



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

RELATÓRIO CIENTÍFICO PARCIAL

Relatório Científico Parcial referente ao período de 1º/12/2011 a 30/11/2012 do Projeto Regular “SP2010 – Construção de uma amostra da fala paulistana” (Proc. FAPESP 2011/09278-6), com vigência de 1º/12/2011 a 30/11/2013.

Pesquisador Responsável: Ronald Beline Mendes

São Paulo
novembro de 2012

1. Introdução

Quando se submeteu à FAPESP o projeto “SP2010 – Construção de uma amostra da fala paulistana”, com a solicitação de auxílio financeiro para a construção de uma (nova) amostra do português da cidade de São Paulo, a principal motivação era contribuir para a mudança de um quadro que, embora explicável, pode ser bastante incômodo: a cidade de São Paulo ainda não teve sua fala amplamente descrita e o paulistano ainda não foi suficientemente estudado como ser sociolinguístico.

Conforme destacou o texto do projeto originalmente submetido, é verdade que trabalhos já foram feitos. É sempre importante lembrar Rodrigues (1987) e Coelho (2006), além de projetos de grande envergadura, como o NURC/SP (Castilho & Preti, 1986, 1987; Preti & Urbano, 1988, 1990) e o Projeto Para a História do Português Paulista (Castilho, 2007), entre outros. Entretanto, a fala paulistana ainda carece de investigação, diferentemente daquilo que se observa para a fala do paulista do interior, do carioca, do gaúcho, do paraibano – graças aos respectivos projetos ALIP, PEUL, VARSUL e VALPB (para citar apenas alguns).

O desenho do presente projeto de coleta de dados se baseou amplamente na experiência piloto do Grupo de Estudos e Pesquisa em Sociolinguística da USP (GESOL), entre os anos de 2009 e 2011, em que foram coletadas e transcritas dezenas de gravações (doravante denominadas Amostra-Piloto). Tais entrevistas tiveram caráter exploratório e, nesse sentido, carecem de qualidade técnica (sobretudo quanto à presença de ruídos); no entanto, elas não só já vêm sendo utilizadas em pesquisas desenvolvidas pelos membros do grupo como também permitiram a previsão de desafios para o presente projeto (p.ex., a localização de certos perfis sociolinguísticos).

Assim, com vistas a permitir uma ampla descrição e análise da variedade paulistana, o objetivo central do presente projeto é construir uma amostra contemporânea da fala de nascidos e criados na cidade de São Paulo, estratificada de acordo com três variáveis sociais (sexo/gênero, faixa etária e nível de escolaridade). As variantes dessas variáveis levam a 12 perfis sociolinguísticos; cada um deles é representado por cinco

informantes (um de cada uma das cinco zonas da cidade de São Paulo), de modo que a amostra, quando concluída, totalizará 60 informantes/entrevistas.

Também é objetivo central do projeto disponibilizar a amostra em site próprio na Internet – de modo que trabalhos de análise possam vir a ser realizados não apenas por atuais ou futuros membros do GESOL, mas também de outros centros. No primeiro ano da vigência do projeto, como não poderia deixar de ser, o foco das atividades foi o treinamento de documentadores e a gravação das entrevistas com a melhor qualidade possível (tanto em termos de áudio quanto em termos de um roteiro que permita amplitude de análises). Em virtude disso, a qualidade dos gravadores e a atenção dos documentadores em seu trabalho de campo podem ser considerados os itens mais importantes nesse primeiro ano.

Essas entrevistas estão sendo gravadas por três bolsistas (TT-1), alunas de graduação e membros do GESOL-USP, que foram e vêm sendo treinadas não só na prática do trabalho de campo, mas também no levantamento de questões teórico-metodológicas caras à pesquisa sociolinguística. De fato, é também objetivo central desse projeto formar pesquisadores no nível da graduação e encaminhá-los na direção da pós-graduação.

A seleção e o trabalho dessas bolsistas, bem como os equipamentos e a metodologia que vêm empregando, estão descritos na seção 2 a seguir, que focaliza os recursos humanos e materiais que o projeto move. Na seção 3, faz-se um relato detalhado do andamento do projeto, seguido, na seção 4, de uma descrição e avaliação dos seus resultados parciais. Finalmente, descrevem-se os passos que serão tomados no segundo e último ano de desenvolvimento do projeto, com uma revisão do cronograma de atividades.

2. Recursos humanos e materiais

Além do pesquisador responsável, trabalham diretamente nesse projeto três bolsistas de Treinamento Técnico-1 (alunas de graduação do curso de Linguística da USP), duas doutorandas do programa de pós-graduação do mesmo departamento e, mais

recentemente (no último mês do período a que se refere este relatório), três mestrados. O item 2.1 apresenta uma descrição mais detalhada da equipe de pesquisadores.

No que toca a materiais, os gravadores e os microfones foram os itens essenciais para o desenvolvimento do projeto nesse primeiro ano. O item 2.2 trata não só deles, mas de todos os materiais que foram adquiridos com verba da FAPESP.

Finalmente, em 2.3 descrevem-se itens para os quais foram utilizadas verbas do Departamento de Linguística ou de seu Programa de Pós-Graduação, como contrapartida.

2.1. Pesquisadores do Projeto

Pesquisador responsável	Ronald Beline Mendes
Gerentes	Livia Oushiro (doutorado) Marília Vieira (doutorado)
Assessores (validação de gravações)	Rafael Stoppa Rocha (mestrado) Wendel dos Santos (mestrado) Mariane Esteves Bieler (mestrado)
Bolsistas TT-1	Ana Paula Piola Araújo (IC) Camila Barbosa (IC) Larissa Soriano (IC)

Quadro 1: Equipe do Projeto SP2010

A função do pesquisador responsável pelo projeto tem sido a de gerenciar seu desenvolvimento, orientar os alunos (tanto prática quanto academicamente) e dar visibilidade ao projeto (através de publicações e de apresentações de trabalhos em congressos – ver subseção 4.3).

O trabalho de gravação de entrevistas, conforme já se mencionou, é feito por três bolsistas TT-1. Sua seleção foi feita criteriosamente, com base no seu histórico escolar na graduação, seu interesse pela Sociolinguística e, especificamente, pelos objetivos do projeto SP2010. Camila Barbosa e Larissa Soriano já vinham trabalhando no GESOL (em reuniões de leituras, disciplinas, gravações e transcrições de entrevistas da Amostra-Piloto), ao passo que Ana Paula Araújo manifestou seu interesse mais recentemente (ao final de 2011, após ter cursado a disciplina de Sociolinguística, obrigatória para os alunos de graduação em Linguística). Todas elas, contudo, passaram por uma fase de treinamento para o início de suas atividades mais práticas, iniciada logo após a aprovação do projeto pela FAPESP (ver seção 3).

Cada uma dessas alunas apresenta seu relatório de atividades individualmente. Convém, contudo, relatar aqui que o trabalho de orientação dessas graduandas, embora gratificante (na medida em que se observam progressos e amadurecimentos tanto pessoais quanto acadêmicos), é também dos mais árduos, sobretudo no que diz respeito à observação de prazos e detalhes técnicos do projeto.

De fato, Ana Paula Araújo, graças ao seu trabalho de seleção de informantes, gravação e cuidados técnicos de toda sorte – essenciais ao trabalho de campo do sociolinguista – começa a dar-se conta de que essa área pode não ser aquela em que se vai firmar academicamente. Essa bolsista manifestou, inclusive, interesse em não renovar a sua bolsa para o segundo ano de vigência do projeto; nesse caso, deverá fazer parte dos próximos passos do projeto selecionar outro bolsista, para os trabalhos programados para 2013.

Por outro lado, a experiência com a bolsista Camila Barbosa, apesar de suas reiteradas afirmações de interesse pela Sociolinguística como campo de atuação e de carreira, tem sido frustrante no que diz respeito ao cumprimento de prazos e à observância de detalhes técnicos do projeto, estabelecidos de comum acordo. Felizmente, o relatório parcial elaborado pela aluna, além de observar o prazo estabelecido pela FAPESP, revela consciência das dificuldades encontradas ao longo do primeiro ano e mostra uma atitude positiva em relação ao desenvolvimento das atividades que definirão o ano de 2013.

Já a bolsista Larissa Soriano, talvez por sua maturidade (havia-se graduado em Jornalismo e, após atuar por um tempo nessa área, decidiu por uma segunda graduação, em Letras), revelou-se extremamente atenta a detalhes do Projeto e bastante pragmática; concluiu todos os trabalhos que lhe foram designados dentro do prazo estabelecido pela agenda interna do grupo e, diante disso, sempre auxiliou suas colegas no cumprimento de sua parte de atividades. Além de todas essas qualidades, tem revelado excelente visão do conjunto do Projeto, de modo que poderá ser muito boa aquisição para o programa de pós-graduação do DL-USP – uma vez que a aluna já manifestou interesse em participar de seu processo seletivo, tão logo termine sua graduação.

Conforme descreve a seção 4, os resultados parciais do Projeto SP2010 são muito positivos: as entrevistas gravadas têm muito boa qualidade e faltam muito poucas a gravar; além disso, as reuniões de leitura, as reuniões para discussão de problemas e caminhos para sua solução no desenvolvimento do projeto, bem como o diálogo acadêmico com professores visitantes (sobretudo o Prof. Sebastião Carlos Gonçalves, da UNESP de São José do Rio Preto, responsável pelo Projeto ALIP) foram atividades que, seguramente, marcaram o ano de 2012 para a formação dessas bolsistas.

Por outro lado, considera-se da maior importância que faça parte da formação desses jovens iniciantes no trabalho acadêmico o aprendizado de que o grupo está “acima” do indivíduo – ou seja, o trabalho individual (sobretudo no caso de um projeto financiado por uma agência de fomento) *precisa* observar constantemente as definições acordadas pelo grupo. Dessa forma, esse trabalho de orientação deverá continuar não só presente, mas também em destaque, na agenda do segundo ano de vigência do Projeto.

Todo o trabalho de gerenciamento prático do trabalho das bolsistas vem sendo feito pelo pesquisador responsável com a valiosíssima assessoria de duas doutorandas, Livia Oushiro e Marília Vieira. Deve ser mencionado neste relatório que seu trabalho tem sido fundamental para o desenvolvimento do Projeto SP2010, não só em termos mais práticos, mas também acadêmicos, na medida em que ambas estão sempre atentas ao cumprimento de prazos internos (além daqueles estabelecidos com a FAPESP) e à excelência do trabalho em curso. De fato, o Projeto não estaria no ponto em que está (muito próximo daquilo que se havia programado no cronograma enviado à FAPESP no momento de sua submissão),

se não fosse pelo trabalho dessas doutorandas. Vale ainda esclarecer que suas pesquisas de doutorado não dependem dos dados que estão sendo coletados e organizados na amostra de 60 entrevistas, já que, em ambos os casos, estão sendo utilizadas entrevistas que foram gravadas na constituição da Amostra-Piloto.

Além da assessoria direta ao pesquisador responsável, essas doutorandas desenvolvem o trabalho de validação das entrevistas (ver subseção 4.2) juntamente com três mestrandos: Rafael Stoppa Rocha, Wendel dos Santos e Mariane Esteves Bieler. Até o início do segundo semestre de 2013, o primeiro foi quem mais trabalhou na validação de entrevistas junto com as doutorandas; a partir de então, em virtude da finalização de sua dissertação de mestrado (que deve defender até fevereiro de 2013), Rafael Rocha foi temporariamente afastado de suas funções no Projeto, que foram assumidas por Wendel dos Santos e Mariane Bieler, recém ingressantes no programa de pós-graduação do DL (ambos já amplamente familiarizados com os interesses e requisitos do Projeto).

Além desses alunos que trabalham na construção da amostra SP2010, deve-se mencionar outros membros do GESOL que, embora não façam parte “oficialmente” da equipe do Projeto, contribuem indiretamente para sua concretização: Dayane Almeida (doutorado), Fernando Gomes da Silva (mestrado) e Rafael Ciancio (iniciação científica). Além de participarem de reuniões do Projeto, esses alunos, graças aos seus trabalhos, dialogam com o Projeto e seus participantes mais diretos, sobretudo no que diz respeito a questões metodológicas: Dayane Almeida está construindo uma amostra de textos a partir de perfis sociolinguísticos para fins de pesquisa forense; Fernando da Silva está gravando alagoanos que moram em São Paulo; Rafael Ciancio está construindo uma amostra especificamente com paulistanos que vivem no bairro do Itaim Bibi.

Dessa forma, esse primeiro ano de desenvolvimento do Projeto contou com a participação de 3 doutorandas, 4 mestrandos e 4 alunos de Iniciação Científica – um grupo bastante interessante (apesar das dificuldades relatadas) que poderá crescer, em número e em qualidade. Nesse sentido, agradecemos à FAPESP, cujo auxílio tem sido essencial para que tais avaliações possam ser feitas na ocasião desse primeiro relatório de atividades.

2.2. Recursos materiais

Ao longo do ano de 2012, a verba disponibilizada pela FAPESP foi empregada nos seguintes itens, cronologicamente organizados:

Data	Modalidade/Montante	Item
28/fev	Material permanente R\$ 5.998,15	5 gravadores de voz TASCAM DR-100
10/abr	Material permanente R\$ 344,50	5 carregadores de pilhas para gravadores
18/abr	Material de Consumo R\$ 905,00	5 cartões de memória 32GB
07/mai	Material permanente R\$ 7.186,30	5 microfones Sennheiser HMD-26
10/jul	Benefícios complem. R\$ 3.411,71	Passagem aérea SP-Berlim-SP
9/ago	Material permanente R\$ 7.973,48	4 laptops DELL Inspiron 14
14/ago	Material permanente R\$ 403,63	HD externo portátil 1TB
28/set	Benefícios complem. R\$ 2.500,00	Passagem aérea SP-Indianapolis-SP

Quadro 2: Utilização da verba disponibilizada pela FAPESP

Todos os itens de material permanente e de consumo foram adquiridos estritamente dentro daquilo que foi aprovado pela FAPESP, tanto no que diz respeito ao número de itens quanto ao seu valor. As passagens aéreas adquiridas foram utilizadas para a participação do pesquisador responsável em dois congressos internacionais, respectivamente o *Sociolinguistics Symposium 19*, na Universidade Livre de Berlim (em agosto) e o *New Ways of Analyzing Variation 41*, na Universidade de Indiana, em Bloomington (Indiana, EUA, em outubro).

Vale ainda relatar que, originalmente, havia-se pedido à FAPESP 5 bolsistas (daí os cinco gravadores e cinco microfones). A agência aprovou os cinco itens de cada material permanente, mas aprovou “apenas” três bolsistas. Na utilização desses materiais, deu-se

conta de que a qualidade das entrevistas ficava ainda melhor com a utilização de dois microfones (um para o documentador e outro para o informante – seis no total). Dessa forma, pediu-se à FAPESP autorização para adquirir um sexto microfone – ao que a agência respondeu afirmativamente. Entretanto, desde então, o preço de tal material aumentou e, até o momento, não se adquiriu tal microfone. Isso deverá ser feito entre dezembro de 2012 e fevereiro de 2013.

2.2.1. Contrapartida da Instituição

Para a participação nos dois congressos elencados acima, o programa de pós-graduação do DL-USP auxiliou o pesquisador responsável com diárias, no valor aproximado de R\$ 4000,00 (8 diárias de aproximadamente US\$250,00)

Além disso, as visitas dos dois professores – Sebastião Carlos Leite Gonçalves (UNESP) e Erez Levon (University of London, Queen Mary) – foram financiadas com verba do programa de pós-graduação e com verba departamental (passagens e diárias totalizando, em conjunto, aproximadamente R\$7600,00).

Por fim, o Departamento de Linguística também cedeu uma de suas salas para o funcionamento do Projeto SP2010 (sala 5 do Prédio Letras/FFLCH), que tem abrigado as reuniões da equipe e o armazenamento do material coletado.

3. Andamento do Projeto

As atividades do Projeto SP2010 iniciaram-se oficialmente em dezembro de 2011. A equipe vem se reunindo regularmente em encontros que têm por objetivo discutir (i) aspectos do andamento do Projeto SP2010 e (ii) leituras relacionadas às tarefas de coleta e organização de *corpus* na Sociolinguística. Destaca-se que têm participado desses encontros não apenas os membros diretamente envolvidos no Projeto SP2010 (professor responsável, duas gerentes e três bolsistas), mas também os demais membros do GESOL,

alguns dos quais têm experiência com a coleta de entrevistas sociolinguísticas e/ou também desenvolvem projetos de pesquisa que envolvem trabalho de campo (ver seção 2.1).

O Quadro 3 apresenta um resumo dos encontros do GESOL relacionados ao desenvolvimento do primeiro ano do Projeto.

	Andamento do Projeto	Leituras
Dez/2011	13/12: Bolsas TT, objetivos do Projeto, obrigações das bolsistas	
Jan/2012	23/01: Preparação para as entrevistas	
Fev/2012		16/02: Tagliamonte 2006
Mar/2012	15/03: Validação de entrevistas 28/03: Distribuição dos gravadores; treinamento sobre seu uso	
Abr/2012	18/04: Distribuição dos carregadores, andamento das gravações, uso da planilha no Skydrive, sistemática	26/04: Labov 2006, cap. 4 e 5
Mai/2012	10/05: Andamento das gravações: experiências de campo, diário de campo 17/05: Reunião com Prof. Sebastião Gonçalves	18/05: Gonçalves 2008 – Projeto ALIP Palestra do Prof. Sebastião Gonçalves (UNESP/IBILCE)
Jun/2012	14 a 18/06: Minicurso do Prof. Erez Levon (University of London, Queen Mary)	
Jul/2012	--	--
Ago/2012	09/08: Andamento das gravações	28/08: Labov 1984

Set/2012	17/09: Andamento das gravações	25/09: Poplack 1989
Out/2012	01/10: Andamento das gravações 15/10: Andamento das gravações	
Nov/2012	14/11: Relatórios parciais individuais das bolsistas	

Quadro 3: Calendário de encontros do grupo de estudos

Paralelamente ao período de aquisição de equipamentos, as reuniões do primeiro quadrimestre (dezembro/2011 a março/2012) se concentraram no treinamento das bolsistas. Ainda que duas das três alunas responsáveis pela coleta de gravações já tivessem experiência nessa tarefa pelo fato de terem participado da fase piloto do projeto, foi fundamental certificar-se de que todas estavam cientes dos objetivos da nova fase de coleta de entrevistas, dos métodos a serem empregados, das funções de cada membro da equipe e de suas obrigações tanto para com a equipe quanto para com a FAPESP.

Na fase de treinamento, buscou-se, tanto quanto possível, o engajamento de todos os membros do GESOL nas decisões metodológicas. Na preparação para as entrevistas, cada membro do grupo de estudos ouviu três gravações da Amostra-Piloto, sendo que pelo menos duas pessoas ouviram uma mesma entrevista. O objetivo foi discutir o que é uma boa entrevista sociolinguística dentro dos objetivos do presente projeto, e definir estratégias boas (a serem emuladas) e estratégias ruins (a serem evitadas) durante a coleta de dados. A partir dessa discussão, começou-se a elaborar uma planilha com os critérios que seriam empregados na validação das entrevistas.

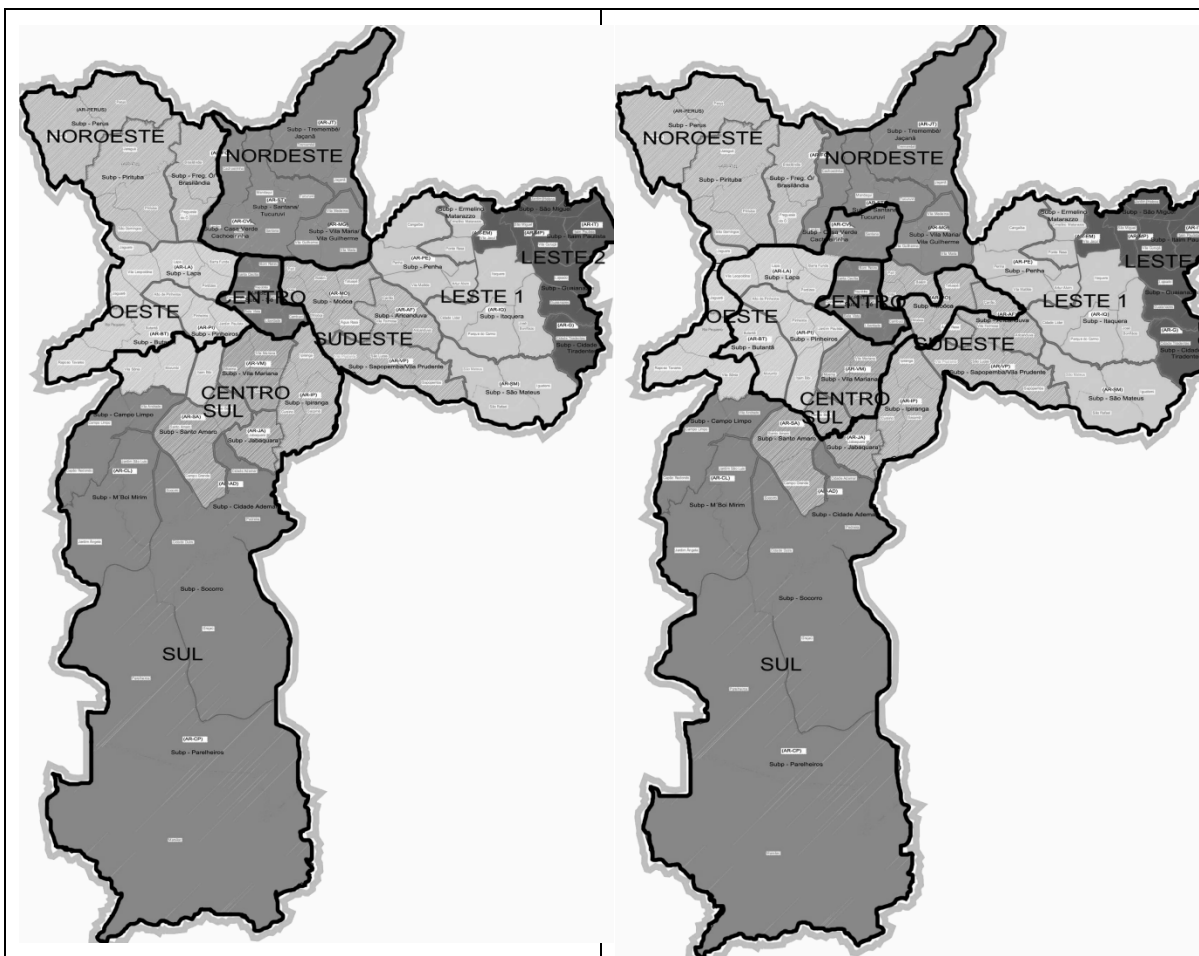


Fig. 1: Divisão da cidade de São Paulo em cinco zonas.

Fig. 2: Divisão da cidade de São Paulo em três regiões.

Para a reunião seguinte, os participantes ouviram três novas entrevistas da Amostra-Piloto e avaliaram cada uma delas a partir dos critérios levantados na reunião anterior. Discutiu-se a pertinência de cada critério, assim como as dificuldades em aplicá-los; em particular, foram extensivamente discutidos os critérios para determinar o perfil sociolinguístico dos informantes quanto ao nível de escolaridade e região de residência – p.ex., um informante com 40 anos de idade que acabou de começar a fazer um curso universitário deveria ser classificado igualmente como uma pessoa de 20 anos que iniciou o nível superior logo após a conclusão do Ensino Médio? Um informante de 25 anos que atualmente vive em Pinheiros (Zona Oeste, Centro expandido) mas que até os 22 anos residiu em São Mateus (Zona Leste, Periferia) deve ser considerado como residente de qual

zona e região? E onde se encontram tais limites geográficos? Tais casos de mobilidade geográfica e socioeconômica parecem constituir a regra na cidade de São Paulo, não exceções. A partir de tal discussão, estabeleceram-se mais rigidamente os limites das zonas e regiões da cidade de São Paulo (ver Fig. 1 e 2); determinou-se que o local de residência do informante deve ser aquele em que residiu por mais tempo nos últimos dez anos (e que informantes que residem há mais tempo no mesmo local são mais “prototípicos” de seu perfil e, portanto, devem ter a “preferência” do Projeto); e decidiu-se que informantes com nível superior em curso só são classificados como de nível superior se estão na primeira faixa etária (20 a 34 anos), mas não se estiverem na segunda ou terceira faixas etárias (acima de 35 anos). O estabelecimento de tais critérios tinha o objetivo de evitar decisões *ad hoc* quanto ao perfil em que cada informante que seria gravado deveria se encaixar.

Tal discussão também gerou a planilha final de validação das entrevistas (ver Anexo V e subseção 4.2). A participação das bolsistas nessa atividade foi importante para que tivessem ciência dos principais pontos aos quais deveriam atentar na seleção de informantes e durante a realização da gravação, já que suas entrevistas seriam avaliadas a partir de tais critérios.

Na fase de treinamento, também foram apresentados e discutidos outros materiais e recursos utilizados no Projeto:

- a Declaração de Consentimento a ser assinada pelo informante (Anexo II);
- a Ficha do Informante, a ser preenchida após a gravação (Anexo IIIa);
- o Questionário Socioeconômico (Anexo IIIb);
- planilha de perfis dos informantes a serem gravados;
- o serviço *online* Skydrive,¹ através do qual os materiais do Projeto são compartilhados entre os membros da equipe.

Ao fim de março de 2012, antes do início da coleta das entrevistas, realizou-se uma reunião para que as bolsistas se familiarizassem com o manuseio dos gravadores TASCAM DR-100 (suas funções, a configuração em que as gravações devem ser realizadas, seus acessórios, a transferência da gravação para um computador) e dos

¹ <https://skydrive.live.com/>. O site Skydrive permite o armazenamento gratuito de até 25 GB de arquivos em qualquer formato de mídia (.doc, .pdf, .jpeg, .wav etc.).

microfones. As gerentes do projeto haviam realizado, previamente, uma série de testes com os gravadores em diferentes ambientes (abertos, fechados, com muito ou pouco barulho, com muito ou pouco vento etc.), a fim de determinar a melhor configuração para as gravações. Durante a reunião, além de realizar diversas gravações-teste, recomendou-se que as bolsistas fizessem novas gravações antes do início efetivo da coleta de dados, a fim de que adquirissem confiança em seu manuseio e que reconhecessem as propriedades de diferentes configurações.

As reuniões do grupo de estudos também determinaram a sistemática de trabalho no projeto. Todo o material necessário para a realização das gravações (roteiro, ficha do informante, declaração de consentimento, questionário socioeconômico, planilha com perfis a serem gravados) foi compartilhado entre os membros da equipe do projeto através do site Skydrive. Ao encontrar um informante, a documentadora deve preencher a célula correspondente com o seu nome; ao realizar a gravação, deve colocar o seu nome em negrito. Desse modo, sinaliza-se às demais documentadoras que tal perfil não precisa, em princípio, de novos informantes e que devem buscar falantes de outros perfis. Os materiais da coleta devem ser repassados a uma das gerentes do projeto tão logo quanto possível. Após checar a boa ordem do material (p.ex., se a declaração do informante está assinada, se a ficha do informante está totalmente preenchida e se o arquivo sonoro abre corretamente), a gravação é atribuída a um validador (coordenador do projeto ou um aluno de pós-graduação). Tais validações são disponibilizadas novamente no Skydrive, de modo que as documentadoras possam se beneficiar dos comentários às gravações e, assim, aprimorar as gravações seguintes.

A fim de garantir maior objetividade na tarefa de validação – ou seja, que o mesmo grau de exigência estava sendo aplicado a cada um dos critérios da planilha –, os “validadores” do projeto avaliaram, todos, um subgrupo de seis entrevistas sociolinguísticas, duas de cada documentadora, logo no início do período de coleta. Para três das entrevistas, houve um grande grau de concordância entre os “validadores”, cujas pontuações finais (de 20 a 100) não diferiram em mais de cinco pontos (p.ex., uma mesma gravação obteve as pontuações 87, 92 e 88.). As divergências quanto as outras três entrevistas foram discutidas, para que se chegasse a um consenso sobre o que deve ser avaliado em cada critério.

A avaliação das gravações e as dificuldades encontradas no trabalho de campo também foram objeto de discussão em reuniões do grupo de estudos. Apesar de ter havido algumas dificuldades relacionadas ao manuseio dos equipamentos durante as primeiras gravações, tais questões foram rapidamente sanadas. A principal dificuldade encontrada pelas documentadoras é a localização de certos perfis sociolinguísticos e/ou de falantes dispostos a ceder parte de seu tempo a um “desconhecido”. Por parte dos validadores, as principais preocupações levantadas se referem ao gerenciamento do tempo de gravação, uma vez que há certas gravações demasiadamente longas ou curtas, e ao fraseamento adequado das perguntas do roteiro, sobretudo daquelas referentes a avaliações linguísticas.

Os encontros para a discussão da leitura de textos teóricos têm o objetivo de complementar a formação das bolsistas e dos membros do GESOL através da reflexão sobre o trabalho de campo do sociolinguista. Nessas reuniões, duas questões recorrentes dizem respeito (i) às diferenças e semelhanças entre o Projeto SP2010 e projetos prévios de coleta de entrevistas sociolinguísticas (p.ex. Projeto ALIP – Gonçalves 2008; Project on Language Change and Variation – Labov 1984; Ottawa-Hull French Project – Poplack 1989); e (ii) as relações entre as problemáticas discutidas nos textos e a própria experiência dos alunos-documentadores. Tal exercício tem permitido aos alunos um olhar crítico quanto às diferentes decisões metodológicas tomadas neste e em outros projetos, bem como a perspectiva de que muitas das dificuldades encontradas em seu trabalho de campo também tendem a ocorrer em outras comunidades.²

Especificamente quanto ao Projeto ALIP, o grupo de estudos teve a oportunidade de discutir questões teórico-metodológicas diretamente com o coordenador de tal projeto, o Prof. Sebastião Carlos Leite Gonçalves (UNESP), que visitou a instituição-sede do Projeto SP2010 por convite do GESOL em 17 e 18 de maio de 2012. No primeiro dia, os membros do GESOL apresentaram seus respectivos trabalhos de Iniciação Científica, mestrado e doutorado, e tiveram a oportunidade de discuti-los com o professor visitante e os demais membros do grupo. No segundo dia, o Prof. Sebastião Carlos Leite Gonçalves proferiu a

² Novas leituras previstas na continuidade do Projeto incluem Hoffman & Walker (2010), sobre o projeto “Contact in the City”, que vem construindo uma amostra da fala de Toronto estratificada em etnoletos; e Mello & Raso (2009), sobre os procedimentos de transcrição no projeto C-ORAL-BRASIL.

palestra “Amostras de fala do interior paulista: aspectos teórico-metodológicos”,³ aberta ao público, em que tratou da experiência da constituição do banco de dados Iboruna.

O grupo de estudos também se beneficiou da visita do Prof. Erez Levon (University of London, Queen Mary), que foi convidado para ministrar um minicurso de uma semana e proferir uma palestra. A temática de suas exposições se concentrou em estudos de sexo/gênero e linguagem e, em especial, permitiu refletir sobre a íntima relação entre a coleta de dados e as perguntas específicas que se buscam responder em cada pesquisa.

4. Resultados parciais

Conforme estabelece o texto submetido originalmente à FAPESP e se resume na introdução deste primeiro relatório, o Projeto SP2010 prevê a gravação, a transcrição e a disponibilização na Internet de uma amostra de 60 entrevistas sociolinguísticas, cujos informantes estarão estratificados de acordo com o seu sexo/gênero, a sua faixa etária e o seu nível de escolaridade, bem como equilibradamente distribuídos pelas cinco zonas da cidade de São Paulo.

A primeira etapa do Projeto previa a coleta das 60 entrevistas e a transcrição de 36 delas. Até o momento, foram realizadas 67 gravações, que preenchem 45 (75%) dos 60 informantes da amostra; das 67 gravações, 50 foram avaliadas pelos validadores (36 das quais de perfis não repetidos), e 39 foram transcritas (34 de perfis não repetidos). A Fig. 3 abaixo permite a visualização do progresso nessas tarefas, comparando as quantidades previstas, realizadas na totalidade e as realizadas de acordo com os objetivos do projeto (ou seja, sem a repetição de certos perfis).

³ Ver <<http://comunicacao.fflch.usp.br/node/1612>>.

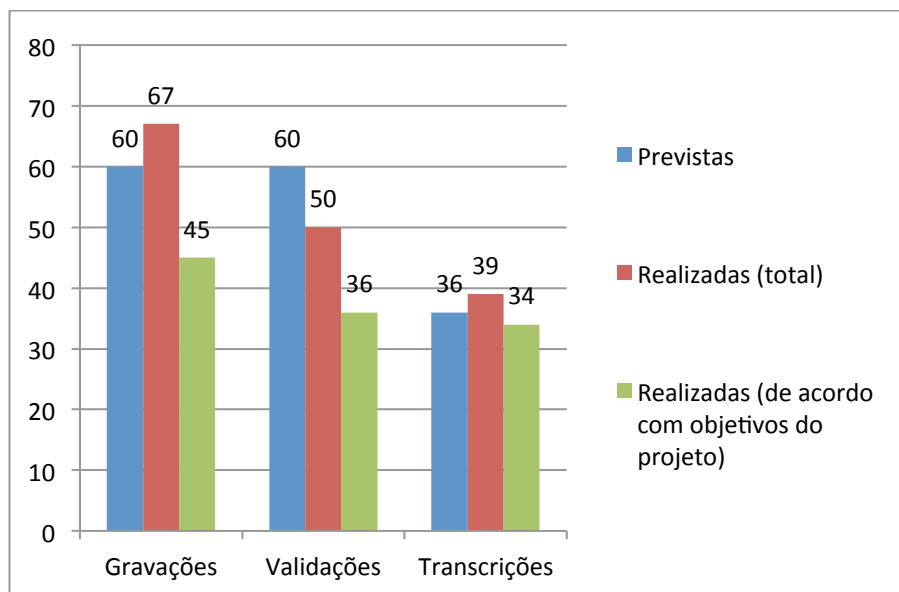


Fig. 3: Número de gravações, validações e transcrições previstas e realizadas na primeira etapa do Projeto SP2010

4.1 Entrevistas gravadas

A gravação de mais de uma entrevista para certos perfis sociolinguísticos numa determinada zona da cidade não foi desincentivada; ainda que a planilha organizadora das entrevistas tenha o objetivo de maximizar a busca por informantes de perfis diferentes, a gravação de mais de uma entrevista em certos casos pode fornecer mais prática de coleta às documentadoras, bem como conduzir a novos informantes através de indicações por parte do entrevistado.

Por outro lado, é necessário avaliar a dificuldade de encontrar certos perfis definidos na amostra, sobretudo em contraste com outros de mais fácil acesso. O Quadro 4 apresenta os 12 perfis estabelecidos pelo projeto, a partir das variáveis sexo/gênero, faixa etária e grau de escolaridade (primeira coluna), bem como as diferentes zonas da cidade (primeira linha)⁴. São portanto 60 os informantes/as entrevistas que compõem a amostra.

⁴ Essa variável, de acordo com o que foi estabelecido no projeto aprovado pela FAPESP, não estratifica a amostra quantitativamente, mas a caracteriza qualitativamente. Se se tratasse de variável definidora do tamanho da amostra, teríamos 2 (sexos) x 3 (faixas etárias) x 2 (graus de escolaridade) x 5 zonas = 60 perfis.

Os nomes (na verdade, pseudônimos) no Quadro 4 indicam os informantes já gravados; as células sem nome indicam informantes que não foram gravados até o momento.

PERFIS	Z. Central	Z. Norte	Z. Sul	Z. Leste	Z. Oeste
F1C	1	2	3 Janaina B. Paula R.	4 Lara N.	5
F1S	6 Valeria R.	7 Silvia B.	8 Raissa O. Viviane P.	9 Fabiana B.	10 Bianca V.
F2C	11 Renata C.	12 Marieta S.	13 Meire C. Tatiana G. Raquel V.	14 Raíssa C.	15
F2S	16	17	18 Morgana S. Luciana N.	19 Elaine S.	20 Fabiana M. Madalena M. Marta V.
F3C	21 Carmen F. Marcia M.	22 Soraia S.	23 Vera D.	24 Edna C.	25 Darcy M.
F3S	26	27 Renata L.	28	29 Sueli O.	30 Vivian T. Marly C. Bianca T.
M1C	31 Lucas S.	32 Douglas A.	33 Robson F.	34 Junior B. Dalson O.	35 Roberto S.
M1S	36 Breno A. Jaime N.	37 Rodrigo R. Fabiano A. Joaquim R.	38 Danilo S. Leandro S.	39 Luan O. Sergio A.	40 Gustavo M. Matheus M. Ricardo T.
M2C	41	42 Sandro S.	43 Nelson F.	44 Mauricio B.	45 Alberto M. Lucas T.

Considerando-se o ideal de 5 informantes por perfil (Labov, 2006), o número total de informantes seria 300 – um número de entrevistas evidentemente inviável para se gravar e transcrever no mesmo tempo de vigência de um projeto de pesquisa regular.

M2S	46 Murilo C.	47 Jose N.	48 Pedro S. Rafael V.	49	50 Frederico V. Marcos B.
M3C	51	52	53 Romulo S.	54 Rodolfo A.	55 Hugo F.
M3S	56	57	58 Nilton R.	59	60 Neimar V. Amadeu C.

Quadro 4: Entrevistas gravadas e faltantes de acordo com os perfis da amostra

As células referentes aos perfis F1S (6-10) e M1S (36-40), de modo esperado, são aquelas de acesso relativamente mais fácil às documentadoras pelo método amigo de amigo (Milroy 2004). Ao mesmo tempo, pelas características sociodemográficas da população, é também relativamente fácil encontrar falantes da terceira faixa etária que não tenham feito o curso superior, principalmente do sexo feminino (perfil F3S, 21-25). Por outro lado, é mais difícil encontrar falantes com menos de 35 anos, sobretudo mulheres, que tenham concluído no máximo o Ensino Médio (F1C, 1-5), assim como o perfil “inverso”, que se refere a homens da terceira faixa com nível superior de escolaridade (M3S, 56-60). Esses dois perfis, F1C e M3S, são aqueles que possuem o menor número de entrevistas e o maior número de células vazias (três).

Tal situação já havia sido vivenciada na coleta da Amostra-Piloto. Uma dificuldade adicional na presente fase de coleta de dados diz respeito à determinação de uma amostra balanceada e mais representativa da cidade através da distribuição dos cinco informantes por perfil pelas cinco zonas de São Paulo.⁵ Para sanar tal questão, os perfis restantes foram atribuídos aleatoriamente a documentadoras específicas, de modo que o seu trabalho de busca por informante se concentre em certos perfis.

⁵ De fato, caso não houvesse tal restrição, faltariam apenas seis entrevistas para a finalização da amostra.

4.2 Critérios de validação das entrevistas

Quanto à qualidade das entrevistas coletadas, foi elaborada uma planilha de validação (ver Anexo V) que contém (i) as informações gerais sobre cada gravação (nome do documentador, perfil do informante e nome dos arquivos); (ii) 20 quesitos a serem avaliados, cuja pontuação vai de 1 a 5 (ver abaixo); (iii) campos para comentários e sugestões do avaliador; (iv) uma avaliação qualitativa geral sobre a entrevista e sobre o informante; e (v) a pontuação final.

Os 20 quesitos a serem avaliados para cada gravação são os seguintes:

1. Perfil do informante (quão prototípico é o informante de acordo com o perfil – p.ex., vive há muito tempo no mesmo local; se fez faculdade, cursou-a logo após o colegial etc.): escala de 1 (menos prototípico) a 5 (mais prototípico)
2. Ficha do informante: 1 – incompleta; 5 – completa
3. Declaração do informante: 1 – não assinada; 5 – assinada
4. Questionário socioeconômico (opcional): 1 – não preenchido; 5 – preenchido
5. Duração da gravação: 5 – de 50 a 70 minutos; 3 – curta (de 40 a 50 minutos) ou longa (de 70 a 80 minutos); 1 – muito curta (menos de 40 minutos) ou muito longa (mais de 80 minutos)
6. Inteligibilidade da voz do informante: escala de 1 (pouco inteligível) a 5 (bastante inteligível)
7. Inteligibilidade da voz do documentador: escala de 1 (pouco inteligível) a 5 (bastante inteligível)
8. Nível de ruído de fundo: 1 – nível muito alto; 2 – nível alto; 3 – nível médio; 4 – nível baixo; 5 – pouco/nenhum ruído de fundo
9. Intervenção de terceiros: escala de 1 (excessivo) a 5 (adequado)
10. Sobreposição de vozes: escala de 1 (excessivo) a 5 (adequado)
11. Intervenções do documentador: escala de 1 (inadequado – interrompe muitas vezes/ deixa a conversa “morrer”/ muda de assunto bruscamente etc.) a 5 (adequado – nem muito, nem pouco)
12. Naturalidade da interação: escala de 1 (pouco natural) a 5 (bastante natural)

13. Aplicação do roteiro: cada subitem abaixo avaliado em uma escala de 1 a 5, de acordo com o desenvolvimento dos tópicos dentro dos objetivos do projeto (ver Anexo I)

- i. Bairro
- ii. Infância
- iii. Família
- iv. Trabalho/Ocupação
- v. Lazer
- vi. A cidade de São Paulo
- vii. Avaliação linguística
- viii. Leituras

A maior parte dos quesitos é escalar (de 1 a 5); três deles (Ficha do Informante, Declaração e Questionário Socioeconômico) são binários (1 ou 5); e um deles (Duração da entrevista) é ternário (1, 3 ou 5). A escolha do informante e o preenchimento da documentação (critérios 1-4) contam 20% da pontuação; aspectos gerais da interação e da qualidade técnica da gravação (critérios 5-12) somam 40%; e a aplicação adequada do roteiro (critérios 13i-13viii), tendo em vista os objetivos do Projeto, contribui para 40% da pontuação final. Visto que cada quesito compreende pontuações possíveis entre 1 e 5, a avaliação dos 20 quesitos produz uma pontuação final entre 20 e 100. A Fig. 4 abaixo apresenta a distribuição das pontuações das 50 gravações validadas até o momento.

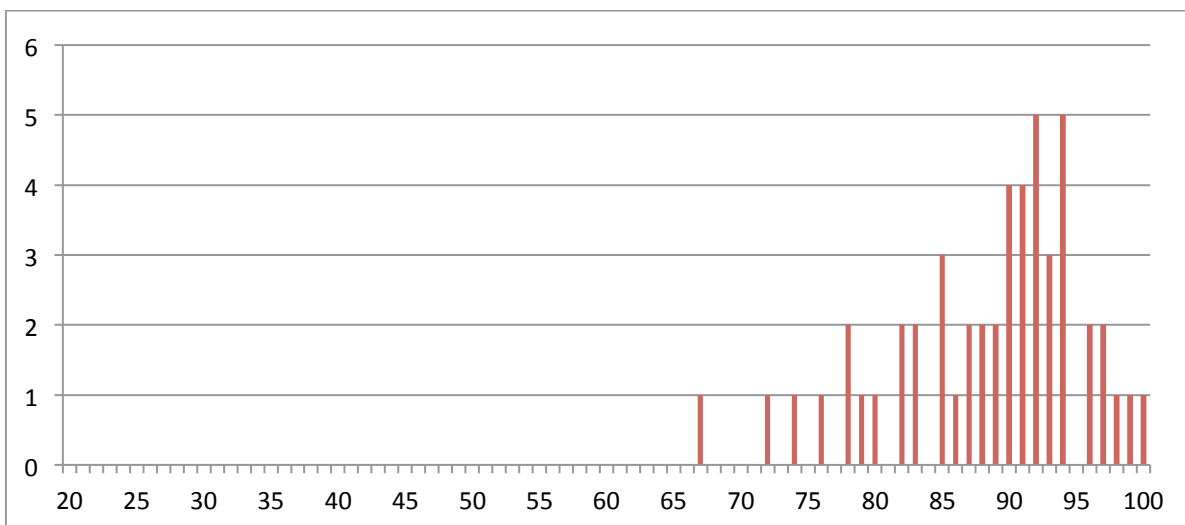


Fig. 4: Distribuição da pontuação das validações

A Fig. 4 mostra que as pontuações atribuídas presentemente vão de 67 a 100, e que a maior parte das entrevistas recebeu pontuações entre 85 e 95. A média das avaliações é de 88,4 e a mediana é 90. Pode-se considerar que, em sua maioria, a qualidade das entrevistas gravadas é muito boa e corresponde ao grau de exigência estabelecido pelo projeto. Possivelmente, ao final da coleta e da validação das gravações, será estabelecido um limite mínimo de 80 para que uma determinada entrevista integre o *corpus* do Projeto SP2010.

Três das entrevistas realizadas (gravação, transcrição, ficha do informante editada, ficha de validação), uma de cada aluna-bolsista, são ora submetidas à apreciação da FAPESP e do parecerista *ad hoc* através da página Windows Live Skydrive.⁶ Como um dos objetivos principais do projeto é disponibilizar a amostra de 60 entrevistas em site próprio – com as informações pessoais dos informantes editadas no áudio – não se disponibilizam todas as entrevistas já gravadas na ocasião deste primeiro relatório (até porque, até o momento, tais edições não foram feitas).

⁶ Acessar a página <https://skydrive.live.com/>. No campo “Windows Live ID”, digitar “SP2010@live.com”. No campo “senha”, digitar “amostra”. As entrevistas e os respectivos materiais encontram-se na pasta “Amostra SP2010”. Os arquivos de áudio aí disponibilizados foram convertidos para o formato .mp3, a fim de diminuir o seu tamanho. As gravações originais, das quais se mantêm cópias, têm cerca de 900 MB cada e ultrapassam o limite do Skydrive de 300 MB por arquivo.

4.3 Divulgação

Durante o seu primeiro ano de desenvolvimento, o Projeto SP2010 foi divulgado através de comunicações orais em quatro encontros científicos e da publicação em anais e em um periódico:

- MENDES, R.; OUSHIRO, L. (2012/no prelo) Mapping Paulistano Portuguese: the SP2010 Project. Comunicação oral apresentada no GSCP 2012 (Belo Horizonte, fev-mar/2012). Trabalho completo a ser publicado no GSCP 2012 Proceedings.

Esse trabalho teve o objetivo de divulgar, para um público amplo de linguistas interessados em corpora de língua oral, os objetivos, métodos e os resultados parciais do Projeto SP2010. O trabalho completo, anexo a este relatório através do SAGe, será publicado nos anais do evento.

- MENDES, R.; OUSHIRO, L. (2012) E o português paulistano? Comunicação oral apresentada no 60º Seminário do GEL (São Paulo, jul/2012).

Essa comunicação foi apresentada no Simpósio “Metodologia de Constituição de Corpus para o Estudo da Variação Linguística”, que contou com a participação de diversos pesquisadores envolvidos em projetos de constituição de banco de dados em diferentes áreas do país: Raquel Meister Ko Freitag (UFSE), “Banco de dados Falaes Sergipanos”; Silvia Figueiredo Brandão (UFRJ) e Silvia Rodrigues Vieira (UFRJ), “Organização de corpora para estudos sociolinguísticos: as experiências com os Projetos Varport e Concordância”; Sebastião Carlos Leite Gonçalves (UNESP), “Iboruna: um banco de dados com amostras de fala do interior paulista”; Marco Antonio Martins (UFRN) e Maria Alice Tavares (UFRN), “Banco de dados Fala-Natal: balanço inicial”. O objetivo foi discutir e comparar aspectos teóricos e metodológicos, como os procedimentos de coleta, armazenamento, disponibilização das amostras.

- MENDES, R.; OUSHIRO, L. (2012) Sampling São Paulo Portuguese. Comunicação oral apresentada no Sociolinguistics Symposium 19 (Berlim-Alemanha, ago/2012).

Essa comunicação integrou a sessão intitulada “Field Methods in Multicultural Megacities”, organizada por Ronald B. Mendes (USP) e James Walker (York University, Toronto), da qual participaram pesquisadores do Canadá, Noruega, Brasil e Alemanha. Aqui, o objetivo central foi discutir os desafios que surgem no estudo da variação e da mudança linguística em centros populosos e multiculturais como Toronto, Oslo, São Paulo, Rio de Janeiro, Teerã, Paris e Barcelona. Em particular, uma das dificuldades de tais empreendimentos sociolinguísticos é o cumprimento de dois requisitos conflitantes entre si: por um lado, a amostragem de um número de falantes representativo da complexidade sociodemográfica da comunidade e, por outro, a coleta de dados etnográficos de cada indivíduo a fim de melhor interpretar os resultados de análises desses *corpora*. Além de permitir a divulgação do Projeto SP2010 a um público amplo e internacional de sociolinguistas, a sessão possibilitou a discussão a respeito de diferentes métodos de amostragem, de representatividade da amostra, do contato linguístico e da percepção sobre identidades sociais em grandes centros urbanos.

- OUSHIRO, L.; MENDES, R. (2012) Does “covert prestige” entail “covert stigma”? Variable (-r) in São Paulo Portuguese. Comunicação oral apresentada no NWAV41 (Bloomington-EUA, out/2012).

O trabalho apresentado no NWAV 41 se baseia em dados coletados na Amostra-Piloto e no Projeto SP2010, de modo que representou nova oportunidade de divulgação do projeto para um público internacional. Tal comunicação apresentou resultados de análises quantitativas de covariação sobre o uso de (-r) em coda no português paulistano atual e de análises qualitativas sobre o metadiscorso dos informantes a respeito da variável, sobretudo através de perguntas do roteiro de entrevistas (ver perguntas 36-53 Anexo I).

- MENDES, R.; OUSHIRO, L. (no prelo) O paulistano no mapa sociolinguístico brasileiro. Alfa, vol. 57(1).

Esse artigo traz notícia dos trabalhos desenvolvidos pelo GESOL, que se organizam em torno da presente construção da amostra do Projeto SP2010. Ao apresentar os objetivos do projeto, os critérios de construção da amostra e os desafios nela envolvidos, o artigo demonstra as potencialidades de pesquisa da amostra já constituída (Amostra-Piloto) e em construção (Amostra SP2010), bem como a agenda de pesquisas do grupo de estudo. Nesse sentido, o Projeto SP2010 tem um papel fundamental na consolidação do grupo de pesquisas e contribui para a expansão da sociolinguística brasileira. Uma cópia do artigo segue anexa ao relatório através do SAGe.

5. Avaliação e Planejamento

A partir do relato formulado nos itens anteriores, é possível avaliar positivamente a condução das atividades do projeto no seu primeiro ano de vigência. A seção 2 deixa claro que o GESOL-USP está funcionando em volta do Projeto SP2010 e que, apesar de determinadas dificuldades individuais e pontuais, faltam poucas entrevistas a serem gravadas – até a data de submissão deste relatório.

O plano é que essas gravações sejam finalizadas até janeiro. A partir daí, o projeto se concentrará, sobretudo, no trabalho de transcrição das entrevistas no ELAN (com sincronização entre voz e texto) e na sua disponibilização em site próprio, a ser criado – de acordo com o cronograma abaixo.

Plano de Atividades	D	J	F	M	A	M	J	J	A	S	O	N
Seleção e treinamento de novo bolsista TT-1	X											
Término das gravações	X	X										
Término das validações		X	X									
Transcrições			X	X	X							
Edição das informações pessoais nos áudios e transcrições					X	X	X					
Criação do site							X	X	X	X	X	
Visita de professores convidados			X						X			

Referências

- 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)
- 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.
- 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>.
- Gonçalves, Sebastião Carlos Leite (2008) Projeto ALIP (Amostra Linguística do Interior Paulista): questões teóricas e metodológicas sobre constituição de um banco de dados de língua falada. In: Tagnin, E.; Vale, Oto Araújo. (Org.). **Avanços da Linguística de Corpus no Brasil**. 1ed. São Paulo: Humanitas, v. 1, p. 217-245.
- Hoffman, M. F. and J. Walker (2010) Ethnolects and the City: Ethnic orientation and linguistic variation in Toronto English. **Language Variation and Change** 22, 37–67.
- Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In: Baugh, J. & J. Sherzer (eds.), **Language in Use: Readings in Sociolinguistics**. Englewood Cliffs: Prentice Hall, p.28-54.
- Labov, W. (2006 [1966]). The isolation of contextual styles/The linguistic interview/Appendix A. In: **The Social Stratification of English in New York City**. Cambridge: Cambridge University Press, p. 58-97; 409-422.

- Mello, H. & T. Raso (2009) Para a transcrição da fala espontânea: o caso do C-ORAL-BRASIL. **Revista Portuguesa de Humanidades/Estudos Linguísticos** 13(1), 301-325.
- Mendes, R.; Oushiro, L. (no prelo) Mapping Paulistano Portuguese: the SP2010 Project. A ser publicado no GSCP 2012 Proceedings.
- Mendes, R.; Oushiro, L. (2012) E o português paulistano? Comunicação oral apresentada no 60º Seminário do GEL (São Paulo, jul/2012).
- Mendes, R.; Oushiro, L. (2012) Sampling São Paulo Portuguese. Comunicação oral apresentada no Sociolinguistics Symposium 19 (Berlim-Alemanha, ago/2012).
- Mendes, R.; Oushiro, L. (no prelo) O paulistano no mapa sociolinguístico brasileiro. **Alfa**, vol. 57(1).
- Oushiro, L.; Mendes, R. (2012) Does “covert prestige” entail “covert stigma”? Variable (-r) in São Paulo Portuguese. Comunicação oral apresentada no NWAV41 (Bloomington-EUA, out/2012).
- Poplack, S. (1989) The care and handling of a mega-corpus: the Ottawa-Hull French Project. In: Fasold, R. & D. Schiffrin (orgs) **Language Change and Variation**. Amsterdam: Benjamins, 411-451.
- 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.
- Tagliamonte, S. (2006) Data Collection/The sociolinguistic interview (cap. 2, 3). In: **Analysing Sociolinguistic Variation**. Cambridge: Cambridge University Press, p.17-49.