

Coleta de dados: dificuldades e desafios

Elizete Maria de Souza

UFMG

Introdução

A ligação entre linguagem e sociedade é indiscutível. Antes mesmo do surgimento da sociolinguística, a tradição de relacionar linguagem e sociedade já fazia parte da reflexão de vários autores do século XX. Estruturalistas ou não, a grande maioria dos lingüistas da época, como Antoine Meillet, Mikhail Bakhtin, Marcel Cohen, Émile Bienveniste e Roman Jakobson já concebiam a linguagem como um fato social. Existia, porém, uma dicotomia entre o caráter homogêneo da língua em oposição ao seu caráter heterogêneo. Os estruturalistas, de um lado, defendiam o caráter homogêneo e estrutural das línguas, as variações eram concebidas como um desarranjo, um verdadeiro ‘caos’ lingüístico. Por outro lado, os lingüistas da corrente não-estruturalista acreditavam que a língua era verdadeiramente um sistema heterogêneo, regulado, porém, por um conjunto de regras do qual a variação é parte inerente (Weinreich, Labov; Herzog, 1968), o que explicava a diversidade e, ao mesmo tempo, justificava a sistematicidade das línguas.

A sociolinguística, entretanto, começou a ganhar espaço como disciplina autônoma por volta da década de 60, quando Willian Labov e outros pesquisadores definiram as dimensões da nova área. Para Labov, o objeto da sociolinguística era a diversidade lingüística, natural e inerente a todas as línguas. Mas faltava demonstrar que existia uma covariação sistemática das variações lingüísticas e sociais. Compartilhando dessa mesma idéia, vários estudiosos vão compor a tradição dos estudos clássicos sobre variação, entre eles: John Gumperz, Einar Haugen, Dell Hymes, John Fisher e Pedro Rona.

Em 1964, o termo Sociolinguística ganhou rótulo próprio e logo em seguida foi denominado de sociolinguística variacionista ou laboviana. Com os estudos de Labov, veio à tona a necessidade de se estudar não só a interferência dos fatores internos (lingüísticos), mas também os fatores externos (extralingüísticos) para a compreensão da diversidade lingüística. O contexto social, a idade, a escolaridade, o

gênero dos informantes, enfim, a identidade social dos falantes e suas atitudes lingüísticas passaram a ser vistos como fatores que interferem diretamente na produção lingüística de uma comunidade de fala¹, não sendo possível compreender as variações e as mudanças lingüísticas sem o estudo dessas variáveis. Esse conjunto de fatores determina o domínio de uma língua particular, que varia diatópica ou diastraticamente.

2. A língua falada como objeto de estudo da sociolingüística

Estudar a língua falada sempre foi um trabalho árduo que requer do pesquisador uma série de habilidades. Penetrar em uma comunidade lingüística sem caracterizar uma ação intromissiva não é tarefa das mais fáceis.² No entanto, não há como estudar a língua falada sem a matéria prima essencial: o vernáculo. Em geral, o acesso aos informantes, não apenas nos grandes centros urbanos, como também em pequenos lugarejos não é fácil. Uma boa forma de minimizar esse problema é formar uma rede de informantes dispostos a colaborar, de forma que o documentador possa coletar ‘bons dados’.³ Outra dificuldade a ser resolvida está relacionada à qualidade das gravações; se, por um lado, o mercado oferece equipamentos de alta precisão e de fácil manuseio, por outro lado, nem sempre o documentador encontra um ambiente favorável à gravação das entrevistas. Por isso mesmo, o objetivo principal desse trabalho é divulgar algumas orientações imprescindíveis ao documentador que vai a campo.

Conforme Oliveira e Silva (2003:117), “a língua falada pouco se presta à experimentação”; ao lingüista, portanto, cabe observar os diferentes fenômenos lingüísticos em variação ou em processo de mudança, buscando sistematizar os estudos para melhor compreender a linguagem humana. Uma boa forma de começar a sistematizar um trabalho é definir o tipo de estudo a ser realizado..

¹ Segundo Gregory Guy (2001), uma comunidade de fala é formada por falantes que “compartilham traços lingüísticos que distinguem seu grupo de outros, comunicam-se relativamente mais entre si do que com outros e compartilham normas e atitudes diante do uso da linguagem.”

² Labov (1972) discute essa questão que ele denomina de “O Paradoxo do Observador”.

³ Na perspectiva laboviana, coletar ‘bons dados’ significa conseguir gravar uma entrevista que se aproxime, ao máximo, da fala espontânea.

Há, basicamente, dois tipos de estudos comumente utilizados em pesquisas sociolingüísticas: um voltado para o estudo da comunidade de fala - estudo de tendência, e outro com o foco no indivíduo, denominado estudo de painel (Labov, 1994).

O estudo do tipo ‘tendência’ se baseia na comparação de amostras aleatórias da mesma comunidade de fala, estratificadas com base nos mesmos parâmetros sociais, em dois momentos de tempo, enquanto o estudo de ‘painel’ consiste no recontato e na obtenção de uma amostra de fala dos mesmos indivíduos em dois pontos separados por um lapso de tempo.

Geralmente, considera-se que o espaço de uma geração (cerca de 20 anos) é suficiente para fornecer subsídios acerca da estabilidade ou mudança no comportamento lingüístico do indivíduo e da comunidade. A vantagem desses dois tipos de estudos é permitir o confronto de duas sincronias do mesmo foco geográfico. Esses tipos de estudos podem ser realizados com base no tempo aparente ou no tempo real.

Labov (op. cit.) afirma que “o primeiro método e o mais simples para se estudar a mudança lingüística em progresso é traçar a mudança no tempo aparente, isto é, a distribuição das variáveis lingüísticas por faixas etárias”, com os mais jovens introduzindo novas alternantes que, gradativamente, substituirão aquelas que caracterizam o desempenho lingüístico dos falantes de faixas etárias mais avançadas. O segundo tipo de estudo, isto é, com base no tempo real, leva em conta intervalos de tempo entre duas sincronias, sendo

um recurso essencial não só para identificar o aparecimento ou morte de uma determinada variante lingüística como também para verificar a regularidade na ação dos princípios que regem a variação e subjagam à implementação da mudança. Essas duas formas de estudo se complementam, fornecendo evidências mais seguras acerca do estatuto dos padrões de variação em um dado recorte sincrônico” (Paiva e Duarte, 2003, p.181-182).

Nos termos de Labov (Op. cit.), “a combinação de evidências no tempo aparente e no tempo real é o método básico para o estudo da mudança em progresso”, pois a conjugação desses dois tipos de estudos permite a comparação de um fenômeno variável em diversos pontos do tempo. O estudo da língua falada na perspectiva laboviana exige,

portanto, a definição do tipo de estudo que melhor se presta à investigação desejada, assim como o conhecimento prévio dos construtos teóricos que fundamentam cada um para que os *corpora* possam ser definidos.

3. Estratégias de entrevistas

Basicamente existem três modelos de entrevistas usados em coleta de dados de fala, a saber: interação livre, entrevista sociolinguística e entrevista subjetiva.

A interação livre se caracteriza pelo diálogo entre dois ou mais interlocutores e, geralmente, é utilizada para o estudo de formas de tratamento, para o estudo dos turnos de fala, etc. Esse tipo de entrevista constitui um recurso primordial para o estudo dos fenômenos citados. A composição de um *corpus* a partir da interação livre é bastante complicada, pois normalmente há muitas sobreposições de fala, o que dificulta a transcrição e o levantamento dos dados. Outro modelo de entrevista utilizado na sociolinguística é a entrevista subjetiva, também chamada de teste subjetivo. Normalmente, os testes subjetivos são elaborados da seguinte forma: o documentador delimita o assunto que será abordado, em seguida ele começa a falar sobre aquele assunto e o informante dá continuidade ao assunto proposto. É comum o pesquisador lançar mão de um teste subjetivo para coletar dados para o estudo de fenômenos que envolvem variações sintáticas, pois fenômenos de natureza sintática são mais difíceis de serem capturados durante a fala espontânea. Há, por fim, um modelo de entrevista tradicionalmente usado para a coleta de dados de fala: a entrevista sociolinguística, ou seja, a entrevista do tipo DID (diálogo entre o documentador e o informante), que é realizada a partir de um roteiro de perguntas que orientam a conversa, permitindo ao informante falar naturalmente sobre assuntos do dia-a-dia.

3.1. A narrativa pessoal como estratégia de entrevista

A narrativa pessoal é o estilo de entrevista que melhor permite emergir a fala espontânea. Ao narrar, o indivíduo volta sua atenção aos fatos e mais facilmente se desvencilha da situação de entrevista, deixando emergir o vernáculo. Durante a narrativa pessoal, a maioria dos informantes discorre naturalmente sobre suas experiências individuais, o que permite a coleta de material precioso para o estudo da língua falada.

Todavia, por maior que seja o grau de intimidade entre o documentador e o informante, a situação de entrevista pode interferir no grau de espontaneidade do informante. Por isso é sempre bom ficar atento à maneira de conduzir uma entrevista.

3.1.1 Conduzindo uma entrevista sociolinguística

Evitar o uso da palavra ‘entrevista’ é um bom começo para dar início a uma conversa informal. Ao estabelecer um diálogo do tipo DID, procure usar termos que sugiram uma situação informal entre os interlocutores, como por exemplo: bate-papo ou simplesmente conversa. O simples fato de não mencionar a palavra entrevista causa um efeito satisfatório na espontaneidade dos informantes.

Outro aspecto importante a ser considerado é a natureza do fenômeno linguístico investigado. A composição de *corpora* para estudos fonológicos, semânticos e lexicais é menos complicada do que a seleção de um *corpus* para estudo de fenômenos sintáticos, pois dados de natureza sintática não costumam aparecer repetidas vezes em uma só entrevista do tipo DID. Isso se deve ao fato de o falante prestar atenção à sua própria fala durante a entrevista; é muito comum entre os informantes a preocupação com o uso do português “correto”, o que dificulta a coleta de dados para estudos de fenômenos de ordem sintática. Por isso, o documentador atento deve ser capaz de conduzir uma entrevista deixando seu informante o mais à vontade possível, sem fazer perguntas que possam deixá-lo desconcertado ou apreensivo com a conversa. Perguntas que façam suscitar assuntos de caráter íntimo podem causar um efeito desastroso na coleta de dados, enquanto a realização de perguntas que permitem ao informante um envolvimento sentimental com o assunto tem se mostrado ideal na coleta de dados de fala. Para obter um diálogo que revele a língua falada no cotidiano, o documentador deve lançar mão de um roteiro de entrevista com orientações muito simples cuja algumas se destacam:

- Pedir ao informante que conte sobre sua infância;
- Pedir ao informante que fale sobre suas conquistas e frustrações;
- Pedir ao informante que fale sobre passeios, momentos descontraídos e de grandes emoções;

- Pedir ao informante que conte sobre situações de perigo, experiências de quase morte consigo próprio ou de morte de pessoas queridas, etc.

Em geral, quando o falante fala de situações de perda de pessoas queridas, doenças na família, momentos de grandes alegrias, ou de indignação, o vernáculo emerge naturalmente. Portanto, é imprescindível que o documentador conduza a conversa de forma que o informante se envolva emocionalmente com o assunto e passe a não se preocupar com a própria fala.

4. Algumas dificuldades e desafios a serem enfrentados na coleta de dados

Alguns problemas são bastante comuns na pesquisa de campo. Em geral, a coleta de dados é um trabalho que apresenta dificuldades naturais, a saber:

- A escolha e seleção dos informantes
- A composição da amostra
- A coleta de “bons dados”
- A seleção dos dados
- A transcrição dos dados

Além dessas dificuldades, o pesquisador se depara ainda com um problema típico da atualidade:

- O fator tempo

4.1. A escolha e a seleção dos informantes para a composição de uma amostra de fala da cidade de Belo Horizonte

Chega a ser um paradoxo dizer que a escolha e a seleção de informantes para a composição de uma amostra de uma comunidade de fala como a de Belo Horizonte pode constituir um problema para os pesquisadores da Teoria da Variação.

Considerando que Belo Horizonte possui uma população com mais de 2 milhões de habitantes, deveria ser extremamente simples constituir um *corpus* a partir de uma amostra de 36 informantes, por exemplo. Mas a realidade é outra, a escolha e a seleção dos informantes é tarefa delicada, pois o pesquisador precisa selecionar seus informantes de acordo com um perfil pré-definido para evitar os possíveis problemas de enviesamento dos dados.

Na constituição de uma amostra de Belo Horizonte para o estudo da indeterminação do sujeito através do pronome ‘eles’, por exemplo, foram respeitados vários critérios para que o falante pudesse fazer parte da amostra. O primeiro critério define que somente pessoas nascidas na capital mineira, ou nela residente por aproximadamente 2/3 de suas vidas, poderiam ser selecionados como informantes. Esse critério foi responsável por excluir um número significativo de falantes, pois principalmente entre as pessoas mais velhas, uma grande parte delas não é natural de Belo Horizonte, são pessoas que vieram de várias cidades do interior de Minas Gerais e não moram na capital mineira há cerca de 2/3 de suas vidas. Associado a outros fatores externos, que geralmente são foco de atenção nos estudos variacionistas, tais como, sexo, escolaridade, faixa etária, a escolha e a seleção de um informante é, às vezes, bastante complicada, mas nada que não possa ser resolvido.

4.2. A coleta e a seleção dos dados

A coleta de bons dados depende de alguns procedimentos básicos. A forma de conduzir a entrevista é muito importante, é imprescindível que o pesquisador saiba ouvir atentamente o seu informante, sem interrompê-lo desnecessariamente, pois o que se deseja é que o informante fale espontaneamente, sem se sentir controlado, observado. O rigor científico na coleta de dados também é fundamental para que os resultados da pesquisa sejam expressivos, somente dados coletados de acordo com a metodologia da pesquisa podem constituir um *corpus* confiável. O local das entrevistas também pode interferir positiva ou negativamente na coleta dos dados. A princípio, as gravações deveriam ser feitas em um local neutro, longe de ruídos ou de quaisquer outras interferências indesejadas, mas geralmente o informante prefere que a entrevista seja gravada em sua própria casa. Deixe que o informante decida pelo lugar onde ele se sinta mais à vontade, mas é prudente que a entrevista seja realizada em um dos

cômodos cujo ambiente seja mais tranquilo. Outro aspecto importante diz respeito à escolha e manuseio dos equipamentos tecnológicos. O uso de um bom gravador auxilia na qualidade do material coletado. Há no mercado vários tipos de gravadores, a maioria digital, o que possibilita a criação de um acervo digitalizado.

4.3. A transcrição dos dados

A transcrição dos dados demanda tempo e exige cuidados especiais. Primeiramente, deve-se escolher uma ficha de transcrição que possibilite a criação de um banco de dados que possa ser facilmente consultado. Em seguida, deve-se decidir por transcrever as entrevistas na íntegra ou apenas parcialmente. É aconselhável que as entrevistas sejam transcritas na íntegra, mas, em muitos casos, o pesquisador decide por transcrever apenas os dados que lhe interessa. Neste caso, o pesquisador fica limitado a poder consultar apenas os dados transcritos, pois se torna difícil recuperar os contextos lingüísticos quando o material é transcrito apenas parcialmente. Por outro lado, uma entrevista transcrita integralmente permite a verificação dos dados, em seus respectivos contextos, a qualquer momento, o que compensa o trabalho de transcrição integral.

4.4. O fator tempo

Por fim, considerando o ritmo acelerado da vida atual, o fator tempo pode acabar se tornando um problema na coleta de dados, pois o pesquisador precisa saber lidar com os seguintes dilemas: (a) a falta de disponibilidade de tempo dos informantes X o tempo de duração das entrevistas, (b) a disponibilidade de tempo do informante X a disponibilidade de tempo do pesquisador.

A princípio, parece difícil lidar com o fator tempo, mas essa dificuldade pode ser minimizada a partir uma interação direta entre pesquisador e informante. O informante precisa ser informado a respeito do tempo de duração da entrevista, que deve ser de 40 à 60 minutos quando se trata de uma entrevista sociolingüística. Outra dificuldade que precisa ser superada é a adequação entre a disponibilidade de tempo do informante e do pesquisador, nem sempre o informante tem disponibilidade de tempo quando o documentador está livre e vice-versa

. O ideal é que se chegue a um consenso de horário que seja bom para os dois lados, pois assim aumentam a possibilidade de uma boa conversa.

Considerações Finais

Nos estudos sobre a língua falada, a coleta de dados é decisiva para a realização de um bom trabalho. Como vimos nesse artigo, não há como estudar a língua falada sem a matéria prima essencial – o vernáculo, o que exige do pesquisador uma série de habilidades. Enfrentar as dificuldades naturais que envolvem o trabalho de campo, como a escolha e a seleção dos informantes, a realização das entrevistas, a transcrição e seleção dos dados, é o primeiro desafio de uma pesquisa em sociolingüística variacionista.

Cabe destacar, que a tarefa de coletar dados é um trabalho que exige do pesquisador rigor científico, sensibilidade e perseverança. Rigor científico para que a metodologia da pesquisa seja seguida adequadamente e os dados coletados sejam confiáveis, sensibilidade para conduzir as entrevistas, de forma que a conversa possa emergir e perseverança para que a etapa de coleta de dados não se torne um empecilho para a realização da pesquisa.

Referências

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

LABOV, William. *Principles of Linguistic Change*. v. 1: Internal Factors. Blackwell: Oxford e Cambridge USA, 1994.

LABOV, William. *Principles of Linguistic Change*. v. 2: Social Factors. Blackwell: Oxford e Cambridge USA, 2001.

SILVA, Giselle Macheline de Oliveira. Coleta de dados. In: Mollica e Braga (Org.). *Introdução à Sociolingüística: o tratamento da variação*. São Paulo: Contexto, 2003.

PAIVA, Maria da Conceição de; DUARTE, Maria Eugenia Lamoglia (Org.). *Mudança lingüística em tempo real*. Rio de Janeiro: FAPERJ, 2003.

TARALLO, Fernando. *A Pesquisa Sociolingüística*. 7. ed. São Paulo: Ática, 2003.

WEINREICH, Uriel.; LABOV, William; HERZOV, Marvin. Empirical Foundations for a Theory of Language Change. In: LEHMANN, W.; MALKIEL, Y. (Ed.). *Directions for Historical Linguistics*. Austin: University of Texas Press, 1968.