



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

SP2010 – Construção de uma amostra da fala paulistana

Ronald Beline Mendes (Departamento de Linguística-USP)

Resumo

O presente projeto objetiva coletar, transcrever e disponibilizar uma amostra recente da fala paulistana, composta de 90 entrevistas sociolinguísticas, com vistas a possibilitar o desenvolvimento de trabalhos de descrição e análise dessa comunidade ainda pouquíssimo estudada de uma perspectiva sociolinguística. Disponibilizar tal amostra (gravações e transcrições), em página própria na internet, significa contribuir para a disseminação e consolidação de trabalhos neste campo de estudos, não apenas na instituição sede, mas também em outros centros de pesquisa. Tendo em mente a complexa composição sociodemográfica da capital paulista, propõe-se a estratificação desse *corpus* a partir de três parâmetros sociais – região da cidade, gênero/sexo e faixa etária do falante –, para que se obtenha, dentro do prazo previsto de dois anos, uma amostra abrangente e representativa da cidade. A construção dessa amostra também estará atenta a distinções de gênero (papéis sociais definidos e construídos em correlação ao sexo) e de classe social, assim como à importância de tais distinções na percepção que os membros dessa comunidade têm acerca do que significa “ser paulistano” e “falar como paulistano”, atendendo à premissa sociolinguística básica de que o interesse de variantes linguísticas reside em seu significado social (Chambers, 1995). A metodologia de coleta segue os preceitos da Sociolinguística Variacionista (Labov, 2006 [1966], 1972; Tagliamonte, 2006) e os critérios de transcrição são elaborados com o intuito de facilitar a manipulação dos arquivos através de novas ferramentas computacionais (programas R, ELAN).

Palavras-chave: sociolinguística; português paulistano; português falado; construção de *corpus*

1. Introdução

Pode-se dizer que a Sociolinguística é bastante desenvolvida no Brasil. Especificamente a sociolinguística cujas bases teóricas e metodológicas são comumente conhecidas como variacionistas é tão presente e forte no cenário brasileiro que tal fato chega a ser reconhecido por Labov 2006, em sua nova edição de “*The Social Stratification of English in New York City*”, revista e comentada pelo próprio autor. De fato, os dialetos brasileiros foram e vêm sendo estudados através de todo o território nacional, através das pesquisas desenvolvidas no interior de vários projetos, dentre os quais podem-se destacar o VARSUL, o VALPB, o PEUL, o ALIP.¹

Por outro lado, o português da cidade de São Paulo, a maior cidade do país, recebeu muito pouca atenção dos sociolinguistas. O falar paulistano é geralmente identificado via certas variantes – tais como a pronúncia da coda silábica /-r/ como tepe em palavras como “caderno” e a ditongação do /e/ nasal em palavras como “setenta” – mas há pouquíssimos trabalhos de natureza quantitativa sobre as variáveis que incluem essas e outras variantes. Na verdade, são numerosos os trabalhos que se valem de dados do português falado em São Paulo, sobretudo aqueles que foram coletados pelo Projeto NURC/SP (Castilho & Preti, 1986, 1987; Preti & Urbano, 1988, 1990), mas o interesse central deles, na grande maioria das vezes, não é descrever “o falar paulistano”. Nas palavras de Oushiro 2011b, “pode-se afirmar que há um grande número de trabalhos *com base* no português paulistano, mas pouquíssimos *sobre* o português paulistano”.²

Exceções podem ser exemplificadas pelos 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 o trabalho de Rodrigues é um dos poucos que organiza uma amostra relativamente recente do português falado na cidade de São Paulo: ainda que circunscrito localmente numa favela, trata-se de um pequeno *corpus* construído a partir de objetivos variacionistas de pesquisa. O segundo investigou a variação no emprego dos pronomes *nós* e *a gente* e a concordância verbal com *nós*, através de 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

¹ Respectivamente, “Variação Linguística Urbana no Sul do País” (<http://www.pucrs.br/fale/pos/varsul/index.php>), que reúne pesquisadores de universidades dos três estados do sul do Brasil; “Variação Linguística no Estado da Paraíba” (Hora, 2004); “Programa de Estudos sobre o uso da Língua” (<http://www.lettras.ufrrj.br/peul/>), no Rio de Janeiro; e “Amostra Linguística do Interior Paulista” (Gonçalves, 2003; <http://www.iboruna.ibilce.unesp.br/index.php>), no estado de São Paulo.

² Grifos no original.

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 uma amostra coletada de acordo com as práticas sociais dos indivíduos em seus grupos, observados na comunidade.

Mais recentemente, o Projeto para a História do Português Paulista (Castilho 2007), financiado pela FAPESP, também vem desenvolvendo estudos mais propriamente *sobre* o português paulistano, variedade que integra sua agenda, de modo que tal projeto deve ser lembrado aqui; no entanto, considerando-se que seus interesses mais centrais são evidentemente de natureza histórica, ainda faz falta um corpo de pesquisas que descreva o português paulistano.

Nesse sentido, este projeto propõe a construção de uma amostra recente da fala paulistana, que permita sua descrição e análise. É claro que amostrar a fala de uma comunidade urbana impõe desafios, sobretudo numa grande cidade como São Paulo. À guisa de introdução, vale apontar que a amostra aqui proposta será organizada com base em três conjuntos de fatores extralinguísticos, devidamente justificados adiante (ver seção 3): o sexo dos informantes (masculino e feminino); três faixas etárias (20-34 anos; 35-59; 60 anos ou mais); e três regiões da cidade (centro velho, centro expandido e periferia). Fazendo-se todas as combinações possíveis desses fatores, tem-se 18 perfis diferentes. A proposta é gravar 5 informantes de cada perfil, de modo que, ao final, a amostra contenha 90 entrevistas sociolinguísticas.

Ao mesmo tempo, conforme a descrição mais detalhada do projeto permite verificar a seguir, a construção da amostra estará atenta a distinções de gênero (papéis sociais definidos e construídos em correlação ao sexo) (Romaine, 2003) e de classe social (Ash, 2004), bem como estará atenta à importância de tais distinções na percepção que os membros dessa comunidade têm acerca do que significa “ser paulistano” (e, de acordo com o interesse mais específico deste projeto, do que significa “falar como paulistano”).

É igualmente importante indicar que tal amostra vai se delimitar a indivíduos nascidos e criados na cidade, ou seja, migrantes de outras regiões do Brasil não serão considerados neste projeto. Não há dúvidas de que os migrantes tenham papel fundamental na constituição sociolinguística da cosmópole paulistana. Da mesma forma, é indubitável que a imigração, talvez sobretudo a italiana (mas também a japonesa, a portuguesa, a árabe), tenha destaque na definição de São Paulo na sua diferenciação sociolinguística diante de qualquer outra grande cidade brasileira. Entretanto, é justamente por ser tão embarçosamente limitado o conhecimento (em termos sociolinguísticos) que se tem da variedade paulistana que, num primeiro momento, faz-se necessária a coleta de uma amostra que permita

verificar, de um lado, fatos de produção e, de outro, fatos de percepção sociolinguística por parte de sujeitos-falantes mais “prototipicamente” paulistanos.

Assim, na constituição da amostra proposta vão-se gerar análises qualitativas acerca do que se percebe como “ser paulistano”, que serão ulteriormente acompanhadas de análises da fala paulistana propriamente. Ambas permitirão, então, propor a construção de novas amostras, que poderão expandir esta que se projeta aqui, complementando-a.³ No atual estado da sociolinguística paulistana, contudo, parece fazer mais sentido construir uma amostra menos ambiciosa e claramente delimitada.

A seguir, após a definição dos objetivos deste projeto, apresentam-se as justificativas e o desenvolvimento dos argumentos acima delineados a favor de tal amostra. A metodologia para a sua construção é brevemente descrita no interior do projeto, mas é complementada com quatro anexos, em que se apresentam: (i) os procedimentos para a gravação das entrevistas sociolinguísticas e o roteiro de perguntas específicas a se fazer aos informantes selecionados (ver Anexo I); uma declaração de livre consentimento, a ser assinada pelos informantes (ver Anexo II); (iii) uma ficha do informante e da gravação, a ser preenchida pelos documentadores (ver Anexo III); e (iv) os critérios que serão empregados na transcrição das gravações (ver Anexo IV). Ao final, apresentam-se os recursos humanos e materiais que se preveem necessários para a execução deste projeto e uma proposta de cronograma para o seu desenvolvimento.

2. Objetivos

Os objetivos centrais deste projeto são:

- (i) construir uma amostra da fala paulistana contemporânea, a partir de três grupos de fatores extralinguísticos (região da cidade, sexo e faixa etária);
- (ii) disponibilizar a amostra construída (áudio e transcrições) em *site* próprio na internet, acessível a partir da página *web* do Departamento de Linguística (DL) da Universidade de São Paulo;
- (iii) formar pesquisadores, em nível de graduação, introduzindo-os no arcabouço teórico-metodológico da Sociolinguística Variacionista (Labov, 1972, 2006 [1996]).

São objetivos ulteriores a este projeto:

- (iv) permitir a descrição e a análise da variedade paulistana, ainda muitíssimo pouco estudada;
- (v) promover a elaboração de propostas de outras amostras da fala paulistana, que complementem aquela que se propõe aqui.

³ Por exemplo, seria interessantíssimo construir um *corpus* de dados da fala de gerações de descendentes de italianos, que permitisse verificar a correlação entre variantes paulistanas e variantes percebidas e avaliadas como “em princípio” italianas. Evidentemente, contudo, a construção de tal *corpus* requereria um outro projeto, que poderia se concentrar apenas nos descendentes de italianos ou, conforme seus objetivos, poderia incluir os descendentes de outros imigrantes.

Acerca do objetivo (iv) acima, alguns trabalhos de descrição e análise começarão a ser desenvolvidos concomitantemente à construção da amostra. De fato, alguns trabalhos já vêm sendo produzidos, com entrevistas coletadas nos últimos cinco anos pelo Grupo de Pesquisa e Estudos em Sociolinguística do DL-USP – cf. p. ex. Mendes (2009), Mendes (2010a, 2010b), Nascimento (2010), Oushiro (2010), Oushiro (2011a), Mendes & Oushiro (2011).⁴ Algumas dessas entrevistas que têm servido de *corpus* para esses trabalhos poderão vir a fazer parte daquela que se propõe organizar aqui, desde que sua qualidade seja compatível com o rigor com que se quer desenvolver este projeto.

Com a construção e a disponibilização desta amostra, espera-se, além da continuação e aprimoramento de trabalhos na linha de pesquisa sobre variação e mudança linguística do DL-USP, tornar possível também o desenvolvimento de trabalhos em outros centros de pesquisa, de modo que se contribua mais amplamente para a descrição sociolinguística da cidade de São Paulo.

3. Desenvolvimento e Justificativa

A proposta de construir e disponibilizar uma amostra robusta da fala paulistana, criteriosamente coletada e organizada, coaduna-se com o plano de fortalecer a linha de pesquisa sobre variação e mudança do Departamento de Linguística da USP. De fato, seria inadmissível que um dos centros de pesquisas linguísticas mais bem avaliados do país não oferecesse um retrato sociolinguístico da cidade em que se localiza. Dessa forma, embora já venham sendo desenvolvidos, conforme se indicou acima, trabalhos de descrição e análise da variedade paulistana nesse departamento, a construção e disponibilização de uma amostra como a que aqui se propõe é o primeiro passo de um programa mais extenso e intenso de pesquisas, a ser desenvolvidas a médio prazo, que contribuam para uma mudança do cenário sociolinguístico – tanto de uma perspectiva interna quanto internacional – ao adicionar São Paulo ao elenco de grandes centros urbanos brasileiros cuja variedade falada tenha sido amplamente dada a conhecer.

Amostrar a fala de uma enorme comunidade urbana como São Paulo impõe, evidentemente, alguns desafios. Para fazer jus à premissa básica dos estudos sociolinguísticos – a saber, o interesse das variantes linguísticas está em sua significação social (Chambers, 1995, 2004) – pode-se começar perguntando: o que significa ser paulistano, sociolinguisticamente? Ou, dito de outro modo: que variantes linguísticas identificam o paulistano? Uma vez que a relação entre variante linguística e

⁴ Vale citar também o trabalho de mestrado (em curso) de Rafael Rocha, sobre o emprego da dupla negação na comunidade paulistana e o projeto de pós-doutorado de Michael Colley (Rice University, Texas), sobre casos de ditongação variável na fala paulistana (a ser submetido à FAPESP).

significado social é estabelecida numa via de mão dupla, a última questão pode ainda ser reformulada do seguinte modo: com que variantes o paulistano opera a fim de se identificar como tal?

Parece haver duas variantes fonéticas com lugar cativo na elaboração de respostas para tais perguntas: a pronúncia tepe para o /-r/ em coda silábica (acompanhada de suposta infrequência da pronúncia retroflexa, mais associada ao interior paulista – Mendes 2009), e a ocorrência de pronúncia ditongada para /e/ nasal em palavras como “setenta”, “pensa”, “entende” – Mendes 2010b). A amostra da fala paulistana que aqui se projeta deverá permitir continuar a extração e a análise de dados como esses, bem como deverá permitir verificar se há outras variantes/variáveis linguísticas que convém estudar. As variáveis sociais que parecem ser mais preponderantes para a constituição sociolinguística da comunidade paulistana – a saber, Região da Cidade, Gênero/Sexo, Faixa Etária, Classe Social, Escolaridade e Geração da Família na cidade – são apresentadas e discutidas a seguir. Embora se reconheça a importância de todas essas variáveis sociais, os argumentos que se formulam a seguir esclarecem que, quantitativamente, a amostra será construída com base nas três primeiras; quanto às outras, elas não definirão o tamanho da amostra, mas farão parte dela qualitativamente – ou seja, também haverá variabilidade de acordo com essas variáveis.

3.1. A cidade e suas regiões

Nos anos 1960, quando Labov começava a publicar os resultados de sua empreitada variacionista, iniciada em Martha’s Vineyard e logo depois em Nova Iorque, ele se valeu do Atlas Linguístico da Nova Inglaterra⁵ e de sua segurança acerca do fato de que o bairro *Lower East Side* era uma localidade que de certa forma espelhava, localmente, a realidade social e linguística da ilha de Manhattan à época – de modo que a amostra da fala novaiorquina que então lhe interessava poderia ser coletada ali. Nas palavras do próprio autor:

Em *Martha’s Vineyard*, os seis mil falantes nativos e ali residentes são como que falantes de um único estilo: eles revelam relativamente pouca alternância no seu comportamento linguístico em face de mudanças na formalidade da situação de fala. Em Nova Iorque, a população a ser amostrada é mil vezes maior e tem muito mais divisões de classes e de castas. Nem as fronteiras interiores, nem as exteriores são fixas em Nova Iorque, como o são em Martha’s Vineyard: dentro dos limites desta última, a aguda distinção entre nativos residentes e novos moradores não dá lugar a equívocos. Em Nova Iorque, a mobilidade é parte do padrão, e os descendentes dos habitantes que ali se estabeleceram há mais tempo não são necessariamente os mais influentes na comunidade de fala de hoje. Grandes números de pessoas vivem na cidade mas ainda assim permanecem fora dos limites da comunidade de fala, e a linha que separa o falante nativo novaiorquino de outros é quebrada por muitos casos duvidosos. A área da cidade de Nova Iorque que foi escolhida para estudo intenso – o *Lower East Side* – não representa uma

⁵ Do inglês *Linguistic Atlas of New England – LANE* (Labov 1963)

simplificação desses problemas. Ao contrário, esta é uma área que exemplifica a complexidade da cidade de Nova Iorque como um todo, com toda a sua variabilidade e inconsistências aparentes.⁶

É claro que São Paulo guarda muitas semelhanças com Nova Iorque: é uma grande metrópole onde habitam milhões de pessoas, parte das quais foi nascida e criada “ali” (havendo quem sempre morou na mesma região da cidade e aqueles que já moraram em vários de seus bairros, em diferentes períodos de suas vidas), enquanto parte delas é representada por migrantes e imigrantes (específicos, em épocas específicas); muitos dos milhões que transitam pela cidade são seus residentes, enquanto outras tantas pessoas são visitantes temporários, ali presentes a trabalho, a estudos ou a lazer; classe e mobilidade são fatores importantes na caracterização da cidade – há pessoas que moram na Zona Sul e trabalham na região mais central, por exemplo; são muitos os bairros em que, ao se caminhar dois quarteirões, tem-se clara impressão de mudança de *status* econômico (visível na manutenção das ruas e calçadas, na qualidade dos imóveis, nos tipos de carros que se veem saindo e entrando nas garagens e estacionados nas ruas).

Há, contudo, diferenças importantes entre a São Paulo da década de 2010 e a Nova Iorque, tanto de hoje em dia quanto aquela contemporânea ao início das pesquisas labovianas. Talvez a mais importante, no que concerne às justificativas ao desenvolvimento deste projeto, seja o fato de que não parece haver, em São Paulo, um bairro que espelhe amplamente a complexidade da cidade como um todo, tal qual *Lower East Side* em Nova Iorque.

É verdade que devem ser vários os bairros cuja constituição é tal que, neles, perceba-se um sentimento de pertencimento à metrópole paulistana: trânsito caótico, muitas pessoas, homens e mulheres de diferentes classes, alguns “claramente” paulistanos nativos, outros “possivelmente” oriundos de outras regiões (na percepção desses mesmos sujeitos) – todos esses são fatos e características comumente associados a uma percepção de São Paulo como cidade cosmopolita. Entretanto, isso não significa que ocorra, num mesmo bairro, a congregação de toda a complexidade social e linguística da cidade – um bairro em que convivam nativos, migrantes e imigrantes observados em outros pontos dela, bem como indivíduos e grupos dos mais variados estratos socioeconômicos.

⁶ Livre-tradução do trecho aqui transcrito: “*On the vineyard, the six thousand native residents are close to single-style speakers: they show relatively little change in their linguistic behavior as the formality of the social context changes. In New York City, the population is a thousand times as large, with many more divisions of social class and caste. Neither the exterior nor the interior boundaries of the New York City community are fixed, as Martha’s Vineyard’s are: for within the limits of the island, the sharp distinction between the native residents and the newcomer permits little equivocation. In New York, mobility is a part of the pattern, and the descendants of the earliest long-term native settlers are not necessarily the most powerful influence in the speech community today. Large numbers of people live within the city yet remain outside the boundaries of the speech community, and the line which divides the native speaker from the foreigner is broken by many doubtful cases. The area of New York City that was chosen for intensive study – the Lower East Side – does not represent a simplification of these problems. On the contrary, it is an area which exemplifies the complexity of New York City as a whole with all its variability and apparent inconsistencies*”. (Labov 2006 [1966]:3-4)

Daí segue que para a cidade de São Paulo talvez fosse necessário organizar a coleta da amostra por bairros. O problema de tal proposta seria definir quais bairros exatamente seriam “contemplados” e com base em que critérios. Qualquer que fosse a decisão, contudo – considerando-se o tamanho da cidade, sua multiplicidade de bairros, as características específicas a eles associadas, por conta das quais vários deles seriam “mandatórios” na amostra –, ela teria que ser tão grande que não seria factível, dentro dos limites de tempo e de financiamento das políticas científicas atuais.

Por essas razões, buscou-se um critério que possibilitasse contemplar a cidade toda, mas sem a necessidade de incluir um grande número de bairros “obrigatórios”. Tal critério é o de uma divisão geográfica da cidade com base na história de ocupação dos bairros e no grau de desenvolvimento urbano em termos de verticalização e serviços (Oushiro, 2011b). Nesse sentido, a proposta é organizar a amostra de acordo com uma espécie de gradiente, que vai da zona mais central (o centro “histórico” e bairros geograficamente centrais, mais antigos e, de um modo ou de outro, percebidos como “tradicionalmente” paulistanos); passando por uma zona também “central”, mas num sentido expandido, que inclui bairros que hoje em dia figuram, tanto na percepção paulistana quanto externa, como bairros “típicos”; chegando, enfim, às áreas mais periféricas.

Para fins de estabelecimento de fatores num grupo, no sentido variacionista, os nomes propostos são, justamente: Centro Velho, Centro Expandido e Periferia. No primeiro, incluem-se bairros como Bela Vista, Consolação, Bom Retiro, Brás, Liberdade, Santa Cecília, Mooca. Dessa forma, o que se chama de Centro Velho representa a São Paulo “mais antiga”, cujos imóveis (bem como valores de aluguel) parecem tender a valorizar menos do que nos bairros do cinturão chamado aqui de Centro Expandido. Nos bairros mais centrais também se encontram mais facilmente informantes cujas famílias estão enraizadas na cidade há mais tempo (há duas ou mais gerações); esses falantes, portanto, podem apresentar uma tendência maior de empregar formas linguísticas consideradas prototipicamente paulistanas.

No Centro Expandido, incluem-se bairros como Pinheiros, Vila Madalena, Lapa, Higienópolis, Santana, Butantã, Vila Mariana, Itaim, Moema, Tatuapé, Vila Carrão. 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, nesses bairros parecem querer habitar os sujeitos de classe média que se percebem em ascensão, ainda que dentro da mesma classe; aí localizam-se os imóveis mais caros da cidade e que relativamente mais parecem se valorizar; aí está a grande maioria dos melhores restaurantes – fato amplamente reconhecido (local, nacional e internacionalmente) como um aspecto definidor da cidade de São Paulo.

Na região mais periférica, incluem-se bairros como Sapopemba, Pirituba, Jabaquara, Freguesia do Ó, Vila Matilde, Artur Alvim. Essa região se assemelha ao Centro Velho no que diz respeito ao fato de que em ambas parecem concentrar-se migrantes que vêm ou que tenham vindo “tentar a vida” na cidade. No entanto, bairros de periferia são historicamente mais recentes, bem como seu processo de verticalização; neles, há ainda uma clara divisão entre uma parte residencial, em que predominam casas em vez de prédios, e uma parte comercial, em geral uma ou duas avenidas em que se localizam bancos, drogarias, papelarias, lotéricas etc.; neles, ainda é possível ver crianças brincando na rua e vizinhos conversando no portão, fatos de sociabilidade que potencialmente têm consequências para o repertório sociolinguístico desses habitantes.

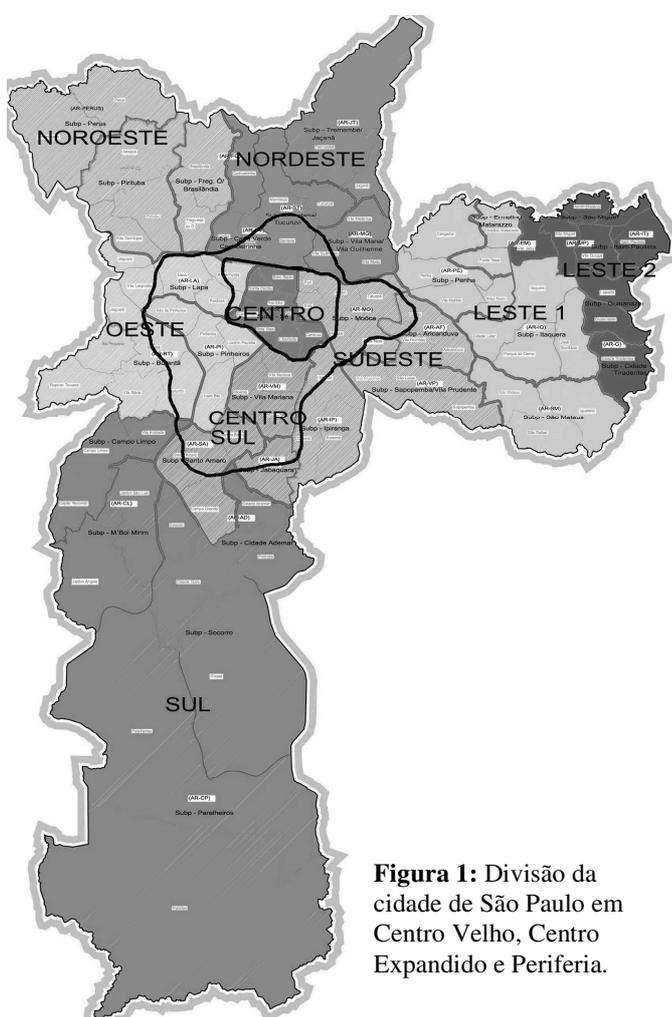


Figura 1: Divisão da cidade de São Paulo em Centro Velho, Centro Expandido e Periferia.

minimamente abrangente. A Figura 1 representa essa divisão da cidade em três regiões aproximadamente concêntricas em relação ao seu centro político e geográfico.

Esta categorização de bairros não é infalível – nem do ponto de vista do povoamento e ocupação histórica, nem do ponto de vista da percepção de distinção de classes. O trabalho de observação etnográfica (que antecede a confecção deste projeto e deverá anteceder a gravação das entrevistas sociolinguísticas propriamente) poderá sugerir ajustes e detalhamentos nos critérios que definem essa divisão. No entanto, é tamanha a complexidade e a heterogeneidade da capital paulista que, possivelmente, qualquer critério de segmentação, com vistas a estudos do tipo que define este projeto, se veria forçado a deixar certos aspectos de lado. A proposta de construir uma amostra da fala paulista que selecione seus sujeitos-informantes dos três conjuntos de bairros acima definidos se justifica pelo fato de que a tarefa se torna muito mais factível, sem deixar de ser

3.2. Gênero/sexo

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 incluísse 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 e, mais especificamente, na construção mesmo da amostra: 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 gênero/sexo 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” (Guy, 1990; Labov, 2001), as mulheres podem se revelar mais conservadoras; (iii) em mudanças “de cima”, as mulheres tendem a se revelar mais inovadoras.

Ora, num trabalho 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 uma atenção maiores 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 a cidade 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 observadas em inúmeros estudos e brevemente lembradas acima se observam também aqui. Além disso, o desenvolvimento deste deve também trazer subsídios para outras pesquisas que se interessem mais

especificamente pela variável gênero/sexo, no formato de um elenco de informações sobre as práticas sociais dos indivíduos, homens e mulheres, que forem selecionados para a construção desta amostra.

3.3. Faixa Etária

Assim como Gênero/Sexo, o grupo de fatores Faixa Etária está entre aqueles mais analisados em estudos sociolinguísticos e não apresenta maiores desafios em sua definição. 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, 2001). Além disso, assim como Gênero/Sexo, 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 permitirá investigar, nos trabalhos que vierem a ser desenvolvidos 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 que se propõe. Como já se adiantou na Introdução, esta amostra incluirá falantes de três faixas etárias: (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.⁷ 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 abarcar pessoas mais

⁷ Assim como para outras variantes sociais já mencionadas, os falantes mais jovens não considerados neste momento poderão ser alvo de projetos futuros com vistas à expansão desta amostra.

plenamente inseridas no mercado de trabalho e relativamente mais estáveis (ou que buscam mais seriamente uma certa estabilidade). Por fim, o grupo de falantes com mais de 60 anos se refere de modo amplo a pessoas da terceira faixa etária, aposentadas ou próximas da aposentadoria.

Essa caracterização, é claro, não se pretende 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, caso algum falante contatado não se encaixe na descrição acima, ele não será considerado na amostra; a proposta de 5 informantes por perfil tem justamente o propósito de 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, podendo, em princípio, ser reorganizado de outras maneiras. Na organização desta amostra, a ficha social de cada informante indicará 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, não-discreta.⁸

3.4 Classe Social, Escolaridade e Geração da família na cidade

Classe Social é uma variável que, pode-se dizer, é “mal” estudada na sociolinguística brasileira, na medida em que as amostras tendem a ser construídas com base em diferenças de escolaridade, mas não em diferenças de classe propriamente (Rodrigues, 2009). Se, por um lado, 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, 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. Ainda que a escolaridade do falante seja uma variável bastante fácil de ser acessada e que, portanto, poderia em princípio ser facilmente levada em consideração na montagem de uma amostra de fala, o aumento geral do grau de instrução da população paulistana, sobretudo dos mais jovens, permite supor que essa variável não mais seja um fator

⁸ O tratamento quantitativo dos dados linguísticos no programa GoldVarb X pressupõe a discretização das variáveis ou, em outras palavras, a análise de variáveis cujas variantes se dividem em categorias discretas e, em princípio, não ordinais (Guy & Zilles, 2007); atualmente, no entanto, o surgimento de outras ferramentas de análise linguística como o Rbrul (Johnson, s/d), que permite o processamento de dados de variáveis contínuas, abre novas possibilidades de análises quantitativas. Como esta amostra pretende servir ao desenvolvimento de uma série de pesquisas sociolinguísticas, a sua constituição e disponibilização deve permitir uma maior flexibilidade no tratamento de dados e nas análises quantitativas.

preponderante na diferenciação de grupos sociais; sugere-se aqui que Escolaridade deva ser tratada como um fator de composição da classe social do falante, mas não como sua substituta.

Por outro lado, 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 deveriam ser interdisciplinarmente estabelecidos e, para isso, seria necessário desenvolver um trabalho extenso de observação etnográfica, que operasse com percepções sociais e fatos socioeconômicos a elas correlacionados. Considerando-se o alegado aumento recente da classe média, conforme vem sendo observado pelas diferentes mídias, seria necessário mesmo um estudo (prévio e suficientemente amplo) que propusesse um desenho atual para as distinções de classe relevantes para o caso da metrópole. De novo, este se constituiria como um outro projeto, em si.

Ao mesmo tempo, incluir tal grupo de fatores entre aqueles que definem a construção da amostra aumentaria obrigatória e significativamente o seu tamanho. Imagine-se que se ficasse com as três áreas da cidade, os dois sexos e as três faixas etárias acima descritos: a inclusão de cinco estratos sociais (classe baixa, média-baixa, média-média, média-alta, alta) aumentaria a amostra mínima igualmente em cinco vezes. Assim, em vez de 18 informantes, seriam necessários no mínimo 90, para se ter um informante de cada combinação de fatores possível. Para se fazer uma amostra representativa, com cinco informantes de cada combinação possível, o total de informantes passaria para 450. Seria maravilhoso construir tal amostra, mas impossível em dois anos.

Neste projeto, Classe Social será uma variável atentamente observada, mas não definirá o tamanho da amostra. Uma vez constituída, ela conterà indivíduos de diferentes classes, e um determinado estudo poderá se valer de tais diferenciações, de acordo com seus objetivos. As diferenciações, que serão reportadas nas fichas de informante, se basearão em observações de natureza etnográfica e em perguntas formuladas no roteiro da entrevista sociolinguística que deverão se repetir com todos os informantes.⁹

Finalmente, propõe-se que esta amostra inclua apenas informantes paulistanos (ainda que seus pais possam não sê-lo), e que não se considerem migrantes e imigrantes. Embora esses habitantes certamente tenham um papel importante na constituição sociolinguística da cidade, sobretudo de uma perspectiva diacrônica, ainda se sabe muito pouco acerca do que constitua a “fala paulistana” – faz-se necessário, então, primeiramente amostrar a fala que seja aquela mais prototipicamente percebida como tal. Ao lado do argumento já invocado a respeito do aumento da amostra mínima, o que tornaria

⁹ Por exemplo, sobre o nível de escolaridade, ocupação e área de residência do informante, suas formas de lazer, lugares a que já viajou etc.

o projeto grande demais para um desenvolvimento a contento em dois anos, a inclusão de migrantes e imigrantes pressuporia responder certas questões que parecem colocar-se em outro momento da pesquisa sociolinguística na cidade, 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 caso dos imigrantes, incluiriam-se “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 coleta desta amostra, os documentadores deverão obter informações acerca de que geração de paulistanos o falante é representante – se os pais não são paulistanos, se um ou ambos os pais são paulistanos, se um ou mais avós são paulistanos – e da(s) origem(ns) da primeira geração da família que veio para cá (de que outra cidade, estado ou país). Tais informações farão parte da Ficha do Informante e da Gravação (ver Anexo III), de modo que, ainda que não constituam um dos parâmetros de estratificação desta amostra, poderão ser objeto de análises linguísticas já a partir deste material (para investigar, por exemplo, se a fala de paulistanos exhibe padrões quanto ao grau de enraizamento da família na cidade).

Este projeto se constitui, então, como um primeiro movimento na amostragem da fala paulistana, com objetivos variacionistas bem delimitados e definidos sobretudo com base na noção laboviana de comunidade de fala. Conforme vem-se apontando, em diversas partes deste projeto, seu desenvolvimento poderá levar ulteriormente a propostas de construção de amostras que se definem diferentemente da que se propõe aqui. Tais diferenças poderão ser definidas não somente por aspectos que por ora estão sendo deixados de lado, mas também por especificidades teóricas – por exemplo, é possível que a amostragem da fala de imigrantes e seus descendentes possa se beneficiar dos conceitos de “redes sociais” (Milroy, 2004) e “comunidades de práticas” (Meyerhoff, 2002). O trabalho de observação etnográfica que será desenvolvido ao longo do presente projeto fornecerá subsídios para o amadurecimento dessas questões.

4. Métodos e Materiais

4.1. Métodos

Do modo como se expôs e se justificou, a amostra da fala paulistana aqui projetada será construída de acordo com as categorias sociais (grupos de fatores extralinguísticos) resumidos no

quadro abaixo. Serão 18 perfis de informantes, para cada um dos quais serão gravados 5 indivíduos diferentes, totalizando 90 entrevistas sociolinguísticas.

Quadro 1 - Grupos de fatores definidores da amostra SP2010 e perfis sociolinguísticos dos informantes

Região da Cidade	Sexo	Faixa Etária	Perfil sociolinguístico
Centro Velho	Feminino	1	1. CF1
		2	2. CF2
		3	3. CF3
	Masculino	1	4. CM1
		2	5. CM2
		3	6. CM3
Centro Expandido	Feminino	1	7. EF1
		2	8. EF2
		3	9. EF3
	Masculino	1	10. EM1
		2	11. EM2
		3	12. EM3
Periferia	Feminino	1	13. PF1
		2	14. PF2
		3	15. PF3
	Masculino	1	16. PM1
		2	17. PM2
		3	18. PM3
18 perfis x 5 informantes = 90 gravações			

A gravação das entrevistas será precedida da seleção e do treinamento de documentadores, a serem escolhidos dentre alunos do curso de Letras da USP, que demonstrem interesse em desenvolver pesquisa em sociolinguística no futuro. O treinamento, que terá continuidade durante todos os semestres de execução deste projeto, compreenderá: (i) a leitura e a discussão de textos teóricos básicos da literatura sociolinguística (p.ex. Mollica & Braga, 2004; Labov, 2006 [1966]; Tagliamonte, 2006; Walker, 2010); (ii) a discussão de questões práticas do trabalho de campo; e (iii) palestras a serem proferidas por professores externos convidados que tenham experiência na formação de *corpora* de língua oral (ver Anexo V).

A amostra será semialeatória, dado que os informantes serão selecionados e contatados a partir do método “amigo de amigo” (Milroy 2004): os membros do GESOL-USP buscarão, a partir de suas redes sociais, a indicação de paulistanos que potencialmente se enquadram num dos 18 perfis acima elencados, observando o critério de que a relação entre documentador e informante não deve ser próxima.¹⁰ A partir desses novos contatos, os documentadores pedirão a indicação de novos

¹⁰ Embora um dos objetivos da entrevista sociolinguística seja o de obter amostras de fala que se aproximem do vernáculo e/ou de estilos de fala menos monitorados e embora se saiba que a conversa entre desconhecidos pode aumentar o Paradoxo do Observador (Labov, 2006 [1966]), esse critério se impõe a fim de que o conjunto final de entrevistas seja coerente: uma

informantes pelas pessoas já gravadas, a fim de ter acesso a um maior número de paulistanos, de perfis variados. Esse método se justifica, com base em experiências-piloto de coleta de entrevistas, pelo fato de que, na cidade de São Paulo, as pessoas dificilmente se dispõem a ceder parte de seu tempo a desconhecidos; a intermediação de um terceiro não só facilita o contato como estabelece um certo grau de confiança entre documentador e informante, o que também tem consequências para uma maior naturalidade da fala.

A partir da seleção dos informantes por esse método, as entrevistas sociolinguísticas procederão de acordo com as instruções e o roteiro descritos detalhadamente no Anexo I. O consentimento para participar das gravações deverá obtido de cada um dos informantes, que estarão cientes do fato de que a conversa será gravada, dos objetivos gerais da pesquisa e da garantia de anonimato na divulgação dos resultados (ver Anexo II). De modo geral, além de objetivar a obtenção de amostras de fala espontâneas de paulistanos de diferentes perfis sociolinguísticos, as entrevistas terão o objetivo de obter mais informações a respeito das condições de vida e das percepções/attitudes sociolinguísticas dos informantes, que poderão auxiliar na análise e na compreensão de fenômenos linguísticos na comunidade. Cada gravação terá duração aproximada de 60 a 70 minutos e será armazenada em formato “.wav” (stereo, 44.100 Hz). Após a coleta de cada entrevista, os documentadores deverão preencher a Ficha do Informante e da Gravação (ver Anexo III), que conterà dados sobre o informante (idade, ocupação etc.), sua família, bairros em que já morou, escolas em que estudou, além de um diário de campo para se registrarem as condições em que se deu a entrevista.

Todas as gravações passarão por validação por parte do coordenador deste projeto e por três alunos de pós-graduação, sendo avaliadas quanto à conformidade do informante com o perfil indicado, qualidade do áudio, naturalidade da interação conversacional e adequação ao roteiro de entrevista. Caso alguma gravação seja invalidada, o documentador deverá coletar nova entrevista com outro informante de mesmo perfil.

Após a validação, cada gravação será transcrita seguindo-se os critérios apresentados no Anexo IV. Este documento, passível de revisão conforme o desenvolvimento do projeto, foi elaborado com vistas a (i) armazenar o material coletado em meio escrito de modo padronizado; (ii) representar a língua falada de forma inteligível e fiel à oralidade; (iii) facilitar o desenvolvimento de análises linguísticas; e (iv) propiciar um tratamento dinâmico do material linguístico em arquivos (em formato .txt e codificação UTF-8) que possam ser facilmente manipulados em concordanceadores, programas

vez que se objetiva a obtenção de amostras de fala de perfis sociolinguísticos bastante variados, dificilmente elas poderiam ser coletadas somente através da gravação de pessoas cuja grau de relação é próximo; desse modo, é mais coerente padronizar as entrevistas a partir de um grau de relação mais distante.

de anotação como ELAN (Hellwig, Tacchetti & Somasundaram, 2011) ou programas como R (Hornik, 2011), que permitem a manipulação de grandes quantidades de dados textuais (por exemplo, para extrair dados automaticamente e realizar análises estatísticas). As revisões das transcrições serão realizadas pelo coordenador do projeto, por três alunos de pós-graduação ou por um dos alunos documentadores que não tenha feito a transcrição.

Antes de o material ser disponibilizado na internet, serão removidos, tanto dos arquivos de áudio quanto de transcrição, certos dados pessoais que por ventura apareçam durante a entrevista, a fim de preservar a identidade dos informantes e outras pessoas não públicas.

4.2. Recursos materiais e humanos

Para a execução deste projeto, prevê-se a necessidade de aquisição de gravadores digitais (modelo TASCAM DR-100), microfones (modelo Sennheiser HMD25-1), computadores (um desktop e três laptops) e equipamentos de armazenamento de dados (HD externa, CDs graváveis), para que os objetivos acima delineados possam ser cumpridos com o nível de qualidade que se impõe. Esses itens são listados, orçados e justificados em formulário próprio.

Prevê-se também a participação de seis bolsistas (Capacitação Técnica – TT-1), a serem selecionados entre alunos de graduação em Letras da USP, cujas atribuições básicas consistirão na coleta, transcrição e revisão de 90 entrevistas (aproximadamente 15 para cada) (ver Anexo V – Plano de Atividades). A criação da página do projeto na internet e a disponibilização adequada do material coletado e transcrito deverão ser realizadas por *webdesigner* contratado.

5. Cronograma

Plano de atividades	1º sem.	2º sem.	3º sem.	4º sem.
Aquisição de material permanente	X			
Seleção de documentadores	X			
Treinamento e formação de documentadores/sociolinguistas	X	X	X	X
Seleção de informantes	X	X		
Gravações-piloto (1/5 – 18 das entrevistas propostas)	X			
Validação das gravações	X			
Apresentação de trabalho no NWAV 40	X			
Gravações de mais 2/5 das entrevistas (36)		X		
Validação das gravações		X		
Transcrição e revisão das 18 primeiras entrevistas		X		
Apresentação de trabalho no GEL		X		
Contratação de técnico de informática		X		
Elaboração e envio do Relatório Parcial à FAPESP		X		
Gravações dos 2/5 restantes (36)			X	
Validação das gravações			X	
Transcrição e revisão das 36 entrevistas gravadas no 2º sem.			X	
Apresentação de trabalho no NWAV 41			X	
Transcrição e revisão das 36 entrevistas restantes				X
Edição dos arquivos de áudio ¹¹				X
Organização e disponibilização da amostra na internet				X
Apresentação de trabalho no congresso da ABRALIN				X
Elaboração e envio do Relatório Final à FAPESP				X

Referências

- Ash, Sharon (2004) Social class. In: Chambers, J.K., P. Trudgill & N. Schilling-Estes (eds.) **The Handbook of Language Variation and Change**. Oxford: Blackwell.
- Boersma, P. & David Weenink (2007) **Praat: doing phonetics by computer**. Disponível em <<http://www.fon.hum.uva.nl/praat/>>. Acesso em 02/02/2011.
- Castilho, Ataliba & Dino Preti (eds.) (1986) **A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. I – Elocuções Formais**. São Paulo: T.A. Queiroz.
- Castilho, Ataliba & Dino Preti (eds.) (1987) **A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. II – Diálogos entre dois informantes**. São Paulo: T.A. Queiroz/FAPESP
- Castilho, A. T. de (2007). Projeto para a História do Português Paulista. Projeto Temático de Equipe (Proc. FAPESP n. 06/55944-0)
- Chambers, Jack K. (1995) **Sociolinguistic Theory**. Oxford: Blackwell

¹¹ Apagamento de dados pessoais que possam identificar pessoas não públicas.

- Chambers, Jack K. (2004) Studying language variation: an informal epistemology. In: Chambers, J.K., P. Trudgill & N. Schilling-Estes (eds.) **The Handbook of Language Variation and Change**. Oxford: Blackwell.
- Chambers, Jack K.; Peter Trudgill & Natalie Schilling-Estes (eds.) (2004) **The Handbook of Language Variation and Change**. Oxford: Blackwell.
- Coelho, Rafael F. (2006) **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**. Dissertação de Mestrado. FFLCH-USP.
- Eckert, Penelope & McConnell-Ginet, S. (1992). “Think practically and look locally.” **Annual Review of Anthropology**, vol. 21 (21), p.461-90.
- Gonçalves, Sebastião Carlos Leite (2003). **O português falado na região de São José do Rio Preto: constituição de um banco de dados anotado para o seu estudo**. Projeto de pesquisa apresentado à FAPESP. Disponível em: <www.iboruna.ibilce.unesp.br/histórico/Projeto>.
- Guy, Gregory R. (1990). “The sociolinguistic types of language change”. *Diachronica*. Vol. 7:1, pp.47-67.
- Guy, Gregory R. & Zilles, Ana (2007). **Sociolinguística Quantitativa: instrumental de análise**. São Paulo: Parábolsa.
- Hellwig, Birgit, M. Tacchetti & A. Somasundaram (2011) ELAN – Linguistic Annotator. Versão 4.1.0. Disponível em <http://www.lat-mpi.eu/tools/elan/manual/>. Acesso em 03/05/2011.
- Holmes, Janet & Miriam Meyerhoff (eds.) (2005) **The Handbook of Language and Gender**. Oxford: Blackwell.
- Hora, D. da (org.) 2004. **Estudos Sociolinguísticos: perfil de uma comunidade**. Santa Maria: Palotti.
- Hornik, Kurt. (2011) R FAQ. Disponível em: <<http://cran.r-project.org/doc/FAQ/R-FAQ.html>>. Acesso em 10/04/2011.
- Johnson, Daniel Ezra (s/d). RBrul Manual. Disponível em <www.ling.upenn.edu/~johnson4/Rbrul_manual.html>. Acesso em 23/01/2011.
- Labov, William (1963) “The social motivation of a sound change.” *Word*, vol. 19, 273–309.
- Labov, William (1972). **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press.
- Labov, William (2001). **Principles of Linguistic Change. Social Factors**. Oxford: Blackwell.
- Labov, William (2006 [1966]). **The Social Stratification of English in New York City**. New York: Cambridge University Press. 2nd Edition.
- Mendes, Ronald Beline (2009). “Who sounds /r/-ful? The pronunciation of coda /-r/ in the city of São Paulo”. Trabalho apresentado no New Ways of Analyzing Variation - NWAV38. University of Ottawa.
- Mendes, Ronald Beline (2010a). “What /-r/ you saying? Retroflex /-r/ in the city of São Paulo”. Trabalho apresentado no Congresso Internacional de Línguas Pluricêntricas. Universidade Católica de Portugal – Braga.
- Mendes, Ronald Beline (2010b). “Sounding Paulistano: variation and correlation in São Paulo”. Trabalho apresentado no New Ways of Analyzing Variation - NWAV39. University of Texas at San Antonio.

- Mendes, Ronald Beline & Livia Oushiro (2011) "Production and perception of retroflex /r/ in São Paulo Portuguese". Trabalho apresentado no Variation and Language Processing – VaLP 2011. University of Chester.
- Meyerhoff, Miriam (2002). Communities of Practice. In: Chambers, Jack K.; Peter Trudgill & Natalie Schilling-Estes (eds.) (2004) **The Handbook of Language Variation and Change**. Oxford: Blackwell.
- Milroy, Leslie (2004) Social networks. In: Chambers, J.K., P. Trudgill & N. Schilling-Estes (eds.) **The Handbook of Language Variation and Change**. Oxford: Blackwell.
- Mollica, Maria Cecilia de M. & Braga, Maria Luiza (eds.) (2003). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto.
- Nascimento, Ivanete Belém do (2010) **O uso variável do pronome de segunda pessoa você(s)/cê(s) na cidade de São Paulo**. Dissertação de Mestrado. FFLCH-USP.
- Oushiro, Livia (2010) "Wh-interrogatives in Brazilian Portuguese: the influence of common ground". Trabalho apresentado no New Ways of Analyzing Variation - NWAV 39. University of Texas at San Antonio.
- Oushiro, Livia (2011a) **Uma análise variacionista para as Interrogativas-Q**. Dissertação de mestrado. FFLCH-USP.
- Oushiro, Livia (2011b) "Identidade na pluralidade: produção e percepção linguística na cidade de São Paulo". Projeto de doutorado.
- Preti, D. & H. Urbano (eds.) (1988) **A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. III – Entrevistas**. São Paulo: T.A. Queiroz/FAPESP.
- Preti, D. & H. Urbano (eds.) (1990) **A linguagem falada culta na cidade de São Paulo: materiais para seu estudo, vol. IV – Estudos**. São Paulo: T.A. Queiroz/FAPESP
- Rodrigues, Angela C. S. (1987) **A concordância verbal no português popular em São Paulo**. Tese de doutorado. FFLCH-USP.
- Rodrigues, Angela C. S. (2009) **Fotografia sociolinguística do português do Brasil: o português popular em São Paulo**. In: Castilho, Ataliba T. (ed.), **História do Português Paulista**. Campinas: Instituto de Estudos da Linguagem/UNICAMP.
- Romaine, Suzanne (2003) Variation in Language and Gender. In: Holmes, Janet & Meyerhoff, Miriam (eds.) **The Handbook of Language and Gender**. Oxford: Blackwell.
- Tagliamonte, Sali A. (2006). **Analysing Sociolinguistic Variation**. São Paulo: Cambridge University Press.
- Walker, James (2010). **Variation in Linguistic Systems**. New York: Routledge.

ANEXO I:

INSTRUÇÕES GERAIS E ROTEIRO DA ENTREVISTA

Construção da amostra SP2010: instruções gerais e roteiro da entrevista

1. Instruções gerais

A coleta de entrevistas sociolinguísticas tem basicamente dois objetivos: (i) obter amostras de fala semiespontânea de paulistanos; e (ii) obter informações sobre o modo/condições de vida e percepções sociolinguísticas dessas pessoas. Esta amostra selecionará falantes paulistanos de maneira semialeatória: eles serão contatados a partir de redes sociais (“amigos de amigos”). A principal justificativa para isso reside no fato de que, de acordo com experiências-piloto desenvolvidas antes da proposta deste projeto, na cidade de São Paulo as pessoas dificilmente se dispõem a ceder parte de seu tempo para “pesquisas feitas por desconhecidos”.

Para proceder à gravação das entrevistas, os documentadores deverão seguir as seguintes instruções:

Antes da entrevista

- Contatar o informante.
 - Não entrevistar ninguém próximo do convívio diário (parentes, amigos, colegas de trabalho). Escolher informantes que, na Ficha do Informante, são definidos pelos graus 4 e 5.¹
 - Entrevistar “amigos de amigo”: procurar, através de seus contatos, outras pessoas que eles conheçam e que sejam informantes potenciais para esta amostra.
 - Entrevistar apenas pessoas que nasceram e foram criadas na cidade de São Paulo.
 - Entrevistar pessoas que conhece “de vista”: o “cara” do mercadinho, do açougue, da padaria... posteriormente pode-se pedir a esses mesmos informantes a indicação de outras pessoas para serem entrevistadas. De acordo com experiências prévias, os informantes, após terem participado de uma entrevista, em geral não apresentam resistências a indicar novos informantes.
- Conversa preliminar
 - Apresentar-se como aluno universitário. Dizer ao informante que está fazendo uma pesquisa sobre vários bairros de São Paulo para saber um pouco sobre como é a vida em cada região da cidade e que o interesse da pesquisa é no dia-a-dia dos paulistanos – ou seja, não é nada teórico, nada elaborado. O objetivo é saber como é a vida no bairro e os pontos positivos e negativos da região em que o informante vive e da cidade de São Paulo como um todo.

¹ Ver Anexo III (Ficha do Informante e da Gravação), p.2, Grau de relação entre informante e documentador.

- Essa conversa prévia também leva à obtenção de informações gerais (p.ex., se, de fato, o informante corresponde a um dos perfis estabelecidos pelo projeto) e ajuda a “quebrar” um pouco o grau de formalidade.
- Dizer que a entrevista dura entre uma hora e uma hora e meia. Perguntar ao possível informante quando é o melhor dia e horário para ele. É melhor não gravar a pessoa no mesmo dia se ela se mostrar pouco disponível. Fazer a entrevista preferencialmente na casa do informante.
- Pouco antes da gravação
 - Antes de iniciar a entrevista, escolher um local adequado para desenvolvê-la: evitar locais barulhentos e/ou abertos, sugerir que a TV seja desligada (caso esteja ligada), evitar sentar perto de janelas ou outros locais em que pode haver barulho (p.ex., perto de um ventilador). De preferência, escolher um local acarpetado – a sala normalmente é o melhor lugar para se fazer uma entrevista.
 - Explicar ao informante os objetivos gerais da pesquisa e obter o seu consentimento para ser gravado.
 - Pedir que o informante coloque o microfone em si e, caso seja necessário ajustá-lo, instruir o informante para que ele mesmo o faça ou pedir licença para ajustá-lo. Apesar de o microfone potencialmente aumentar o grau de formalidade da situação, ele é importante para garantir uma boa qualidade de gravação, o que posteriormente facilitará a tarefa de transcrição e possibilitará o desenvolvimento de trabalhos a respeito de variáveis fonético-fonológicas.
 - Se sentir que o seu informante está um pouco tenso ou desconfiado, conversar um pouco mais com ele antes de começar a gravação. Responder todas as dúvidas que ele possa ter, deixar claro que sua identidade não será divulgada nos resultados da pesquisa e esclarecer que a conversa pode ser interrompida a qualquer momento.

Durante a entrevista

- A regra mais básica é falar POUCO! O objetivo é proporcionar as condições para que o informante fale, de forma mais natural possível, dentro da situação de entrevista sociolinguística. Para tanto:
 - Escutar o que o seu informante tem a dizer. Mostrar um interesse genuíno pelo que ele está dizendo. Se ele mencionar alguma coisa interessante/pitoresca, pedir que ele fale mais: “Nossa! E como foi isso?”... A sua postura, durante toda a entrevista, deve ser a de alguém que está lá para aprender.
 - Fazer perguntas curtas. Ao mesmo tempo, evitar fazer perguntas cuja resposta é simplesmente “sim” ou “não”; se for fazê-las, ter outra pergunta na

ponta da língua para que o informante continue falando – p.ex. “Você gosta de morar aqui? (deixar o informante responder) Por que você escolheu morar nesse bairro?”

- Evitar interromper o informante enquanto ele está falando e evitar muitas sobreposições de vozes.
- Os tópicos abaixo são apenas para a sua organização, mas a conversa deve fluir. Não falar coisas como: “Agora vamos falar sobre a sua família”. As perguntas da primeira parte do roteiro são sugestões para manter o fluxo da conversa e, pela experiência das gravações-piloto, fornecem uma sequência “natural” de tópicos a serem tratados. Ser flexível para explorar temas relacionados que se revelem de interesse para o informante. Por outro lado, as perguntas da segunda parte do roteiro são obrigatórias e devem ser feitas, na medida do possível, na mesma sequência em que são apresentadas.
- As gravações devem durar entre 60 e 70 minutos. Administrar o tempo de forma que dê para cobrir todos os tópicos.
- Se for à casa do informante, posteriormente fazer anotações que julgar interessantes e que possam levar a inferências a respeito da classe social, infraestrutura do bairro (se há uma associação de bairro, locais de lazer) etc. Isso também pode facilitar o contato com novos informantes.

Depois da entrevista

- Pedir ao informante que assine a declaração.² Explicar que ela é uma declaração de que ele sabia que estava sendo gravado e que serve para proteger os seus dados pessoais. Deixar claro que nenhum dado pessoal será publicado.
- Pegar informações adicionais que podem não ter surgido durante a entrevista. É melhor coletar os dados da Ficha do Informante depois da gravação – o informante se sentirá mais seguro de que seus dados pessoais não foram gravados.
- Em casa, preencher a Ficha do Informante o mais cedo possível. Vai ser mais fácil encontrar/organizar as informações. Depois provavelmente será mais difícil lembrá-las!

² Ver Anexo II (Declaração).

2. Roteiro da entrevista

I. Primeira parte

BAIRRO (aprox. 10 min.)

Objetivos: descobrir o “grau de enraizamento” do informante no bairro onde vive/outros bairros; descobrir o “grau de mobilidade” da pessoa

1. Há quanto tempo você mora na (Mooca, Pinheiros, Bexiga...)?
2. Você gosta de morar aqui?
3. Por que você escolheu morar neste bairro? (manter em mente que o informante pode não ter “escolhido” morar ali: pode ter sido porque a família já morava lá etc.)
 - a. Se o informante mora há bastante tempo: Como era o bairro antes/quando você se mudou pra cá? Mudou muito?
 - b. Se o informante não mora lá há muito tempo, ou se nem sempre viveu ali: Em que outros lugares você já morou? Como era lá comparado com esse bairro aqui? Onde você preferia morar?
4. Você tem bastante contato com as pessoas aqui do bairro?
5. Hoje, tem algum outro bairro em que você gostaria de morar?

INFANCIA (aprox. 5 min.)

Objetivos: relaxar o informante (em geral, as pessoas gostam de falar de sua infância); obter informações sobre mudanças no bairro/cidade de São Paulo; grau de mobilidade do informante; obter informações sobre escolaridade

6. E como foi a sua infância (no bairro X)? Você pode contar um pouco de como foi, o que você fazia...?
 - a. brincava na rua/dentro de casa? Do que vocês brincavam?
 - b. como eram os seus pais? eram rígidos...?
7. Você foi pra escola no mesmo bairro? Como era a escola? Você sempre estudou na mesma escola? Você gostava de ir pra escola? Tem algum professor que te marcou? Até que série você estudou?
8. Enquanto ainda era criança/adolescente, você ia pra outros lugares dentro da cidade de São Paulo? (pra onde, pra fazer o quê...)

FAMILIA (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante, grau de enraizamento no bairro/cidade

9. Você tem irmãos? (é possível que esta informação já tenha aparecido na parte sobre infância; neste caso, falar algo como: “Você disse que tem uma irmã... Você tem outros irmãos?”) Quantos anos eles têm?

10. Seus pais são daqui de São Paulo mesmo? (Se sim, perguntar também sobre avós, bisavós... até encontrar a primeira geração da família que veio pra cá). Quantos anos eles (pais) têm? Quando seus (pais/avós/bisavós) vieram pra São Paulo? Você sabe por que eles vieram?

11. Você tem filhos? Quantos anos eles têm?

12. E o resto da família, tios, primos, também vivem aqui em São Paulo? (Se sim, em que bairros? Moram próximo? Se não, onde? Tem contato sempre?)

TRABALHO/OCUPAÇÃO (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante; características socioeconômicas

13. Você trabalha aqui por perto? (se não souber onde a pessoa trabalha)

14. O que você faz?

15. Você gosta do seu serviço? (Se não, o que a pessoa preferiria fazer?)

LAZER (aprox. 5 min.)

Objetivos: obter informações sobre rede social do informante; mobilidade na cidade; características socioeconômicas

16. E nas horas de lazer, o que você e sua família gostam de fazer? (Se saem, vão pra que lugares?) Você acha que a cidade de São Paulo tem boas opções de lazer? Quais?

17. A maioria dos seus amigos mora aqui nesse bairro mesmo? (Se não, onde?)

18. Você tem algum atividade de recreação em grupo, algum clube...?

19. Quais são seus amigos mais antigos? Você mantém contato com os amigos de colégio?

20. Você costuma viajar? Pra que lugares já viajou? Que lugares gostaria de conhecer?

II. Segunda parte

A CIDADE DE SÃO PAULO (aprox. 20 min.)

21. Uma pesquisa recente mostrou que 57% das pessoas em São Paulo deixariam a cidade caso pudessem. O que você acha disso? (se o informante perguntar, a pesquisa é do Ibope e foi publicada em janeiro/2010³)

³ Disponível em <http://www.nossasaopaulo.org.br/portal/arquivos/Pesquisa_IRBEM_Ibope_2010_

22. Você gosta de morar em São Paulo? (Se não, em que lugar preferiria morar?)
Por quê?
23. O que você acha que caracteriza a cidade (tanto as coisas boas quanto ruins)?
24. O que você acha que caracteriza o paulistano (tanto as coisas boas quanto ruins)?
25. Olhando pra mim, você diria que eu sou paulistano(a)? Por quê?
26. O que você mais gosta em São Paulo?
27. O que você não gosta em São Paulo? (a depender do tópico mencionado pelo informante - violência, trânsito, poluição etc. - procurar explorar mais o assunto. P.ex.: você já foi assaltado? O que aconteceu? (para obter narrativa pessoal) O que o governo poderia fazer pra solucionar esse problema? (para obter uma fala mais distanciada)) (explorar o subtópico por aproximadamente 10 min.)
28. Pras pessoas que não vivem em São Paulo, como você acha que elas imaginam que seja a cidade? Qual é a imagem que as pessoas de fora de São Paulo têm da cidade?
29. Você falou que já foi pra (X, Y, Z). Quando você foi pra esses lugares, as pessoas percebiam que você era paulistano? (Se sim) como elas percebiam?
30. Quando você conhece alguém, você percebe se a pessoa não é daqui de São Paulo?
- a. (Se sim) como você percebe? (Se o informante mencionar o modo de falar, seguir nessa linha e tentar conseguir informações mais precisas. É normal as pessoas não saberem definir o porquê, mas devemos tentar tirar mais informações).
- b. (Se não) Quando você ouve uma pessoa falando, por exemplo, você percebe que ela é de fora pelo sotaque?

PRODUÇÃO/PERCEÇÃO/AVALIAÇÃO LINGUÍSTICA (aprox. 10 min.)

31. Já que a gente está falando de sotaque... qual o sotaque do Brasil que você mais gosta? E tem algum que te irrita? Como é que (o gaúcho/o carioca/o mineiro/o caipira etc. – a depender dos sotaques mencionados) fala(m)?
32. Aqui em São Paulo tem muito (i)migrante, né? De onde vêm a maioria das pessoas? Tem algum bairro específico em que eles (italianos, coreanos, nordestinos etc.) se concentram?
33. E como é que as pessoas falam na cidade de São Paulo? (evitar usar a palavra “sotaque”)

34. E aqui dentro da cidade de São Paulo, você vê diferenças entre as diferentes regiões, diferentes bairros? (Se sim) Você poderia dar alguns exemplos? (É normal as pessoas responderem que não, mas se responderem que sim, tente obter informações mais precisas)

(Imprimir a lista de palavras, a notícia e o trecho, todos abaixo, para mostrar ao informante)

LISTA de PALAVRAS:

Apresentar ao informante a lista de palavras e dizer: “Agora eu queria te pedir pra ler algumas coisas. Eu tenho aqui uma lista de palavras, e eu queria que você lesse cada uma delas. Você pode dar uma olhada primeiro, antes de começar a ler.” Esperar que o informante leia a lista de palavras. Depois perguntar:

35. A gente estava falando de sotaques agora há pouco... Como você acha que uma pessoa do interior falaria algumas dessas palavras? Tem diferenças?
36. E como um carioca falaria algumas dessas palavras?
37. Tem mais algum sotaque no Brasil que você conhece?

LEITURA DE NOTÍCIA:

Dizer ao informante: “Agora eu queria que você lesse esse texto. Você pode ler em voz baixa antes de falar em voz alta.” Esperar que o informante leia a notícia.

LEITURA DE TRECHO:

Dizer ao informante: “Por último, queria que você lesse esse trechinho.” Esperar que o informante leia o trecho.

38. O que você acha desse modo de falar: “Você tá entendendo o que eu tô dizendo?” (com “en” ditongado e exagerado). Se a pessoa manifestar uma atitude negativa, perguntar: Como você acha que deveria ser? O que tem de errado aí?
39. Quem você acha que fala assim? (se falar “paulistanos”, perguntar: “você acha que todos os paulistanos falam assim ou é uma coisa de uma região mais específica na cidade?”)
40. Você fala desse modo?
41. E o que você acha de “Me vê dois pastel e um chopps?” (Fazer as mesmas perguntas acima)
42. E o que você acha de “A porta tá aberta.” (com retroflexo exagerado). Fazer as mesmas perguntas acima.
43. E tem mais algum modo de falar que você acha assim bem paulistano?

Palavras

alma
amargo
animação
argola
atitude
barqueiro
biscoito
cacique
carteiro
cerca
chácara
circo
cisne
curto
defender
discoteca
elefante
enchente
entender
entretenimento
erguer
fazenda
felicidade
firme
fornalha
furgão
fusquinha
geleira
gérmen de trigo
gordo

gula
Hércules
hilário
irmã
justiça
lento
mortadela Marba
mosca
necessidade
noite
orca
Ordem e progresso
orgânico
órgão
ostracismo
penteado
pertencimento
perto
porto
presente
rapidez
riqueza
sabor de menta
soberba
trabalho
turco
urgente
utilidade
vulto
zebra

Notícia

08/02/2010

Moradores cobram solução para área alagada em SP e aguardam encontro com prefeito

Moradores da região do Jardim Pantanal protestaram nesta segunda-feira em frente à Prefeitura de São Paulo contra a inundação em bairros da zona leste, que completa dois meses hoje. Uma comissão de manifestantes foi recebida pela Secretaria de Relações Institucionais e apresentou reivindicações, mas o prefeito Gilberto Kassab (DEM) não participou do encontro.

Representantes dos manifestantes esperam que o governador José Serra (PSDB) também participe da reunião de sexta. De acordo com os integrantes da comissão, não havia integrantes do governo estadual no encontro realizado hoje na prefeitura.

Além de cobrar a limpeza imediata das águas nos bairros, os moradores cobram uma solução para a falta de moradia. A maioria dos desalojados continua em escolas municipais e, com o início do ano letivo, temem não ter para onde ir.

Trecho

Tá chovendo muito! Choveu tanto, tanto na semana passada que ficou uma piscina na minha casa. Ó, pra você ver: Molharam todos os armários, a cama, os colchões, tudo... Foi um sacrifício... O que a gente fez? Nós tivemos que erguer os móveis pra limpar tudo: a geladeira, o forno... minha irmã até veio me ajudar, sabe? E meus filhos compraram umas cadeiras novas, mas é aquela coisa, assim... quando chover de novo, vai molhar tudo outra vez. Você fica sem ter o que fazer. E tem um rio lá perto que sempre alaga... quer dizer, é água dentro e fora de casa! Daí, o que acontece? Fica aquele trânsito, os carros todos parados, a gente demora um tempão pra chegar em casa... Não aguento mais enchente nessa cidade... Agora que eu vou fazer? Os políticos falam, falam, mas eles tinham que fazer alguma coisa urgente. Você tá entendendo o que eu tô dizendo?

ANEXO II:

DECLARAÇÃO DE LIVRE CONSENTIMENTO



UNIVERSIDADE DE SÃO PAULO

FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS

Av. Prof. Luciano Gualberto, 403 - Cx. Postal: 26097 - CEP: 01060-970 - São Paulo

Fone: (011) 3091.4298 - e-mail: fl@usp.br

Projeto de Pesquisa: SP2010 – Construção de uma amostra da fala paulistana

Docente responsável: Prof. Dr. Ronald Beline Mendes

Contato: rbeline@usp.br

Declaração

Declaro ter consentido em ter gravada minha conversa com o aluno universitário _____, como parte do projeto acima identificado, em execução pelo Grupo de Estudos e Pesquisa em Sociolinguística da USP. Estou ciente de que a gravação será transcrita e disponibilizada na Internet para o desenvolvimento de estudos sociolinguísticos na cidade de São Paulo. Entendo que essa gravação é idônea e que meus dados pessoais não serão tornados públicos na divulgação de resultados da pesquisa.

São Paulo, _____ de _____ de _____.

Nome do informante: _____

Assinatura: _____

ANEXO III:

FICHA DO INFORMANTE E DA GRAVAÇÃO

Ficha do Informante e da Gravação

Nome do Informante:

Pseudônimo:

Data de Nascimento:

Idade:

Endereço:

Telefones para contato:

E-mail:

Ocupação:

Pessoas com quem mora:

Documentador:

Arquivo de gravação e transcrição:

Bairros em que já morou na cidade de São Paulo, em ordem cronológica, e tempo aproximado (em anos) de quanto tempo morou em cada bairro:

P.ex.: Tatuapé – 10 anos

Número de irmãos:

() mais velhos

() mais novos

Nome, idade e escolaridade dos irmãos:

Geração na cidade: () 0 – pais não nascidos na cidade de São Paulo

() 1 – pai ou mãe nascido na cidade de São Paulo

() 2 – pai e mãe nascidos na cidade de São Paulo

() 3 – um avô ou uma avó nascido na cidade de São Paulo

() 4 – dois ou mais avós nascidos na cidade de São Paulo

Nome do Pai:

Idade do Pai:

Grau de Escolaridade do Pai:

Ocupação do pai:

Naturalidade: () São Paulo (capital)

() outra: _____

Nome da Mãe:

Idade da Mãe:

Grau de Escolaridade da Mãe:

Ocupação da mãe:

Naturalidade de: () São Paulo (capital)

() outra: _____

Escolas em que estudou e tipo de escola (pública municipal (M), pública estadual (E) ou particular (P)):

P.ex.: Fundamental 1: Escola Estadual President Roosevelt (E)

Fundamental 1: _____

Fundamental 2: _____

Ensino Médio: _____

Faculdade: _____

Data da Entrevista:

Horário:

Local da Entrevista e breve descrição do local:

Presença e atuação de terceiros:

Como o documentador conheceu o informante:

Grau de relação entre informante e documentador:

() 1 – Bastante próximo. O entrevistado faz parte do meu círculo imediato de amigos/familiares e conversamos frequentemente.

() 2 – Próximo. Conversamos frequentemente, mas o entrevistado não faz parte do meu círculo imediato de amigos/familiares.

() 3 – Próximo, mas não conversamos frequentemente.

() 4 – Neutro. Ele é meu conhecido, mas não nos falamos com frequência.

() 5 – Distante. Não o conhecia anteriormente e praticamente só conversamos na ocasião da entrevista.

Outras informações:

ANEXO IV:

CRITÉRIOS PARA A TRANSCRIÇÃO DE ENTREVISTAS DA AMOSTRA SP-2010

Critérios para a Transcrição das Entrevistas da Amostra SP-2010

1. Introdução

O objetivo principal do sistema de transcrição aqui descrito é transpor a língua falada para o texto escrito de uma forma fiel à língua oral, mas inteligível, de modo a armazenar o material coletado em meio escrito de maneira padronizada para facilitar a sua análise. É necessário lembrar, antes de tudo, que nenhum sistema de transcrição é capaz de reproduzir a fala tal e qual, e que qualquer sistema proposto sofrerá de limitações; com esse fato em mente, os critérios devem sempre levar em conta os objetivos do grupo de pesquisa que, no presente caso, definem-se pelo desenvolvimento de estudos sociolinguísticos a respeito de fenômenos linguísticos em variação entre falantes da cidade de São Paulo.

Por um lado, prevê-se que quanto maior o número de critérios, mais heterogêneas serão as transcrições das gravações do mesmo *corpus*, algo que não é desejável quando um dos principais objetivos é padronizar o material coletado. Essa observação, aliada à proposta deste projeto de disponibilizar não apenas as transcrições, mas também as gravações coletadas – de modo que os pesquisadores que venham a utilizar esse material poderão recorrer ao material sonoro –, justifica a opção por um sistema mais simples de transcrição em relação a outros sistemas já propostos anteriormente (Cf. p.ex. Castilho & Preti, 1986; Tenani & Gonçalves, s/d; Mello & Raso, 2009).

Ao mesmo tempo, os critérios abaixo estipulados têm a preocupação de produzir arquivos de texto que serão compatíveis com a posterior manipulação através de concordanceadores, programas de anotação como o ELAN (Hellwig, Tacchetti & Somasundaram, 2011) ou programas como o R (Hornik, 2011), que permitem o processamento de grandes quantidades de dados textuais, a extração automática de dados e a realização de análises estatísticas. Assim, evitam-se, por exemplo, caracteres que formarão novas palavras, visto que isso pode alterar a sua contagem e “inflacionar” o tamanho do *corpus*, e caracteres que podem prejudicar a busca e a extração automática de palavras ou trechos de fala (por exemplo, não se usará “:” para indicar alongamento de vogal, tampouco “-” entre sílabas para indicar silabação).¹

Ressalta-se ainda que estes critérios serão objeto de discussão entre os membros do projeto (quanto à sua objetividade, exequibilidade e pertinência) e que poderão ser revisados ou complementados à medida que novas questões surjam ao longo do desenvolvimento desta tarefa.

2. Formato e nome do arquivo de transcrição

O arquivo de transcrição deve ser salvo em formato .txt e com codificação UTF-8. Desse modo, a transcrição não deve conter nenhuma formatação “especial”, como **negrito**, *itálico*, TAB, colunas, tabelas, Shift + Enter.

Cada arquivo de transcrição deve corresponder a uma gravação de entrevista sociolinguística e deve ser salvo com o mesmo nome que identifica a gravação, de acordo com a seguinte convenção:

¹ Por exemplo, na realização de buscas automáticas por sílabas com /r/ em coda silábica (em palavras como “porta”, “caderno”) no programa R, a inserção de :: para indicar alongamento de vogal pode prejudicar a localização de ocorrências como “po::rta”.

SP(ano de gravação)-(número da gravação)-(sexo)(idade)(escolaridade)(região da cidade)(zona da cidade)

sexo:	M (masculino) / F (feminino)
idade:	em número
escolaridade:	G (até E. Fundamental II); C (até E.Médio); S (E. Superior completo ou não)
região de SP:	V (centro velho); E (centro expandido); P (periferia)
zona de SP:	L (leste); O (oeste); N (norte); S (sul); C (centro)

Ex.: SP2011-001-F25CEO (i.e., gravação 001, referente a informante do sexo feminino, com 25 anos, com escolaridade até Ensino Médio, que vive na região do centro expandido na zona oeste)

3. Cabeçalho

O arquivo deve se iniciar pelo cabeçalho, que se delimita pelo símbolo #:

#

S1: (perfil e pseudônimo do informante)

S2: (S3, S4... perfil e pseudônimo de eventual terceiro falante durante a gravação)

D1: (D2... nome do documentador)

Transcritor: (nome do transcritor)

Revisor: (nome do revisor)

Duração total: (no formato 00h00min00seg)

Número de palavras: (excetuando-se cabeçalho, “D1:”, “S1:” e comentários do transcritor entre colchetes [])

#

4. Corpo da transcrição

Orientações gerais: a transcrição deverá ser semiortográfica, utilizando os seguintes critérios:

- Não se trata de uma transcrição fonética. Portanto, NÃO altere a ortografia das palavras. Mesmo que, por exemplo, o falante pronuncie “minino”, “muleque”, “vrido”, deve-se transcrever “menino”, “moleque” e “vidro”, respectivamente, EXCETO em casos de comentários metalinguísticos. P.ex.:

S1: eles falam tudo errado lá sabe... eles falam... ‘vrido’
‘pobrema’...

- Por outro lado, NÃO devem ser feitas “correções” gramaticais. Por exemplo, se o informante disser “eles gosta de música”, sem concordância verbal, deve-se transcrever dessa forma, sem concordância; se o informante disser “o que que aconteceu?”, devem-se transcrever as duas palavras “que”.

Quadro de sinais de transcrição

Ocorrências	Critério/ sinal	Exemplos ²
Identificação dos falantes	Os turnos do documentador e do informante devem ser indicados, respectivamente, por D1 e S1 seguidos de dois pontos e um espaço simples. No caso de haver mais de um documentador ou mais de um informante, eles devem ser identificados por D1, D2 etc. (para os documentadores) e S1, S2, S3 etc. (para os informantes).	D1: você mora há quanto tempo nesse bairro? S1: eu moro aqui desde que eu nasci S2: os pais dele são daqui também D2: e (vo)cê gosta de morar aqui?
Linhas	Pular apenas uma linha para indicar a mudança de turno de fala, i.e. não deve haver uma linha em branco entre os turnos.	D1: (vo)cê gostou? S1: ô e como!
Grafia das palavras	Digitar a transcrição em letras minúsculas, exceto nos casos especificados a seguir.	
Nomes próprios de pessoas não públicas	Indicar apenas a inicial em maiúscula, seguida de ponto.	daí a M. veio aqui e...
Nomes próprio de pessoas públicas	Transcrever por extenso, com letra inicial maiúscula	quando o Lula se reelegeu...
Demais nomes próprios	Transcrever por extenso, com letra inicial maiúscula (exceto preposições):	Rio Grande do Sul Cem Anos de Solidão Banco do Brasil
Números e nomes de letras	Transcrever sempre por extenso:	eu já (es)tou com sessenta e três anos... o melhor filme que eu já vi... é... Dois Mil e Um Uma Odisseia no Espaço eu não gosto daquele erre carioca
Siglas e abreviaturas	Grafar conforme a pronúncia do informante. Se pronunciada letra a letra (ex.1), grafar em caixa alta, separando as letras por ponto. Se pronunciada como palavra (ex.2), seguir a grafia prevista pela ortografia, em caixa alta e sem pontos entre as letras.	Ex. 1: B.O., I.N.S.S., I.N.P.S., U.F.R.J., R.G., C.P.F. Ex. 2: USP, IAMSP, TAM, SUS, UFSCAR, CIC.
Pontuação	NÃO usar ponto final e vírgula. Ponto de exclamação e ponto de interrogação devem ser usados para indicar entonação de certas sentenças.	
Pausas curtas	Indicar por reticências ...	mas... sei lá eu acho que...
Pausas longas (acima de 3 segundos)	Além das reticências, deve-se especificar a duração em segundos entre colchetes	deixa eu ver... [4] não acho que não

² Nos exemplos abaixo, as partes em negrito servem apenas para destacar o critério exemplificado. Na transcrição, esse recurso não será empregado.

Interjeições e marcadores conversacionais	Usar as seguintes formas:	hein, ahn, aham [concordando], aham [negando], uhn, uhum, tsc tsc, né?, tá?, vixe, ixe, poxa, ó...
Palavras estrangeiras	Transcrever de acordo com a ortografia padrão da língua utilizada.	eu gosto bastante de Friends... CSI...
Silabação	Digitar logo após a palavra [silab.].	ele ficou apavorado [silab.]
Hipótese do que se ouviu	Quando houver incerteza quanto ao que o falante disse, digitar a palavra ou expressão entre parênteses.	(hoje em dia) a situação em São Paulo
Trechos ininteligíveis	Quando não entender o que foi falado, indicar por (xxx). No caso de se tratar de um trecho com mais de 3 segundos, indicar o tempo em segundos (xxx-5).	mas aí não tem condições né (xxx) não dá...
Truncamentos	Indicados por / (barra) após a última sílaba da palavra proferida.	ca/ cachorro
Apagamento de alguns segmentos	O apagamento de (i) /r/ em coda silábica; (ii) sílaba /es/ do verbo “estar”, em todos os tempos e modos verbais; e (iii) sílaba “vo” de “você(s)” deve ser indicado entre parênteses. Outros apagamentos (p.ex. “pe(i)xe”, “falamos(s)”) não devem ser indicados.	come(r), mulhe(r) (es)tou, (es)tivemos (vo)cê, (vo)cês
Dados contextuais	Notificar entre colchetes.	[risos] [ruído] [tosse] [informante se levanta para buscar água]
Hesitação	Indicar por [hes.]	ele ahn... [hes.] nem sempre
Sobreposição de vozes	Marcar com chaves { }, no turno de ambos os falantes, os trechos em que há sobreposição de vozes.	S1: o Rio é lindo né... D1: {você já esteve lá?} S1: {sempre gostei de ver o Rio nas} imagens da novela da Globo...
Metalinguagem do informante	‘aspas simples’	carioca fala ‘porta’... [r aspirado] ‘biscoito’... [s palatoalveolar] paulistano não fala assim
Citações/ discurso direto	“aspas duplas”	aí ele disse assim “(vo)cê vai?” e eu disse... “depende”...

Referências

- Castilho, Ataliba T. & Dino Preti (eds.) (1986). **A linguagem falada culta na cidade de São Paulo: materiais para seu estudo**, vol. I – Elocuções formais. São Paulo: T.A. Queiroz.
- Hellwig, Birgit, M. Tacchetti & A. Somasundaram (2011) **ELAN – Linguistic Annotator**. Versão 4.1.0. Disponível em <http://www.lat-mpi.eu/tools/elan/manual/>. Acesso em 03/05/2011.
- Hornik, Kurt. (2011) **R FAQ**. Disponível em <http://cran.r-project.org/doc/FAQ/R-FAQ.html>. Acesso em 10/04/2011.
- Mello, Heliana & Tommaso Raso (2009). Para a transcrição da fala espontânea: o caso do C-ORAL-BRASIL. In: **Revista Portuguesa de Humanidades**. Estudos Linguísticos, 13-1: 301-325.
- Tenani, Luciani E. & S. C. L. Gonçalves (s/d). **Manual do sistema de transcrição de dados** (v.5) – Projeto ALIP (Amostra Linguística do Interior Paulista). Ms.

ANEXO V:

PLANO DE ATIVIDADES PARA BOLSISTAS TT-1

SP2010 – Construção de uma amostra da fala paulistana

Plano de Atividades para Bolsistas TT-1

Resumo

As atribuições básicas do bolsista TT-1, no âmbito deste projeto regular de construção de uma amostra de fala paulistana, consistirão na coleta e na transcrição de 15 entrevistas sociolinguísticas, a serem gravadas com informantes paulistanos, ao longo de um período de dois anos (12 meses, prorrogáveis por mais 12). Anterior e paralelamente a essas tarefas, o bolsista TT-1 também deverá participar de reuniões de treinamento e de discussão.

Objetivos

As atribuições do bolsista TT-1, abaixo discriminadas, têm os objetivos de:

- Contribuir para a construção de um *corpus* de língua oral do português paulistano que conterà, ao final, 90 entrevistas sociolinguísticas; as gravações e suas respectivas transcrições serão disponibilizadas em *site* próprio na internet e deverão permitir o desenvolvimento de pesquisas de descrição e de análise dessa variedade de fala;
- Introduzir o bolsista no arcabouço-teórico metodológico da Sociolinguística Variacionista, de modo que o aluno possa vir a desenvolver posteriormente estudos nessa linha de pesquisa, com sólida formação básica.

Plano de Trabalho

Serão obrigações do bolsista TT-1:

- Participar das reuniões de treinamento (orientações gerais para coleta da entrevista; discussão de textos básicos da literatura sociolinguística; palestras com professores visitantes);
- Participar de reuniões de discussão sobre questões práticas do trabalho de campo e reportar à equipe como um todo (coordenador, alunos de pós-graduação e demais bolsistas TT-1) a experiência de coleta, com o objetivo de possibilitar a revisão dos procedimentos e a aprimoração dos métodos de coleta;
- Coletar 15 entrevistas sociolinguísticas;
- Transcrever 15 entrevistas sociolinguísticas e anotar, juntamente à tarefa de transcrição, dados pessoais dos informantes que possam identificá-los;
- Remover, dos arquivos de áudio de 15 entrevistas sociolinguísticas, dados pessoais que possam identificar o informante ou outras pessoas não-públicas.

Cronograma

Plano de atividades	Semestres			
	1º	2º	3º	4º
Treinamento para coleta de entrevistas	X			
Reuniões para discussão de textos teóricos e de questões sobre trabalho de campo	X	X	X	
Gravação de 3 entrevistas (1/5) – grupo A	X			
Transcrição das entrevistas do grupo A		X		
Gravação de mais 6 entrevistas (2/5) – grupo B		X		
Transcrição das entrevistas do grupo B			X	
Gravação de mais 6 entrevistas (2/5) – grupo C			X	
Transcrição das entrevistas do grupo C				X
Edição das entrevistas (remoção de dados pessoais de pessoas não-públicas)				X

Justificativa

As bolsas TT-1 serão destinadas a alunos de graduação em Letras a serem selecionados dentre aqueles que demonstrem interesse em desenvolver pesquisa em Sociolinguística no futuro. Nesse sentido, o treinamento e a prática de coleta de dados é fundamental como passo inicial em sua formação: por um lado, a obtenção de amostras de fala é imprescindível para o desenvolvimento de pesquisas neste campo e, por outro, a experiência com essas tarefas permitirá ao aluno entrever os objetivos e questões mais gerais que envolvem o trabalho do sociolinguista (por que coletar gravações, de que modo, com quais pessoas, que consequências esses métodos podem ter para análises posteriores etc.). As atribuições básicas desses bolsistas – coletar e transcrever entrevistas sociolinguísticas de maneira padronizada – estão plenamente de acordo com a finalidade da bolsa. Ao mesmo tempo, as atribuições adicionais de leitura e discussão de textos teóricos básicos, bem como a reflexão e o debate acerca do trabalho de campo com demais colegas, apesar de, em princípio, ultrapassar o caráter técnico de suas tarefas, não podem ser desvinculadas da prática de coleta e de transcrição. Para que a qualidade final do material coletado e transcrito, que se pretende disponibilizar publicamente à comunidade de linguistas, esteja de acordo com o rigor que se impõe, é necessário que os documentadores-bolsistas tenham plena consciência do contexto em que se inserem suas tarefas.