

O BANCO DE DADOS FALA-NATAL: UMA AGENDA DE TRABALHO

Maria Alice Tavares
Marco Antonio Martins

INTRODUÇÃO

Estudos feitos na perspectiva da Sociolinguística Variacionista vêm fomentando a ampliação do conhecimento sobre o português brasileiro, desde a década de 1970, através da descrição e da análise de fenômenos variáveis nos âmbitos fonológico, morfológico, sintático, semântico e discursivo. Contudo, há estados da federação em que tais pesquisas são ainda incipientes ou mesmo inexistentes. É o caso do Rio Grande do Norte (RN), que não conta com um banco de dados de fala com as características necessárias para a pesquisa sociolinguística. Para suprir essa lacuna, propusemo-nos a organizar um *corpus* de fala denominado Banco de Dados da Fala do Rio Grande do Norte (FALA-RN), que contará com amostras representativas de diferentes comunidades de fala norte-rio-grandenses.

O marco inicial da organização do Banco de Dados FALA-RN é a constituição do Banco de Dados FALA-Natal, que congrega entrevistas sociolinguísticas feitas com membros da comunidade de fala e, na medida do possível, de diferentes comunidades de prática, do município de Natal, que é a capital e o maior centro urbano do estado potiguar. Posteriormente, serão coletadas entrevistas sociolinguísticas em comunidades de fala representativas do interior do RN.

Apresentamos, neste texto, um balanço geral da constituição do banco de dados FALA-Natal para o qual as entrevistas estão em fase final de coleta.

Apresentamos ainda, as dificuldades práticas com as quais temos nos deparado na constituição do *corpus*, assim como as soluções encontradas.

1. O BANCO DE DADOS FALA-NATAL

Os informantes do Banco de Dados FALA-Natal são socialmente estratificados de modo similar aos informantes de bancos de dados já existentes no país, a exemplo do Programa de Estudos sobre o Uso da Língua (PEUL) e do projeto Variação Linguística na Região Sul do Brasil (VARISUL). Além disso, buscamos considerar aspectos relacionados à identificação de comunidades de prática em que estejam engajados os informantes.

Inicialmente, o Banco de Dados FALA-Natal será composto por 48 entrevistas sociolinguísticas com cerca de 60 minutos de duração. Essas entrevistas serão distribuídas, em termos de estratificação social, quanto ao sexo (24 informantes de sexo feminino e 24 informantes de sexo masculino); idade (12 informantes de 8 a 12 anos, 12 informantes de 15 a 21 anos, 12 informantes de 25 a 50 anos e 12 informantes de mais de 50 anos) e nível de escolaridade (12 informantes com ensino fundamental I completo, 12 informantes com ensino fundamental II completo e 12 informantes com ensino médio completo, além de 12 informantes cursando o ensino fundamental I – os indivíduos de 8 a 12 anos). Serão entrevistados informantes de diferentes bairros das quatro zonas de Natal.

Essa constituição inicial do Banco de Dados busca similaridade com os de dados sociolinguísticos já existentes no Brasil, com o objetivo de facilitar a execução de estudos sociolinguísticos comparativos. Entre as sugestões para futuras investigações e desdobramentos, frequentemente apontadas em pesquisas concluídas na perspectiva da Sociolinguística Variacionista, destaca-se a possibilidade de realização de análises comparativas dos resultados obtidos para o fenômeno estudado com os de pesquisas efetuadas em outras regiões do Brasil, que tenham o mesmo ou semelhante objeto de estudo.

No entanto, essas análises sociolinguísticas comparativas de grande extensão são pouco realizadas, talvez por conta das dificuldades que sua execução implica. Conforme Guy (1999), muito é perdido ao se deixar de empreender comparações entre resultados obtidos para fenômenos variáveis dentro de uma mesma língua ou mesmo interlinguísticas, pois um dos objetivos centrais da Sociolinguística Variacionista é o estabelecimento de princípios gerais e, na medida do possível universais, que estariam subjacentes à variação e à mudança, e seriam válidos para todas ou grande parte das comunidades de fala.

Segundo Tagliamonte (2003, p. 729), a “comparação sempre esteve na raiz da sociolinguística”, permitindo a construção de generalizações através do

cotejamento de amostras de dados em tempo real e em tempo aparente. Todavia, análises comparativas interdialetais apenas recentemente vêm recebendo um destaque crescente no cenário mundial, o que tem levado à proposição de generalizações e mesmo de princípios universais de variação e mudança. Como afirma Chambers (2004, p. 128), “à medida que a sociolinguística se torna menos restrita a eventos locais, se torna comparativa e, à medida que o aspecto comparativo ganha peso, generalizações interlinguísticas não apenas se tornam possíveis, mas inevitáveis”.

Assim, para fazer avançar ainda mais a sociolinguística no Brasil, é de fundamental importância a realização de estudos comparativos que visem buscar semelhanças e diferenças quanto ao comportamento de uma mesma variável linguística, em diferentes dialetos brasileiros. Quais os ganhos que adviriam desse tipo de estudo?

Entre muitas vantagens, a principal parece ser a possibilidade de observar se as restrições linguísticas e sociais à variação e à mudança para um dado fenômeno são as mesmas em todas as regiões do Brasil e, se não, em que diferem, aventando explicações que abranjam resultados provindos de diversas comunidades de fala. A partir de tais observações e explicações, podemos chegar a estabelecer generalizações, base dos princípios gerais tão procurados pela Sociolinguística Variacionista para a construção de sua teoria. Contudo, não é apenas sobre o que é comum às comunidades de fala que recai o interesse de um estudo comparativo: a comparação pode auxiliar na descoberta de especificidades e de idiosincrasias em comunidades particulares, revelando o jogo local *versus* universal típico da língua.

Uma vez finalizadas e armazenadas, as entrevistas integrantes do Banco de Dados FALA-Natal poderão servir de *corpus* para pesquisas que objetivem: i) a descrição e a análise da fala de Natal; ii) a comparação com outros dialetos brasileiros, com o intuito de descrever o português brasileiro de modo mais abrangente e detalhado, e de observar as diferenças e semelhanças interdialetais; iii) a comparação com outras vertentes do português; iv) a testagem de teorias linguísticas.

2. SOBRE AS DIFICULDADES ENCONTRADAS E O ENCAMINHAMENTO DE SOLUÇÕES

Na constituição do banco de dados FALA-Natal, que ainda está em desenvolvimento (em fase final de realização das entrevistas), temos nos defrontado com uma série de questões para as quais temos buscado soluções. Entre essas questões, apontamos as seguintes:

- **Representatividade da amostra:** De acordo com o Censo de 2010 do IBGE, a capital norte-rio-grandense tem 803.739 habitantes. Se considerarmos uma amostragem na condição metodológica ideal, aplicando o corte de 0,5% da população, o banco de dados FALA-Natal deveria contar com 4.019 entrevistas/falantes.

Em condições reais, o desenvolvimento de um banco de dados com esse número de entrevistas demandaria anos de realização. Com o significativo crescimento da população, se adotássemos essa condição para o desenvolvimento do banco, quando a última entrevista fosse realizada, a comunidade, com certeza, já não seria a mesma. Além disso, a quantidade de informantes também depende de financiamento e de quanto tempo se dispõe para a organização do banco de dados, fatores que, em geral, impedem a coleta de um grande número de entrevistas.

De qualquer forma, um número menor de entrevistas pode ser representativo de tendências gerais da comunidade. Segundo Sankoff (1988, apud TAGLIAMONTE, 2006, p. 23), é necessário “não que a amostra seja uma versão em miniatura da população, mas apenas que tenhamos a possibilidade de fazer inferências sobre a população com base na amostra”. Cada banco de dados deve ter um mínimo de representatividade com base em idade, sexo, classe social e/ou nível de educação, o que assegura que a diversidade linguística da comunidade de fala esteja representada na amostra.

Lembramos que a maior coleta de entrevistas sociolinguísticas já feita foi dirigida por Shuy et al. (1968), tendo sido gravadas 702 entrevistas em Detroit, nos Estados Unidos. No entanto, as análises mais detalhadas desse *corpus* utilizaram apenas 48 dessas entrevistas, com os informantes distribuídos simetricamente em quatro classes sociais, em um total de 12 informantes por classe (cf. WOLFRAM, 1969; LABOV, 2008; CHAMBERS, 1995; TAGLIAMONTE, 2006).

No caso do Brasil, os bancos de dados costumam ter de 2 a 3 informantes por célula social, o que tende a ser suficiente para a obtenção dos padrões gerais de variação de uma comunidade de fala no que diz respeito aos diversos fenômenos variáveis. Quanto ao Banco de Dados FALA-Natal, caso algumas características de uso linguístico variável chamem, por alguma razão, a atenção no conjunto das 48 entrevistas sociolinguísticas iniciais, outras entrevistas poderão ser realizadas – com os mesmos ou outros informantes – no sentido de possibilitar uma análise mais refinada desses usos.

- **Dificuldade de localização de informantes com certos traços socioeconômicos:** A esse respeito, nossa maior dificuldade atualmente está na localização

de indivíduos com mais de 50 anos, nascidos em Natal e com pais natalenses. Boa parte dos indivíduos que temos contatado são oriundos do interior do Rio Grande do Norte. Para tentar resolver esse problema, estamos utilizando a metodologia de “bola de neve”, pedindo a cada informante com mais de 50 anos que nos indique amigos e/ou conhecidos com as mesmas características sociais;

- **Necessidade de maior diferenciação de faixas etárias para testar hipóteses relativas à mudança linguística:** No Banco de Dados FALA-Natal, estamos considerando quatro faixas etárias, prevendo 12 informantes de 8 a 12 anos, 12 informantes de 15 a 21 anos, 12 informantes de 25 a 50 anos e 12 informantes de mais de 50 anos. A proposta de termos informantes com menos de quinze anos é motivada pela possibilidade de descoberta, para casos de mudança linguística, de padrões caracterizados por um pico de uso na fala dos adolescentes.

Como vários estudos sociolinguísticos vêm constatando a existência do uso intenso de formas inovadoras por indivíduos em torno de dezesseis a vinte anos de idade, Labov (2001) propôs a existência de um pico de uso de formas inovadoras no período final da adolescência. Segundo Labov, é esperado que ocorra, nos processos de mudança, após o pico de uso da forma inovadora, uma retração de seu aparecimento: ela é incorporada, ainda com índices de grande frequência, à gramática dos falantes do grupo em que teve seu uso fortemente acelerado, mas passa a recorrer menos, em comparação com a fase de pico de uso. Desse modo, a mudança adquire matizes não tão radicais, e sim, uma maior gradualidade: passa a haver uma distribuição linear crescente ou decrescente entre faixas etárias adultas, agora ocupadas pelos mesmos indivíduos que levaram a forma inovadora a seu ápice. A forma poderá vir a derrotar as demais concorrentes com o passar do tempo, mas com uma menor velocidade do que a prevista, considerando-se somente seu estágio de pico de uso.

Para que seja possível a verificação, em cada fenômeno variável para o qual tenhamos indícios de mudança, dessa possibilidade de ocorrência da forma inovadora na fala adolescente, é preciso que levemos em conta faixas etárias menores. No caso do Banco de Dados FALA-Natal, estamos contando com uma faixa etária de 8 a 12 anos, adicionando assim, os pré-adolescentes ao banco de dados.

- **Validade da comparação entre análises realizadas em dados extraídos de entrevistas sociolinguísticas feitas recentemente** (o que será o caso do Banco de Dados FALA-Natal), e **dados extraídos de entrevistas sociolinguísticas**

feitas há dez ou vinte anos: Essa é outra questão que nos preocupa, afinal, na comparação de resultados provenientes de bancos de dados constituídos em diferentes períodos de tempo, sempre será necessário ter em mente que cada comunidade de fala comparada pode estar representando uma etapa diferente de variação e mudança linguística. Por exemplo, se comparássemos dados extraídos do VARSUL referentes à comunidade de fala de Florianópolis (cujo banco foi organizado ao longo da última década do século XX) com os bancos constituídos recentemente em outras comunidades de fala, seria preciso levar em conta, na análise dos resultados, que os bancos mais novos podem estar representando um período mais recente de evolução para certos fenômenos variáveis em processo de mudança em comparação com o banco de Florianópolis.

- **Observação de aspectos relacionados à questão da análise estilística pelo viés da terceira onda:** Em relação a esse último tópico, nosso objetivo inicial foi a coleta de entrevistas sociolinguísticas em uma comunidade de fala ampla para que seja possível a realização de mapeamentos de tendências gerais de variação e mudança em relação a essa comunidade.

Todavia, nossa comunidade de fala alvo abriga, naturalmente, inúmeras comunidades de prática. Com a intenção de aprofundarmos nosso conhecimento acerca das comunidades de prática em que se engajam cada um dos informantes a serem selecionados para o banco de dados, elaboramos uma ficha social a ser preenchida previamente à entrevista na qual constam, entre outras, questões que permitam a obtenção de informações a respeito das diferentes comunidades de prática em que se engaja o informante em sua vida cotidiana. Nessa ficha social, solicitamos, por exemplo, para os informantes de 15 a 21 anos, que respondam às seguintes questões: (i) Como ocupa seu tempo livre? e (ii) Participa de algum grupo (igreja/ jovens/ esporte/ clube)? Se sim, com que frequência?

Também foram propostos, nas entrevistas, tópicos que estimulassem o informante a discorrer sobre as diferentes comunidades de prática das quais faz parte. Elaboramos um roteiro para as entrevistas com sugestões de perguntas que o entrevistador poderia fazer ao entrevistado. Entre elas, estão questões do tipo: (i) Com quem você passa o tempo, além das pessoas da sua família? O que vocês fazem juntos? Que tipo de lazer vocês têm?; (ii) Você participa de algum trabalho voluntário? Como é?; (iii) Você participa de algum grupo de jovens? O que vocês fazem juntos?; (iv) Você participa de algum grupo da igreja? Como é?; (v) Você frequenta algum clube? Qual? Como é?; (vi) Algo interessante já aconteceu no clube/grupo de jovens/

grupo da igreja quando você estava? O que aconteceu? e (vii) Descreva o que você faz em um dia, desde que acorda até ir dormir.

Esse maior conhecimento sobre as comunidades de prática em que se engaja cada informante, que será obtido através das fichas sociais e das próprias entrevistas poderá ser considerado na análise dos fenômenos variáveis. Tanto as informações extraídas das fichas sociais e das entrevistas, não apenas fornecerão subsídios para uma análise mais aprofundada de cada informante, no que tange às características sociais e de prática, como também trarão indícios a respeito de quais comunidades de prática – entre as inúmeras em que se integra cada indivíduo – são mais interessantes para a realização de estudos nos moldes da terceira onda, o que pode, naturalmente, ensejar a coleta de novas entrevistas com grupos de indivíduos pertencentes a tais comunidades.

Ou seja, para a organização do Banco de Dados FALA-Natal, estamos conscientes da necessidade de contemplar não apenas pesquisas sociolinguísticas afiliadas à abordagem variacionista de Labov (um retrato amplo de comunidades de fala definidas geograficamente), mas também à abordagem etnográfica alinhada ao Milroy (um retrato local, etnográfico, de comunidades de fala definidas geograficamente) e à abordagem da identidade social alinhada à Eckert (um retrato do(s) indivíduo(s) integrante(s) de comunidades de prática, pelo viés do estilo como elemento central de constituição da *persona*).

3. UM BALANÇO DA CONSTITUIÇÃO DO CORPUS E UMA AGENDA DE TRABALHO

Com a organização do Banco de Dados FALA-RN, do qual a composição do Banco de Dados FALA-Natal representará a primeira etapa, será possível a descrição de dialetos do português brasileiro falados em um estado nordestino no qual, sob a perspectiva da variação e da mudança linguística, pouco foi feito. Para preencher essa lacuna, o Banco de Dados FALA-RN fomentará o desenvolvimento de projetos voltados para a pesquisa, o ensino e a extensão, tanto nos cursos de pós-graduação quanto nos de graduação, bem como oferecerá aos interessados em geral uma fonte de dados linguísticos contemporâneos. Estamos em fase de conclusão das entrevistas que constituirão o banco FALA-Natal. As entrevistas, em formato digital, serão tratadas e ficarão à disposição da comunidade acadêmica para pesquisas.

REFERÊNCIAS

- CHAMBERS, J. Dynamic typology and vernacular universals. In: KORTMANN, B. (Ed.). *Dialectology meets typology: dialect grammar from a cross-linguistic perspective*. Berlin: Mouton de Gruyter, 2004. p. 127-145.
- GUY, G. R. Notas do curso Sociolinguística Comparativa, ministrado de 22/02 a 05/03, na UFSC, por ocasião do XIV Instituto Linguístico da ABRALIN, 1999.
- LABOV, W. *Padrões sociolinguísticos*. Trad. M. Bagno, M. M. P. Scherre, C. R. Cardoso. São Paulo: Parábola Editorial, 2008. Título original: *Sociolinguistic Patterns*, 1972.
- _____. *Principles of linguistic change: social factors*. Oxford: Blackwell, 2001.
- TAGLIAMONTE, S. A. Comparative sociolinguistics. In: CHAMBERS, J. K.; TRUDGILL, P.; _____. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press, 2006.
- SHILLING-ESTES, N. (Eds.). *The handbook of language variation and change*. Cambridge: Blackwell, 2003. p. 729-763.
- SHUY, R.; WOLFRAM, W.; RILEY, W. *Field techniques in an urban language study*. Washington, DC: Center for Applied Linguistics, 1968.
- WOLFRAM, W. *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics, 1969.