

Dados geolinguísticos sob uma perspectiva estatística: a variação lexical no Centro-Oeste, Sudeste e Sul do Brasil

Geolinguistic Data under a Statistical Perspective: Lexical Variation in Midwest, Southeast and South of Brazil

Valter Pereira Romano^{1*}
Universidade Estadual de Londrina
Londrina - Paraná / Brasil

Rodrigo Duarte Seabra^{2**}
Universidade Federal de Itajubá
Itajubá - Minas Gerais / Brasil

Resumo

Este trabalho utiliza como *corpus* de análise os dados coletados pela equipe do Projeto Atlas Linguístico do Brasil em três regiões: Centro-Oeste, Sudeste e Sul. A pesquisa objetiva discutir a distribuição diatópica das variantes lexicais para a questão 133 – *Criança pequenininha, a gente diz que é bebê. E quando ela tem de 5 a 10 anos, do sexo feminino?* – do Questionário Semântico-Lexical (COMITÊ NACIONAL, 2001, p. 32). A metodologia utilizada envolveu análises descritivas e inferenciais pertinentes ao estudo. Para tanto, cinco hipóteses foram testadas a fim de averiguar o comportamento e a distribuição das cinco variantes mais produtivas no conjunto de respostas. Os principais resultados indicam que as variantes apresentam comportamento distinto nas três regiões: (i) a

* valter.romano@hotmail.com

** rodrigo@unifei.edu.br

variante *menina* apresenta distribuição homogênea nos dez estados; (ii) a forma lexical *guria* possui distribuição heterogênea nas regiões Centro-Oeste e Sul; (iii) a variante *garota* apresenta homogeneidade no Sudeste do país.

Palavras-chave

Projeto ALiB; Dialectologia; Variantes Lexicais.

Abstract

This paper uses as its corpus of analysis the data collected by the project Brazilian Linguistic Atlas in three regions: Midwest, Southeast and South. The research aims at discussing the diatopical distribution of lexical variants concerning question 133 – *Criança pequeninha, a gente diz que é bebê. E quando ela tem de 5 a 10 anos, do sexo feminino?* – from the Lexical-Semantic Questionnaire (COMITÊ NACIONAL, 2001, p. 32). The methodology involved descriptive and inferential analysis appropriate for the study. Five hypotheses were tested in order to analyze the behavior and distribution of the five most productive variants in the set of answers. The main results indicate that variants exhibit different behavior in the three regions analyzed: (i) the variant *menina* shows homogeneous distribution in ten states; (ii) the lexical form *guria* has heterogeneous distribution in the Midwest and Southern regions; (iii) the variant *garota* shows homogeneity in the Southeast of the country.

Keywords

ALiB Project; Dialectology; Lexical Variants.

1 INTRODUÇÃO

Este trabalho, inserido no âmbito dos estudos dialetais do Brasil, visa a discutir a distribuição diatópica das variantes lexicais para designativos de *menina* em dez estados brasileiros situados nas regiões Centro-Oeste, Sudeste e Sul. Partindo de dados empíricos coletados para o Projeto Atlas Linguístico do Brasil, formulam-se hipóteses que retratam o comportamento linguístico das variantes documentadas submetendo-as a testes estatísticos.

O trabalho pauta-se no léxico, pois este é o “nível da língua que melhor documenta o modo como um povo vê e representa a realidade em que vive” (ISQUERDO, 2003, p. 165). De acordo com Isquerdo (2007, p. 533), o léxico representa ainda um aspecto diferenciador no que se refere à variação linguística, sobretudo, a geográfica, pois além de evidenciar diferenças de uma região para outra, demonstra também a conseqüente mobilidade dessas diferenças de um espaço para outro. Dessa forma, verificar as especificidades lexicais de cada área geográfica é uma tarefa desafiadora, dada a dinamicidade da língua e a disseminação de variantes.

Nas próximas seções, apresentam-se algumas considerações sobre os estudos geolinguísticos no Brasil, com o intuito de traçar o panorama dos atlas linguísticos estaduais, bem como caracterizar, sumariamente, o Projeto Atlas Linguístico do Brasil. Dando continuidade ao trabalho, resgata-se o registro das principais variantes documentadas em seis obras lexicográficas. Na seção 4, tecem-se comentários sobre os materiais e métodos empregados no estudo e, na seqüência, a análise dos dados. Ao final do artigo, apresentam-se as considerações finais, seguidas das referências bibliográficas.

2 SITUANDO A GEOLINGUÍSTICA BRASILEIRA

O estudo da variedade linguística de determinada comunidade evidencia características próprias de seus falantes, visto que cada indivíduo imprime as suas particularidades na língua. Assim, dependendo do sexo, da faixa etária,

da etnia, do nível social do indivíduo, entre outras variáveis, encontram-se diferentes comportamentos linguísticos.

A variação diatópica, no entanto, subjaz aos diferentes tipos de variação, seja ela diastrática, diassexual, diagenérica ou diafásica³, pois o indivíduo encontra-se situado em determinada área geográfica, o que também determina a variedade linguística que utiliza. Nesse sentido, os atlas linguísticos⁴ têm contribuído, sobremaneira, para a descrição diatópica das línguas.

No que se refere ao português brasileiro, a elaboração de atlas linguísticos tem alcançado inúmeros avanços desde 1963, ano em que Nelson Rossi, grande expoente da Dialetoologia brasileira, publicou o primeiro atlas linguístico regional, o Atlas Prévio dos Falares Baianos (APFB). Passados catorze anos da publicação do APFB, outro atlas linguístico foi publicado, o Esboço de um Atlas Linguístico de Minas Gerais (RIBEIRO *et al.*, 1977), seguindo-se o Atlas Linguístico da Paraíba (ARAGÃO e BEZERRA DE MENEZES, 1984), o Atlas Linguístico de Sergipe (FERREIRA *et al.*, 1987) e o Atlas Linguístico do Paraná (AGUILERA, 1994).

Os estudos dialetológicos brasileiros, em especial os da área da Geolinguística, tomaram novo fôlego com o lançamento, em 1996, do Projeto Atlas Linguístico do Brasil (ALiB)⁵, à medida que propôs uma sólida metodologia no que range à definição de rede de pontos, ao perfil dos informantes, ao modelo de transcrições e de instrumentos de coleta de dados. Assim, a esses cinco primeiros atlas estaduais publicados somam-se o Atlas Linguístico Sonoro do Pará (RASKY, 2004), o Atlas Linguístico do Amazonas (CRUZ, 2004), o Atlas Linguístico do Sergipe II (CARDOSO, 2005), o Atlas Linguístico do Mato Grosso do Sul (OLIVEIRA, 2007), o Atlas Linguístico do Paraná II (ALTINO, 2007) e o Atlas Linguístico do Ceará (BESSA, 2010), dois destes, até o presente momento não publicados, o ALAM e o ALPR II. Além desses 11 atlas estaduais, soma-se, ao conjunto de atlas brasileiros divulgados, o Atlas Linguístico Etnográfico da Região Sul (ALTENHOFEN; KLASSMAN, 2011), o único que recobre uma região administrativa do país.

Dessa forma, os atlas linguísticos fornecem um amplo material de cartas fonéticas, léxicas e morfossintáticas que retratam o falar de determinada área geográfica e comunidade linguística, constituindo-se, portanto, como documentos históricos, que trazem informações não somente linguísticas, mas, principalmente, socioculturais.

Os dados coletados para a elaboração do Atlas Linguístico do Brasil têm contribuído para a descrição do português brasileiro sob a perspectiva lexical,

fonética, morfossintática e prosódica, conforme se atesta em Aguilera, Altino e Isquerdo (2012).

No nível lexical, dentre os trabalhos mais recentes, citam-se como exemplos os de Isquerdo (2009), Romano e Aguilera (2009), Silva e Aguilera (2010), Paim (2011), Yida (2011), Silva-Costa e Isquerdo (2012), Freitas Marins (2012), Ribeiro (2012), entre outros. Esses trabalhos apontam para as diferenças regionais do léxico do português brasileiro, ora confirmando a proposta de divisão dialetal de Nascentes (1953), sob a perspectiva lexical, ora evidenciando particularidades que caracterizam cada uma das regiões.

Interessa-nos, neste estudo, verificar a variação lexical dos designativos para ‘a criança que tem entre cinco e dez anos do sexo feminino’ relativa à questão 133 do Questionário Semântico Lexical do ALiB (COMITÊ NACIONAL, 2001, p. 32) junto a informantes naturais das três regiões brasileiras supracitadas.

3 RESGATE DAS VARIANTES EM OBRAS LEXICOGRÁFICAS

Para a questão 133 do QSL, foram registradas nove variantes lexicais, das quais discutimos as ocorrências das seis mais produtivas no universo das respostas: *menina*, *guria*, *garota*, *mocinha*, *moleca* e *piveta*⁶. Cada um desses itens apresenta diferentes formações etimológicas e distintas datações na língua portuguesa. Assim, foram selecionados seis dicionários para verificar as acepções, a datação e a etimologia de cada um dos vocábulos, a saber: três dicionários antigos, Bluteau (1728), Moraes Silva (1813) e Pinto (1832); dois dicionários contemporâneos gerais do português, Houaiss (2001) e Ferreira (2004); e um dicionário etimológico, Cunha (2010).

Bluteau (1728) registra o verbete *menina* como um sinônimo de *rapariga*, registrado com a acepção de “mocinha, Puella” (BLUTEAU, 1728, p. 107). O lexicógrafo abona o verbete *menina* da seguinte forma: “São senhoras da primeira qualidade, & moças; ouvi dizer que lhe chamão Meninas, porque andão com calçado baixo, & sem chapins” (BLUTEAU, 1728, p. 421). Moraes Silva (1813), por sua vez, apresenta o vocábulo *menina* como um substantivo feminino “A femea de tenra idade” (MORAES SILVA, p. 288). Pinto (1832), além de apresentar a locução *menina dos olhos* como designativo da pupila, remete o leitor à forma masculina do vocábulo, *menino* e, para este verbete, “Diz do homem, ou mulher até a idade de sete anos” (PINTO, 1832, s/p). Para

Cunha (2010, p. 420), o vocábulo data sua introdução na língua portuguesa, aproximadamente, no século XIII, derivada das formas históricas “meni)a, menynna XIII etc”, com a acepção de “criança do sexo feminino”. Ferreira (2004) apresenta três acepções para o vocábulo, todas classificadas como substantivo feminino, indicando também outras variantes: “1. Criança do sexo feminino; garota. 2. Mulher nova e/ou solteira; mocinha, garota. 3. Tratamento familiar e afetuoso dado às pessoas de sexo feminino, crianças e adultos” (FERREIRA, 2004). Já Houaiss (2001) apresenta seis acepções acrescentando outras não apontadas pelos lexicógrafos:

1. criança ou adolescente do sexo feminino; garota; 2. Jovem do sexo feminino até a idade núbil; moça jovem e/ou solteira; mocinha; 3. descendente do sexo feminino; filha; 4. maneira carinhosa e familiar de tratar um parente ou amiga, mesmo quando já adulta; 5. moça que se namora; namorada, garota, pequena; 6. Uso: informal. mulher da vida; meretriz, prostituta (HOUAISS, 2001).

O item lexical *guria* relaciona-se diretamente com a sua forma no masculino, *guri*. Nenhum dos três dicionários antigos a que tivemos acesso, Bluteau (1728), Moraes Silva (1813) e Pinto (1832) registram *guri* ou *guria*. Em Cunha (2010, p. 329), encontra-se o registro do verbete *guri* e, de acordo com o etimologista, este vocábulo veio do tupi *ü i'ri* com a acepção que designa o *bagre novo* (tipo de peixe), por extensão de sentido, a criança. Ainda nesta entrada, o dicionarista remete o usuário ao verbete “*guiri* - ‘sm bagre’ / *curi* 1587, *guori* datada aproximadamente 1631”. Houaiss (2001), em contrapartida, registra a datação de 1890, documentada no Dicionário Etimológico da Língua Portuguesa de José Pedro Machado (1952). Para o lexicógrafo, entre outras acepções, trata-se de um regionalismo brasileiro para designar o *menino, a criança*. Apresenta o verbete *guria* como um regionalismo do Brasil sem especificar a área em que ocorre, definindo como “1. criança do sexo feminino; menina; 2 moça com quem se namora; namorada, garota”. Segundo o lexicógrafo, este vocábulo está registrado na 10ª edição do Dicionário da Língua Portuguesa de Moraes Silva (1958); portanto, sua introdução nos dicionários de português não é tão antiga quanto o vocábulo correlato masculino, *guri*. Ferreira (2004), apesar de apresentar a mesma acepção para o vocábulo *guri*, traz outra etimologia. Segundo o dicionarista, este item lexical

também vem do tupi, porém com o sentido de pequeno, não fazendo, portanto, alusão ao peixe. No verbete *guria*, apresenta as mesmas acepções de Houaiss (2001), tratando-se de um brasileirismo, para o feminino de *guri*: “Menina, namorada, garota” (FERREIRA, 2004).

O vocábulo *garota* não está dicionarizado nem em Bluteau (1728) nem em Pinto (1832). Esses dicionários também não registram o vocábulo na sua forma masculina. Somente em Moraes Silva (1813) se encontra o primeiro registro na língua portuguesa que, por sua vez, também não registra a forma no feminino. Para Moraes Silva (1813), *garoto* é o “rapaz brejeiro, mal criado, petulante” (MORAES SILVA, 1813, p.80). Cunha (2010, p.311) considera incerta a origem do vocábulo, porém, de acordo com Houaiss (2001), José Pedro Machado relaciona o vocábulo ao “fr. *gars* (sXII) ‘rapaz’, do fr. *garçon*, seguido do sufixo dim. *-oto*”. O vocábulo *garçon*, por sua vez, segundo o Dicionário Larousse (GÁLVEZ, 2005, p.159), é usado na língua francesa para designar o *menino*, assim como *gamin*. Houaiss (2001) apresenta três acepções para o verbete *garota*, das quais duas referem-se ao referente em pauta: “1. criança ou adolescente do sexo feminino; 2. moça que se namora; namorada, pequena” (HOUAISS, 2001). Ferreira (2004) traz as mesmas acepções de Houaiss (2001), além de uma forma sinonímica para *ônibus, jardineira*.

Na entrada do vocábulo *mocinha*, Bluteau (1728) remete o leitor ao verbete *moçasinha*: “moça pequena. Mocinha” (BLUTEAU, 1728, p.522). Moraes Silva (1813, p.307) também remete ao verbete *moçasinha*, porém, nas duas entradas, apresenta-o como um substantivo feminino, indicando o diminutivo de *moça*. Pinto (1832) não registra o verbete *mocinha* nem *moçasinha*, somente em *moça* o usuário é remetido ao verbete *moço*, classificado como “adj. que está nos annos da mocidade” (PINTO, 1832). Cunha (2010, p. 431), no verbete *moço* não reconhece a origem do vocábulo e, segundo o etimologista, data o século XIII. Houaiss (2001), no entanto, apesar de salientar que a origem do vocábulo é controversa, acrescenta que “Corominas associa o port. *moço* ao esp. *mozo* e, lembrando que originalmente ambos designam ‘criança, homem jovem’, supõe que se ligue a *mocho* ‘raspado, pelado, sem chifres’, tido como voc. expressivo; seria costume manter crianças e jovens rapazes de cabelos raspados” (HOUAISS, 2001) e das quatro acepções para o verbete *mocinha*, apenas a primeira se aproxima do referente em pauta: “moça muito nova; moçoila, jovem, rapariga” (HOUAISS, 2001). Ferreira (2004)

apresenta apenas uma acepção para *mocinha*, como um vocábulo formado pela palavra *moça* + o sufixo *-inha* tratando-se de um brasileirismo na mesma acepção apontada por Houaiss (2001), “moça muito jovem; moçoila” (FERREIRA, 2004). Conforme se verifica, nas obras lexicográficas, não há registros da variante *mocinha* como forma sinônímica para a criança do sexo feminino, apenas referências ao caráter jovial da pessoa.

Bluteau (1728) não registra o verbete *moleca* e, na entrada do vocábulo *moleque*, afirma que “Veyo-nos esta palavra do Brasil” (p. 541), não o considerando, portanto, como uma forma típica do português europeu. Cunha (2010, p. 433) registra que esse vocábulo veio do quimbundo mu’leke, para designar o menino, rapazote com a datação de 1731. No século XIX, essa forma lexical já se encontrava dicionarizada, pois Moraes Silva (1813) registra que *moleque* é o “pretinho, negro pequeno” (p. 310). Pinto (1832) não registra *moleca* e na entrada de *moleque* aponta como o “preto, escravo, pequeno” (PINTO, 1832), portanto, no mesmo sentido de Moraes Silva (1813). Nos dicionários do século XXI, no entanto, a forma *moleca* encontra-se dicionarizada, pois tanto Ferreira (2004) quanto Houaiss (2001) a trazem como o feminino do vocábulo *moleque*, porém este último lexicógrafo acrescenta: “1 menina nova, de raça negra ou mista; molecote; 2 qualquer garota; molecota; 3 garota dada a travessuras; 4 menina brincalhona, jocosa; 5 menina criada à solta; 6 Regionalismo: Moçambique. criada doméstica jovem” (HOUAISS, 2001).

Nenhum dos seis dicionários consultados registraram a forma *piveta*. No entanto, verificou-se que o vocábulo *pivete*, nos dicionários mais antigos, possui apenas uma acepção. Bluteau (1728), por exemplo, registra *pivete* como o “perfume da feyçao) de hum paosinho redondo, que aceso exhala hu) sumo odorifero até que fica convertida em cinzas toda a materia de sua fragrância” (BLUTEAU, 1728, p.539). A acepção de Moraes Silva (1813) no verbete *pivete* é mais específica. Para o lexicógrafo trata-se de “s. m. Um pedacinho de droga aromática para perfumar; fino, e roliço (MORAES SILVA, 1813, p.456). Pinto (1832), por sua vez, restringe ainda mais a designação de *pivete* com uma única acepção “s. m. perfume”. Cunha (2010, p. 501), no entanto, esclarece que *pivete* é uma “substância aromática que se queima para perfumar” datada do século XVI e também como designativo para a “criança esperta, menino ladrão e/ou que trabalha para ladrões” já implementada na língua portuguesa no século XX. Para este lexicógrafo, o vocábulo veio do castelhano *pebete*, derivado do catalão *pevet*. Houaiss (2001), acerca da etimologia do vocábulo acrescenta:

“cast. pebete (1575) ‘pasta que, uma vez queimada, exala aroma’ < cat. pevet (ant. peuet) ‘incensário’, der. do cat. peu ‘pé’, porque sustenta o piveteiro; segundo Corominas, “ironicamente, empregou-se no sentido de (1612) ‘objeto mal cheiroso’, donde ‘menino em cueiros’ e ‘menino algo maior’, no esp. americano” (HOUISS, 2001).

O lexicógrafo apresenta três acepções para o verbete como a criança ou menino esperto que pratica furtos, sentido este empregado na língua portuguesa a partir do século XX, informação também constada em Ferreira (2004) e Cunha (2010) para esta acepção.

Conforme se verifica, os dicionários consultados apresentam diferentes acepções das variantes documentadas. Algumas não se encontram dicionarizadas como *piveta* e *moleca* remetendo-as, por vezes, à sua forma no masculino *pivete* e *moleque*. O contraste das acepções constantes de dicionários antigos da língua portuguesa (BLUTEAU, 1728; MORAES SILVA, 1813 e PINTO, 1832) com os modernos (HOUISS, 2001 e FERREIRA, 2004) revela que novos significados foram acrescentados ao vocábulo *pivete*, bem como destaca as influências de outras línguas no processo de formação do português no Brasil como a indígena, no caso de *guria*, africana, para *moleque* e espanhola, para *pivete*.

4 MATERIAIS E MÉTODOS

Os dados analisados foram coletados pela equipe do ALiB em 135 municípios brasileiros nas regiões Centro-Oeste, Sudeste e Sul. Em cada uma das unidades da Federação, foram selecionadas, como pontos de inquéritos, diferentes cidades, considerando-se a importância histórico-cultural da localidade e, também, a densidade demográfica do estado. Assim, o número de pontos linguísticos em cada unidade difere de uma para a outra. Em cada ponto, foram entrevistados quatro informantes de ambos os sexos com nível de escolaridade fundamental⁷, perfazendo um total de 540 informantes. Desse modo, tivemos acesso aos dados relativos a nove localidades do estado do Mato Grosso; seis do Mato Grosso do Sul; nove de Goiás; 17 municípios de Minas Gerais⁸; 38 de São Paulo; cinco localidades do Espírito Santo; nove do Rio de Janeiro⁹; 17 municípios do Paraná; 10 localidades de Santa Catarina; e, por fim, 15 do Rio Grande do Sul¹⁰.

Para o armazenamento e análise dos dados coletados foi utilizado o *software* SPSS¹¹ (*Statistical Package for the Social Sciences*) versão 17.0 (SPSS, 2008), o que possibilitou estabelecer as análises descritivas e inferenciais pertinentes ao estudo.

As variáveis numéricas, ou seja, o número de ocorrências de cada variante, foram expressas em média e desvio padrão. Quanto às variáveis categóricas, aqui representadas pelas variantes lexicais, foram utilizadas, para sumarizá-las, frequências simples e relativas (números e porcentagens) e intervalo de confiança de 95%. Utilizou-se o teste estatístico de Kolmogorov-Smirnov (KS) para avaliar o pressuposto de normalidade, isto é, verificar a aderência das variáveis estudadas à distribuição de probabilidade normal. Valores de p-valor¹² maiores que 0,05 indicam que a variável tem adesão à distribuição normal. Para a comparação de médias entre grupos, necessária para os testes de hipóteses relativas às variáveis categóricas, foi considerada a adesão das variáveis à distribuição normal. Para tanto, utilizou-se o teste de Kruskal-Wallis para comparar mais que dois grupos independentes, no caso os estados e, em alguns casos, as mesorregiões desses estados. Os dados referentes a cada variante serão analisados segundo sua ocorrência. Os resultados são considerados estatisticamente significantes se p-valor < 0,05.

5 ANÁLISE DOS DADOS

5.1 Contextualização da Análise Estatística

A análise estatística dos resultados obtidos constitui um importante instrumento na validação desses dados. Resumidamente, o objetivo consiste em extrapolar os resultados da análise para a população estudada, neste caso, os informantes representativos do falar de cada região brasileira. Ademais, no estudo lexical, carecem pesquisas que abordam tratamentos estatísticos sofisticados na análise dos dados adentrando à formulação de hipóteses, e não somente pautando-se na descrição de percentuais em função de frequências absolutas e relativas. Aspectos teóricos estatísticos relevantes para o estudo da variação linguística, por exemplo, podem ser encontrados em Oliveira (2009).

No estudo, cinco hipóteses foram testadas com base nos dados coletados: (1) as ocorrências das variantes das três regiões apresentam comportamento homogêneo; (2) a variante *menina* apresenta homogeneidade em sua distribuição nos dez estados; (3) a variante *guria* apresenta distribuição homogênea na região Centro-Oeste; (4) a variante *garota* apresenta distribuição homogênea na região Sudeste; (5) a variante *guria* apresenta distribuição homogênea na região Sul.

5.2 Tratamento dos Dados

Para a questão 133 do QSL foram documentados 711 registros, distribuídos em seis variantes lexicais, conforme se observa na Tabela 1.

TABELA 1
Distribuição geral das variantes lexicais para a questão 133 do QSL

Variantes	N	% total
<i>Menina</i>	471	66,24
<i>Guria</i>	90	12,66
<i>Garota</i>	86	12,10
<i>Mocinha</i>	42	5,91
<i>Moleca</i>	19	2,67
<i>Piveta</i>	3	0,42
<i>Total</i>	711	

Fonte: Banco de dados do ALiB (2013).

A variante mais produtiva, no universo de todas as respostas, é *menina*, que representa 66,24% do *corpus*, seguida de *guria* (12,66%), *garota* (12,10%), *mocinha* (5,91%), e *moleca* (2,67%). Com menos de um por cento de representatividade, foi documentada a variante *piveta* (0,42%). A figura 1 apresenta a produtividade dessas variantes por região administrativa.

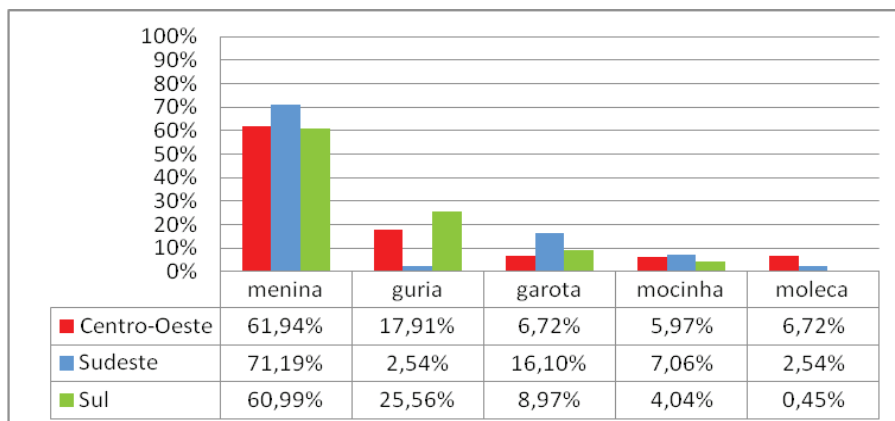


FIGURA 1 - Produtividade das variantes lexicais para a questão 133 do QSL segundo a região administrativa - Fonte: Banco de dados do ALiB (2013).

Menina apresentou-se como a variante mais produtiva nas três regiões: Sudeste (71,19%), Centro-Oeste (61,94%) e Sul (60,99%). No Centro-Oeste e no Sul, a segunda variante mais produtiva foi *guria* com 17,91% e 25,56%, respectivamente. A forma lexical *garota* foi a segunda variante mais produtiva nos municípios da região Sudeste, representando 16,10% das respostas, seguindo-se da variante *mocinha* (7,06%). As formas *garota* e *moleca*, no Centro-Oeste, representam a terceira variante mais produtiva, ambas com 6,72% de produtividade, enquanto *mocinha* ocupa a quarta colocação, com 5,97% de produtividade. Na Região Sudeste, as variantes *guria* e *moleca* ocupam a quarta posição, ambas com 2,54% de produtividade. Na Região Sul, por sua vez, destaca-se a variante *garota* (8,97%) como a terceira mais representativa. Observa-se, nesta análise inicial, que a variante *moleca* apresentou baixo índice percentual na Região Sul (0,45%).

Com vistas a facilitar a representação cartográfica e o tratamento estatístico dos dados, a análise apresentada no estudo se atém mais detalhadamente a cinco variantes mais produtivas: *menina*, *guria*, *garota*, *mocinha* e *moleca*. Para o teste das hipóteses formuladas na pesquisa, fez-se necessário analisar o comportamento das referidas variantes caracterizadas como variáveis aleatórias dependentes em diferentes contextos. Para isso, foi utilizado o teste KS objetivando verificar se os dados apresentam aderência à distribuição de probabilidade normal. Neste momento, é importante destacar que a distribuição normal é uma das mais importantes distribuições da estatística, pois muitas variáveis aleatórias de ocorrência natural obedecem a essa distribuição. Também conhecida como distribuição de Gauss ou Gaussiana, a normal possui grande uso na estatística inferencial. A distribuição normal tem como características fundamentais a média e o desvio padrão, e a curva que representa a distribuição de suas variáveis apresenta a forma de sino, sendo unimodal e simétrica. A simetria pode ser verificada se passarmos uma linha exatamente pelo centro da curva, que a dividirá em duas metades, sendo que cada uma delas é a imagem especular da outra. A tabela 2 apresenta os resumos estatísticos segundo os grupos de regiões.

TABELA 2
Resumo estatístico segundo a região administrativa

		<i>menina</i>	<i>Guria</i>	<i>garota</i>	<i>mocinha</i>	<i>moleca</i>
Centro-Oeste	Respostas: média (N*)	3,46 (24)	1,00 (24)	0,38 (24)	0,33 (24)	0,38 (24)
	Desvio padrão	0,658	1,022	1,964	0,565	0,576
	KS (p-valor)**	1,648 (0,009)	1,239	(p < 0,001)	(p < 0,001)	2,005 (p < 0,001)
Sudeste	Respostas: média (N*)	3,64 (69)	0,14 (69)	0,81 (69)	0,39 (69)	0,13 (69)
	Desvio padrão	0,727	0,430	0,845	0,521	0,339
	KS (p-valor)**	3,692 (p<0,001)	4,287	2,092 (p < 0,001)	3,298 (p < 0,001)	4,313 (p<0,001)
Sul	Respostas: média (N*)	3,24 (42)	1,36 (42)	0,48 (42)	0,21 (42)	0,02 (42)
	Desvio padrão	1,078	1,394	0,634	0,470	0,154
	KS (p-valor)**	1,995 (p < 0,001)	1,581	2,391 (p < 0,001)	3,144 (p<0,001)	3,483 (p<0,001)

Fonte: Banco de dados do ALiB (2013) Nota: *N = Quantidade de localidades ** Teste Kolmogorov-Smirnov

Com base nos dados apresentados na tabela 2, nota-se que, na região Centro-Oeste, apenas a variante *guria* apresenta aderência à distribuição normal (p-valor > 0,05). No Sudeste e no Sul, nenhuma variante adere à normal. Portanto, considera-se que os itens lexicais estudados não apresentam aderência à normal, sendo necessária a aplicação de testes estatísticos não paramétricos para testar as hipóteses formuladas no estudo.

Outro aspecto importante a ser verificado diz respeito à análise de variância, que consiste na comparação de dois ou mais grupos em relação à localização. Para exemplificar, pode-se considerar a situação em que se pretende avaliar se três grupos apresentam desempenhos similares em determinada avaliação. A análise de variância procura responder a questões deste tipo por meio da comparação das localizações dos diferentes grupos. Essa comparação é feita com base na dispersão presente no conjunto de dados (VIEIRA, 2006).

Com o objetivo de verificar se as variâncias das ocorrências são homogêneas nas três regiões analisadas, foi aplicado o teste de Kruskal-Wallis (TAB. 3). O resultado mostra que o teste alcançou um nível significativo, podendo-se assumir que as variâncias dos grupos não são homogêneas. Com base na análise de variância, nota-se que há diferença significativa entre as médias de todas as variantes

(p -valor $< 0,05$), exceto *mocinha*. Assim, de acordo com os dados apresentados, conclui-se que as regiões não apresentam comportamento homogêneo e a hipótese nula 1 pode ser rejeitada.

TABELA 3
Teste de homogeneidade das variantes com base nas regiões.

	Centro-Oeste	Sudeste	Sul	df	Teste*(p-valor)
	Média do rank (N)	Média do rank (N)	Média do rank (N)		
<i>menina</i>	62,48 (24)	74,81 (69)	59,96 (42)	2	6,114 (0,047)
<i>guria</i>	82,58 (24)	50,41 (69)	88,56 (42)	2	39,314 (p < 0,001)
<i>garota</i>	57,31 (24)	75,78 (69)	61,33 (42)	2	7,042 (0,030)
<i>mocinha</i>	67,60 (24)	72,67 (69)	60,56 (42)	2	3,912 (0,141)
<i>moleca</i>	81,71 (24)	67,74 (69)	60,60 (42)	2	12,839 (0,002)

Fonte: Banco de dados do ALiB (2013) *Teste de Kruskal-Wallis N = quantidade de localidades

Devido ao fato de *menina* ser a variante mais produtiva, presente em todos os estados, a próxima hipótese averiguada consistiu em verificar se a referida forma lexical apresenta distribuição homogênea nos dez estados, considerando as mesorregiões contempladas pela rede de pontos¹³. A análise compreendeu a aplicação do teste de Kruskal-Wallis (TAB. 4).

TABELA 4
Análise de variância de *menina* nos estados

Estados		<i>menina</i>	Estados		<i>menina</i>
Mato Grosso	χ^2	3,848	Espírito Santo	χ^2	2,333
	df	4		df	3
	p-valor (N)	0,427 (9)		p-valor (N)	0,506 (5)
Mato Grosso do Sul	χ^2	1,250	Rio de Janeiro	χ^2	6,756
	df	3		df	5
	p-valor (N)	0,741 (6)		p-valor (N)	0,239 (9)
Goiás	χ^2	3,800	Paraná	χ^2	9,956
	df	4		df	9
	p-valor (N)	0,434 (9)		p-valor (N)	0,354 (17)
Minas Gerais	χ^2	11,733	Santa Catarina	χ^2	1,800
	df	8		df	4
	p-valor (N)	0,164 (17)		p-valor (N)	0,772 (10)
São Paulo	χ^2	13,893	Rio Grande do Sul	χ^2	6,210
	df	15		df	4
	p-valor (N)	0,534 (38)		p-valor (N)	0,184 (15)

Fonte: Banco de dados do ALiB (2013) *Teste de Kruskal-Wallis N = Quantidade de localidades

Os resultados mostram que o teste não alcançou um nível significativo, podendo-se assumir que as variâncias dos estados são homogêneas. Com base na análise de variância (TAB 4), nota-se que não há diferença significativa entre as médias da variante *menina* (p-valor > 0,05). Assim, de acordo com os dados apresentados, conclui-se que os estados apresentam comportamento homogêneo em relação à distribuição de *menina* e a hipótese nula 2 pode ser aceita. A seguir, passa-se a apresentar os dados considerando individualmente cada uma das três regiões do país, bem como os estados que as constituem.

5.2.1 Região Centro-Oeste

Dos 711 registros documentados no cômputo geral das ocorrências, 134 referem-se aos dados coletados nos municípios da região Centro-Oeste, distribuídos pelos três estados: Mato Grosso, Mato Grosso do Sul e Goiás. Em cada um desses estados, as variantes lexicais obtiveram diferentes índices de ocorrência. Em dados percentuais, apresenta-se a distribuição visualizada na figura 2.

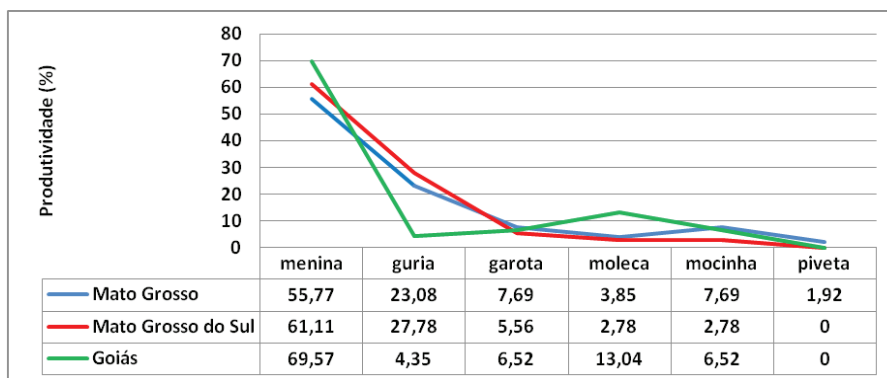


FIGURA 2 - Produtividade das variantes lexicais para a questão 133 do QSL nos estados da região Centro-Oeste - Fonte: Banco de dados do ALiB (2013)

Observa-se que a variante *menina* predominou nos três estados, obtendo maior produtividade em Goiás (69,57%). A variante *moleca* também foi mais produtiva nesse estado (13,04%), ao passo que *guria* não foi tão significativa (4,35%) se comparada à produtividade dessa variante em Mato Grosso do Sul (27,78%) e Mato Grosso (23,08%).

A variante *menina* esteve presente em todas as localidades que compõem essa região. Assim, para a representação cartográfica, não foi considerada¹⁴ essa forma lexical por apresentar uma distribuição diatópica homogênea na região Centro-Oeste. A Figura 3 mostra a distribuição diatópica de quatro variantes lexicais, a saber: *guria*, *garota*, *moleca* e *mocinha*.

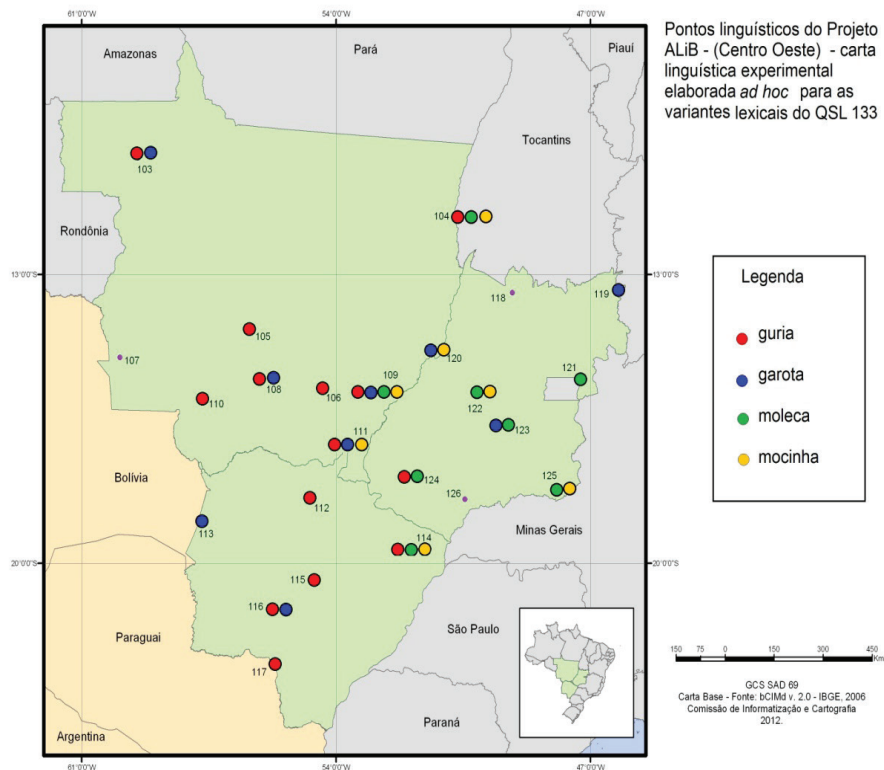


FIGURA 3 - Distribuição diatópica das quatro variantes lexicais mais produtivas na rede de pontos da região Centro-Oeste - Fonte: Banco de dados do ALiB (2013)

Observa-se que nos pontos 107 – Bela Vista/MT, 118 – Porangatu/GO e 126 – Quirinópolis/GO não há registro de nenhuma das quatro variantes elencadas na legenda. Isso se deve ao fato de que nessas localidades houve a ocorrência exclusiva da variante *menina*.

Os dados da carta linguística (FIG 3) ratificam as informações constantes da figura 2. Verifica-se que há uma continuidade na distribuição diatópica da

variante *guria* na rede de pontos do Mato Grosso e do Mato Grosso do Sul (não ocorrendo apenas no ponto 113 - Corumbá), adentrando apenas a uma localidade goiana, ponto 124 (Jataí), localizada na Mesorregião do Sul Goiano. Assim, a forma lexical *guria* não esteve presente em oito das nove localidades que compõem a rede de pontos em Goiás, que, por sua vez, apresenta uma maior distribuição da variante *moleca*.

O resultado referente à análise da segunda hipótese do estudo revela que *guria* não apresenta homogeneidade em sua distribuição na região Centro-Oeste, pois os dados obtidos pelo teste Kruskal-Wallis revelam que a hipótese nula 3 pode ser rejeitada ($\chi^2 = 9,767$; $df = 2$; $p\text{-valor} = 0,008$). Depreende-se, portanto, uma diferença linguística do estado de Goiás em relação aos demais estados que compõem a região Centro-Oeste no que se refere à distribuição diatópica da variante *guria*. Resultado análogo foi encontrado no trabalho de Romano e Seabra (no prelo) acerca do emprego da variante *guri* nesse estado federativo.

Assim, os testes estatísticos empreendidos ratificam a diferença linguística de Goiás, quanto ao uso da variante *guria*, em relação aos outros dois estados que compõem a região Centro-Oeste, evidenciando aspectos sócio-históricos dos processos de formação daquele estado. Sabe-se que o povoamento do território goiano processou-se rápido no período colonial, não somente por paulistas, mas também por bandeirantes e aventureiros da Bahia, Pernambuco, Maranhão, Pará e Minas Gerais (DIÉGUES, Jr., 1960, p. 272), o que propiciou o desenvolvimento de uma sociedade instável, caracterizada pela miscigenação de povos e diferentes padrões linguísticos, conforme atesta o trabalho de Freitas Marins (2012) acerca da variação lexical nessa região administrativa.

A variante *garota* encontra-se esparsamente distribuída nos três estados, o que não possibilita o traçado de uma linha de isoléxica¹⁵. Em contrapartida, as variantes *moleca* e *mocinha* encontram-se diatopicamente marcadas em Goiás e em cidades limítrofes a esse estado. Dessa forma, um estudo acerca do emprego dessas duas últimas variantes em localidades nortistas e nordestinas poderá confirmar a hipótese sobre as influências dos processos de ocupação e povoamento nos diferentes comportamentos linguísticos observados no estado de Goiás em relação ao Mato Grosso e Mato Grosso do Sul. Nesse sentido, o desenvolvimento do Projeto ALiB pode contribuir, sobremaneira, na definição dessas áreas convergentes e divergentes.

5.2.2 Região Sudeste

Nos municípios da região Sudeste, foram documentados 354 registros com diferentes índices de ocorrência. Em dados percentuais, apresenta-se a produtividade dessas ocorrências por meio da figura 4.

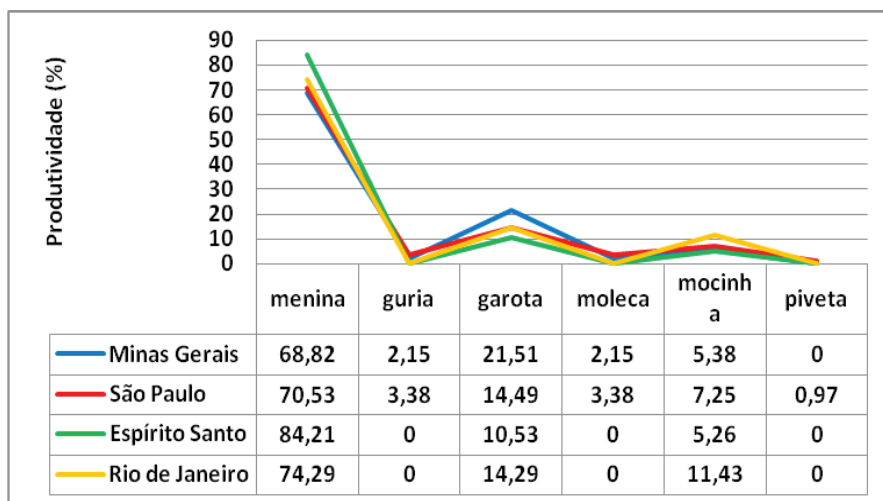


FIGURA 4 - Produtividade das variantes lexicais para a questão 133 do QSL nos estados da região Sudeste - Fonte: Banco de dados do ALiB (2013)

O gráfico revela um comportamento semelhante das variantes em cada um dos quatro estados que compõem a região. Assim, pode-se apontar três pontos principais: (i) a predominância da variante *menina* nos quatro estados, com produtividade acima de 68% em todos eles; (ii) a baixa produtividade da variante *guria* em dois estados: São Paulo (3,38%) e Minas Gerais (3,45%), não ocorrendo nos municípios do Espírito Santo e Rio de Janeiro; (iii) a presença da variante *garota* com percentual próximo em São Paulo (14,49%) e Rio de Janeiro (14,29%) e diferentes índices no Espírito Santo (10,53%) e Minas Gerais (21,51%).

Assim como no Centro-Oeste, no Sudeste, a variante *menina* também ocorreu em todos os pontos linguísticos. Dessa forma, para a cartografia, essa variante lexical foi desconsiderada, haja vista que a sua distribuição diatópica é homogênea em toda a região. A figura 5 apresenta a distribuição das quatro variantes (*guria*, *garota*, *moleca* e *mocinha*) nos pontos linguísticos do Sudeste.

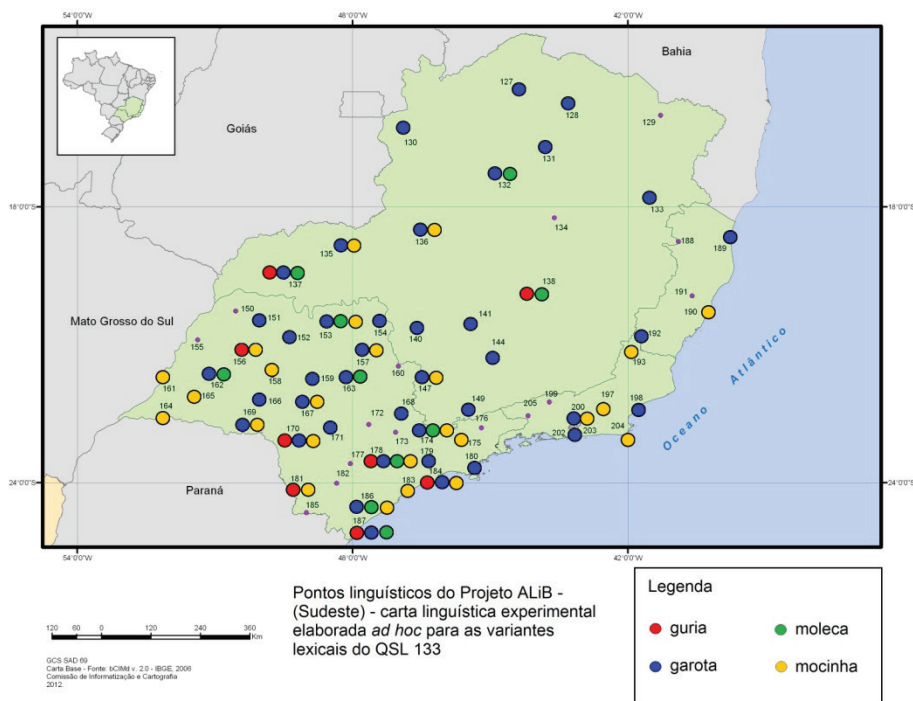


FIGURA 5 - Distribuição diatópica das quatro variantes lexicais mais produtivas na rede de pontos da região Sudeste – Fonte: Banco de dados do ALiB (2013)

A princípio, observa-se que, no mapa, há quinze pontos linguísticos em que não ocorre nenhuma das quatro variantes elencadas na legenda, dos quais nove são municípios paulistas (ponto 150 – Jales, 155 – Andradina, 160 – Mococa, 172 – Piracicaba, 173 – Campinas, 176 – Guaratinguetá, 177 – Itapetininga, 182 – Capão Bonito, 185 – Ribeira); dois mineiros (129 – Pedra Azul e 134 – Diamantina); duas cidades cariocas (ponto 199 – Valença e 205 – Barra Mansa) e dois municípios capixabas (ponto 188 – Barra de São Francisco e 191 – Santa Teresa). Isso se deve ao fato de que, nesses pontos, houve a ocorrência exclusiva da variante mais produtiva do *corpus*, *menina*.

Dentre as quatro variantes escolhidas para representação cartográfica, destaca-se a forma lexical *garota*, que apresenta uma ampla distribuição no estado de São Paulo (presente em 21 dos 38 pontos linguísticos do estado) e Minas Gerais (constante de 13 dos 17 pontos investigados). No Rio de

Janeiro, *garota* apresentou-se em três dos nove pontos linguísticos e, no estado do Espírito Santo, essa forma lexical foi registrada em duas das cinco cidades investigadas. O teste de homogeneidade de variâncias indicou que a hipótese 4 pode ser aceita ($\chi^2 = 7,402$; $df = 3$; $p\text{-valor} = 0,06$), ou seja, os resultados do teste evidenciam a homogeneidade da distribuição da variante *garota* na rede de pontos a que tivemos acesso. Vale, porém, ressaltar que a coleta de dados em Minas Gerais e no Rio de Janeiro ainda não foi concluída pela equipe do Projeto AliB; portanto, tal hipótese se aplica aos dados analisados, e a sua possível generalização dependerá do *corpus* completo dos referidos estados.

Apesar de não estar dicionarizada na acepção para a criança do sexo feminino e sim para a adolescente ou pessoa adulta, conforme se atestou na consulta às obras lexicográficas, o uso da variante *mocinha* é generalizado, principalmente nas localidades paulistas. Assim, dos quatro estados da região Sudeste, *mocinha* esteve presente em 17 dos 38 pontos linguísticos do Estado de São Paulo, ao passo que, em Minas Gerais, essa variante esteve em apenas duas localidades (pontos 135- Uberlândia e 136 – Patos de Minas). No Espírito Santo, foi registrada apenas na capital capixaba – ponto 190 – e, no Rio de Janeiro, apresentou-se nos pontos 193 – Itaperuna, 197 – Nova Friburgo, 200 – Petrópolis e 204 – Arraial do Cabo.

A variante *moleca*, por sua vez, tem a sua área de ocorrência mais restrita, pois foi registrada apenas em São Paulo e Minas Gerais, presente em sete municípios paulistas e em três cidades mineiras. A baixa produtividade de *moleca* nas localidades do Sudeste justifica-se pelos relatos dos informantes, como o transcrito *ipsis litteris*, a seguir¹⁶:

INF.- Minina [mi' ninø].

INQ.-Tem mais algum além de menina?

INF.- Baiano fala muléca, né.

Além da identificação diatópica sobre o uso da variante *moleca* nas localidades do Sudeste, pertencente, portanto, conforme o relato de alguns informantes a outras localidades, o emprego dessa variante está atrelado também

ao caráter comportamental da criança. Recuperando as acepções 3 e 4 do verbete *moleca* constantes de Houaiss (2001) encontra-se a seguinte definição: “3 garota dada a travessuras; 4 menina brincalhona” (HOUAISS, 2001). Essa informação é ratificada no discurso de um informante¹⁷:

INF.- Menina [me'ninɐ].

INQ.- Tem algum outro nome que vocês chamam?

INF.- Garota [ga'rotɐ], muleca [mu'lekɐ].

INQ.- Qual que é o mais comum?

INF.- Se a menina fô bem levadinha, chama de muleca. Mai' quietinha chama de menina mesmo.

Moleca trata-se de uma forma derivada da variante *moleque*, de étimo africano. Assim, por extensão, criou-se, no séc. XX, essa forma lexical no feminino equivalente ao masculino.

Observa-se também, analisando os relatos de mais alguns informantes, que o uso da forma lexical *moleca* pode evidenciar diferenças diastráticas. Utilizando-se do método de testagem, no trecho apresentado a seguir, o inquiridor não obtendo a variante *moleca* pergunta ao informante se ele a utiliza. Como resposta obtém o seguinte discurso¹⁸:

INQ.- Moleca chama?

INF.- Chama moleca também, pessoal da roça gosta de pôr esses nome, às vez não é brigano, o modo das pessoa falá, falá isso

Possivelmente, pode-se justificar também a baixa incidência dessa variante, no *corpus* analisado, a esse caráter diastrático empregado por alguns dos informantes no uso dessa forma lexical. Se *moleca* é uma forma típica para nomear o referente pelas pessoas de origem rural, é evidente que, no *corpus* do Projeto ALiB, a produtividade deste item será menor, haja vista que o perfil dos informantes do atlas nacional é, fundamentalmente, urbano.

Ainda no Sudeste, analogamente a essa situação, encontra-se a distribuição diatópica da variante *guria*, presente em seis pontos linguísticos do estado de São Paulo e em duas localidades mineiras. O uso dessa variante também evidencia diferenças diastráticas conforme se observa na fala de um dos informantes¹⁹:

INF.- Ah, a gente fala minina [mi'niɲə], tem outros fala gurizinha [guri'ziɲə], né. O povo... mais caipira, mais assim, eles falam... usam muito.

INQ.- Ah é?

INF.- É?

INQ.- Qual que usa muito?

INF.- É né, assim, falá gurizinho né, se é minino fala gurizinho, e se é minina, gurizinha, né, o povo mais caipira eles usam muito, né.

Depreende-se, desse contexto, que a informante se refere ao 'povo caipira' como as pessoas pertencentes à zona rural. Dessa forma, os caipiras, portanto, possivelmente, empregariam a variante *guria*, o que revela uma forma mais antiga de denominar o referente.

Observa-se a ocorrência da variante *guria*, além de pontos isolados em São Paulo como em Araçatuba (ponto 156) e Minas Gerais (pontos 137 e 138 - Campina Verde e Belo Horizonte), em uma área antiga do estado, próxima ao vale do Ribeira, nos pontos 170, 181 e 178 (Bernardino de Campos, Itararé,

Sorocaba) e em duas localidades no litoral paulista, pontos 187 – Cananéia e 184 – Santos, onde alguns ainda empregam essa variante. Nota-se que o uso de *guria* está atrelado também à origem geográfica do informante, conforme se atesta no discurso a seguir:

INF.- Garota [ga'rotə], garoto [ga'rotu], guri [gu'ri],
guria [gu'riə].

INQ.- Mas dizem esses aqui também?

INF.- Não. Aqui fala, às vezes, garoto, garota.

INQ.- E o guri e a guria quem que fala?

INF.- No sul²⁰.

Assim, a ocorrência de *guria*, identificada como uma variante diatópica típica de gaúchos, em algumas cidades paulistas deve-se a influências vindas do Rio Grande do Sul, pois, conforme se observa, duas das cidades em que este item lexical ocorreu – Itararé e Sorocaba – foram pontos de passagem dos antigos Caminhos das Tropas, pela consequência do Tropeirismo no Paraná²¹. Nesse sentido, o primeiro autor desta pesquisa, em sua tese de doutoramento (em andamento), pautando-se em dados do Projeto ALiB, vem buscando respaldo na sócio-história do português para interpretar os aspectos lexicais ocorridos em determinadas áreas, por exemplo, no estado de São Paulo.

Conforme se observa, embora não seja possível traçar uma área de isoléxica para as variantes no Sudeste, pois ocorrem em áreas descontínuas do território, muito do que se tem documentado revela aspectos sócio-históricos da formação do português. Os testes estatísticos empreendidos para algumas das variantes mais produtivas nessa região confirmam informações constantes das cartas linguísticas, pois ratificam a homogeneidade na distribuição espacial de determinadas variantes lexicais.

5.2.3 Região Sul

No Sul do Brasil, foram documentados 223 registros com diferentes índices de representatividade para cada uma das variantes. Assim, destacam-se duas variantes principais, *menina* e *guria*, conforme se observa na figura 6.

Contrariamente ao que ocorreu no Sudeste, a variante *menina* não predominou em todos os estados que compõem a região Sul, pois essa forma lexical obteve maior produtividade no Paraná (73,33%) e Santa Catarina (70%). No Rio Grande do Sul, por sua vez, *guria* se apresentou como a forma majoritária (48,19%).

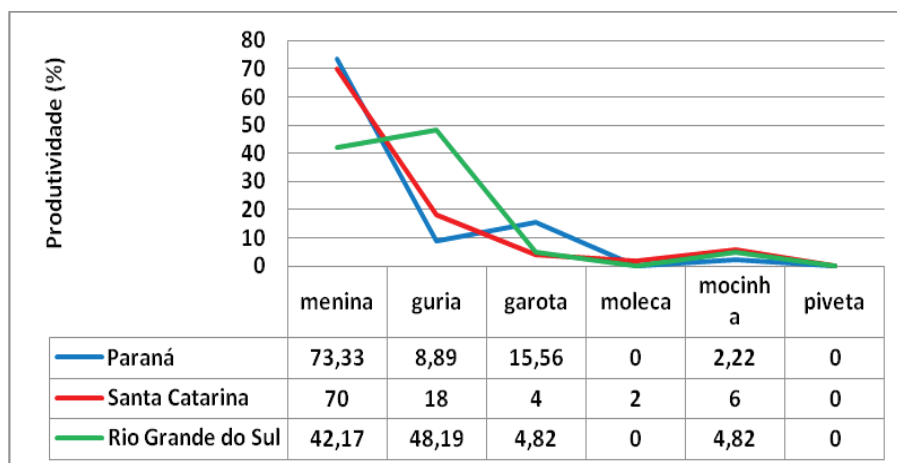


FIGURA 6 - Produtividade das variantes lexicais para a questão 133 do QSL nos estados da região Sul - Fonte: Banco de dados do ALiB (2013)

A figura 7 apresenta a distribuição diatópica das quatro principais variantes documentadas no *corpus*, excetuando-se a variante *menina*, pois, das 42 localidades consultadas, em apenas duas a referida variante não foi registrada: pontos 237 – Vacaria e 247 – Santana do Livramento, no estado do Rio Grande do Sul.

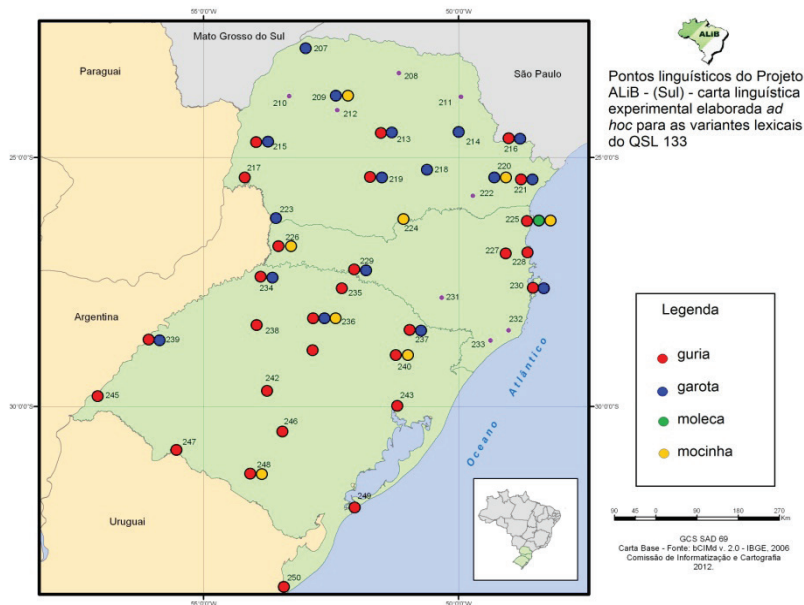


FIGURA 7 - Distribuição diatópica das quatro variantes lexicais mais produtivas na rede de pontos da região Sul - Fonte: Banco de dados do ALiB (2013)

Observa-se que, em oito pontos linguísticos do mapa, não há nenhuma das quatro variantes elencadas na legenda. Isso se deve ao fato de que nessas localidades houve a ocorrência exclusiva da variante *menina*. Assim, a figura 7 apresenta uma concentração da variante *guria* no Rio Grande do Sul, adentrando-se ao estado de Santa Catarina e ao centro-sul do estado do Paraná. O teste de homogeneidade de variâncias indicou que a hipótese 5 pode ser rejeitada ($\chi^2 = 20,808$; $df = 2$; $p\text{-valor} < 0,001$), ou seja, os resultados do teste evidenciam a heterogeneidade da distribuição da variante *guria* na rede de pontos da região Sul.

Os resultados estatísticos indicaram ainda que a distribuição diatópica da variante *guria* em seu sentido setentrional diminui à medida que adentra ao estado de Santa Catarina (18%), obtendo representatividade ainda menor no Paraná (8,89%). Nesse último, observa-se que *guria* não ocorreu no norte do estado, confirmando a possível influência linguística do estado de São Paulo no norte paranaense. Tal prognóstico é ratificado pela concentração de *garota*

no estado do Paraná, amplamente distribuída pelo território paulista. No Paraná esta variante obteve 15,56% de produtividade (figura 6), ao passo que em Santa Catarina e Rio Grande do Sul apresentou-se com baixo percentual, 4% e 4,82%, respectivamente.

Confirma-se, portanto, sob a perspectiva lexical com base nos dados analisados e nos testes estatísticos empreendidos, a hipótese já sustentada por Koch (2000) e Altenhofen (2005) acerca da divisão dialetal da região Sul em duas grandes áreas linguísticas, a do norte (denominada como paranaense) e a sul (a rio-grandense). Estas duas grandes áreas estão separadas por um feixe intermediário, no estado de Santa Catarina, denominado por Koch (2000) como *leque catarinense*. Nos dados analisados, é evidente a presença das duas áreas: a do norte, caracterizada pela produtividade de *garota* e, a do sul, com o aumento da ocorrência da variante *guria*. O estado de Santa Catarina, neste caso, ora se identifica com o Rio Grande do Sul (com o emprego de *guria*) ora com o Paraná (uso de *mocinha* e *garota*).

Conforme relatos de alguns informantes, identifica-se que o uso do item lexical *guria*, no estado do Paraná, é uma forma que caminha para o desuso, sendo substituída, progressivamente, por *menina* ou *garota*. Como exemplo, apresenta-se o relato de um informante de Curitiba²²:

INQ.- E se for do sexo feminino nós vamos dizer que é...?

INF.- Menina [me'ninɐ], ou a guria [gu'riɐ].

INQ.- Falam guria também aqui em Curitiba?

INF.- Guria. Falavam muito. Hoje a gente quase num vê mais essa pronúncia (=pronúncia), mais... falava-se muito "uma guria!" Não é menina, é uma guria. Como guri, no caso. Seria eh... mema coisa. Guri e a guria.

Essa variante, portanto, para esse informante é identificada como uma forma diacrônica, utilizada antigamente para denominar a criança do sexo feminino, não empregada com tanta frequência na atualidade.

Observa-se também, analisando os relatos dos informantes, outra variante diacrônica que não se utiliza mais, como a do relato do informante apresentado a seguir²³:

INQ.- E se for uma menina de cinco a dez anos do sexo feminino... ?

INF.- É menina [mi'ninɐ]. Daí esse num tinha, porque o menino que vinha c'a história de que piá... e o... a menina era menina memo, daí num tinha outro jeito. Não, lembrei, (inint.) bem mais antigo. Num é bem meu tempo, mais eu ouvia falá. Sabe como é que diziam pr'uma menina? Daí tem, a senhora tá certa em fazê a pergunta, lembrei, lembrei, se num mexê num vai. Então tinha o... o do sexo masculino que era, no caso, o piá, né, qu'eles chamava, e a menina era piona, agora lembrei a piona. Que daí a piona da comadre, "chama a piona da comadre lá." Piona [pi'onɐ]. Piá, piona, era assim que chamava a menina, né, piona.

INQ.- E hoje em dia se ouve?

INF.- Não, nunca mais ovi isso mais não. Hoje o piá ainda falam muito: "Óh, o piá do fulano" esse ficô, né, mais a piona não existe, qu'eu saiba não.

A forma lexical *piona*, derivada do correlato no masculino *piá*, não é uma variante típica do vocabulário ativo dos informantes do projeto ALiB e, nesse contexto, pode ser identificada como uma variante diacrônica. O Atlas Linguístico do Paraná, (AGUILERA, 1994; ALTINO, 2007) não contempla esta pergunta em seu questionário, não possibilitando a comparação com o *corpus* de informantes rurais da década de 80. Também o Atlas Linguístico-Etnográfico da Região Sul (ALTENHOFEN *et al.*, 2011) na carta linguística 272 – Menina – não documenta esta variante, o que não possibilita fazer

comparação entre dados rurais (ALERS) e urbanos (ALiB). Dessa forma, pode-se concluir que o uso de *piona* restringe-se mais ao idioleto²⁴ do informante do que se configura como uma marca dialetal da comunidade linguística como um todo.

6 CONSIDERAÇÕES FINAIS

O tratamento estatístico em dados geolinguísticos torna-se uma importante abordagem para a formulação de hipóteses que extrapolam possíveis inferências pautadas somente em dados percentuais. Assim, embora não se utilize o *corpus* do Projeto ALiB em sua totalidade, esse estudo comprova, por meio de testes, que há diferenças linguísticas entre os estados e as regiões investigadas.

O resgate das principais variantes em obras lexicográficas (históricas e contemporâneas) possibilitou verificar a introdução dos itens na língua portuguesa bem como comparar diferentes acepções. Ademais, evidenciou as influências de outras línguas no processo de formação do português, como a indígena, no caso de *guria*, africana, para *molequel/moleca*, e espanhola, para *pivete/piveta*.

As análises estatísticas empreendidas no estudo evidenciam que as regiões não apresentam comportamento homogêneo quanto à distribuição diatópica das variantes *menina*, *guria*, *garota* e *moleca*. De acordo com os resultados, somente a forma lexical *mocinha* apresentou distribuição homogênea nas regiões administrativas investigadas. Dessa forma, a variante mais difundida na norma lexical do português brasileiro para a questão analisada, considerando o *corpus* investigado, é *menina*, registrada em quase todos os pontos linguísticos. ,

A pesquisa evidenciou que, no Centro-Oeste, há diferenças linguísticas entre o Mato Grosso do Sul e o Mato Grosso em relação ao estado de Goiás. Nos dois primeiros, além da predominância da variante *menina*, há um índice considerável de registros da variante *guria*, típica da região Sul, o que ressalta a semelhança desses estados em alguns aspectos com o Rio Grande do Sul devido a possíveis fluxos migratórios. Em Goiás, por sua vez, o índice de ocorrência da variante *guria* é menor, evidenciando as diferenças linguísticas deste estado em relação aos outros dois. Assim, essas diferenças decorrem, possivelmente, de fatores sócio-históricos no processo de ocupação e povoamento do território.

A variante *garota* difunde-se pelas três regiões, constituindo-se como a segunda mais produtiva na região Sudeste, amplamente distribuída no estado de São Paulo e Minas Gerais, com a homogeneidade estatisticamente

comprovada por meio dos testes. Essa forma lexical, no entanto, apresenta baixa produtividade em Santa Catarina e Rio Grande do Sul.

A ocorrência esparsa das variantes lexicais no Centro-Oeste e Sudeste não permite o traçado de linhas de isoléxicas. Cabe notar, porém, que as variantes *moleca* e *mocinha* encontram-se diatopicamente marcadas em Goiás e em cidades limítrofes a esse estado.

Por fim, na região Sul, apesar da maciça presença de *guria* em Santa Catarina e, principalmente, no Rio Grande do Sul, os resultados dos testes evidenciam a heterogeneidade da distribuição dessa variante nessa região. À medida que *guria* avança em direção ao Paraná, a sua produtividade diminui, não ocorrendo no norte desse estado, o que ratifica a possível influência do falar paulista no norte do Paraná. O trabalho confirma também, sob o ponto de vista lexical, as hipóteses de divisão dialetal da região Sul em duas grandes áreas (KOCH, 2000; ALTENHOFEN, 2005), a paranaense e a rio-grandense, intermediadas por uma faixa de transição localizada no território de Santa Catarina.

NOTAS

¹ Pesquisador do Projeto Atlas Linguístico do Brasil (regional PR). Mestre em Estudos da Linguagem pela Universidade Estadual de Londrina. Doutorando em Estudos da Linguagem nesta mesma instituição. Tem experiência em pesquisa nas áreas de Sociolinguística, Dialetoлогия e Geolinguística.

² Doutor em Ciências pela Universidade de São Paulo, é Professor Adjunto A no Instituto de Matemática e Computação da Universidade Federal de Itajubá. Tem experiência em pesquisa nas áreas de Informática na Educação, Engenharia de Software e Pesquisa Experimental.

³ Nomenclatura utilizada pelo pesquisador Thun (1998) para se referir às diferentes dimensões da variação linguística.

⁴ “Los atlas lingüísticos se distinguen de los atlas geográficos corrientes también porque no contienen mapas de vários territorios, sino una serie de mapas del mismo territorio, y precisamente un mapa para cada concepto o para cada fonema (o serie de fonemas) cuya expresión o cuya realización concreta se han comprobado por el investigador en una red de puntos (localidades) previamente establecida. (...) En otras palabras, los atlas lingüísticos son colecciones cartográficas de material lingüístico.” (COSERIU, 1991, p.111-112)

⁵ O Projeto Atlas Linguístico do Brasil, doravante ALiB, é um projeto de pesquisa nacional e interinstitucional, com sede na Universidade Federal da Bahia, sob a

presidência da Dr^a. Suzana Alice Marcelino Cardoso e dirigido por um Comitê Nacional formado por pesquisadores de universidades de oito estados. O Projeto ALiB visa a descrever a variante brasileira da língua portuguesa nos níveis fonético-fonológico, semântico-lexical e morfossintático, por meio da aplicação de questionários a informantes de todo território nacional, para, assim, formar um banco de dados que, posteriormente, será utilizado para a elaboração de um Atlas Linguístico de âmbito nacional. Mais informações podem ser obtidas em: <http://twiki.ufba.br/twiki/bin/view/Alib/WebHome>

⁶ Não consideramos as variantes que apresentaram ocorrência única no *corpus*: *filha*, *pirralha* e *mulherzinha*.

⁷ Nas capitais, foram entrevistados também quatro informantes de nível superior. Não nos atemos, neste estudo, à análise considerando as variáveis sociolinguísticas, pois o foco principal do trabalho é a distribuição diatópica.

⁸ Faz parte ainda da rede de pontos do ALiB mais seis localidades mineiras (Ipatinga, Ouro Preto, Viçosa, São João Del Rei, Muriaé e Juiz de Fora) cujos inquéritos não foram finalizados até a elaboração deste trabalho, portanto, não fazem parte do *corpus* analisado.

⁹ No estado do Rio de Janeiro, a coleta de dados não foi concluída em cinco localidades: São João da Barra, Campos dos Goytacazes, Três Rios, Nova Iguaçu e Parati, não integrantes do *corpus* em análise.

¹⁰ Não constam do *corpus* os dados referentes a Osório e Santa Cruz do Sul também pelo fato de que nessas localidades a coleta de dados ainda não foi concluída.

¹¹ *Software* de análise estatística amplamente utilizado em pesquisas experimentais aplicadas nas mais variadas áreas do conhecimento humano.

¹² Valor estatístico utilizado para sintetizar o resultado de um teste de hipóteses.

¹³ Em Minas Gerais, três mesorregiões não são contempladas pela rede de pontos a que tivemos acesso: Central Mineira, Zona da Mata e Vale do Rio Doce. No estado do Rio Grande do Sul, apenas a mesorregião Centro Oriental Rio-grandense não é contemplada.

¹⁴ A variante *piveta* também não foi considerada na cartografia, pois ocorreu somente no ponto 108/3 – Cuiabá.

¹⁵ Uma linha de isoléxica é um tipo de isoglossa de base lexical. “Por isoglossa entende-se uma linha virtual, que marca o limite, também virtual, de formas e expressões linguísticas (...) uma isoglossa pode ser lexical, ou seja, isoléxica; pode ser fônica, isófono; pode ser morfológica, isomorfa e pode ser sintática.” (FERREIRA; CARDOSO, 1994, p. 13)

¹⁶ Transcrição grafemática e fonética. As siglas INQ. e INF. indicam os turnos da conversação do inquiridor e do informante, respectivamente. Esse trecho refere-se ao

discurso do informante 3 de Jales/SP – ponto 150 – (homem pertencente à segunda faixa etária - 50 a 65 anos - com ensino fundamental).

¹⁷ Esse trecho refere-se ao discurso do informante 1 de Cananéia/SP – ponto 187 – (homem pertencente à primeira faixa etária - 18 a 30 anos - com ensino fundamental).

¹⁸ Esse trecho refere-se ao discurso do informante 3 de Januária/MG – ponto 127 – (homem pertencente à segunda faixa etária - 50 a 65 anos - com ensino fundamental).

¹⁹ O trecho refere-se ao discurso da informante 2 de Bernardino de Campo/SP – ponto 170 – (mulher pertencente à primeira faixa etária - 18 a 30 anos - com ensino fundamental).

²⁰ Trecho referente ao discurso da informante 4 de Franca/SP – ponto 154 (mulher pertencente à segunda faixa etária – 50 a 65 anos – com ensino fundamental).

²¹ O chamado Caminho das Tropas consistiu em uma antiga via terrestre de acesso ao Rio Grande do Sul, na época do Brasil colônia usado principalmente para o transporte de muares. O Caminho do Viamão era o mais utilizado, partia de Viamão-RS e atravessava os campos do Paraná adentrando ao estado de São Paulo, passando por Itararé, Itapetininga até o destino final, Sorocaba.

²² Trecho referente ao discurso da informante 3 de Curitiba/PR – ponto 220 (homem pertencente à segunda faixa etária – 50 a 65 anos – com ensino fundamental).

²³ Trecho referente ao discurso da informante 3 de Cândido de Abreu/PR – ponto 213 (homem pertencente à segunda faixa etária – 50 a 65 anos – com ensino fundamental).

²⁴ Por idioleto entende-se: “conjunto de usos de uma língua própria de um indivíduo, num momento determinado” (DUBOIS *et al*, 1999, p. 329).

REFERÊNCIAS

AGUILERA, V. de A. *Atlas Linguístico do Paraná*. Curitiba: Imprensa Oficial, 1994.

AGUILERA, V. de A. De onde vieram e por onde andam as nossas libélulas e jacintas? Um estudo da etimologia popular com base em dados do Atlas Linguístico do Brasil (ALiB). *Estudos Linguísticos e Literários*. n. 41. Salvador: Programa de Pós-graduação em Língua e Cultura da Universidade Federal da Bahia, p. 291-309, 2010.

AGUILERA, V. de A.; ALTINO, F. C.; ISQUERDO, A. N. *Atlas Linguístico do Brasil: descrevendo a língua, formando jovens pesquisadores II*. CD vol.2. Londrina: UEL, 2012.

ALTENHOFEN, C. Áreas linguísticas do português falado no sul do Brasil: um balance das fotografias geolinguísticas do ALERS. In.: AGUILERA, V. de A.(org.) *A geolinguística no Brasil: trilhas seguidas, caminhos a percorrer*. Londrina: Eduel, 2005, p. 177-208.

- ALTENHOFEN, C. V.; KLASSMANN, M. S. (orgs.). *Atlas Linguístico-Etnográfico da Região Sul do Brasil – ALERS: cartas semântico-lexicais*. Porto Alegre: Editora da UFRGS; Florianópolis: Ed. UFSC, 2011.
- ALTINO, F. C. *Atlas Linguístico do Paraná II*. 2007. 223 p. Tese. (Doutorado em Estudos da Linguagem) – Universidade Estadual de Londrina, Londrina, 2 v.
- ARAGÃO, M. do S. Silva de; BEZERRA DE MENEZES, C. *Atlas Linguístico da Paraíba*. Brasília: UFPB; CNPq, Coordenação Editorial, 1984.
- BESSA, J. R. F. (coordenador). *Atlas Linguístico do Ceará*. Universidade Federal do Ceará. Fortaleza: Edições UFC, 2010.
- BLUTEAU, R. *Vocabulario portuguez & latino: aulico, anatomico, architectonico*. Coimbra: Collegio das Artes da Companhia de Jesu, 1712 - 1728. 8 v. Disponível em: <http://www.brasiliana.usp.br/pt-br/dicionario/edicao/1> Acesso: 21 jun. 2013.
- CARDOSO, S. A. M. *Atlas Linguístico de Sergipe II*. Salvador: EUFBA, 2005.
- COMITÊ NACIONAL do Projeto ALiB. *Atlas Linguístico do Brasil: questionários 2001*. Londrina: UEL, 2001.
- COSERIU, E. *El hombre y su lenguaje*. 2. ed. Madrid: Editorial Gredos, 1991.
- CRUZ, M. L. de C. *Atlas linguístico do Amazonas*. 2004. 2 v. Tese (Doutorado em Língua Portuguesa) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- CUNHA, A. G. *Dicionário Etimológico da Língua Portuguesa*. 4 ed. Rio de Janeiro: Lexikon, 2010.
- DUBOIS, Jean et al. *Dicionário de Lingüística*; trad. Frederico Pessoa de Barros [et al]. SP: Cultrix, 1999.
- DIÉGUES Jr. M. *Regiões Culturais do Brasil*. v. 2, série V. Rio de Janeiro: INEP/ Centro Brasileiro de Pesquisas Educacionais, 1960.
- FERREIRA, A. B. de H. *Novo dicionário da língua Portuguesa*. 3. ed., Versão eletrônica 5.0, 2004.
- FERREIRA, C. et al. *Atlas linguístico de Sergipe*. Salvador: Universidade Federal da Bahia; Fundação de Cultura de Sergipe, 1987.
- FERREIRA, C.; CARDOSO, S. *A Dialetoлогия no Brasil*. São Paulo: Contexto, 1994.
- FREITAS MARINS, L. G. *O rural e o urbano: novos e velhos falares na região Centro-Oeste do Brasil*. 2012. 310 p. Dissertação (Mestrado em Estudos de Linguagem) – Universidade Federal do Mato Grosso do Sul, Campo Grande, 2012.
- GALVEZ, J. A. [ed.] *Dicionário Larousse Francês/Português, Português/Francês*. 1.ed. São Paulo: Larousse do Brasil, 2005.

HOUAISS, A.; VILLAR, M. de S. *Dicionário eletrônico Houaiss da língua portuguesa*. Versão 1.0, 2001.

ISQUERDO, A. N. Designações para estilingue em atlas lingüísticos brasileiros: perspectivas diatópica e sócio-histórica. In.: TROTTER, David (Ed.) *Actes du XXIVe Congress International de Linguistique et de Philologie Romanes*. Tome I. Tübingen: Max Niemeyer Verlag, 2007. p. 533-546.

ISQUERDO, A. N. Léxico em tempo e espaço: a questão dos Regionalismos. In.: MARIN, J. R.; VASCONCELOS, C. A. (orgs.). *História, região e identidades*. Campo Grande: Ed. UFMS, 2003, p.165-181.

ISQUERDO, A. N. O caminho do rio, o caminho do homem, o caminho das palavras. In.: RIBEIRO, S. S. C.; COSTA, S. B. B.; CARDOSO, S. A. M. (orgs.) *Dos sons às palavras: nas trilhas da língua portuguesa*. Salvador: EDUFBA, 2009, p. 41-59.

KOCH, W. O povoamento do território e a formação de áreas linguísticas. In.: GÄRTNER, E.; HUNDT, C. SCHÖNBERGER, A. (eds.) *Estudos de geolinguística do português Americano*. Franckfurt am Maim: TFM, 2000, p.55-69.

KOCH, W.; ALTENHOFEN, C. V.; KLASSMANN, M. S. (orgs.). *Atlas Linguístico-Etnográfico da Região Sul do Brasil – ALERS: cartas fonéticas e morfossintáticas*. 2.ed. Porto Alegre: Editora da UFRGS; Florianópolis: Ed. UFSC, 2011.

MACHADO, J. P. *Dicionário etimológico da língua portuguesa*. 2 Tomos. Lisboa, 1952.

MORAES SILVA, A. *Dicionário da Língua Portuguesa*. 2. ed. Lisboa: 1813.

NASCENTES, A. *O Linguaajar Carioca*. 2. ed. Rio de Janeiro: Organizações Simões, 1953.

OLIVEIRA, A. J de. Análise quantitativa no estudo da variação linguística: noções de estatística e análise comparativa entre Varbrul e SPSS. *Revista de Estudos da Linguagem*. v.17, n.2, 2009, p-93-119.

OLIVEIRA, D. G. de (org.) *ALMS – Atlas Linguístico do Mato Grosso do Sul*. Campo Grande: Editora UFMS, 2007.

PAIM, M. M. T. A variação lexical nos campos semânticos corpo humano e ciclos da vida: o que revelam os dados do Projeto ALiB. *Diadorim: Revista de Estudos Linguísticos e Literários*, n. 8. Rio de Janeiro: UFRJ, 2011, p. 143-159.

PINTO, L. M. da S. *Diccionario da Língua Brasileira*. Ouro Preto: Typographia da Silva, 1832. Disponível em: <<http://www.brasiliana.usp.br/pt-br/dicionario/edicao/3>> Acesso em: 22 de jun. 2013.

RAZKY, A. *Atlas linguístico sonoro do Estado do Pará* (ALiSPA 1.1). Belém: s/ed. 2004 (Programa em CD-ROM).

- RIBEIRO, J. et al. *Esboço de um atlas linguístico de Minas Gerais*. Rio de Janeiro: Casa de Rui Barbosa, 1977.
- RIBEIRO, S. S. C. *Brinquedos e brincadeiras infantis na área do falar baiano*. 2012, 466 p. Tese de Doutorado (Programa de Pós-graduação em Letras e Linguística). Universidade Federal de Bahia.
- ROMANO, V. P.; AGUILERA, V. de A. Distribuição diatópica para as variantes para *tangerina*: um estudo geo-sociolinguístico.: In.: ISQUERDO, A. N.; ALTINO, F. C.; AGUILERA, V. de A. (orgs.) *Atlas linguístico do Brasil*: descrevendo a língua, formando jovens pesquisadores. CD vol1. Londrina: UEL, 2009, p. 148-157.
- ROMANO, V. P.; SEABRA, R. D. Menino, guri ou piá? Um estudo diatópico nas regiões Centro-Oeste, Sudeste e Sul a partir dos dados do Projeto Atlas Linguístico do Brasil. *Alfá: Revista de Linguística* (no prelo).
- ROSSI, Nelson et al. *Atlas Prévio dos Falares Baianos*. Rio de Janeiro: INL, 1963.
- SILVA, L. M. da; AGUILERA, V. de A. Variantes lexicais para o *gambá*: um estudo geolinguístico nas capitais brasileiras. In: *Anais do 1º CIELLI – Colóquio Internacional de Estudos Linguísticos e Literários*. Maringá: Universidade Estadual de Maringá, 2010. p. 1-13.
- SILVA-COSTA, D. de S.; ISQUERDO, A. N. Um estudo etnolinguístico de designativos para ‘gambá’ no Brasil Central: contribuições do Projeto ALiB. *Estudos Linguísticos*. N. 41.2, 2012, p. 779-792.
- SPSS for Windows. Version 17.0. IBM, 2008. CD-ROM.
- THUN, H. La géographie linguistique romane à la fin du XX siècle. In.: RAENDONCK, D. V. et all. Orgs. *Actes du XXII Congrès International de Linguistique e Philologie Romanes*. Bruxelles, 1998, 367-409.
- VIEIRA, S. *Análise de Variância: Anova*. 1. ed., São Paulo: Atlas, 2006.
- YIDA, V. *O campo semântico da alimentação e cozinha no Atlas Linguístico do Brasil*: um estudo lexical nas capitais. 2011, 191 p. Dissertação (Mestrado em Estudos da Linguagem) – Universidade Estadual de Londrina, Londrina, 2011.

Submissão do artigo: 18/05/2013

Aprovação do artigo: 30/09/2013