

PERCURSO TEÓRICO-METODOLÓGICO DE UMA PESQUISA GEOLINGÜÍSTICA

Lígia Maria Campos Imaguire¹

RESUMO

PALAVRAS-CHAVE: (Geolingüística, cartas, sujeitos, questionário, gravações)

No final do século XIX e nas primeiras décadas do século XX surge o método cartográfico da geografia lingüística, ou da chamada geolingüística, que tem sido amplamente utilizado para a descrição das variantes de uma língua, distinta em termos sociais ou regionais e identificadas por um conjunto particular de palavras e estruturas gramaticais.

Esse método, de acordo com Coseriu (1982:79), pressupõe o registro em mapas especiais de um número elevado de formas lingüísticas (fônicas, lexicais ou gramaticais) ou em outras palavras, constitui um acervo de diferentes realizações formado por diversas normas comprovadas mediante pesquisa direta e unitária numa rede de pontos de um determinado território.

Um mapa/carta retrata as migrações de certos fenômenos lingüísticos horizontalmente, como convivem e se sobrepõem, sendo dessa forma possível reconstituir estados antigos de uma determinada língua, que podem remontar até o idioma original.

A comunicação que iremos apresentar faz parte de uma pesquisa desenvolvida no curso de pós-graduação em Semiótica e Lingüística Geral na Universidade de São Paulo: Estudo geolingüístico de oito municípios do litoral sul paulista, são eles: Peruíbe, Itanhaém, Mongaguá, Praia Grande, São Vicente, Cubatão, Guarujá e Bertioga. Essa pesquisa teve como um dos objetivos elaborar um estudo de aspectos semântico-lexicais.

Faremos uma breve exposição dos aspectos teóricos-metodológicos empregados no trabalho, salientando os pontos lingüísticos, a busca e a seleção dos sujeitos/informantes, o questionário lingüístico, a entrevista, as gravações, a análise do *corpus*, a forma de transcrição, a criação e o manuseio do banco de dados, a preparação das tabelas e as cartas lingüísticas.

¹ Lígia Maria Campos Imaguire. Colégio Santa Cruz. Endereço: Rua Carlos Weber, 457, ap.82 A São Paulo – SP – Brasil. 05303-000 - Endereço eletrônico: ligia@filosofos.com.br

Métodos e procedimentos

A seleção dos pontos lingüísticos

Fizemos pesquisa direta (*in loco*), isto é, fomos a campo coletar os dados. Selecionamos, dentro do território paulista, uma região administrativa, incluindo aí um ponto de entrevista do ALiB. Optamos pela região de Santos.

Para o estabelecimento dos pontos, levamos em consideração a densidade demográfica, o processo de povoamento, os aspectos sociopolítico e econômico, a importância de cada município, além da equidistância entre eles. Fizemos pesquisa em toda a região administrativa de Santos. Quando possível, iniciamos a pesquisa no primeiro núcleo populacional (bairro) existente no município e seguimos, de forma concêntrica, até o bairro mais recente; esse procedimento foi utilizado quando o número de sujeitos entrevistados nos permitiu fazer esse percurso.

Excluimos de nossa pesquisa apenas o município de Santos, onde nossa orientadora pretende fazer futuramente uma pesquisa sobre a fala local.

Na região administrativa de Santos foram objeto de pesquisa os sujeitos dos seguintes pontos lingüísticos:

Pontos	Descrição dos pontos
01	Peruíbe
02	Itanhaém* ²
03	Mongaguá
04	Praia Grande
05	São Vicente
06	Cubatão
07	Guarujá
08	Bertioga

A seleção dos sujeitos

No capítulo referente à história dos municípios pesquisados, notamos que muitas foram as influências socioculturais, raciais e lingüísticas que compuseram o perfil dos habitantes da região pesquisada.

Para que pudéssemos observar detalhadamente algumas variáveis lingüísticas diatópicas, diastráticas, diafásicas, diageracionais e diagenérica, procedemos ao estudo

² O município assinalado com asterisco é ponto também do ALiB.

do perfil do sujeito que constituiu a amostra lingüística da rede de pontos pesquisada.. Esses dados sobre os sujeitos são apresentados em uma tabela na tese (no capítulo “Métodos e procedimentos”) e também no banco de dados, do qual constam: número, correspondente à ordem das entrevistas e à ordem no *corpus*; o ponto, designando os municípios, apresentados por números na ordem em que foram visitados: 1-Bertioga, 2 - Cubatão, 3-Guarujá, 4-Itanhaém, 5-Mongaguá, 6-Peruíbe, 7-Praia Grande, 8-São Vicente; as iniciais do nome completo do sujeito; o sexo: F para feminino e M para masculino; a idade, expressa em números; o estado civil; as profissões; e o salário (em salários mínimos no período em que realizamos a nossa pesquisa o valor referencial do salário mínimo era de R\$ 181,00, ou U\$ 75,42 valor cambial do dólar em U\$ 2,40 – novembro de 2001).

Na seleção dos sujeitos, utilizamos os mesmos critérios que estão sendo usados na elaboração do Atlas Lingüístico do Brasil – ALiB: ser natural da localidade ou aí ter residido 1/3 (um terço) de sua vida, quando procedente de outra localidade. Quanto à escolaridade, o nível exigido foi o fundamental 1 completo. Foram entrevistados adultos nas seguintes faixas etárias: 18-30, representando os mais jovens (normalmente utilizam variantes mais inovadoras), 31-49 (essa faixa etária o ALiB não utilizou), representando a faixa etária intermediária (às vezes essa faixa etária mantém um padrão mais conservador de linguagem, normalmente de maior prestígio, outras vezes utiliza variantes mais inovadoras) e 50-65, representando os mais velhos (lembrando que as pessoas mais velhas, normalmente, são usuárias de variantes mais conservadoras).

Para a seleção do número de sujeitos entrevistados em cada ponto, utilizamos a seguinte equação:

$$\text{Número de entrevistados} = [(\text{população}/150.000)+1] \times 3 \text{ faixas etárias} \times 2 \text{ sexos}$$

Equação matemática utilizada para a escolha do número dos sujeitos

Cidade	População	div./150000	arredondado+1	x3 faixas etárias x2 sexos
Bertioga	30.903 ³	0,206	1,21	6
Cubatão	107.904	0,719	1,72	6
Guarujá	265.155	1,768	2,77	12
Itanhaém	71.947	0,480	1,48	6
Mongaguá	35.106	0,234	1,23	6
Peruíbe	51.384	0,343	1,34	6

³ Dados pesquisados no *site* do IBGE – www.ibge.gov.br, disponível em 22/10/2000.

Praia Grande	191.811	1,279	2,28	12
São Vicente	302.678	2,018	3,02	18
Total	1.056.888	7,047	15,05	72

Em Bertioiga, Cubatão, Itanhaém, Mongaguá e Peruíbe foram entrevistados seis sujeitos, três de cada sexo e dois em cada faixa etária. Na Praia Grande e no Guarujá foram entrevistados doze sujeitos, seis de cada sexo e quatro em cada faixa etária. Em São Vicente foram entrevistados dezoito sujeitos, nove de cada sexo e seis em cada faixa etária. No total foram entrevistados 72 sujeitos.

Devido à exigüidade do tempo, não contemplamos todas as variáveis; priorizamos o gênero, a faixa etária, o nível de escolaridade e a situação socioeconômica.

O instrumento da coleta: o questionário lingüístico

De acordo com (Santos, 2003) “Desde o surgimento da Geolingüística, o questionário tem sido um instrumento largamente utilizado para a coleta de dados/experimentação (acréscimo nosso). Na verdade, não existe apenas um instrumento para a coleta de dados em Geolingüística. O questionário semântico-lexical, objeto do presente trabalho, consiste num conjunto de questões de cunho descritivo, cujo objetivo se constitui em investigar a designação atribuída pelo entrevistado, sujeito da pesquisa, a determinados objetos do mundo referencial ou imaginário previamente selecionados”.

Configura-se como um instrumento de cunho onomasiológico, pois se parte do ‘significado (‘conceito’ ou ‘noção’) para estudar suas manifestações no plano dos signos ou lexemas” (Greimas; Courtés, 1993, p. 261). Dito em outras palavras, faz-se uma pergunta padrão, que consiste na descrição do objeto do mundo referencial ou imaginário, com vistas a se obter o(s) lexemas(s) utilizado(s) em um determinado ponto. Segmentado em domínios, o questionário é aplicado em uma entrevista. As respostas emitidas pelos entrevistados são, em geral, lexemas ou frases breves. Ocasionalmente, há diálogos mais extensos, constituídos de vários turnos para uma mesma questão. Originam-se de uma indagação do entrevistados quanto à existência de outra resposta para um objeto do mundo referencial ou imaginário, à complementação de um dado ou a uma explicação factual. Na maioria dos atlas lingüísticos, os lexemas correspondentes às respostas a uma dada questão são o alvo da pesquisa. Conseqüentemente, são

destacados e dispostos em listas de palavras e, mais recentemente, em tabelas (apresentado no trabalho de Imaguire, 1999 — acréscimo nosso). Posteriormente, as listas e tabelas dão origem às cartas lexicais, que se caracterizam não apenas como base dos atlas lingüísticos, como também de inúmeras monografias dialetais”.

O questionário semântico-lexical (QSL) que utilizamos tem como base a versão constante do II Workshop de Preparação dos Inquiridores para o “Atlas Lingüístico do Brasil”, realizado em Londrina, de 3 a 7 de julho de 2000. Existe uma versão posterior a essa, porém, quando iniciamos nossa pesquisa piloto, era essa versão que estava em vigência. De acordo com o objetivo do trabalho, usamos apenas o questionário semântico-lexical, que é o instrumento básico para pesquisas dessa natureza, e é composto de 210 questões, distribuídas em 15 campos semânticos:

- acidentes geográficos (6 questões)
- fenômenos atmosféricos (14 questões)
- astros e tempo (17 questões)
- flora (6 questões)
- atividades agropastoris (22 questões)
- fauna (26 questões)
- corpo humano (32 questões)
- cultura e convívio (6 questões)
- ciclos da vida (19 questões)
- religião e crenças (9 questões)
- festas e divertimentos (19 questões)
- habitação (8 questões)
- alimentação e cozinha (10 questões)
- vestuário (6 questões)
- vida urbana (11 questões)

De acordo com o comitê do Projeto Atlas Lingüístico do Brasil (Projeto ALiB:14), a seleção dos itens inclusos no QLS levou em conta, além da orientação onomasiológica, o objetivo de documentar o registro coloquial do falante, buscando as formas de emprego mais gerais da comunidade, sem com isso priorizar regionalismos, arcaísmos ou linguagens especiais de grupos.

A necessidade de uniformização levou à formulação prévia de todas as questões, como pode ser observado no próprio questionário.

Assim, não se incluem, por exemplo, perguntas a respeito de flora, fauna, acidentes geográficos, costumes ou objetos característicos exclusivamente de determinadas regiões. As perguntas existentes no QSL são de caráter geral.

Estão inclusas, no entanto, perguntas referentes a formas que se revelaram de interesse do ponto de vista lexical nos atlas já publicados, quer pela natureza sinonímica que apresentam, quer pela indicação de áreas dialetais.

Esse questionário não pretende abarcar todas as variantes utilizadas em cada localidade, mas apresentar uma amostra do falar desse lugar.

A aplicação do questionário: a entrevista

A técnica utilizada normalmente para recolher os dados e garantir a homogeneidade do trabalho é a aplicação de um questionário lingüístico.

Além das entrevistas, gravamos também o diálogo que tínhamos com os entrevistados, no momento de preenchimento da ficha do sujeito e o diálogo pós-entrevista, em que os sujeitos contavam detalhes da vida deles. Trudgill (31-51:1983) propõe que se apliquem duas técnicas de coleta de dados: um questionário e gravações de elocuições livres. Para esta pesquisa usamos apenas o questionário, porque pretendemos fazer comparações com atlas e trabalhos da área de dialetologia e a análise de uma mesma lexia.

Acreditamos que, somente com elocuições livres, não teríamos elementos suficientes para o trabalho que pretendemos executar. De modo geral, gravamos também elocuições livres dos sujeitos, mas nesse momento não iremos trabalhar com esses dados.

As entrevistas foram feitas pela própria pesquisadora. Foram gravadas e realizadas no hábitat do sujeito, nas prefeituras, padarias, postos de saúde, em praças públicas, nas ruas, no local de trabalho de alguns sujeitos, com roteiro semelhante ao questionário do ALiB, visto que utilizamos apenas o questionário semântico-lexical (QSL). As entrevistas realizadas na casa dos sujeitos possibilitava a desinibição e a descontração, no entanto algumas vezes deparamos com barulhos domésticos.

Foram feitas pequenas alterações no questionário para que ele se adaptasse à realidade das regiões em que fizemos a pesquisa (região administrativa de Santos).

Transcrição dos dados

Transcrição é a cópia, a reprodução por escrito das entrevistas feitas na pesquisa de campo. O *corpus* desta pesquisa foi transcrito de forma grafemática e algumas lexias, de forma fonética. Fizemos transcrição fonética das lexias que apresentaram alofones. Para a transcrição grafemática, utilizamos as notações definidas por Marcuschi (1986) e, para a transcrição fonética, o Alfabeto Fonético Internacional (IPA).

Equipamentos técnicos utilizados

Para recolhermos os dados de nossa pesquisa, utilizamos um gravador da marca *Panasonic*, cassette, com microfone acoplado, movido por bateria, modelo nº RQ – 319. Foram usadas fitas cassette de marca Sony, tipo EF-X 60, *tipe* I normal, grande parte

cedidas à pesquisadora pelo Departamento de Lingüística da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

Para fazer a digitação, compilação e a elaboração do trabalho, utilizamos um PC-MDK62, 350Hzt, um scanner Gennius, uma impressora Epson Stylus Color 660 e um gravador de CD acoplado ao computador. O tempo de gravação das entrevistas variou de 60 min a 1h e 20 minutos.

A Análise dos dados

Após a gravação das entrevistas, seguimos as seguintes etapas: ouvimos as fitas para nos familiarizarmos com as gravações; transcrevemos as entrevistas de forma grafemática. No *corpus* não foram consideradas as variantes de gênero, número e grau.

Fizemos essa opção, pois, se colocássemos todas as variáveis, a pesquisa ficaria muito extensa; assim, decidimos arquivar esses dados para análises posteriores. Depois revisamos as transcrições, elaboramos o banco de dados, aprimoramos o banco de dados, digitamos os dados no banco, imprimimos as tabelas para conferência dos dados e revisão das tabelas, analisamos os campos semânticos individualmente, criamos, com base nas tabelas geradas pelo banco de dados, as tabelas: número de variantes, número de ocorrências pela variável sexo e abstenções de cada campo semântico. Com base nessas tabelas construímos, no programa Excel, os seguintes histogramas: variantes lexicais e ocorrências da variável sexo em cada um dos campos semânticos. As abstenções não foram consideradas em nenhuma contagem de variantes nem de ocorrências. Para analisar os campos semânticos e as perguntas que apresentaram o maior e o menor número de variantes criamos alguns tópicos: Em cada campo semântico e em cada pergunta analisaremos:

- o título do campo semântico ou a pergunta;
- a formulação das perguntas;
- a pergunta que apresentou o maior e o menor número de variantes;
- a pergunta que apresentou o maior número de ocorrências dentre as perguntas de um determinado campo;
- o número de ocorrências proferidas por homens e por mulheres em que perguntas os homens proferiram mais ocorrências e em quais as mulheres proferiram mais;
- se houve variantes comuns nos pontos lingüísticos, destacando esses pontos;

— o número de abstenções apresentado em cada campo, detalhando quando possível, os motivos.

Em seguida, mostraremos duas tabelas, uma com as perguntas do campo e o número de variantes e a outra com as perguntas do campo e as ocorrências, detalhando se foram proferidas por homens ou por mulheres. Separadamente mostraremos dois histogramas⁴, que complementam as tabelas mencionadas nesse parágrafo. A partir do histograma, selecionamos duas perguntas: a que apresentou o maior e a que teve o menor número de variantes lexicais no *corpus* da pesquisa.

Na análise das perguntas, como já salientamos anteriormente, em alguns casos, tecemos comentários sobre elas, o número de variantes apresentadas, o número de ocorrências proferidas pelos sujeitos do sexo masculino e feminino e o número de abstenções. Fizemos algumas considerações semelhantes às que foram feitas em cada campo semântico, classificamos as variantes de acordo com a proposta de Pottier (1972:27) — lexia simples, complexa, composta e textual — pesquisamos nos dicionários apenas a variante/resposta proposta pelo ALiB que teve o maior e a que apresentou o menor número de variantes em cada campo, em dicionários de especialidade de cada área, no *Dicionário do Folclore Brasileiro* de Câmara Cascudo, quando a palavra estava registrada lá, em seguida, nos seguintes dicionários gerais: *Dicionário Caldas Aulete*, *Novo Aurélio Século XXI: o Dicionário da Língua Portuguesa*, *Dicionário Houaiss da Língua Portuguesa* e no *Dicionário Etimológico Nova Fronteira da Língua Portuguesa de Cunha*; por último, pesquisamos todas as palavras encontradas no *corpus*, para identificação e comparação em atlas lingüísticos publicados no Brasil: APFB — *Atlas Prévio dos Falares Baianos*, ALS — *Atlas Lingüístico de Sergipe*, ALPB — *Atlas Lingüístico da Paraíba*, EALMG — *Esboço de um Atlas Lingüístico de Minas Gerais*, ALPR — *Atlas Lingüístico do Paraná*, em dissertações de mestrado de Aguilera: EALL — *Esboço de um Atlas Lingüístico de Londrina*, de Berti Santos: EGMS — *Estudo Geolingüístico de Aspectos Semânticos-Lexicais do Campo Semântico “Alimentação e cozinha” (Questionário do ALiB) no Município de Sorocaba*, e de Imaguire: EALISC — *Estudo com vistas ao Atlas Lingüístico da Ilha de Santa Catarina – Abordagem dos aspectos lexicais* e na tese de

⁴ Histograma: representação gráfica de uma distribuição de freqüência em que as freqüências de classes são representadas pelas áreas de retângulos contíguos e verticais, com as bases colineares e proporcionais aos intervalos das classes.

doutorado de Cardoso da Silva: ALMP — *Estudo Semântico- Lexical com vistas ao Atlas Lingüístico da Mesorregião do Marajó/Pará.*

A seguir, fizemos alguns comentários sobre a carta lexical que representa a pergunta que teve o maior e a que apresentou o menor número de variantes de cada campo semântico. Isso foi feito em todos os campos semânticos.

O banco de dados

Conforme mencionamos anteriormente, especificamos um programa para manipulação do banco de dados. O programa é constituído basicamente de três módulos: 1) interface para entrada de dados dos entrevistados; 2) interface para entrada das respostas das entrevistas; 3) geração de tabelas e estatísticas. Utilizamos esse programa para inserir os dados do *corpus*. No total, resultaram aproximadamente 500 horas de digitação dos dados. O banco gerou as tabelas automaticamente, porém fomos nós quem fizemos o *layout* das tabelas e esse processo passou por algumas reelaborações até chegar à forma atual. O *software* foi desenvolvido pela empresa Void Caz.

De acordo com Vieira (1998:208) “o tempo que se dispensa com a criação de um banco de dados informatizado é compensado com grande vantagem, pois nos permite a elaboração de cartas e a publicação do Atlas Lingüístico em tempo muito menor e, como consequência, liberação do banco para acesso aos dados por outros pesquisadores interessados em aprofundar seus estudos”.

Concordamos com a citação acima e, por esse motivo, utilizamos em nosso trabalho o *software* STDPSL (*Software* para Tratamento de Dados de Pesquisa Semântico-Lexical), que gerou o banco de dados, a partir das informações coletadas na pesquisa de campo. Os dados coletados nas fitas foram transcritos, revisados e digitados no STDPL, que gerou as tabelas e as consultas.

No banco de dados estão inseridos, além das respostas dos sujeitos, o questionário lingüístico separado por campo semântico; informações sobre o sujeito: abreviatura do nome, sexo, faixa etária, estado civil e outros; dados históricos e geográficos sobre os municípios pesquisados, número da entrevista, número de fitas utilizadas etc.

Utilização do STDPSL (*Software* para Tratamento de Dados de Pesquisa Semântico-Lexical)

Basicamente o STDPSL é dividido da seguinte forma:

Pasta 1: “Cadastro”. Esta pasta é subdividida em três: “ponto lingüístico”, “ficha do sujeito” e “questionário lingüístico”.

Pasta 2: “Entrevista”.

Pasta 3: “Estatística”.

A seguir, explicaremos cada pasta e suas funções.

Pasta 1: Cadastro

Nessa pasta encontramos:

a) Ponto lingüístico

No ponto lingüístico são cadastrados os municípios e as observações sobre eles.

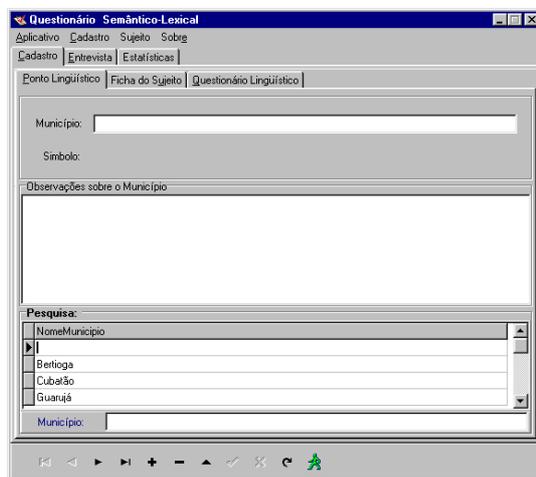


Figura 1

b) Ficha do sujeito

A “ficha do sujeito” foi usada para cadastramento dos dados pessoais das pessoas entrevistadas

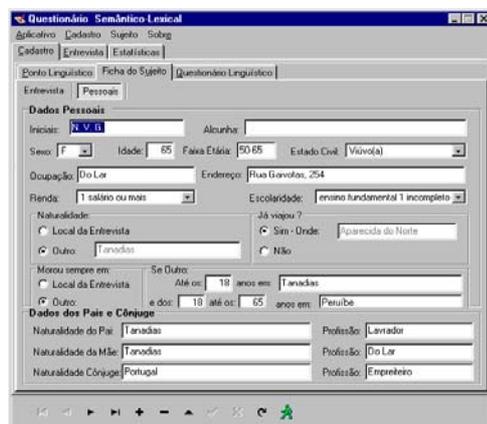


Figura 2

Estes são os campos pertencentes à “ficha do sujeito”, ou pessoa entrevistada:

Iniciais: são as iniciais do nome. Ex: João Silva é J. S.

- 1) Alcinha: se houver; é o apodo da pessoa entrevistada.
- 2) Sexo: masculino e feminino.
- 3) Idade: número que equivale à idade do entrevistado.
- 4) Faixa etária — existem três faixas predefinidas: 18 a 30, 31 a 49 e 50 a 65. Esse campo é autocompletado a partir do preenchimento da idade.
- 5) Estado civil — existem seis opções: casado, desquitado, divorciado, solteiro, viúvo e outro.
- 6) Ocupação: profissão ou ocupação do entrevistado.
- 7) Escolaridade — existem nove opções: Analfabeto, Ensino Fundamental 1 Completo ou Incompleto, Ensino Fundamental 2 Completo ou Incompleto, Ensino Médio Completo ou Incompleto, Ensino Superior Completo ou Incompleto.
- 8) Renda — existem cinco opções: 1 salário ou mais, de 2 a 5 salários, de 6 a 10 salários, mais que 10 salários, menos que 1 salário.
- 9) Naturalidade: cidade onde o entrevistado nasceu, se nasceu na própria cidade da entrevista ou em outra; em caso de outra, especifica-se a cidade e qual o período da vida em que morou nela.
- 10) Dados dos pais e cônjuge: naturalidade e profissão dos pais e do cônjuge.

c) Questionário lingüístico

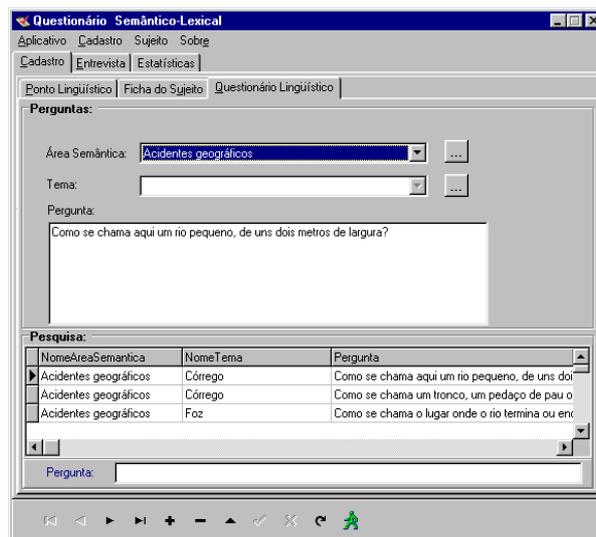


Figura 3

Nessa subpasta são definidas as “áreas semânticas”, o “nome” (tema) e a “pergunta”.

Pasta 2: Entrevista: A pasta “entrevista” foi usada para cadastramento das respostas dos entrevistados.

Entrevista	Área	Tema	Fita
1	Religião e crenças	Presépio	
1	Fenômenos atmosféricos	Redemoinho (de vento)	
1	Fenômenos atmosféricos	Relâmpago	
1	Fenômenos atmosféricos	Raio	

Figura 4

São campos da “entrevista”:

- 1) Número da entrevista.
- 2) Número da fita.
- 3) Campo semântico: deve ser escolhido uma área semântica declarada no questionário lingüístico.
- 4) Tema: deve ser escolhido um tema declarado no questionário lingüístico; os temas são restritos à área semântica escolhida.
- 5) Pergunta: aparece automaticamente na escolha do tema.
- 6) Respostas dessa pergunta: digita-se a resposta dada pelo entrevistado.
- 7) Pesquisa: tela de consulta das respostas já dadas.

Pasta 3: Estatísticas

Funcionamento da “ficha estatística”: seu propósito é possibilitar a recuperação das informações digitadas no sistema, possibilitando a apresentação desses dados de duas formas:

- a) Impressão da tabela em formato pré-definido.
- b) Geração dinâmica de consultas à base de dados com exportação para o Excel.

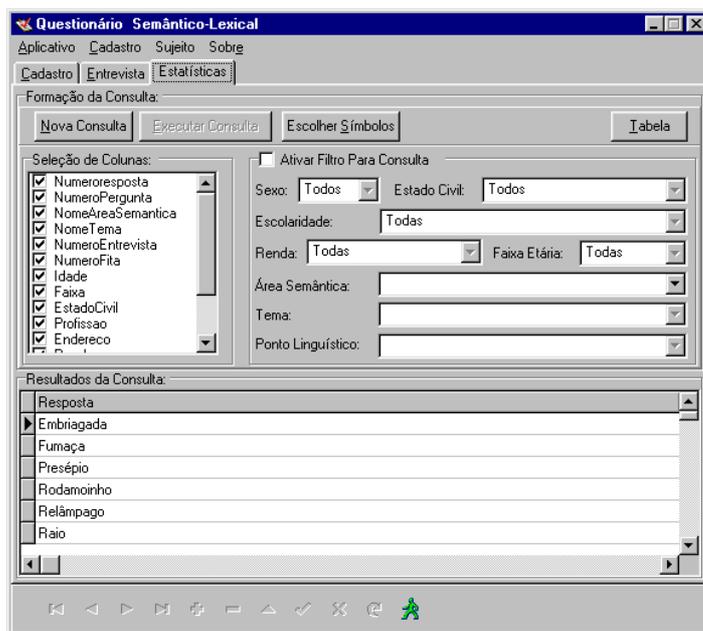


Figura 5

A seleção dos itens das colunas serve para filtrar dados da pesquisa; por exemplo, pode-se fazer uma pesquisa por faixa etária, por faixa salarial e outros. É possível ativar e desativar itens do filtro conforme a necessidade ou interesse da pesquisa.

Na pasta estatística encontra-se a subpasta “escolher símbolos”, na qual podem ser acrescentados e retirados os símbolos que farão parte das tabelas.

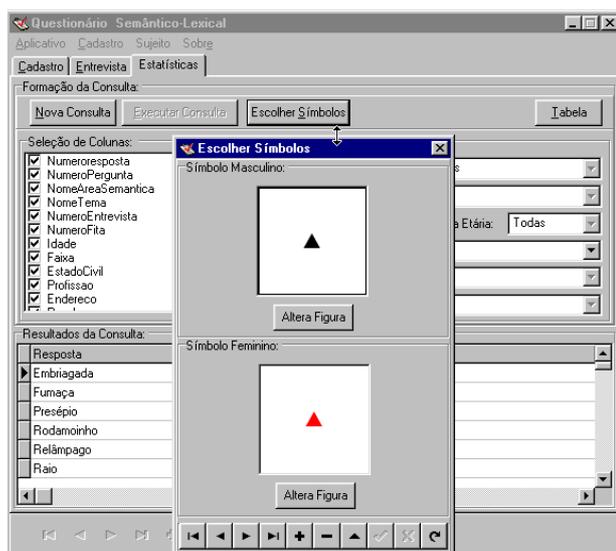


Figura 6

Outra subpasta importante é a “tabela”, por meio da qual podemos gerar uma tabela das variantes de uma pergunta, a partir dos cruzamentos dos dados que foram inseridos no programa. Essas tabelas podem ser gravadas em vários formatos gráficos ou textos tabulados.

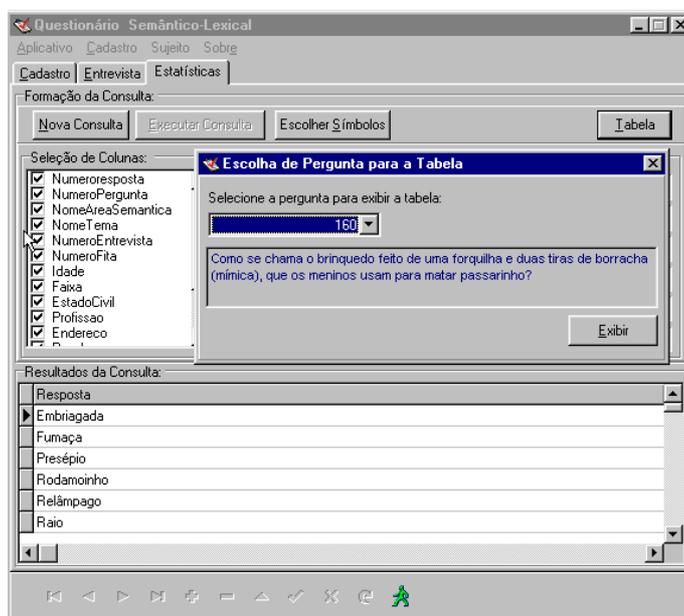


Figura 7

Impressão da tabela

Essa opção, acessível por meio do clique no botão “tabela” permite a exibição, impressão e/ou salvamento de uma tabela em formato predefinido.

Importante: a tabela sempre mostra todos os dados da pergunta selecionada, independente do que foi selecionado na ficha de estatísticas. Os dados de seleção de colunas e filtro para consultas somente são válidos para a tela de “resultados da consulta”, conforme é demonstrado no próximo item.

Geração dinâmica de consultas à base de dados

Conforme mostra a figura 5, o sistema permite a geração de consultas diretamente à base.

Para realizar uma nova consulta, devemos seguir os seguintes passos:

- 1) Clique no botão “nova”.
- 2) Selecione as colunas que você deseja que apareçam no resultado da consulta.

A coluna “resposta” não está disponível para seleção e é obrigatória.

3) Para restringir os dados que serão apresentados, habilite o “filtro” para consulta. Por exemplo: se você quer que apareçam somente as entrevistas realizadas com sujeitos do sexo masculino, selecione o sexo “M”, caso você deseje ainda que somente apareçam as entrevistas de sujeitos cujo estado civil seja “outros”, selecione a opção correspondente, e assim por diante.

4) Após elaborar os “filtros” e selecionar as colunas, clique em “executar consulta”, para que as informações sejam, buscadas no banco de dados.

5) Assim que o sistema terminar a busca, você poderá exportar o resultado para o Excel, clicando com o botão direito sobre os dados que aparecem e escolhendo a opção “exportar para o Excel.

As tabelas

As tabelas foram geradas a partir do STDPSL — *Software* para Tratamento de Dados de Pesquisa Semântico-Lexical, porém, como citado acima, fomos nós quem as criamos. Elas estão organizadas da seguinte maneira: na primeira linha consta a pergunta do questionário do ALiB, na primeira célula da segunda linha, à esquerda, estão as variantes, ou seja, as várias unidades lingüísticas que figuram no mesmo âmbito diatópico. As variantes foram registradas ortograficamente e seguem a ordem em que apareceram na pesquisa.

Na segunda célula, da esquerda para direita, temos a frequência absoluta e percentual; a frequência absoluta no *corpus* indica o número de vezes que a resposta apareceu e a frequência percentual indica a porcentagem de aparecimento do fato na pesquisa.

Na célula seguinte temos o sexo, Masc., para o sexo masculino, e Fem., para o sexo feminino; na mesma coluna temos uma outra subdivisão representada por Qtd. que indica a quantidade de variantes apresentadas e o Simb. que indica os símbolos — os pretos indicam as respostas das mulheres e os vermelhos, as dos homens. Esses servirão também para registrar, na carta lexical, as respostas dos sujeitos ocorridas em cada ponto lingüístico.

Na última célula da tabela, à direita, aparece, na primeira linha, “pontos”, que se refere aos pontos lingüísticos pesquisados, embaixo aparecem colunas numeradas (o significado da numeração é encontrado na última linha da tabela, onde descrevemos o nome dos municípios pesquisados, ou seja, os pontos lingüísticos).

Na confecção das cartas que representam as perguntas com o maior número de variantes, em cada campo semântico, usamos uma avaliação estatística para a observação das palavras de maior incidência e eliminamos as palavras com ocorrência menor que 5% no total da pesquisa. Utilizamos esse critério por falta de espaço para a confecção da legenda.

A tabela nos permite fazer a análise quantitativa, confeccionar os histogramas, elaborar as cartas e fornecer os elementos para a análise qualitativa. Para esta última, agruparemos as lexias de acordo com a pesquisa efetuada em dicionários. Rossi (1965: 15-21) também efetuava triagem em dicionários, mas para certificar-se da inclusão ou exclusão da palavra no questionário. Pesquisava as formas obtidas nas sondagens preliminares nos seguintes dicionários, que ele denominava de léxicos gerais: Bluteau, Viterbo, Moraes, Faria, Cortesão, Figueiredo, Aulete e Domingos Vieira.

Histograma

Os histogramas foram elaborados no programa Excel, a partir das tabelas “Número de variantes lexicais obtidas em cada pergunta” e “Número de ocorrências relativas à variável sexo”, foram geradas em auto-escala. Por esse motivo, cada histograma tem intervalos diferentes, definidos pelo próprio programa, de maneira a apresentar melhor os dados.

No histogramas das variantes lexicais, as colunas representadas pela cor roxa indicam as variantes encontradas para cada uma das perguntas. O histograma das ocorrências da variável sexo estão organizados da seguinte forma: as ocorrências mencionadas pelos sujeitos do sexo masculino estão representados pela coluna na cor roxa e as ocorrências mencionadas pelos sujeitos do sexo feminino estão representadas na cor marrom. Ambos estão organizados por ordem crescente de variantes e de ocorrências.

No eixo das ordenadas, eixo y, vertical, encontramos a escala numérica e, no eixo das abscissas, eixo horizontal, os temas das perguntas. De acordo com o questionário do ALiB, optamos por colocar o tema e não as perguntas por extenso devido ao espaço pequeno e à estética da página.

Nos histogramas, analisamos e comparamos as ocorrências mencionadas por homens e mulheres, de acordo com o pico máximo (coluna mais alta) citada por homem e mulher. Apontamos também os percentuais ocorridos em cada pergunta, por faixa de

variantes e por faixa de ocorrências por sexo em cada campo semântico; para isso, aplicamos a seguinte fórmula:

$$\text{valor \%} = \frac{\text{número de perguntas pertencente a um intervalo} \times 100\%}{\text{número total de perguntas}}$$

Conforme exposto na fórmula, multiplicamos o número de colunas pertencentes a um determinado intervalo por cem (100%) e dividimos o resultado pelo número total de colunas do histograma. Os resultados dessa fórmula estão expostos na análise de cada histograma.

As cartas lingüísticas

Muitos autores, como é o caso de Coseriu (1982:83), utiliza a palavra “mapa” para “carta lingüística”; outros pesquisadores, da área de Geografia principalmente, chamam de “carta”. Fomos pesquisar o significado de “carta” no dicionário Aurélio (1986:360): é a “arte ou ciência de compor cartas geográficas” e, de acordo com a Associação Cartográfica Internacional (1993:2) apud Rosa, “uma imagem convencional, representando feições e características da realidade geográfica, construída para uso quando as relações espaciais são de relevância fundamental”. Segundo Salichtchev (1973:1) apud Rosa “cartografia é a ciência da retratação, é o estudo da distribuição espacial dos fenômenos naturais, suas relações e suas mudanças ao longo do tempo, por meio das representações cartográficas – modelos – imagens – símbolos que reproduzem este ou aquele aspecto da realidade de forma gráfica e generalizada”.

Utilizamos o texto de Coseriu apenas para exemplificar os diversos tipos de “mapas/cartas” existentes, entretanto nós, em nossa pesquisa, usamos o termo “carta”.

De acordo com Coseriu (1982:83) os mapas lingüísticos mostram a variedade de estudo que a pesquisa da fala pode proporcionar. Conforme o enfoque dado aos fatos lingüísticos, os mapas lingüísticos podem ser:

a) **fonéticos** — registram as variantes de um fonema, ou de vários fonemas correspondentes a um único fonema mais antigo, ou determinadas séries de fonemas que se encontram na mesma situação do ponto de vista histórico;

b) **lexicais** — registram as palavras empregadas para expressar o mesmo conceito, não considerando as variações fônicas, isto é, a pronúncia de cada ponto;

c) **propriamente lingüísticos** — registram em sua integridade fônica e morfológica as expressões comprovadas em cada ponto investigado;

d) **sintéticos** — necessitam de uma elaboração, pois estabelecem os limites das áreas correspondentes às formas típicas comprovadas;

e) **pontuais** — registram fielmente as formas comprovadas em todos e em cada um dos pontos investigados.

Elaboramos dois tipos de carta. As primeiras, geográficas, com o objetivo de situar a região administrativas no Estado de São Paulo e no Brasil e as lexicais com os registros comuns para expressar a mesma lexia.

A escolha das palavras que integrarão as cartas lexicais basear-se-á nos critérios:

- pergunta que apresentou o maior número de variantes
- pergunta que apresentou o menor número de variantes

Para a confecção das cartas que representam as perguntas com o maior número de variantes na pesquisa, ou a maior frequência no campo semântico, de acordo com Muller, (1977:81), criamos uma avaliação estatística e suprimimos da carta as lexias com ocorrência menor que 5% no total da pesquisa. No entanto, esse procedimento não foi utilizado na carta que representa a pergunta com o menor número de variantes, pois nesse caso, foram consideradas todas as variantes para a elaboração da carta. As cartas lexicais serão apresentadas por campos semânticos, sua ordem seguirá a das perguntas do questionário e, dentro de cada campo semântico, seguirá a ordem maior número e menor número de variantes.

A numeração das cartas lexicais segue a ordem das perguntas do questionário; portanto o número da carta corresponde ao número da pergunta.

Considerações parciais

Alguns métodos e procedimentos utilizados em nossa pesquisa estão sendo usados para a confecção do ALiB.

Fizemos pesquisa direta *in loco*. Selecionamos dentro do território paulista uma região administrativa, incluindo aí um ponto de entrevista do ALiB. Optamos pela região de Santos e fizemos pesquisa em oito municípios (oito pontos lingüísticos).

Os critérios utilizados para a escolha dos sujeitos priorizaram: ser natural da localidade ou estar residindo ali um terço da sua vida, faixa etária, nível de escolaridade e sexo.

A equação matemática utilizada para a escolha do número de sujeitos em cada ponto foi de grande valia, pois ajudou-nos a encontrar a relação mais adequada, entre o número de habitantes do município e o número de sujeitos que iríamos entrevistar.

O questionário utilizado foi elaborado pela equipe que está executando o atlas lingüístico do Brasil, composto de duzentas e dez perguntas subdivididas em quinze campos semânticos.

As tabelas foram geradas a partir do STDPSL — *Software* para Tratamento de Dados de Pesquisa Semântico Lexical, desenvolvido pela empresa Void Caz e cujos quesitos básicos foram apresentados pela própria pesquisadora e o software foi especificado por Mário Imaguire.

O banco de dados possibilitou-nos armazenar os dados pesquisados e fazer diversas consultas, gerando tabelas, enfim, facilitou o nosso trabalho.

Nos histogramas conseguimos visualizar e analisar as ocorrências da variável sexo e as variantes de cada pergunta.

As gravações foram imprescindíveis para a pesquisa, pois nos permitiram captar e armazenar o material do *corpus*, isto é, a fala dos sujeitos.

De acordo com Coseriu (1982:83): “As cartas lexicais têm como objetivo registrar as palavras utilizadas para expressar o mesmo conceito, independentemente das variações fônicas, isto é, a pronúncia peculiar comprovada em cada ponto

Bibliografia

AMARAL, Amadeu. *O dialeto caipira*. Gramática – Vocabulário. São Paulo: Editora Anhembi Limitada, 1955.

ARAGÃO, Maria do Socorro Silva de & MENEZES, Cleusa P. Bezerra de. *Atlas Lingüístico da Paraíba*. Brasília: CNPq/UFPB, 1984. 2v.

AGUILERA, Vanderci de Andrade. *Esboço de um Atlas Lingüístico de Londrina*. Assis, São Paulo: 1987.

_____. *Atlas Lingüístico do Paraná*. Curitiba: Editado pela Imprensa Oficial do Paraná, 1994.

_____. (org.) *A geolingüística no Brasil: caminhos e perspectivas*. Londrina: Editora da UEL, 1998.

BRANDÃO, Sílvia Figueiredo. *A geografia lingüística no Brasil*. São Paulo: Ática, 1991.

CARUSO, Pedro. *Atlas Lingüístico do Estado de São Paulo*. Questionário. Assis, Instituto De Letras, História e psicologia/UNESP, Prefeitura Municipal de Assis, 1983.

_____. *Atlas lingüístico de São Paulo: Histórico*. Mimeo, S/D

CHAMBERS, J. K. et TRUDGILL, Peter. *Dialectology*. Cambridge: University Press, 1988.

COMITÊ NACIONAL DO PROJETO ALiB. *Atlas Lingüístico do Brasil*. Mimeo, S/D.

- COMITÊ NACIONAL DO PROJETO ALiB. *Atlas Lingüístico do Brasil: questionários*. Londrina: Ed. UEL, 1998.
- COMITÊ NACIONAL DO PROJETO ALiB. *Atlas Lingüístico do Brasil: II Workshop de preparação de inquiridores para o Atlas Lingüístico do Brasil*. Londrina: Ed. UEL, 2000.
- COMITÊ NACIONAL DO PROJETO ALiB. *Atlas Lingüístico do Brasil: questionário 2001*. Londrina: Ed. UEL, 2001.
- FERREIRA, Carlota et CARDOSO, Suzana Alice. *A dialetologia no Brasil*. São Paulo: Contexto, 1994.
- FERREIRA, Carlota; FREITAS, Judith; MOTA, Jacyra; ANDRADE, Nadja; CARDOSO, Suzana; ROLLEMBERG, Vera; ROSSI, Nelson. *Atlas Lingüístico de Sergipe*. Salvador: Universidade Federal da Bahia/ Fundação Estadual da Cultura de Sergipe, 1987.
- IMAGUIRE, Lúgia Maria Campos. *Estudo com vistas a um Atlas Lingüístico da Ilha de Santa Catarina: Abordagem de aspectos semânticos lexicais*. 1999. 554 f. (Mestrado em Lingüística). Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo.
- NASCENTES, Antenor. *O linguajar carioca*. 2ª ed., Rio de Janeiro: Editora da "Organização Simões", 1953.
- _____. *Bases para a elaboração do Atlas Lingüístico do Brasil*. Rio de Janeiro: MEC, Casa de Rui Barbosa, 1958.
- _____. *Bases para a elaboração do Atlas Lingüístico do Brasil*. Rio de Janeiro: MEC, Casa de Rui Barbosa. V 2, 1961.
- RIBEIRO, José et alii. *Esboço de um Atlas Lingüístico de Minas Gerais*. Rio de Janeiro: MEC, Fundação Casa de Rui Barbosa, UFJF, V 1, 1977.
- RODRIGUES, Ada Natal. *O dialeto caipira na região de Piracicaba*. São Paulo: Ática, 1974.
- ROSSI, Nelson. *Atlas Prévio dos Falares Baianos*. Rio de Janeiro: Instituto Nacional do Livro, MEC, 1963.
- _____. *Atlas Prévio dos Falares Baianos; introdução, questionário comentado, Elenco das respostas transcritas*. Rio de Janeiro: Instituto Nacional do Livro, MEC, 1965.
- SANTOS, Irenilde Pereira dos. A fala da comunidade do Jardim São Francisco: análise de alguns aspectos lingüísticos. In: *A Geolingüística no Brasil*. Vanderci de Andrade Aguilera. (org.). Londrina: Editora da UEL, 1998.
- SANTOS, Irenilde Pereira dos. Elementos para o estudo do gênero em Geolingüística. (mimeo), 2003.
- _____. *Geolingüística vertentes*. (mimeo), 1999.
- SILVA NETO, Serafim da. *Guia para estudos dialetológicos*. 2ª ed., Belém: Instituto Nacional de Pesquisas da Amazônia, CNPq, 1957.
- VARIA. SEMINÁRIO NACIONAL CAMINHOS E PERSPECTIVAS PARA A GEOLINGÜÍSTICA NO BRASIL. Fotocopiado, Bahia: UFB, 1996.