

A LÍNGUA FALADA EM ALAGOAS: COLETA E TRANSCRIÇÃO DOS DADOS

Elyne Giselle de Santana Lima Aguiar Vitória

INTRODUÇÃO

Ao se interessar pelo estudo da língua dentro do contexto social da comunidade de fala, a Sociolinguística Variacionista não só a vê como um fator importante na identificação e na demarcação de diferenças sociais na comunidade, como também sugere um modelo que analisa o uso variável dos fenômenos e a interferência dos condicionamentos linguísticos e sociais, proporcionando descrições mais adequadas da língua em uso pelos falantes (LABOV, 2008).

Para tanto, faz-se necessário constituir amostras sincrônicas e/ou diacrônicas da língua usada em comunidades de falas heterogêneas, tendo em vista que, para a sistematização de uma regra variável, o pesquisador sociolinguista, de acordo com Campoy e Almeida (2005), Tagliamonte (2006) e Guy e Zilles (2007), precisa definir a variável dependente e as independentes, delimitar a amostra e constituir o *corpus*, transcrever, codificar e quantificar os dados e por fim, interpretar e explicar os resultados obtidos.

Neste texto, apresentamos a metodologia e os pressupostos norteadores do trabalho de campo sociolinguístico empreendido para a constituição de uma amostra sincrônica da língua falada no estado de Alagoas. Para tanto, seguimos os pressupostos teórico-metodológicos básicos, apresentados em Guy e Zilles

(2007), e focalizamos nossa discussão em quatro momentos distintos desse processo, a saber: delimitação da comunidade estudada; constituição do *corpus* e estratificação da amostra; coleta dos dados e transcrição dos dados.¹

1. COMUNIDADE PESQUISADA

Como não é possível compreender o processo de variação e de mudança linguística fora do contexto social de uma comunidade de fala, uma vez que, para a Sociolinguística Variacionista, a língua é uma forma de comportamento social – ou seja, a língua não é propriedade do indivíduo, mas da comunidade e, portanto, social – selecionamos a comunidade de fala alagoana e assumimos a definição de comunidade de fala proposta por Labov.

A comunidade de fala não é definida por nenhuma concordância marcada no uso de elementos lingüísticos, mas sim pela participação num conjunto de normas compartilhadas; estas normas podem ser observadas em tipos de comportamento avaliativo explícito e pela uniformidade de padrões abstratos de variação que são invariantes no tocante a níveis particulares de uso. (LABOV, 2008, p. 150).

Para a constituição da amostra sincrônica da fala alagoana, partimos do pressuposto de que “[...] existe um conjunto uniforme de atitudes frente à linguagem que são compartilhadas por quase todos os membros da comunidade de fala, seja no uso de uma forma estigmatizada ou prestigiada da língua em questão” (LABOV, 2008, p. 176), delimitando assim, a que tipo de comunidade de fala pertence determinado indivíduo.

Conhecido como Paraíso das Águas, o estado de Alagoas está situado à leste da região Nordeste, fazendo fronteiras com o Oceano Atlântico, divisa ao Norte e ao Noroeste com o estado de Pernambuco, ao Sul com o estado de Sergipe e ao Sudeste com o estado da Bahia. Ocupa uma área de 27.779.343 km², sendo considerado um dos menores estados do Brasil, mais extenso apenas do que Sergipe. O estado de Alagoas é formado por 102 municípios e apresenta as seguintes características: o relevo é composto por planície litorânea, planalto ao norte e depressão ao centro, tendo como ponto mais elevado a serra Santa Cruz com 844 m; a vegetação é formada por floresta tropical, mangues litorâneos e caatinga; seu clima se caracteriza por ser tropical na costa e semi-árido no interior; possui o São Francisco, o Mundaú e o Paraíba do Meio como os principais rios.

1 A coleta de dados subsidiou a pesquisa *Ter/haver existenciais na fala alagoana: variação estável ou mudança em progresso?* (VITORIO, 2012).

As cidades alagoanas mais populosas, de acordo com o Censo 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), são: Maceió (capital de Alagoas), Arapiraca, Palmeira dos Índios, Rio Largo, União dos Palmares, Penedo, São Miguel dos Campos, Coruripe, Campo Alegre e Delmiro Gouveia, sendo Maceió, Maragogi, Japaratinga, Barra de São Miguel, Piaçabuçu, Marechal Deodoro e Penedo os destinos turísticos mais procurados.

A população alagoana é composta por 3.120.494 habitantes, que segundo os dados da Pesquisa Nacional por Amostras de Domicílios (PNAD) – 2009, estão distribuídos entre 1.511.767 homens e 1.608.727 mulheres, apontando a predominância de habitantes do sexo feminino (Gráfico 1).

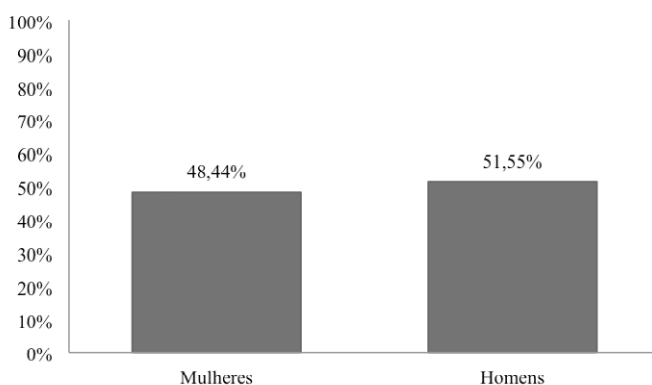


Gráfico 1 – Estratificação da população alagoana por sexo

De acordo com a estratificação por faixa etária, apresentada pela PNAD-2009, a população alagoana, em 15 grupos etários, conforme gráfico 2, apresenta percentual maior de pessoas entre o grupo de 10 a 14 anos, representando 11,1% da população, seguido do grupo de 15 a 19 anos, com um percentual de 10,7%. Já o menor grupo é formado por pessoas com idade entre 65 e 69 anos, representando 2,4% da população alagoana, seguido do grupo de 60 a 64 anos com 3% (Gráfico 2).

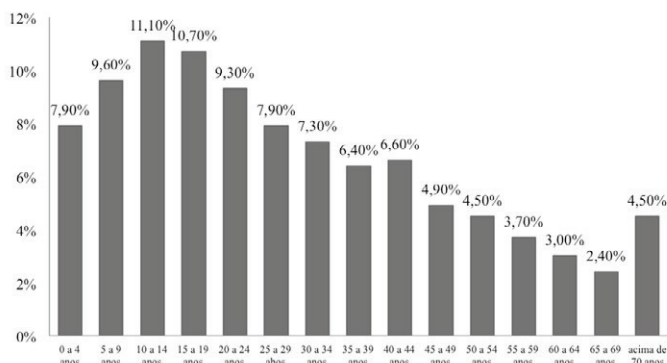


Gráfico 2 – Estratificação da população alagoana por faixa etária.

Quanto aos anos de escolarização, o PNAD – 2009, considerando apenas as pessoas de 10 anos ou mais de idade, aponta que pessoas com 4 a 7 anos de estudos constituem o maior grupo da população alagoana, representando 30,1%; já as pessoas com 15 anos ou mais de estudos representam apenas 4,1% (Gráfico 3).

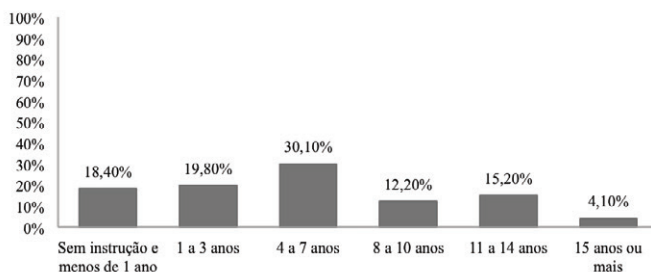


Gráfico 3 – Estratificação da população alagoana por anos de estudos.

Os dados do PNAD – 2009 também apontam que mais de 68,3% da população alagoana com 10 anos ou mais de idade não possuem o ensino fundamental completo, e que 18,4% estão sem instrução ou possuem menos de um ano de estudo. Esses dados sociodemográficos nos embasam para o dimensionamento da amostra sociolinguística.

2. ESTRATIFICAÇÃO DA AMOSTRA

Selecionada a comunidade de fala, o passo seguinte foi a estratificação da amostra a ser coletada. Para tanto, partimos do pressuposto de que o termo

amostra “refere-se ao grupo de indivíduos [...] selecionados para *representar*, no estudo, a população ou o universo do qual fazem parte e que o pesquisador deseja estudar” (GUY; ZILLES, 2007, p. 109, grifo nosso).

Há, pois, uma pressuposição de que o comportamento lingüístico dos indivíduos, cujo discurso examinamos reflete regularidades ligadas ao fato de que aderem às normas de seus respectivos grupos sociais; é nesse sentido que os resultados do estudo do comportamento de certo número de indivíduos (a amostra) são generalizados para os grupos sociais aos quais eles pertencem (e representam). (GUY; ZILLES, 2007, p. 109).

De acordo com Guy e Zilles (2007, p. 109), algumas perguntas norteiam a constituição de uma amostra, a saber: “Como definir, identificar ou delimitar os grupos sociais que constituem uma comunidade? Quais deles devem ser incluídos na amostra? Como relacionar os indivíduos necessários para ter uma amostra representativa nesse sentido estatístico?”.

Embora não haja uma resposta simples e única para essas questões, pois muitas alternativas têm sido adotadas por diferentes pesquisadores para a realização do trabalho de campo sociolinguístico, o que direcionam as respostas a esses questionamentos são os objetivos propostos em cada pesquisa sociolinguística. Os critérios de constituição de uma amostra devem ser coerentes com a pesquisa que se pretende realizar. Para a constituição de nossa amostra, estabelecemos, logo de início, dois parâmetros para a seleção dos informantes: os falantes deveriam ser pessoas nascidas em Alagoas e que não tivessem se afastado do Estado por tempo superior a cinco anos.

Em seguida, estratificamos a amostra em três células sociais: sexo, faixa etária e escolaridade, e as subdividimos nos seguintes fatores: sexo (masculino / feminino), faixa etária (F1 – 15 a 29 anos / F2 – 30 a 44 anos / F3 – mais de 44 anos) e escolaridade (E1 – Ensino Fundamental / E2 – Ensino Médio / E3 – Ensino Superior). Com isso, obtivemos um total de 18 células, conforme Quadro 1.

Masculino	F1	E1	Feminino	F1	E1
Masculino	F1	E2	Feminino	F1	E2
Masculino	F1	E3	Feminino	F1	E3
Masculino	F2	E1	Feminino	F2	E1
Masculino	F2	E2	Feminino	F2	E2
Masculino	F2	E3	Feminino	F2	E3
Masculino	F3	E1	Feminino	F3	E1
Masculino	F3	E2	Feminino	F3	E2
Masculino	F3	E3	Feminino	F3	E3

Quadro 1 – Estratificação da amostra. Fonte: Vitória (2012, p. 68).

A partir da estratificação, delimitamos o número de informantes necessários para obtermos uma amostra representativa da comunidade estudada. Selecionamos quatro informantes por células e obtivemos um total de 72 ($4 \times 18 = 72$) a serem entrevistados, constituindo, assim, uma amostra sincrônica composta por 72 entrevistas.

De acordo com Guy e Zilles (2007), em uma pesquisa de cunho variacionista, o ideal é selecionar quatro ou cinco informantes em cada célula, para evitar, durante o momento da entrevista e constituição do *corpus* da pesquisa, um comportamento linguístico idiossincrático ou enviesado, caso trabalhássemos com um ou dois informantes por células.

O acréscimo de uma terceira pessoa já nos daria chance de identificar as tendências de uso para aquele grupo, mas ainda assim poderíamos enfrentar dúvidas relacionadas com diferenças (se são por acaso, por idiossincrasia ou por razões de outra ordem) e ter pouca base para fazer qualquer tipo de generalização. Por isso, diz-se que, com 4 ou 5 indivíduos em cada célula, aumentamos substancialmente as chances de identificar *tendências* através da constatação de regularidades no comportamento dessas pessoas, em contraste com o de outras pessoas da amostra. (GUY; ZILLES, 2007, p. 112-113, grifo nosso).

Quanto à seleção desses informantes, seguimos o método aleatório estratificado, levando em consideração não só que eles deveriam ser pessoas nascidas em Alagoas e que não tivessem se afastado do Estado por tempo superior a cinco anos, mas também deveriam corresponder às células sociais definidas na estratificação da amostra.

Nesse método, o pesquisador sociolinguista divide a amostra em células sociais (no nosso estudo, dividimos em sexo, faixa etária e escolaridade) compostas, cada uma delas, por indivíduos que apresentam as mesmas características sociais e em seguida, procede-se, a partir de uma seleção aleatória, ao preenchimento de cada célula.

Também consideramos, conforme Milroy e Milroy (1992), a abordagem “bola de neve”, em que um informante, a partir do contato com o pesquisador, indica outro para fazer parte da pesquisa. A adoção dessa abordagem também facilitou o acesso aos informantes, uma vez que mediava a interação entre entrevistador e entrevistado.

3. COLETA DOS DADOS

Delimitada a amostra da pesquisa, o passo seguinte foi a coleta dos dados. Guy e Zilles (2007, p.20) apontam essa atividade lida com as seguintes perguntas:

“Como obtemos os dados? Os dados são válidos para refletir o fenômeno investigado? Os procedimentos para a obtenção dos dados são confiáveis e reproduzíveis? O que pode ser feito para minimizar a parcialidade dos dados?”.

Para a obtenção dos dados, elaboramos uma ficha da amostra sociolinguística contendo os dados dos informantes a serem entrevistados: nome, naturalidade, profissão, sexo (masculino / feminino), faixa etária (F1 – 15 a 29 anos / F2 – 30 a 44 anos / F3 – mais de 44 anos) e escolaridade (E1 – Ensino Fundamental / E2 – Ensino Médio / E3 – Ensino Superior). A elaboração da ficha social, além de fornecer informações sobre os informantes, o familiariza o com gravador e mapeia seus possíveis interesses, auxiliando no momento da entrevista.

Também elaboramos um questionário-guia de entrevistas, com os tópicos da conversa:

1. Fale-me da sua profissão/curso.
2. Como é o seu dia de trabalho/estudo? O que você faz?
3. Pretende fazer algum outro curso? Qual? Por quê?
4. Fale-me um pouco da sua cidade e da administração do atual prefeito.
5. Fale-me um pouco do nosso estado e da administração do atual governador.
6. Qual a sua opinião sobre essa violência toda que está havendo aqui?
7. O que você faria para amenizá-la?
8. Você (amigo/parente/conhecido) já sofreu algum tipo de violência? O que aconteceu?
9. Fale-me de um passeio/viagem que você fez e achou interessante.
10. Já passou por alguma situação que pôs sua vida em risco? O que aconteceu?

O questionário-guia de entrevistas tem como principal objetivo homogeneizar os dados para posterior comparação, controlar os tópicos da conversa e provocar narrativas de experiências pessoais, tendo em vista que estudos com narrativas de experiências pessoais têm demonstrado que, ao relatá-las, o informante está tão envolvido com *o que* fala que presta o mínimo de atenção ao *como* fala, o informante “fica envolvido na narrativa a ponto de parecer estar revivendo aquele momento.” (LABOV, 2008, p. 119).

Em seguida, entramos em contato com os informantes e realizamos as entrevistas, que foram feitas, na maioria das vezes, em nosso primeiro encontro, devido à disponibilidade e ao interesse dos falantes em participar da pesquisa².

2 Como nosso objeto de estudo era a realização dos verbos *ter* e *haver* em construções existenciais na fala alagoana, verificamos, durante o período de coleta de dados, que as entrevistas realizadas, tanto no primeiro contato com os informantes quanto no segundo, não interferiam na realização da variação em estudo.

Nossas entrevistas aconteceram ou nas residências, ou nos locais de trabalho dos informantes, no período de fevereiro a julho de 2010.

Nesse período de coleta de dados, entrevistamos aproximadamente 15 informantes por mês. No entanto, convém ressaltar que, ao considerarmos o método aleatório estratificado na seleção dos informantes, realizamos mais entrevistas no início do que no final da coleta de dados, uma vez que à medida que as células sociais foram sendo preenchidas, aumentavam ainda mais as especificidades dos informantes a serem entrevistados.

Como equipamento de gravação, utilizamos um gravador digital padrão e armazenamos nossas entrevistas no formato de arquivo .wav. Para obtermos um material linguístico em que predominasse a espontaneidade da fala dos entrevistados, nossas interferências geralmente ocorriam para estimular a continuidade da fala. O resultado de todas as gravações, de mais ou menos 15 minutos por falantes, corresponde a aproximadamente 1080 minutos de falas, totalizando quase 18 horas de entrevistas³.

[...] nosso objetivo é observar o modo como as pessoas usam a língua quando não estão sendo observadas. Todos os nossos métodos envolvem uma aproximação a esse objetivo: quando fazemos uma abordagem a partir de duas direções diferentes e obtemos o mesmo resultado, podemos ter certeza de que conseguimos vencer o paradoxo do observador no sentido de que a estrutura existe independentemente do analista. (LABOV, 2008, p. 83).

Realizadas todas as entrevistas, obtivemos uma amostra da comunidade de fala alagoana composta de 72 informantes, oriundos das regiões destacadas no mapa da Figura 1.

3 Entendemos que entrevistas sociolinguísticas de base laboviana duram em média de 50 a 80 minutos, desconsiderando, na maioria das pesquisas, os 15 minutos iniciais de coleta de dados, devido ao estranhamento em relação ao contexto de entrevista. No entanto, como nosso foco nessa coleta era a variação dos verbos *ter* e *haver* em construções existenciais, a gravação de mais ou menos 15 minutos por informantes nos forneceu um *corpus* estatisticamente relevante para a análise dos dados.

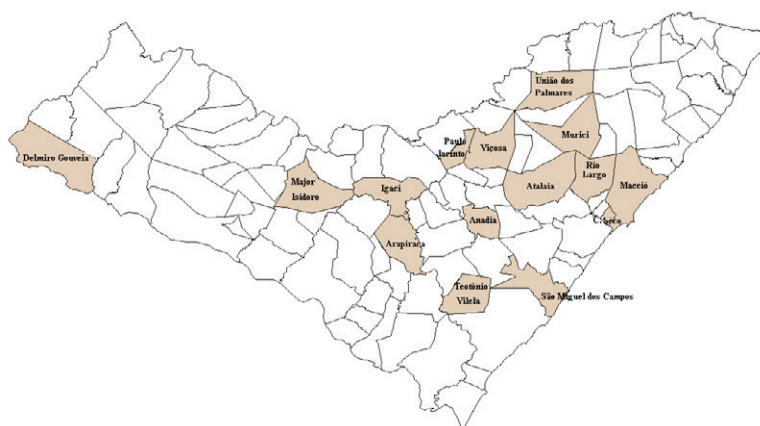


Figura 1 – Distribuição geográfica dos informantes. Fonte: Vitório (2012)

Embora não consideremos, na estratificação da amostra, a origem geográfica dos informantes, entrevistamos, no momento da coleta de dados, falantes que residiam em várias regiões do estado, para que pudéssemos ter uma visão geral da língua usada em Alagoas.

4. TRANSCRIÇÃO DOS DADOS

Após a realização das entrevistas, o passo seguinte foi transcrevê-las. Na prática, as transcrições dos dados ocorreram simultaneamente às entrevistas, ou seja, ao fim de cada, procurávamos logo iniciar o trabalho de transcrição dos dados. Procedemos dessa forma com o objetivo de capturar de forma mais fidedigna possível os fatos relatados.

Para tanto, seguimos o Protocolo de Transcrição do Projeto A Língua Usada em Alagoas (LUAL)⁴, segundo o qual, todas as entrevistas gravadas tiveram transcrição ortográfica, ou seja, procuramos seguir a ortografia oficial, mas registrando o máximo de questões características da fala coletada, conforme podemos observar no excerto transcrito a seguir.

4 As convenções de transcrição foram adotadas da adaptação do modelo da Equipe do Groupe Aixois de Recherches en Sociolinguistique (GARS), dirigido por Blanche-Benveniste, para o português, realizadas pela Prof^a Dr^a Denilda Moura, para o projeto A Língua Usada em Alagoas – LUAL. Tanto as convenções de transcrição quanto o Protocolo de Transcrição do Projeto LUAL estão disponíveis no banco de dados do Programa de Extensão em Ensino de Línguas (PRELIN), no Programa de Pós-Graduação em Letras e Linguística da Universidade Federal de Alagoas.

L27— cara eu acho que a principal causa é:: é a falta de imprego e oportunidade pru povo porque não tem educação de qualidade – você num tem educação de qualidade você num tem incentivo – porque as pessoas xxx porque como lá nu colégio que eu tava dando aula você via – os alunos gritava – ah vô istudá pra que? vô mi formá e num vô tê imprego – vô mi formá e num máximo que vô consegui é vendê no shopping – aí eu vô tá feliz cum isso – eu acho que isso é uma das grandes causas – num tem um alvo – a educação num presta – você num tem um objetivo pra xxx

As transcrições dos dados foram feitas com o auxílio do programa computacional Express Scribe, que pode ser obtido gratuitamente na internet e objetiva auxiliar o pesquisador na tarefa de transcrição do registro de áudio. A escolha desse programa foi motivada por possibilitar o uso de teclas de atalhos para executar diversas funções, permite recortar pedaços registrados da fala, facilitando o processo de edição, e visualiza as informações trabalhadas em uma única janela, conforme Figura 2.

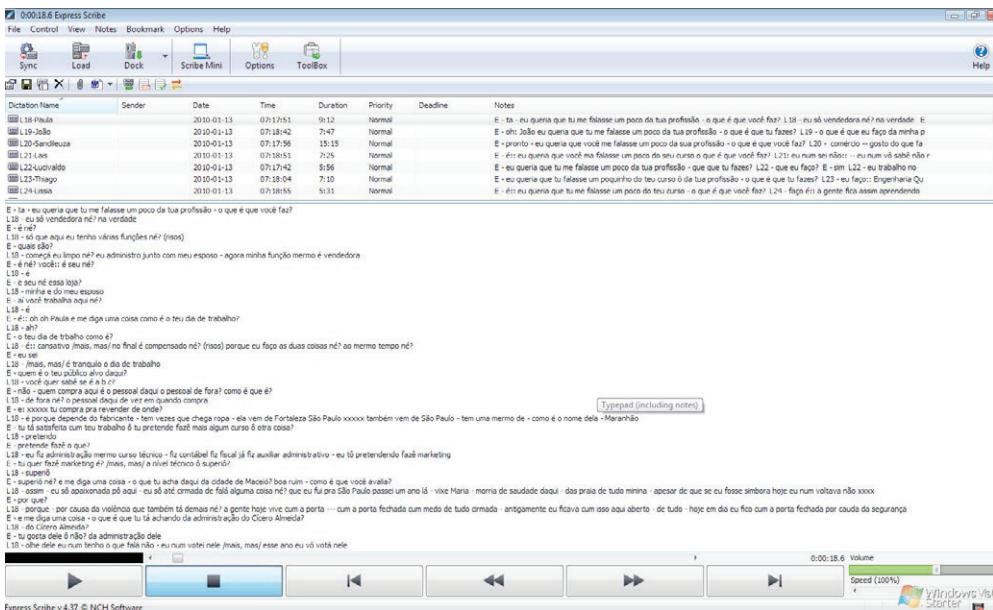


Figura 2 – Janela de trabalho do Express Scribe

Realizadas as transcrições das entrevistas, fizemos uma leitura de revisão para checar se os dados transcritos estavam de acordo com as falas coletadas.

O pesquisador deve ter em mente que, embora a transcrição seja o resultado do que é percebido e, portanto, também é consequência do cruzamento das opções teóricas do

pesquisador, da interpretação atribuída aos dados e dos objetivos da pesquisa, não existe uma transcrição irretocável, por isso é conveniente que a transcrição realizada possa ser ‘avaliada’ por outra(s) pessoa(s) e, quando necessário, submetida a revisões. (DE PAULA, 2011, p. 37).

Atualmente, está sendo realizada mais uma nova checagem das transcrições realizadas e, quando terminada, a amostra sincrônica da fala alagoana, coletada e transcrita, será disponibilizada à comunidade acadêmica científica como mais uma fonte para estudos descritivos de variedades do português brasileiro falado.

CONCLUSÕES

Adotando os pressupostos teórico-metodológicos da Teoria da Variação e Mudança, apresentamos os passos metodológicos seguidos para a coleta de uma amostra sincrônica da língua falada em Alagoas. Para tanto, focalizamos nossa discussão na delimitação da comunidade estudada, na estratificação da amostra a ser coletada e nas tarefas de coleta e transcrição dos dados. Vale ressaltar que, mesmo seguindo os passos da Sociolinguística Quantitativa, os procedimentos aqui adotados tinham como objetivo a pesquisa sobre a variação dos verbos “ter” e “haver” em construções existenciais na fala alagoana (VITORIO, 2012). Dessa forma, consideramos que “cada pesquisa impõe determinados limites, obrigando o pesquisador à contorná-los como uma forma de obter dados pertinentes, confiáveis e representativos do universo pesquisado” (DE PAULA, 2011, p. 39).

REFERÊNCIAS

- CAMPOY, J; ALMEIDA, M. *Metodología de la investigación sociolingüística*. Granada: Editorial Comares, 2005.
- BRASIL. Ministério do Planejamento, Orçamento e Gestão. Instituto Brasileiro de Geografia e Estatística. *Censo 2010*. Brasília, 2010. Disponível em: <www.Ibge.gov.br>. Acesso em: set. 2010.
- _____. Ministério do Planejamento, Orçamento e Gestão. Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional por Amostras de Domicílios – PNAD/2009*. Brasília: 2009. Disponível em: <www.Ibge.gov.br>. Acesso em: set. 2010.
- DE PAULA, A. O trabalho de campo sociolinguístico. In: COSTA, J.; SANTOS, R.; VITÓRIO, E. (org.). *Variação e mudança linguística no estado de Alagoas*. Maceió: EDUFAL, 2011, p. 29-41.

GUY, G.; ZILLES, A. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola Editorial, 2007.

LABOV, W. *Padrões sociolinguísticos*. Trad. M. Bagno, M. M. P. Scherre, C. R. Cardoso. São Paulo: Parábola Editorial, 2008. Título original: *Sociolinguistic Patterns*, 1972.

MILROY, L.; MILROY, J. Social network and social class: toward an integrated sociolinguistic model. *Language in Society*, 1992, v. 21, n. 1, p. 1-26.

TAGLIAMONTE, S. *Analysing sociolinguistic variation: key topics in sociolinguistic*. Cambridge: Cambridge University Press, 2006.

VITORIO, E. *Ter/haver existenciais na fala alagoana: variação estável ou mudança em progresso?*. Alagoas, 2012. Tese (Doutorado em Linguística). Faculdade de Letras, Universidade Federal de Alagoas.