

DEBATE

ENFOQUES SOBRE AMOSTRAGEM EM SOCIOLINGÜÍSTICA

Cecília MOLLICA (Universidade Federal do R.de Janeiro) e
Cláudia RONCARATI (Universidade Federal Fluminense)

ABSTRACT: Sociolinguistic research depends on real data of the language produced in the heart of the speech community. In his analysis the scientist is very often faced with some crucial problems related with the linguistic raw material. In this paper, we describe some properties of random stratified samples, in order to raise some related problems and to offer some methodological alternatives for enriching the researcher's work.

1. Pré-condições da pesquisa sociolingüística¹

A área de pesquisa sociolingüística é marcada pela tônica da indissociabilidade e/ou interdependência entre o analista e os dados emitidos numa situação real de língua falada ou escrita. E tal se aplica a todos os pesquisadores voltados para o uso lingüístico, seja qual for a dimensão do universo de análise ou subsistema enfocado na descrição, seja qual for a metodologia adotada ou a especialização dentro da lingüística.

Discorrer a respeito da matéria-prima de que se lança mão em pesquisa sobre o uso da língua não constitui tópico propriamente inusitado. Na perspectiva da sociolingüística não há como se colocar em dúvida a busca de maior fidelidade à realidade da língua, numa análise voltada para o uso concreto. Todavia, há que se ter unanimidade quanto a limitações de amostras aleatórias estratificadas com parâmetros previamente decididos.

Neste artigo, objetivamos identificar algumas propriedades de amostras aleatórias (cf. Lakatos e Marconi, 1988: 43-6) e alguns problemas delas decorrentes no co-

tidiano do pesquisador. Não esgotamos, aqui, todos os enfoques possíveis sobre amostragem em sociolinguística. Deixamos de lado, por exemplo, questões como tamanho da amostra, coleta e transcrição de dados (cf. Stubbs, 1984: 230-38).

2. Propriedades de amostras aleatórias e suas consequências para a pesquisa

O primeiro aspecto a mencionar diz respeito ao caráter estático de uma amostra aleatória qualquer. Enquanto coleta de um certo estado da língua, uma amostra é uma espécie de congelamento linguístico, interrompendo seu dinamismo inerente, muito embora se possa admitir que o dinamismo do passado deixe vestígios: o futuro da língua já está pré-configurado.

Esse é um ponto que contraria, de certa forma, a visão que o sociolinguista assume em relação à variação e à mudança. Enquanto retrato de um momento da língua, uma amostra de dados é sempre o passado. Corre-se o risco de analisarem-se continuamente etapas linguísticas transcorridas: os dados caducam, insinuando o compromisso do retorno a dados atualizados.

Enquanto visão plana e horizontal da língua, pode-se admitir também que uma amostra aleatória dificulta o diagnóstico de estruturas iniciais e terminais no sistema, ainda que se suspeitem de mecanismos de mudança em curso via tempo aparente (cf. Labov, 1981). Processos de entrada e saída de inovações não são facilmente xerografados numa amostra, a não ser uma amostra tirada na hora certa, e um oportuno exemplo disso é o fenômeno do dequeísmo, senão vejamos.

Até onde foi possível investigar em Mollica (1989), casos de dequeísmo como:

"Essa viagem, ela foi muito importante para o partido dos trabalhadores e foi muito importante para mim, porque pudemos descobrir, constatar, ver e

sentir de que todos nós da América Latina somos vítimas de uns inimigos." (Político, Entrevista Debate em Manchete, agosto 1988),

já habitavam o português desde os anos 60. Possivelmente, esse é um marco passível de ser abalado. No entanto, as amostras existentes e disponíveis no Brasil, confeccionadas por pesquisadores da área do uso, ou contêm o dequeísmo numa taxa tão baixa quanto imperceptível, como é o caso das amostras do projeto NURC, ou o possuem atípica e esporadicamente quase ao sabor de idiosincrasias dos falantes, como é o caso da amostra CENSO.

Note-se que o contexto do dequeísmo típico se configura por uma estrutura de sentença complexa em que há pelo menos uma matriz e uma sentença encaixada que, normalmente, é uma oração substantiva objetiva direta ou uma substantiva subjetiva. Há outros tipos de estruturas dequeístas, inclusive casos em cláusulas relativas, mas as mencionadas são as mais frequentes (cf. Bentivoglio, 1976; Boretta de Macchia, 1988). Tais contextos ocorrem em baixíssima taxa nos *corpora* do projeto NURC e, simplesmente, inexistem na amostra CENSO. Neste acervo, há casos de dequeísmos menos típicos que, enquanto raros, aparecem em locuções conjuntivas como "*na medida de que*", considerado um tipo de dequeísmo por substituição.

O dequeísmo é marcado pragmaticamente. Ele emerge em contexto de fala formal e de alta tensão discursiva, haja vista o fato de ocorrer quase exclusivamente em entrevistas na mídia, em conferências, ou em situações de fala similares. Eis um bom motivo por que a sua frequência é quase nula na amostra CENSO, composta de enunciados de fala semi-espontâneos. Talvez a explicação para aparecer mais nos dados do NURC repouse no fato de haver nesse acervo um conjunto de entrevistas, denominadas de elocução formal, de discurso planejado.

Eis, em síntese, a questão que se pretende destacar a respeito de propriedades e limites de uma amostra aleatória estratificada. Teoricamente, acredita-se que

não haja variação aleatória e, desta feita, qualquer uso é motivado por variáveis internas e externas ao sistema. Ora, se uma amostra, por definição, constrói-se a partir de parâmetros cruciais como condição indispensável à existência de certa estrutura, e se esta estrutura não apresenta esses parâmetros, o resultado será a inexistência da estrutura buscada. Caso a estrutura que se busque apareça provavelmente apresentará uma taxa de ocorrência tão baixa a ponto de tornar-se imperceptível e/ou de não dar margem para ser analisada com os instrumentos metodológicos normalmente adotados. Isso, todavia, não invalida o fato de que a própria amostra esteja refletindo fielmente a realidade da língua, onde a estrutura é rara e usada apenas sob circunstâncias específicas. Em casos desse tipo, será necessário, por conseguinte, constituir uma amostra em que as circunstâncias desejadas se encontrem concentradas (cf. Lakatos e Marconi, *op. cit.* p.47). Portanto, para buscar estruturas raras e/ou controladas por parâmetros cruciais, há que se optar pela utilização de amostras intencionalmente elicítadas.

Servem também de exemplo para este ponto de reflexão as estruturas condicionais do português, estudadas por Gryner (1990). Em razão de serem marcadas discursivo-pragmaticamente, elas não emergem em amostras de fala semi-espontânea e sem um tipo de técnica especial de elicitação, fato que reafirma a abrangência e suficiência relativas de uma amostra aleatória em face da comunidade de fala que se deseja ver refletida.

Esses casos ilustram a possibilidade da instauração de um paradoxo que pode dificultar ou até mesmo obstaculizar a desejada adequação entre os postulados sociolinguísticos e a metodologia de que se servem. O diagnóstico da sistematicidade da variação acha-se apensado aos próprios dados, não por eles inexistirem na língua, mas em razão da eventualidade da amostra.

Outro ponto a ser relevado diz respeito à limitação natural dos dados de fala. Por exemplo, a fala mais casual é a mais difícil de ser obtida. Conforme Stubbs (op.cit. p.234), sublinha, é mais susceptível de contaminação por observação, seja pelo efeito do estranhamento e distanciamento entre o informante e a naturalidade da situação, seja, como justifica Wolfson (1976, apud Stubbs (1974: 234), pela própria impossibilidade de se obter um registro puro e autêntico da fala natural em um sentido absoluto. Os usos se ajustam a diferentes situações ...

Grosso modo, qualquer amostra fotografa "ângulos" da língua, pois o espectro de possibilidades comunicacionais do falante é muito amplo. Os dados fornecem necessariamente um recorte da língua, constituem uma instância verbal em que, ao lado de fatores lingüísticos, imbricam-se fatores históricos, sociais e pragmáticos.

3. Alternativas do pesquisador

Em face das características mencionadas e de seus conseqüentes problemas, o pesquisador tem diante de si algumas alternativas. A primeira delas consiste em assumir previamente as limitações operacionais da pesquisa e atingir metas menos ambiciosas, tais como: (a) aceitar resultados referentes estritamente ao universo analisado, conformando-se com um grau menor de generalização; (b) formular interpretações a nível de hipóteses. A segunda opção para o pesquisador é a de não assumir as limitações operacionais da pesquisa e tentar transpô-las usando algumas estratégias metodológicas.

Trabalhar com pluralidade de amostras constitui uma alternativa a ser praticada em decorrência do direcionamento assumida por uma dada análise. Quando se buscam fenômenos característicos da fala mais espontânea, o uso de amostras tipologicamente diversificadas torna-se mais produtivo. Em amostras marcadas por uma relação de assimetria de papéis interacionais, o mapeamento de

perguntas que suscitam relatos de vida pessoal e de vida em comunidade registre e até certo ponto artificializa os eventos da interação verbal. Diante da necessidade de obter o máximo de produção linguística, o entrevistador não atua como interlocutor ativo, antes registre sua participação a breves intervenções que se destinam à cooperação e estímulo convencionais e/ou à mudança de tópico. Por sua vez, o informante aceitando as regras da entrevista, dispõe de liberdade relativa para considerar ou não-considerar as solicitações que lhe são feitas. (cf. Paiva, 1989; Bortoni-Ricardo, 1984; Lavandera, 1985).

Uma motivação natural para se advogar a pluralidade de amostras respalda-se na riqueza em comparar amostras diversificadas. Roncarati *et alii* (1988) só puderam captar o comportamento sistemático das variáveis estudadas correlacionadas ao uso de negativas do tipo: "Sei não" e "Quero não", porque se valeram tanto de entrevistas semi-informais, quanto de situações de interação médico-paciente e de conversas espontâneas. Da mesma forma, só foi possível obter-se certeza quanto à existência do dequeísmo no português em Mollica (op. cit.) em razão do trabalho de rastreamento de dados em inúmeras amostras.

Há que se levar em conta ainda as dificuldades de se obterem estilos diferentes em amostras aleatórias. Para contornar o problema, Labov (op.cit.: 9-10 e 70-109) recorre a outros métodos de elicitación de dados, posteriormente aplicados em diferentes pesquisas, a exemplo de Oliveira e Silva (1982). Mollica e Matos (1989), ao estudarem a assimilação mb b, sublinham que a amostra CENSO não favorece a testagem do grau de formalidade, tensão e monitoração do discurso.

Neste artigo, sustentamos, então, que a utilização de amostras pluridimensionadas e a elicitación provocada de enunciados instrumentam o pesquisador com recursos que lhe possibilitam contornar deficiências inerentes às amostras com as características aqui apontadas. Uma amos

tra nem vai fotografar a complexidade lingüística nem vai prever as possibilidades de análise a respeito da língua. Ela estará sempre a aquêm da inesgotável curiosidade científica e dificilmente dará conta da acentuada margem de imprevisibilidade dos fatos da língua.

NOTA:

1. Este texto é uma versão revista da comunicação apresentada no V Encontro Nacional da ANPOLL, realizado em Recife, em julho de 1990. Agradecemos a leitura e comentários dos professores que compõem o PEUL, e em especial ao professor Anthony Julius Naro, pelas sugestões e críticas.

REFERÊNCIAS BIBLIOGRÁFICAS

- BENTIVOGLIO, P. (1976) Queísmo y dequeísmo en el habla culta de Caracas. IN: AID, F. & RESNICH, M.C. (orgs.), (1975) *Colloquium on Hispanic Linguistics*. Washington: Georgetown University Press, p. 1-18.
- BORETTI DE MACCHIA, S. (1988) (De)queísmo en el habla culta de Rosário. Rosário: Universidade Nacional de Rosário, Consejo de Investigaciones. 29 p. Mimeo.
- BORTONI-RICARDO, S.M. (1984) Problemas de comunicação interdialetoal. *Tempo Brasileiro*, 79: 9-32.
- GRYNER, H. (1990) *A variação de tempo-modo e conexão nas condições em português*. Tese de doutorado, UFRJ: Faculdade de Letras. 550 p. Mimeo.
- LABOV, W. (1981) What can be learned about change in progress from synchronic descriptions? IN: SANKOFF, D. & CEDERGREEN, H. (orgs.) *Variation omnibus*. Canada: Linguistic Research Inc., p. 177-99.
- _____. (1972) *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press. 43-59.
- LAKATOS, E.M. & MARCONI, M.A. (1988) *Técnica de pesquisa*. São Paulo, Atlas.

- LAVANDERA, B.R. (1985) *Curso de lingüística para el análisis del discurso*. Buenos Aires: Centro Editor de América Latina.
- MOLLICA, M.C.M. (1988) *Queísmo e dequeísmo no português do Brasil*. Tese de Doutorado, UFRJ: Faculdade de Letras: 308 p. Mimeo.
- _____ & MATTOS, P.B. (1989b) Dois processos de assimilação fonológica no português falado semi-espontâneo do Rio de Janeiro. IN: *Relatório final FINEP do Projeto Mecanismos funcionais do uso lingüístico*. Rio de Janeiro: UFRJ p. 524-559.
- OLIVEIRA SILVA, G.M. & MACEDO, A.V.T. (1989) Análise Sociolingüística de alguns marcadores conversacionais. IN: *Relatório final FINEP do projeto Mecanismos funcionais do uso lingüístico*. Rio de Janeiro: UFRJ. p.127-42.
- PAIVA, M.C.A. (1989) Codificação da relação causa-consequência. IN: *Relatório final FINEP do Projeto Mecanismos Funcionais do uso lingüístico*. Rio de Janeiro: UFRJ
- RONCARATI, C.N.S. et alii (1988) As negativas na fala cearense. IN: *Relatório final FINEP do Projeto Dialetos sociais cearenses*. Vol II, Fortaleza: UFC. p.1/175.
- STUBBS, M. (1984) Collecting conversational data: notes on sociolinguistics methodology. IN: STUBBS, M. *Discourse analysis*. England: Blackwell. p. 218-246.