

O “social” da sociolinguística: o controle de fatores sociais

Raquel Meister Ko. Freitag¹

1. Introdução

A Sociolinguística Variacionista, impulsionada especialmente pelos projetos que desenvolveram bancos de dados, tem se mostrado um campo de estudos altamente produtivo no cenário nacional dos estudos linguísticos, contribuindo para uma ampla descrição do português brasileiro. O quanto esta descrição reflete de fato a realidade social do Brasil em termos linguísticos é uma questão que merece atenção, já que os bancos de dados sociolinguísticos nem sempre refletem a estratificação da sociedade. Assim é pertinente tecer reflexões sobre a forma de controle dos “fatores sociais clássicos” em fenômenos de variação e mudança, especialmente naqueles dos níveis gramaticais mais altos (cf. Freitag, 2009), em que nem sempre os efeitos do controle dos fatores sociais são estatisticamente significativos nos estudos. Para ilustrar as reflexões, toma-se o banco de dados VARSUL, a partir do qual se centra em discutir a questão da homogeneização dos resultados em função da homogeneização da amostra; alertar para a sobreposição de papéis sociais agrupados no rótulo “faixa etária” (cf. Eckert, 1997; Freitag, 2005); e discutir um problema encontrado no banco de dados VARSUL no que se refere à escolaridade: o comportamento anômalo da faixa “ginasial”, com dados de trabalhos que focam a amostra de Florianópolis e fenômenos semântico-discursivos.

¹ Professora da Universidade Federal de Sergipe (rkofreitag@uol.com.br)

Não é objetivo discutir as influências do social (a noção de social de Durkheim, por exemplo), mas apenas as implicações deste “social” da metodologia da Sociolinguística Variacionista, e em que medida os resultados obtidos são relativizados. Para iniciar, são apresentados bancos de dados sociolinguísticos, com especial atenção ao VARSUL e aos aspectos sociais de sua constituição.

2. Banco de dados e homogeneização da amostra

Para viabilizar os estudos sociolinguísticos em termos de prazos e otimização de recursos, ao invés de cada pesquisador realizar a sua própria coleta de dados para a sua investigação, tem se tornado prática a constituição de bancos de dados constituídos de acordo com uma metodologia de coleta muito específica (Labov, 1984). No Brasil, o método da entrevista sociolinguística é o mais difundido. O Programa de Estudos sobre o Uso da Língua – PEUL <<http://www.lettras.ufrj.br/peul/index.html>>, conhecido originalmente como Projeto Censo da Variação Linguística do Estado do Rio de Janeiro, foi pioneiro em adotar a metodologia da Sociolinguística Variacionista no Brasil, com o objetivo de estudar o português falado no Rio de Janeiro. Nos mesmos moldes, o banco de dados VARSUL <<http://varsul.cce.ufsc.br>> é resultado do projeto Variação Linguística Urbana na Região Sul do Brasil, cujos objetivos são o armazenamento e a disponibilização de amostras de fala de habitantes característicos de áreas urbanas representativas de cada um dos três estados da região sul do Brasil. Nos bancos de dados elencados, a dimensão “sócio” da Sociolinguística é implementada por meio do controle de estratificações sociais que podem ser aferidas, como a faixa etária, a escolaridade, o sexo/gênero do informante. Porém, 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?

A amostra sociolinguística deve ser representativa de um determinado grupo, denominado “comunidade de fala”; para Labov (2001, p. 38), uma amostra verdadeiramente representativa de uma comunidade de fala precisa tomar como base uma coleta aleatória em que cada um dos muitos falantes que a constituem tenha a mesma chance de ser selecionado. Nem sempre é possível controlar todos os fatores sociais, e, às vezes, o controle de um fator pode quebrar a ortogonalidade dos demais fatores, gerando células sociais impossíveis de serem

preenchidas, como, por exemplo, a relação entre escolarização e faixa etária, em que certos níveis de escolarização geram células vazias quando cotejados à faixa etária: salvo raras exceções, não existem crianças e adolescentes que têm curso superior (Freitag, 2005). Ou, ainda, no atual cenário brasileiro, é provável que se encontrem mais adultos analfabetos do que jovens, ou que se encontrem mais analfabetos na região rural do que na região urbana. Apesar destes "buracos" na amostra, as análises sociolinguísticas costumam homogeneizar o tamanho da amostra e a distribuição por célula. Idealmente, seriam necessários entre quatro e cinco informantes por célula social (Oliveira e Silva, 2004; Guy e Zilles, 2007), entretanto, alguns bancos de dados, como o VARSUL, foram constituídos em sua fase inicial com apenas dois indivíduos por célula.

A possibilidade de homogeneização da amostra, equiparando-se o número de indivíduos das células sociais independentemente da sua representatividade no conjunto real, é uma decorrência da assunção da premissa de que a natureza do sistema linguístico é probabilística (Labov, 2001). Embora os bancos de dados tenham apresentado resultados empíricos efetivos para a descrição do português, não se pode negar que existem problemas na sua constituição. No caso do VARSUL, a amostra básica de cada cidade é formada por um conjunto de 24 entrevistas, correspondentes a 12 perfis sociais (sexo masculino e feminino, três níveis de escolarização e duas faixas etárias) de duas entrevistas (exceto Florianópolis, que dispõe de uma amostra de informantes da faixa etária 15 a 21 anos), totalizando 288 entrevistas (96 em cada Estado). A meta inicial do projeto era contar com um número mínimo de cinco informantes por célula social, mas por conta de questões financeiras, a amostra ficou restrita a apenas dois informantes por célula. Este recorte fomenta questões tais como: A amostra é representativa? A homogeneização da amostra é pertinente? Ou seja, será que os mesmos dois informantes são suficientes para representar uma cidade como Porto Alegre, com aproximadamente 3.900.000 habitantes, e Panambi, com cerca de 36.000 habitantes? A homogeneização dos perfis sociais não distorceria os resultados? E se, por algum descuido, um dos dois informantes não atenda aos critérios? Um informante equivocadamente alocado no banco pode alterar os resultados. A falta de proporcionalidade entre a amostra do banco de dados e a população efetiva da cidade torna o resultado da análise destoante da realidade.

As entrevistas que constituem o banco de dados estão estratificadas em três níveis de escolarização. Novamente, cabem questionamentos, motivados por constatações empíricas: será que a homogeneização dos

três níveis de escolaridade reflete a realidade social de todas as cidades da amostra? Até que ponto a amostra do banco de dados VARSUL reflete a realidade socioeconômica da cidade considerada? Até que ponto os resultados de uma amostra homogeneizada podem ser generalizados para um grupo maior, heterogêneo? Calvet (2002) questiona se é válida a relação entre a heterogeneidade social e a homogeneização da amostra, o que pode vir a produzir efeitos de interação entre as variáveis sociais e linguísticas; Mollica e Roncarati (2001) também consideram esta questão ao tratar de uma agenda de trabalho para a área no Brasil. A homogeneização da amostra do VARSUL, especialmente no que se refere à estratificação por faixas de escolaridade, pode ser a responsável pelos resultados distorcidos e incongruentes, como tem acontecido em muitos dos trabalhos realizados com o banco de dados. Diante deste quadro, o que fazer com os resultados obtidos nas análises? Como interpretá-los? O que eles indicam? Em fenômenos variáveis nos níveis gramaticais mais altos, nem sempre os fatores sociais se mostram relevantes; e, quando se manifestam, faixa etária e tempo de escolarização costumam se mostrar mais frequentemente significativos. Nas seções a seguir, são discutidos os resultados associados aos fatores sociais *faixa etária* e *escolarização* em estudos nos domínios gramaticais mais altos, com base nos dados coletados na amostra do Banco de Dados VARSUL de Florianópolis.

3. Idade e mudança linguística

A idade é uma das três supercategorias sociais nas sociedades industrializadas modernas, junto com a classe e o sexo, e seu atributo social é a correlação primária com a mudança linguística. Intuitivamente, percebemos a influência da idade nos processos de variação e mudança linguística: uso de uma expressão “fora de moda”, gírias desatualizadas, enfim, percebemos que o tempo passou e ainda guardamos traços daquela época em nosso repertório linguístico.

Para operacionalizar essas intuições acerca da relação entre língua e faixa etária, Labov (1994) propõe uma metodologia que se resume à observação de dois estados de língua e a garantia de que haja continuidade entre eles. Em um dado momento, são coletados dados do fenômeno de uma amostra x . E, passado um período y , repete-se a coleta de dados, na mesma amostra x . A observação de um estado de uma língua é feita através de estudo quantitativo de uma amostra aleatória e representativa de todos os segmentos de uma comunidade de fala. Estudos desse tipo, chamados estudos em tempo real, se subdividem em *estudo de tendência* e *estudo de painel*. O estudo de tendência (*trend study*) é mais simples: requer uma

amostra aleatória da mesma comunidade de fala em um período y, posterior ao da primeira coleta. Já o estudo de painel (*panel study*) é mais complexo, pois requer o recontato com os mesmos indivíduos informantes da primeira coleta, com a aplicação do mesmo instrumento. É possível estabelecer algumas correlações entre estudos de tempo real e de tempo aparente no que se refere à estabilidade/instabilidade da mudança e a relação entre comunidade e indivíduo. Se o comportamento linguístico dos indivíduos é estável durante toda a sua vida e o comportamento linguístico da comunidade também, não há variação a analisar. Já se os indivíduos mudam seu comportamento linguístico durante o decorrer da sua vida e a comunidade não mostra a mesma mudança, o padrão é caracterizado como gradação etária. Mudança geracional e mudança comunitária não são transparentes, requerem um controle mais refinado para serem identificados.

A outra estratégia para identificar, descrever e analisar um dado fenômeno de variação ou de mudança linguística em um período de tempo reduzido proposta por Labov (1994) é que a mudança pode ser observada em tempo aparente. Essa saída metodológica pressupõe que a idade cronológica dos indivíduos represente uma "passagem no tempo", e se apoia na hipótese clássica de que a língua de um indivíduo se constitui até cerca de seus quinze anos de idade.

A análise em tempo aparente considera a distribuição das ocorrências do fenômeno em estudo em função das faixas etárias para caracterizar uma situação de estabilidade, mudança incipiente, mudança em progresso ou mudança completa. Eckert (1997), porém, vê problemas em estudos que consideram só tempo aparente: a estratificação etária pode refletir mudança em uma comunidade de fala em relação ao tempo (mudança histórica) e também a mudança na fala de um indivíduo em relação ao tempo de sua vida (gradação etária). Segundo a autora, considerar o tempo refletido na idade cronológica dos indivíduos pode levar a equívocos entre mudança em tempo aparente de fato e gradação etária. Isso porque o comportamento linguístico de todos os indivíduos muda no decorrer de sua vida e mudanças linguísticas individuais não são exclusivamente decorrentes de mudanças linguísticas históricas, são mudanças decorrentes da história do indivíduo: nascemos, crescemos, nos tornamos adultos, envelhecemos. A cada etapa do ciclo vital, mudanças de ordem biológica e social ocorrem e se refletem na língua do indivíduo: a aquisição da língua, a entrada na escola, a aplicação da rede de relações sociais, a entrada e a saída do mercado de trabalho são fatores que se refletem diretamente nas faixas etárias. Para Eckert (1997), a faixa etária

é apenas um rótulo que agrupa vários fatores de ordem social e biológica do indivíduo.

Para efeitos da constituição da amostra é preciso definir quantas e quais as faixas etárias que podem ser controladas e que fornecem pistas significativas para a compreensão real do fenómeno de variação e de mudança linguística. Labov (1994) propõe duas faixas extremas: a dos mais velhos e a dos mais jovens. Chambers (2003) propõe três: crianças, adolescentes e adultos. Eckert (1997), por sua vez, propõe que as faixas etárias representem o curso da vida linguística: infância, adolescência, vida adulta e velhice. Vejamos os efeitos do recorte das faixas etárias em um fenómeno variável do nível semântico-discursivo no português: a variação entre as formas de pretérito imperfeito e de passado progressivo na expressão do passado imperfectivo (Freitag, 2007). Do ponto de vista semântico-discursivo, o passado imperfectivo refere-se a uma situação anterior ao momento de fala e simultânea ao ponto de referência; em português, pode se realizar de duas formas: pretérito imperfeito do indicativo (IMP) e o passado progressivo (PPROG):

(1) Na época que eu mais precisei dele, que eu mais *precisava* (IMP) de um apoio, foi quando a minha mãe morreu. (SC FLP 03)

(2) Aí também foi na época que a gente voltou, a gente *estava precisando* economizar pra começar nossa vida. (SC FLP 01)

As formas são intercambiáveis quando assumem o valor semântico-discursivo passado imperfectivo, funcionando como variantes de uma mesma variável linguística (Freitag, 2007). Na variação na expressão do passado imperfectivo na fala de Florianópolis, a variável faixa etária foi apontada como estatisticamente significativa, como podemos observar na Tabela 1.

Tabela 1: Expressão do passado imperfectivo quanto à faixa etária em função de IMP (variante conservadora)

Faixa etária	Peso relativo	%	Aplicação/total
15 a 21 anos	0,34	42,7	91/213
25 a 49 anos	0,57	70,1	356/508
Mais de 50 anos	0,48	61,5	99/161

A distribuição dos dados da função está concentrada na faixa etária 25 a 49 anos. Quais as motivações para explicar esta distribuição centralizada? Há uma série de fatores que devem ser considerados. Primeiramente, a característica da constituição da amostra: entrevistas sociolinguísticas, que privilegiam narrativas. Indivíduos que, apesar das instruções do roteiro da entrevista, produziram poucas narrativas em suas entrevistas influenciam nesse resultado. Ter o que contar também tem relação com o papel social do indivíduo. Uma dona de casa pode ter menos fatos reportáveis do que um indivíduo que é ativo no mercado de trabalho. Já entre indivíduos ativos, aqueles que lidam diretamente com o público, como auxiliar de escritório, bancário ou professora primária, possivelmente são mais desinibidos do que aqueles colocados em funções que não exigem interação direta com o público, como cozinheira, ou auxiliar de serviços. A inibição/desinibição pode se refletir nas entrevistas: inibidos fariam menos do que os desinibidos, com a consequência de menor probabilidade de realizar o fenômeno analisado. Há que se considerar ainda que a extensão da faixa etária 25 a 49 anos é muito ampla, especialmente se comparada à faixa de 15 a 21 anos. Talvez se fosse desmembrada em duas faixas (25 a 35 anos e 36 a 49 anos) os resultados ficassem distribuídos de maneira mais equilibrada.

Considerando o contexto mais específico de uso das formas – duratividade com pares mínimos –, o padrão de distribuição das frequências da expressão do passado imperfectivo na fala de Florianópolis é quase linear. Há uma forte correlação entre o uso de PPROG e a faixa etária mais jovem, em oposição às faixas mais velhas, que apresentam um comportamento estatisticamente mais próximo.

Os resultados apontam para uma leve e incipiente tendência à polarização entre faixa etária mais jovem, implementando a variante inovadora, PPROG. Apesar das restrições consideradas, os jovens assumem papel de vetores na implementação de PPROG como forma de expressão de passado imperfectivo, configurando um quadro de *mudança em tempo aparente*. Evidentemente, estudos comparando outras sincronias são necessários para averiguar se de fato se trata de mudança histórica ou apenas efeitos da gradação etária, conforme Eckert (1997), ou se a mudança se dá na comunidade ou no indivíduo, por meio de gradação etária ou por mudança geracional, respectivamente, conforme o modelo de Labov (1994).

4. A escolaridade

O controle da variável escolaridade é bastante recorrente na sociolinguística brasileira. De acordo com Bortoni-Ricardo (2004, p. 48), “os anos de escolarização de um indivíduo e a qualidade das escolas que frequentou também têm influência em seu repertório sociolinguístico. Observe-se que esses fatores estão intimamente ligados ao estatuto socioeconômico, na sociedade brasileira”.

Sobre os efeitos da escolaridade nas investigações do PEUL, no Rio de Janeiro, Paiva e Scherre ponderam: “É possível também que a influência da variável escolaridade reflita, na verdade, a ação da variável classe social. Se assim for, as consequências são ainda mais perversas: não se modificam variantes linguísticas, mas, sim, se excluem os indivíduos que não possuem determinadas variantes linguísticas.” (Paiva; Scherre, 1999, p. 217-218)

A *escolaridade*, ou nível de escolarização, é um fator de estratificação social do banco de dados VARSUL que tem apresentado um comportamento irregular e pouco previsível em fenômenos de variação em níveis gramaticais mais altos (cf. Freitag, 2009). Apesar do comportamento irregular, as pesquisas realizadas partem do pressuposto de que a escolarização afeta os fenômenos de variação. Tomem-se os casos da variação na expressão do passado imperfectivo (já apresentado na seção 3), variação do passado anterior, ordem do quantificador, expressão do futuro, variação entre presente do indicativo e subjuntivo, concordância com o pronome *tu*, que foram analisados no banco de dados VARSUL considerando a cidade de Florianópolis, com três faixas etárias e três faixas de escolarização.

A expressão do passado anterior, analisada por Coan (1997), trata da variação no uso de formas do pretérito perfeito *vs.* pretérito mais-que-perfeito composto para codificar uma situação passada em relação a outra, como em (3) e (4).

(3) Aí eu peguei, telefonei pra Macarronada e descobri que *aconteceu* um acidente. (SC FLP FAB 03)

(4) Aí eu peguei, telefonei pra Macarronada e descobri que *tinha acontecido* um acidente.

Back (2000) analisa a variação da ordem do quantificador universal (QU) no sintagma nominal (SN) em contextos em que o fenômeno se realiza nas formas *tudo*, *todo/ toda* e *todos/ todas*, como em (5) e (6).

- (5)a. *todos* os meus tios ficaram muito chateados. (SC FLP FAC 20)
b. os meus tios *todos* ficaram muito chateados.
c. os meus tios ficaram *todos* muito chateados.

- (6)a. porque ela conhecia as famílias *todas*. (SC FLP FBC 24)
b. porque ela conhecia *todas* as famílias.
c. porque as famílias, ela conhecia *todas*.

A expressão de futuro enquanto categoria linguística, de acordo com a análise de Gibbon (2000), tem, entre outras, três variantes: *presente do indicativo*, *forma perifrástica*, composta pelo auxiliar IR (*presente do indicativo*) + *infinitivo*, e o *futuro do presente*, como em (7), (8) e (9).

(7) A seleção que *vai ter* em março, de repente, *faço* carreira. (SC FLP MAP 04)

(8) Tu não *vais ter* matéria pra estudar e *chega* no dia da prova tu não *consegues* a média. (SC FLP MAG 10)

(9) Eu acho que o dia que o povo der conta de que a educação é a base de tudo, acho que nós não *teremos* guerra, não *teremos* briga, não *teremos* nada, pelo contrário, o mundo *vai viver* em paz. (SC FLP MBG 13)

Pimpão (1999) analisa o uso variável entre a forma de presente de subjuntivo e de presente do indicativo em contextos de subjuntivo (dúvida, incerteza, futuridade), como em (10).

(10) Ela tem muitos que ela não prefere, né? Aí é. Professor de Física porque quer que ela *vá* de short curto: "Ah, mãe, não sei porque ele quer que eu *vou* de short curto". Porque ela vai de short mais comprido, ele acha que tem que ser mais curto (SC FLP FAC 11).

E Loregian (1996) analisa a variação na concordância verbal com o pronome *tu*, como ilustrado em (11) e (12).

(11) Assim tu *queres* parecer igual aquelas pessoas (SC FLP MJG 21)

(12) Tu *descasca* o camarão, depois tu *bota* tudo na fervura. (SC FLP FAP 02)

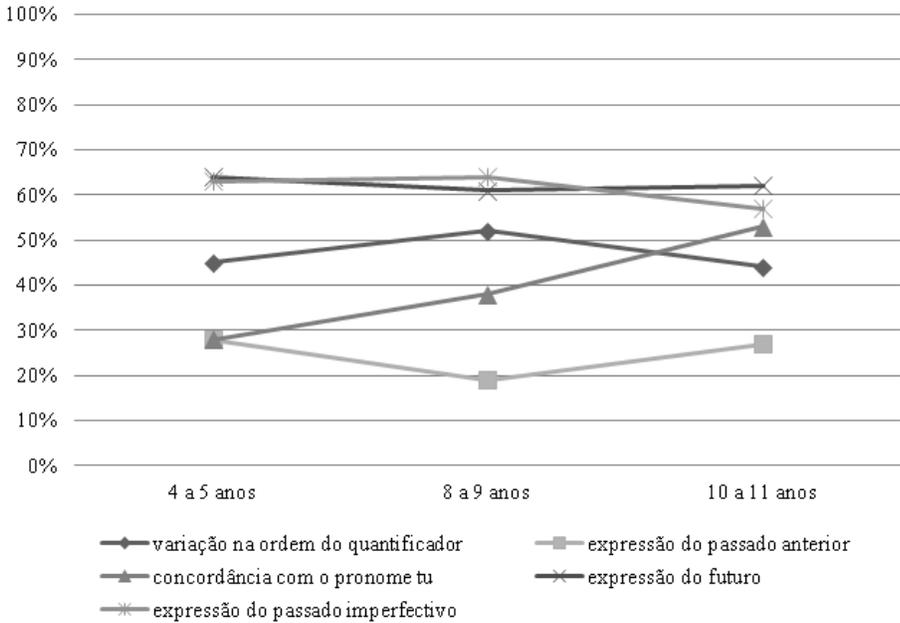
Para todos os fenômenos, foram postuladas hipóteses relativas ao fator escolaridade que previam uma distribuição escalar. Back (2000) correlaciona a ordem [QU SN], prevista na gramática normativa, com a faixa de maior escolaridade: “a hipótese que cerca esta variável entende que os informantes devem apresentar comportamento diferenciado conforme o grau de escolaridade de cada um, de modo que esperamos encontrar o uso da ordem [QU SN] na fala de pessoas de nível colegial, pois esse uso reflete aquele tido como ‘correto’, segundo os parâmetros da gramática normativa abordados nas escolas.” (Back, 2000, p. 93)

Pimpão também prevê a associação entre o uso da forma de subjuntivo, prevista pela gramática normativa, aos níveis mais altos de escolaridade, pois “a escola concentra o domínio do padrão culto tentando preservar as imposições da tradição normativa para o uso linguístico não-estigmatizado e de prestígio. O efeito da escolaridade a variáveis linguísticas prevê o uso mais próximo à norma gramatical para os níveis escolares mais elevados. A maior permanência na escola pressupõe o contato mais direto e intenso do ensino prescritivista sobre o idioleto do aluno.” (Pimpão, 1999, p. 90). Já Coan (1997) vê com mais cautela a correlação entre o uso de uma forma e a escolaridade dos informantes:

Por se tratar de um fenômeno não claramente percebido, o uso da variante pretérito perfeito em lugar do mais-que-perfeito para codificar a função de anterioridade é bem pouco discriminado e corrigido nas escolas. [...] O nosso objeto de estudo não parece ser tão marcante em termos de “certo” ou “errado”. A todo o momento, encontramos indivíduos de todos os níveis de escolaridade e de todas as classes fazendo variação. A possibilidade de itens linguísticos variarem sem que lhes seja atribuído qualquer estigma pode facilitar uma mudança linguística (Coan, 1997, p. 12).

Os resultados quanto à escolaridade, entretanto, não corroboraram as hipóteses levantadas, com exceção da variação na concordância com o pronome *tu*, que se mostra distribuída escalarmente em função da escolaridade. Na variação na ordem do quantificador, expressão do futuro, expressão do passado imperfeito e expressão do passado anterior, a faixa de escolarização intermediária, “ginásio”, desvia a trajetória da linha formada pelas faixas etárias extremas, conforme aponta o Gráfico 1.

Gráfico 1: Influência da variável escolaridade em fenômenos variáveis semântico-discursivos em função da variante considerada como conservadora (Florianópolis/VARSUL)



Tomando apenas as faixas de escolarização extremas, com exceção do fenômeno de variação na concordância verbal com o pronome *tu*, os demais fenômenos não parecem ser sensíveis ao nível de escolarização do indivíduo. Como explicar a anomalia produzida pelos informantes pertencentes ao nível intermediário de escolaridade no projeto VARSUL? Não se trata de um comportamento específico da amostra de Florianópolis; para Porto Alegre, Loregian (1996, p. 13) encontra resultados na variação da concordância verbal com o pronome *tu* que são alterados por causa da faixa etária intermediária. Como lidar com a variável escolaridade no projeto VARSUL? O que a escolaridade indicia? Há duas hipóteses que podem ser exploradas: i) a possibilidade de interação entre escolaridade e faixa etária; e ii) como a variável se correlaciona com os mecanismos de promoção ou resistência à mudança linguística?

Na primeira hipótese, é possível considerar que o comportamento incongruente do nível de escolarização seja decorrente da sua interação com as faixas etárias que compõem a amostra, principalmente por conta dos informantes que cursaram o antigo ginásio, que é diferente do nível

de 8 a 9 anos de estudo (5^a a 8^a série do Ensino Fundamental). O antigo ginásio vigorou no período de 1931-1969 e exigia que o estudante prestasse um rigoroso exame de admissão, além de uma diversidade de disciplinas que hoje sequer são vistas no Ensino Médio, como latim, retórica e francês. Indivíduos que cursaram o antigo ginásio teriam, por hipótese, um comportamento muito diversificado do daqueles que cursaram o equivalente à 5^a a 8^a série do Ensino Fundamental.

Outra possibilidade, também relacionada com a interação de fatores, é que a homogeneização das faixas de escolaridade do banco de dados VARSUL pode estar interferindo na incongruência dos resultados. A amostra Censo, do PEUL, por exemplo, não preenche as células relativas à escolaridade com o mesmo número de informantes. Segundo Oliveira e Silva e Paiva (1996, p. 60), “como uma amostra deve refletir até certo ponto as características da população, e considerando que nesta última não há distribuição equilibrada entre os níveis de escolarização, havendo muito mais pessoas no 1^o grau do que no 2^o grau, optou-se por escolher 19 falantes do *primário*, 16 do *ginásio* e 13 do 2^o grau.” Os resultados do PEUL que consideram a amostra Censo apresentam uma distribuição escalar para a escolaridade; talvez se o número de informantes por estrato fosse o mesmo, os resultados seriam diferentes.

É necessário refletir sobre os aspectos que estão envolvidos na escolarização e como estes aspectos interagem com o ensino e a valoração social das formas que expressam passado imperfectivo no português. Votre (2004, p. 51-54) enumera quatro aspectos da escolarização que podem ser verificados na dinâmica social em que a escola interage e que podem estar associados à variação e mudança linguística: i) formas de prestígio e formas relativamente neutras; ii) fenômeno socialmente estigmatizado e fenômeno imune à estigmatização; iii) fenômenos objeto do ensino escolar e fenômenos que escapam à atenção normativa; e iv) fenômeno discursivo (ou no mais alto nível gramatical) ou fenômeno da gramática (níveis gramaticais intermediários entre o fonológico e o discursivo).

Os três primeiros aspectos não parecem influenciar na variação em domínios gramaticais mais altos, cujas formas são relativamente neutras e os fenômenos parecem imunes à estigmatização. O fato de serem neutros socialmente torna-os pouco salientes ao ensino normativo. Porém, o último aspecto parece ser relevante, pois se relaciona com a opcionalidade (ou não obrigatoriedade) do fenômeno. As categorias verbais estão a serviço da intenção dos propósitos comunicativos do falante; são recursos estilísticos à disposição no repertório de estratégias comunicativas. A variação das categorias que expressam tempo e aspecto

está relacionada com fatores internos e cognitivos, como a relação entre figura/fundo e valores inerentes ao item lexical, como aspecto (Coan, 1997) ou movimento (Gibbon, 2000). O controle da variável escolaridade na variação entre IMP e PPR OG na expressão de passado imperfectivo, apesar de estatisticamente significativo, não mostrou um resultado pontual e clarificador da situação, mas motivou uma reflexão sobre a pertinência do fator na análise do fenômeno. Talvez a escolaridade não seja um fator influente na variação entre formas para a expressão de tempo e aspecto, e controlá-lo sobrecarregaria o modelo estatístico, além de frustrar expectativas. Ou a escolaridade é apenas a ponta do *iceberg* dos fatores não estratificados (como poder aquisitivo, rede de relações sociais, engajamento social etc.) e seus resultados devem ser avaliados com uma lente multifocal. Ou, ainda, a estratificação homogênea da escolaridade na amostra do VARSUL é a causa das incongruências nos resultados.

5 Considerações finais

A adoção de bancos de dados sociolinguísticos é uma maneira mais otimizada de subsidiar estudos sociolinguísticos. As ponderações tecidas neste texto, tomando por base a atuação de "fatores sociais clássicos", como a escolarização e a faixa etária, em fenômenos de variação e mudança nos níveis gramaticais mais altos, podem auxiliar a interpretação de resultados de outros estudos, nos mesmos moldes, assim como podem balizar a constituição de novos bancos de dados e a ampliação dos já existentes, na medida em que colaboram para que estes reflitam, de fato, a realidade social do Brasil.

Referências

- BACK, Angela C. Di Palma. *O uso variável do quantificador universal no sintagma nominal na língua falada de Florianópolis*. Dissertação (Mestrado em Linguística) — Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, 2000.
- BORTONI-RICARDO, Stella Maris. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2004.
- CALVET, Louis-Jean. *Sociolinguística: uma introdução crítica*. São Paulo: Parábola, 2002.
- CHAMBERS, J. K. *Sociolinguistics*. 2. ed. Oxford: Blackwell, 2003.
- COAN, Márluce. *Anterioridade a um ponto de referência passado: pretérito (mais*

que) perfeito. Dissertação (Mestrado em Linguística) — Programa de Pós-graduação em Linguística, Universidade Federal de Santa Catarina, 1997.

ECKERT, Penelope. "Ages as a sociolinguistic variable". In: COULMAS, Florian (org.). *The handbook of sociolinguistics*. Oxford: Blackwell, 1997, pp. 151-167.

FREITAG, Raquel Meister Ko. *A expressão do passado imperfeito no português: variação/gramaticalização e mudança*. Tese (Doutorado em Linguística) — Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2007.

FREITAG, Raquel Meister Ko. "Problemas teórico-metodológicos para o estudo da variação linguística nos níveis gramaticais mais altos". *Matraga*, v. 16, pp. 115-132, 2009.

FREITAG, Raquel Meister Ko. "Idade: uma variável sociolinguística complexa". *Línguas & Letras*, v. 6, pp. 105-121, 2005.

GIBBON, Adriana. *A expressão do futuro na língua falada em Florianópolis: variação e gramaticalização*. Dissertação (Mestrado em Linguística) — Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, 2000.

GUY, Gregory; ZILLES, Ana Maria Stahl. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola, 2007.

LABOV, William. "Field methods of the project on linguistic change and variation". In: BAUGH, J.; SHERZER, J. (orgs.). *Language in use*. Englewood Cliffs: Prentice Hall, 1984, pp. 28-54

LABOV, William. *Principles of linguistic change: internal factors*. Oxford: Blackwell, 1994.

LABOV, William. *Principles of linguistic change: social factors*. Oxford: Blackwell, 2001.

LOREGIAN, Loremi. *Concordância verbal com o pronome tu na fala do Sul do Brasil*. Dissertação (Mestrado em Linguística) — Programa de Pós-graduação em Linguística, Universidade Federal de Santa Catarina, 1996.

MOLLICA, Maria Cecília; RONCARATI, Cláudia Nívea. "Questões teórico-descritivas em Sociolinguística e em Sociolinguística Aplicada e uma proposta de agenda de trabalho". *DELTA*, v. 17, n. esp., pp. 45-55, 2001.

OLIVEIRA e SILVA, Gisele Machline. "Coleta de dados". In: MOLLICA, Maria Cecília; BRAGA, Maria Luíza (orgs.). *Introdução à sociolinguística: o tratamento da variação*. 2. ed. São Paulo: Contexto, 2004, pp. 117-133.

OLIVEIRA e SILVA, Gisele Machline; PAIVA, Maria Conceição. "Visão de conjunto das variáveis sociais". In: OLIVEIRA e SILVA, Gisele Machline; SCHERRE, Maria Marta. *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996, pp. 335-378.

PAIVA, Maria da Conceição; SCHERRE, Maria Marta Pereira. "Retrospectiva sociolinguística: contribuições do PEUL". *DELTA*, v. 15, n.esp., pp. 201-232, 1999.

PIMPÃO, Tatiana Schwochow. *Variação no presente do modo subjuntivo: uma abordagem discursivo-pragmática*. Dissertação (Mestrado em Linguística) — Programa de Pós-graduação em Linguística, Universidade Federal de Santa Catarina, 1999.

VOTRE, Sebastião Josué. "Relevância da variável escolaridade". In: MOLLICA, Maria Cecília; BRAGA, Maria Luíza (orgs.). *Introdução à sociolinguística: o tratamento da variação*. 2. ed. São Paulo: Contexto, 2004, pp. 51-57.

Resumo

Neste texto, são discutidos aspectos relacionados ao controle dos fatores sociais “faixa etária” e “escolarização” em estudos sociolinguísticos de cunho variacionista, focando a estratificação social de bancos de dados sociolinguísticos brasileiros. São discutidos aspectos da constituição dos bancos de dados dos projetos PEUL e VARSUL, com atenção aos efeitos da homogeneização metodológica imposta à amostra, especialmente no que se refere à faixa etária e escolarização. Resultados de estudos de variação e mudança em domínios gramaticais mais altos são utilizados para respaldar a discussão.

Palavras-chave: Banco de dados sociolinguístico. Homogeneização da amostra. Fatores sociais.

Abstract

In this paper aspects related to control of social factors “age” and “education” in variacionist sociolinguistic studies are discussed, focusing the social stratification of sociolinguistic Brazilian databases. Aspects of constitution of PEUL’s and VARSUL’s databases are discussed with special attention to methodological effects of homogenization that are imposed upon sample, especially in relation to age and education. Results of studies of variation and change in higher grammatical domains are used to support the discussion.

Keywords: Sociolinguistic databank. Sample homogenization. Social factors.