

CORPUS DE LÍNGUA ORAL

Os critérios rigorosos utilizados para a constituição do *corpus* de língua oral para fins do doutorado¹, mediante o controle de variáveis linguísticas – relativas às especificidades da língua falada – e variáveis extralinguísticas – região de origem, sexo, escolaridade, faixa etária, nível socioeconômico, condições extraverbais de interação dialógica – na seleção dos informantes e nos critérios de armazenamento dos dados, permitiram a obtenção de uma amostra representativa do português falado paulista, passível de ser objeto de estudo em diferentes áreas dos estudos da linguagem e de áreas afins. Trata-se de *corpus* compilado, também conhecido como *corpus* de amostragem, porque é fixo, uma vez que foi compilado através de amostras pré-selecionadas.

Selecionamos 216 informantes de três regiões do Estado de São Paulo – a Capital e duas regiões do interior, Campinas e Itu.

A fim de tornar as amostras de cada região o mais possível comparáveis, estabelecemos alguns critérios que presidiram o levantamento dos informantes para a constituição da amostra. Alguns informantes foram selecionados dentro das escolas e outros, fora. Na escolha de uns e de outros, atendemos ao controle de variáveis extralinguísticas, consideradas relevantes na literatura sobre o assunto:

- região de origem – informantes naturais da região, tendo sempre nela residido, com pais também naturais da região;
- sexo – informantes de ambos os sexos;
- escolaridade – avaliada numa escala de seis pontos, desde o informante com superior completo até o analfabeto;
- faixa etária – em função do nível de escolaridade;

¹ ZAPPAROLI CASTRO MELO, Zilda Maria. *Análise do comportamento fonológico da junctura intervocabular no português do Brasil (variante paulista). Uma pesquisa linguística com tratamento computacional*. São Paulo, 1980. Tese (Doutorado em Linguística) – Programa de Pós-Graduação em Linguística do Departamento de Linguística da Universidade de São Paulo. /Disponível na Coleção Didática da Biblioteca Central da FFLCH/USP/. v.1, t.1.

- nível socioeconômico – determinado pela escolaridade e *status* profissional do chefe de família.

As amostras das falas dos informantes foram recolhidas de 26/11/1972 a 02/05/1973, por meio de diálogos gravados entre o informante e o entrevistador, num total de 54 horas e de 432 diálogos, visto que incluem dois tipos de interação dialógica – entrevistas e conversações. Gravamos quinze minutos de cada informante, assim distribuídos:

- cinco minutos iniciais – dados com os quais os entrevistados foram cadastrados.

Esses dados foram desprezados para a transcrição, a fim de possibilitar a acomodação psicológica do sujeito, uma vez que a ansiedade e o fato de o indivíduo estar ou não descontraído são fatores que influem na sua realização fonética. Na ficha do informante, nos dados pessoais, fizemos anotações relativas ao seu timbre de voz (grave, intermediário, agudo, outro), ritmo de fala (rápido, médio, lento, outro), espontaneidade de elocução (total, grande, média, fraca), por serem aspectos relevantes numa pesquisa fonética;

- do sexto ao décimo minutos – diálogo em situação de entrevista (formal), através de um guia de trabalho à maneira de um questionário único para todos os informantes: diálogo dirigido pelo documentador. A fim de lograr unidade no critério linguístico, entrevistamos os informantes das três regiões seguindo as diretrizes contidas num questionário previamente elaborado. Observe-se que, completados cinco minutos, a entrevista formal era interrompida, mesmo que o questionário não tivesse sido inteiramente aplicado. Para a entrevista formal, utilizamos o microfone de colete, a uma distância de 30 cm da cavidade bucal, com sensibilidade de gravação nº 2.
- cinco minutos finais: diálogo em situação de conversação (informal) – conversa livre/pate-papo.

Dessa forma, foram controladas, também, as condições extraverbais – formal e informal – do contexto interacional – condições que cercaram o diálogo no momento de sua realização. Como o comportamento verbal de um mesmo falante varia em função das circunstâncias em que se realiza o diálogo – variações por influência da situação² –, o controle das condições – formal/informal – possibilita verificar a variação da fala dos informantes em suas atitudes *formal* (fala esmerada) e *informal* (fala familiar).

Visando ao alcance de um bom controle experimental, para a composição dos grupos, adotamos a técnica de emparelhamento em função das variáveis – um grupo exerce controle sobre o outro –, de modo a haver seis informantes em cada subgrupo, o que permite que os resultados sejam comparáveis.

O Diagrama de Distribuição dos Informantes apresenta a distribuição dos informantes nas variáveis e nos diversos níveis de cada uma delas, demonstrando as várias possibilidades de relações contrastivas e de constituição de *subcorpora*.

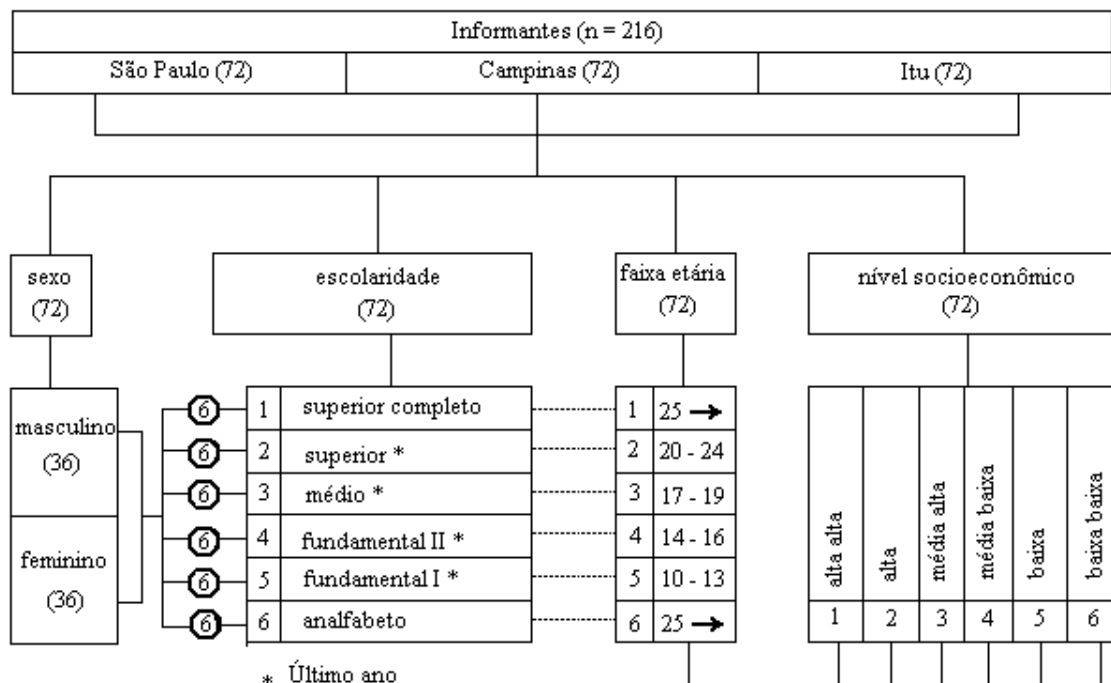


Figura – Diagrama de Distribuição dos Informantes

² Empregamos o termo situação no sentido que lhe atribui Bally: "a realidade extralinguística em que se coloca o discurso, as circunstâncias gerais ou particulares que cercam as palavras pronunciadas" (BALLY, 1962, p. 127, *apud* PRETI, 1975, p. 21).