

## DO QUE TRATA ESTE LIVRO

Um dia, quando estava na fase inicial da pesquisa que deu origem a este livro, uma amiga advogada foi me visitar no Rio de Janeiro e viu um pequeno fichário de madeira cheio de gavetinhas, que servia de mesinha de canto ao lado de almofadões em um típico apartamento de estudante na década de 1980. Curiosa, perguntou-me o que tinha em tanta gavetinha. Expliquei que eram fichas com a transcrição fonética das vogais que eu estava estudando para a minha Tese de Doutorado. Espantada com a especificidade do estudo, ela arregalou os olhos e me questionou: “Cinco vogais dão para escrever uma tese?!”. Rindo, eu respondi: “Cinco vogais não! São quatro!”.

São, pois, essas quatro vogais médias, /o/, /ɔ/, /e/ e /ɛ/, objeto da minha Tese de Doutorado, *As pretônicas no falar baiano: a variedade culta de Salvador*, defendida em 1989 na Universidade Federal do Rio de Janeiro, que trago a público agora depois de trinta anos!

Esse trabalho ficou restrito ao conhecimento de alguns poucos que tiveram acesso a ele, ou pela consulta à Biblioteca da UFRJ ou ao Banco de Teses da CAPES, instituição que presta um relevante trabalho aos pesquisadores brasileiros. Dificultava a sua divulgação por meios eletrônicos a sua impressão no velho mimeógrafo. Incentivada por Jacyra Mota, pioneira na pesquisa das vogais

pretônicas brasileiras e ainda estudiosa delas, resolvi voltar a rever esse trabalho para disponibilizá-lo em forma de livro. Esse projeto não previu modificar o seu conteúdo, mas corrigir algumas imprecisões e erros da edição anterior. Isso foi feito, exceto na seção 4.2, onde registrei os títulos que vieram a lume nesse lapso de cerca de 30 anos em que estive afastada do tema.

Por que estudar as vogais pretônicas no português americano?

As vogais pretônicas<sup>2</sup> constituem, ao lado da entoação, uma das principais marcas dos dialetos brasileiros e, por que não dizer, entre a variedade brasileira e a portuguesa pelo menos desde o fim do século XVIII, quando, no português europeu (PE), as vogais átonas sofreram uma grande mutação que resultou na centralização e elevação de [e] pretônico e na elevação do [o] para [u] na mesma posição.

A língua portuguesa, na variedade brasileira, de fato apresenta uma alternância fonética que caracteriza o falar de cada região. Trata-se da tripla possibilidade de realização na fala das vogais orais /O/ e /E/<sup>3</sup> quando precedem a sílaba acentuada da palavra. Podem-se ouvir, por exemplo, *c[u]légio*, *c[o]légio*, *c[ɔ]légio*, *r[i]vista*, *r[e]vista* ou *r[ɛ]vista*<sup>4</sup>.

Essa diferença de altura das vogais pré-acentuadas se constitui numa forte marca regional, mas não caracteriza o sujeito socialmente, como parece ser o caso da variação entre os traços [recuado] e [não-recuado], antiga na língua, e

<sup>2</sup> Procurou-se, dentro do possível, adotar a nomenclatura da linguística moderna, admitindo-se, porém, termos já consagrados. É o caso, por exemplo, dos termos *tônica* e *pretônica*, que muitas vezes ocorrerão neste texto alternando com *acentuada* e *pré-acentuada*, apesar da imprecisão do termo *tônico* para referir-se ao acento de intensidade, caracterizador da língua portuguesa.

<sup>3</sup> Usam-se neste trabalho símbolos maiúsculos para indicar segmentos fônicos cujos traços não estão inteiramente especificados na base.

<sup>4</sup> Optou-se por só transcrever foneticamente toda a palavra quando absolutamente necessário à compreensão do texto. Assim o vocábulo onde ocorre a pretônica é transcrito ortograficamente, restringindo-se a transcrição fonética apenas à vogal em tela. Entre os exemplos, já citados nesta introdução, aparece o vocábulo *colégio*, que é escrito *c[u]légio*, *c[o]légio* e *c[ɔ]légio*, devendo o leitor entender a informação de que sua primeira vogal se realiza foneticamente ora como [u], ora como [o], ora como [ɔ]. De acordo com a conveniência, indicou-se a nasalidade da vogal da sílaba vizinha com til (*m[ɛ]n[ĩ]no*, *d[ɛ]n[õ]mina*). Convencionou-se ainda informar, depois de cada dado, separando-se por vírgulas, o número de vezes em que ele ocorre dentro da amostra (ex.: *c[u]légio*, 4; *c[o]légio*, 2; *c[ɔ]légio*, 6), assinalando com um asterisco a realização possível, mas não documentada no *corpus* (*\*c[ɔ]legiado*). Para simplificar a transcrição, foram registrados os nomes no masculino singular, eliminando distinções não pertinentes ao interesse da pesquisa, mas fazendo-se exceção aos casos em que essa resolução modificaria o contexto vocálico da pretônica, como nos adjetivos terminados em *-oso* (*n[ɛ]rv[o]so*, *n[ɛ]rv[ɔ]sa*, *n[ɛ]rv[ɔ]sos*) e nos plurais dos nomes em *-ão* (*f[u]gão*, *f[u]gões*). Por essa mesma razão, foram respeitadas as formas em que ocorreram os verbos.

que se encontra em exemplos como \**pr[ɔ]guntar* :: \**p[ɛ]rguntar*, \**tr[ɔ]vuada* :: \**tr[ɛ]vuada*, \**p[u]rc[ɛ]lana* :: \**p[u]rç[u]lana*, que revelam a classe social do locutor além de sua origem rural. O presente trabalho se restringe ao primeiro caso, isto é, ao estudo da variação entre vogais alta, média e baixa da série recuada ([u] :: [o] :: [ɔ]) e da série não-recuada ([i] :: [e] :: [ɛ]) na classe média de Salvador, na Bahia.

Antenor Nascentes, no primeiro quartel do século passado, exatamente em 1922 (data de publicação da primeira edição de *O linguajar carioca*), chamou atenção para a importância da variação entre vogais abertas (baixas) e fechadas (médias), quando estabeleceu essa isoglossa como limite dialetal do *falar brasileiro* e dividiu o território nacional em subfalares do Norte e subfalares do Sul. Segundo o autor, “o que caracteriza estes dois grupos é a cadência e a existência de pretônicas abertas em vocábulos que não sejam diminutivos nem advérbios em *-mente*” (NASCENTES, 1953 [1922], p. 25).

Apesar disso, o *subfalar baiano*, que compreendia Bahia e Sergipe e parte de Minas, de acordo com essa classificação, estaria incluído entre os *subfalares do Sul*. Neste trabalho apresenta-se um estudo quantitativo desse traço no dialeto urbano culto da cidade de Salvador-Bahia.

Utilizando uma amostra de oito horas de fala, com 12 informantes<sup>5</sup> graduados por escola superior, nascidos na Cidade do Salvador<sup>6</sup>, descreve-se a realização de /E/ e /O/, estabelecendo as regras que determinam a sua altura. Entram em discussão as variantes documentadas em sílaba inicial aberta ([i]xiste :: [e]xiste) ou fechada ([i]sgoto :: [e]sgoto), em sílabas internas fechadas (*prop[u]rção* :: *prop[o]rção* :: *prop[ɔ]rção*) ou abertas (*s[i]m[i]lhantes* :: *s[e]m[e]lhantes* :: *s[ẽ]m[ɛ]lhantes*). Por último são examinadas também as vogais pré-acentuadas que constituem ditongos (*el[ey]ção* :: *el[ɛy]ção*) e hiatos (*r[i]almente* :: *r[ɛ]almente* :: *r[ɛ]almente*).

Não é propósito deste trabalho discutir questões teóricas. Procura-se tão somente descrever esse conjunto de dados, levantar hipóteses explicativas para o funcionamento sincrônico das regras no dialeto, ampliando, assim, o conhecimento que se tem da língua portuguesa. Além disso, buscaram-se também in-

<sup>5</sup> Participam dessa amostra 12 homens e 12 mulheres divididos em três faixas etárias: F<sub>1</sub>, que alcança locutores de 25 a 35 anos; F<sub>2</sub>, formada por indivíduos de 36 a 55 anos, e, finalmente, F<sub>3</sub>, que congrega os maiores de 55 anos, distribuídos igualmente. Todos os informantes tinham o curso universitário completo, restrição já imposta pelo Projeto NURC.

<sup>6</sup> Exigiu-se também que fossem filhos de indivíduos nascidos na Bahia, seja capital ou interior, a fim de se evitar a interferência de outro dialeto no seu vernáculo.

formações na história da língua, consultando velhos gramáticos e ortógrafos, arriscando explicar a origem dessa diferença dialetal.

Seguindo essa trilha e a das minhas próprias observações como falante nativa do “subfalar baiano”, busquei dividir a matéria, considerando o padrão silábico onde se situam as pré-acentuadas. Assim procedi porque se anteviam diferenças marcantes no comportamento dessas vogais de acordo com o contexto silábico. Notava-se, por exemplo, que a vogal /E/ nesse *corpus* se elevava quase categoricamente quando, em posição inicial absoluta, era seguida por um /S/ pós-vocálico (cf. 2.2.1: [i]scola, [i]spécie), confirmando o que ocorre em outros dialetos, mas que, diante de outras consoantes, ela se harmonizava com a altura da vogal seguinte (cf. 2.2.1.1.2: [ɛ]l[ɛ]mento; [e]leitor; [i]xistia). Observava-se também, continuando com o mesmo exemplo, que, no padrão CVC, não havia nenhum contexto em que /E/ se elevasse categoricamente. Por outro lado, o manuseio dos dados fazia notar que /O/, na posição inicial absoluta, raramente se elevava (cf. 2.2.1.2). Portanto a reunião desses dados produziria o efeito de aumentar demais o índice de elevação de /E/, porque incluiria numerosas ocorrências de [is], e de diminuir demais o índice de elevação de /O/, porque incluiria dados em que a elevação raramente ocorria. A única desvantagem desse recorte da amostra era não se poder quantificar os dados segundo o modelo escolhido<sup>7</sup>, quando estivessem em jogo as pretônicas em posição inicial absoluta, as pretônicas em hiato e os ditongos, cujas ocorrências eram muito reduzidas.

Para eliminar tal desvantagem, bastaria ampliar essa parte da amostra, solução pela qual não optei por entender que a pesquisa é como um novelo, cujo fio pode ser puxado sempre mais, e que deve ser cortado em algum ponto mais ou menos arbitrário, para que o trabalho seja concluído. Preferi, aqui, pois, deixar por conta do tempo de gravação fixado inicialmente para a realização desta pesquisa o estabelecimento do número de dados nos diversos padrões silábicos.

Por essas razões, foram discutidas, em primeiro lugar, as vogais precedidas de consoante, isto é, nos padrões CV e CVC, cujos dados, mais numerosos, foram submetidos aos programas computacionais que forneceram a percentagem e a probabilidade de aplicação da regra estudada (cf. seção 2.1).

---

<sup>7</sup> Essa pesquisa utilizou dois programas para os estudos apresentados nos capítulos 2 e 3, em que o número de dados permitiu a sua utilização: o SWAMINC, criado por Naro em 1977, que calcula as porcentagens da ocorrência do fenômeno na presença de cada fator e prepara os dados para serem submetidos ao segundo programa, o VARBRUL-2, responsável pelos cálculos probabilísticos, ou seja, por fornecer a probabilidade de a regra estudada se aplicar diante de um certo contexto. No Capítulo 3 recorreu-se à percentagem pela razão inversa.

Na segunda parte do Capítulo 2, as pretônicas foram examinadas no padrão VC, V\$V (antes de vogal, em hiato) e V\̇ ou \̇V (antes e depois de semivogal, num ditongo), considerando-se apenas os percentuais de aplicação das regras em um contexto determinado.

O Capítulo 3 não tinha sido previsto no plano inicial do trabalho. Ele foi necessário para descrever a realização de vogais médias [o] e [e], que escaparam às descrições mencionadas nos parágrafos anteriores. Nesse ponto discutiu-se o que se denominou provisoriamente *Regra variável de timbre*.

Segue a análise dos dados o Capítulo 4, em que se busca reunir o maior número de informações possível sobre a trajetória das pretônicas no tempo e no espaço geográfico, especialmente no brasileiro. Essa especulação foi feita, inicialmente, com base numa bibliografia que inclui tanto estudos modernos quanto gramáticas e ortografias antigas, sobre a origem das variantes em questão. Em seguida, considerando a divisão de Nascentes (1953 [1922]), busquei reunir informações sobre os falares brasileiros com base em trabalhos da primeira metade do século XX e entre os contemporâneos. Na presente edição, essa parte foi atualizada com a relação das pesquisas acadêmicas produzidas no lapso de tempo entre a defesa da minha tese de Doutorado (1989) e o ano de 2018.

O princípio básico que orientou a investigação foi o de que a variação é uma condição natural das línguas, portanto a descrição pressupõe não só o conhecimento interno de sua estrutura invariante mas também do funcionamento de sua estrutura variante.

O conceito de língua como um sistema organizado começou, no início do século XX, com Saussure<sup>8</sup>, que distinguia a *linguística da língua* – a *linguística propriamente dita* – e a *linguística da fala*. No recorte que prioriza, a tarefa do linguista consiste na explicitação desse corpo organizado, a língua, quer na perspectiva sincrônica, quer na diacrônica, reconhecendo os fenômenos de variação apenas quando decorrentes das diversidades geográfica e temporal. Essa visão da língua não se contrapunha, assim, à doutrina linguística, anterior, fundada principalmente no interesse histórico de estudar a mudança linguística, para o que muito contribuiu a Dialectologia, que encontrava na variação de formas dialetais (entenda-se, aqui, dialetos geográficos) um meio para a reconstituição histórica.

---

<sup>8</sup> Em 1916 surge o texto inaugural da linguística moderna, o *Curso de linguística geral*, compilação feita, poucos anos depois de sua morte, por dois de seus alunos (Charles Bally e Albert Sechehaye), das notas de aulas do mestre e dos apontamentos de alunos presentes aos cursos que ele dera.

A dicotomia saussureana *langue x parole* atravessou o século XX, tanto quanto a separação que se estabeleceu entre *sincronia e diacronia*, e está na base da doutrina estruturalista. A doutrina de Saussure teve, porém, alguns opositores, como o filósofo Benedetto Croce, na Itália, e os chamados idealistas alemães, como Karl Vossler, Eugene Lerch e Leo Spitzer. Jespersen, outro de seus opositores, desprezou a dicotomia *língua x fala*, preocupado com a relação entre o indivíduo e a utilização coletiva da linguagem. Mais modernamente, Coseriu (1979) retomou a discussão sobre esse par dicotômico para propor a sua substituição pelos conceitos de *sistema, norma e fala*, melhor dizendo, por *sistema, norma coletiva, norma individual e fala*, considerando que o termo *língua*, muito geral, se presta a diversas interpretações. Pondo-se ao lado daqueles que consideram a fala “a única realidade investigável da linguagem”, Coseriu (1979, p. 71) estabelece uma escala de abstrações em cujo extremo mais concreto estaria a *fala* (descrita como atos de criação inéditos) e no mais abstrato, o *sistema* (sistema funcional de oposições). Entre *língua e fala* estariam as *normas: individual* (conjunto de formas linguísticas habituais do indivíduo) e *coletiva* (modelo segundo o qual todos os membros de um mesmo grupo linguístico realizam as possibilidades oferecidas pelo sistema), Já que, para o autor, os atos linguísticos, embora inéditos, são atos de “re-criação” e não “invenções ex-novo e totalmente arbitrárias do falante, mas se estruturam sob modelos precedentes, que os novos atos contém e, ao mesmo tempo, superam” (COSERIU, 1979, p. 72), parece justo inferir-se que o saber linguístico do indivíduo não corresponde apenas ao *sistema*, mas à *norma* do seu grupo social e à sua própria (que não são estáticas).

Essa concepção coincide em parte com as noções da *competência* linguística do falante ou do seu *desempenho*. A teoria gerativa padrão – que data dos meados do século passado –, à qual pertencem esses conceitos, distingue como parte do saber linguístico, da competência, portanto, o dicionário de itens lexicais e o sistema de regras interiorizadas pelo falante, que lhe permitem produzir e compreender um número infinito de frases inéditas, desfazer ambiguidades e julgar a propriedade e adequação (gramaticalidade) dos enunciados. E essas regras, que podem ser obrigatórias e opcionais, fazem parte da sua gramática e deverão ser o objeto de investigação do linguista. Assim as variações a que todas as línguas estão sujeitas são fenômenos do desempenho, não previsíveis pelas regras da gramática. Ficou assim afastada do interesse dos gerativistas a variação linguística.

Nesse sentido, a teoria padrão, pelo menos nisso, se aproxima do estruturalismo, que a precede. O interesse dos estruturalistas se limitava à descrição das

regras do sistema, bastando-lhes rotular os fenômenos variáveis como *variação livre*. Nenhuma dessas correntes linguísticas, portanto, reconheceu o caráter *sistemático* da variação, que constitui o objeto da Sociolinguística.

Como chamam à atenção Cedergren e Sankoff (1974), alguns linguistas – e entre eles, até certo ponto, Labov – tentaram explicar a variação através da existência de muitas gramáticas dentro de um contínuo, ou de descrições estruturais com regras cada vez mais pormenorizadas no intuito de, se não anular o número de ambientes onde ocorre variação, pelo menos torná-lo reduzido ao extremo. Ao contrário deles, esses autores argumentaram que a variação é fenômeno inerente às línguas e que, cada vez mais, as pesquisas revelam “[...] que as possibilidades representadas pelas regras opcionais abstratas são distribuídas de um modo bem padronizado em um dado falante e em uma determinada comunidade linguística” (CEDERGREIN; SANKOFF, 1974, p. 333)<sup>9</sup>.

Tendo em vista a regularidade dentro da variação, concluem:

Embora o desempenho possa ser considerado somente um reflexo aproximado da competência [...], é difícil escapar da conclusão de que aqueles aspectos do desempenho que são inteiramente sistemáticos no indivíduo e dentro da comunidade de falantes sejam reflexo da competência linguística. (CEDERGREIN; SANKOFF, 1974, p. 333-334)<sup>10</sup>

Essa nova abordagem da língua amplia, por isso, a concepção de competência do falante, considerando parte do saber linguístico também as regras que operam essa sistematicidade. Desse ponto de vista, a nova maneira de encarar a gramática do falante, longe de constituir um contraponto à teoria gerativa, parece uma ampliação de um de seus pressupostos<sup>11</sup>. As regras variáveis postuladas nesse modelo substituem as regras facultativas da GG<sup>12</sup>, especificando a ten-

---

<sup>9</sup> Do original: “[...] that the possibilities represented by abstract optional rules are distributed in a reproducible and well patterned way in a given speaker and in a given speech community” (tradução minha).

<sup>10</sup> Do original: “Although performance may be considered only an approximate reflection of competence [...], it is difficult to escape the conclusion that those aspects of performance that are found to be thoroughly systematic in an individual and throughout a community are reflections of linguistic competence” (tradução minha).

<sup>11</sup> Cf. Cedergren e Sankoff (1974, p. 333): “Essa competência é moldada nos termos gerativos convencionais, exceto que às regras optativas são atribuídas probabilidades de aplicação[...]”. Do original: “This competence is modeled in conventional generative terms, except that optional rules are assigned application probabilities[...]” (tradução minha).

<sup>12</sup> Cf. Labov (1978, p. 45): “Desse ponto de vista, pode-se dizer que uma regra variável nada mais é que uma regra facultativa prevista em um quadro probabilístico [...]”. Do original: “De ce point de vue, on peut dire qu’une règle variable n’est autre qu’une règle facultative envisagée

dência de aplicação da regra de acordo com restrições linguísticas e/ou sociais. Assim a Sociolinguística busca aprofundar o conhecimento que se tem sobre o que o falante “sabe”, para que seu comportamento linguístico se aproxime do comportamento dos outros membros do seu grupo (social, regional, etário etc.). Enfim, essa disciplina considera como princípio fundamental o da *variação inerente*, ou seja, o de que a variação faz parte do sistema.

Sobre essa questão ensina Labov (1982 [1969]):

A condição normal da comunidade de fala é a heterogeneidade: podemos esperar encontrar uma ampla gama de variantes, estilos, dialetos e línguas faladas por seus membros. Além disso, essa heterogeneidade é uma parte integrante da economia linguística da comunidade, necessária para satisfazer à demanda linguística da vida diária<sup>13</sup>. (LABOV, 1982 [1969], p. 17)

Decorrem desse entendimento da gramática do falante os três tipos de regras usadas no presente trabalho, propostas por Labov (1977, 1978)<sup>14</sup> para descrever o fenômeno linguístico: as regras categóricas, as regras semicategóricas e as regras variáveis.

As regras categóricas têm a mesma configuração que lhes deu a gramática gerativa, ou seja,  $X \rightarrow Y / A\_B$ , pela qual se entende que, dado o contexto  $A\_B$ ,  $X$  obrigatoriamente se torna  $Y$ . Já a formalização da regra variável difere da primeira pela inclusão de colchetes angulares em torno do elemento à direita da seta, que representam a sua natureza variável, com o sentido matemático de mais ( $>$ ) e de menos ( $<$ ). Desse modo  $X \rightarrow <Y> / A\_B$  é a formulação correta de uma regra variável<sup>15</sup>. As regras semicategóricas se distinguem das categóricas porque as violações das segundas produzem perturbação na comunicação, e se distinguem das regras variáveis pela raridade dessas violações. Mas todos os três tipos de regras são passíveis de violações por parte dos locutores na situação real de fala. O que as distingue são, tão somente, a frequência e o tipo de violação.

---

dans un cadre probabiliste [...]” (tradução minha).

<sup>13</sup> Do original: “The normal condition of the speech community is a heterogeneous one: we can expect to find a wide range of variants, styles, dialects, and languages used by members. Moreover, this heterogeneity is an integral part of linguistic economy of the community, necessary to satisfy the linguistic demands of every-day life” (tradução minha).

<sup>14</sup> As edições originais são, respectivamente, de 1972 (LABOV, W. The internal evolution of linguistic rules. In: STOCKWELL, R.; MACAULAY, R. (Eds.). *Historical linguistics and generative theory*. Bloomington: Indiana University Press, 1972, p. 101-171) e de 1969 (LABOV, W. Contraction, deletion and inherent variability of English copula. *Language*, 45, p. 715-762, 1969).

<sup>15</sup> Para maiores detalhes sobre esse ponto, cf. Labov (1978).

A formulação do *princípio da variação inerente* traz consequências não só para a abordagem sincrônica da língua como também para a diacrônica. Como a Dialetoлогия no passado, a Sociolinguística impulsionou os estudos sobre a mudança linguística. O relacionamento da Dialetoлогия com os estudos históricos baseava-se no fato de a mudança linguística não observar o mesmo ritmo em todas as regiões. Isso fica claro no seguinte trecho de Câmara Jr. (1979), que resumiu bem a questão:

A geografia linguística é, ademais, importante como uma nova abordagem ao estudo histórico comparativo. Em vez de ter que recorrer aos textos antigos de fases passadas extintas, *o investigador apenas focaliza os aspectos vivos, contemporâneos, da língua apreendendo as formas linguísticas no intercâmbio oral*. Obtém uma corrente evolucionária pela comparação das muitas variantes de cada forma, *cuja distribuição no espaço pode ser traduzida numa distribuição através de tempo de acordo com regras metodológicas*. (CÂMARA JR., 1979, p. 102) (grifos meus)

Guardadas as devidas proporções, e considerada a variação não mais no eixo espacial mas no eixo social, a Sociolinguística moderna continua, de certo modo, a proposta da Geografia Linguística, que, com os instrumentos da época, vislumbrou apenas a ponta do *iceberg*. Do mesmo modo que a variação geográfica, a variação sociolinguística reflete a mudança, porque esta não atinge ao mesmo tempo todas as classes sociais, todas as faixas geracionais, enfim todos os grupos construídos na sociedade. Resta, pois, aos linguistas recolher as formas que representam, num tempo determinado, os diferentes estágios da mudança e traçar-lhes a cronologia.

Isso trouxe de volta questões antigas, propostas pelos neogramáticos, sobre o modo como a mudança se processa: se gradual ou abruptamente, se atinge ou não todos os vocábulos do léxico. Qualquer que seja a conclusão de cada linguista sobre a questão de como mudam as línguas, a correlação estabelecida entre a variação observada sincronicamente nas comunidades linguísticas e os processos de desenvolvimento linguístico importou um avanço considerável da linguística histórica.

A ampliação do modelo de gramática do falante, discutida na seção anterior, impõe, portanto, ao linguista uma nova tarefa: descrever o comportamento linguístico variável, mas sistemático, do falante, o que tem reflexos fundamentais na metodologia empregada para acercar-se do objeto de estudo. O primeiro desses reflexos é o abandono da intuição como base para abordar o funcionamento da gramática da comunidade, o que vale dizer, a renúncia da utilização do idioleto

do autor. Essa prática do modelo gerativista mereceu de Labov (1978) a crítica que se lê a seguir:

Ora, por úteis que elas pudessem ser ao teórico que formula suas hipóteses, está claro que as intuições pessoais representam bem o único tipo de dados que não poderia servir de provas, pois ninguém pode estimar precisamente a que ponto elas estão sob influência do desejo universal e compreensivo de demonstrar que se tem razão<sup>16</sup>. (LABOV, 1978, p. 80)

O estudo ora apresentado se beneficiou desses avanços e do progresso que o Estruturalismo, primeiro, e, depois, a Teoria Gerativa imprimiram ao campo da Fonologia, progresso ampliado agora pela Sociolinguística, ao estender o conceito de regra de gramática a regras variáveis. É assim que esta pesquisa deve muito aos estudiosos do passado, filólogos, gramáticos e linguistas, sem o trabalho dos quais essa geração teria de começar do nada. Como disse Labov (1981a, p. 305): “*Insights* se acumulam gradualmente e somente para aqueles que podem ler seus predecessores com compreensão”<sup>17</sup>.

Com a Sociolinguística, a “intuição” do linguista foi substituída pelo trabalho de campo, aperfeiçoando as técnicas de coleta de dados que foram a prática dos estruturalistas e mais, dos dialetólogos. A Sociolinguística deslocou o foco das observações do eixo horizontal – o geográfico – para o eixo vertical dos dialetos sociais. Contando agora com modernos e eficientes aparelhos para a gravação de trechos da fala, pode hoje obter um número representativo de dados de diversas situações de interação linguística e analisar um número considerável deles, com a ajuda de programas computacionais.

Na nova metodologia, o processo de recolha da amostra é visto como condição essencial para conclusões satisfatórias. Alerta Labov (1978, p. 8) que para se obterem conclusões satisfatórias o cuidado deve começar na recolha da amostra: “o fato de se registrar nunca é, bem entendido, sem efeito”<sup>18</sup>. Ao recolher os dados, o linguista é paradoxalmente um falante que participa da conversa e o observador interessado em obter certo tipo de informação. Consciente disso, ele terá de se precaver contra essa situação, que pode interferir nessa recolha,

---

<sup>16</sup> Do original: “Or, pour utiles qu’elles puissent être au théoricien qui formule ses hypothèses, il est clair que les intuitions personnelles représentent bien le seul type de données qui *ne sauraient* servir de preuves, car nul ne peut estimer précisément à quel point elles sont sous l’influence du désir universel, et compréhensible, de démontrer qu’on a raison” (tradução minha).

<sup>17</sup> Do original: “*Insights* accumulate gradually, and only for those who can read the work of their predecessors with understanding” (tradução minha).

<sup>18</sup> Do original: “Le fait d’enregistrer n’est, bien entendu, jamais sans effet” (tradução minha).

produzindo efeitos indesejáveis. A amostra utilizada nesta investigação não foi recolhida com o propósito de recolher um dado específico. Ela faz parte de um grande projeto que visava a estabelecer a norma de falantes cultos de grandes cidades brasileiras (Projeto NURC)<sup>19</sup>.

A segunda consequência metodológica do estudo das regras variáveis é o tratamento quantitativo dos dados, uma vez que se passou a lidar com fenômenos que ora ocorrem, ora não. A princípio foram realizados estudos de frequência da regra, calculando-se sua percentagem de aplicação na presença de um determinado fator. Mas a regra variável é governada por um conjunto de fatores, cujo efeito não é possível de ser isolado com o cálculo de porcentagem, já que em cada ocorrência do fenômeno estudado há sempre a concorrência de vários fatores, aumentando as chances de aplicação ou as diminuindo. Depois de variadas tentativas de quantificação dos dados, Rousseau e Sankoff (1978) conseguiram obter resultados mais satisfatórios que as tentativas antecessoras<sup>20</sup> com o *modelo logístico*, que formou a base para o programa computacional VRBRUL 2, utilizado no tratamento da maior parte dos dados deste trabalho. Esse *software* fornece os resultados em probabilidades de a regra estudada se aplicar na presença de um determinado fator.

Finalmente, o título deste livro tornou-se mais amplo que o da tese que lhe deu origem. Embora o centro da pesquisa continue sendo a descrição de uma amostra do subfalar baiano, a pretensão dele se modificou. Muitas informações sobre outras variedades do português brasileiro surgiram nos últimos anos. Espero que ele seja considerado dentro desse cenário como uma contribuição para formar um quadro completo das pretônicas no Brasil e colocar mais luz sobre a origem dessa diversidade, honrando o legado dos dialetólogos brasileiros que se debruçaram sobre o tema, especialmente Antenor Nascentes, que nos legou a primeira divisão dos dialetos nacionais.

---

<sup>19</sup> O estudo das pretônicas na variedade culta de Salvador conta com uma amostra principal e duas secundárias, que foram constituídas para responder a questões levantadas ao longo desta investigação. A amostra principal é formada de um segmento do *corpus* do Projeto da Norma Urbana Culta de Salvador (NURC-SSA), cuja constituição atende às exigências da pesquisa sociolinguística. Uma das amostras secundárias foi tomada de empréstimo ao *Atlas prévio dos falares baianos* (ROSSI, 1965), cuja *Introdução* fornece dados transcritos foneticamente. A segunda resultou de um levantamento do exemplário do trabalho de Mota (1979) sobre uma localidade sergipana, Ribeirópolis. As razões que me levaram a lançar mão dessas duas amostras complementares e o tratamento que a elas foi dado estão consignados na seção 2.1.1.2.

<sup>20</sup> Para a descrição desses modelos cf. Naro (1980).

