

BANCO DE DADOS DE LÍNGUA FALADA DE BAGÉ

Táise Simioni (UNIPAMPA)
taise.simioni@unipampa.edu.br

Introdução

O presente trabalho tem por objetivo apresentar as bases para a formação de um banco de dados de língua falada constituído por entrevistas com falantes da cidade de Bagé (RS). Os pressupostos teóricos que norteiam as etapas de construção deste banco de dados encontram-se na sociolinguística quantitativa (Labov, 2008 [1972]). Os dados assim coletados permitirão que se realizem pesquisas que descrevam as variedades linguísticas encontradas na região. A divulgação de tais pesquisas pode abrir caminho para a discussão sobre a variação linguística, de maneira que se desfaçam mitos e se construa o respeito às diferentes variedades linguísticas.

O projeto de pesquisa intitulado “Banco de Dados de Língua Falada de Bagé” propõe a constituição de um banco de dados de língua falada formado a partir de entrevistas realizadas com 36 falantes bajeenses. No presente trabalho, mostramos como se justifica a constituição de tal banco de dados, bem como explicitamos seus objetivos. Na sequência, apresentamos a metodologia adotada para que se alcance tais objetivos. Por fim, os resultados esperados são descritos.

1 Por que a formação de um banco de dados desta natureza?

Com a formação do “Banco de Dados de Língua Falada de Bagé”, pretendemos alcançar os seguintes objetivos:

- coletar dados de língua falada em Bagé que permitirão a realização de análises descritivas e explicativas das variedades aqui encontradas;
- disponibilizar tais dados e as pesquisas deles resultantes à comunidade acadêmica e não acadêmica;
- fornecer subsídios para a discussão sobre a variação linguística com diversos públicos, de maneira que se crie um ambiente de respeito às diferentes variedades linguísticas.

O projeto aqui descrito parte do inegável fato de que as línguas variam. Citamos como exemplo de variação a concordância nominal em português brasileiro, objeto de estudo de vários pesquisadores, como, por exemplo, Scherre (1996), que realizou sua pesquisa com dados de falantes do Rio de Janeiro. Em sintagmas nominais, o plural pode estar marcado (pelo morfema *-s*) ou pode estar ausente (\emptyset). Assim, podemos ter sintagmas como “os meninos” ou “os menino \emptyset ”.

Apesar de a variação linguística ser inegável, o tratamento que ela recebe pode variar muito dependendo da abordagem teórica que se utiliza. A presente pesquisa toma como pressuposto teórico a sociolinguística quantitativa (Labov, 2008 [1972]), como mencionamos anteriormente.

Inseridos em tal perspectiva teórica, assumimos que o aparente caos linguístico da variação pode ser sistematizado. Mais do que isto, tomamos a heterogeneidade como inerente a qualquer sistema linguístico. Nas palavras de Labov (2008, p. 238):

A existência de variação e de estruturas heterogêneas nas comunidades de fala investigadas está certamente bem fundamentada

nos fatos. É a existência de qualquer outro tipo de comunidade de fala que deve ser posta em dúvida. [...] essa é a situação *normal* – a heterogeneidade não é apenas comum, ela é a consequência natural de fatores linguísticos fundamentais.

Desta forma, cabe ao pesquisador sociolinguista buscar a sistematicidade por trás da variação, ou seja, é preciso investigar de que maneira os falantes fazem as escolhas que fazem. Como explica Beline (2003, p. 125), “de uma perspectiva variacionista quantitativa, [...], a sociolinguística ocupa-se em desvendar como a heterogeneidade – ou seja, a variação – se organiza. O sociolinguista tem como principal interesse compreender de que modo a variação é regulada”. No caso do nosso exemplo de concordância nominal, Scherre (1996) mostra, entre outros resultados, que (i) mulheres apresentam mais plurais marcados, ou seja, fazem mais a concordância, e (ii) mulheres “são mais sensíveis à atuação da escola do que os homens, no sentido de favorecer o uso da forma socialmente prestigiada” (Scherre, 1996, p. 254).

Sendo a variação um aspecto tão importante da língua, é preciso buscar subsídios para analisá-la e compreendê-la. O que propõe o projeto que está sendo descrito aqui, então, é a criação de um banco de dados de língua falada representativa da cidade de Bagé (RS). Tal banco de dados será constituído por entrevistas com falantes bajeenses que sempre residiram nesta cidade. Segundo Labov (2008, p. 244):

Não importa que outros métodos possam ser usados para obter amostras de fala (sessões em grupo, observação anônima), a única maneira de obter bons dados de fala em quantidade suficiente é mediante a entrevista individual, gravada, ou seja, por meio do tipo mais óbvio de observação sistemática.

Este banco de dados permitirá dar visibilidade à variação linguística que ocorre em Bagé e possibilitará, também, que se verifiquem as diferenças linguísticas entre as variedades aqui faladas e as variedades faladas em outras regiões do estado.

A divulgação destes dados poderá ter consequências interessantes para a sociedade, tendo em vista que será um instrumento importante no combate ao preconceito linguístico. Como alerta Scherre (2005, p. 112), “nós, falantes-linguistas e linguistas nem sempre muito falantes, não podemos nos omitir. Temos, todos nós, o dever de participar do debate público contra o preconceito linguístico”. Uma vez que se torne explícito que a variação linguística é inerente a qualquer língua e que se pode compreender as motivações por trás desta variação, isto pode auxiliar a se desfazer os mitos de que o português brasileiro é uma língua homogênea, de que existe um padrão que é a língua portuguesa (e não uma variedade da língua portuguesa tomada como exemplar) e de que tudo que se distancie deste padrão deve ser eliminado por ser um desvio que destrói a língua. Mostra-se, desta maneira, que qualquer julgamento de um fato linguístico como “certo” e “errado”, “bonito” ou “feio” está assentado em bases extralinguísticas. Julga-se, desta forma, o falante, e não seu modo de falar. Concordamos com Naro e Scherre (2006, p. 236) quando afirmam que “reconhecer e aceitar a diversidade linguística é uma questão de cidadania”.

O banco de dados cuja proposta de formação apresentamos aqui será constituído por 36 informantes, divididos de acordo com as seguintes variáveis: sexo, idade e escolaridade. No que diz respeito à variável sexo, os informantes serão divididos em masculino e feminino. Com relação à idade, dois fatores serão considerados: entre 25 e 50 anos e mais de 50 anos. Por fim, no que se refere à

escolaridade, os informantes serão divididos considerando-se as seguintes faixas: de um a quatro anos de escolaridade; de cinco a oito anos de escolaridade; e de nove a onze anos de escolaridade.

O diagrama abaixo mostra a distribuição dos informantes.

Homens	25 a 50 anos	1 a 4 anos de escolaridade	3 informantes
		5 a 8 anos de escolaridade	3 informantes
		9 a 11 anos de escolaridade	3 informantes
	mais de 50 anos	1 a 4 anos de escolaridade	3 informantes
		5 a 8 anos de escolaridade	3 informantes
		9 a 11 anos de escolaridade	3 informantes
Mulheres	25 a 50 anos	1 a 4 anos de escolaridade	3 informantes
		5 a 8 anos de escolaridade	3 informantes
		9 a 11 anos de escolaridade	3 informantes
	mais de 50 anos	1 a 4 anos de escolaridade	3 informantes
		5 a 8 anos de escolaridade	3 informantes
		9 a 11 anos de escolaridade	3 informantes

A estratificação dos dados foi feita em conformidade com o que foi realizado inicialmente no Projeto Varsul (Variação Linguística Urbana na Região Sul)¹, porém com um informante a mais por célula. Optamos por tal estratificação porque isto permitirá que as pesquisas realizadas a partir do banco de dados idealizado no projeto que descrevemos aqui sejam comparadas com pesquisas realizadas com os dados do projeto Varsul, que são inúmeras e cobrem diversas localidades dos três estados da região sul. Para exemplificar, o fato de Bagé estar localizada em uma região de fronteira com um país de língua espanhola certamente influencia a fala de seus habitantes. Assim, será possível verificar, por exemplo, a variação entre os sufixos de diminutivo *-inho/-zinho* e *-ito/-zito*, levando em consideração a hipótese de que *-ito* ocorrerá com mais frequência aqui do que em outras regiões do estado por influência do espanhol.

Em uma etapa futura de constituição do banco de dados, idealiza-se ampliar o número de informantes por célula, incluir o Ensino Superior nas faixas de escolaridade consideradas e refinar as faixas etárias analisadas, diminuindo o tempo contemplado por cada faixa.

Assim, a realização do projeto aqui descrito permitirá que se conheçam melhor as variedades faladas em Bagé, com consequências positivas para o desenvolvimento científico e social desta comunidade.

2 Metodologia

A etapa inicial da constituição do banco de dados é a preparação das pessoas envolvidas no projeto. Tal preparação envolve a leitura e a discussão de textos teóricos

1 Maiores informações sobre o projeto Varsul podem ser obtidas no site <http://www.pucrs.br/fale/pos/varsul/>.

sobre a sociolinguística quantitativa. Neste momento, serão formados os entrevistadores.

A etapa seguinte é a realização das entrevistas, que possuem dois momentos: a construção da ficha social e a entrevista propriamente dita. Esta última terá duração aproximada de 60 minutos e será realizada, preferencialmente, na residência do informante. Em um primeiro momento, será realizada uma entrevista-piloto, que será analisada detalhadamente pelo grupo envolvido na pesquisa, a fim de que se reformulem, se necessário, os parâmetros para as próximas entrevistas. Será realizado um total de 36 entrevistas, com informantes que sempre residiram em Bagé. Estes serão divididos segundo as categorias de sexo (feminino e masculino), idade (de 25 a 50 anos e mais de 50 anos) e escolaridade (de 1 a 4 anos de escolaridade; de 5 a 8 anos de escolaridade; e de 9 a 11 anos de escolaridade), como foi mencionado anteriormente. Trata-se, portanto, de uma amostra aleatória estratificada. Para esta etapa, serão utilizados um gravador digital (como o Zomm H4) e um microfone de lapela.

Fazem-se necessárias algumas observações sobre a metodologia empregada para a realização das entrevistas. O objetivo a ser atingido com tais entrevistas é a coleta de dados representativos do vernáculo dos informantes. Segundo Labov (2008, p. 244), o vernáculo é “o estilo em que se presta o mínimo de atenção ao monitoramento da fala. A observação do vernáculo nos oferece os dados mais sistemáticos para a análise da estrutura linguística”. Isto significa que se buscará uma fala em que o informante não preste atenção ao modo como está falando e se concentre no que está falando, como se estivesse em um evento que pertencesse ao seu dia a dia de comunicação.

Neste momento, instaura-se o que Labov chamou de *paradoxo do observador*: “o objetivo da pesquisa linguística na comunidade deve ser descobrir como as pessoas falam quando não estão sendo sistematicamente observadas – no entanto, só podemos obter tais dados por meio da observação sistemática” (Labov, 2008, p. 244). Ou seja, procura-se que os entrevistados ajam de maneira natural em uma situação que não é natural: há um microfone na sua frente, há um gravador e há um entrevistador.

Segundo Tarallo (2007, p. 21), “o pesquisador deverá tentar neutralizar a força exercida pela presença do gravador e por sua própria presença como elemento estranho à comunidade”. Autores como Labov (2008), Tarallo (2007), Oliveira e Silva (2010) e Tagliamonte (2006) mostram como fazer isto e, assim, minimizar os efeitos do paradoxo do observador. É preferível utilizar um microfone de lapela, pois este fica longe dos olhos do informante, mas, ao tempo tempo, é capaz de captar a fala com uma boa qualidade. Deve-se evitar, ao longo da entrevista, a menção à palavra “língua”, uma vez que isto fará com que o informante passe a prestar atenção ao modo como está falando. Como nos ensina Tarallo (2007, p. 23), “a narrativa de experiência pessoal é a mina de ouro que o pesquisador-sociolinguísta procura”. Ao relatar eventos nos quais o informante esteve envolvido, este se concentra tanto no que está contando que não se preocupa com o modo como fala. O fato de a entrevista ser realizada na própria residência do informante também tem o objetivo de deixá-lo mais à vontade. Uma coleta de dados realizada em um laboratório terá a vantagem da ausência de ruídos que podem atrapalhar a qualidade sonora da gravação. Entretanto, nestas condições, dificilmente se obterá do informante sua forma vernacular de se expressar.

O preenchimento da ficha social, que antecede a entrevista propriamente dita, terá dois objetivos. O primeiro é a obtenção de informações relevantes sobre o

informante, como a confirmação de sua idade e de sua escolaridade, a coleta de dados sobre sua classe sócio-econômica (mesmo que esta não seja uma variável controlada na pesquisa), a confirmação de que sempre morou em Bagé. Neste momento, serão obtidas, também, informações sobre, por exemplo, as atividades profissionais e de lazer do informante e o tipo de relações sociais que estabelece com a comunidade. Estas informações atenderão ao segundo objetivo do preenchimento da ficha social: auxiliar na elaboração de um questionário que conduzirá a entrevista a ser realizada posteriormente.

Tal questionário não deverá ser seguido rigidamente pelo entrevistador, porque isto trará obstáculos para a naturalidade que se busca nas entrevistas. Ele servirá apenas como um guia para o entrevistador, de maneira que, ao perceber que um assunto está se esgotando, ele tenha um outro tópico a introduzir na conversa. Como afirma Oliveira e Silva (2010, p. 125), “esta entrevista, longe de ser um questionário, deve se constituir de uma conversa, a mais informal possível, apesar das circunstâncias adversas, posto que o que se quer é a fala casual, habitual, dos falantes”.

Os manuais de sociolinguística ensinam que perguntas cuja resposta possa ser “sim” ou “não” devem ser evitadas, uma vez que tais respostas não trarão os dados necessários que são buscados. Tagliamonte (2006, p. 41) esclarece que este tipo de pergunta é, muitas vezes, inevitável, mas basta, logo em seguida, formular uma pergunta que exija uma resposta mais elaborada, como nesta sequência: “Você já foi acusado por alguma coisa que não fez? *O que aconteceu?*”. Esta autora traz várias sugestões de tópicos e de perguntas para a realização de uma entrevista sociolinguística.

Ao mesmo tempo em que se realizam as entrevistas, será formulado um manual para a transcrição dos dados. Esse manual será impresso. Conforme Paiva (2010, p. 135), “o objetivo básico de uma transcrição é transpor o discurso falado, da forma mais fiel possível, para registros gráficos mais permanentes”. Isto requer uma série de tomadas de decisões. A observação de manuais de transcrição de outros bancos de dados nos auxiliará nesta tarefa.

A etapa seguinte será a de transcrição das entrevistas, com base no manual previamente elaborado. A transcrição de cada entrevista resultará em um caderno. Em seguida, serão revisadas as transcrições, e os 36 cadernos com as entrevistas serão impressos.

Na etapa final do projeto, será planejada a disponibilização dos dados à comunidade, que poderá ser feita através da criação de uma página na internet contendo links para o acesso às entrevistas.

3 Resultados esperados

A constituição do banco de dados de língua falada de Bagé permitirá que se realizem pesquisas a fim de descrever as variedades encontradas nesta cidade. Além da descrição, os dados possibilitarão a verificação de hipóteses sobre os fenômenos linguísticos pesquisados, contemplando, desta maneira, uma análise explicativa da língua.

Os resultados das pesquisas realizadas serão divulgados por meio da participação dos pesquisadores em eventos acadêmicos e da produção de artigos. Desta forma, serão conhecidas pelo meio acadêmico as características das variedades linguísticas encontradas em Bagé, tanto no que diz respeito aos aspectos que as

diferenciam das variedades encontradas em outras regiões, quanto no que diz respeito aos aspectos que há em comum com outras variedades.

A participação em projetos de extensão, que envolvam, por exemplo, professores da rede básica de ensino, permitirá que os conhecimentos construídos a partir deste bando de dados sejam divulgados para a comunidade em geral, de maneira a criar um espaço propício para a discussão sobre a variação linguística. Tal iniciativa possibilitará que se crie um ambiente em que sejam (re)conhecidas e, conseqüentemente, respeitadas as variedades linguísticas características da cidade de Bagé. Cabe, aqui, mencionar trabalhos que discutem sobre essa relação entre os estudos sociolinguísticos e o ensino, como Bortoni-Ricardo (2004), Gorski e Coelho (2006) e Bagno (2007), entre outros.

Considerações finais

Este trabalho teve por finalidade apresentar o projeto “Banco de Dados de Língua Falada de Bagé”, cujo objetivo é a formação de um banco de dados com entrevistas gravadas. Tal banco de dados permitirá a realização de pesquisas sociolinguísticas, de maneira que se conheçam as variedades linguísticas presentes na região.

Os resultados obtidos a partir de tais pesquisas serão um instrumento importante para a ampliação das discussões sobre variação linguística, preconceito linguístico, políticas linguísticas, dentre outros assuntos.

Para finalizar, destacamos a participação neste projeto de alunos de graduação do Curso de Licenciatura em Letras, da Universidade Federal do Pampa. Nossa expectativa é a de que sua atuação neste projeto de pesquisa lhes permita, em um primeiro momento, participar ativamente de debates teóricos sobre um tema bastante atual em linguística, que é o da variação linguística. Isto lhes possibilitará conhecer os pressupostos teóricos e metodológicos da sociolinguística quantitativa, de maneira a se capacitarem para analisar criticamente pesquisas na área e, logo após, realizá-las.

Referências bibliográficas

BAGNO, Marcos. *Nada na língua é por acaso*. São Paulo: Parábola, 2007.

BELINE, Ronald. A variação linguística. In: FIORIN, José Luiz (org.). *Introdução à linguística*. v. I. São Paulo: Contexto, 2003. p. 121-140.

BORTONI-RICARDO, Stella M. *Educação em língua materna*. São Paulo: Parábola, 2004.

GORSKI, Edair Maria; COELHO, Izete L. (orgs.). *Sociolinguística e ensino*. Florianópolis: Editora da UFSC, 2006.

LABOV, William. *Padrões sociolinguísticos*. São Paulo: Parábola, 2008.

NARO, Anthony J.; SCHERRE, Maria Marta P. Variação linguística, expressividade e tradição gramatical. In: GORSKI, Edair Maria; COELHO, Izete L. (orgs.). *Sociolinguística e ensino*. Florianópolis: Editora da UFSC, 2006. p. 235-245.

OLIVEIRA E SILVA, Giselle M. de. Coleta de dados. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (orgs.). *Introdução à sociolinguística*. São Paulo: Contexto, 2010. p. 117-133.

PAIVA, Maria da Conceição de. Transcrição de dados linguísticos. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (orgs.). *Introdução à sociolinguística*. São Paulo: Contexto, 2010. p. 135-146.

SCHERRE, Maria Marta P. Sobre a influência de variáveis sociais na concordância nominal. In: OLIVEIRA E SILVA, Giselle Machline de; SCHERRE, Maria Marta Pereira. *Padrões sociolingüísticos*. Rio de Janeiro: Tempo Brasileiro, 1996. p. 239-264.

SCHERRE, Maria Marta P. *Doa-se lindos filhotes de poodle*. São Paulo: Parábola, 2005.

TAGLIAMONTE, Sali. *Analysing sociolinguistic variation*. Cambridge: CUP, 2006.

TARALLO, Fernando. *A pesquisa sociolingüística*. São Paulo: Ática, 2007.