

**UNIVERSIDADE FEDERAL DE SANTA CATARINA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS/LINGÜÍSTICA**

**A CONCORDÂNCIA DE NÚMERO**  
**NOS PREDICATIVOS E NOS PARTICÍPIOS PASSIVOS**  
**NA FALA DA REGIÃO SUL:**  
**Um Estudo Variacionista**

Por

**JUÇÁ FIALHO VAZZATA DIAS**

**Dissertação apresentada ao curso de Pós-Graduação em Letras/Lingüística da Universidade Federal de Santa Catarina, como parte dos requisitos para obtenção do grau de Mestre em Lingüística.**

**Florianópolis**

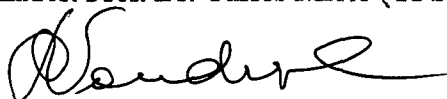
**1996**

**A CONCORDÂNCIA DE NÚMERO  
NOS PREDICATIVOS E NOS PARTICÍPIOS PASSIVOS  
NA FALA DA REGIÃO SUL:  
Um Estudo Variacionista**

Esta dissertação foi julgada adequada para a obtenção do grau de Mestre em Letras/Linguística e aprovada em sua forma final pelo Programa de Pós-Graduação em Linguística da Universidade Federal de Santa Catarina.



Coordenador: Prof. Dr. Carlos Miotto (UFSC)



Orientador: Prof. Dr. Paulino Vandresen (UFSC)

Banca Examinadora:



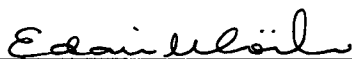
Prof. Dr. Paulino Vandresen (UFSC)



Prof. Dr. Attila Louzada Jr. (FURG) (Co-orientador)



Profa. Dra. Maria Marta Pereira Scherre (UFRJ-UnB)



Profa. Dra. Edair Maria Görski (UFSC)

## AGRADECIMENTOS

Professor Paulino Vandresen, orientador dedicado e grande incentivador à pesquisa.

Professor Attila Louzada Jr., co-orientador exemplar.

Projeto VARSUL e todos que o constróem.

Professoras Edair Maria Görski e Izete Lehmkuhl Coelho, pelo incentivo e ajuda constantes; sem dúvida, amigas que não medem esforços para nos verem crescer.

Professora Maria Marta Pereira Scherre, pelo carinho e ensinamentos.

Colega Marisa Fernandes, pelo apoio dado no uso do Programa VARBRUL.

Márcia Fernandes, Maria de Lourdes Novaes, Loremi Loregian, Janete Cristina Stülp, Luciane Baretta, Brigitte de Lima Goulart, Márluce Coan, Cristiano dos Passos, Miguel Augusto C. P. Ribeiro, Maria Lúcia Miranda, Diane Dal Mago, Maria Alice Tavares, Sergio Luis Duarte, Isabel Monguilhott, Ana Luzia Dias Pereira, Emília Marta Schweitzer e Rosa Maria Somavilla, amigos do Projeto VARSUL, por todo apoio e carinho.

Suzana Aparecida D. de O. da Rocha, secretária do Curso de Pós-Graduação em Lingüística, pela amizade e dedicação.

CAPES, pela concessão da bolsa.

## DEDICATÓRIA

A

Minha família, meu porto seguro;

Antonio Carlos;

Meu orientador;

Meu co-orientador;

Amigos;

Projeto VARSUL e a todos que fizeram e fazem parte dele;

Prof. Dr. Domingos Bernardes Morey Filho (In Memoriam);

Informantes, donos da língua,

Dedico este trabalho.

**(...) faz parte da formação para a ciência o desenvolvimento de uma atitude intelectual capaz de separar os fatos claramente dos valores de senso comum que recaem sobre eles. [Neste sentido é que] o estudante tem de acostumar-se a submeter à crítica rigorosa e permanente os juízos sociais sobre a língua, procurando se livrar dos preconceitos e respaldando sempre suas próprias afirmações com dados empíricos. (Faraco, 1991:17)**

## SUMÁRIO

LISTA DE QUADROS E TABELAS	viii
1. INTRODUÇÃO	1
2. PRESSUPOSTOS TEÓRICO	6
2.1. A Teoria Sociolingüística Variacionista	10
3. CONTRIBUIÇÕES DA SOCIOLINÜÍSTICA	16
4. METODOLOGIA	22
4.1. Pressupostos básicos	22
4.2. Constituição da amostra	26
4.2.1. As regiões e seus grupos étnicos	27
4.2.1.1. Florianópolis (SC)	27
4.2.1.2. Chapecó (SC)	28
4.2.1.3. Irati (PR)	29
4.3. Processo de coleta e armazenamento dos dados	30
4.3.1 - A coleta dos dados	32
4.3.2 - O armazenamento dos dados	33
4.4. Levantamento e codificação dos dados	36
5. SUPORTE QUANTITATIVO	41
5.1. Introdução	41
5.2. Os modelos quantitativos	42
6. ANÁLISE DAS VARIÁVEIS LINGÜÍSTICAS	51
6.1. Variáveis lingüísticas estatisticamente significativas	52

6.2. Variáveis lingüísticas estatisticamente não significativas	74
7. ANÁLISE DAS VARIÁVEIS SOCIAIS	84
7.1. O comportamento de variáveis sociais na variação estável	85
7.2. O comportamento de variáveis sociais na mudança em progresso	86
8. ANÁLISE DE CRUZAMENTOS	94
9. CONSIDERAÇÕES FINAIS	101
REFERÊNCIAS BIBLIOGRÁFICAS	109

## LISTA DE QUADROS E TABELAS

Quadro 1	Estratificação	27
Tabela 1	Paralelismo formal das seqüências no discurso	54
Tabela 2	Características formais do verbo da construção	57
Tabela 3	Características formais do sujeito da construção	59
Tabela 4	Estrutura do Predicativo	62
Tabela 4a	Tabulação cruzada entre estrutura do predicativo e características formais do verbo	65
Tabela 5	Tonicidade dos itens singulares	67
Tabela 6a	Distância do sujeito correferente	71
Tabela 6b	Distância do sujeito correferente (com pesos relativos da rodada sem a fala do entrevistador)	72
Tabela 6c	Distância do sujeito correferente	73
Tabela 7	Ordem dos elementos na estrutura	75
Tabela 8	Material interveniente entre o verbo e o pred./part.	76
Tabela 9a	Processos morfofonológicos de formação do plural	78
Tabela 9b	Processos morfofonológicos de formação do plural	80
Tabela 10	Tipo de estrutura	82
Tabela 10a	Tabulação cruzada entre tipo de estrutura e escolaridade	83
Tabela 11	Variáveis sociais	91



Tabela 12	Escolaridade com etnia	94
Tabela 13	Escolaridade com faixa etária	95
Tabela 14	Escolaridade com sexo	96
Tabela 15	Etnia com sexo	97
Tabela 16	Etnia com faixa etária	98
Tabela 17	Sexo com faixa etária	99

## RESUMO

Este trabalho objetiva a desenvolver, sob o enfoque da Sociolinguística Variacionista Quantitativa, uma análise descritiva do comportamento da concordância de número nos predicativos/particípios passivos, na fala de moradores de três cidades da região sul do Brasil (Florianópolis, Chapecó e Irati).

O corpus analisado faz parte do Banco de Dados do **Projeto Variação Linguística Urbana da Região Sul (VARSUL)** e consiste em 24 entrevistas de cada cidade, num total de 72, todas coletadas e transcritas conforme metodologia laboviana.

Na análise, descreve-se, primeiramente, o conjunto de variáveis linguísticas que atuam na variação do fenômeno em estudo, enfocando dois princípios já incorporados pelos estudos variacionistas: **Processamento Paralelo** e **Saliência Fônica**. Além destes, aborda-se também o **Subprincípio da Quantidade** (Princípio da Iconicidade) da Teoria Funcionalista. Depois, analisa-se o conjunto de variáveis sociais que condicionam a presença de marcas formais de plural, buscando evidências de variação sociolinguística estável e mudança em progresso.

Este estudo deixa sua contribuição na medida em que descreve o comportamento da concordância de número nos predicativos/particípios passivos, o que vem colaborar para um melhor entendimento do Português falado.

## ABSTRACT

The aim of this thesis is to develop, through the methodology of Quantitative Sociolinguistics, a descriptive analysis of the variable phenomenon "**number agreement in the predicate/passive participle**", in the speech of residents of three cities in the South of Brazil (Florianópolis, Chapecó and Irati).

The corpus is extracted from the VARSUL (**Projeto Variação Lingüística Urbana da Região Sul**) Database and consist of 24 interviews in each city, all of them collected and transcribed according to Labovian methodology.

First of all, the linguistic variables that determine systematic tendencies of variation in number agreement in the predicate/passive participle are described, focusing on two Principles: **Parallel Processing** and **Phonic Salience**. Furthermore, the **Quantity Principle** from Functionalist Theory is also approached. The influence of social variables is then discussed, looking for evidences of stable sociolinguistic variation and change in progress.

This study contributes in the way it describes the variation of number agreement in the predicate/passive participle, which helps provide a better understanding of oral Portuguese.

## 1- INTRODUÇÃO

No presente trabalho, apresentamos uma análise quantitativa do fenômeno da **concordância de número nos predicativos e nos participios passivos** com base em dados reais de fala de moradores de três cidades do sul do Brasil: Florianópolis, Chapecó e Irati. Os dados analisados fazem parte do Banco de Dados do Projeto VARSUL<sup>1</sup> (Variação Lingüística Urbana da Região Sul).

Partindo dos pressupostos da Teoria da Variação Lingüística, segundo a qual existe uma relação sistemática entre língua e pressões internas do sistema lingüístico, de um lado, e forças sociais sobre a comunidade, de outro, procuramos confirmar as hipóteses estabelecidas por Scherre (1991), bem como confrontar seus resultados com os alcançados neste estudo.

Em seu artigo, Scherre analisa o fenômeno da concordância de número nos predicativos e nos participios passivos a partir de dados reais de fala de informantes do Rio de Janeiro, dados esses coletados e armazenados pelo Programa de Estudos sobre o Uso da Língua (PEUL), originado do Projeto Censo da Universidade Federal do Rio de Janeiro.

Além desta pesquisa, Scherre desenvolveu também trabalhos sobre a **concordância de número no sintagma nominal** (1978 e 1988) e, juntamente com Naro, vem se dedicando ao estudo da **concordância verbal** (Scherre e Naro, 1991; Naro, 1981).

<sup>1</sup> V. capítulo 4 - Metodologia

Outros estudos envolvendo fenômenos de concordância ainda podem ser citados.

Braga (1977 - apud Scherre, 1988) analisa a **concordância de número no sintagma nominal no triângulo Mineiro**; Ponte (1979, apud id. ibid) desenvolve um estudo sobre a **concordância nominal de uma comunidade de Porto Alegre**; Carvalho Nina (1980, apud id. ibid.) trabalha com a **concordância nominal/verbal do analfabeto na micro-região de Bragantina**, e Saraiva & Bittencourt (1987, apud id. ibid.) estudam a **concordância verbal em estruturas com SN complexo em português**.

Utilizando dados de informantes da região Sul (Banco de Dados do Projeto VARSUL), temos Fernandes analisando a **concordância de número no sintagma nominal** e Loregian estudando a **concordância verbal com o pronome tu**.

Fora do Brasil, Poplack (1980) analisa o fenômeno da queda do /s/ no espanhol falado em Porto Rico, e também a queda da nasal e a queda do r em final de palavra (Tese de Doutorado, 1979).

Com a investigação, pretendemos alcançar três objetivos.

O primeiro deles é o de apresentar a descrição do conjunto de variáveis lingüísticas e extra-lingüísticas que regem a variação da concordância de número nos predicativos/participios passivos. Vamos mostrar que a variação é sistemática e que, dependendo das circunstâncias lingüísticas, o predicativo/participio passivo aparecerá ou não com todas as marcas formais de plural.

O segundo objetivo é o de trazer à tona mais evidências que venham corroborar os Princípios do Processamento Paralelo, da Saliência Fônica e da Quantidade.

O terceiro objetivo diz respeito à relação entre variação e mudança, no sentido de se verificar se o fenômeno em estudo reflete um estágio de variação sociolingüística estável ou um processo de mudança lingüística em progresso.

Pretendemos, portanto, através desta análise, trazer contribuições acerca do funcionamento do português, através da descrição das variáveis lingüísticas e sociais que regem a presença ou ausência de marcas formais de plural no predicativo/particípio passivo.

A exposição do trabalho seguirá a seguinte ordem:

No segundo capítulo, apresentamos os pressupostos teóricos básicos da Teoria Variacionista Laboviana, dentro de cujos princípios teórico-metodológicos se desenvolve toda a pesquisa. Introduzindo este capítulo, traçamos uma breve retrospectiva que enfoca precipuamente a Teoria Estruturalista Saussureana e a Teoria Gerativo-Transformacional.

No terceiro capítulo, discutimos algumas contribuições que os trabalhos de descrição sociolingüística trazem em seu bojo, tanto no que se refere a uma melhor compreensão da configuração social das comunidades quanto ao ensino do português.

No quarto capítulo, tratamos da metodologia seguida neste estudo. Iniciamos fazendo algumas considerações básicas a respeito da Metodologia criada por Labov para, em seguida, apresentarmos a amostra de informantes cuja fala analisamos. Neste momento, descrevemos as características sociais dos falantes, bem como apresentamos algumas informações sobre cada uma das três cidades e grupos étnicos que estamos considerando na pesquisa. Tratamos também neste capítulo sobre o processo de coleta e armazenamento dos dados (Projeto VARSUL), além do levantamento e codificação dos mesmos.

No quinto capítulo, descrevemos o suporte quantitativo utilizado na análise dos dados. Fazemos uma breve apresentação do desenvolvimento de modelos quantitativos para, em seguida, abordarmos os passos que envolvem a aplicação dessa orientação metodológica.

No sexto capítulo, apresentamos a análise das variáveis lingüísticas. Abordamos aqui os Princípios do Processamento Paralelo e da Saliência Fônica e o Princípio da Quantidade da Teoria Funcionalista.

No sétimo capítulo, apresentamos a análise das variáveis sociais, com destaque ao comportamento das mesmas no que diz respeito à variação estável e mudança em progresso.

No oitavo capítulo, procedemos à análise e discussão de alguns cruzamentos de variáveis.

No nono e último capítulo, apresentamos as considerações finais onde retomamos os grupos de fatores mais atuantes para o favorecimento ou desfavorecimento da aplicação da regra de concordância nos predicativos e nos participios passivos, bem como mostramos as semelhanças e diferenças em termos dos resultados alcançados por Scherre e os desta pesquisa. Por fim, apresentamos algumas sugestões de continuidade da pesquisa.

O estudo de análise variacionista sobre o fenômeno da concordância de número nos predicativos e nos participios passivos, o qual nos propomos desenvolver, está apenas germinando. A semente foi lançada por Scherre (1991) e começamos a tratá-la aqui. Desta forma, alguns pontos que expomos ainda se revestem de incertezas. A pesquisa tem muito o que avançar e, neste momento, buscamos apontar os primeiros resultados referentes aos contextos lingüísticos e sociais onde haverá maior ou menor chance de ocorrer a concordância de número nos predicativos/participios passivos.



## 2- PRESSUPOSTOS TEÓRICOS

Pretendemos neste capítulo estabelecer a base teórica sobre a qual este trabalho terá sustentação, a fim de que possamos caminhar ao longo do mesmo já com determinadas perspectivas em mente. Mostraremos, aqui, o momento do surgimento, características, e a que veio a **Teoria Sociolinguística Variacionista**.

Antes, porém, façamos uma breve retrospectiva<sup>2</sup> e reportemo-nos ao início do século XX, momento que marca o surgimento de um dos grandes paradigmas da Linguística: o **Estruturalismo**, que tem como personagem principal **Ferdinand de Saussure**.

Influenciado pela Sociologia de Durkheim, Saussure estabeleceu uma distinção entre *língua* (*langue*) e *fala* (*parole*), considerando aquela um sistema abstrato, homogêneo, social, supra-individual, ao passo que a fala seria a realização concreta e individual da língua. Para Saussure, a linguística constitui-se em "uma ciência que estuda os signos no seio da vida social"<sup>3</sup>.

<sup>2</sup> A opção por uma apresentação retrospectiva tem caráter puramente didático. Queremos, com isso, situar a Sociolinguística dentro do universo da ciência linguística.

<sup>3</sup> "Saussure conceived of linguistics as one part of 'une science qui étudie la vie des signes au sein de la vie sociale'." (Labov, 1972b:185;) (v. Saussure, 1992:15-25.)

Entretanto, mesmo fazendo menção a uma visão da língua como realidade social, essa relação serviu apenas para fazer o recorte e a delimitação do objeto de estudo lingüístico. Saussure e seus discípulos não lidavam de forma alguma com o elemento social em suas pesquisas lingüísticas: "eles trabalhavam com um ou dois informantes em seus laboratórios, ou examinavam seus próprios conhecimentos sobre a língua"<sup>4</sup> (id. *ibid.*). Ou seja, a fala como realização concreta da língua pelo indivíduo, utilizada nos mais variados contextos da vida social, não é tomada como objeto de estudo pela escola estruturalista saussureana.

Não considerando, então, o estudo da fala, e, sim, do sistema abstrato e homogêneo<sup>5</sup> da *langue*, tal corrente passa a não dar importância às **variações lingüísticas** em suas análises, voltando suas preocupações para a "investigação de fatos lingüísticos com base na idéia fundamental de que língua é sistema e de que cada elemento desse sistema possui um valor especial, compreendido, principalmente, por suas *oposições em relação a outros elementos*" (Suassuna, 1995)<sup>6</sup>.

A partir da década de 50, esboça-se nova corrente teórica – a da **gramática gerativo<sup>7</sup>-transformacional**, cujo grande condutor é o lingüista norte-americano Noam Chomsky que procurou tecer, através de sua teoria, uma gramática formal, de base

<sup>4</sup> "(...) the linguists who work within the saussurian tradition (...) do not deal with social life at all : they work with one or two informants in their offices, or examine their own knowledge of langue." (Labov, 1972:185)

<sup>5</sup> No período saussureano a homogeneidade da linguagem era tomada como base e pré-requisito para a análise lingüística. (cf. Tarallo, 1990)

<sup>6</sup> O *itálico* é nosso.

<sup>7</sup> Gramática gerativa, pois a Teoria está centrada na criatividade da linguagem humana, ou seja, a característica que faz com que todo falante-ouvinte nativo produza ("gere") e interprete um número infinito de enunciados com base em um conjunto finito de unidades lingüísticas.

fundamentalmente sintática. Seu objetivo mais geral era explicar os mecanismos cognitivos da mente humana, para o que o caminho mais adequado seria o estudo da linguagem. Para Chomsky, a estrutura das línguas seria justificada com base em pressupostos inatistas<sup>8</sup> (biológicos), sendo a gramática universal uma entidade inata, programada no cérebro da criança como parte de sua herança genética. Segundo essa concepção, as gramáticas das línguas são fortemente restringidas por estruturas cerebrais e são da forma que são em razão de o cérebro ter a forma que tem.

Cartesiano tradicional, Chomsky rejeita veementemente o empiricismo Skinneriano, que se utiliza da manipulação e controle social em suas formas mais variadas. Critica duramente esta abordagem, pois a mesma nega a existência de capacidades inatas, ao mesmo tempo em que acredita única e exclusivamente no estudo do comportamento observável. Para ele, a *Teoria Behaviorista* peca ao explicar o comportamento humano em termos de respostas a certos estímulos presentes no meio ambiente, perspectiva esta que dá margens a uma *visão do homem como um animal essencialmente governado por fatores e condições externas, totalmente fora do seu controle.*

Em sua abordagem, Chomsky distingue **competência lingüística** (*conhecimento abstrato das regras da língua*) de **desempenho lingüístico** (*seleção e execução dessas regras, sujeito a falhas e imperfeições por se tratar do uso individual que cada falante faz da competência*), e estabelece que a lingüística deve se preocupar com o estudo da competência.

<sup>8</sup> Nas palavras de Chomsky, "I'm convinced that language is species-specific as a biological attribute, like vision, and that some of the deepest properties are really genetically determined." (Mussad, 1991)

Isto é, para ele, o dado lingüístico a ser estudado não é a expressão verbal (elocução) do indivíduo mas, sim, suas intuições sobre a língua – seus julgamentos sobre que sentenças são gramaticais e quais não são e que sentenças significam a mesma coisa. Como Saussure, Chomsky também acredita que as teorias lingüísticas podem ser desenvolvidas somente a partir do dado homogêneo e uniforme e que qualquer variação lingüística importa apenas a um ponto de vista aplicativo, não sendo, portanto, útil a uma lingüística teórica.

Tais abordagens, tanto a de Saussure quanto a de Chomsky, não consideram a possibilidade de um estudo do comportamento social ou um estudo da fala em uso. Para eles, tanto quanto para seus seguidores, a fala não caracteriza boa evidência uma vez que é constituída (do ponto de vista deles) de estruturas mal formadas. Rigidamente imanentistas, defendem a idéia de que se deve explicar fatos lingüísticos somente por meio de outros fatos lingüísticos, nunca por meio de fatos extralingüísticos também. Qualquer passo neste sentido é considerado desvio metodológico.

Trata-se a língua, segundo essas perspectivas, como se ela não tivesse falantes concretos. O falante que entra em cena é um curioso indivíduo idealizado, isolado, desconectado da história e da sociedade.

## 2.1- A Teoria Sociolinguística Variacionista

Na década de 60, entretanto, um novo quadro se delinea, com uma perspectiva de mudança substantiva através da consolidação de uma concepção de linguagem essencialmente social, que trata a língua como correlacionada de forma sistemática com a história social dos falantes. Trata-se da Teoria Sociolinguística Variacionista, de William Labov, que se constitui numa espécie de reação às teorias estruturalista e gerativo-transformacional.

A sociolinguística<sup>9</sup>, como ramo específico da ciência da linguagem, é relativamente recente, mas as preocupações com o entrelaçamento linguístico x social são mais antigas. Antes de Labov, alguns lingüistas já preconizavam uma visão sociolinguística, tecendo algumas relações língua-sociedade e distanciando-se, assim, de uma visão imanentista da língua como a de Saussure e Chomsky.

Como exemplos, Labov (1972) cita os nomes de Whitney 1901 (*"a fala não se constitui em posse individual, mas social; ela pertence não ao indivíduo, mas aos membros da sociedade."*); Vendryes 1951 (*"...A língua é, portanto, o fato social por excelência, o resultado do contato social."*); Jespersen 1946 (*"A língua de uma nação é o conjunto de*

<sup>9</sup> O termo "sociolinguística" não pode ser confundido aqui com "sociologia da linguagem". De acordo com Hudson (1980) (apud Wardhaugh, 1986), "Sociolinguística é o estudo da língua em relação à sociedade, enquanto a sociologia da linguagem é o estudo da sociedade em relação à língua. (...) A diferença entre sociolinguística e a sociologia da linguagem é muito mais uma questão de ênfase, conforme o interesse do investigador esteja na linguagem ou na sociedade, e também conforme ele apresente mais habilidades em análises das estruturas linguísticas ou sociais". O termo sociolinguística se consolidou apesar de o próprio Labov ter resistido, por vários anos, a empregá-lo por entender que o mesmo daria margens a se pensar numa linguística que não fosse social, ou seja, que não estudasse a língua no seio da sociedade que a cria e emprega na comunicação do dia-a-dia.

*hábitos pelos quais os membros de uma nação estão acostumados a se comunicar uns com os outros.*"); e Sturtevant 1947 ("*Uma língua é um sistema de símbolos vocais arbitrários pelos quais os membros de um grupo social interagem*") (a tradução é minha).

Segundo Suassuna (1995), os historiadores da Lingüística reconhecem que os primórdios da Sociolingüística estariam na "escola sociológica francesa", constituída por Saussure e seus herdeiros. Entre eles, Meillet reveste-se de importância especial, dadas as suas considerações acerca da relação estrutural lingüística x estrutura social, ainda que marcadas pela supremacia desta sobre aquela. Foi com este que uma concepção mais sociológica do falante e da língua encontrou uma formulação mais consistente e sólida. Não obstante aparecer como seguidor direto de Saussure, Meillet estabeleceu as raízes de uma *Lingüística da fala* e, no seu ponto de vista, o mestre pecou ao insistir na idéia de sistema, colocando a presença do homem no processo lingüístico em segundo plano. Nas palavras de Faraco (1991), "em Meillet, a língua não é mais concebida como um organismo vivo e autônomo (...) nem como uma realidade eminentemente psíquico-subjetiva (...), mas como um fato social: 'Tout fait de langue manifeste un fait de civilisation' ('Todo fato de língua manifesta um fato de civilização')"<sup>10</sup>.

Mas, é com a Teoria Variacionista Quantitativa de Labov que a Sociolingüística alcança maturidade e consistência, no sentido em que "repisa os caminhos traçados pelos lingüistas que propugnaram a necessidade de se estudar a língua sem dissociá-la da estrutura social, sem fazer abstrações de sua normal heterogeneidade" (id. *ibid*).

<sup>10</sup> O grifo é meu.

A grande questão com que Labov se depara (cf. Weinreich, Labov, e Herzog, 1968; Labov, 1972 e Tarallo, 1990) é a de que se uma língua tem que ser estruturada para funcionar eficientemente, segundo declaravam os estruturalistas da primeira metade do nosso século, como continuam então as pessoas a se comunicar durante os períodos de mudança do sistema lingüístico?

Para Labov, a resposta a esse paradoxo está em equacionar a heterogeneidade à noção de estrutura e funcionamento. Nas palavras de Weinreich, Labov e Herzog:

(...) se uma língua tem de ser estruturada de forma a funcionar eficientemente, como as pessoas continuam a se comunicar enquanto a língua sofre mudanças? (...) A resposta está no fato de não haver igualdade entre estrutura e homogeneidade. A solução para uma concepção racional de mudança lingüística está na possibilidade de se descrever sistematicamente a variação existente na língua de uma comunidade. (...) Um dos corolários de nossa abordagem é que em uma língua de uma determinada comunidade complexa (i. e. real), é a falta de heterogeneidade estruturada que seria considerada disfuncional. (a tradução é minha)<sup>11</sup>

Neste sentido, a Teoria Sociolingüística laboviana considera a associação entre estrutura e homogeneidade uma ilusão. Para ela, "a estrutura lingüística inclui a diferença [heterogeneidade] sistemática de falantes e de estilos através de regras que governam a variação na comunidade de fala; e, a seu ver, o domínio de uma língua pressupõe o controle de tais estruturas heterogêneas" (Tarallo, 1990). Na concepção laboviana, a heterogeneidade se constitui em propriedade inerente ao próprio sistema lingüístico, entendendo por sistema lingüístico aquele manifesto no *uso real de falantes reais em processo de comunicação*.

<sup>11</sup>(...) if a language has to be structured in order to function efficiently, how people continue to talk while the language changes? (...) The solution lies in the direction of breaking down the identification of structuredness with homogeneity. The key to a rational conception of language change is the possibility of describing orderly differentiation in a language serving a community. (...) One of the corollaries of our approach is that in a language serving a complex (i.e., real) community, it is absence of structured heterogeneity that would be dysfunctional. (Labov, 1972b:15-6).

Partindo, então, dos empregos concretos da língua falada<sup>12</sup> e focalizando como objeto de estudo a **variação** (diversidade<sup>13</sup>) lingüística, entendendo-a como "um princípio geral e universal das línguas, passível de ser descrita e analisada" (Mollica, 1992), a Sociolingüística pressupõe que toda variação é motivada, ou seja, controlada tanto por fatores internos ao sistema lingüístico quanto por fatores externos ao mesmo (fatores sociais como **escolaridade, sexo, idade, classe social, etnia**, para citar os mais considerados), de modo que "a heterogeneidade se delinea sistemática e previsível, sendo o emprego das formas lingüísticas de maneira alguma aleatório" (id. *ibid.*).

Dessa forma, a Sociolingüística Variacionista vem romper incisivamente com a tendência lingüística de tratar as línguas como sendo completamente uniformes, homogêneas, em que "as diferenças encontradas nos hábitos de fala de uma comunidade eram encobertas como 'variação livre' (Bright, 1974).

Outro conceito que a Sociolingüística laboviana vem sistematizar é o de mudança em progresso<sup>14</sup>. Segundo Faraco (1991:117),

<sup>12</sup> Língua falada é o vernáculo, não no sentido que habitualmente encontramos nos meios educacionais brasileiros em que o termo parece significar "norma culta"; ao contrário, aqui refere-se às variedades coloquiais informais: "é a língua que usamos em nossos lares ao interagir com os demais membros de nossas famílias; é a língua usada nos botequins, clubes, parques, rodas de amigos" (Tarallo, 1986)

<sup>13</sup> Conforme Bright (1974), "o termo diversidade lingüística não deve ser compreendido como referente a medidas puramente geográficas, nem tampouco a medidas lingüísticas simples, tais como o número de palavras compartilhadas. Ao contrário, refere-se à diferença entre partes de uma única sociedade ou nação, oposta à diferença entre sociedades e nações distintas, e a diferença entre variedades de uma única língua oposta à diferença entre línguas distintas".

<sup>14</sup> Tarallo e Duarte (1988) vêem a "mudança lingüística em progresso" como uma das contribuições mais significativas da sociolingüística laboviana ao pensamento contemporâneo. Segundo eles, este conceito vem contra os pressupostos bloomfieldianos sobre a regularidade da mudança fonológica e, em particular, sobre a impossibilidade de se observar seu processo em progresso.



a Sociolinguística dá nova força empírica ao princípio de que a mudança não se dá por mera substituição discreta de um elemento por outro, mas que o processo histórico, pressupondo sempre um quadro sincrônico de variação, envolve fases em que as variantes coexistem, ao cabo da qual uma termina por vencer a outra, podendo – por vicissitudes do processo – subsistirem áreas sociais e/ou geográficas em que a mudança não se dá.

Para Labov, fatores linguísticos e sociais encontram-se intimamente relacionados no desenvolvimento da mudança linguística, ou seja, a língua como um sistema muda em associação com mudanças na estrutura social. Diferente de Labov, Martinet (1955)<sup>15</sup>, por exemplo, só admite a intervenção de fatores externos ao sistema na interpretação da mudança linguística sob duas condições: que as possibilidades de condicionamentos estritamente internos estejam esgotadas e que os fatores externos sejam apenas de natureza linguística, isto é, advindos do contato entre línguas, dialetos e usos diferentes (Faraco, 1991).

Deve ficar claro, contudo, que não é qualquer diferença que pode estar indicando mudança. Muitas dessas diferenças são apenas variantes ("*diversas maneiras de se dizer a mesma coisa em um mesmo contexto, e com o mesmo valor de verdade*" – Tarallo, 1986) características da fala de cada grupo e nada têm a ver, em princípio, com mudança. Daí se dizer que nem toda variação implica mudança, mas que toda mudança pressupõe variação; em outras palavras, as mudanças emergem da heterogeneidade própria da língua, embora de nem todo fato heterogêneo resulte necessariamente mudança.

<sup>15</sup> MARTINET, A. 1955. *Économie des changements phonétiques; traité de phonologie diachronique*. Berna, Francke.

Finalmente, posto isso, podemos sintetizar a abordagem Sociolingüística Variacionista em alguns pontos que sustentam este estudo: para Labov, a língua individual é concreta e heterogênea; estrutura e heterogeneidade caminham sempre juntas; a variação tem caráter sistemático; mudança lingüística implica sempre variação ("mudança é variação"); a investigação lingüística é mais rigorosa quando se leva em conta o uso da língua no contexto social; a Sociolingüística Variacionista assume a existência de forças externas interagindo com forças internas na interpretação do funcionamento de uma dada língua; a aptidão para o emprego de regras variáveis é um dos aspectos da competência lingüística; o conhecimento da língua corresponde não só ao domínio da gramática dessa língua como também ao conhecimento do valor social atribuído às formas lingüísticas.

### 3- CONTRIBUIÇÕES DA SOCIOLINGÜÍSTICA

A sociolingüística tem comprovado a real heterogeneidade das línguas e demonstrado que cada variedade, além de sofrer condicionamentos lingüísticos, também está sujeita a condicionamentos sócio-histórico-culturais do grupo que a usa.

Ou seja, a maneira como cada grupo se constitui, como é sua posição na estrutura socioeconômica, como ele se organiza socialmente, quais seus valores e visão de mundo, quais suas possibilidades de acesso à escola e aos meios de informação constituem-se em categorias que influenciam na existência de uma diversidade de estilos sociolingüísticos.

Neste sentido, a abordagem sociolingüística traz em seu bojo duas grandes contribuições: uma refere-se à possibilidade de se verificar que a língua pode ser tomada como uma das balizas da compreensão da configuração social das comunidades; a segunda, vista de certa forma como decorrência da primeira, diz respeito a questões do ensino de língua materna nas escolas, ou seja, de que maneira a sociolingüística pode ajudar na solução ou diminuição de alguns problemas que enfrentamos nesta área.

Para ilustrar, vejamos dois trabalhos pioneiros de Labov.

O primeiro deles refere-se ao estudo que o linguista realizou na ilha de Marta's Vineyard (costa do Estado de Massachusetts, EUA)<sup>16</sup>, uma comunidade constituída basicamente de pescadores e agricultores, mas que vinha sendo invadida por veranistas do continente, o que, de alguma forma, interferia em padrões tradicionais de vida do ilhéu nativo.

Pesquisando o fenômeno de centralização dos primeiros elementos dos ditongos /ay/ e /aw/, Labov verifica uma forte correlação entre a atitude positiva em relação à ilha (fator sócio-cultural) e a pronúncia centralizada dos ditongos.

Em outras palavras, os falantes com grande identificação com suas raízes históricas e culturais e que mostravam uma atitude positiva em relação à ilha intensificaram um processo de centralização dos primeiros elementos dos ditongos /ay/ e /aw/, "numa espécie de reação à presença dos veranistas de fora e como forma de marcar sua identidade como membros da comunidade da ilha" (Faraco, 1991). Por outro lado, os jovens menos apegados à ilha e suas peculiaridades locais e com intenções de deixá-la, por exemplo, centralizavam menos.

O segundo trabalho constitui-se no estudo realizado por Labov sobre o inglês vernacular falado pelas comunidades negras ("*black English vernacular*" – BEV) das áreas da "*inner city*" de Nova Iorque, Boston, Detroit, Philadelphia, Washington, Cleveland, Chicago, St. Louis, São Francisco, Los Angeles, e outros centros urbanos<sup>17</sup>.

<sup>16</sup> V. Labov, 1972b.

<sup>17</sup> V. Labov (1972a).

Num primeiro momento de suas pesquisas sobre o BEV<sup>18</sup>, Labov se preocupa em analisar a fala dos negros no sentido de verificar as diferenças estruturais entre o inglês vernacular destes e a linguagem padrão, a fim de levantar novas discussões sobre propostas alternativas de ensino do inglês básico, considerando-se que ocorriam altos índices de repetência e abandono dos bancos escolares por parte desse grupo minoritário segregado sociolinguisticamente.

Desta forma, Labov reúne em seu trabalho sobre o dialeto negro informações que pudessem auxiliar principalmente os professores alfabetizadores bem como autores de livros didáticos que, ao que tudo indicava, ignoravam o uso de regras do sistema não padrão na mesma proporção em que os falantes do inglês não padrão desconheciam o uso de regras da norma culta (Tarallo, 1995).

Com relação ao português falado no Brasil, a exemplo do trabalho realizado por Labov sobre o inglês vernacular do negro, estudos sociolinguísticos também têm apontado as questões de linguagem como uma das maiores causas do fracasso escolar.

Com a expansão da rede escolar e a ampliação do número de cursos profissionalizantes, mesmo que em quantidade e qualidade insuficientes para responder às necessidades do país, mais estudantes, provenientes dos mais variados grupos sociais e étnicos e "usuários" de uma ampla gama de variações dialetais, vêm ocupando nossos bancos escolares.

<sup>18</sup> Esta pesquisa desenvolveu-se em três momentos. A respeito de quando e como se deu cada um deles, ver Tarallo (1995).

Todavia, o que se observa, mesmo com todas as mudanças positivas ocorridas no sistema educacional brasileiro nos últimos anos, é que a escola ainda não está preparada para enfrentar o desafio de instruir falantes do português não padrão, a fim de habilitá-los ao uso da norma culta, sem, no entanto, excluí-los das escolas e da sociedade através da segregação lingüística.

A escola, sob a influência do mito de que "somos um país privilegiado, pois do ponto de vista lingüístico tudo nos une e nada nos separa" (Bortoni-Ricardo, 1984, apud Assis, 1988), não reconhece (ou desconhece) a realidade heterogênea da língua portuguesa. Somando-se a isso (ou em virtude disso), a escola ainda opta somente pela veiculação da língua cultivada pela tradição gramatical, cristalizando a variedade padrão como a única correta e excluindo as demais como "formas incorretas", "erros", "desvios". Em outras palavras, a escola considera a norma padrão-culta lingüisticamente superior a todas as outras variedades, o que faz dela um local de "conversão da cultura e da linguagem dos grupos dominantes em saber escolar *legítimo* e local de imposição desse saber (bens simbólicos, na visão de Bourdieu) aos grupos dominados" (Soares, 1991).

Mas, através de trabalhos de descrição lingüística, a Sociolingüística Variacionista pode modificar tal posição, ou seja, mudar a visão de "erro" e de deficiência das formas não-padrão e, com isso, conscientizar a escola de que existem variedades (diferenças) lingüísticas e que não basta trabalhar a forma padrão da língua como se ela fosse a única existente mas, sim, formar seus alunos no sentido de adquirir uma flexibilidade lingüística necessária para o desempenho adequado nos mais variados atos lingüísticos.

Em outros termos, a Sociolinguística Variacionista pode mostrar à escola que o seu papel central com relação ao ensino de português é levar o educando a **conhecer e usar a norma culta sim** (e não somente reconhecê-la<sup>19</sup>), desde que a mesma se **acrescente ao português que ele fala** (e não que venha tentar substituir ou menosprezar sua variedade), mostrando a esse aluno quando uma forma ou outra deva ser utilizada, conforme as circunstâncias<sup>20</sup>.

No Brasil, estudos descritivos da fala rural e urbana e a formação de uma gramática da fala<sup>21</sup> têm sido implementados nas duas últimas décadas, contribuindo, assim, para o conhecimento da realidade linguística brasileira. A partir desses estudos, os sociolinguistas brasileiros pretendem expor um quadro da realidade de fala e do dado natural, reunindo um material de análise do vernáculo que pode ser aproveitado em pesquisas:

1) sobre **alfabetização**, na medida em que se procure "contrapor a gramática falada na comunidade à aprendizagem da escrita e da leitura" (Tarallo, 1995);

2) sobre o **ensino de língua portuguesa** nas séries mais avançadas, no sentido de promover um ensino com vitalidade, flexibilidade e criatividade de produção linguística, que faça o aluno "apreciar criticamente os diversos usos linguísticos e vivenciar experiências que ampliem a competência na manipulação das variedades" (Suassuna, 1995);

<sup>19</sup> Confira a Teoria da "economia das trocas linguísticas" de Bourdieu em Soares, 1991.

<sup>20</sup> Confira a Teoria do "bidialetalismo funcional" em Soares, 1991.

<sup>21</sup> Detalhes sobre a apresentação do *Projeto de Gramática do Português Falado* são apresentados em Castilho (1990).

3) e em pesquisas sobre a **organização de material didático** adequado que auxilie o professor no planejamento de um trabalho "dentro de uma situação marcada pela variação" (Moraes, 1984).

No que se refere aos trabalhos de descrição e análise do português falado no Brasil, as pesquisas com dados de falantes da região sul ainda são em número relativamente pequeno, se comparadas com de outras regiões do país, o que faz crescer a necessidade de mais estudos sociolingüísticos no sul (seja a nível fonético/fonológico, morfo-sintático, discursivo ou lexical), que venham dar subsídios para um ensino de língua portuguesa capaz de levar nossos alunos a usar a língua mais adequadamente em relação às diferentes instâncias da vida.

Reconhecemos, também, que apesar do esforço dos sociolingüistas brasileiros em registrar o Português falado no Brasil, pouco foi feito até agora com relação ao aproveitamento desse material sociolingüístico para a formulação de propostas que promovam novos direcionamentos do ensino da língua portuguesa<sup>22</sup>.

Como primeiro passo, poder-se-ia começar por (in)formar os professores, principalmente os de língua portuguesa, quanto às regras lingüísticas que fazem parte da competência comunicativa dos falantes de uma comunidade, ou seja, habilitá-los para "promover o necessário ajustamento entre a linguagem aprendida fora e dentro da escola" (Assis, 1988), num processo harmonioso de comunicação interdialetoal.

<sup>22</sup> Vale citar a proposta de trabalho de Suassuna (1995), desenvolvida em Recife com uma turma de 8ª série e que teve como "eixo do programa a variação lingüística". É inegável o valor desse trabalho, mesmo que pautado mais por princípios teóricos da abordagem sociolingüística variacionista.



## 4- METODOLOGIA

### 4.1. - Pressupostos básicos

Nesta pesquisa, adotamos os princípios metodológicos da Teoria Variacionista Laboviana, cujos objetivos são "estabelecer, via procedimento estatístico-quantitativo<sup>23</sup>, a gramática [da comunidade de fala]" (Tarallo, 1995). Percorrendo tais princípios, o pesquisador sociolinguista tem como tarefa empírica primária correlacionar a variável dependente (o fenômeno que está sendo estudado) com as variáveis independentes (grupos de fatores postulados como possíveis condicionadores da variável dependente, tais como o ambiente linguístico, estilo ou categorias sociais) (Chambers, 1995).

Segundo Labov (1972c), são cinco os princípios básicos que caracterizam a metodologia variacionista: princípio de mudança de estilo; efeito do monitoramento na fala; importância do vernáculo; formalidade na entrevista; primazia dos dados.

O primeiro princípio (o da mudança na fala) diz que não existem falantes de estilo único. Segundo ele, todo falante mostra uma variabilidade linguística conforme mudanças no contexto social e de tópico<sup>24</sup>.

<sup>23</sup> V. Capítulo 5 (Suporte Quantitativo).

<sup>24</sup> V. capítulo 2 (Suporte Teórico).

O princípio do efeito do monitoramento, por sua vez, estabelece que os estilos podem ser medidos pelo nível de monitoramento. Neste sentido, o menor grau de monitoramento é obtido em falas casuais e que envolvam assuntos de experiência pessoal. Decorre daí o princípio da importância do vernáculo, ou seja, o estilo cujo monitoramento na fala é mínimo, e do qual obtemos os dados mais sistemáticos à análise da estrutura lingüística.

Do princípio da formalidade na entrevista, depreendemos o fato de que qualquer observação sistemática por parte do entrevistador cria um contexto formal, fazendo com que o informante monitore sua fala. E, por fim, o princípio da primazia dos dados declara que as respostas às questões lingüísticas somente podem ser obtidas por entrevistas diretas, com o melhor equipamento de gravação.

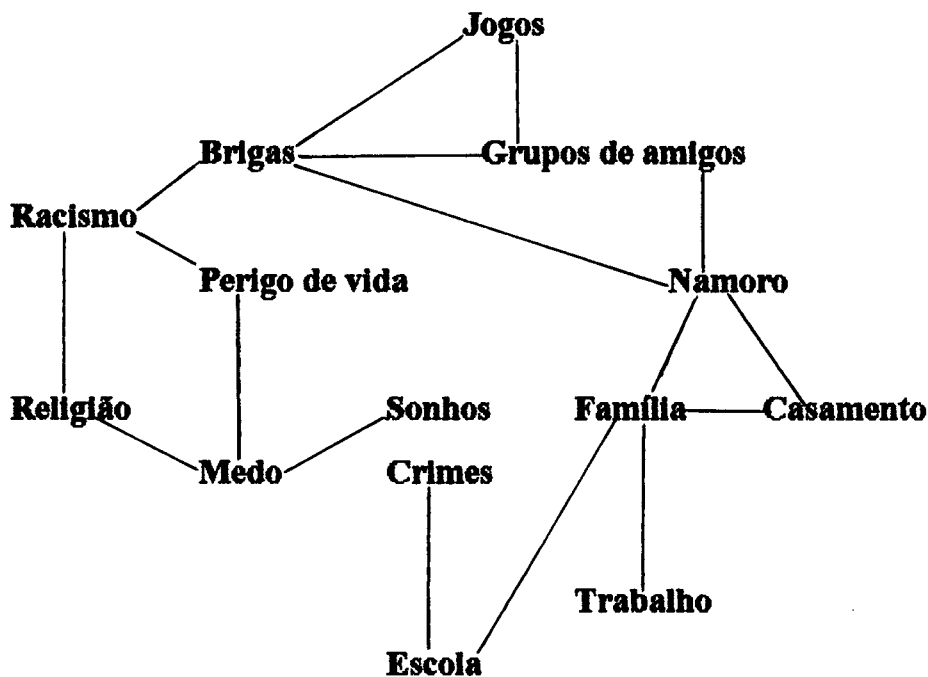
A partir desses princípios, Labov instaura um modelo de análise cuja metodologia se destina primordialmente a instigar, segundo ele, o único alvo de análise a ser considerado pela lingüística: a fala natural, não monitorada, em situações normais de interação social, isto é, o **vernáculo da comunidade de fala**. Diante disso, porém, surge um problema fundamental à pesquisa sociolingüística, o qual Labov chamou de **Paradoxo do Observador** e que estabelece o seguinte:

o objetivo da pesquisa lingüística na comunidade deve ser o de verificar como as pessoas falam quando não estão sendo sistematicamente observadas; ainda que só possamos obter tais dados através da observação sistemática" (a tradução é minha).<sup>25</sup>

<sup>25</sup> "the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation." (Labov, 1972b:209).

Em contrapartida, a metodologia laboviana desenvolve maneiras de se diminuir ao máximo o efeito de estranhamento causado pela observação sistemática, tentando conduzir os sistemas de entrevistas, por exemplo, de modo que o informante preste o mínimo de atenção possível a sua fala, permitindo, desta forma, que o vernáculo venha à tona. O principal mecanismo, então, é fazer com que o falante se entregue a assuntos que o envolvam emocionalmente, através das **narrativas de experiência pessoal**, levando-o a se preocupar tão somente com o que está narrando.

Para auxiliar na condução dos sistemas de entrevistas, Labov (1984) cria um módulo conversacional que se constitui em um grupo de questões focalizando um tópico particular, isto é, *brincadeiras de infância, premonições, perigo de vida, aspirações, etc.* Vejamos abaixo um esquema desse módulo.



O módulo funciona, então, como um recurso conversacional do qual o entrevistador se utiliza para elaborar um plano de entrevista, sendo que as perguntas formuladas se constituem em uma espécie de guia para se obter o estilo coloquial. Obviamente, este guia poderá sofrer adaptações, conforme as características da comunidade em contexto, bem como o entrevistador poderá administrar a ordenação dos tópicos e a ênfase dada a cada um deles conforme o gosto do informante por este ou aquele tema<sup>26</sup>.

Segundo Labov (id. *ibid.*), alguns cuidados são necessários no momento da realização da entrevista sociolinguística. O autor aconselha, primeiramente, que o entrevistador "conduza" a entrevista de modo que o falante não se detenha apenas a responder às perguntas, pois "é justamente o material adicional, aquilo que o informante fala a mais além da questão inicial, que se constitui na principal substância da entrevista" (a tradução é minha).

Labov sugere, ainda, que a postura do entrevistador, tanto no momento de primeiro contato com os informantes quanto no momento da entrevista propriamente dita, seja a de "aprendiz", ou seja, que o entrevistador assuma uma posição de menos autoridade do que a pessoa entrevistada. O pesquisador deve rever suas escolhas gramaticais e lexicais a fim de afastar quaisquer evidências de formação superior que venham a inibir o falante ou levá-lo a policiar sua fala. Nas palavras de Labov,

o entrevistador deve mostrar autoridade somente no que diz respeito à realização da gravação, no sentido em que pode sugerir a localização das pessoas no ambiente para que se possa obter o melhor som; se a entrevista for feita em ambiente aberto, o entrevistador deve impedir que o vento e o ruído do trânsito,

<sup>26</sup> A respeito disso, confira Labov (1984) em que o autor descreve uma entrevista encaminhada de forma a focar temas da vida cotidiana da informante, observando bem seus tópicos prediletos e voltando aos mesmos por inúmeras vezes.

por exemplo, prejudiquem a qualidade da gravação; se feita em lugar fechado, o entrevistador deve se preocupar com a possível interferência de ruídos de aparelhos eletrônicos como televisão, [telefone, liquidificador, etc.]. (a tradução é minha).<sup>27</sup>

Ademais, Labov propõe uma política de proteção dos dados e da identidade dos informantes. Segundo ele, o informante precisa estar ciente de que a entrevista está sendo gravada, bem como o entrevistador deve esclarecer aos informantes que o material gravado não será nunca divulgado e, caso seja utilizado em pesquisas, nunca a identidade do informante será revelada.

Feitas essas considerações, damos continuidade apresentando a amostra selecionada para a pesquisa.

#### 4.2. - Constituição da amostra

Para este estudo, especificamente, consideramos uma amostra de 72 informantes do Banco de Dados do Projeto VARSUL, distribuídos igualmente segundo quatro categorias sociais estratificadas pela equipe: etnia, sexo, faixa etária e escolarização.

Os grupos étnicos escolhidos foram o açoriano, o italiano e o eslavo, sendo 24 informantes (12 homens e 12 mulheres) de cada grupo. Duas faixas etárias estão sendo consideradas, uma de 25 a 49 anos e outra de + de 50 anos. Em termos dos anos de

<sup>27</sup> In one respect, the interviewer should retain his authority: in his own area of expertise in making recordings. He should feel free to suggest where the others might sit so that he can get the best sound; if outdoors, to move out of the wind or away from the street; if indoors, to turn off the sound of television set, turn off electric fans, or move away from a noisy motor. (Labov, 1984:40-1).

escolarização dos informantes, três recortes são estabelecidos: **primário, ginásio e segundo grau.**

Abaixo, temos o quadro 1 com o esquema de nossa estratificação.

**QUADRO 1 (ESTRATIFICAÇÃO)**

<b>ETNIA</b>	<b>Açoriana(24)</b>		<b>Italiana(24)</b>		<b>Eslava(24)</b>	
<b>SEXO</b>	M(12)	F(12)	M(12)	F(12)	M(12)	F(12)
<b>ESCOL.</b>	P/G/C	P/G/C	P/G/C	P/G/C	P/G/C	P/G/C
<b>IDADE</b>	-50 +50	-50+50	-50 +50	-50 +50	-50 +50	-50 +50

A seguir, tecemos algumas considerações sobre os grupos étnicos (e respectivas localidades) considerados em nossa análise.

#### **4.2.1 - As regiões e seus grupos étnicos<sup>28</sup>**

##### **4.2.1.1. – Florianópolis (SC)**

Florianópolis está situada na ilha de Santa Catarina, com uma extensão territorial de 451 km<sup>2</sup> e com uma população de 244.941 habitantes, sendo 229.566 na área urbana e 15.375 na área rural.

<sup>28</sup> Knies & Costa, 1995.

Fundada por Domingos Dias Velho, a cidade foi chamada inicialmente de Nossa Senhora do Desterro e teve como primeira etapa de colonização a vinda de moradores da Capitania de São Vicente. Seguindo a essa colonização, a ilha foi povoada por imigrantes das Ilhas dos Açores e Madeira, principalmente no período de 1748 a 1756, com o objetivo de ocupar efetivamente a região e impedir, com isso, que tropas espanholas dominassem o território.

Atualmente, a ilha de Santa Catarina tem sua população acrescida principalmente de gaúchos, argentinos, uruguaios e pessoas de outros municípios catarinenses.

#### 4.2.1.2. – Chapecó (SC)

A cidade de Chapecó está situada no oeste de Santa Catarina a 640 km de Florianópolis. Possui uma extensão territorial de 990 km<sup>2</sup> e uma população de 122.889 habitantes, sendo 96.599 na zona urbana e 26.290 na zona rural.

O território de Chapecó foi colonizado no início deste século, quando se deu o fim da revolução do Contestado, em 1910. Neste período, a Companhia de Colonização Bertaso e Maia atraiu, principalmente, colonos vindos de antigas colônias italianas do Rio Grande do sul para comprar seus lotes na região.

Inicialmente, Chapecó caracterizou-se como grande centro madeireiro, chegando a exportar para o Rio Grande do Sul e Argentina. Atualmente, a região se destaca pela atividade industrial na área de alimentos, sediando grandes indústrias como a Sadia, Ceval, Chapecó Avícola e outras. Tem como atividades básicas a agropecuária e a agroindústria.

Mesmo com o crescimento acentuado e a vinda de novos habitantes de diferentes etnias, Chapecó é uma cidade que prima pela manutenção de suas raízes italianas, através de suas organizações culturais e incentivo pelo ensino do italiano nas escolas.

#### 4.2.1.3. – Irati (PR)

O município de Irati está localizado no Planalto de Ponta Grossa, região sudoeste do Paraná e seu território é de 947 km<sup>2</sup>. Inicialmente, a localidade se chamava Covalzinho e só em 1899, com a inauguração da estrada de ferro São Paulo/Rio de Janeiro e com desenvolvimento da região, é que surgiu Irati. O nome Irati, na língua dos iraitins ou iratins (tribo indígena que habitava inicialmente a região), significa "rio de mel" (ira [mel]; ty [rio]).

Nascida no terceiro ciclo da economia paranaense, o ciclo da erva-mate, Irati também beneficiou-se do comércio de madeira de pinho. Atualmente, com o declínio dessas duas fontes de renda, Irati tem como base econômica a agricultura e o comércio voltado para esta.



O governo federal foi o responsável pela colonização da cidade, enviando para lá imigrantes poloneses, ucranianos, alemães e italianos, sendo que os de origem eslava (poloneses e ucranianos) impõem suas características culturais aos demais.

A seguir, apresentamos a descrição do processo de coleta e armazenamento dos dados.

#### **4.3. - Processo de coleta e armazenamento dos dados**

Como dissemos acima, nossa amostra faz parte do Banco de Dados do Projeto VARSUL (Variação Lingüística Urbana na Região Sul), que tem como meta "armazenar e colocar à disposição dos pesquisadores interessados amostras de realizações da fala de habitantes enraizados em áreas urbanas sócio-culturalmente representativas de cada um dos três estados da Região Sul do Brasil" (Knies & Costa, 1995).

É objetivo primeiro do Projeto VARSUL fornecer subsídios para que se desenvolvam pesquisas da variedade lingüística falada nos centros urbanos da região sul, a fim de que se possa repensar políticas da linguagem, replanejar o ensino da língua portuguesa, bem como testar modelos teóricos em vigor na área de lingüística.

Em áreas como a Sociologia, Antropologia e Psicologia, o Banco de Dados também será útil na medida em que contribuirá para uma melhor compreensão do morador do sul do país, considerando que a linguagem reflete todos os aspectos do ser humano: seus estados mentais, sua cultura, seus valores e suas relações sociais.

Integram o Projeto VARSUL quatro universidades do sul: Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Federal do Paraná (UFPR) e a Pontifícia Universidade Católica do Rio Grande do Sul (PUC - RS), as quais se unem na elaboração do Banco que, até 1996, constará da amostra representativa da fala de 24 moradores de 12 cidades, quatro em cada estado da Região Sul (SC – Florianópolis, Chapecó, Blumenau e Lages; RS – Porto Alegre, Flores da Cunha, Panambi e São Borja; PR – Curitiba, Londrina, Pato Branco e Irati), totalizando 96 entrevistas por estado, e 288 no acervo total.

Ao final dessa etapa do Projeto, cada uma dessas instituições responsáveis pela implementação do Banco terá à disposição de seus professores, alunos de graduação e pós-graduação, enfim, pesquisadores que se interessem pela pesquisa variacionista o seguinte material:

- a) cópia em fita cassete da entrevista com cada um dos informantes;
- b) cópia em disquete das transcrições dessas entrevistas;
- c) cópias impressas e encadernadas das referidas transcrições;
- d) cópia das fichas sociais dos informantes;
- e) programa Editor VARSUL - Engesis para leitura e impressão das entrevistas;
- f) programa Interpretador VARSUL - Engesis para a localização automática de ocorrências de interesse do pesquisador<sup>29</sup>.

<sup>29</sup> As instruções de como utilizar esses programas computacionais estão detalhadamente descritas no Manual do Usuário. (Knies & Costa, 1995)

#### 4.3.1 – A coleta dos dados

Todo o processo de coleta dos dados que compõem o banco de dados VARSUL segue os princípios da metodologia sociolinguística laboviana.

A coleta se dá em dois momentos. No primeiro deles, o entrevistador vai a campo para localizar o informante que preencha os requisitos básicos necessários (além da faixa etária, escolaridade e descendência étnica) para compor a amostra.

Assim, o possível entrevistado deve falar apenas o português (exigência para os informantes das capitais mas não nas áreas bilingües); deve ter morado na cidade pelo menos 2/3 de sua vida; e não ter morado fora da região por mais de um ano no período de aquisição da língua nativa (2 a 12 anos). (Knies & Costa, 1995)

Detectado o informante com um perfil compatível, o entrevistador se apresenta como um pesquisador interessado em desenvolver um trabalho sobre "como vive, o que pensa, o que faz, como se diverte, em que acredita o verdadeiro morador da cidade"<sup>30</sup> (id. *ibid*).

<sup>30</sup> Em momento algum o entrevistador dirá ao informante que o objetivo do trabalho será o de análise da fala, pois estará pondo em risco toda a pesquisa, uma vez que o informante seguramente se policiará.

Uma vez obtida a permissão para a realização da entrevista, o entrevistador inicia imediatamente a coleta das características sociais do falante<sup>31</sup>, as quais ficarão registradas na ficha social do mesmo. É nesse momento que o entrevistador reunirá as primeiras informações sobre o falante que o auxiliarão no trabalho de entrevista propriamente dito (na elaboração de um roteiro de assuntos, por exemplo), além do que esse contato já estará contribuindo para minimizar o conhecido **Paradoxo do Observador**.

Coletados esses dados, é marcada a segunda entrevista que dura em média uma hora, e que se realiza preferivelmente na casa do próprio falante para que o mesmo se sinta o mais à vontade possível, não obstante a interferência normal da presença do gravador.

Nessa segunda entrevista, é aplicada a técnica pela qual o falante é levado a discorrer sobre sua cidade, sua história de vida, seus valores, suas crenças, seu folclore (**narrativas de experiência pessoal**) e outros assuntos gerais/particulares que venham a ocorrer durante a gravação.

#### 4.3.2 – O armazenamento dos dados

Após a coleta das 24 entrevistas, são feitas cópias das fitas e inicia-se a transcrição das mesmas a qual se dá num sistema trilinear (somente a fala do informante)<sup>32</sup>.

<sup>31</sup> Os temas abordados nesta primeira fase da coleta, bem como o modelo das fichas sociais, aparecem devidamente esclarecidos no Manual do Usuário. (Knies & Costa, 1995)

<sup>32</sup> No capítulo 3 do Manual do Usuário, tem-se uma descrição completa do Roteiro de Transcrição utilizado pelos bolsistas e professores do Projeto VARSUL.

Na **primeira linha**, registra-se a sintaxe real da fala dos informantes, considerando-se todas as hesitações e interrupções. A ortografia oficial, inclusive os sinais de pontuação, são utilizados nessa primeira linha. Na **segunda linha** da transcrição, são marcadas as pausas e registrados os aspectos fonéticos variáveis previsíveis ou inesperados detectados nas entrevistas. Na **terceira linha**, procede-se a uma classificação morfossintática dos itens lexicais, bem como à marcação de aspectos supra-segmentais do tipo ênfase e velocidade da fala.

A seguir, transcrevemos um trecho de uma das entrevistas trabalhadas em nossa pesquisa (SCFLP08FBPRI), caracterizando o sistema de transcrição em três linhas.

E \*Eu queria que a senhora me contasse como é que era a sua vida lá no Rio Vermelho quando a senhora era criança.

F \*Eu nasci lá e me criei aqui em Florianópolis, né? (est) \*De lá não sei muita

0 00 1 d u~ :0 0  
n v a c n v a p s av,pi p a a v n

coisa. \*E depois [que eu]- [que eu]- eu vim pra cá com uns dois anos, né? (est)

3 0\$ 0+ 0+ 00+ j j0 1  
s c a c n c n n v p a p d d s av,pi

\*Morava com a minha avó e o meu avô, (inint) minha mãe e pai. \*Meu pai

w0+ 0+ :5 1 1  
v p d n s c d n s n s c s n s

\*Meu pai faleceu e eu fiquei com a minha mãe. \*Eram [três]- três irmãos. \*Eu e

5 w0+ 1 0 \$ + \$1  
n s v c n v p d n s v d d s n c

mais uma irmã e um irmão. \*Daí a minha avó quis vir pra Florianópolis, aí

\$5 x x l 0+ 5 \$ m 0 00 0+  
n d s c d s pa,pi d n s v i p s a,pi

vieram. \*Mas eles chegaram aqui, nesse tempo era água de torneira, eles estavam

u0 3 + \$ u0 0 0+ d:5 x 0 0 \$ 00 u 0  
v c n v a pn s v s p s n v

acostumados lá com água corrente, né? \*Aí chegaram aqui, (inint) foram pro

u\$ 0 u x' t 0 l 00+ 0 u0 0  
i a p s j av,pi a,pi v a v pd

Hospital de Caridade. \*Lá, eles morreram. (...)

\$ w00 d l 0 j x' u0 l  
s# p# s,s a n v

Terminada a transcrição, é iniciada a etapa de digitação do material em editor de texto especialmente desenvolvido para o projeto VARSUL. Finalmente, o material é impresso, revisado, corrigido, reimpresso e encadernado. A partir daí, todo o material está pronto para consulta.<sup>33</sup>

<sup>33</sup> No Manual do Usuário, encontramos as normas para o uso do Banco de Dados Lingüísticos VARSUL.

#### 4.4. - Levantamento e codificação dos dados

Após a delimitação do fenómeno a ser analisado, procedemos à identificação e ao levantamento dos dados relevantes.

Foram localizados nas 72 entrevistas um total de 938 ocorrências de predicativo/participio, sendo abordados casos do seguinte tipo: "eles eru muito *bons* pra mim" (FLPSCMPA); "nóys somos *casado*" (FLPSCFGA); "O Mauricio e o Marcelo são *gêmeos*" (CHPSCMCB); "São *dura*" (FLPSCMCA); "dois foru *criado* ali" (FLPSCFPA); "meus pais eram *peessoas honestas*" (IRTPRFGB); "são *tudo casado*" (CHPSCMGB); "é muito *maroto* aqueles ensaio" (CHPSCMCA); "são até *loucos*" (IRTPRMCA); "os piá eram *horrrível*" (IRTPRFPA); "os piquininho ficum *sentadinho*" (IRTPRFGA); "meus pais eru bem *religioso*" (FLPSCMGA); "são *holandeses*" (CHPSCMGB).

Na análise probabilística, trabalhamos com 768 ocorrências, sendo desconsideradas as 170 restantes: dados com sujeito coletivo ou genérico (*a gente, o povo, todo mundo*); casos em que ocorriam elisões ("era mais bem *dada as* matéria") ou neutralizações do tipo "nem *presos são*"; casos de audição duvidosa; e, ainda, casos em que eram detectadas possíveis influências do entrevistador. Estes 170 dados poderão ser retomados em análises futuras.

Uma vez identificados os dados, organizamos os mesmos de forma apropriada ao processo de codificação, o qual consiste em transformar em código identificável pelos

programas computacionais (VARBRUL)<sup>34</sup> tudo o que queremos que seja analisado. Antes, porém, escolhemos símbolos (motivados) para as variantes da variável dependente e para cada um dos fatores dos grupos de fatores ou variáveis independentes. Concluída a codificação, o formato de entrada dos dados no sistema VARBRUL assume as formas exemplificadas a seguir:

(1iePArpCiaAFPa/	Nós ficamo tão <b>frustrados</b>
(ØSØPtrpCØpIFGb2	foro tudo <b>preso</b>
(1icPArpCiaEMCb/	os resultados não eram tão <b>grandes</b> assim

Na primeira coluna, encontra-se um parêntesis de abrir que indica início de um dado para o programa de correção dos dados<sup>35</sup>. Na segunda coluna, temos a codificação da variável dependente que, no nosso caso, é binária, com os símbolos Ø e 1, indicando, respectivamente, ausência e presença de marca formal de plural nos predicativos/participios analisados.

Da 3ª à 16ª coluna, apresentamos a codificação dos 14 grupos de fatores ou variáveis independentes. Por exemplo: na 3ª coluna, 1 indica que o predicativo analisado se encontra em construção isolada (não-seriada) e S indica que o predicativo analisado é precedido de outro predicativo sem nenhuma marca formal de plural (no caso de construção seriada).

Na 4ª coluna, e significa que o sujeito é explícito com marca semântica de plural (nós e numerais isolados); c significa que o sujeito da construção é explícito com marcas em

<sup>34</sup> V. capítulo 5 (Suporte Quantitativo).

<sup>35</sup> Confira a descrição dos programas que compõem o sistema VARBRUL (Suporte Quantitativo).



todos os elementos (nos últimos elementos ou último neutralizado); Ø significa que o sujeito da construção não é explícito. Na 5ª coluna, P indica que o verbo apresenta marca formal de plural.

Na 6ª coluna, A indica que o predicativo analisado é do tipo adjetivo/formas participiais de um elemento e t quer indicar que o predicativo apresenta mais de um elemento com os itens *tudo/todo*.

Nas 7ª, 8ª e 9ª colunas, r, p e C significam, respectivamente, que a formação de plural do predicativo é do tipo regular, que a forma singular do predicativo é paroxítona e que a ordem dos elementos na estrutura é canônica.

Na 10ª coluna, I significa que há presença de intensificadores entre o verbo e o predicativo e Ø indica a ausência de material interveniente. Na 11ª coluna, a indica que o predicativo está em uma estrutura ativa e p indica que o predicativo está em uma estrutura passiva.

Na 16ª coluna temos a codificação da última variável lingüística, sendo que os símbolos 2 e "/" indicam, respectivamente, distância do sujeito correferente ao sujeito zero de 6 a 10 cláusulas e fator não se aplica (caso em que o sujeito da construção é explícito).

Nas colunas 12, 13, 14 e 15, temos a codificação das variáveis sociais.

Na 12ª, A, I e E indicam o grupo étnico dos informantes que são, respectivamente, açoriano, italiano e eslavo. Na 13ª, os símbolos M e F indicam que o dado foi produzido, respectivamente, por um informante do sexo masculino ou feminino.

Na 14ª coluna, P, G e C significam, respectivamente, que o dado foi produzido por um informante com escolarização primária, ginásial ou colegial. E, finalmente, na 15ª coluna, o código a indica que o informante é da faixa etária de 25 a 50 anos e b, que o informante faz parte da faixa etária de + de 50 anos.

Uma vez digitados, os dados foram, então, submetidos ao programa de análise de regra variável (VARBRUL).

Na análise, identifiquei a entrevista da qual retiro os dados segundo a codificação estabelecida pelo VARSUL, preservando, dessa forma, a identidade dos informantes. Assim, o dado é seguido de uma seqüência de cinco letras e dois algarismos, mais a identificação do informante quanto ao seu sexo, faixa etária e escolaridade, conforme exemplificamos a seguir:

O meu livro era *tudo cortado* (FLPSC01FPRIA)

eles tão *ispalhado* (CHPSC11MGINA)

eles tão *disponíveis* (IRTPR23MCO LB)

Sendo que:

FLP = Florianópolis

CHP = Chapecó

IRT = Irati

SC = Santa Catarina

PR = Paraná

01, 11 e 23 = números atribuídos a cada entrevista na cidade

F = feminino

M = masculino

PRI = primário

GIN = ginásial

COL = colegial<sup>36</sup>

A = faixa etária de 25 a 50 anos

B = faixa etária de + de 50 anos

<sup>36</sup> Por questões práticas, estou mantendo a antiga identificação estabelecida para informantes com segundo grau, ou seja, "COL" (de colegial), ao invés de "SEG", como ficou determinado pela equipe do VARSUL.

## 5- SUPORTE QUANTITATIVO

### 5.1. - Introdução

Vimos no capítulo 2 que a Sociolinguística Variacionista vem firmar o tratamento da variabilidade linguística, sistematizando-a de modo a desmistificar a visão que se tinha de que a fala concreta é caótica e, por isso, impossível de ser analisada. Com sua abordagem teórico-metodológica, Labov quer mostrar justamente o contrário: que a heterogeneidade linguística é sistemática, não aleatória e que, ao contrário do que se pensava, a aparente confusão do dado real de fala é, na verdade, ordenada, regulada, seguindo padrões sociais e linguísticos de variação.

Tradicionalmente, "todas as unidades linguísticas – fones, fonemas, morfemas, frases e cláusulas – eram tratadas como invariantes, discretas e qualitativas" (Labov, 1966, apud Chambers, 1995). A Teoria Variacionista, contudo, reverte esta situação na medida em que trata a variável linguística como uma *estrutura variante* (no sentido em que é realizada diferentemente em situações diversas), *contínua* (certas variantes têm seu valor social atrelado ao distanciamento ou diferenciação com relação à variante padrão) e *quantitativa* (no sentido em que a variável tem sua significação determinada não simplesmente pela presença ou ausência de suas variantes mas por suas frequências relativas) (id. ibid.). Daí denominar-se a abordagem laboviana de Teoria Variacionista Quantitativa.

Com o advento da metodologia variacionista quantitativa tem-se, então, a oportunidade de depreender e avaliar em termos quantitativos o efeito de fatores lingüísticos e extralingüísticos que condicionam os fenômenos de variação e mudança na fala, ou seja, "avaliar o *QUANTUM* cada categoria postulada contribui para a realização de uma ou outra variante das formas em competição" (Naro, In Mollica, 1992). Trata-se, portanto, de um modelo de análise de grande número de dados da fala concreta que possibilita ao lingüista descrever, estatisticamente, a variedade do fenômeno em estudo.

Entretanto, não queremos dizer com isso que a Teoria Variacionista opta apenas por uma descrição quantitativa do seu objeto de estudo. Ao contrário, ela combina as abordagens quantitativa e qualitativa na descrição de fenômenos lingüísticos, sendo ambas complementares uma à outra.

## 5.2. - Os modelos quantitativos

Inicialmente, alguns modelos matemáticos foram criados para substituírem o então falacioso modelo de frequências brutas, cujos cálculos não levavam em conta as inter-relações existentes entre os fatores que atuam numa regra variável. A questão central com que se deparavam os pesquisadores na época era a seguinte: como eles poderiam tratar o efeito de uma categoria separadamente se tal categoria nunca se apresenta isolada nos dados? Era necessário, dessa forma, que se criasse um "instrumental capaz de produzir o efeito global que se verifica nos dados empíricos" (id. *ibid.*).

Em 1969, Labov propõe o primeiro modelo que viria, então, tentar solucionar o impasse: o **modelo aditivo** do tipo  $F_t = F_0 + F_1 + F_2 + \dots + F_n$ , onde  $F_t$  é a soma dos fatores contextuais,  $n$  é o número total de categorias relevantes e  $F_0$  é a média global de aplicação da variante sob estudo (Naro, In Mollica, 1992). Mas, devido a problemas de natureza técnica, o mesmo foi abandonado e substituído pelo **modelo de efeito simultâneo de fatores independentes**, elaborado por Henrietta Cedergren e David Sankoff em 1974, cuja formulação é do tipo multiplicativo:  $P_t = P_0 \times P_1 \times P_2 \times \dots \times P_n$  ( $P_t$  representa a probabilidade de aplicação da regra,  $P_0$  é uma probabilidade "input" cujo papel é o mesmo do  $F_0$  do modelo aditivo e  $P_1, P_2$ , etc. são as "probabilidades fatoriais" (Naro, 1981).

Este último sofreu algumas reformulações até que em 1975 Pascale Rousseau e David Sankoff chegam ao **modelo logístico**<sup>37</sup> que vem, finalmente, ao encontro das expectativas dos pesquisadores variacionistas. Trata-se, então, de um modelo cuja base é a "estatística destinada a trabalhar com categorias que se inter-relacionam" (id. ibid.). Como uma síntese dos modelos anteriores, o modelo logístico apresenta a seguinte fórmula:

$$\frac{P_t}{(1 - P_t)} = \frac{P_0}{(1 - P_0)} \times \frac{P_1}{(1 - P_1)} \times \frac{P_n}{(1 - P_n)}$$

Neste modelo, os símbolos são interpretados da mesma forma que o são os do modelo anterior, com uma diferença: este modelo, ao contrário do seu antecessor, "não segue o modelo de eventos independentes da teoria da probabilidade, não se justificando mais a utilização do termo "probabilidade" para os  $P_i$ , sendo preferível o termo 'peso relativo' (Naro, In Mollica, 1992).

<sup>37</sup> Para maiores detalhes acerca do modelo logístico, confira Labov, 1972a e Naro, 1981 e 1992.

A partir do modelo logístico e para a implementação do mesmo, David Sankoff desenvolve o programa computacional VARBRUL (**V**ariable **R**ule **A**nalysis), cuja função é medir os efeitos de variáveis independentes sobre a variável dependente, ou seja, o fenômeno em estudo. Como instrumento de análise, o VARBRUL "quantifica a influência relativa de cada variável, atribuindo pesos probabilísticos devidos aos diversos fatores das diversas variáveis" (Scherre, 1988) consideradas na pesquisa.

O sistema VARBRUL já passou por várias versões desde sua criação e, atualmente, as utilizadas são as de 1988 e 1992, com essa segunda podendo ser executada sem o co-processador aritmético e podendo receber pelo menos o dobro de dados que as versões anteriores. A versão de 1992<sup>38</sup>, utilizada em nosso trabalho, compõe-se de dez programas: CHECKTOK, READTOK, MAKECELL, IVARB, TVARB, MVARB, TSORT, TEXTSORT, CROSSTAB e COUNTUP.

Há, ainda, o editor QUICH EDITOR<sup>39</sup> em cujos arquivos ficam armazenados os dados brutos de entrada (os dados do fenômeno analisado) codificados adequadamente<sup>40</sup> – **arquivo de dados** –; os dados de controle (de especificação de todos os símbolos usados na codificação – **arquivo de especificação** – e os de condição – **arquivo de condições**<sup>41</sup>); e os dados gerados pelos outros programas.

<sup>38</sup> Como leitura complementar, sugerimos Scherre et. al., 1992 (mimeo); Scherre, 1993 (mimeo); Pinto & Fioretti, 1992 (mimeo).

<sup>39</sup> Este editor de dados não faz parte do pacote VARBRUL. O mesmo foi cedido pelo Professor Anthony J. Naro.

<sup>40</sup> Cf. o capítulo 4 (Metodologia).

<sup>41</sup> Para maiores detalhes sobre os arquivos de especificação e condição, ver Scherre, 1993:13-9.

Com exceção do programa IVARB, todos os outros três (CHECKTOK, READTOK e MAKECELL) designam-se a preparar os dados que serão submetidos àquele. O CHECKTOK toma como entrada o arquivo de especificação e o arquivo de dados e fornece um **arquivo corrigido** como saída. Caso haja erros de codificação no arquivo de dados, o programa fornece o arquivo de erros *checktok.err* e se houver erros no arquivo de especificação, o programa cria o arquivo *facspec.err*.

O READTOK, por sua vez, tem como arquivo de entrada o arquivo corrigido pelo checktok e fornece na saída um **arquivo de ocorrências**.

O MAKECELL toma o arquivo de ocorrências mais o arquivo de condições como entrada e fornece na saída o **arquivo de células** "em que constam as percentagens associadas a cada fator e as especificações das células resultantes das combinações dos grupos de fatores postulados" (Votre, 1991). Se houver fatores categóricos (100% de aplicação ou não da regra [*knockout*] ou grupos constituídos de um só fator [*singleton groups*]), o programa acusa erro e alerta que o IVARB somente processará os dados após os ajustes devidos. É neste arquivo que o usuário opta pelo valor de aplicação da variável dependente, se 1 (aplicação da regra) ou  $\emptyset$  (não aplicação da regra). No nosso caso, estamos trabalhando com a presença da marca formal de plural, logo valor de aplicação 1.

Por último, temos o IVARB que recebe o arquivo de células (sem fatores knockout) e fornece como saída o **arquivo de probabilidades** com os pesos relativos associados a cada fator das variáveis independentes, correlacionados à variável dependente em sua forma binária de variantes.



O TVARB e o MVARB são as versões enéarias, o primeiro para variáveis dependentes com três variantes e o segundo para variáveis dependentes com quatro ou cinco variantes.

Como já mencionamos, o Pacote VARBRUL apresenta, além dos já discutidos, mais quatro Programas: o TSORT, o TEXTSORT, o CROSSTAB e o COUNTUP. Os dois primeiros procuram e reordenam dados ("o TSORT efetua procura em função da cadeia de codificação e o TEXTSORT, em função do que foi digitado após a cadeia de codificação" – cf. Scherre, 1992 e 1993).

O CROSSTAB, ou **Programa de tabulação cruzada**, realiza o cruzamento das **percentagens** atribuídas a dois grupos de fatores. Como entrada, ele recebe um arquivo de células e produz, na saída, um arquivo contendo o cruzamento dos dois grupos de fatores especificados. Este Programa nos possibilita detectar possíveis interferências entre dois ou mais grupos de fatores.

O COUNTUP, ou **Programa de contagem geral de dados**, fornece a frequência geral dos dados por variável e por fator dentro de cada variável, inclusive do fator não se aplica. Ele recebe como entrada um arquivo de ocorrências gerado pelo READTOK e produz um arquivo de saída (**arq. tup**). Além de permitir uma visão global da distribuição dos dados, este programa pode apontar possíveis erros na codificação dos dados.

O programa trabalha com nível de significância de 5 por cento (*threshold* .05) e os resultados que apresentam nível superior a esse são considerados estatisticamente não significativos.

O programa gera um log likelihood que mede a diferença entre as frequências esperadas e as observadas, associado a um determinado nível de significância. Desta forma, o nível de significância .000 é considerado o ideal uma vez que "indica uma certeza estatística de os valores gerados pelo modelo estarem adequados aos valores observados" (Scherre, 1988).

Segundo Sankoff (1988 – apud Scherre, 1993), "um dos aspectos mais interessantes do IVARB consiste no fato de que ele trabalha com níveis diversos de análises, efetuando comparações progressivas entre os pesos relativos atribuídos aos diversos fatores das variáveis independentes e fazendo seleção estatística de variáveis a cada passo da análise".<sup>42</sup>

Neste sentido, o IVARB (na opção **STEP UP/STEP DOWN**) iniciaria a rodada com a sequência **step up** a partir do nível zero, apresentando, desta forma, o input global da regra. Neste nível, portanto, "o programa calcula a média global de aplicação da regra corrigida que tem sido vista como 'a probabilidade de aplicação da regra quando o efeito de todos os fatores de todas as variáveis é neutro'" (Scherre, 1993).

<sup>42</sup> O grifo é nosso.

No nível 1, o programa calcula os pesos relativos dos fatores de cada variável, atribuindo-lhes os respectivos valores de log likelihood e nível de significância e finaliza selecionando o primeiro grupo de fatores estatisticamente mais relevante.

Após a seleção da primeira variável, o programa vai para o segundo nível de análise onde combina, par-a-par, a variável selecionada com cada uma das outras variáveis consideradas na pesquisa. No nível 2, portanto, os valores de log likelihood e de nível de significância são atribuídos a cada par de variáveis. Ao final, após testar a relevância estatística de cada combinação feita, o programa seleciona a segunda variável mais significativa estatisticamente.

Em seguida, no nível 3, o programa combina as duas variáveis já selecionadas com as restantes, três a três, executando o mesmo processo descrito acima até chegar ao término, quando seleciona a terceira variável mais relevante e assim, sucessivamente, até não restar mais nenhum grupo de fatores que o programa considere significativo na aplicação da regra. Ao final da rodada **step up**, o programa nos informa que mais nenhuma variável será selecionada como relevante e nos mostra a síntese dos grupos de fatores apontados como os mais significativos em ordem de seleção.

O programa, então, inicia uma rodada inversa ao **step up**, ou seja, o **step down** que vai do nível N, onde constam todos os grupos de fatores, até o nível 1, descendo um nível de cada vez. No **step down**, o programa testa novamente todas as variáveis - as selecionadas e as não selecionadas no **step up** -, ou seja, ele verifica se as variáveis selecionadas não são eliminadas e se as variáveis não selecionadas são eliminadas.

A análise em diversos níveis nos permite verificar a interferência entre variáveis, provocada por codificação superposta ou enviesada. Não ocorrendo sobreposições entre variáveis, as probabilidades do nível 1 ao nível N permanecem as mesmas, ou bastante próximas, o que se constituiria na situação ideal.

Todavia, a realidade lingüística nem sempre é essa, podendo ocorrer variáveis sobrepostas. Neste caso, "o programa atribui pesos probabilísticos **relativos em função da importância estatística [de cada variável]**, com base, por exemplo, numa distribuição equilibrada dos dados entre os diversos fatores de uma variável" (Scherre, 1988).<sup>43</sup>

Para efeitos de análise, bem como para montar suas tabelas e gráficos, o pesquisador deve considerar os resultados (pesos relativos) que aparecem no conjunto de todas as variáveis selecionadas, quase ao final da rodada **step up**.

A leitura dos pesos relativos associados às variantes de uma variável binária (duas variantes) supõe dois pólos, um próximo a 100, fortemente favorecedor da regra de aplicação do fenômeno em estudo, e um próximo a zero, que desfavorece a aplicação da regra. No meio desta linha (.50 de peso relativo), bem como em suas imediações, localiza-se a área neutra; e todo fator em estudo que tiver seu peso relativo nessa área é considerado neutro ou irrelevante.<sup>44</sup>

<sup>43</sup> O **negrito** é nosso.

<sup>44</sup> Sankoff enfatiza a importância da relação entre as grandezas e não das grandezas em si. Assim, um fator com peso relativo próximo a .50 não necessariamente indica influência neutra. Isto vai depender dos valores associados ao(s) outros(s) fatores. (afirmação feita à viva voz, por ocasião de uma palestra proferida por Shana Poplack na UFRJ).

Vale ressaltar, por fim, que os resultados numéricos fornecidos pelo VARBRUL só têm valor estatístico, não sendo, desta forma, absolutos. A estatística é apenas um instrumental (valioso e seguro, sem dúvida) que nos auxilia a entender um pouco mais o comportamento dos fenômenos lingüísticos variáveis. Cabe ao lingüista, e somente a ele, atribuir um valor lingüístico aos valores numéricos alcançados, ou seja, interpretar os números segundo uma visão teórica da língua. É a análise correta dos números, e não os números em si, que leva a um entendimento do sistema lingüístico.

Mostramos, aqui, um pouco do instrumental estatístico utilizado nas pesquisas variacionistas, procurando dar uma idéia geral de como os dados são processados, bem como a importância que tem o sistema VARBRUL de computadores para a agilização do trabalho quantitativo que envolve a metodologia laboviana. Procuramos, também, apontar as obras que consideramos imprescindíveis ao manuseio do Programa Computacional. Em seguida, apresentamos a análise das variáveis trabalhadas neste estudo, submetidas ao tratamento estatístico da metodologia variacionista.

## 6- ANÁLISE DAS VARIÁVEIS LINGÜÍSTICAS

Vamos desenvolver, neste capítulo, a análise das variáveis lingüísticas<sup>45</sup> que julgamos atuar sobre a concordância de número no predicativo e no particípio passivo. Discutiremos aqui as variáveis selecionadas como mais significativas pelo Programa Computacional VARBRUL, bem como as menos significativas, traçando um paralelo com os resultados obtidos por Scherre (1991).

Para nossa análise, consideramos um número de dez variáveis lingüísticas, listadas abaixo<sup>46</sup>:

- 1) Paralelismo formal das seqüências de predicativos/particípios no discurso;
- 2) Características formais do sujeito;
- 3) Características formais do verbo;
- 4) Estrutura do predicativo;
- 5) Processos morfofonológicos de formação de plural;
- 6) Tonicidade dos itens singulares;
- 7) Ordem dos elementos na estrutura;
- 8) Material interveniente entre o verbo e o predicativo;

<sup>45</sup> Executamos rodadas com as variáveis lingüísticas e sociais juntas e separadamente. Apresentamos aqui os resultados das rodadas separadas. Os resultados de uma e outra rodada não apresentam diferenças estatísticas significativas.

<sup>46</sup> As variáveis 1-9 são consideradas com base em Scherre (1991).

- 9) Tipo de estrutura;
- 10) Distância do sujeito correferente.

A seguir, mostraremos cada variável lingüística e seus grupos de fatores, o objetivo da análise de cada uma delas, as hipóteses que estão por trás das mesmas e os resultados estatísticos alcançados. Apresentaremos, primeiramente, as variáveis lingüísticas estatisticamente relevantes (em ordem de significância) para, em seguida, abordarmos aquelas selecionadas como não significativas ao fenômeno em estudo.

#### **6.1. – VARIÁVEIS LINGÜÍSTICAS ESTATISTICAMENTE SIGNIFICATIVAS**

Foram selecionadas pelo Programa VARBRUL como estatisticamente relevantes ao fenômeno em análise as variáveis: paralelismo formal das seqüências de predicativos/participios no discurso, características formais do verbo da construção, características formais do sujeito da construção, estrutura do predicativo, tonicidade dos itens singulares e distância do sujeito correferente.<sup>47</sup>

##### **Paralelismo formal das seqüências de predicativos/participios no discurso**

Tradicionalmente, o princípio da economia lingüística vem sendo evocado para explicar fenômenos que envolvem variação na concordância. Segundo esse princípio, a língua tende a eliminar ou a codificar menos o que é mais previsível e reter ou codificar mais o que é menos previsível.

<sup>47</sup> Esta ordem é a mesma em que as variáveis são selecionadas pelo programa VARBRUL.

Todavia, estudos têm mostrado que, no uso real, há uma tendência de formas gramaticais semelhantes ocorrerem juntas (Schiffrin, 1981, apud Scherre & Naro, 1993). Na literatura, conhecemos essa tendência pelo nome de **paralelismo formal**, cuja importância tem sido atestada em vários trabalhos de variação (cf., por exemplo, Poplack, 1980; Naro, 1981; Scherre, 1988, 1991; Scherre & Naro, 1992).

O **paralelismo formal** determina que "o emprego de formas antecedentes num discurso influencie o emprego de formas subsequentes, favorecendo a semelhança fonomorfo-gramatical entre os elementos na cadeia discursiva" (Mollica, 1992). Scherre (1988) chega a postular a variável paralelismo formal como o **Princípio do Processamento Paralelo**, que, segundo Poplack (1980), pode ser sintetizado assim : *marcas levam a marcas e zeros levam a zeros*.

Partindo, então, dessa hipótese, pretendemos, através da análise da variável paralelismo formal das seqüências de predicativo/particípio no discurso, buscar, em níveis supra-sentenciais, maiores evidências ao Princípio do Processamento paralelo proposto por Scherre (1988). Pretendemos, também, verificar se a influência ocorre apenas da esquerda para a direita ou se existe uma influência recíproca no interior das ocorrências em série. Neste estudo, consideramos uma série aquela em que os predicativos/particípios passivos se refiram ao mesmo sujeito ou tenham a mesma forma caso se refiram a sujeitos diferentes, e que a construção analisada não esteja separada da construção anterior por mais de dez cláusulas e nem pelo discurso do interlocutor. (cf. Scherre, 1991; Scherre e Naro, 1993)



A variável paralelismo formal das seqüências no discurso foi subdividida em cinco categorias ou fatores, conforme verificamos a seguir :

- 1) predicativo/particípio em construção isolada;
- 2) primeiro de uma série;
- 3) predicativo/particípio precedido de predicativo/particípio com todas as marcas de plural ("Eles foram absowvidos (...) os coitados foram *levados*" – CHPSCFGB);
- 4) predicativo/particípio precedido de predicativo/particípio sem marca(s) de plural ("eles são rebewde mesmo (...) ficam *revowtado*" – IRTPRFGA);
- 5) casos mistos<sup>48</sup> ("Meus irmãos eru bem cawmo (...) *istudiosos* (...) filhos exemplares" – FLPSCFPA).

Na tabela 1 abaixo, apresentamos os resultados da variável paralelismo formal das seqüências no discurso.

**TABELA 1**  
**Paralelismo formal das seqüências no discurso**

FATORES	Apl/Total	%	P.R.
1) Pred./part. em const. isolada.	209/510	41	.50
2) Primeiro de uma série	52/117	44	.49
3) Pred./part. precedido de pred./part. com todas as marcas de plural	50/61	82	.85
4) Pred./part. precedido de pred./part. sem marca(s) de plural	9/72	13	.15
5) Casos mistos	7/8	88	.90
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

<sup>48</sup> Consideramos casos mistos as ocorrências em série cujo predicativo/particípio é precedido por formas marcadas e seguido por formas não marcadas, ou vice-versa.

Os resultados mostrados na tabela 1 vêm confirmar a hipótese de que marcas levam a marcas e zeros levam a zeros. Similarmente aos resultados encontrados por Scherre (1991), as formas precedidas de formas marcadas são mais marcadas (.85 de peso relativo – .70 em Scherre) se comparadas com as formas precedidas de formas não marcadas (.15 – .13 em Scherre).

As formas isoladas e os casos que são os primeiros de uma série, por sua vez, apresentam-se neutros, (.50 e .49, respectivamente) não aparecendo, portanto, como fatores com influência na presença da marca de plural no predicativo/particípio passivo.

No primeiro caso, ou seja, o das formas isoladas, temos um resultado que se aproxima bastante do encontrado por Scherre (.54); já com relação aos casos que são primeiros de uma série, encontramos pesos relativos distintos. Enquanto nossos resultados apontam esse fator como neutro (.49), os de Scherre, por outro lado, apontam-no como relevante (.66).

O peso relativo de .66 em Scherre (1991), pode estar indicando a ocorrência de uma influência recíproca entre as diversas ocorrências em série (não só da esquerda para a direita), ao passo que neste estudo tal influência ainda não é muito clara, com o fator em questão não se mostrando relevante. Esta hipótese, entretanto, ainda é prematura, podendo estar em jogo aqui o condicionamento de outros fatores.

Os casos mistos apresentam-se como fortes condicionadores à presença de marcas de plural (.90). Nos resultados de Scherre, entretanto, esse fator, ao lado das formas isoladas, apresenta-se neutro (.55).

O resultado de .90 pode estar indicando que a influência não ocorre somente da esquerda para a direita nas construções em série mas também que pode se dar de maneira recíproca no interior da mesma. Por exemplo, nas construções em série do tipo Ø 1 1 ("foi *aglomerado* muitos escravos (...) eles foram *abolidos* (...) eles foram *libertos* – CHPSCMGA), o processamento paralelo pode estar se dando reciprocamente entre a segunda e a terceira construção da série. Contudo, essa é ainda uma hipótese. Precisaríamos de mais dados deste tipo para tirarmos alguma conclusão mais consistente.

Ao contrário do fator "primeiro de uma série", que possivelmente esteja indicando uma influência recíproca entre as ocorrências em série no estudo de Scherre (.66), conforme colocamos mais acima, os **casos mistos**, que são neutros nos resultados da autora (.55), não aparecem, se considerarmos o raciocínio do parágrafo anterior, como condicionadores a uma influência recíproca. Mas, como afirma Scherre em seu artigo (e que se aplica a este estudo), tanto os casos que são os **primeiros de uma série** quanto os **casos mistos** precisam ser retomados.

### **Características formais do verbo da construção**

Com a análise da variável *características formais do verbo da construção*, visamos buscar mais evidências ao princípio do processamento paralelo, só que agora em níveis sentenciais.

A variável em análise se divide em três fatores listados a seguir com seus respectivos exemplos.

- 1) Zero verbal ("Meus irmãos eru bem *cawmo*, *Ø istudiosos*, *Ø filhos exemplares*" – FLPSCFPA);
- 2) Verbo com marca de plural ("eles são muito *pobres*" – IRTPRMCB);
- 3) Verbo sem marca de plural ("os vizinho da gente é *melhó* que os parente" – CHPSCFGA);

Os resultados da análise dessa variável se encontram abaixo, na tabela 2.

**TABELA 2**  
Características formais do verbo da construção

FATORES	Apl/total	%	P.R.
1) Zero verbal	16/27	59	.59
2) Verbo com marca de plural	308/694	44	.53
3) Verbo sem marca de plural	3/47	6	.10
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Nossos resultados indicam que também a níveis setenciais o paralelismo formal está funcionando. As construções com verbo marcado apresentam mais marcação formal no predicativo/participio (.53 – em Scherre, .61); diferente do que ocorre com as construções com verbos sem marca formal de plural que têm seus predicativos/participios pouco marcados (.10 – .18 para Scherre).

Entretanto, para responder a hipótese de que marcas levam a marcas e zeros levam a zeros, o que fazer com os casos de sujeito zero que apresentam, nos dados analisados, peso relativo de .59 (Scherre = .74) o que é mais alto que o dos casos de verbos com marcas de plural?

Seria de se esperar, pelo Princípio do Processamento Paralelo, que, por se tratar de zero verbal, sua influência sobre a marcação formal no predicativo/participio fosse baixa, próxima dos casos de verbo sem marca de plural. O que tudo indica, neste caso, é que o *zero* verbal tem comportamento distinto do *zero* ausência de marca formal no verbo e poderia estar em jogo, aqui, um outro princípio, o da *recuperação da informação*. Segundo esse princípio, a total ausência de informação do zero verbal seria recuperada no predicativo/participio e, por isso, um peso relativo de presença de marca de plural tão alto nesse caso.

Seguindo a visão de Du Bois (1984), poderíamos trabalhar com a possibilidade de o **Processamento Paralelo** e a **Recuperação da Informação** estarem em competição, o que não foi atestado ainda.

### **Características formais do sujeito da construção**

A exemplo da variável características formais do verbo da construção, buscamos, também, com a análise dessa variável, sustentar a hipótese de que marcas levam a marcas e zeros levam a zeros a nível setencial. A variável constitui-se dos fatores descritos e exemplificados abaixo.

1) Sujeito zero ("elas são *capitã* ... Ø São *dura*..." – FLPSMCA);

2) Sujeito explícito :

→ com todos os elementos nominais flexionáveis marcados ("os avós por parte de pai são *mortos*" – CHPSMCA);

→ com a última marca neutralizada ("meus pais São *descendentes de italiano*" – CHPSCMGA);

3) Sujeito explícito :

→ sem a(s) última(s) marca(s) formal(is) de plural ("As festa eram *boa*" – CHPSCMPB);

→ coordenado singular ("O Mauricio e o Marcelo são *gêmeos*" – CHPSCMCB);

4) Sujeito explícito com marca semântica de plural (nós e numerais isolados) ("nós fomos toda vida assim *criado*" – flpscfa) ("dois foram *criado* ali" – FLPSCFPA);

5) Sujeito explícito com marca formal de plural totalmente neutralizada ("vocêS São *idiota*" – IRTPRFGA).

Apontados os fatores que fazem parte dessa variável, apresentamos a tabela 3 com os resultados obtidos.

**TABELA 3**  
Características formais do sujeito da construção

FATORES	Apl/total	%	P.R.
1) Sujeito zero	109/246	44	.49
2) Sujeito explícito com todas as marcas formais de plural ou com a última marca neutraliz.	145/287	51	.59
3) Sujeito explícito sem a(s) última(s) marca(s) formal(is) de plural	17/74	23	.35
4) Sujeito explícito com marca semântica de plural (nós e numerais)	33/84	39	.50
5) Sujeito explícito com a marca de plural totalmente neutralizada	23/77	30	.34
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Os resultados mostrados na tabela 3 acima vêm comprovar novamente que **marcas levam a marcas e zeros levam a zeros**: os casos de sujeito predominantemente com todas as marcas de plural levam a uma maior marcação do predicativo/participio (.59 para .63 em Scherre), enquanto que os casos de sujeito predominantemente sem a(s) última(s) marca(s) formal(is) de plural têm suas chances de marcar o predicativo/participio diminuídas para .35 (em Scherre esse valor é de .30).

Em Scherre (1991), os fatores **sujeito explícito com marca semântica de plural (nós e numerais)** e **sujeito explícito com marca de plural totalmente neutralizada** foram considerados de atuação neutra sobre a presença das marcas de plural nos predicativos/participios (.48 e .49, respectivamente). Nossos dados, contudo, revelaram um resultado um pouco diferente.

Com relação ao primeiro fator, ou seja, *nós e numerais*, sua atuação também mostrou-se neutra (.50), resultado muito próximo àquele alcançado por Scherre. Já nos casos de *sujeitos com marcas de plural totalmente neutralizada*, o quadro é outro: aqui, o fator mostrou-se negativo quanto ao favorecimento da marcação formal de plural nos predicativos/participios (.34 de peso relativo), aproximando-se muito dos casos de sujeito explícito sem a(s) última(s) marca(s) formal(is) de plural (cf. acima). A aproximação entre os pesos relativos desses dois fatores pode estar relacionada ao fato de que um não apresenta sua marca formal e o outro a tem totalmente neutralizada pelo fonema do item que o segue.

Os casos de sujeito zero mostram um comportamento neutro ao condicionamento de marca de plural com .49 de peso relativo. Contudo, supunha-se, pelo Princípio do Processamento Paralelo, que, nesses casos, o peso relativo fosse bastante baixo para corroborar a hipótese de que marcas levam a marcas e zeros levam a zeros. A exemplo do fator "zero verbal", Scherre (1991) levanta a hipótese de o Princípio da recuperação da informação estar em jogo para interpretar a influência do sujeito zero. Mesmo apresentando-se neutro, voltaremos a mencionar esse fator mais adiante.

### **Estrutura do predicativo**

Com a análise da variável estrutura do predicativo, objetivamos verificar qual é a relação entre a configuração sintagmática do predicativo e o condicionamento na marcação formal de plural. Nossas hipóteses eram de que a **configuração nominal** favoreceria mais a presença de marcas de plural do que a **configuração adjetiva**, e que os predicativos do tipo **de mais de um elemento com os itens *todo/tudo*** desfavoreceriam a marcação formal de plural nos predicativos/participios.

Antes de elencarmos o conjunto de fatores que compõem a variável estruturas do predicativo, vale ressaltar que consideramos marcados somente os predicativos que exibiam marcas formais explícitas em **todos os elementos flexionáveis da estrutura**.

Vejamos, então, como esta variável ficou subdividida.



- 1) Predicativo adjetivo e formas participiais de um elemento ("eles são *duro*" – IRTPRMGB; "as leis são *feita...*" – IRTPRMGB);
- 2) Predicativo nominal de um elemento ("elas são *intérpretes*" – IRTPRFCB);
- 3) Predicativo de mais de um elemento ("meus pais eram *peessoas honestas*" – IRTPRFGB);
- 4) Predicativo de mais de um elemento com os itens *todo/tudo* ("são *tudo casado*" – CHPSCMGB).

Sendo assim, apresentamos a tabela 4 com os resultados obtidos da análise da variável.

**TABELA 4**  
**Estrutura do predicativo**

FATORES	Apl./Total	%	P.R.
1) Predicat. adj. e formas particip. de 1 elem.	222/548	41	.48
2) Predicat. nom. de 1 elemento.	53/111	48	.58
3) Predicat. de mais de 1 elemento.	47/86	55	.61
4) Predicat. de mais de 1 elem. com os itens <i>tudo/todo</i> .	5/23	22	.26
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Conforme apontam os valores numéricos da tabela acima, nossas expectativas quanto à variável estrutura do predicativo foram correspondidas.

Com relação à hipótese de que a configuração do tipo predicativo de mais de um elemento com os itens *tudo/todo* desfavoreceria a marcação formal de plural nos

predicativos/participios, o resultado que apresentamos na tabela 4 é inequívoco (.26). Scherre obteve um resultado ainda mais expressivo (.07). Acreditamos que este tipo de configuração não favoreça marca por influência dos itens **tudo/todo** que são singulares.

Da mesma forma, nossa hipótese de que a configuração nominal favoreceria mais marcas de plural do que a configuração adjetiva confirmou-se. A tabela mostra que as **formas adjetivas e participiais** comportam-se como menos condicionadoras à presença de marcas de plural nos predicativos/participios passivos (.48), do que **formas nominais de um elemento** (.58). Esses resultados parecem indicar uma espécie de saliência<sup>49</sup> (Scherre, 1991), em que a configuração adjetiva por ser menos saliente (mais comum) do que as formas nominais é, por isso, menos marcada.

Nossos resultados indicam os **predicativos de mais de um elemento** como os mais marcados, com um peso relativo de .61.

Scherre, entretanto, encontrou valores numéricos relativamente distintos dos nossos.

Para o fator **formas adjetivas e participiais**, ela obteve um peso relativo de .66 e para as formas nominais de um elemento .77., com as formas nominais também favorecendo mais a presença de marca de plural do que as formas adjetivas.

<sup>49</sup> Levanta-se a hipótese de um sentido da saliência mais amplo do que a própria saliência fônica, falando-se até em um **Princípio da Saliência**.

Nos resultados da autora, o fator predicativo **nominal** de mais de um elemento (ela o denomina dessa forma uma vez que a maioria dos casos deste tipo em seu corpus são nominais, o que não é o meu caso) tem peso relativo de .68, menor que o das formas nominais de um elemento.

Voltando aos nossos resultados, no caso dos **predicativos de mais de um elemento** (que nos nossos dados são na maioria **formas adjetivas**), não era esperado, pela nossa hipótese, que tal configuração favorecesse mais a presença de marcas de plural, chegando a apresentar o peso relativo mais alto de todo o grupo (.61).

Uma explicação possível para isso poderia ser a de que o contexto antecedente estaria influenciando, isto é, provavelmente os grupos de fatores **características formais do sujeito** e **características formais do verbo** estariam interferindo no resultado. Em vista disso, rodamos o CROSSTAB (CROS3000) e constatamos que o contexto mais próximo, ou seja, o grupo **características formais do verbo** parece interferir mais diretamente. Vejamos a tabela 4a abaixo.

**TABELA 4a**  
**Tabulação cruzada entre Estrut. do Predic. e Carac. Formas do Verbo**

ESTRUT. PRED.	Verbo C/ marca	Zero verbal	Verbo S/ marca	TOTAL
<i>Pred. adj. e formas partíc. de 1 elemento</i>	1 206 42% 0 281 58% T 487	14 64% 8 36% 22	2 5% 37 95% 39	222 41% 326 59% 578
<i>Pred. de + de 1 elem. c/ tudo/todo</i>	1 5 24% 0 16 76% T 21	0 0% 1 100% 1	0 0% 1 100% 1	5 22% 18 78% 23
<i>Pred. nom. de 1 elemento</i>	1 53 50% 0 53 50% T 106	0 0% 1 100% 1	0 0% 4 100% 4	53 48% 58 52% 111
<i>Pred. de + 1 elem.</i>	1 44 55% 0 36 45% T 80	2 67% 1 33% 3	1 33% 2 67% 3	47 55% 39 45% 86
<b>TOTAL</b>	1 308 44% 0 386 56% T 694	16 59% 11 41% 27	3 6% 44 94% 47	327 43% 441 57% 768

Se verificarmos os resultados grifados na tabela, veremos que nos casos em que o predicativo é do tipo **mais de um elemento**, a grande maioria é precedido por **verbos com todas as marcas (80 – 44 marcados [55%])**, enquanto que há apenas 3 ocorrências de **verbos sem marca formal de plural** no contexto precedente a este tipo de predicativo. Isto, possivelmente, venha explicar o porquê de um peso relativo alto de presença de marca na **configuração predicativo de mais de um elemento**.

### **Tonicidade dos itens singulares**

Clássica na literatura (cf. Lemle & Naro, 1977; Naro, 1981 e Scherre, 1988), a variável **tonicidade dos itens singulares** tem sido largamente considerada na testagem do **Princípio da Saliência Fônica**.

Este Princípio foi introduzido por Lemle e Naro em estudos realizados no período de 1974 a 1976 sobre o Português do Brasil. Depois deles, outros lingüistas também estudaram a saliência fônica: em 1976, Braga e Scherre verificam a influência desse princípio sobre a **concordância de número entre os elementos do sintagma nominal**; outros estudos, também nesta linha, são realizados por Braga (1977), Scherre (1978); Ponte (1979); Carvalho Nina (1980) e Scherre (1988); e, ainda, Naro (1981), com estudos sobre a concordância verbal, e Guy (1981) (V. Scherre, 1988).

O Princípio da Saliência Fônica estabelece que "as formas **mais salientes**, e por isso **mais perceptíveis**, são **mais prováveis de serem marcadas** do que as menos salientes" (Scherre, 1992)<sup>50</sup>. E é esta hipótese que pretendemos verificar em nosso estudo sobre a concordância de número no predicativo e particípio passivo, ou seja, a hipótese de que quanto maior a saliência fônica dos itens que compõem o predicativo/particípio, maior o favorecimento à presença de marcas formais de plural.

Estabelecemos três fatores para a variável tonicidade.

<sup>50</sup> O negrito é nosso.

- 1) Itens oxítonos;
- 2) Itens paroxítonos;
- 3) Itens proparoxítonos.

Os resultados do comportamento desse grupo de fatores podem ser verificados a seguir, na tabela 5.

**TABELA 5**  
**Tonicidade dos itens singulares**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>P.R.</b>
1) Oxítonos	57/105	54	.64
2) Paroxítonos	249/625	40	.47
3) Proparoxítonos	21/38	55	.57
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Nossa expectativa com relação à tonicidade era de que os itens oxítonos, por apresentarem a sílaba mais forte na posição em que se posiciona o morfema de pluralidade, ou seja, por serem mais salientes, seriam mais marcados do que os itens paroxítonos e proparoxítonos. A tabela 5 revela resultados incontestáveis quanto a isso, mostrando-nos o maior peso relativo de presença de marca formal de plural nos predicativos/participios com itens oxítonos (.64), enquanto os outros dois, paroxítonos e proparoxítonos têm seus pesos relativos de .47 (neutro) e .57, respectivamente.

Scherre (1991) encontrou resultados similares a estes, embora essa variável, em sua análise, não tenha sido selecionada como estatisticamente significativa. Segundo seus resultados, os predicativos/participios oxítonos também são mais marcados (.64). Os itens paroxítonos tiveram peso relativo de .46 e os proparoxítonos (e aqui a diferença entre os resultados), peso relativo de .40.

É de se esperar um resultado como esse de Scherre em que quanto mais distante do local em que se marcará a pluralidade está a sílaba mais forte (quanto menos saliente é o item), menos marcas de plural ocorrerão. Segundo nossas expectativas, portanto, o resultado de .57 para os itens proparoxítonos seria considerado atípico.

Uma interpretação possível para este resultado, conforme Scherre (1988), poderia ser a de que "os proparoxítonos e os oxítonos, enquanto um todo, têm a característica comum de serem mais salientes do que os paroxítonos, pois eles são itens marcados na língua portuguesa, que é constituída predominantemente de palavras paroxítonas".

### **Distância do sujeito correferente**

O grupo de fatores **distância do sujeito correferente** não faz parte das variáveis lingüísticas analisadas por Scherre (1991). Esta variável está sendo lançada nesta investigação em função do resultado referente ao fator **sujeito zero** do grupo de fatores **características formais do sujeito**.

Vimos que, mesmo apresentando um peso relativo neutro (de .49), e não estando muito próximo dos casos de sujeito explícito com todas as marcas de plural (.59), o sujeito zero comportou-se de modo não esperado, se pensarmos no Princípio do Processamento Paralelo que, como verificamos até agora, tem se mostrado consistente como hipótese.

Esperávamos que os casos de sujeito zero se aproximassem muito daqueles em que os sujeitos não apresentam as últimas marcas formais de plural, os quais mostraram um peso relativo condizente ao Princípio do Processamento Paralelo (.35).

Poderíamos explicar esse resultado através do Princípio da Recuperação da Informação, mas decidimos ir por um outro caminho: já que estamos diante de zeros sujeitos, podemos recuar no discurso à procura de seus correferentes e verificar a influência que esses sujeitos correferentes ao sujeito zero da construção em análise estão tendo sobre os predicativos/participios.

Segundo Givón (1983), há três maneiras de se medir a continuidade tópica<sup>51</sup> no discurso: a) *distância referencial*; b) *interferência potencial*; e c) *persistência*.

Interessa-nos a primeira, *distância referencial*, (DR) que, nas palavras do autor, é aquela

que nos permite avaliar, calcular a distância entre a ocorrência prévia de um referente/tópico no discurso e sua atual ocorrência na cláusula em análise (...). A distância é, portanto, expressa em termos de *número de cláusulas para a esquerda*. O valor mínimo que pode ser atribuído é, então, de *1 cláusula* (...) e fica estabelecido um número arbitrário de *20 cláusulas para a esquerda*. Quando a referência de um tópico excede essa escala, o valor 20 é automaticamente atribuído. (Givón, 1983:13) (A tradução é minha)

A partir disso, buscamos verificar a hipótese de que quanto maior for a DR, maiores serão as chances de os predicativos/participios serem marcados ou vice-versa.

<sup>51</sup> Tópico para Givón é o tema, a informação velha. (Givón, 1983)



Para a análise, dividimos a variável distância do sujeito correferente em 4 grupos listados abaixo.

- 1) 1 a 5 cláusulas para a esquerda;
- 2) 6 a 10 cláusulas para a esquerda;
- 3) 11 a 15 cláusulas para a esquerda;
- 4) 16 a 20 cláusulas para a esquerda.

Numa primeira análise, estamos considerando, como Givón, um número máximo de 20 cláusulas para a esquerda. Pretendemos, também, verificar esse grupo de fatores com um número máximo de 10 cláusula para trás, diminuindo, assim, o número de cláusulas em cada subdivisão. Decidimos trabalhar com esse valor pelo fato de termos poucas ocorrências com sujeitos correferentes localizados a mais de 10 cláusulas para a esquerda.

É importante colocarmos que realizamos primeiro uma rodada considerando os dados em que a correferência está na fala do entrevistador e, também, com dados em que a fala do entrevistador aparece intercalada entre a correferência do sujeito zero e o sujeito zero. Num outro momento, verificamos o comportamento da variável sem esses dados mencionados.

Na tabela 6a, mostramos os resultados da análise dessa variável .

**TABELA 6a**  
**Distância do sujeito correferente**

FATORES	Apl/Total	%	P.R.
1) De 1 a 5 cláus. para a esquerda	79/196	40	.45
2) De 6 a 10 cláus. para a esquerda	21/36	58	.66
3) De 11 a 15 cláus. para a esquerda	5/6	83	.90
4) 16 ou mais cláus. para a esquerda	4/7	57	.74
<b>TOTAL</b>	<b>109/245</b>	<b>44</b>	

Os resultados obtidos correspondem às nossas expectativas. O sujeito correferente ao sujeito zero, quando próximo a ele leva o predicativo/participio a ser menos marcado (.45 – peso relativo neutro). Porém, à medida em que as distâncias referenciais aumentam, aumentam também as chances de o predicativo/participio aparecer com marcas formais de plural (.66 para o fator distância média; .90 e .74 para as distâncias maiores).

Na tabela, verificamos uma inversão de pesos relativos entre o terceiro e quarto níveis. Esperava-se que este último nível fosse aparecer com o peso relativo maior, numa progressão. Contudo, isto não ocorreu, talvez pelo fato de termos poucos dados nestes dois últimos níveis da escala, o que é mais um motivo para reajustarmos esse grupo de fatores.

Tudo indica que nossos resultados confirmam o *subprincípio da quantidade* (princípio da iconicidade) da teoria Funcionalista de Givón que diz que quanto maior, menos previzível e mais importante for a informação, mais código será necessário (Givón, 1991: 87-9).

Sendo assim, podemos considerar, analogicamente, que quanto maior for a quantidade de informação entre o sujeito zero e o seu correferente (quanto maior for a DR), mais material morfossintático será necessário, no nosso caso, mais marcas de plural no predicativo/participio passivo. Por outro lado, se a quantidade de informação intercalada entre a anáfora zero e sua correferência é menor, ou seja, quanto mais próximo estiver o sujeito, menor será também a necessidade de se marcar o predicativo/participio.

Conforme dissemos acima, realizamos também uma rodada dessa variável sem os dados em que a fala do entrevistador é considerada. Nesta rodada, os pesos relativos pouco diferem dos da rodada anterior (.43 para 1 a 5 cláusulas para a esquerda; .67 para 6 a 10 cláusulas; .90 para 11 a 15 cláusulas; e .76 para 16 a 20 cláusulas), conforme verificamos na tabela 6b a seguir.

**TABELA 6b**  
Distância do sujeito correferente (com pesos relativos da rodada sem a fala do E)

FATORES	Apl./Total	%	P.R.	P.R.
1) De 1 a 5 cláus. para a esquerda	79/196	40	.45	.43
2) De 6 a 10 cláus. para a esquerda	21/36	58	.66	.67
3) De 11 a 15 cláus. p/ a esq.	5/6	83	.90	.90
4) 16 ou mais cláus. p/ a esq.	4/7	57	.74	.76

Dois pontos, entretanto, chamam a atenção. Nesta segunda rodada, a variável **distância do sujeito correferente** passa a ser a quinta variável selecionada como mais significativa, antes da variável **tonicidade dos itens singulares**.

Além disso, houve uma mudança significativa no comportamento do fator **sujeito zero**: o peso relativo de presença da marca formal de plural aumentou em **.10** em relação à primeira análise (de **.49** para **.59** favorecendo a concordância).

Realizando a rodada com a variável **distância do sujeito correferente** modificada (numa contagem de até dez cláusulas para a esquerda<sup>52</sup>, como colocamos acima), obtivemos os resultados que seguem.

**TABELA 6c**  
**Distância do sujeito correferente**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>P.R.</b>
1) De 1 a 2 cláus. para a esquerda	52/131	40	.45
2) De 3 a 4 cláus. para a esquerda	17/46	37	.44
3) De 5 a 6 cláus. para a esquerda	10/19	53	.66
4) De 7 a 8 cláus. para a esquerda	6/14	43	.44
5) 9 ou mais cláus. para a esquerda	12/18	67	.78
<b>TOTAL</b>	<b>97/228</b>	<b>43</b>	

A retestagem da variável nos mostra um comportamento que vem ao encontro das nossas expectativas, ou seja: à medida que a distância aumenta, maior é a incidência de concordância, salvo pelo fator **distância de 7 a 8 cláusulas para a esquerda**, em que se esperava um peso relativo mais alto. Neste caso, é imprudente tirar quaisquer conclusões uma vez que o número de ocorrências ainda é pequeno (apenas 14).

<sup>52</sup> A distância máxima considerada é de dez cláusulas. A correferência situada numa distância além dessa foi considerada como se estivesse a dez cláusulas do sujeito zero.

## 6.2. – VARIÁVEIS LINGÜÍSTICAS ESTATISTICAMENTE NÃO SIGNIFICATIVAS

A seguir, apresentaremos as variáveis cujos resultados são considerados estatisticamente não significativos. São elas: ordem dos elementos na estrutura, material interveniente entre o verbo e o predicativo/particípio, processos morfofonológicos de formação de plural e tipo de estrutura.<sup>53</sup> Apesar de não terem sido selecionadas como relevantes ao fenómeno em estudo, decidimos mostrar os resultados dessas variáveis pela seguinte razão: as variáveis em questão, tais como as variáveis selecionadas como estatisticamente relevantes, estão presentes nos dados e também estão relacionadas a hipóteses a serem confirmadas ou não.

### **Ordem dos elementos na estrutura**

A variável ordem dos elementos na estrutura está sendo considerada com o objetivo de se verificar se ordem canônica tende a favorecer mais marcas de plural do que ordem não canônica. Nossa hipótese é de que sim, a ordem canônica, por apresentar um fluxo normal de processamento, sem interrupção, favorecerá mais marcas de plural no predicativo/particípio do que a ordem não canônica dos elementos.

Dividimos a variável, então, nos seguintes fatores.

<sup>53</sup> A apresentação das variáveis será feita na mesma ordem de retirada da variável na rodada.

- 1) Ordem canônica : sujeito + verbo + predicativo ("elas já são mais *metidas*" – FLPSCFPB);
- 2) Ordem não canônica ("é muito *maroto* aqueles ensaio" – CHPSCMCA).

A seguir temos a tabela 7 com os resultados desse grupo de fatores.

**TABELA 7**  
**Ordem dos elementos na estrutura**

FATORES	Apl/Total	%	P.R.
1) Ordem canônica	313/732	43	.50
2) Ordem não canônica	14/36	39	.56
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

As percentagens indicam uma tendência um pouco maior de a ordem canônica (que não interrompe o fluxo do processamento linear) ocorrer na presença de marcas formais de plural (43%), quando em comparação com a ordem não canônica (39%). Entretanto, os pesos relativos nos mostram um outro quadro: a **ordem canônica** aparece como neutra ao condicionamento de marcas de plural (.50), enquanto que a **ordem não canônica** apresenta-se condicionadora à presença de marcas (.56). Este resultado, contudo, não é conclusivo por estarmos diante de distribuição de dados bastante assimétrica (732 ocorrências de estruturas canônicas contra apenas 36 de não canônicas).

### Material interveniente entre verbo e predicativo/particípio

No que se refere à variável em questão, buscamos verificar até que ponto a ocorrência de elementos entre o verbo e o predicativo/particípio influencia no sentido da ocorrência de marcas formais de plural.

A variável consta dos seguintes fatores exemplificados abaixo.

- 1) Ausência de material interveniente ("dois foru Ø *criado* ali" –FLPSCFPA);
- 2) Presença de intensificadores ("as passage tão muito *cara*" – FLPSCFPA);
- 3) Demais materiais intervenientes ("são até *loucos*" – IRTPRMCA).

A tabela 8 indica o comportamento da variável.

**TABELA 8**  
Material interveniente entre o verbo e o predicativo/particípio

FATORES	Apl/Total	%	P.R.
1) Sem material interv.	253/584	43	.50
2) Presença de intensific.	58/143	41	.53
3) Demais elementos intervenientes	16/41	39	.44
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Nossos resultados apontam neutralidade da variável material interveniente, no que se refere à influência na presença/ausência de marcas de plural, em dois fatores: fator ausência de material interveniente (.50 - .48 em Scherre) e fator presença de intensificadores (.53 - .54 para Scherre).

No caso do fator **demais elementos intervenientes**, em Scherre, o mesmo apresenta peso relativo neutro (.48), diferente do nosso resultado que apontou o fator como não condicionador à presença de marca de plural (.44).

Este grupo de fatores poderá ser analisado ainda quanto ao número de sílabas dos elementos.

### **Processos morfofonológicos de formação do plural**

Como a variável **tonicidade dos itens singulares**, a variável **processos morfofonológicos de formação do plural** também é considerada clássica na literatura variacionista sobre estudos do português falado no Brasil.

Da mesma forma que a variável **tonicidade**, esta variável também está relacionada ao Princípio da Saliência Fônica, que determina que quanto mais saliente é o item analisado mais chances ele tem de ser marcado.

A princípio, a variável **processos** foi subdividida em três categorias elencadas e exemplificadas abaixo.

- 1) Plural regular ("nós somo **obrigado** a votá" – FLPSMCA);
- 2) Plural não regular ("as duas são **iguais**" – FLPSFCB);
- 3) Plural misto (diminutivos e aumentativos) ("os piquinho ficam **sentadinho**" – IRTPFGA).



Nossa hipótese com relação ao comportamento dessa variável é a seguinte: das três categorias mencionadas, a dos plurais não regulares, por ser mais saliente, será mais marcada. Vejamos então os resultados apresentados na tabela 9a.

**TABELA 9a**  
**Processos morfofonológicos de formação do plural**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>Prob.</b>
1) Plural regular	273/660	41	.49
2) Plural não regular	51/94	54	.58
3) Plural misto (diminut. e aumentativos)	3/14	21	.30
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Nossa hipótese de que os não regulares, por serem mais salientes, seriam mais marcados, confirmou-se (.58 de peso relativo). O fator "plural regular" apresentou um comportamento neutro à presença de plural (.49). Em Scherre (1991), a diferença entre um fator e outro é ainda mais polarizada, com .67 para os não regulares e .40 para os regulares.

Os casos de plural misto (diminutivos e aumentativos) mostraram um resultado também esperado. Esses dados englobam plural regular e não regular, além de envolverem os graus aumentativo e diminutivo, e, segundo Scherre (1988), comportam-se como elementos **informais**, o que explicaria o peso relativo de apenas .30 de influência sobre a presença de marcas de plural.

Poderíamos analisar essa questão da informalidade dos itens lexicais aumentativos e diminutivos em uma variável separada para verificar como a mesma se comporta. Contudo, isso se torna inviável, por enquanto, em vista da pequena quantidade de dados disponíveis (apenas 14 ocorrências).

Numa segunda análise da variável **processos morfofonológico de formação do plural**, subdividimos a mesma em mais fatores:

- 1) Plural regular;
- 2) Plural duplo (itens que apresentam duas possibilidades de marcas: religioso-religiosos) ("meus pais eram bem *religioso*" – FLPSCMGA);
- 3) Plural de itens terminados em -l ("as pessoas são assim tão *agradáveis*" – IRTPRFGB);
- 4) Plural de itens terminados em -ão ("aqueles são *alemão* puro" – IRTPRMCA);
- 5) Plural de itens terminados em -R ("eles eru *aprontadores*" – CHPSCMCA);
- 6) Plural de itens terminados em -S ("são *holandeses*" – CHPSCMGB);
- 7) Plural misto.

De acordo com o "Princípio da Saliência Fônica", nossas expectativas são de que os itens lexicais que apresentam mais diferença fônica entre as suas respectivas formas singular e plural (mais salientes) serão mais marcados.

Na análise da variável processos assim subdividida, encontramos os seguintes resultados, como mostra a tabela 9b abaixo.

**TABELA 9b**  
**Processos morfofonológicos de formação do plural**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>P.R.</b>
1) Plural regular	273/660	41	.49
2) Plural duplo	6/14	43	.47
3) Plural de itens terminados em -l	19/31	61	.72
4) Plural de itens terminados em -ão	2/5	40	.31
5) Plural de itens terminados em -R	21/38	55	.55
6) Plural de itens terminados em -S	3/6	50	.56
7) Plural misto	3/14	21	.31
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Nesta análise mais detalhada da variável processos, o plural regular e o plural misto praticamente mantiveram seus pesos relativos inalterados em relação à análise menos detalhada (tabela 9a): .49 em ambas as rodadas no caso dos regulares e .31 para .30 para o plural misto.

O comportamento do plural não regular, agora subdividido em categorias, mostrou que nossas expectativas foram correspondidas em parte. Pelo Princípio da Saliência Fônica, esperava-se que todas as categorias não regulares (plural duplo, itens terminados em -l, -ão, -R e -S, com maior diferença material fônica na relação singular/plural) fossem mais marcadas do que o plural regular. Os itens terminados em -l, -S e -R corresponderam a nossas expectativas, favorecendo mais marcas do que os itens regulares (.72, .56 e .55, respectivamente).

Entretanto, o plural duplo e os itens terminados em *-ão* fogem do previsto, comportando-se de forma bastante diferente do restante do grupo dos não regulares, na medida em que são menos marcados do que os itens regulares (os menos salientes na escala): .47 e .31, respectivamente.

Poderíamos supor que o baixo peso relativo de os itens terminados em *-ão* aparecerem marcados se deve ao fato de que na língua portuguesa estes itens são pluralizados de três formas diferentes: em *-ãos*, *ões* e *ães*, o que poderia causar dúvidas ao falante. Nos nossos dados, por exemplo, há uma ocorrência do tipo: "são *alemões*" (CHPSCMGB) e a respeito do mesmo fica um questionamento: Até que ponto esse enunciado não criaria mais estranhamento ao ouvinte do que a forma não marcada *são alemão*?

De qualquer modo, é precipitado tirar alguma conclusão mais consistente, uma vez que ainda temos poucos dados para cada categoria dos não regulares, principalmente os itens terminados em *-ão* e em *-S* e o plural duplo.

### Tipo de estrutura

Pretendemos, com a análise dessa variável, verificar a influência da estrutura passiva sobre a marcação de plural no predicativo/particípio. Nossa hipótese é de que a estrutura passiva (menos saliente) desfavorecerá a presença de marcas formais de plural.

A variável está subdividida em dois fatores exemplificados abaixo.

- 1) Estrutura ativa ("eles eru muito *bons* pra mim" – FLPSCMPA);
- 2) Estrutura passiva ("os preso foram *mandado* pra Ponta Grossa" – IRTPRFGA)

Os resultados acerca do comportamento da variável estão na tabela 10.

**TABELA 10**  
**Tipo de estrutura**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>P.R.</b>
1) Estrutura ativa	258/625	41	.48
2) Estrutura passiva	69/143	48	.58
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Conforme a tabela acima, os resultados não evidenciam nossa hipótese. A estrutura ativa apresenta influência neutra com peso relativo de .48 e a estrutura passiva, contrariando nossas expectativas, favorece mais a presença de marcas formais de plural em .58.

Diferente dos resultados de Scherre que apontaram a estrutura ativa como favorecedora à presença de marca (.56 contra .44 para a estrutura passiva), nossos valores indicam que a estrutura passiva está condicionando mais marcas de plural do que a estrutura ativa. A expectativa era outra em função de todos os casos de passiva serem regulares (menos salientes).

Acreditamos que outro(s) fator(es) esteja(m) interferindo nesse resultado. Um deles, intuitivamente, parece ser **escolaridade**. Para nos certificarmos disso, rodamos o CROSSTAB (CROS3000), chegando ao seguinte resultado, conforme mostra a tabela 10a.

**TABELA 10a**  
**Tabulação cruzada entre tipo de estrutura e escolaridade**

TIPO ESTRUT.	DE	Primário	Ginásial	Colegial	TOTAL
<b>Ativa</b>	1	35 20%	78 37%	145 62%	258 41%
	0	143 80%	135 63%	89 38%	367 59%
	T	178	213	234	625
<b>Passiva</b>	1	10 26%	12 30%	47 73%	69 48%
	0	29 74%	28 70%	17 27%	74 52%
	T	39	40	64	143
<b>TOTAL</b>	1	45 21%	90 36%	192 64%	327 43%
	0	172 79%	163 64%	106 36%	441 57%
	T	217	253	298	768

Pelos resultados, verificamos que a estrutura passiva é mais marcada nos informantes **primários** (26% contra 20% de presença de marca na estrutura ativa) e nos informantes **colegiais** (73% contra 62% de presença de marca na estrutura ativa – e aqui a diferença é um pouco maior [11%], se comparado com os 6% no caso dos informantes primários).

Por outro lado, nos informantes **ginásiais**, a estrutura passiva aparece menos marcada que a estrutura ativa (30% contra 37%). Esta distribuição "dois para um" parece esclarecer o resultado inesperado de passivas aparecerem com mais marcas que estruturas ativas.

## 7- ANÁLISE DAS VARIÁVEIS SOCIAIS

No capítulo 2, tratamos, entre outras coisas, da variação e mudança nos sistemas lingüísticos, máxima da Teoria Sociolingüística Variacionista. Vimos, por exemplo, que nem sempre um quadro lingüístico variável pode levar a mudanças lingüísticas. Contudo, em determinados casos, em que as relações sócio-interacionais levam duas ou mais variedades a se confrontarem dialeticamente, levando uma forma a derrotar a(s) outra(s), pode-se ter configurado, aí, um processo de mudança na língua.

Vimos também que fatores lingüísticos e sociais concorrem para o emprego de formas variantes e encontram-se intimamente relacionados no desenvolvimento da mudança lingüística. Entre os fatores lingüísticos, encontram-se arrolados os fatores fono-morfo-sintáticos, os semânticos, os discursivos e os lexicais. No segundo grupo, encontram-se os fatores que são inerentes ao indivíduo (**sexo, idade, etnia**), os sócio-geográficos (**região, escolarização, nível de renda, profissão, classe social**), e os contextuais (**grau de formalidade e tensão discursiva**).

No capítulo anterior, mostramos e analisamos as variáveis lingüísticas do nosso estudo. Abordaremos, agora, as variáveis sociais estratificadas da nossa amostra, com o objetivo de verificar indícios de variação estável ou de mudança em progresso envolvendo o fenômeno em estudo.

Começaremos mostrando a relação entre **variação estável** e **mudança em progresso** a partir de estudos variacionistas para, em seguida, apresentarmos os fatores extralingüísticos considerados em nossa análise, juntamente com seus resultados probabilísticos.

### **7.1. - O COMPORTAMENTO DE VARIÁVEIS SOCIAIS NA VARIAÇÃO LINGÜÍSTICA ESTÁVEL**

Normalmente, caracteriza-se como uma situação de **variação estável** aquela em que se detecta uma relação diretamente proporcional entre classe social ou escolarização e uso de formas de prestígio.

Mas não basta somente essa relação para se estar diante de um caso de **variação estável**. O fator sexo também estaria envolvido na relação entre classe social/escolarização e uso de formas de prestígio. Neste sentido, os falantes do sexo feminino (mesmo quando crianças) falariam da forma mais próxima da norma estabelecida socialmente, principalmente em contextos mais formais de fala.

Com relação à variável idade, nas situações de **variação estável**, ela apresenta ou uma *distribuição plana* sem gradação etária ou uma *distribuição curvilinear*, indicando gradação etária, com o ápice das formas de prestígio sendo encontrado nas faixas etárias intermediárias.



## 7.2. - O COMPORTAMENTO DE VARIÁVEIS SOCIAIS NA MUDANÇA EM PROGRESSO

A mudança em progresso só pode ser detectada em estudos sincrônicos pela constatação da diferença etária (*tempo aparente*), o que por si só não seria suficiente, uma vez que, como foi visto acima, diferenças etárias podem estar indicando apenas gradação etária (padrão curvilíneo). Para se caracterizar, então, uma mudança em progresso, o fator idade deve seguir uma *distribuição inclinada*, com os jovens e formas inovadoras numa extremidade e os mais velhos e as formas conservadoras na outra.

O primeiro estudo apontado pela literatura a considerar o fator idade como condicionador da mudança lingüística é o de Gauchat (1905) (Naro, In Mollica, 1992:83). Analisando a fala de uma aldeia suíça, o pesquisador

constatou que nas palavras em que as pessoas mais velhas usavam a lateral palatal [λ] (lh em português), os mais jovens preferiam [y] enquanto os de meia idade usavam ambos os sons. Esta escala em tempo aparente levou o pesquisador a concluir que o som [λ] estava em processo de extinção no dialeto e que futuramente seu substituto seria [y]. Vinte e cinco anos mais tarde, esta previsão foi confirmada por outro pesquisador (Hermann, 1929) que visitou a mesma aldeia e notou que [λ] já tinha praticamente desaparecido da fala local.

Labov (1972b), em seu estudo de Martha's Vineyard sobre o fenômeno de centralização do ditongo /aw/, revelou uma situação parecida com a da aldeia suíça, "com os informantes mais velhos preservando mais a forma original não-centralizada e os mais jovens utilizando cada vez mais centralização" (Naro, In Mollica, 1992).

No português do Brasil, Tarallo (1986) estudou as orações relativas considerando três variantes: uma padrão – **preposição mais pronome relativo** ("Este é o homem *com quem eu falei ontem.*"), e duas não-padrão – a **relativa com pronome-lembrante** ("Este é o homem *que eu falei com ele ontem.*") e a **cortadora** ("Este é o homem *que eu falei ontem.*"). Seus resultados indicam que a primeira forma (a variante **padrão**) já está praticamente desaparecendo da língua falada, sendo substituída, por sua vez, pela **relativa cortadora**, e que a **relativa com pronome-lembrante** não goza de prestígio na comunidade.

Além da faixa etária, o fator sexo também concorre para a caracterização da mudança em progresso. Verifica-se, por exemplo, que nas situações de mudança em direção à forma de prestígio geralmente são as mulheres (e não os homens) que lideram. Ora, considerando que a literatura variacionista em geral considera as mulheres mais sensíveis à forma padrão do que os homens, em se tratando de variação estável, qualquer distanciamento da norma pelo sexo feminino poderia ser interpretado como mudança lingüística.

Mas, o que se observa é que a questão não é tão simples assim, na medida em que os estudos evidenciam as mais diversas conclusões acerca da influência desse fator sobre a mudança em progresso (Scherre, 1988). Todavia, um fato parece ser certo: variação estável e mudança em progresso à parte, tudo indica que há uma tendência de as mulheres se aproximarem mais da forma padrão do que os homens.

Vejamos o que nos coloca Chambers (1995) acerca dessa questão, ao citar outros estudiosos do assunto:

(...) as mulheres, mais do que os homens, demonstram uma sensibilidade maior a traços lingüísticos socialmente avaliados. (Wolfram, 1969)  
 Na linguagem formal, as mulheres utilizam menos formas estigmatizadas que os homens, e são mais sensíveis que eles à forma de prestígio. (Labov, 1972)  
 As mulheres demonstram mais consciência das normas de prestígio tanto em suas falas quanto em suas atitudes com relação à fala. (Wolfram & Fasold, 1974)  
 As mulheres (...) são claramente mais preocupadas com a pressão exercida por normas locais e mais preocupadas em asseverar seu status de acordo com (...) a estrutura social. (Romaine, 1978)<sup>54</sup>

Fischer (1958, apud Paiva In Mollica, 1992 )<sup>55</sup>, analisando a variação entre *-ing* e *-in*, terminações verbais em inglês, verifica que a forma *-ing* (a mais prestigiada) é mais frequente entre as mulheres do que entre os homens, que preferem a forma *-in* de menos prestígio social.

Quanto ao português, Scherre (1985, apud id.ibid.), com o estudo da concordância nominal na fala carioca, também verifica diferenças entre homens e mulheres, com essas favorecendo a marca de plural (forma padrão) mais do que os homens.

<sup>54</sup> (...) females show a greater sensitivity to socially evaluative linguistic features than do males (Wolfram, 1969).

In careful speech, women use fewer stigmatized forms than men, and are more sensitive than men to the prestige pattern (Labov, 1972).

Females show more awareness of prestige norms in both their actual speech and their attitudes toward speech (Wolfram & Fasold, 1974).

The females (...) are clearly more concerned with the pressure exerted by local norms and asserting their status within the (...) social structure (Romaine, 1978).

<sup>55</sup> Fischer (1958), com o estudo intitulado *Influências sociais na escolha de variantes lingüísticas*, é a primeira referência à influência do fator sexo sobre um fenômeno de variação.

Um outro estudo sobre o português refere-se ao fenômeno da passagem de vibrante múltipla para fricativa velar na fala carioca desenvolvido por Callou (1987, apud id.ibid.). Segundo a pesquisadora, "também neste caso a mudança se dá em direção a um novo padrão, com o sexo feminino favorecendo a pronúncia inovadora".

Labov (1966, apud id.ibid.), estudando a variação na ausência ou presença de /r/ em posição final ou pré-consonantal, verificou que a mudança se processava em direção à implementação de uma forma socialmente prestigiada: na pronúncia de /r/, padrão nova-iorquino. Neste caso, as mulheres lideravam a implementação da nova pronúncia.

Feitas essas breves considerações, passamos a apresentar as variáveis sociais estratificadas de nossa amostra e os resultados estatísticos.

Conforme já explicitamos no capítulo 4, analisamos no trabalho três variáveis sociais convencionais: **Sexo, Faixa etária e Grau de escolarização**. Além dessas variáveis utilizadas tradicionalmente, estamos considerando também a variável **Etnia**, que compõe, juntamente com as anteriores, a estratificação social da amostra do Banco de Dados do Projeto VARSUL. Scherre (1991) trabalha apenas com as variáveis convencionais e analisa quatro faixas etárias (07 - 14 anos; 15 - 25 anos; 26 - 49 anos e 50 - 71 anos), enquanto consideramos apenas duas: **25-49 anos e mais de 50 anos**.

Os três primeiros fatores sociais são considerados com a finalidade de se verificar sinais de variação estável ou de mudança em progresso envolvendo a concordância de número no predicativo e no participio passivo. A variável etnia, por sua vez, é analisada com o objetivo de se constatar a influência dos grupos étnicos sobre a presença de marcas de plural.

Seguindo a hipótese tradicional, esperamos que as mulheres realizem mais as marcas formais de plural do que os homens. Nossas expectativas na análise da variável **idade** é de que os informantes mais velhos marquem mais o plural nos predicativos/participios. No grupo de fatores **grau de escolarização**, nossa hipótese é de que a presença de marcas de plural (forma de prestígio) é diretamente proporcional aos anos de escolarização. Com relação à variável etnia, não temos, por enquanto, qualquer hipótese.

Abaixo, temos a tabela 11 com os resultados da análise dessas variáveis sociais.

**TABELA 11**  
**Variáveis Sociais**

<b>FATORES</b>	<b>Apl/Total</b>	<b>%</b>	<b>P.R.</b>
<b>1) Sexo</b>			
Feminino	194/417	47	.57
Masculino	133/351	38	.42
<b>2) Grau de escolarização</b>			
Primário	45/217	21	.27
Ginasial	90/253	36	.43
Colegial	192/298	64	.72
<b>3) Faixa etária</b>			
25-49 anos	123/364	34	.41
+50 anos	204/404	50	.59
<b>4) Etnia</b>			
Açoriana	89/247	36	.44
Italiana	150/284	53	.57
Eslava(Ucraniana)	88/237	37	.48
<b>TOTAL</b>	<b>327/768</b>	<b>43</b>	

Todas as variáveis sociais foram selecionadas como estatisticamente significativas, nesta ordem: grau de escolarização, idade, sexo e etnia.

Como esperávamos, as mulheres favorecem mais a forma de prestígio (.57) do que os informantes do sexo masculino (.42). Da mesma forma, a expectativa era de que os informantes mais velhos realizassem mais marcas de plural do que os mais jovens, o que ocorreu, com os primeiros concordando em .59 para .41 entre os mais jovens.

A hipótese de que os informantes com maior grau de escolarização utilizariam mais a forma de prestígio (norma culta ensinada na escola) do que os informantes com menos anos de escolarização também foi confirmada. Os resultados mostram que há uma relação direta entre o aumento dos anos de escolarização, com uma maior exposição à norma padrão culta, e o peso relativo de presença da marca formal de plural no predicativo/particípio passivo: os informantes primários concordam em apenas .27, os ginásiais em .43, enquanto os colegiais em .72.

Na análise da variável etnia verificamos que o grupo étnico **italiano** favorece mais marcas de plural no predicativo/particípio (.57 de peso relativo). A etnia **eslava** aparece como o grupo intermediário quanto à utilização da marca formal de plural (.48). E, por fim, o grupo étnico **açoriano** comporta-se como o menos favorecedor à forma de prestígio (.44).

Acreditamos aqui tratar-se de uma diferenciação de nível de ensino, sem falar, é claro, da possível interferência dos dialetos. Sugerimos que seja feito um estudo sobre a língua de origem de cada grupo para se verificar até que ponto a estrutura da mesma possa estar influenciando na concordância de número nos predicativos/particípios passivos.

Pela configuração acima, seríamos levados a concluir que o fenômeno de concordância no predicativo/particípio passivo reflete uma **variação lingüística estável**, com as mulheres favorecendo mais marcas de plural e com a utilização de formas de prestígio diretamente proporcional à escolaridade dos falantes.

Com relação à variável **idade**, não podemos ainda tirar conclusões com relação à possibilidade de estar indicando alguma mudança em progresso. O fato de termos à disposição apenas duas faixas etárias que oferecem, obrigatoriamente, uma distribuição inclinada não quer dizer, necessariamente, que o fenômeno em estudo esteja passando por um processo de extinção de uma das variantes.

Precisaríamos ir além com a pesquisa e acrescentar mais uma faixa etária para daí, sim, podermos trabalhar com resultados mais consistentes com relação ao fenômeno de mudança, além, é claro, da realização de uma investigação em tempo real (v. Labov, 1994).

A análise dos fatores sociais em separado mostrou alguns comportamentos frente ao fenômeno em estudo.

Entretanto, uma análise mais detalhada, com o cruzamento desses fatores entre si (bem como de variáveis lingüísticas entre si), é necessária para que se tenha uma fotografia mais real da regra. Através dos cruzamentos, buscamos mostrar a estreita correlação entre alguns grupos de fatores.

Primeiramente, apresentaremos os resultados e a análise dos cruzamentos de variáveis sociais entre si, para, em seguida, abordarmos os outros tipos de cruzamentos que fizemos neste trabalho.



## 8 - CRUZAMENTO DE VARIÁVEIS

Nosso objetivo principal em efetuar rodadas cruzando os fatores sociais entre si é o de verificar o grau de correlação que cada variável extra-lingüística tem com as demais variáveis desse tipo, bem como observar como essas correlações interferem no fenômeno em estudo. Para tanto, executamos todos os cruzamentos possíveis, num total de seis: **escolaridade com etnia/faixa etária/sexo, etnia com sexo/faixa etária, e sexo com faixa etária.**

Na tabela 12 apresentamos os resultados obtidos do cruzamento da variável **escolaridade com etnia.**

**TABELA 12**  
**Escolaridade com Etnia**

<b>ETNIA</b>	<b>PRIMÁRIO</b>	<b>GINASIAL</b>	<b>COLEGIAL</b>
<b>Açoriana</b>	F. 19/84=23% P. .28	F. 22/76=29% P. .38	F. 48/87=55% P. .61
<b>Italiana</b>	F.10/55=18% P. .23	F. 41/96=43% P. .48	F. 99/133=74% P. .82
<b>Eslava</b>	F. 16/78=21% P. .26	F. 27/81=33% P. .41	45/78=58% P. .69

Pelos resultados da Tabela 12, podemos verificar que os anos de escolaridade mantêm-se diretamente proporcional à presença de marcas formais de plural, independente da etnia considerada. Se, por outro lado, fizermos a leitura do cruzamento **escolaridade com etnia**, observaremos alguns pontos interessantes ao estudo.

Com relação aos informantes primários, os pesos relativos não se distinguem muito, sendo que os de etnia italiana concordam menos (.23). Já com relação aos informantes ginasiais, o quadro se modifica, com os pesos relativos se distanciando e com o grupo dos italianos concordando mais (.48) do que os informantes eslavos (.41) e mais ainda do que os falantes açorianos (.38). Quanto aos informantes colegiais, novamente os italianos utilizam a marca formal de plural muito mais do que os os eslavos e açorianos: .82 para .69 e .61, respectivamente.

Os resultados da influência do fator etnia (cf. tabela 11 acima) nos mostram que o grupo étnico italiano é o que apresenta um peso relativo maior de presença de marcas de plural, com .57, enquanto os eslavos vêm em segundo lugar (.48) e os açorianos em terceiro, concordando menos (.44). Contudo, verificamos acima que, dos informantes primários, o grupo dos italianos é o que menos utiliza a concordância, o que não seria esperado. Neste caso, é possível que outros fatores estejam motivando este resultado.

A seguir, mostramos a Tabela 13 com os resultados da rodada cruzando escolaridade com faixa etária.

**TABELA 13**  
**Escolaridade com Faixa Etária**

<b>FAIXA ETÁRIA</b>	<b>PRIMÁRIO</b>	<b>GINASIAL</b>	<b>COLEGIAL</b>
<b>25-49 anos</b>	F. 23/107=21% P. .27	F. 37/121=31% P. .38	F. 63/136=46% P. .54
<b>+ de 50 anos</b>	F. 22/110=20% P. .26	F. 53/132=40% P. .46	F. 129/162=80% P. .84

Conforme verificamos na tabela acima, os informantes primários concordam menos, independente da faixa etária: .27 para os falantes de 25 a 49 anos e .26 para os de mais de 50 anos. Já a partir do nível ginásial, o quadro é outro, ou seja, os informantes ginásiais e colegiais de mais de 50 anos apresentam pesos relativos de presença de marca de plural maiores (principalmente os colegiais mais velhos) (.46 e .84, respectivamente) se comparados com os dos informantes de 25 a 49 anos: (.38 e .54, respectivamente).

Estes resultados nos indicam que o aumento dos anos de escolarização, tanto para os informantes mais jovens quanto para os mais velhos, acarretou num aumento no uso da concordância, só que em maiores proporções para esses últimos.

Uma interpretação para estes resultados pode ser a de que estaria atuando aqui o fator idade, com os mais jovens apresentando menos concordância por serem eles mais sensíveis à mudança e mais influenciáveis pelos mais diversos segmentos da sociedade. É muito provável que uma terceira faixa etária mais jovem apresentaria probabilidades de concordância ainda menores nestes dois níveis de escolaridade (ginásial e colegial).

Apresentamos agora a Tabela 14 com os resultados do cruzamento de escolaridade com sexo.

**TABELA 14**  
**Escolaridade com Sexo**

<b>SEXO</b>	<b>PRIMÁRIO</b>	<b>GINASIAL</b>	<b>COLEGIAL</b>
<b>Feminino</b>	F. 38/123=31% P. .42	F. 55/144=38% P. .48	F. 101/150=67% P. .76
<b>Masculino</b>	F. 7/94=7% P. .10	F. 35/109=32% p. .41	F. 91/148=61% p. .70

Os resultados da tabela acima vêm corroborar a hipótese de que as mulheres, mais propensas ao uso de formas da norma culta, concordam mais do que os homens. Contudo, chama-nos a atenção a combinação **primário/masculino** em que o peso relativo de concordância, se comparado ao das mulheres com escolaridade primária, é consideravelmente baixo (apenas .10 para .42 das mulheres primárias).

Esta configuração nos mostra que a escolarização, nos níveis ginásial e colegial, está pesando muito mais para os informantes do sexo masculino do que para as mulheres, ou seja, na medida em que vão aumentando os anos de escolarização os homens praticamente vão se igualando às mulheres (ginásial – .41 para os homens e .48 para as mulheres; colegial – .70 para os homens e .76 para as mulheres), o que não ocorre no nível primário, conforme vimos, parecendo revelar aí uma influência maior do fator sexo.

Abaixo temos os resultados obtidos do cruzamento de **etnia com sexo**.

**TABELA 15**  
**Etnia com Sexo**

<b>SEXO</b>	<b>AÇORIANO</b>	<b>ITALIANO</b>	<b>ESLAVO</b>
<b>Feminino</b>	F. 64/152=42% P. .51	F. 74/142=52% P. .61	F. 56/123=46% P. .58
<b>Masculino</b>	F. 25/95=26% P. .36	F. 76/142=54% P. .53	F. 32/114=28% p. .36

A tabela 15, com os resultados do cruzamento de etnia com sexo, mostram alguns pontos que devem ser considerados.

Em cada grupo étnico, as mulheres apresentam pesos relativos de presença de plural maiores do que os homens.

Fazendo uma leitura horizontal, observamos que as mulheres italianas concordam mais (.61), ao passo que os homens açorianos e eslavos concordam menos (.36). A configuração mostrada aqui segue a hipótese de que as mulheres concordam mais, além de corroborar os resultados da tabela 11 que traz a etnia italiana com maior probabilidade de concordância, em segundo a etnia eslava e, por último, a etnia açoriana.

Um ponto, entretanto, foi surpreendente: no grupo dos italianos, a etnia parece se sobrepôr ao fator sexo, ou seja, os homens italianos apresentam frequência (e principalmente essa) e peso relativo maiores (54% e .53) do que as mulheres açorianas (.42% e .51).

Na Tabela 16 abaixo, mostramos os resultados da rodada cruzando os fatores etnia e faixa etária.

**TABELA 16**  
**Etnia com Faixa Etária**

<b>FAIXA ETÁRIA</b>	<b>AÇORIANOS</b>	<b>ITALIANOS</b>	<b>ESLAVOS</b>
<b>25 a 49 anos</b>	F. 39/125=31% P. .42	F. 55/115=48% p. .50	F. 29/124=23% P. .28
<b>+ de 50 anos</b>	F. 50/122=41% P. .45	F. 95/169=56% P. .64	F. 59/113=52% P. .67

Os informantes de mais de 50 anos, independente da etnia, concordam mais, confirmando as expectativas. Na faixa etária dos 25 a 49 anos, temos os italianos concordando mais (.50) e os eslavos concordando menos (apenas .28). Já na faixa dos mais de 50 anos, os eslavos e italianos concordam mais (.67 e .64, respectivamente).

O peso relativo de .28 para os escravos mais jovens possivelmente esteja indicando uma influência maior do fator idade, uma vez que os escravos apresentam peso relativo de concordância maior do que os açorianos (cf. Tabela 11).

Finalmente, temos a Tabela 17 mostrando os resultados do cruzamento de sexo com faixa etária.

**TABELA 17**  
**Sexo com Faixa Etária**

<b>FAIXA ETÁRIA</b>	<b>FEMININO</b>	<b>MASCULINO</b>
<b>25 a 49 anos</b>	F. 69/203=34% P. .43	F. 54/161=34% P. .38
<b>+ de 50 anos</b>	F. 125/214=58% P. .68	F. 79/190=42% P. .46

Os resultados dos cruzamento dos fatores sexo e faixa etária vêm confirmar nossas expectativas de que mulheres mais velhas concordam mais e homens mais jovens concordam menos. Os valores que encontramos mostram isso, conforme a tabela acima: peso relativo maior para as mulheres com mais de 50 anos (.68) e peso relativo menor para os homens de 25 a 49 anos (.38).

Se fizermos uma leitura horizontal da tabela, veremos que a diferença entre o peso relativo das mulheres e o dos homens mais jovens é de apenas .05, enquanto que essa diferença passa para .22 se compararmos as mulheres e os homens mais velhos. O que já era esperado, então, se confirma: as mulheres mais velhas concordam muito mais que os homens mais velhos, ao passo que entre os jovens o fator sexo não está pesando tanto.

Além dos cruzamentos entre as variáveis sociais, realizamos, também, conforme já foi colocado, rodadas cruzando variáveis lingüísticas entre si.

Interessa-nos, a priori, dois tipos: o primeiro, **processos com tonicidade**, devido ao fato de estudos apontarem a existência de interferência entre estes dois grupos de fatores (Scherre, 1988); o segundo, **distância do sujeito correferente com paralelismo formal**, porque, olhando para os dados, verificamos um possível condicionamento desta sobre aquela, ou seja, casos em série em que o dado anterior com todas as marcas de plural levava o dado subsequente a ser marcado também, mesmo que o seu sujeito correferente estivesse próximo.

Não foi possível, entretanto, executar a rodada do cruzamento da variável **distância do sujeito correferente com paralelismo formal** em virtude de o Programa VARBRUL não ter aceito o imput para o IVARB.

Acreditamos que isto se deu devido à utilização do fator "não se aplica" na codificação desta variável para os dados com sujeito explícito. Precisamos, portanto, rever esta questão.

O cruzamento das variáveis **processos morfofonológicos de formação do plural e tonicidade dos itens singulares** (que denominamos de **variável saliência**), não foi selecionado pelo Programa VARBRUL como estatisticamente relevante. Dando continuidade ao trabalho, poder-se-á retomar este ponto.

## 9- CONSIDERAÇÕES FINAIS

Conforme explicitamos na introdução deste trabalho, três foram os objetivos básicos da investigação realizada:

- 1 – apresentar a descrição do conjunto de variáveis lingüísticas e extralingüísticas que regem a variação da concordância de número nos predicativos/participios passivos;
- 2 – trazer à tona maiores evidências que venham corroborar os Princípios do Processamento Paralelo, da Saliência Fônica e da Quantidade;
- 3 – verificar se o fenômeno em estudo reflete um estágio de variação sociolingüística estável ou um processo de mudança lingüística em progresso.

Além disso, procuramos confirmar as hipóteses estabelecidas por Scherre (1991), bem como traçar um quadro comparativo entre os resultados alcançados pela autora e os obtidos neste estudo.

Por tudo que vimos, fica a certeza de que alcançamos em boa parte os objetivos propostos. Podemos constatar que os resultados lingüísticos de nossa análise com falantes de três cidades da região sul do Brasil e os de Scherre, com falantes cariocas, mostram uma convergência significativa.



Utilizando o mesmo instrumental de análise (VARBRUL), vimos que as variáveis escolhidas pelo Programa como mais significativas foram em sua grande maioria as mesmas, com uma única exceção: ao invés da variável processos morfofonológicos de formação do plural, que em Scherre (1991) foi selecionada como estatisticamente relevante, neste estudo, a variável tonicidade dos itens singulares aparece mais atuante para o favorecimento ou desfavorecimento da aplicação da regra de concordância de número nos predicativos/participios passivos.

Sendo assim, com relação aos grupos de fatores que se mostraram mais relevantes, tivemos as seguintes variáveis selecionadas conforme ordem da significação estatística em termos de pesos relativos:

- 1) Paralelismo formal das seqüências de predicativos/participios no discurso;
- 2) Características formais do verbo da construção;
- 3) Características formais do sujeito da construção;
- 4) Estrutura do predicativo;
- 5) Tonicidade dos itens singulares; e
- 6) Distância do sujeito correferente.

Se compararmos, então, com as variáveis selecionadas na análise de Scherre (1991), vemos que houve uma considerável convergência em termos de grupos lingüísticos mais significativos.

Em sua análise, os grupos lingüísticos selecionados foram: *Paralelismo formal das seqüências de predicativos/participios no discurso*, *Características formais do sujeito da construção*, *Características formais do verbo da construção*, *Estruturas do predicativo* e *Processos morfofonológicos de formação do plural*. É importante observar que, tanto em Scherre (1991) quanto neste estudo, a variável lingüística estatisticamente mais forte é o paralelismo formal no discurso.

Nesta investigação, não se mostraram relevantes as variáveis:

- 1) Ordem dos elementos na estrutura;
- 2) Material interveniente entre verbo e predicativo/particípio;
- 3) Processos morfofonológicos de formação do plural;e
- 4) Tipo de estrutura.

Quanto às variáveis sociais, todas foram selecionadas como estatisticamente significativas, sendo que a escolaridade se constitui no grupo de fatores mais relevante.

A análise das variáveis *Paralelismo formal das seqüências de predicativos/participios no discurso*, *Características formais do verbo da construção* e *Características formais do sujeito da construção* evidenciam, a níveis supra sentenciais no primeiro caso e sentenciais, nos dois últimos, que *marcas levam a marcas e zeros levam a zeros*.

Tal fato vem corroborar o *Princípio do Processamento Paralelo* proposto por Scherre (1988) que diz que "formas semelhantes se agrupam, através de processos mentais associativos presentes no desempenho lingüístico". Com relação às ocorrências de predicativos/participios em série, ainda não temos evidências suficientes para afirmar se a influência ocorre apenas da esquerda para a direita ou se existe uma influência recíproca no interior das mesmas.

De qualquer forma, esta investigação vem somar-se a todos os trabalhos até aqui desenvolvidos e que têm demonstrado, consistentemente, que o Princípio do Processamento Paralelo se mostra eficaz para explicar o funcionamento de um expressivo conjunto de fenômenos do Português do Brasil e de outras línguas naturais (Scherre, 1988:378-425). Neste sentido, podemos reafirmar as palavras da autora de que o "Princípio do Processamento Paralelo é um candidato a universal (...) e, como tal, deve ser incorporado pela teoria lingüística".

Na análise das variáveis Tonicidade dos itens singulares, Processos e Estrutura do predicativo, pudemos trazer maiores evidências ao *Princípio da Saliência Fônica* que estabelece que *formas mais salientes são mais perceptíveis e, por isso, mais marcadas*. Sendo assim, os predicativos/participios com itens *oxítonos não regulares nominais*, ou seja, mais salientes, apresentam mais chances de conter marcas formais de plural, conforme nossos resultados evidenciaram.

Pelos resultados que obtivemos com relação à variável distância do sujeito correferente, não trabalhada por Scherre (1991), verificamos a aplicabilidade do *Subprincípio da Quantidade* da Teoria Funcionalista de Givón, que diz que quanto menos previsível/predizível é a informação, mais código é necessário. Os dados mostraram que quanto mais distante estiver o sujeito zero de seu correferente, isto é, quanto maior a distância referencial (DR), menos predizível é a informação de plural nos predicativos/participios passivos, necessitando-se, então, de mais explicitação da marca.

Apesar de os resultados corresponderem às expectativas e de a variável ter sido selecionada como relevante, conclusões mais consistentes sobre a mesma ainda são prematuras. É necessário que se busque maiores evidências em análises futuras.

Com relação às variáveis ordem dos elementos na estrutura, tipo de estrutura e material interveniente entre o verbo e o predicativo, constatamos que as mesmas se comportaram como não significativas ao fenômeno em estudo e, quanto às hipóteses levantadas, fazemos as seguintes considerações.

A variável ordem dos elementos na estrutura foi a primeira que o Programa VARBRUL desconsiderou, e, ao contrário do que esperávamos, a ordem não canônica (que interrompe o fluxo normal do processamento) apareceu com peso relativo maior de presença de marcas de plural do que a ordem canônica. Da mesma forma, contrariando nossas expectativas, nesta análise a estrutura passiva (menos saliente pelo fato de todos os casos serem regulares) apresentou peso relativo maior do que a estrutura ativa.

Quanto à variável Material interveniente entre o verbo e o predicativo/particípio, buscávamos verificar até que ponto a ocorrência desses elementos influencia na presença de marcas formais de plural, e constatamos, a exemplo de Scherre, que a variável pouco interfere na presença/ausência de marcas nos elementos analisados. Contudo, a análise desta variável pode ser retomada na tentativa de se verificar a presença do material interveniente em termos de número de sílabas e o que isso pode influenciar na presença ou ausência de marcas de plural.

No que concerne às variáveis sociais, todas as nossas expectativas foram correspondidas. Assim, o aumento dos anos de escolarização (com uma maior exposição à norma padrão culta) mostrou-se diretamente proporcional à presença de marcas de plural nos predicativos/particípios passivos (forma de prestígio). Os informantes do sexo feminino, seguindo a hipótese tradicional de que as mulheres são mais sensíveis às formas de prestígio, realizam mais marcas de plural do que os homens. Os informantes com mais de cinquenta anos marcam mais o plural nos predicativos/particípios do que os informantes mais jovens (25-49 anos).

Não tínhamos uma hipótese quanto à variável etnia e nossos resultados mostraram que o grupo dos italianos marca mais o plural, seguidos dos eslavos e por último os açorianos. As investigações, entretanto, precisam prosseguir no sentido de verificar o que estaria por trás desse resultado. Poderíamos tentar encontrar respostas através da análise da estrutura das línguas de origem destes grupos, bem como os estágios de evolução das mesmas.

No que se refere à verificação de o fenômeno em estudo refletir um estágio de variação estável ou um processo de mudança em progresso, ainda não podemos concluir muita coisa. Dando prosseguimento à pesquisa, teremos que incluir uma terceira faixa etária para verificarmos se o fenômeno terá uma distribuição inclinada, indicando, assim, a possibilidade de um processo de mudança na língua, ou, ao contrário, curvilíneo, mostrando que o fenômeno passa por um processo de variação estável.

Tentando fazer uma espécie de generalização, somos levados a prever, então, que o protótipo do falante que faz muita concordância é do sexo feminino, com segundo grau, com mais de 50 anos e pertencente ao grupo étnico italiano. Provavelmente, ele não colocará marcas de plural justamente nos ambientes lingüísticos menos propícios à existência de marcas, ou seja:

- 1) em predicativos/participios precedidos de predicativos/participios sem marca(s) de plural;
- 2) em estruturas contendo verbo sem marca de plural;
- 3) em estruturas com sujeito explícito sem a(s) última(s) marca(s) formal(is) de plural e com sujeito explícito com a marca de plural totalmente neutralizada;
- 4) em predicativos de mais de um elemento com os itens *tudo/todo*;
- 5) em predicativos/participios com itens paroxítonos regulares e com itens diminutivos/aumentativos.

Já o informante que faz pouca concordância, possivelmente do sexo masculino, açoriano ou eslavo, de nível primário e pertencente à faixa etária de 25 a 49 anos, quando a fizer, realizará marcas de plural justamente nos ambientes que mais provocam a presença das mesmas, isto é:

- 1) em predicativos/particípios precedidos de predicativos/particípios com todas as marcas de plural;
- 2) em estruturas contendo verbo com marcas de plural;
- 3) em estruturas com sujeito explícito com todas as marcas formais de plural;
- 4) em predicativo nominal de 1 elemento;
- 5) em predicativos com itens oxítonos não regulares.

Em suma, pelo que pudemos depreender de nossos resultados, algumas coisas precisam ser revistas e outras analisadas, conforme explicitamos ao longo deste capítulo. Além disso, consideramos essencial que se investigue o fenômeno também no texto escrito, bem como em estilos formais de fala.

Esperamos que a verificação das hipóteses aqui testadas, bem como as interrogações levantadas, possam servir para futuras investigações que levem a um melhor entendimento dos condicionadores lingüísticos e sociais que envolvem a concordância de número nos predicativos e nos particípios passivos no português falado no Brasil.

## REFERÊNCIAS BIBLIOGRÁFICAS

ASSIS, Rosa Maria. "Variações Lingüísticas e Suas Implicações no Ensino do Vernáculo: Uma abordagem sociolingüística". In: **Sociolingüística**. Revista Ilha do Desterro. Florianópolis: Editora da UFSC, nº 20, 2º semestre de 1988 (pp.59-81).

BRIGHT, William. "As Dimensões da Sociolingüística". In: FONSECA, Maria Stella V. & NEVES, Moema F. (Org.). **Sociolingüística**. Rio de Janeiro: Eldorado, 1974.

CASTILHO, Ataliba Teixeira de. "Apresentação do Projeto de Gramática do Português Falado". In: CASTILHO, Ataliba T. de (Org.). **Gramática do Português Falado – A Ordem**. Vol. 1. Campinas: Editora da UNICAMP /FAPESP, 1990.

CHAMBERS, J. K. **Sociolinguistic Theory – Linguistic Variation and its Social Significance**. Cambridge, Massachussets: Basil Blackwell, 1995.

DU BOIS, John W. "Competing Motivations". In: HAIMAN, John. **Iconicity in syntax**. Amsterdan: John Benjtiming, 1984 (pp. 342-65).

ELIA, Silvio. **Sociolingüística**. Niterói: EDUF, 1987.

FARACO, Carlos Alberto. **Lingüística Histórica**. São Paulo: Editora Martins Fontes, 1988.

FASOLD, Ralph. **Sociolinguistics of Language**. Cambridge, Massachussets: Basil Blackwell, 1990.

GIVÓN, T. (Ed.). **Topic Continuity in Discourse: A Quantitative Cross-Language Study**. Philadelphia: John Benjamins Publishing Company, 1983.

-----, "Isomorphism in the Grammatical Code: Cognitive and Biological Considerations"  
In: **Studies In Language - International Journal Sponsored by the Foundation**



"Foundations of Language". Philadelphia: John Benjamins North America, Inc. vol. 15.

n° 1 (1991:85-115).

KNIES, Clarice & COSTA, Iara Bemquerer. **Manual do Usuário - Banco de Dados Lingüísticos VARSUL**. 1995 (Mimeo).

LABOV, William. "Field Methods of the Project on Linguistic Change and Variation". In: BAUGH, John & SHERZER, Joel (Eds.). **Language In Use - Readings in Sociolinguistics**. New Jersey: Prentice Hall, 1984.

----- . **Language in the inner city**. Philadelphia: University of Pennsylvania Press, 1972a.

----- . **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press, 1972b.

----- . **The design of a sociolinguistic research project**. 1972c (Mimeo).

LEMLE & NARO, Anthony J. **Competências básicas do português**. Relatório final de pesquisa apresentado às instituições patrocinadoras. Fundação Movimento Brasileiro de alfabetização (MOBRAL) e Fundação Ford. Rio de Janeiro, 1977.

MOLLICA, Maria Cecília (Org.) **Introdução à sociolingüística variacionista**. Rio de Janeiro. UFRJ, 1992.

MORAES, Euzi Rodrigues. "Alguns Aspectos da Variação Sintática no Português do Brasil". In: **Sociolingüística e Ensino do Vernáculo**. Rio de Janeiro: Editora Tempo Brasileiro. n° 78/79. julho-dezembro de 1984 (pp. 122-146).

MUSSAD, Albert E. "Human freedom in Chomsky's system of ideas". In: **The Georgetown Journal of Languages & Linguistics**. Vol. 2, n° 01, 1991 (pp.113-21).

NARO, Anthony J. "The social and structural dimensions of a syntactic change. In: **Language**, LSA, vol. 57, n° 01, 1981 (pp. 63-98).

----- et. al. **Relatório final de pesquisa: Projeto Subsídios do projeto Censo à Educação**. Rio de Janeiro. UFRJ, 1986 (Mimeo).

PASCALE, Rousseau & SANKOFF, David. "Advances in Variable Rule Methodology" In:

SANKOFF, D. (ed.). **Linguistic Variation – Models and Methods**. New York: Academic Press, 1978.

PINTO, Ivone I. & FIORETTI, Maria Thereza G. **Tutorial para o pacote VARBRUL**, 1992  
(Mimeo).

POPLACK, Shana. **Function and process in a variable phonology**. Ph.D dissertation, University of Pennsylvania, 1979.

-----, "The notion of the plural in Puerto Rico Spanish: competing constraints on /s/ deletion". In: LABOV, William (Ed.). **Locating language in time and space**. Philadelphia, University of Pennsylvania, 1980.

SAUSSURE, Ferdinand de. **Curso de Linguística Geral**. São Paulo: Cultrix, 1992.

SCHERRE, Maria Marta P. "A concordância de número nos predicativos e nos participios passivos". In: **Organon**, nº 18, 1991 (pp. 52-70).

-----, **Reanálise da concordância nominal em português**. Tese (Doutorado em Letras) UFRJ, 1988.

----- et. al. (Orgs.). **Programas VARBRUL – dicas para o uso do computador – versão 2**. Rio de Janeiro, 1992. (Mimeo)

-----, **Introdução ao Pacote VARBRUL para Microcomputadores**. Rio de Janeiro, 1992 – Brasília, 1993. (Mimeo)

----- & NARO, Anthony J. "Marking in discourse: Birds of a feather". In: **Language Variation and Change**, vol. 03, 1991 (pp. 23-32).

----- & NARO, Anthony J. "The serial effect on internal and external variables". In: **Language Variation and Change**, vol. 04, 1992 (pp. 1-13).

----- & NARO, Anthony J. "Duas Dimensões do Paralelismo Formal na Concordância Verbal no Português Popular do Brasil". In: **D.E.L.T.A.**, vol. 09, nº 01, T1993 (pp. 1-14).

SOARES, Magda. **Linguagem e Escola - Uma perspectiva social**. São paulo: Ática, 1991

SUASSUNA, Livia. **Ensino de Língua Portuguesa: Uma abordagem pragmática**.  
Campinas: Papyrus, 1995.

TARALLO, Fernando. **A pesquisa sociolingüística**. São Paulo: Ática, 1986.

----- (Org.). **Fotografias Sociolingüísticas**. Campinas: Fontes: Editora da UNICAMP,  
1989.

----- **Tempos Lingüísticos – Itinerário histórico da língua portuguesa**. São Paulo:  
Ática, 1990.

----- "Reflexões sobre o conceito de mudança lingüística". In: **Organon**, nº 18, 1991 (pp.  
11-22).

----- "A Sociolingüística Na (E Da) Alfabetização". In: **D.E.L.T.A.**, vol. 11, nº 01, 1995  
(pp.91-132).

----- & DUARTE, M. E. L. "Processos de Mudança Lingüística em Processo: A  
Saliência vs. não Saliência de Variantes". In: **Sociolingüística**. Ilha do Desterro.  
Florianópolis: Editora da UFSC. nº 20, 1988 (pp. 44-58).

WARDHAUGH, Ronald. **An Introduction to Sociolinguistics**. New York: Basil Blackwell,  
1986.

WEINREICH, U. LABOV, W. and HERZOG, M. "Empirical foundations for a theory of  
language change". In: LEHMANN, W.P. and MALKIEL, Y. (eds.). **Directions for  
historical linguistics**. Austin: University of Texas Press, 1968.