentam esse padrão?

O estudo desenvolve-se na perspectiva da variação linguística como prática social (ECKERT, 2000), uma abordagem que aproxima sociolinguística de antropologia, articulando análise quantitativa e análise qualitativa. Para dar conta dessa articulação, são realizadas análise de regra variável (LABOV, 1972, 1994, 2001), análise de rede social (MILROY, 1980) e estudo etnográfico, que contempla observação participante e análise discursiva das entrevistas sociolinguísticas.

Mostraremos que a palatalização em Flores da Cunha é condicionada tanto por variáveis linguísticas quanto sociais. O processo correlaciona-se ao *status* da vogal-gatilho da regra, à qualidade da consoante-alvo, à idade dos informantes e a seu local de residência. A menor aplicação da regra na zona rural relaciona-se a diferenças na estrutura e no conteúdo da rede social. As formas palatalizadas são salientes para os florenses, que as percebem como não locais, num quadro de práticas sociais tradicionais mescladas a tendências sociais globais.

O artigo se estrutura em seis seções, cinco além desta. Na seção seguinte, Fundamentação Teórica, descreve-se a palatalização das plosivas alveolares do ponto de vista fonético e em suas repercussões sociais e abordam-se os fundamentos das análises quantitativa e qualitativa de processos variáveis. Na Revisão de Literatura, revisam-se alguns estudos sobre a palatalização das plosivas alveolares com dados do Sul do Brasil. Na seção Metodologia, detalham-se os procedimentos seguidos nas análises realizadas: de regra variável, de rede social, etnográfica. Os resultados são apresentados e discutidos na quinta seção do trabalho, este finalizado com a Conclusão.

A palatalização das plosivas alveolares

Palatalização é um processo assimilatório, envolve a adoção, por um segmento, de características de segmentos vizinhos. A característica em questão é o emprego da lâmina/corpo da língua na articulação do segmento. Na palatalização das plosivas alveolares, /t/ e /d/ assimilam tal característica articulatória da vogal alta seguinte - /i/, como em *tia*, *dia*, ou [i] derivado de /e/, como em *parte*, *onde*. Além da palatalização propriamente dita, as consoantes sofrem africatização, isto é, há um pequeno escape de ar ao final da articulação.

Em estudos sobre o português brasileiro em contato com o italiano (FROSI; MIORANZA, 1983), registra-se a baixa aplicação dessa regra variável, o que constitui um traço característico desse falar em oposição aos falares monolíngues-português, em que a palatalização é mais difundida. É o que se verifica em Flores da Cunha: a não palatalização das plosivas alveolares caracteriza o português local. Isso é percebido pelos próprios habitantes, o que contrasta com a pronúncia palatalizada dos indivíduos “de fora”, os monolíngues-português.

Pagotto (2001), em estudo sobre o português falado em Florianópolis, verificou a existência de questões de identidade envolvidas no emprego de plosivas palatalizadas e não palatalizadas. Naquela localidade, as questões resultam da tensão entre os habitantes da ilha e os “de fora”, veranistas e novos habitantes vindos de outros lugares do Brasil e exterior, que invadiram a ilha a partir do processo de reurbanização ocorrido nas últimas décadas. Empregar uma ou outra forma adquire valor social, reveste-se de significados que nascem do conjunto de práticas que o uso da língua integra.

É nesse sentido que se estuda a regra variável de palatalização, como processo variável dirigido por restrições linguísticas e sustentado por valores ou significados que se originam nas práticas sociais.

Variação linguística e práticas sociais

Os estudos de variação linguística conformes à Teoria da Variação (LABOV, 1972, 1994, 2001) têm como objeto as formas alternativas de dizer a mesma coisa, naquilo que informam sobre as regras que estão por trás da variação. Tais estudos buscam generalidades preditivas sobre a realização linguística em uma comunidade de fala, para o que uma considerável quantidade de dados variáveis e seu tratamento quantitativo são necessários. Seguem o pressuposto de que

a variação linguística, além da motivação interna (linguística), tem motivação externa (social): correlaciona-se à classe social, etnia, idade, sexo/gênero dos indivíduos, entre outras características.

Nessa perspectiva, a língua não é propriedade do indivíduo, mas da comunidade. O indivíduo herda da comunidade o sistema da língua (a variação, inclusive). Os informantes são representantes da comunidade e a comunidade é definida pelo analista através das categorias sociais que propõe e controla na análise. O que define a comunidade de fala são os padrões de uso da língua e não o indivíduo ou a fala individual. Por exemplo, a realização e os padrões de uso da palatalização das plosivas alveolares é uma das características que distinguem a comunidade de falantes de português brasileiro da comunidade de falantes de português europeu.

Além dessa análise, a investigação do valor social atribuído às variantes, na linha de Eckert (2000), requer considerar que os falantes constituem, e não apenas representam, categorias sociais. O significado social da variação não é apenas reflexo da pertença ou forma de reclamar pertença a categorias, mas está relacionado às práticas que dão origem e mantêm as categorias e que tornam a pertença a elas significativa: as formas alternantes são valoradas conforme o prestígio da comunidade, o indivíduo é valorado conforme o grau de adesão à comunidade. Esta não é mais vista apenas como comunidade de fala, mas como comunidade de prática. É na comunidade de prática, e no mercado linguístico (BOURDIEU, 2008) aí estabelecido, que os falantes são avaliados, que se dá significado para as suas diferenças.

Num município como Flores da Cunha, como em qualquer outro, muitas são as comunidades de prática, onde símbolos como a origem familiar e a linguagem, entre outros, convertem-se em patrimônio, em capital cultural que compete no mercado de bens simbólicos locais, regionais e nacionais e contribui para a distinção simbólica e para a visibilidade do grupo.

Acreditamos que, não só em Flores da Cunha, mas na região como um todo, a baixa palatalização seja recurso estilístico que permita aos descendentes de italianos (e de outras etnias que porventura residam hoje na comunidade) realizarem práticas sociais locais como forma de diferenciarem-se no cenário estadual e nacional e, assim, ganharem visibilidade, mesmo que elementos não locais sejam também experimentados na comunidade.

As relações dos falantes em redes sociais têm efeitos sobre seu comportamento linguístico, o que sustenta e explica a emergência das variantes. Milroy (1980) faz uma análise da modalidade não padrão do irlandês falado em três bairros de classe trabalhadora¹ de Belfast, no que se refere aos padrões variáveis de realização vocálica, predominantemente. A cuidadosa

¹ Do inglês *working-class neighborhoods*.

quantificação da variação, correlacionada à rede social dos informantes, revelou que a predominância de alternantes não padrão reflete os padrões de interação social das comunidades em redes. Não é a classe baixa o que determina o emprego de variantes não padrão, mas a identidade de grupo que, por sua vez, nasce de uma coesão originada nas próprias condições de sobrevivência do grupo, em suas práticas sociais, que se dão basicamente num único território. O sentimento de pertença ao bairro e o valor social (positivo) a ele atribuído Milroy denomina *localismo*.

Conforme Evans (2004), as redes apresentam diferenças em sua configuração em duas dimensões, a da estrutura (ou densidade) e a do conteúdo (*plexidade*)². A densidade refere-se à quantidade de contatos dos indivíduos: quanto maior o número de pessoas em rede que se conhecem umas às outras, maior sua densidade. Uma rede em que poucos indivíduos conhecem-se mutuamente é uma rede frouxa, de baixa densidade. À estrutura da rede importa o fato de os membros conectarem-se uns aos outros por mais de um tipo de relacionamento (colega de trabalho e, ao mesmo tempo, vizinho). Uma rede na qual os indivíduos se relacionam de múltiplas maneiras é uma rede *multiplexa*, uma rede na qual os indivíduos se relacionam de uma única maneira é *uniplexa*.

Revisão de literatura

Battisti e Guzzo (2010) fazem uma extensa revisão de estudos sobre a palatalização das plosivas alveolares com dados do Sul do Brasil realizados na perspectiva sociolinguística laboviana, em análises de regra variável. A maior parte deles diz respeito a falares do Rio Grande do Sul e a constatação das autoras é a de que, exceto Porto Alegre, em que a aplicação da palatalização tem frequência superior a 90%, as comunidades gaúchas apresentam taxa de palatalização de baixa à moderada. Isso poderia ser indício de a regra estar se expandindo da capital para o interior, não fosse o comportamento de algumas comunidades em que a palatalização é variação estável e aplica-se em níveis moderados.

É o caso de Antônio Prado, estudado por Battisti et al. (2007) com dados do BDSer³. Os autores verificaram que a frequência total de aplicação da palatalização é de 30% e o processo é favorecido por vogal alta fonológica /i/ (a redução de /e/ átono em Antônio Prado é baixa) como também por jovens habitantes da zona urbana do município. Embora haja um aumento na frequência de palatalização

² Em inglês, os termos correspondentes a densidade e plexidade são *density* e *plexity*, respectivamente.

³ Banco de Dados de Fala da Serra Gaúcha, da Universidade de Caxias do Sul. É um acervo de entrevistas sociolinguísticas iniciado, em agosto de 2000, por um grupo de pesquisadores da UCS integrado pelos autores deste artigo. O acervo contém 55 entrevistas de Caxias do Sul, 55 de São Marcos, 57 de Antônio Prado e 56 de Flores da Cunha.

com o declínio da idade do informante, as taxas se estabilizam nas faixas etárias mais jovens, o que leva a crer que a regra em Antônio Prado não seja variação num processo de mudança, mas alternância que tenda a estabilizar-se no sistema linguístico em índices modestos. A análise da rede social dos informantes revela que a relação em rede de informantes da zona rural de Antônio Prado sustenta o emprego de alternantes conservadoras, no caso, as variantes não palatalizadas. Os de zona urbana participam de redes densas, porém os laços inter-individuais são menos íntimos, assim menos reforçadores de um falar local, possibilitando o emprego de alternantes menos conservadoras, as variantes palatalizadas na análise.

Nas demais comunidades, a regra está em progressão, mas em ritmos diferentes. É o que constata Bisol (1986) em seu estudo pioneiro. A autora analisou a palatalização das plosivas /t/ e /d/ na fala de 60 informantes de quatro grupos geográficos do Rio Grande do Sul: 15 de Porto Alegre, monolíngues-português; 15 de Livramento, fronteira do Rio Grande do Sul com o Uruguai, onde se vivencia o contato português-espanhol diariamente; 15 de Taquara, bilíngues português-alemão; e 15 de Monte Bérico, Veranópolis, bilíngues português-italiano. A frequência total de aplicação foi de 65%, com a seguinte distribuição: 90% em Porto Alegre, 79% em Livramento, 60% em Taquara e 18% em Monte Bérico. A autora verificou que a palatalização é favorecida pelos metropolitanos e fronteiriços, desfavorecida por italianos e alemães, oposição que se mantém nos resultados da variável Sílaba: nos metropolitanos e fronteiriços, o processo é favorecido nas posições relativamente mais fortes da palavra; nos italianos e alemães, nas posições mais fracas. Jovens palatalizam mais em todos os grupos étnicos. A autora registra a interação entre a palatalização e a regra de elevação de /e/ átono, a última alimentadora da primeira.

A palatalização em Flores da Cunha foi investigada por Almeida (2000) em dados do VARSUL⁴. O autor levantou contextos de 24 entrevistas sociolinguísticas e, neles, a frequência total de aplicação da regra foi de 47%. O autor acredita que o processo esteja em aquisição na comunidade. Foram controladas as variáveis linguísticas Contexto Fonológico Precedente e Seguinte, Sonoridade, Tonicidade, Tipo de Vogal Alta; e as variáveis extralingüísticas Idade, Escolaridade e Sexo/Gênero. Favorecem a palatalização os fatores vogal, vibrante, fricativa alveolar precedente, lateral e labial seguinte, como também oclusiva surda /t/, sílaba postônica final e vogal alta derivada de /e/ átono. As três variáveis sociais mostraram-se relevantes para o processo: promovem a palatalização as mulheres, os indivíduos que completaram o ensino médio e aqueles com menos de 50 anos

⁴ O projeto VARSUL (Variação Linguística na Região Sul do Brasil) tem por objetivo geral a descrição do português falado e escrito de áreas socioculturalmente representativas do Sul do Brasil. Constituem seu acervo o Banco de Dados VARSUL, a Amostra Digital VARSUL e o Banco de Dados Diacrônico. Disponível em: <<http://www.varsul.org.br/>>. Acesso em: 8 fev. 2012.

de idade, o que permite ao autor afirmar a existência de mudança em progresso na comunidade, conduzida pelas mulheres ao adotarem a variante de maior prestígio.

Matté (2009) testou a hipótese de que Caxias do Sul (RS) difundiria a palatalização de /t d/ a municípios vizinhos. Fez análise de regra variável de dados levantados de 16 entrevistas sociolinguísticas do BDSer. Foram 9.006 os contextos, cuja análise revelou uma proporção de 35% de palatalização. Como essa taxa é próxima à de Antônio Prado para o processo, a hipótese inicial não se confirmou. A maior diferença entre as duas cidades está no fator 18-29 anos da variável Idade: 72% em Caxias do Sul, 42% em Antônio Prado. O autor discutiu a ideia de que os jovens de Caxias do Sul talvez usem a regra temporariamente e, mais tarde, voltem a falar como os mais velhos, com o que se mantém a palatalização moderada.

Metodologia

Flores da Cunha

A comunidade investigada no presente estudo é, como se vem afirmando desde a Introdução, Flores da Cunha, situada na antiga região colonial italiana do Rio Grande do Sul (RCI-RS)⁵. Conde D'Eu (Garibaldi), Dona Isabel (Bento Gonçalves) e Campo dos Bugres (Caxias do Sul) foram as três primeiras colônias italianas no Rio Grande do Sul, fundadas a partir de 1875. Inicialmente, Flores da Cunha se chamava São José e era um distrito de Caxias do Sul. Em 1924, por um decreto estadual, criou-se o município, que tinha por sede a antiga vila de Nova Trento. Em 1935, substituiu-se a denominação Nova Trento por Flores da Cunha, em homenagem ao ex-governador do Estado, José Antônio Flores da Cunha.

Hoje, o município conta com uma população pequena (em torno de 27.000 habitantes⁶), mas não em redução. É vizinho a Caxias do Sul, maior cidade do Rio Grande do Sul após a capital, Porto Alegre. Na área de 272,66 km², a zona rural é relativamente grande e o pequeno centro urbano dá sinais de crescimento. Sua economia é diversificada, inclui indústria, comércio, serviços, agricultura. O

⁵ Sabbatini e Franzina (1977), que empregam termo correspondente (Região de Colonização Italiana), explicam que a RCI-RS corresponde especificamente às áreas das ex-colônias de natureza pública, fundadas entre 1875 e 1892 na Encosta Superior do Nordeste do Rio Grande do Sul. Atualmente, 55 municípios localizam-se nesse território, entre eles Bento Gonçalves, Caxias do Sul, Farroupilha, Flores da Cunha, Garibaldi, São Marcos, Veranópolis, Antônio Prado. De acordo com Frosi e Mioranza (1983) e Frosi (1989), o grupo de imigrantes era misto relativamente à província italiana de origem, sendo quatro as regiões da Itália setentrional de que veio a maioria deles: Lombardia, Vêneto, Friuli Venezia-Giulia e Trentino-Alto Ádige.

⁶ Confira os dados do IBGE, censo 2010. Disponíveis em: <www.ibge.gov.br>. Acesso em: 13 mar. 2011.

município é um dos maiores produtores de vinho do Brasil e sua indústria moveleira exporta para a América Latina, Estados Unidos e Europa, principalmente.

Como outros municípios da RCI-RS, Flores da Cunha celebra suas origens em festivais e movimentos culturais, o que reifica a identidade étnica italiana da comunidade.

Figura 1 - Flores da Cunha no Rio Grande do Sul e no Brasil.



Fonte: Elaboração própria.

A análise de regra variável

Para a análise de regra variável (LABOV, 1972, 1994, 2001), utilizaram-se 48 entrevistas sociolinguísticas do BDSer com informantes de Flores da Cunha. Essas entrevistas, de que se levantaram os contextos de palatalização, foram realizadas entre 2008 e 2009.

A variável dependente é a palatalização das plosivas alveolares. As variáveis independentes são de dois tipos, linguísticas e extralingüísticas.

As variáveis linguísticas controladas são Contexto Fonológico Precedente, Contexto Fonológico Seguinte, *Status da Vogal Alta*⁷, Posição da Sílaba na Palavra, Tonicidade da Sílaba, Qualidade da Consoante-Alvo. Considerando o que se constatou nos estudos revisados (seção acima), nossas hipóteses foram as de que vogal alta fonológica /i/, sílabas postônicas e átonas e consoante-alvo /t/ condicionariam a palatalização. O papel dos contextos fonológicos precedente e seguinte estaria por ser constatado.

⁷ O controle dessa variável tem consequência sobre a ortogonalidade entre grupos de fatores, principalmente os que se referem à tonicidade da sílaba e sua posição na palavra: uma vogal alta fonética é átona, exclusivamente; uma vogal alta fonológica átona praticamente não se verifica em sílaba final de vocábulo. Assim, para viabilizar a operação adequada dos programas do pacote VARBRUL, versão Goldvarb-X, a variável *Status da Vogal Alta* não foi controlada em rodadas juntamente com as variáveis Posição da Sílaba na Palavra e Tonicidade da Sílaba e vice-versa.

Os informantes são dos sexos/gêneros masculino e feminino, residentes na zona urbana e rural e estratificados em quatro grupos etários: 18 a 30 anos de idade; 31 a 50 anos; 51 a 70 anos; 71 ou mais anos. Essas características configuram as três variáveis extralingüísticas controladas na análise: Sexo/Gênero, Local de Residência e Faixa Etária. As hipóteses foram as de que falantes do sexo/gênero feminino, idosos e de zona urbana seriam os condicionadores da aplicação da regra.

O Quadro 1 traz o conjunto de variáveis independentes controladas na análise, com exemplos de contextos de palatalização.

Quadro 1 - Variáveis independentes controladas na análise de regra variável.

Linguísticas	Extralingüísticas
Contexto fonológico precedente Vogal oral: metida, fedido Ditongo: oitiva , Neide Consoante nasal: mentira , bandido Consoante lateral: malte, balde Consoante sibilante: poste, desde Vibrante ou tepe: arte, morde Zero: tipo, dia	Sexo/Gênero Feminino Masculino
Contexto fonológico seguinte Vogal: teatro, diabo Consoante sibilante e fricativa anterior: disse, disco, dúvida, atiçar, artista, tivesse Consoante sibilante posterior: tijolo, digital Consoante oclusiva: tipo, dica Consoante lateral anterior: dilema, mutilado Consoante lateral posterior: gatilho, armadilha Consoante nasal anterior: time, rudimentar, rotina, ladino Consoante nasal posterior: dinheiro, tinhoso Vibrante, tepe: tira, direção, Valdir, repartir Consoante africada: te [dʒ]isse, se [tʃ]inha Zero: gente, onde	Faixa etária 18-30 anos 31-50 anos 51-70 anos 71 ou mais anos
Status da vogal alta Alta fonológica: mentira, dica Alta fonética: gente, bonde	Local de residência Zona urbana Zona rural
Posição da sílaba na palavra Inicial: teatro, diário Medial: antigo, ardido Final: pente, acorde Monossílabo: diz , TIM Inicial de locução (clítico): de noite, te queria Inicial em locução (hospedeiro): pra tirar, se dizer Monossílabo em locução: se diz , pra ti	
Tonicidade da sílaba ⁸ Átona não final: atitude, médico Átona final: vinte, bonde Tônica: dia, tia Átona (clítico): de , te	
Qualidade da consoante alvo Vozeada: dia Desvozeada: tia	

Fonte: Elaboração própria.⁸

⁸ O fator Átona não final reúne pretônicas e postônicas não finais em razão da raridade das últimas na amostra.

Os contextos levantados das entrevistas foram codificados e posteriormente submetidos aos programas computacionais do pacote VARBRUL, versão Goldvarb X⁹ para ambiente Windows, que realizam análise estatística de regressão logística.

Os valores obtidos na análise quantitativa são apresentados em percentuais e em pesos relativos. Os percentuais expressam a distribuição dos dados e a aplicação da regra variável por fator dos diferentes grupos (variáveis) controlados. Os pesos relativos expressam a tendência de o processo estudado ocorrer, como efeito dos diferentes fatores considerados na análise. Os pesos relativos estão compreendidos no intervalo de 0 a 1: valores em torno de 0,5 indicam a neutralidade do fator em relação ao processo estudado, valores abaixo de 0,5 indicam que o fator não condiciona (desfavorece) o processo, valores acima de 0,5 indicam que o fator condiciona (favorece) o processo.

A análise da variação linguística como prática social

Para a análise da palatalização variável como prática social (ECKERT, 2000), realizaram-se análise de rede social e estudo etnográfico.

A análise de rede social considerou os 48 informantes de Flores da Cunha de cujas entrevistas sociolinguísticas levantaram-se os contextos de palatalização considerados na análise de regra variável. A rede foi analisada nas dimensões da estrutura e do conteúdo. Essa última dimensão foi contemplada, como fizeram Blake e Josey (2003), pelo controle dos graus de relacionamento interindividual. As informações consideradas, de natureza cultural e voltadas às práticas, foram obtidas junto aos informantes e complementadas com o conteúdo das entrevistas sociolinguísticas e das fichas sociais dos informantes. Relacionamentos familiares e entre colegas de trabalho são os mais relevantes na comunidade florense, variando neles o grau de intimidade dada a natureza/frequência da interação. O Quadro 2 traz os graus de relacionamento utilizados na análise: 1, 2 e 3, do mais íntimo ao menos íntimo, em conformidade com os padrões locais.

⁹ Disponível em: <<http://individual.utoronto.ca/tagliamonte/goldvarb.htm>>. Acesso em: 21 dez. 2010.

Quadro 2 - Graus de relacionamento em rede em Flores da Cunha.

1. Primeiro grau
1A – Marido/mulher
1B – Pais/filhos
1C – Colegas de trabalho com interação
2. Segundo grau
2A – Tios/sobrinhos/primos/cunhados
2B – Amigos íntimos
2C – Vizinhos íntimos
2D – Colegas de associação com interação
3. Terceiro grau
3A – Amigos não íntimos
3B – Vizinhos não íntimos
3C – Colegas de trabalho sem interação
3D – Colegas de associação sem interação
3E – Tios/sobrinhos/primos/cunhados

Fonte: Elaboração própria.

A rede social dos informantes foi formada a partir da realização das próprias entrevistas sociolinguísticas, quando um informante indicava outro com as características sociais de interesse¹⁰. Quando não foi possível obter essa indicação, os pesquisadores, através de seus contatos na comunidade, entrevistaram florenses com o perfil exigido, a eles perguntando, subsequentemente, se conheciam os demais informantes já entrevistados e que espécie de relacionamento mantinham com cada um.

A rede social foi modelada por um grafo (BOAVENTURA NETTO, 1996; FRUCHTERMAN; REINGOLD, 1991; MATHEWS, 1992; GERHARDT; CORSO; LEMKE, 2006). Cada informante é um vértice e cada relação de interação é uma aresta. Efetuou-se um estudo do problema de posicionamento dos vértices de modo que o desenho resultante tivesse bons atributos estéticos e de visibilidade, computacionalmente elaborados, usando-se uma adaptação do algoritmo de Fruchterman e Reingold (1991) e o método de minimização do gradiente. Cruzaram-se, com análise de correlação, a frequência de aplicação individual das realizações variáveis com a aplicação média dos contatos do informante na rede e com características sociais que se mostraram relevantes na análise de regra variável.

No estudo etnográfico, utilizaram-se dois procedimentos: análise crítica de ideologia, de natureza discursiva, e observação participante. Tanto na execução de um quanto na do outro procedimento, a técnica básica é a da inferenciação, como forma de verificar e compreender mudanças sociais ligadas a mudanças linguísticas.

¹⁰ O BDSer selecionou os informantes em cada município conforme os critérios Sexo/Gênero (Masculino, Feminino), Faixa Etária (18 a 30 anos; 31 a 50 anos; 51 a 70 anos; 71 ou mais anos), Escolaridade (0 a 4 anos; 5 a 8 anos; 9 a 11 anos; 12 ou mais anos) e Local de Residência (Zona urbana e Zona rural).

A inferenciação, conforme Spradley (1979), é um processo de aprendizagem realizada pelo pesquisador no convívio com as pessoas, com os grupos pesquisados. São três as grandes fontes de inferências: o que as pessoas dizem, o que elas fazem (suas práticas sociais) e os artefatos que elas usam.

Os resultados referentes à análise crítica de ideologia foram gerados das 48 entrevistas sociolinguísticas de que se levantaram os contextos de palatalização também considerados na análise de regra variável. Segundo Cerqueira Filho (1982), voltam-se ao que as pessoas dizem, aos assuntos de que falam, reveladores não só das razões tácitas para inovações, inclusive linguísticas, como também de visões de mundo motivadoras de mudanças ou de resistência a elas. Levantamos das entrevistas afirmações das pessoas sobre o que elas fazem diariamente, afirmações ligadas às dimensões do trabalho, da família, do lazer, dos estudos, da religião, da política e da própria comunidade.

Do ponto de vista ideológico e com foco na tensão entre local e global, a hipótese seguida foi a de que, apesar de as elites¹¹ regionais tenderem a manter padrões de dominação que não sirvam diretamente à globalização, os padrões locais são cooptados pela globalização, isto é, há dominação sem destruição. Isso corresponderia à tendência à produção da variante não palatalizada (forma tradicional), a despeito das frequências maiores de palatalização em outros falares de português brasileiro.

Outros são dados gerados de momentos de observação participante em festas de capela com almoços ou jantares, feiras, celebrações religiosas como a de *Corpus Christi*, em que se procurou fazer o que os outros fazem. Assim, observaram-se as ações das pessoas, o protocolo seguido na realização dessas ações, vestimentas, artefatos utilizados.

No que se refere às práticas sociais, nossa hipótese é a de que os florenses em geral, assim como os habitantes de outras comunidades ítalo-brasileiras da região, realizam práticas sociais inovadoras, como a utilização de computadores pessoais e da Internet para comunicação pessoal e para a realização de ações do cotidiano como pagamentos, cadastramentos, matrículas. No entanto não abrem mão de práticas sociais tradicionais: seguem fazendo o mesmo que seus antepassados, por exemplo, ao optar por comer galete com polenta aos domingos, frequentar missas e outras manifestações religiosas católicas nos finais de semana e em dias santos, visitar parentes nos momentos de lazer e, se habitantes da zona rural, dedicar-se a atividades agrícolas nas pequenas comunidades familiares, atividades essas ligadas ao setor vitivinícola.

¹¹ Sobre a teoria das elites, ver Bottomore (1974) e Mills (1962).

Resultados

Análise de regra variável

Foram 23.163 os contextos de palatalização levantados das entrevistas sociolinguísticas e considerados na análise de regra variável. A proporção total de aplicação da regra foi de 29%, índice moderado e próximo ao de outros municípios da região, como Antônio Prado (BATTISTI et al. 2007), com 30%, e Caxias do Sul (MATTÉ, 2009), com 35%.

Todas as variáveis foram selecionadas após refinamento da análise estatística, com amalgamação de fatores em alguns grupos. Os resultados das variáveis serão apresentados e discutidos pela ordem de relevância estatística dos valores obtidos, da maior para a menor. Destaca-se que, dentre as quatro primeiras variáveis, estão duas linguísticas e duas extralingüísticas, o que revela o caráter tanto estrutural quanto social da palatalização.

O primeiro grupo de fatores selecionado pelo programa na análise dos dados de Flores da Cunha é social e indica variação com mudança em progresso. Trata-se de Faixa Etária, controlado segundo a hipótese de que a palatalização, regra nova na comunidade, estaria em progresso, sendo introduzida pelos mais jovens e por esses difundida aos demais grupos etários. Essa hipótese se confirmou.

Tabela 1 - Aplicação da regra de palatalização e faixa etária.

Fatores	Aplic./Total	%	Peso relativo
18 a 30 anos	3.114/6.146	51	0,86
31 a 50 anos	1.599/5.766	28	0,50
51 a 70 anos	1.530/6.080	25	0,42
71 ou mais anos	515/5.171	10	0,14
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

O grupo etário mais jovem apresenta a maior proporção de aplicação da regra. Seu peso relativo (0,86) indica não só que os falantes mais jovens condicionam a palatalização mas também que há a tendência de o processo se aplicar e progredir com eles. É o que Bisol (1986) verificou para os descendentes de italianos de Veranópolis e Matté (2009), para os de Caxias do Sul. Mas não é o que Battisti et al. (2007) verificaram em Antônio Prado, onde a palatalização aplica-se numa taxa de 30%, mas em que há sinais de estabilização do processo em índices modestos.

Tanto a relação em rede dos informantes de Antônio Prado quanto a orientação positiva dos jovens pradenses explica o fato de a regra, inovadora, não estar progredindo no município, haja vista a estabilização nos grupos de 18 a 30 anos e de 31 a 50 anos. Já em Caxias do Sul e em Flores da Cunha, em que o índice do grupo de 18 a 30 anos aponta o avanço da palatalização, o que há de diferente parece ser uma maior mobilidade territorial dos indivíduos. Há mais oportunidades de contato com pessoas de fora da comunidade, o que abriria a fala a inovações como a palatalização e poderia explicar a tendência ao progresso da regra, a despeito de sua moderada aplicação.

O grupo *Status* da Vogal Alta contém dois fatores, *vogal alta fonológica* (*mentira*) e *vogal alta fonética* (*gente*), isto é, vogal alta que pode ser elevada de /e/ átono. Esperava-se encontrar resultados similares aos de Battisti et al. (2007) e Matté (2009), que apontam a vogal alta fonológica como condicionadora da regra. A hipótese foi confirmada, como mostra a Tabela 2.

Tabela 2 - Aplicação da regra de palatalização e *status* da vogal alta.

Fatores	Aplic./Total	%	Peso relativo
Alta fonológica (<i>mentira</i>)	5.053/8.183	62	0,89
Alta fonética (<i>gente</i>)	1.705/14.980	11	0,23
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

O mais forte condicionador da regra é a própria vogal gatilho, /i/. Os contextos com vogal alta fonológica correspondem a um terço dos dados e a frequência de palatalização neles é alta: 62%. Já nos contextos com vogal /e/ candidata a elevar-se a [i], que reúnem a maioria dos dados, a frequência é baixa, de apenas 11%. Isso explica a tendência atestada pelos valores de peso relativo obtidos: em Flores da Cunha, a regra tende a aplicar-se com /i/ e a ser inibida por /e/.

Confirma-se a tendência de a vogal alta fonológica condicionar a palatalização e de a vogal alta fonética inibir o processo. Se a elevação de /e/ átono é baixa, não se alimenta a palatalização. É essa baixa elevação de /e/ átono o que está em jogo em Flores da Cunha. Bisol (1986) atribui a tendência de palatalização com vogal alta fonológica à interação do processo com a variável Etnia, não no que se refere aos aspectos sociais envolvidos, mas a motivações predominantemente internas da variação e mudança, decorrentes do contato do português com o italiano, o alemão e o espanhol, no caso das comunidades por ela estudadas¹². A autora afirma:

¹² As etnias alemã, italiana e fronteiriça, contempladas pela autora, são representadas na amostra por informantes dos municípios gaúchos de Taquara, Veranópolis e Livramento, respectivamente.

Parece que o contato do dialeto gaúcho com línguas em que a palatalização da oclusiva não existe ou não segue os cânones da Língua Portuguesa [...] vem embargando o caminho de expansão da regra, reforçando-lhe a restrição peculiar nas comunidades monolíngues e acentuando-lhe o caráter de regra adquirida nas comunidades bilíngues. (BISOL, 1986, p.170).

O contato entre português e falares dialetais italianos se verifica ainda hoje na RCI-RS, mesmo que na forma de bilinguismo não generalizado e, em muitos casos, passivo¹³. Africadas como as que resultam da palatalização das plosivas alveolares em português existem nos falares italianos, como nos vocábulos **ciao** (olá) e **oggi** (hoje). Mas neles as africadas não são variantes. Além disso, as africadas opõem-se a /t/ e /d/. O que pode transferir-se dos falares italianos ao português é, sim, o reflexo da distinção gramatical veiculada pelas vogais /e/ e /i/ quando morfemas de plural específicos de gênero: **-e** é morfema feminino plural (*donna-donne*, mulher-mulheres), **-i**, morfema masculino plural (*bambino-bambini*, menino-meninos). Falantes bilíngues português-italiano, ou oriundos de comunidades onde ainda se verifica o bilinguismo, tendem a não elevar a média átona /e/, como reflexo de seus hábitos de fala em língua italiana, que requer a clara emissão e manutenção da vogal para garantir a veiculação de informação gramatical. Esse foi o raciocínio de Bisol (1981) no estudo de harmonia vocálica, em que a etnia italiana aparece como a que menos eleva as vogais médias átonas. Entendemos que tal raciocínio também se aplique ao estudo em questão: ainda há bilíngues na comunidade de descendentes de italianos que é Flores da Cunha. A elevação de /e/ átono tende a ser de baixa a moderada no município (ROVEDA, 1998; GUZZO, 2010), o que reduz o número de contextos com vogal alta fonética, que alimentariam a regra de palatalização.

O terceiro grupo selecionado foi Local de Residência. Noll (2008), sobre o português brasileiro, afirma que a palatalização é um processo urbano. É o que verificaram Battisti et al. (2007) em Antônio Prado e que se esperava verificar também em Flores da Cunha. A análise confirmou essa hipótese. Os resultados referentes a Flores da Cunha estão na Tabela 3.

Tabela 3 - Aplicação da regra de palatalização e local de residência.

Fatores	Aplic./Total	%	Peso relativo
Zona urbana	4.336/11.494	38	0,67
Zona rural	2.422/11.669	21	0,33
TOTAL	6.758/23.163	29	
Input 0,174	Significância 0,000		

Fonte: Elaboração própria.

¹³ Entende-se por bilinguismo passivo o de indivíduos que apenas comprehendem uma das línguas, não falam, não escrevem, tampouco leem essa língua, que é não dominante.

A frequência de palatalização, embora moderada na zona urbana (38%), é superior à verificada na zona rural (21%). Os pesos relativos obtidos indicam que o processo tende a aplicar-se na zona urbana de Flores da Cunha, favorecedora da regra, e é inibido na zona rural. Essa é a tendência também em Antônio Prado e em Caxias do Sul.

A razão para esse comportamento distinto e polarizado dos dois locais de residência pode ter relação, como já observaram Battisti et al. (2007) em Antônio Prado, com práticas sociais distintas realizadas pelos indivíduos e com a rede social que integram. As zonas rurais de municípios da RCI-RS abrigam pequenas propriedades rurais produtivas, quase autossuficientes e geridas por um núcleo familiar que trabalha e vive na terra. Afora os integrantes mais jovens da família, que se deslocam diariamente a escolas próximas para estudar, os demais membros apresentam pequena mobilidade territorial e, assim, travam contato com poucos desconhecidos. Sendo a palatalização aspecto inovador na fala em língua portuguesa, mas cuja difusão depende de contato interpessoal e de práticas compartilhadas, é de se esperar que a alternante palatalizada tenha pequeno emprego, pois os indivíduos da zona rural expõem-se pouco a indivíduos palatalizadores, a interação entre eles é quase nula. As redes sociais que formam são densas e multiplexas, impedindo a palatalização de nelas entrar e se propagar.

Outro aspecto que merece destaque é o fato de o bilinguismo português-italiano persistir na zona rural. Pelas razões abordadas acima quando da apresentação dos resultados da variável *Status* da Vogal Alta, esse contato preserva da elevação a vogal média que desencadearia a palatalização e que representa boa parte dos contextos analisados.

Os habitantes da zona urbana realizam práticas sociais diferenciadas, o que contribui para a difusão da regra. Tanto o setor de comércio quanto o da indústria e o de serviços propiciam interação com indivíduos de municípios vizinhos e, mais importante, de fora da RCI-RS. O próprio deslocamento para realizar curso superior leva, principalmente os jovens, a práticas compartilhadas com indivíduos diversos, flexibilizando laços que preveniriam mudanças de comportamento, inclusive o linguístico.

No quarto grupo de fatores selecionado, Qualidade da Consoante-Alvo, controlou-se a consoante candidata à palatalização, se a vozeada /d/ ou a desvozeada /t/. A hipótese era a de que em Flores da Cunha a consoante /t/ condicionaria a regra de palatalização, como visto em Battisti et al. (2007) e Matté (2009) para Antônio Prado e Caxias do Sul, respectivamente. Novamente, a hipótese foi confirmada pela análise. Os resultados de Flores da Cunha estão na Tabela 4.

Tabela 4 - Aplicação da regra de palatalização e qualidade da consoante-alvo.

Fatores	Aplic./Total	%	Peso relativo
Desvozeada (<i>tia</i>)	3.817/10.573	36	0,60
Vozeada (<i>dia</i>)	2.941/12.588	23	0,40
TOTAL	6.758/23.163	29	
Input 0,174	Significância 0,000		

Fonte: Elaboração própria.

A frequência de palatalização de /t/ (36%) não é tão distinta da palatalização de /d/ (23%). No entanto os pesos relativos são reveladores: contextos com /t/ tendem a ser palatalizados; já a consoante vozeada /d/ desfavorece o processo. A mesma tendência se verifica nos outros municípios, o que, retomando a hipótese de Guy (2000), se pode esperar de condicionamentos de natureza interna ou estrutural que refletem a unidade maior do português do Brasil. A explicação para o fato pode ser a de Dutra (2007, p.105-106):

[...] durante a articulação de consoantes [-voz], a energia articulatória se concentra nos órgãos supraglóticos [...] o processo de palatalização leva a uma grande concentração de energia na parte anterior da cavidade bucal e é justamente a articulação das consoantes [-voz] que propicia tal configuração articulatória.

Há, portanto, uma semelhança na articulação da consoante desvozeada /t/ que aproxima do processo de palatalização, explicando seu papel favorecedor.

Os demais grupos de fatores não foram descartados da análise estatística durante as rodadas dos programas computacionais, mas os valores obtidos ou giram em torno do ponto neutro, caso em que se atesta a ausência de papel da variável frente ao processo investigado, ou apresentam enviesamento na relação entre frequências e pesos relativos¹⁴. Por essa razão, a exposição dos resultados segue obedecendo a ordem de relevância estatística, mas a discussão será breve.

¹⁴ Enviesamentos ocorrem quando os maiores pesos não correspondem às maiores frequências e vice-versa.

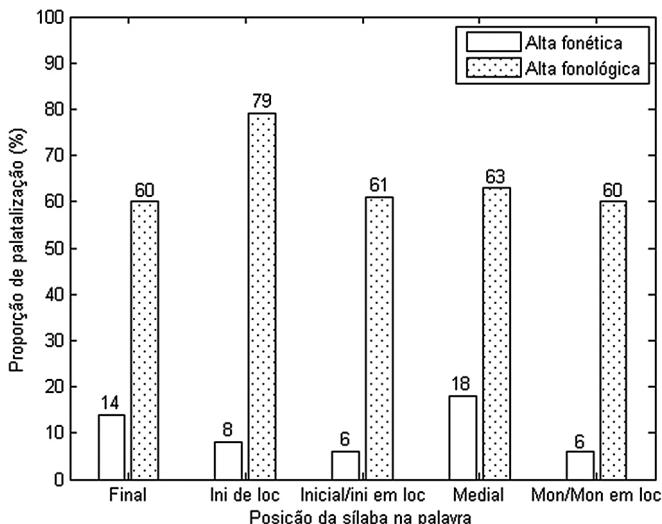
Tabela 5 - Aplicação da regra de palatalização e posição da sílaba na palavra.

Posição da sílaba na palavra	Aplic./Total	%	Peso relativo
Final (<i>pente</i>)	1.416/8.321	17	0,54
Inicial de locução (<i>de noite</i>)	390/4.743	8	0,50
Inicial/inicial em locução (<i>teatro, de teatro</i>)	3.028/6.517	46	0,47
Medial (<i>antigo</i>)	1.843/3.361	55	0,46
Monossílabo/monossílabo em locução (<i>diz</i>)	81/221	37	0,38
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

O maior peso relativo é o do fator Final, ainda assim em torno do ponto neutro. Além disso, há enviesamentos nos resultados. Como afirmamos na nota 7, fatores da variável *Status* da Vogal Alta se sobrepõem aos da variável Posição da Sílaba na Palavra como também aos de Tonicidade. O Gráfico 1 traz o cruzamento de *Status* da Vogal com Posição e talvez explique os resultados desse último grupo de fatores.

Gráfico 1 - Palatalização por *Status* da Vogal Alta x Posição da Sílaba na Palavra.



Fonte: Elaboração própria.

Todos os fatores considerados combinam-se e em todas as posições a palatalização é desencadeada por vogal alta fonológica. Praticamente não há diferença de aplicação entre as posições. É, portanto, o efeito do *status* da vogal, e não propriamente a posição da sílaba na palavra, o que conta para a palatalização.

A Tabela 6 traz os resultados da variável Contexto Fonológico Seguinte. Enviesamentos se verificaram novamente, mas há ao menos um valor de peso relativo que se destaca, o do fator Vogal Oral, que condiciona a aplicação da regra.

Tabela 6 - Aplicação da regra de palatalização e contexto fonológico seguinte.

Fatores	Aplic./Total	%	Peso relativo
Vogal oral (<i>teatro</i>)	1.115/2.552	44	0,59
Zero (<i>gente</i>)	1.216/7.884	15	0,53
Consoante sibilante/fricativa (<i>disco/tive</i>)	1.426/4.188	34	0,50
Consoante oclusiva/africada (<i>ditado, de/dʒ/inho</i>)	1.147/4.119	28	0,45
Consoante nasal/líquida (<i>tinha, diretora</i>)	1.854/4.420	42	0,45
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

Os demais fatores da variável Contexto Fonológico Seguinte, com pesos relativos em torno do ponto neutro, não têm papel atuante.

O controle estatístico não sugeriu descartar a variável Tonicidade, mas também aqui os resultados não são expressivos, todos os pesos relativos aproximam-se do ponto neutro. Veja-se a Tabela 7:

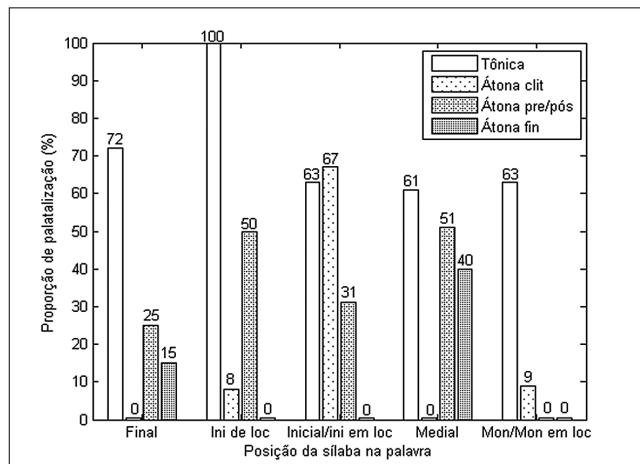
Tabela 7 - Aplicação da regra de palatalização e tonicidade.

Tonicidade da sílaba	Aplic./Total	%	Peso relativo
Tônica (<i>dia</i>)	3.004/4.768	63	0,53
Átona não final (<i>atirado, médico</i>)	2.104/5.452	39	0,52
Átona: clítico (<i>de, te</i>)	400/4.849	8	0,49
Átona final: (<i>vinte</i>)	1.250/8.094	15	0,47
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

O cruzamento da variável Tonicidade com Posição da Sílaba na Palavra sugere a razão dessa inexpressividade, como se vê no Gráfico 2:

Gráfico 2 - Palatalização por Tonicidade x Posição da Sílaba na Palavra.



Fonte: Elaboração própria.

O Gráfico 2 permite visualizar a inexistência de combinação de certos fatores: não poderia haver sílaba átona final a não ser na posição final da palavra; de forma complementar, as demais sílabas átonas ocupam as outras posições. Apenas a sílaba tônica ocupa todas as posições da palavra. Vale observar que a frequência de palatalização em sílaba tônica mantém-se significativa, independentemente da posição que ocupa na palavra. Mostra-se condicionadora da palatalização a Tonicidade da Sílaba, antes que Posição da Sílaba na Palavra.

Pelo número insuficiente de dados em alguns fatores da variável Contexto Fonológico Precedente, como também por enviesamentos entre frequências e pesos relativos, verificados na primeira rodada e em outras subsequentes, fizeram-se amalgamações, de que resultaram os quatro fatores registrados na Tabela 8.

Tabela 8 - Aplicação da regra de palatalização e contexto fonológico precedente.

Fatores	Aplic./Total	%	Peso relativo
Consoante sibilante (<i>poste</i>)	310/641	48	0,56
Vogal oral/ditongo (<i>cedido, leite</i>)	1.446/4.376	33	0,54
Zero (<i>tipo</i>)	3.398/11.088	31	0,49
Consoante lateral/vibrante/nasal (<i>malte, parte, mentira</i>)	1.604/7.058	23	0,48
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

O fator Consoante Sibilante aparece como condicionador da regra, com peso relativo de 0,56, mas o número total de contextos para esse fator não é significativo do ponto de vista estatístico¹⁵. Os pesos relativos dos demais fatores agrupam-se em torno do ponto neutro.

A variável Sexo/Gênero, embora não descartada na análise estatística, ocupa a última posição na ordem de relevância dos grupos de fatores. As proporções de aplicação, de 29% para cada sexo/gênero, e a neutralidade dos fatores indicada pelos pesos relativos, em torno de 0,50, permitem-nos afirmar que sexo/gênero não tem papel destacado na palatalização das plosivas alveolares em Flores da Cunha.

Tabela 9 - Aplicação da regra de palatalização e sexo/gênero.

Fatores	Aplic./Total	%	Peso relativo
Masculino	3.348/11.361	29	0,53
Feminino	3.410/11.802	29	0,47
TOTAL	6.758/23.163	29	
Input 0,174			Significância 0,000

Fonte: Elaboração própria.

A análise de regra variável da palatalização das plosivas alveolares em Flores da Cunha permite afirmar que o processo é condicionado pelos jovens, por vogal alta fonológica, por habitantes da zona urbana do município e por consoante-alvo desvozeada. Aos fatores das demais variáveis controladas não é possível atribuir papel claro, embora os resultados tenham demonstrado que exerçam algum efeito.

A análise de rede social dos informantes permitirá abordar por outro ângulo, o das práticas sociais, a correlação de jovens e habitantes urbanos com a palatalização. O conteúdo das entrevistas sociolinguísticas e o que se observou na participação em práticas da comunidade também contribuirão para essa abordagem.

A palatalização variável como prática social

Análise de rede

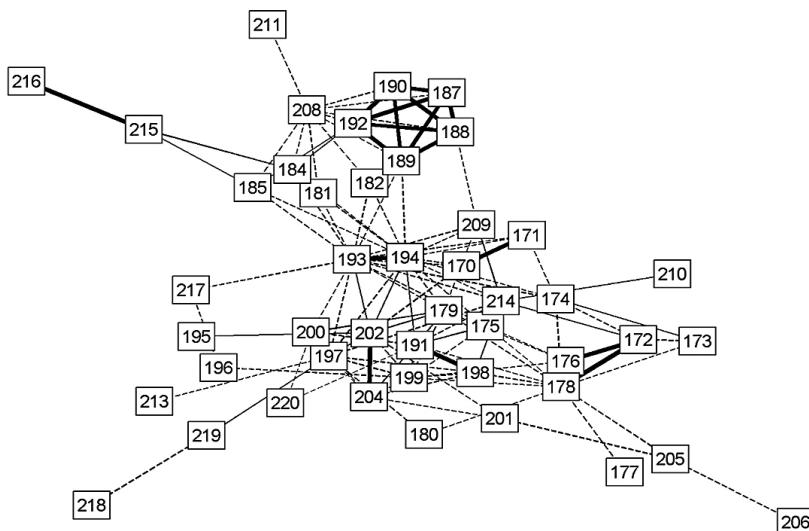
A rede social dos informantes de Flores da Cunha é densa e nela predominam os relacionamentos de terceiro grau. A Figura 2 traz a configuração da rede.

¹⁵ Não reúne o mínimo de 10% do total de dados da análise.

Os informantes são representados por retângulos contendo o seu número de identificação no BDSer. O relacionamento entre dois informantes é representado por uma linha unindo os retângulos, e os graus de relacionamento, por diferentes linhas: relacionamentos de primeiro grau são representados por uma linha contínua mais grossa; os de segundo grau, por uma linha contínua mais fina, e os de terceiro, por uma linha tracejada.

Figura 2 - Rede social dos informantes de Flores da Cunha.

Rede FC: Número do informante



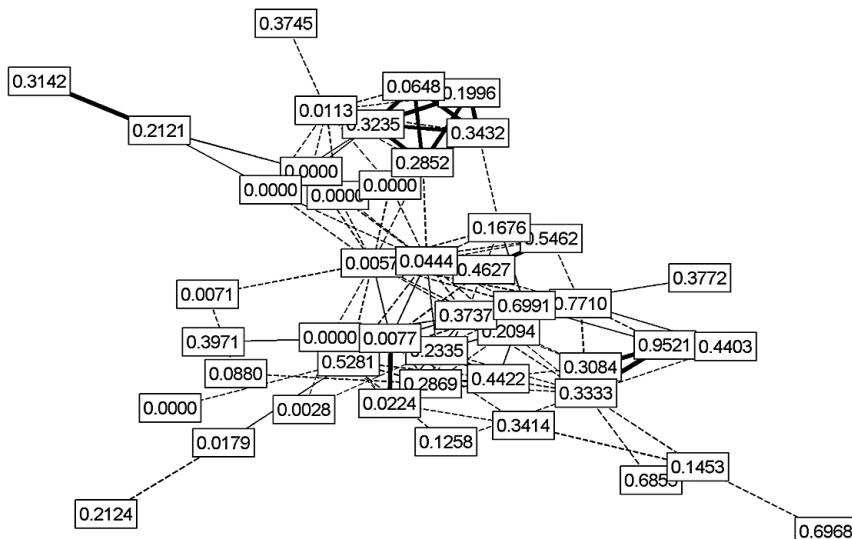
Fonte: Elaboração própria.

A rede possui 48 informantes ligados por 121 contatos. A proporção de conexão é de 0,1169 e está associada à densidade da rede social. Pode ser entendida como a probabilidade de que, escolhidos ao acaso, dois informantes tenham contato. A rede social dos informantes de Flores da Cunha é densa, com uma média de 5,26 contatos por informante. Alguns informantes, como os informantes 206, 211, 216 e 218, possuem contato com apenas um outro informante na rede. Outros informantes, como o 193 e o 194, apresentam contatos com muitos informantes da rede: 16 e 17 contatos, respectivamente. São elementos centrais na rede. Foram os primeiros contatos na comunidade e serviram de ponte para novos contatos na pesquisa.

A Figura 3 mostra a rede social de Flores da Cunha com o registro, nos retângulos, da proporção de aplicação da palatalização de cada informante.

**Figura 3 - Rede social de Flores da Cunha:
proporção de aplicação da regra variável.**

Rede FC: Proporção de aplicação RV

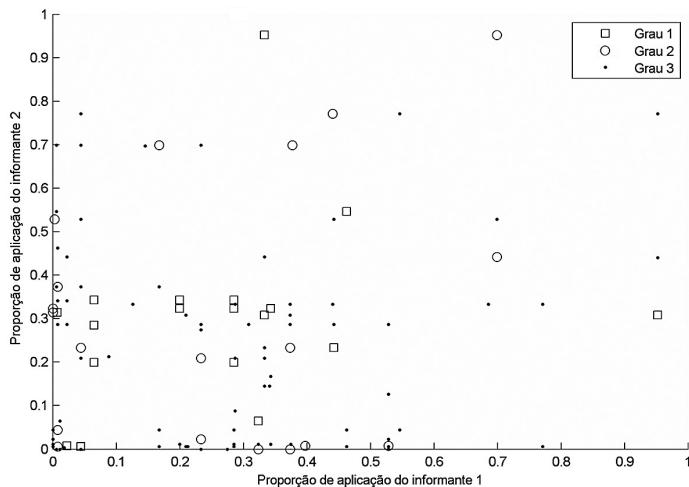


Fonte: Elaboração própria.

Correlacionadas a proporção de aplicação de um dado informante e a proporção de aplicação de seus contatos, verificou-se que os relacionamentos em rede têm efeito sobre o comportamento linguístico: sujeitos palatalizadores relacionam-se predominantemente com sujeitos palatalizadores e vice-versa. A Figura 4 representa, num diagrama de dispersão, essa correlação. Cada ponto representa um par de informantes. A forma do ponto representa o grau de relacionamento entre os informantes, conforme a legenda. Observa-se uma leve, porém significativa, correlação positiva ($0,1920 < r < 0,2129$; $0,0044 < \text{sig.} < 0,0105$)¹⁶ entre as proporções de aplicação dos informantes.

¹⁶ O coeficiente de correlação foi calculado levando em consideração o grau de relacionamento: pesos 3, 2 e 1 para graus de relacionamento 1, 2 e 3, respectivamente. O coeficiente de correlação foi calculado 1.000 vezes, alternando-se aleatoriamente os informantes 1 e 2 em cada relacionamento. Esse procedimento fez com que os valores do coeficiente de correlação e significância variassem. Os valores de r e sig. correspondem a um intervalo de confiança de 95%, seguindo um procedimento *bootstrap*.

Figura 4 - Proporção de aplicação de um dado informante e a proporção de aplicação de seus contatos.

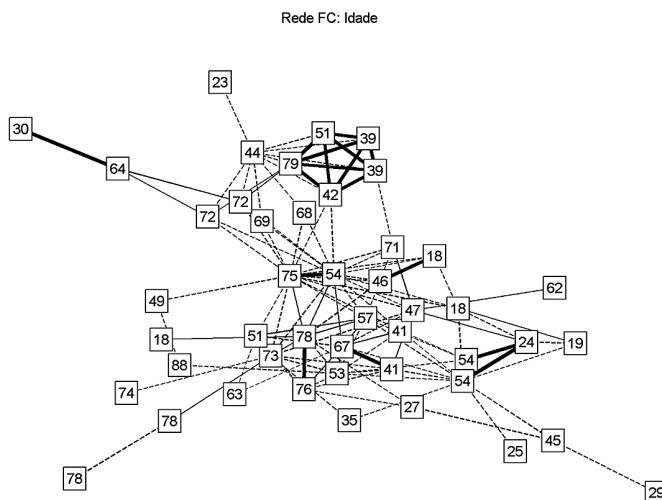


Fonte: Elaboração própria.

A Figura 5 mostra a rede social de Flores da Cunha com o registro da idade de cada informante. A Figura 6 mostra a rede social de Flores da Cunha com o local de residência dos informantes (U para urbano, R para rural).

Em relação à idade (Figura 5), chama a atenção o fato de não se visualizarem, na rede, aglomerações de sujeitos de um mesmo grupo etário.

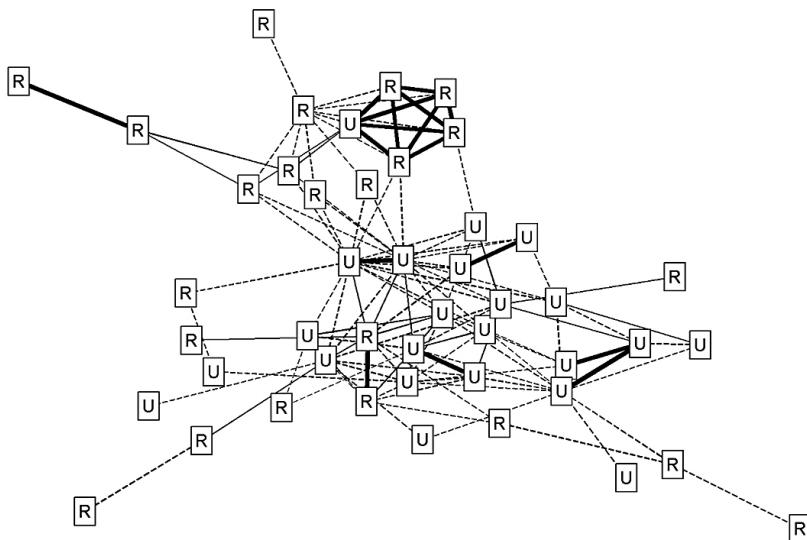
Figura 5 - Rede social de Flores da Cunha: idade.



Fonte: Elaboração própria.

Figura 6 - Rede social de Flores da Cunha: local de residência.

Rede FC: Local de residência



Fonte: Elaboração própria.

Sujeitos situados nos quatro grupos etários na análise de regra variável antes realizada não se relacionam exclusivamente com sujeitos do mesmo grupo etário. Em muitos casos, há relacionamentos familiares em jogo: de pais e filhos, tios, sobrinhos, avós e netos, como também de vizinhos e colegas de trabalhos de variadas idades, o que contribui para a ausência de aglomerações por similaridade etária. Relevante é o fato de os jovens, que condicionam a palatalização, situarem-se na periferia da rede, posição em que os efeitos de seu comportamento não são tão fortes como poderiam ser se ocupassem uma posição central na rede.

Quanto ao local de residência (Figura 6), são visíveis as aglomerações: sujeitos de zona urbana relacionam-se predominantemente com sujeitos de zona urbana e vice-versa. Os laços mais fortes, os de relacionamentos de primeiro grau, estabelecem-se na mesma zona; os mais fracos, de terceiro grau, ultrapassam os locais de residência. Isso sustenta a ideia de que as práticas sociais diárias em Flores da Cunha são localizadas. A mobilidade territorial não é acompanhada de relacionamentos fortes, que poderiam ter efeitos sobre o comportamento verbal dos sujeitos. Especialmente na zona rural, o trabalho na terra requer a permanência no local, o que não fomenta inovações.

Análise das entrevistas e observação participante

Os relatos dos entrevistados de Flores da Cunha sobre seu cotidiano repetem o que já se havia verificado em Antônio Prado, também situado na RCI-RS. Neles, há indícios de que a comunidade, embora não tenha aberto mão de práticas sociais tradicionais, ligadas à história da imigração italiana, abre-se a práticas inovadoras, o que se faz presente sobretudo na fala dos jovens. Por exemplo, considerem-se as faixas etárias usadas na análise quantitativa (18 a 30 anos, 31 a 50 anos, 51 a 70 anos e 71 ou mais anos). Quando os entrevistados falam sobre lazer, o peso dos meios de comunicação de massa é maior na fala dos entrevistados jovens e da faixa de 31 a 50 anos, diminuindo nas demais faixas etárias. Assistir a filmes no cinema e usar computador com acesso à Internet é mais frequente entre os jovens, principalmente os da zona urbana que se deslocam a outras comunidades para realizar estudos em nível médio e superior. Informantes de todas as faixas etárias assistem a programas de televisão e ouvem rádio. Todos afirmam ler, muda o tipo de texto lido: os mais jovens afirmam ler livros, os demais leem jornais e revistas, usualmente.

Percebe-se a importância da família como referência para as práticas do trabalho e, de variadas maneiras, para os momentos de lazer. Todos os informantes, se não afirmam textualmente que gostam do que fazem, dão a entender isso. Até os jovens urbanos, aqueles que têm maior trânsito fora dos limites de Flores da Cunha e relatam não haver na cidade *shopping*, teatro e cinema para frequentar, passam momentos agradáveis com amigos e familiares nos finais de semana, com quem praticam esportes e têm contato com a natureza.

Em se tratando dos jovens, que condicionam a palatalização e, assim, podem ser tomados como inovadores, a observação participante revela uma tendência similar: inovação com preservação. Percebemos nos jovens florenses os *jeans*, tênis, jaquetas, bonés e celulares que acompanharia os jovens em qualquer outro lugar do Rio Grande do Sul e do Brasil. Um pouco mais cuidados e “arrumadinhos”, é verdade, mas todos esses itens estão lá, no visual dos jovens, numa demonstração de que a prática é local, mas os artefatos são supralocais.

Os informantes são católicos em sua grande maioria. Entre os grupos mais jovens (18 a 30 anos e 31 a 50 anos), apesar da afirmação, na entrevista, de que não são praticantes, observa-se algum envolvimento em práticas religiosas, principalmente as festivas. Na celebração de *Corpus Christi*, por exemplo, que envolve uma missa na igreja matriz e, em seguida, a procissão do Senhor Morto sobre tapetes de serragem, constatamos a presença de número significativo de jovens. Não pareciam estar lá forçados, sua participação demonstrou-se fervorosa. Constatamos também que muitos tapetes haviam sido confeccionados por grupos de jovens – juventude católica, escoteiros, organizações antidrogas, grêmios desportivos.

Identificadas com o tradicional, as práticas religiosas são o exemplo claro de elemento local que faz parte da representação da relação imaginária que os indivíduos mantêm com suas condições de existência, que é assim componente da identidade construída na comunidade, mas que tem sua força minimizada por elementos não locais. É em cenários como esse que o novo, o não tradicional, entra, embora não com muita liberdade ou intensidade.

Talvez reforçada pelas festas locais e pelo turismo, que celebra e comercializa as raízes italianas, há uma consciência étnica que serve de explicação, para o próprio jovem, da razão de realizar certas práticas, como se pode captar de uma afirmação assim:

[...] a maioria (dos jovens) aqui da cidade tem descendência italiana. (Tu) sabe que descendência italiana sempre tem a matriarca e o patriarca. Eu acho que a maioria sai, assim, bastante “família” e sempre tem que jantar junto com o pai e com a mãe, salvo exceções. [...] Chega essa idade assim e “ah, tenho que casar, tenho que ter minha família”. (Inf.172, 24 anos, masculino, zona urbana).

A vida em família não parece ser algo que o jovem gostaria de evitar, tampouco um real desejo seu, mas algo que acontece com ele. O jovem florense nem se rebela contra a família nem adere incondicionalmente a ela, o que pode estar na base das graduais e pequenas mudanças diárias por ele promovidas, embora as tradições sejam seguidas, inclusive nos momentos de lazer:

Aqui (se) segue... um pouco de tradição da cidade, que no caso é, no domingo, sair e dar voltas ao redor da praça. É uma coisa que a cidade pequena tem. Ficar parado olhando o movimento passar. (Inf.172, 24 anos, masculino, zona urbana).

Essa fala não revela propriamente uma opção do jovem, mas uma consequência das tradições e, mais importante, do fato de a cidade não oferecer opções de lazer, comentário que soa como queixa na fala de muitos deles e que é razão para buscarem, em outras cidades, diferentes formas de distração:

Eu gostaria de viver em Caxias... Acho que desde pequena... uma vez eu fui na Festa da Uva lá e nossa!... sempre gostei de Caxias, tanto pequena quanto adolescente, sempre quis ir pra Caxias, mas minha mãe não deixava e, nossa... eu ia no cinema, ficava feliz, sempre gostei da cidade. (Inf.174, 18 anos, feminino, zona urbana).

Deslocar-se a outras localidades, dessa forma, parece fazer parte de ser jovem em Flores da Cunha, inclusive para cursar faculdade (não há instituição de Ensino Superior em Flores da Cunha), mas em um roteiro de ida e volta: o jovem realiza

práticas fora da comunidade e retorna a Flores da Cunha. Em termos linguísticos, tem contato com outros padrões de fala e pode, nessa circunstância, sentir-se pressionado a evitar formas reveladoras de sua identidade local, como empregar formas não palatalizadas. É possível então que, buscando sintonia com a fala do outro, abra mão momentaneamente de marcas locais e palatalize as plosivas alveolares. Depois, esse mesmo jovem retorna a Flores da Cunha e, lá, volta a orientar-se pela identidade e pelos valores locais, embora não realize práticas sociais exatamente da mesma forma que os florenses mais velhos. Sobre sair da comunidade e retornar a ela, e também sobre mudanças geracionais, observe-se a afirmação de outra jovem sobre trabalho:

A gurizada daqui não quer mais trabalhar na colônia [...] Vão pra Caxias estudar. Todos fazem faculdade, a grande maioria. Aí, então, ninguém mais quer trabalhar na colônia como acontecia antigamente, as famílias eram numerosas, as pessoas ficavam na colônia e não estudavam, né [...] Eu acho que mais pessoas continuam morando aqui e vão e voltam, do que se mudam. São poucos que se mudam. (Inf.211, 23 anos, feminino, zona rural).

A jovem começa afirmando que “a gurizada” não quer mais trabalhar na terra, nas propriedades rurais, junto à família. A suposição imediata seria a de que os jovens da zona rural desejariam migrar para a cidade, mas não é o que vai na última afirmação da jovem, repetida a seguir: “Eu acho que mais pessoas continuam morando aqui e vão e voltam, do que se mudam. São poucos que se mudam”. É o tipo de trabalho que jovens de zona rural criticam: muito duro, segundo alguns. Mas não criticam o lugar onde moram, que não abandonam, se possível. É o que observamos em campo. Nas propriedades rurais onde se cultiva uva, por exemplo – setor bastante lucrativo para os pequenos proprietários rurais da região –, jovens que, na adolescência, planejaram abandonar a zona rural afirmam ter se dado conta, alguns anos depois, de que a viticultura lhes daria um retorno financeiro interessante. E de fato trabalham na terra, acompanhando seus pais.

Sobre práticas linguísticas e o valor das formas alternantes, são raras as afirmações nas entrevistas sociolinguísticas, como é de se esperar. Mas veja-se uma interessante, de uma jovem que falava sobre padrões familiares:

Assim: começa sempre pelos avós, então, a criação dos meus tios, né, era uma e veio vindo... e a gente sempre pegava [...] o próprio falar: tu fala “erado”, e quem convive contigo que fala “erado” vai falar sempre “erado”. (Inf.171, 18 anos, feminino, zona urbana).

A jovem tem ciência de que, ao falarmos, tendemos a reproduzir práticas linguísticas. Ela seleciona uma alternante, o emprego de vibrante simples em lugar de múltipla, saliente para muitos florenses e para “os de fora” da comunidade, e

deixa transparecer que há algo de negativo que a acompanha. A jovem entende o emprego de vibrante simples como um erro que se poderia ou deveria evitar. Não é o mesmo o que ocorre com as formas palatalizadas. Um exemplo disso é o que vem a seguir. Em campo, após realizarmos uma entrevista sociolinguística com uma florense de zona rural do grupo etário de 51 a 70 anos, essa conta-nos que sua neta mais velha cursa Direito na universidade do município vizinho. A outra neta, de apenas oito anos de idade e que ouvia a conversa, corrige nossa réplica, em função de uma forma palatalizada:

- “... e a minha neta faz Direito [sem palatalização] em Caxias”, diz a entrevistada.
- “[Dj]ireito [com palatalização]? Que legal, bem importante”, replica a entrevistadora.
- “Não é **[Dj]**ireito, é **Direito** [com ênfase nas sílabas iniciais]”, corrige a neta.

O que pode ter originado essa correção? Lições de não palatalização não há na escola, a menina não teria reproduzido ali um ajuste aprendido em aulas de Língua Portuguesa. A forma palatalizada é, sem dúvida, saliente para a menina. Interpretamos a correção como uma solicitação de ajuste ao comportamento linguístico local, como um depoimento de que “não se fala assim aqui”. A variante palatalizada não é um erro com sanções sociais, mas uma forma pouco frequente ou quase inexistente nas interações cotidianas da menina com os seus e, quem sabe, nas interações de muitos outros florenses. As formas palatalizadas são entendidas como não locais e seu prestígio é relativo ao grupo social a que pertence o florense (rural ou urbano) e à situação de uso da linguagem (se entre florenses, ou entre florenses e não locais).

Em termos de práticas sociais e cultura local, então, vimos, com o estudo etnográfico, indícios de um hibridismo em que aspectos tradicionais e inovadores convivem, em que mudanças linguísticas (entre outras) ocorrem, mas lentamente.

Conclusão

A palatalização variável das plosivas alveolares, processo inovador no português falado em Flores da Cunha, foi analisada quantitativa e qualitativamente. Verificou-se que a palatalização tem condicionamento linguístico e social e produz formas salientes aos florenses, valoradas conforme suas práticas sociais locais.

A vogal alta fonológica condiciona a palatalização, a vogal alta fonética inibe o processo, porque a elevação de /e/ átono é baixa em Flores da Cunha, o que não alimenta a palatalização. Outro condicionador linguístico é a consoante-

alvo desvozeada /t/, cuja articulação concentra energia na parte anterior da cavidade bucal, à semelhança do processo de palatalização. Os condicionadores sociais são jovens e habitantes da zona urbana, sujeitos cujas práticas sociais de lazer, estudo e trabalho misturam elementos locais e não locais, expondo-os a grupos palatalizadores de outras comunidades e, assim, tornando sua fala porosa a mudanças.

A densidade da relação em rede dos informantes e o papel central que nele ocupam os sujeitos dos grupos etários mais velhos explicam por que a palatalização, introduzida no grupo pelos mais jovens, nele progride lentamente. As afirmações dos informantes nas entrevistas sociolinguísticas e os fatos observados na participação em atividades em Flores da Cunha possibilitaram verificar, em termos de práticas sociais, reprodução dos padrões tradicionais, ligados à história da imigração italiana, na gastronomia, na religião, nas formas de trabalho e de lazer, mas com algumas inovações. Em específico, os jovens, que tendem a palatalizar e fazem a regra progredir na comunidade, deslocam-se diariamente a outras localidades para frequentar faculdade e se divertir, mas retornam a Flores da Cunha, seguem morando nela. Observações como essas revelam um localismo peculiar, permitem-nos compreender o padrão de palatalização da comunidade e apontam a tendência de mudança fônica de progressão moderada no português falado em Flores da Cunha, a ser concretizada pelas gerações futuras.

BATTISTI, E.; DORNELLES FILHO, A. A. The palatalization of dental stops in Flores da Cunha: language variation and social practices. Alfa, São Paulo, v.56, n.3, p.1107-1139, 2012.

■ *ABSTRACT: The variable palatalization of dental stops in Brazilian Portuguese in the speech community of Flores da Cunha is moderate and contributes to the characterization of the local speech. The variable rule analysis (LABOV, 1972, 1994, 2001, 2010) of data collected from 48 sociolinguistic interviews of the corpus BDSer reveals that the total rate of rule application is 29%. The palatalization is conditioned by both linguistic and social variables and the process progresses in the community: underived high front vowel, voiceless target consonant, young people, inhabitants of the urban area condition palatalization. In the analysis of variation as social practice (ECKERT, 2000), the informants network analysis shows that its high density and the centrality of old subjects refrain palatalization. The ethnographic study reveals that people in Flores da Cunha, especially the young ones, engage in both innovative and traditional social practices and introduce palatalization in the speech community. The palatalization is valued as a non-local practice and a relatively prestigious one.*

■ *KEYWORDS: Language variation. Palatalization. Dental stops. Social practices.*

REFERÊNCIAS

- ALMEIDA, M. A. B. de. *A variação das oclusivas dentais na comunidade bilíngue de Flores da Cunha: uma análise quantitativa*. 2000. 106 f. Dissertação (Mestrado em Linguística Aplicada) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2000.
- BATTISTI, E. et al. Palatalização das oclusivas alveolares e a rede social dos informantes. *Revista Virtual de Estudos da Linguagem: ReVEL*, Porto Alegre, v.5, n.9, p.01-29, ago. 2007. Disponível em: <http://www.revel.inf.br/files/artigos/revel_9_palatalizacao_das_oclusivas_alveolares.pdf>. Acesso em: 27 fev. 2008.
- BATTISTI, E.; GUZZO, N. B. Palatalização das oclusivas alveolares: O caso de Chapecó. In: BISOL, L.; COLLISCHONN, G. (Org.). *Português no sul do Brasil: variação fonológica*. Porto Alegre: EDIPUCRS, 2010. p.114-140.
- BISOL, L. A palatalização e sua restrição variável. *Estudos: Lingüísticos e Literários*, Salvador, v.5, p.151-162, 1986.
- _____. *Harmonização vocálica: uma regra variável*. 1981. 335f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- BLAKE, R.; JOSEY, M. The /ay/ diphthong in Martha's Vineyard community: what can we say 40 years after Labov? *Language in Society*, Cambridge, v.4, n.32, p.451-485, 2003.
- BOAVENTURA NETTO, P. O. *Grafos: teoria, modelos, algoritmos*. São Paulo: Ed. Edgard Blücher, 1996.
- BOTTOMORE, T. B. *As elites e a sociedade*. Rio de Janeiro: Zahar, 1974.
- BOURDIEU, P. *A economia das trocas linguísticas: o que falar quer dizer*. 2.ed. São Paulo: EDUSP, 2008.
- CERQUEIRA FILHO, G. *A questão social no Brasil: crítica do discurso político*. Rio de Janeiro: Civilização Brasileira, 1982.
- DUTRA, E. de O. *A palatalização das oclusivas dentais /t/ e /d/ no município de Chuí, Rio Grande do Sul*. 2007. 129f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2007.
- ECKERT, P. *Linguistic variation as social practice*. Oxford: Blackwell, 2000.
- EVANS, B. The role of social network in the acquisition of local dialect norms by Appalachian migrants in Ypsilanti, Michigan. *Language Variation and Change*, Cambridge, n.16, p.153-167, 2004.

FROSI, V. M. *Provérbios dialetais italianos: uma linguagem em extinção*. 1989. 244f Dissertação (Mestrado em Letras) – Faculdade de Letras. Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1989.

FROSI, V. M.; MIORANZA, C. *Dialetos italianos: um perfil linguístico dos ítalobrasileiros do nordeste do Rio Grande do Sul*. Caxias do Sul: EDUCS, 1983.

FRUCHTERMAN, T. M. J.; REINGOLD, E. Graph drawing by force-directed placement. *Software: Practice and Experience*, London, v.21, n.11, p.1129-1164. 1991.

GERHARDT, G. J. L.; CORSO, G.; LEMKE, N. Network clustering coefficient approach for DNA sequences. *Chaos Solitons & Fractals*, Oxford, v.28, n.4, p.1037–1045, 2006.

GUY, G. R. A identidade linguística da comunidade de fala: Paralelismo interdialetal nos padrões de variação linguística. *Organon*, Porto Alegre, v.14, n.28/29, p.17-32, 2000.

GUZZO, N. B. *A elevação da vogal média anterior átona em Flores da Cunha*. 2010. 158f. Dissertação (Mestrado em Letras) – Universidade de Caxias do Sul, Caxias do Sul, 2010.

LABOV, W. *Principles of linguistic change: social factors*. Oxford: Blackwell, 2001.

_____. *Principles of linguistic change: internal factors*. Oxford: Blackwell, 1994.

_____. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press, 1972.

MATHEWS, J. H. *Numerical methods for mathematics, science and engineering*. Eglewood Cliffs: Prentice Hall, 1992.

MATTÉ, G. D. A palatalização variável de /t d/ em Caxias do Sul. *Cadernos do IL*, Porto Alegre, n.38, p.43-55, jun. 2009.

MILLS, C. W. *A elite no poder*. Rio de Janeiro: Zahar, 1962.

MILROY, L. *Language and social networks*. Oxford: Blackwell, 1980.

NOLL, V. *O português brasileiro: formação e contrastes*. São Paulo: Globo, 2008.

PAGOTTO, E. G. *Variação é identidade*. 2001. 454 f. Tese (Doutorado em Letras-Linguística) – Universidade Estadual de Campinas, Campinas, 2001.

ROVEDA, S. D. *Elevação da vogal média átona final em comunidades bilíngues: português e italiano*. 1998. 81f. Dissertação (Mestrado em Letras) – Faculdade

de Letras. Pontifícia Universidade Católica do Rio Grande do Sul, Rio Grande do Sul, 1998.

SABBATINI, M.; FRANZINA, E. *I veneti in Brasile nel centenario dell'emigrazione (1876-1976)*. Vicenza: Accademia Olimpica, 1977.

SPRADLEY, J. P. *The ethnographic interview*. New York: Holt, Rinehart and Winston, 1979.

Recebido em março de 2012

Aprovado em junho de 2012

ÍNDICE DE ASSUNTOS

- Alternância pronominal, p.1035
Aplicações, p.739
Atlas Linguístico do Brasil, p. 855, p.871
Atlas Linguístico, p.891
Banco de Dados de Fala, p.771
Banco de dados, p.917
Brasil, p.891
Composição do Comitê Nacional, p.871
Concordância de gênero, p.835
Concordância nominal, p.1065, p.1099
Concordância verbal, p.1035, p.1099
Corpus, p.945
Dialectologia pluridimensional, p.891
Diatopia, p.855
Divisão dialetal do Brasil, p.871
Estilo, p.917
Fala capixaba, p.807
Fatores sociais, p.917
Focalização dialetal, p.835
Formalismo, p.1065
Funcionalismo, p.1065
Gêneros textuais, p.739
Geolinguística, p.855
Identidade, p.945
Informantes, p.891
Leitura e Ensino, p.1117
Língua escrita, p.739
Língua falada, p.739
Mudança linguística, p.807
Palatalização, p.1003
Paulistanos, p.945
Plosivas alveolares, p.1003
Pluralidade, p.1065
Português brasileiro, p.807, p.973
Português paulista, p.1035
Práticas sociais, p.1003
Predicativo, p.1065
Produção científica, p. 871
Projeto PortVix, p.807
Pronomes de segunda pessoa, p.835
Região Centro-Oeste, p.835
Região Norte, p.973
Sociolinguística, p.739, p.917, p.1099, p.1117
Variação e mudança linguística, p.917
Variação Fonológica, p.771
Variação linguística, p.807, p.835, p.855, p. 945, p. 973, p.1003
Variação regional, p.807
Variação, p.1117
Variedade brasiliense, p.835
Variedade Cuiabana, p.835
Variedades do Português, p.1099
VARSUL, p.771
Vogais átonas, p.973

SUBJECT INDEX

- Applications, p. 739
Brazil, p.891
Brazilian Portuguese, p.807, p.973
Capixaba Speech (Speech from the state of Espírito Santo) , p.807
Constitution of the National Committee, p.871
Corpus, p.945
Databases, p.917
Dental stops, p.1003
Dialect focusing, p.835
Dialectal Division of Brazil, p.871
Diatopic variation, p.855
Formalism, p.1065
Functionalism, p.1065
Gender agreement, p.835
Geographical variation, p.807
Geolinguistics, p.855
Identity, p.945
Informants, p.891
Language variation, p.945, p.1003
Linguistic Atlas of Brazil, p.855, p.871
Linguistic Atlas, p.891
Linguistic Change, p.807
Linguistic variation and change, p.917
Linguistic Variation, p.807, p.855, p.973
Linguistic variety from Brasília, p.835
Linguistic variety from Cuiabá, p.835
Linguistics Variation, p.835
Midwest region, p.835
Multidimensional dialectology, p.891
Nominal agreement, p.1065
North of Brazil, p.973
Noun agreement, p.1099
Palatalization, p.1003
Paulista Portuguese, p. 1035
Paulistanos, p.945
Phonological Variation, p.771
Plurality, p.1065
PortVix Project, p.807
Predicate, p.1065
Pronominal Alternation, p. 1035
Reading and Teaching, p.1117
Scientific Production, p. 871
Second person pronouns, p.835
Social factors, p. 917
Social practices, p.1003
Sociolinguistics, p.739, p.917, p.1099, p.1117
Speech Database, p.771
Spoken Language, p.739
Style, p.917
Text Genre, p.739
Unstressed vowels, p.973
Variation, p.1117
Varieties of Portuguese, p.1099
VARSUL, p.771
Verb agreement, p.1099
Verbal Agreement, p. 1035
Written Language, p.739

ÍNDICE DE AUTORES
AUTHOR INDEX

- AGUILERA, V. DE A.; ALTINO, F. C., P. 871
BATTISTI, E.; DORNELLES FILHO, A. A., P.1003
BRANDÃO, S. F.; VIEIRA, S. R., P.1099
CAMACHO, R. G.; SALOMÃO, M. H., P.1065
CARDOSO, S. A.; MOTA, J. A., P.855
COLLISCHONN, G.; MONARETTO, V. DE O., P.771
CRUZ, R., P.973
DETTONI, R. DO V.; PACHECO, C. DA S.; ANDRADE, C. Q.; SCHERRE, M. M. P. P.835
FREITAG, R. M. K.; MARTINS, M. A.; TAVARES, M. A., P.917
HORA, D. DA.; AQUINO, M. DE F. S., P.1117
ISQUERDO, A. N.; ROMANO, V. P., P.891
MENDES, R. B.; OUSHIRO, L., P.945
PAIVA, M. DA C. DE; SILVA, V. L. P., P.739
RUBIO, C. F.; GONÇALVES, S. C. L., P.1035
YACOVENCO, L. C.; SCHERRE, M. M. P.; TESCH, L. M.; BRAGANÇA, M. L. L.; EVANGELISTA, E. M.; MENDONÇA, A. K. DE.; CALMON, E. N.; CAMPOS JÚNIOR, H. DA S.; BARBOSA, A. F.; BASÍLIO, J. O. S.; DEOCLÉCIO, C. E.; SILVA, J. B.; BERBERT, A. T. F.; BENFICA, S. DE A., P.807

NORMAS PARA APRESENTAÇÃO DE ORIGINAIS DA

Alfa: Revista de Linguística

1. Informações gerais

A *Alfa: Revista de Linguística*, financiada pela Pró-Reitoria de Pesquisa da UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho” – publica trabalhos inéditos de professores e pesquisadores, vinculados a instituições de ensino e pesquisa nacionais ou internacionais. Não serão aceitos, trabalhos produzidos por pesquisadores que não tenham o título de doutor. No caso de trabalhos em co-autoria, mesmo que haja autor(es) que esteja(m) cursando o doutorado, um dos autores, necessariamente, deve já possuir o título de doutor. A revista edita artigos, retrospectivas, resenhas, entrevistas e traduções vinculados a todas as linhas de pesquisa dos Estudos Linguísticos.

São aceitas apenas as resenhas de livros que tenham sido publicados no Brasil nos dois últimos anos e, no exterior, nos quatro últimos anos.

Sem o conhecimento da autoria, dois membros do Conselho Editorial emitem parecer sobre os trabalhos. Em caso de um parecer ser favorável e outro contrário, o trabalho será enviado a um terceiro parecerista, que também não terá acesso ao nome autor. Depois da análise, cópias dos pareceres serão encaminhadas aos autores juntamente com instruções para modificações, quando for o caso.

No caso dos textos produzidos por autores convidados, o *peer review* será realizado pelos membros do Conselho Editorial da revista.

Os trabalhos poderão ser redigidos em português, francês, inglês, espanhol ou italiano. Para artigos escritos em português, *TÍTULO*, *RESUMO* e *PALAVRAS-CHAVE* precedem o texto e *TITLE*, *ABSTRACT* e *KEYWORDS* sucedem o texto. Para artigos escritos em outros idiomas, *TÍTULO*, *RESUMO* e *PALAVRAS-CHAVE* que precedem o texto devem ser escritos no idioma do artigo; os que sucedem o texto deverão ser em português, no caso de artigos em inglês; e em inglês, no caso de artigos em francês, espanhol ou italiano.

Os trabalhos que não se enquadram nas normas da revista serão devolvidos aos autores, ou serão solicitadas adaptações, indicadas em carta pessoal.

Dados e conceitos emitidos nos trabalhos, bem como a exatidão das referências bibliográficas, são de inteira responsabilidade dos autores.

2. Apresentação dos trabalhos

Encaminhamento: Os autores devem realizar o cadastro (Login/Senha) no site da revista (www.alfa.unesp.br), na

seção Submissões Online, preencher corretamente o perfil e escolher a opção “AUTOR”. Após haver realizado esses passos, deve ir para “SUBMISSÕES ATIVAS” e iniciar o processo de submissão através do link “CLIQUE AQUI PARA INICIAR O PROCESSO DE SUBMISSÃO”, no qual irá realizar os cinco passos básicos:

1. Início: Iniciar o processo de submissão, confirmando se está de acordo com as condições estabelecidas pela revista (marcando as caixas de seleção das condições e da declaração de direito autoral) e selecionar a seção artigos;
2. Inclusão de metadados: indicar os dados principais – nome, sobrenome, e-mail, instituição, resumo da biografia, título e resumo;
3. Transferência de manuscritos: realizar a transferência do arquivo para o sistema;
4. Transferência de documentos suplementares: realizar a transferência de arquivos com informações suplementares, que funcionam como um apêndice ou anexo ao texto principal, tais como instrumentos de pesquisa, conjuntos de dados e tabelas, que seguem os padrões de ética de avaliação, fontes de informação normalmente não disponíveis para leitores, ou figuras e/ou tabelas que não podem ser integradas ao texto em si.
5. Confirmação: Concluir a submissão.

Após concluir os cinco passos acima descritos, o autor deve aguardar o e-mail do editor e, nesse ínterim, pode acompanhar todo o fluxo de seu trabalho, da submissão, aceite, avaliação, re-edição do original até a publicação. Os artigos, após a submissão, são designados aos avaliadores definidos pelo conselho ou editores da revista. A política de seleção dos artigos é definida pelos membros do Conselho Editorial, Consultivo e os Editores da revista, disponibilizadas na seção “Sobre a Revista”, “Processo de Avaliação por Pares”.

3. Preparação dos originais

3.1. Apresentação

A preparação do texto deve obedecer aos seguintes parâmetros: *Word for Windows*, fonte *Times New Roman* de tamanho 12, espaçamento um e meio entre linhas no corpo do texto, papel tamanho A4 (21 cm x 29,7 cm), margens esquerda e superior 3,0 cm, direita e inferior 2,0 cm e extensão total de no mínimo 15 páginas e no máximo 30, incluindo referências bibliográficas e anexos e/ou apêndices.

3.2. Estrutura do trabalho

Para elaboração do artigo, o autor(es) deve(m) obedecer à seguinte sequência:

1. **título** em caixa alta e em negrito, centralizado no alto da primeira página, em espaçamento simples entrelinhas;

2. texto, de, no mínimo, 150 palavras e, no máximo, 200, contendo resumo do artigo, que indique seus objetivos, referencial teórico utilizado, resultados obtidos e conclusão, precedido da palavra RESUMO, em caixa alta, duas linhas abaixo do nome do autor, sem adentramento e em espaçamento simples;
 3. palavras-chave, no máximo de sete, separadas por ponto, precedidas do termo PALAVRAS-CHAVE, em caixa alta, mantendo-se o espaçamento simples, duas linhas abaixo do resumo. Para maior facilidade de localização do trabalho em consultas bibliográficas, a Comissão Editorial sugere que as palavras-chave correspondam a conceitos mais gerais da área do trabalho;
 4. o corpo do texto inicia-se duas linhas abaixo das palavras-chave, em espaçamento um e meio entrelinhas;
 5. subtítulos correspondentes a cada parte do trabalho, referenciados a critério do autor, devem estar alinhados à margem esquerda, em negrito, sem numeração, com dois espaços de um e meio depois do texto que os precede e um espaço um e meio antes do texto que os segue;
 6. agradecimentos, quando houver, seguem a mesma diagramação dos subtítulos, precedidos da palavra **Agradecimentos**;
 7. título do artigo, em inglês (para artigos redigidos em português,
- francês, espanhol e italiano), inserem-se duas linhas abaixo do final do texto, em espaçamento simples, sem caixa alta e negrito.
 8. versão do resumo, em inglês (para artigos redigidos em português, francês, espanhol e italiano), em itálico, precedida da palavra *ABSTRACT*, em itálico e caixa alta, em espaçamento simples entre linhas, duas linhas depois do título do artigo em inglês;
 9. versão das palavras-chave, em inglês (para artigos redigidos em português, francês, espanhol e italiano), em itálico, precedida da expressão *KEYWORDS*, em itálico e caixa alta, em espaçamento simples entre linhas, duas linhas depois do abstract;
- OBS.: No tocante às três últimas instruções, artigos redigidos em inglês devem seguir a versão em português do título, do *RESUMO* e das *PALAVRAS-CHAVE*:
10. referências, sob o subtítulo **REFERÊNCIAS** alinhado à esquerda, em negrito e sem adentramento, devem ser mencionadas em ordem alfabética e cronológica, indicando-se as obras de autores citados no corpo do texto, separadas por espaço simples, duas linhas abaixo das palavras-chave em inglês (cf. 3.3.1 abaixo);
 11. duas linhas abaixo das referências, se considerado imprescindível, sob o subtítulo Bibliografia consultada, alinhado à esquerda, em negrito

e sem adentramento, podem ser indicadas, também em ordem alfabética e cronológica, obras consultadas ou recomendadas, não referenciadas no texto.

3.3. Outras instruções

3.3.1. Normas para referências

As referências devem ser dispostas em ordem alfabética pelo sobrenome do primeiro autor e seguir a NBR 6023 da ABNT, de agosto de 2002: espaço simples e um espaço entre cada obra. Caso a obra seja traduzida, solicita-se que se informe o nome do tradutor.

Exemplos:

Livros

Authier-Revuz, J. **Palavras incertas:** as não coincidências do dizer. Tradução de Cláudia Pfeiffer et al. Campinas: Ed. da UNICAMP, 1998.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia do trabalho científico.** 2. ed. São Paulo: Atlas, 1986.

Coracini, M. J.; BERTOLDO, E. S. (Org). **O desejo da teoria e a contingência da prática.** Campinas: Mercado das Letras, 2003.

Capítulos de livros

PECHEUX, M. Ler o arquivo hoje. In: Orlandi, E. P. (Org). **Gestos de leitura: da história no discurso.** Tradução de Maria das Graças Lopes Morin do

Amaral. Campinas: Ed. da UNICAMP, 1994. p.15-50.

Dissertações e teses

BITENCOURT, C. M. F. **Pátria, civilização e trabalho:** o ensino nas escolas paulista (1917-1939). 1988. 256 f. Dissertação (Mestrado em História) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1998.

Artigos em periódicos

SCLIAR-CABRAL, L.; RODRIGUES, B. B. Discrepâncias entre a pontuação e as pausas. **Cadernos de Estudos Linguísticos,** Campinas,n.26, p.63-77, 1994.

Artigos em periódicos on line

SOUZA, F. C. Formação de bibliotecários para uma sociedade livre. **Revista de Biblioteconomia e Ciência da Informação,** Florianópolis, n.11, p.1-13, jun. 2001. Disponível em: . Acesso em: 30 jun. 2001.

Artigos em jornal

BURKE, Peter. Misturando os idiomas. **Folha de S. Paulo,** São Paulo, 13 abr. 2003. Mais!, p.3.

EDITORA plagiou traduções de clássicos. **Folha de S. Paulo,** São Paulo, p.6, 4 nov. 2007.

Documento eletrônico

UNIVERSIDADE ESTADUAL PAULISTA. Coordenadoria Geral de Bibliotecas. Grupo de Trabalho Normalização

Documentária da UNESP. Normalização Documentária para a produção científica da UNESP: normas para apresentação de referências. São Paulo, 2003. Disponível em: . Acesso em: 15 jul. 2004.

Trabalho de congresso ou similar (publicado)

MARIN, A. J. Educação continuada. In: CONGRESSO ESTADUAL PAULISTA SOBRE FORMAÇÃO DE EDUCADORES, 1., 1990. **Anais**...São Paulo: UNESP, 1990. p.114-8.

3.3.2. Citação no texto

O autor deve ser citado entre parênteses pelo sobrenome, em letras maiúsculas, separado, por vírgula, da data de publicação (BARBOSA, 1980). Se o nome do autor estiver citado no texto, indica-se apenas a data entre parênteses: “Morais (1955) assinala...”.

Quando for necessário especificar página(s), estas deverão seguir a data, separadas por vírgula e precedidas de p. (MUNFORD, 1949, p.513).

As citações de diversas obras de um mesmo autor, publicadas no mesmo ano, devem ser discriminadas por letras minúsculas após a data, sem espaçamento (PESIDE, 1927a), (PESIDE, 1927b). Quando a obra tiver dois ou três autores, todos poderão ser indicados, separados por ponto e vírgula (OLIVEIRA; MATEUS; SILVA, 1943), e quando houver mais de 3 autores, indica-se o primeiro seguido de et al. (GILLE et al., 1960).

Citações diretas em mais de três linhas deverão ser destacadas com recuo de 4 cm da margem esquerda, com letra tamanho 11, sem aspas e espaço simples entrelinhas. Citações com menos de três linhas devem seguir o fluxo normal do texto e virem destacadas apenas entre aspas.

3.3.3. Notas

Notas devem ser reduzidas ao mínimo e colocadas no pé da página; remissões para o rodapé devem ser feitas por números, na entrelinha superior, após o sinal de pontuação, quando for o caso.

3.3.4. Ilustrações

Ilustrações compreendem figuras, desenhos, gráficos, quadros, mapas, esquemas, fórmulas, modelos, fotografias, radiografias. As legendas devem ser inseridas abaixo das ilustrações, numeradas consecutivamente em algarismos arábicos e centralizadas. As figuras, os desenhos, os gráficos, os quadros, os esquemas, as fórmulas e os modelos devem ser enviados em arquivo separado, no programa em que foram gerados. Os mapas, as fotografias e as radiografias também devem ser enviadas em arquivos separados e em alta resolução (300 dpi).

3.3.5. Tabelas e quadros

Tabelas devem ser usadas para apresentação de informações tratadas estatisticamente e quadros para

sintetizar e organizar informações textuais. O título de tabelas e quadros insere-se na parte superior, centralizado, e a numeração é consecutiva, em algarismos arábicos; caso seja necessário especificar a fonte dos dados, esta deverá ser colocada abaixo da tabela ou do quadro e o texto, alinhado à esquerda. Tabelas devem ser construídas com as bordas laterais abertas e sem linhas de separação de colunas.

go 5º da Lei 9610/98. O ARTIGO poderá ser acessado tanto pela rede mundial de computadores (WWW – Internet), como pela versão impressa, sendo permitidas, A TÍTULO GRATUITO, a consulta e a reprodução de exemplar do ARTIGO para uso próprio de quem a consulta. ESSA autorização de publicação não tem limitação de tempo, FICANDO A UNESP responsável pela manutenção da identificação DO AUTOR do ARTIGO.

3.3.6. Anexos e/ou Apêndices.

Quando imprescindíveis à compreensão do texto, e inclusos no limite de 30 páginas, **Anexos e/ou apêndices**, seguindo formatação dos subtítulos, devem ser incluídos no final do artigo, após as referências bibliográficas ou a bibliografia consultada.

3.3.7. Transferência de direitos autorais – Autorização para publicação

Caso o artigo submetido seja aprovado para publicação, JÁ FICA ACORDADO QUE o autor AUTORIZA a UNESP a reproduzi-lo e publicá-lo na *Alfa: Revista de Linguística*, entendendo-se os termos “reprodução” e “publicação” conforme definição respectivamente dos incisos VI e I do arti-

STAEPE – Seção Técnica de Apoio ao Ensino, Pesquisa e Extensão
Laboratório Editorial
Rodovia Araraquara-Jaú, km 01
14800-901 – Araraquara
Fone: (16) 3334-6275 ou 3334-6234
e-mail: laboratorioeditorial@fclar.unesp.br
site: <http://www.fclar.unesp.br/laboratorioeditorial>

Produção Editorial:



Impressão:

