

A IDENTIDADE LINGÜÍSTICA DA COMUNIDADE DE FALA: PARALELISMO INTERDIALETAL NOS PADRÕES DE VARIAÇÃO LINGÜÍSTICA

Gregory R. Guy

RESUMO: *The basic sociolinguistic model for relationships between idiolects and dialects is the speech community, defined by shared linguistic features and attitudes and relatively high internal density of communication. Since these definitions are relative, speech communities can nest and overlap, so that local subdialects, sociolects, ethnic lects, and personal networks can form smaller speech communities that share local traits and are locally high in communication density, while still belonging to a broader speech community which shares wider traits, and whose internal communication density is high, relative to, say, other geographic regions. In variation studies, it has long been assumed that some of the shared linguistic traits that define a community are certain constraints on variable processes. The conceptual problem that arises with this model is, how far are constraints shared, and how much can they differ? If some constraint was due, for example, to a universal process, then it would be expected to be shared by all speakers of all speech communities, while at the other extreme, the existence of idiosyncratic differences in language usage raises the possibility that at least some constraints may differ for each individual. This paper takes a cross-dialectal comparative approach. Two variable processes are studied in four communities drawn from the VARSUL corpus, each with distinctive ethnic and sociolinguistic characteristics. A socially diversified sample of 8-12 speakers is investigated in each community. The variables investigated include one syntactic process (noun phrase Agreement, NPA) and one phonological process (final -s deletion, SDEL). The constraints on NPA are mainly morphosyntactic in nature while the constraints on SDEL are straightforwardly phonological. In each case, the constraint effects are broadly similar across communities and speakers. Between-speaker differences within communities are mainly the result of either statistical noise (smaller sample sizes lead to larger differences), or of predictable social differentiation (e.g. speakers with less formal education use more of the*

Gregory R. Guy é professor na York University e na New York University.
Artigo traduzido por Leonardo Z. Maya (bolsista voluntário de Iniciação Científica 98-2000)

nonstandard variants.) And, strikingly, the main constraint effects are highly consistent across the different communities. The results generally lend support to the model of Cedergren & Sankoff (1974), that "performance is a statistical reflection of competence", and competence, here dealing with the knowledge of what varies where, is powerfully shared across a language community.

PALAVRAS-CHAVE: comunidade de fala, concordância nominal, apagamento de /s/

Para tentar entender a organização social da língua, grande parte da pesquisa sociolingüística usa uma estrutura básica de referência para além do falante individual, entendido como a base em relação à qual cada idioleto é delimitado. Essa unidade social é a comunidade de fala, que tem duas funções na teoria sociolingüística. Fornece, em primeiro lugar, uma base fundamentada para explicar a distribuição social de semelhanças e diferenças lingüísticas, a razão por que certos grupos de falantes compartilham traços lingüísticos que os distinguem de outros grupos de falantes. Em segundo lugar, a noção de comunidade de fala fornece uma justificativa teórica para unir os idioletos de falantes individuais (que são os únicos objetos lingüísticos cuja existência se pode realmente observar), em objetos maiores, as línguas (que são, na verdade, construções abstratas). Consideremos esses dois pontos separadamente.

Primeiro, tratamos da comunidade de fala como um modelo explicativo de semelhanças e diferenças no uso da língua. Embora haja várias definições de comunidade de fala na literatura sociolingüística, podemos identificar um grupo comum de características sobre o qual parece haver consenso. Essas características estão resumidas em (1):

(1) *Definição de comunidade de fala*

- características lingüísticas compartilhadas; isto é, palavras, sons ou construções gramaticais que são usados na comunidade, mas não o são fora dela.
- densidade de comunicação interna relativamente alta; isto é, as pessoas normalmente falam com mais frequência com outras que estão dentro do grupo do que com aquelas que estão fora dele.
- normas compartilhadas; isto é, atitudes em comum sobre o uso da língua, normas em comum sobre a direção da variação estilística, avaliações sociais em comum sobre variáveis lingüísticas.

É a primeira dessas características que organiza as semelhanças e as diferenças lingüísticas no uso da língua. A participação como membro em uma comunidade de fala é definida por contraste, em função do uso de traços específicos da comunidade: usá-los mostra que você é um membro, e não os usar mostra que você é um intruso. A participação como membro da comunidade de

fala da Filadélfia é, dessa maneira, definida em parte pela centralização do /ay/ em *fight*, pela fusão de *sure* e *shore*, pela vocalização do /l/ em *dollar*, pelo uso de *anymore* em construções afirmativas (*anymore, there's a lot of stuff to learn*, por exemplo), e pelo complexo padrão de tensionamento do /a/ (com pronúncia tensa em *man, half, mad*, mas não-tensa em *manner, ran, dad*). A participação na comunidade de fala de Toronto, por outro lado, é definida, entre outras coisas, pela centralização do /aw/ em *house*, pela fusão de *cot* e *caught*, por itens lexicais como *zed* e *washroom*, e pela mudança em cadeia das vogais frontais não-tensas, quando o /a/ é retrocedido e o /i/ e o /e/ são rebaixados.

Esse compartilhamento de características lingüísticas em uma comunidade de fala inclui as restrições de processos de variação. Em Guy (1980), mostrei, por exemplo, que uma das diferenças fonológicas entre Nova Iorque e a Filadélfia é o efeito do contexto *pausa seguinte* no apagamento de oclusivas coronais finais. Na Filadélfia, em palavras como *west* e *bold* em posições anteriores à pausa, as oclusivas finais são preferencialmente mantidas, ao passo que em Nova Iorque elas são preferencialmente apagadas (*wes', bol'*). Trabalho subsequente mostrou que outros dialetos adotam uma escala de valores para o efeito desse contexto no processo. Salienta-se que, apesar de o valor do contexto diferir de uma comunidade de fala para outra, dentro de uma mesma comunidade os falantes são extremamente coerentes. Esse é, então, um gesto lingüístico compartilhado por todos os falantes em uma comunidade, como todos os nova-iorquinos compartilham *standing ON line*, e todos os porto-riquenhos aspiram seus /s/ na coda silábica.

Outro exemplo de restrição variável que define e distingue comunidades de fala pode ser encontrado no trabalho de Cameron (1994) sobre a expressão do pronome reto em espanhol. Em Porto Rico, o *tu* genérico favorece o preenchimento, mas na Espanha ocorre o oposto. Essa restrição, de fato, parece separar o espanhol latino-americano de todos os dialetos europeus: deste lado do Atlântico é provável que o *tu* genérico seja dito, enquanto do outro lado do oceano é provável que ele seja apagado. Mesmo nesse caso, um processo sintático variável, encontramos, então, valores característicos de restrições contextuais que ajudam a definir uma comunidade de fala.

Finalmente, apresento um terceiro exemplo, tirado do trabalho de Labov (1973) sobre o apagamento da cópula em AAVE¹. A observação de adolescentes que eram membros de grupos semi-organizados, e a comparação destes com os que, além de não serem membros, eram socialmente isolados na comunidade – os *lames* em inglês –, revelaram que ambos os conjuntos de garotos usavam o apagamento da cópula. Porém, o que distinguia os membros do grupo dos marginalizados era que estes não apresentavam as condições contextuais corretas

¹ A sigla significa *African American Vernacular English*. Trata-se da variedade usada por muitos negros americanos, chamada antigamente de BEV (Black English Vernacular) ou simplesmente Black English. AAVE é a denominação mais usada atualmente.

de apagamento: eles não variavam a taxa de apagamento de acordo com o contexto sintático seguinte. O domínio desse detalhe mínimo dependia da participação verbal ativa na minicomunidade de fala local, que nesse caso era constituída das redes de companheiros de grupo em que os adolescentes passavam grande parte de seu tempo.

Entretanto, há um ponto crucial sobre as restrições lingüísticas na variação que deve ser considerado. Muitas, ou de fato, a maioria dessas restrições parece ter efeitos bem gerais, ou mesmo universais. Assim, todos os dialetos do inglês mostram que o apagamento da oclusiva coronal é favorecido quando esta é seguida por consoante, e o mesmo tipo de efeito é encontrado no processo de apagamento de segmento final em outras línguas. Isso se deve, sem dúvida, a um universal fonológico: a preferência universal por seqüências CV a seqüências CC. Se uma restrição em variação é devida a universais estruturais, obviamente, não deveriam existir diferenças entre comunidades. Contudo, a partir de evidências como o efeito da pausa no apagamento da oclusiva coronal em inglês e o efeito do *tu* genérico no apagamento do pronome em espanhol, parece claro que nem todas as restrições são universais, e aquelas que não o são, podem ser definidas localmente para cada comunidade.

Voltando agora às características definidoras de comunidade de fala, apresentadas em (1), vimos exemplos de traços lingüísticos compartilhados que ajudam a identificar contrastivamente os membros de uma comunidade. A pergunta que realmente quero fazer é: POR QUE todos os membros de uma comunidade deveriam fazer a mesma coisa, e os não-membros não? A meu ver, o modelo de comunidade de fala tem algum valor explicativo para dar conta da distribuição social de semelhanças e diferenças no uso da língua. Onde está a explicação? A resposta reside nos outros dois elementos da definição em (1): densidade de comunicação e normas compartilhadas. Uma breve reflexão mostrará que o que explica traços lingüísticos compartilhados são esses dois elementos. Consideremos, primeiro, a densidade de comunicação. Uma densidade de comunicação relativamente alta em um grupo significa que os falantes têm mais acesso e exposição aos usos lingüísticos de outros membros do grupo. Logo, podem mais prontamente ouvir o que outros membros da comunidade estão fazendo com a linguagem, o que cria a possibilidade de adquirir deles certos traços lingüísticos. Em contraste, o nível relativamente baixo de comunicação com não-membros torna menos provável que se adquiram usos de fora da comunidade de fala. Em outras palavras, a comunidade de fala é o domínio no qual os processos sociolingüísticos de acomodação e convergência ocorrerão. Dito sucintamente, fala-se como as pessoas COM as quais se fala.

A mera exposição a um traço lingüístico, todavia, não é suficiente para propiciar a acomodação e a aquisição do mesmo. Os falantes de cidades fronteiriças do Canadá como Windsor e St. Catherine's têm acesso imediato e maciço às comunidades de fala americanas de Detroit e Buffalo, muito maiores e mais poderosas economicamente, mas não mostram sinais de abandonar a maneira

tipicamente canadense de falar. De modo semelhante, os falantes de Patterson, Nova Jérsei, e do leste de Long Island são maciçamente expostos ao inglês da cidade de Nova Iorque, mas fogem dos traços lingüísticos característicos da *Big Apple* como o diabo foge da cruz. Logo, devemos concluir que há também a questão das atitudes e vontades: os falantes QUEREM se acomodar a outros?

É aqui que entra a terceira característica definidora de comunidade de fala: os membros de uma comunidade compartilham normas e atitudes em comum sobre o uso da língua: o que é apropriado para contextos formais e o que é apropriado para os informais, que taxa de uso de uma variável sociolingüística é apropriado para que grupo social, etc. Uma das coisas que define um nova-iorquino é a interpretação de pronúncias sem *r* como indicativas de classe trabalhadora, estilo casual e forma não-padrão, enquanto pronúncias com *r* denotam classe média-alta, estilo cuidado e forma padrão. Do mesmo modo, uma característica que define um falante porto-riquenho é saber que pronunciar o *-s* final tem prestígio e denota formalidade, enquanto a aspiração e o apagamento do *-s* final indicam familiaridade e informalidade. No entanto, todas essas interpretações e essas normas de comportamento são estritamente locais, conhecidas apenas pelos membros dessas comunidades, não sendo, certamente, compartilhadas pelos moradores de Londres, Lima, Lisboa ou Los Angeles.

Mas o que dizer sobre a segunda função teórica da comunidade de fala, a de modelar a conexão entre idioleto e língua, entre o indivíduo e o grupo, entre o observável concretamente e a hipótese abstrata, entre, resumidamente, *parole* e *langue*. A definição de comunidade de fala dada acima implica certos aspectos sobre a diversidade dos idioletos. Primeiro, o modelo de comunidade de fala postula que os falantes não variam aleatoriamente uns dos outros, ou ao menos não variam aleatoriamente em todos os tipos de variáveis. A participação em uma comunidade e o processo de acomodação garantem que muitos traços lingüísticos serão mantidos em comum nessa comunidade. Segundo, o modelo de comunidade fornece uma base social e externa para investigar quem compartilha que traço com quem. Na falta de tal modelo, a busca empírica de generalizações inter-idioletais, a busca de objetos como 'línguas' ou dialetos, bem poderia começar por comparar o meu falar com o de alguém de Birmingham, Berlim ou Bangcoc.

Finalmente, o modelo de comunidade de fala implica graus de semelhança e diferença lingüísticas, bem como distribuições concentradas, cruzadas ou sobrepostas de traços compartilhados por falantes. Assim, uma comunidade de fala local, com características localmente distintas, pode, no próximo nível ascendente, compartilhar características dialetais regionais, a seguir, pertencer a uma comunidade de fala nacional mais ampla e, finalmente, no nível mais alto, participar de uma comunidade internacional de falantes de uma mesma língua. Em tal caso, as comunidades estariam encaixadas umas dentro das outras como bonecas russas, e os falantes compartilhariam o maior número de características com seus vizinhos imediatos, seguidos por compatriotas regionais e nacionais, terminando com um mínimo de características compartilhadas com outros falantes

da mesma língua onde quer que se encontrassem no globo terrestre. Outras comunidades, porém, podem estar interligadas: subcomunidades definidas por vizinhança e redes comunicativas pessoais, por classe social, etnia, religião, ocupação, etc., uniriam falantes a várias comunidades sobrepostas e cruzadas.

A esse respeito, a questão para os variacionistas é como os efeitos das restrições variáveis se encaixam nisso? Podemos discernir algum tipo de padrão significativo ou interpretável nos pesos relativos dos fatores para diferentes restrições em processos variáveis? Por exemplo, há alguma significância na oposição entre fatores que estão em *knock-out* e os que são variáveis, ou então, quais restrições favorecem e quais desfavorecem um processo? A ordem relativa das restrições (X favorece o apagamento mais do que o Y) é a única coisa importante, ou há significância nos valores absolutos dos pesos relativos, ou nas diferenças escalares entre eles? Tais questões são na maioria abertas e empíricas, mas algumas linhas preliminares de indagação podem ser úteis. Primeiro, as restrições universais devem, por isso mesmo, ser universais e não admitir diferença entre comunidades ou entre falantes. Segundo, os efeitos lexicais devem ser altamente locais e assistemáticos, dada a axiomática “arbitrariedade do signo lingüístico”. Terceiro, então, os resultados mais interessantes talvez sejam aqueles que envolvam diferenças não-universais, mas estruturais (não-lexicais) entre comunidades. A busca de padrões de distribuição dessas diferenças pode ser uma linha de investigação produtiva, e esse é o meu propósito com esta pesquisa.

Olhemos agora alguns dados. Meu projeto de pesquisa atual, financiado pelo Conselho de Pesquisa das Ciências Sociais e Humanísticas do Canadá, trata de investigar a estrutura dessa relação universal/particular, *langue/parole* em um nível quantitativo. A principal fonte dos dados que estou reportando aqui é o banco de dados sociolingüístico do português brasileiro compilado pelo Núcleo VARSUL, em quatro universidades do sul do Brasil, as universidades federais do RS, SC e PR, e pela PUC-RS. Gostaria de expressar meu agradecimento ao VARSUL como um todo e, individualmente, a Leda Bisol, Paulino Vandresen, Edair Gorski e Ana Zilles por me envolverem com o projeto e seus dados, por sua generosidade em me dar acesso ao banco de dados e por toda a ajuda com o meu projeto nos últimos anos.

O *corpus* do VARSUL inclui entrevistas sociolingüísticas com amostras estratificadas de no mínimo 24 falantes em cada uma de 12 cidades localizadas nos três estados do Sul do Brasil. O *corpus* para cada cidade está igualmente distribuído por idade, sexo e nível de escolaridade. Algumas das cidades envolvidas têm histórias importantes de línguas em contato, estando localizadas em áreas colonizadas no século passado por imigrantes da Itália, Alemanha, Polônia e Ucrânia. Há uma cidade que está localizada na fronteira com a Argentina e duas outras que foram fundadas por imigrantes dos Açores.

Deste *corpus*, retirei um subgrupo de sete cidades dos dois estados do extremo sul brasileiro, com 8 a 12 falantes por cidade. Essa amostra possibilitará a investigação de vários aspectos dos padrões de encaixamento e sobreposição das

comunidades de fala por meio do estudo dos efeitos de várias restrições sobre processos variáveis. Poderemos investigar questões como as seguintes:

- até que ponto as várias cidades são similares em restrições variáveis?
- até que ponto as diferenças entre as cidades são devidas a fatores sociais como proximidade, região dialetal, constituição étnica, história de línguas em contato, etc.?
- até que ponto diferenças entre cidades podem ser atribuídas a diferenças de gramática, em oposição a diferenças meramente estatísticas, ou a flutuação aleatória da amostra?
- até que ponto os falantes de uma cidade são semelhantes entre si?
- até que ponto as diferenças entre falantes de uma cidade são atribuíveis a fatores sociais como idade, escolaridade e gênero?
- até que ponto diferenças entre falantes podem ser atribuídas a diferenças de gramática, em oposição a diferenças meramente estatísticas, ou a flutuação aleatória da amostra?
- finalmente, que efeitos restritivos podem ser atribuídos a fatores lingüísticos gerais ou universais, e quais devem ser vistos como dialetalmente arbitrários ou lexicalmente específicos?

Essa é uma empreitada ambiciosa e não será relatada em sua totalidade aqui. O foco estará nas diferenças entre cidades, e estarei considerando apenas quatro das sete cidades das quais possuo dados, todas no estado do RS: Porto Alegre (a capital estadual e maior metrópole), Panambi (em área de colonização alemã), Flores da Cunha (colonização italiana) e São Borja, na fronteira com a Argentina. Nos gráficos que seguem, elas aparecem abreviadas como POA, PAN, FLC e SBO, respectivamente. As demais questões levantadas acima, relacionadas com diferenças sociais e lingüísticas em cada cidade, serão tratadas em trabalhos futuros.

A fim de maximizar a abrangência dos fenômenos variáveis, estamos investigando duas variáveis sociolingüísticas, uma fonológica e uma morfossintática. O processo fonológico é o apagamento variável do *-s* final como, por exemplo, *menos ~ meno*. O processo morfossintático é a marca variável de plural no sintagma nominal como, por exemplo, *as nações unidas ~ as nação unida*. Os dois processos estão intimamente conectados no caso de plurais regulares, em que a única marca de pluralidade é o sufixo *-s*. Logo, em um caso como *as casa*, pode ser impossível dizer se o *-s* está ausente no final de *casa* porque ele não foi marcado pela regra sintática de marcação de plural, ou porque ele foi apagado pela regra fonológica de apagamento do *-s*. Há, no entanto, casos suficientes de plurais irregulares, como *nação ~ nações*, que nos permitem distinguir as duas derivações. Marcação de plural seguida por apagamento de *-s* resulta em *naçãoe*, e não marcação fornece *nação*. Ambos os casos ocorrem, de modo que podemos estar confiantes de que ambos os processos variáveis existem independentemente.

Minha abordagem será comparar resultados para as principais restrições lingüísticas nesses dois processos para as quatro cidades de POA, PAN, FLC e SBO, buscando similaridades, diferenças e generalizações. Quão semelhantes em processos variáveis são essas diferentes regiões de um mesmo estado?

Primeiramente, consideremos as restrições no processo fonológico, o apagamento do *-s*. Incluirei apenas aquelas palavras cujo *-s* final não é um marcador de plural, porque, como observado, a marcação de plural também é variável. Logo, estamos analisando apenas os *-s* finais provenientes de entradas lexicais. As principais restrições nesse processo são fonológicas: a tonicidade da sílaba e a natureza do segmento seguinte. Há também uma restrição morfolexical: palavras puramente monomorfêmicas, como *mais*, *menos*, *ônibus*, apresentam índices menores de apagamento do que formas verbais de primeira pessoa do plural, as quais são regularmente flexionadas com o sufixo *-mos* (por exemplo, *falamos*; em oposição a *falo* e *fala*.)

Como são os resultados das cidades aqui consideradas? Para esses dados, as taxas de apagamento são baixas, de modo que vários itens do *corpus* apresentaram 0% de apagamento, tornando-se *knock-outs* negativos. Não efetuamos, portanto, análises com o Varbrul, e apresentaremos apenas os percentuais de apagamento para as variáveis fonológicas.

Olhemos primeiro o efeito métrico. O principal efeito aqui é o acento: outros estudos do português brasileiro mostraram que é muito mais provável que sílabas átonas sofram apagamento do *-s* do que as tônicas. E quanto aos monossílabos? A maioria das palavras monossilábicas recebe acento lexical primário, mas algumas, como a conjunção *mas*, podem não estar acentuadas no discurso. Conseqüentemente, os monossílabos foram mantidos separados, embora esperássemos que fossem acentuados na sua maioria.

Os resultados das quatro cidades estão na Figura 1. Os efeitos principais são bem coerentes: para todas as cidades, sílabas átonas sempre têm uma taxa de apagamento significativamente maior, e nossas dúvidas sobre os monossílabos estão resolvidas, uma vez que não são significativamente diferentes dos polissílabos acentuados. A surpreendente coincidência de valores aqui sugere uma gramática compartilhada, comum a todas as cidades da região. A explicação final desses resultados poderia ser um universal fonológico: sílabas acentuadas são mais salientes e menos suscetíveis a vários tipos de reduções. O tratamento sistemático de monossílabos mostra que qualquer redução do acento frasal dessas palavras é leve, e sem muito efeito no apagamento do *-s*.

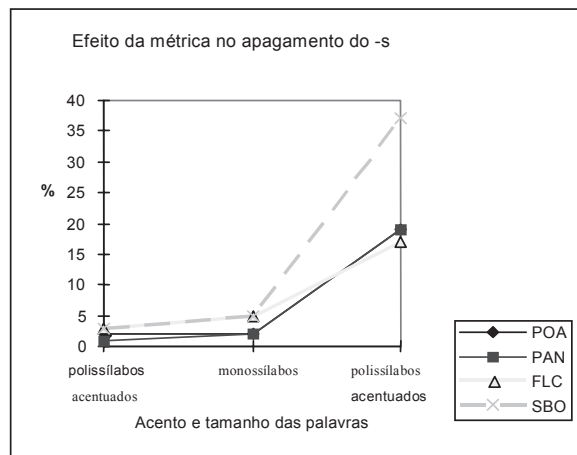


Fig.1

Há, no entanto, um ponto divergente no gráfico: a cidade de SBO tem uma taxa notavelmente maior de apagamento em polissílabos não acentuados, e a relação entre esse contexto e os outros dois é maior do que nas outras cidades, como pode ser visto pela inclinação mais acentuada da linha. Isso implica que essa cidade deveria ter, para essa restrição, um peso relativo diferente e maior do que o das outras cidades? Na verdade, eu penso que não, mas a razão para isso requer que primeiro consideremos uma outra restrição, morfolexical, que compara o apagamento do *-s* final no sufixo verbal *-mos* com outros *-s* monomorfêmicos.

Os resultados para esse grupo de fatores estão ilustrados no Figura 2. Aqui, novamente, os resultados são bem coerentes para as quatro comunidades, com maior taxa de apagamento nos casos de *-s* verbal. Mais uma vez, a única diferença notável é que SBO tem uma taxa muito alta de apagamento na categoria verbal, aproximadamente o dobro da taxa das outras três cidades. Isso é compatível com os resultados para o grupo de fatores tonicidade, porque o sufixo verbal *-mos* é sempre uma sílaba átona em uma palavra polissilábica.

Conseqüentemente, todas as ocorrências de *-mos* recaem na categoria de polissílabos átonos na variável tonicidade. Tendo em mente que estamos lidando com percentuais, não pesos relativos do programa Varbrul, teríamos que nos perguntar qual é o efeito principal aqui: SBO está apresentando esses apagamentos porque a sílaba é átona ou porque são sufixos verbais? Cruzamentos sugerem fortemente a segunda opção. Separando-se os casos de verbos dos dados de não-verbos não-acentuados, os índices de apagamento para não-verbos não-acentuados são de 3, 2, 6 e 7% para POA, PAN, FLC e SBO, respectivamente, todos mais altos do que os índices das categorias acentuadas, mas muito inferiores aos casos de *-s* verbal. É, então, no tratamento desse sufixo verbal específico *-mos* que devemos buscar uma explicação de por que SBO é diferente das outras cidades.

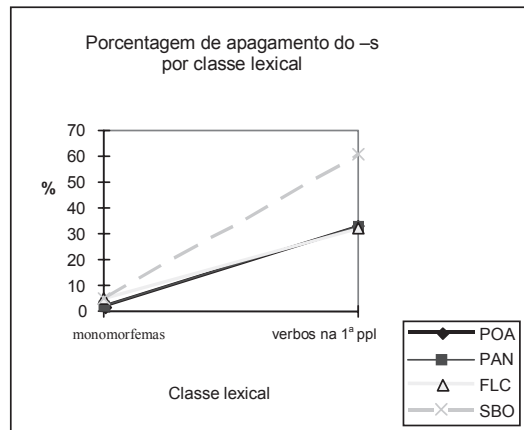


Fig.2

Uma resposta para essa pergunta depende de um breve desvio da linha principal do meu trabalho, para considerar como essa restrição específica, ou restrições similares que têm aparecido em outros estudos, devem ser representadas em uma gramática variacionista. O *-s* verbal que é preferencialmente apagado aqui não carrega nenhuma carga funcional especial; é apenas parte de um afixo maior *-mos*; seu apagamento, portanto, não afeta a categoria número-pessoal do verbo: *falamo* ou *falamos* ainda são identificáveis como significando *nós falamos*.

Esse apagamento seria como apagar o *-s* final do afixo *-ness* do inglês. Um condicionamento funcional, portanto, não pode dar conta dessa restrição. Tampouco pode ser tratada por uma restrição de fronteira morfológica: o *-s* não é imediatamente precedido por uma fronteira morfológica. Então, esse caso não pode ser tratado da maneira usada para modelar casos como o baixo índice de apagamento de *-t, d* nas formas verbais do tempo passado em inglês. No inglês, podemos dizer efetivamente que a regra de apagamento de *-t, d* não se aplica ou tem uma taxa reduzida de aplicação no contexto **#_#*. Aqui, porém, teria pouco valor explicativo dizer que a regra em português aplica-se mais no contexto *#mo_#*. Quais são as alternativas então? Para mim parece mais plausível tratar tais casos como variáveis lexicais. Postularíamos que o sufixo *-mos* tem duas entradas lexicais, uma com e outra sem o *-s* final, e a taxa mais elevada de ausência do *-s* na superfície dessas palavras é devida à seleção lexical variável da forma *-mo* desse sufixo. Acredito que mecanismos similares seriam necessários para explicar as taxas preferencialmente altas de apagamento de *-t, d* na palavra *and* em inglês, e de *-s* no *entonces* do espanhol argentino, entre outros casos semelhantes. Se isso for correto, e tais casos realmente indicarem uma forma de variação lexical, então podemos prever que eles variarão mais de comunidade para comunidade. Isso, então, nos dá um princípio para explicar por que SBO diverge nesse fator, mas não em outros. Os itens lexicais são por definição arbitrários, da

mesma forma que o seu tratamento variável também é arbitrário e, portanto, menos sistemático, mais mutável do que o de outros elementos de estrutura linguística.

Consideremos agora o efeito de segmento seguinte, apresentado na Figura 3. Parece, à primeira vista, que há aqui maior variabilidade entre as cidades. Um exame mais cuidadoso, entretanto, revela que há mais similaridade aqui do que pode parecer. Muito da disparidade dos resultados parece ser devida à flutuação fortuita em pequenas amostras. Os valores em todas as cidades agrupam-se firmemente para certos segmentos: vogais seguintes (com ou sem acento), pausa e /k/ seguinte por exemplo. Acontece que esses são alguns dos contextos seguintes de maior frequência. Alguns outros segmentos seguintes, entretanto, são extremamente diversos em efeito aparente, especialmente /b,g,r,l/. Resulta que esses segmentos apareceram em números muito baixos. Assim, o grau de convergência em porcentagem de apagamento está correlacionado com a quantidade de dados. Marquei com um asterisco, na Figura 3, todos os contextos seguintes que têm uma taxa de ocorrência de 50% ou mais abaixo da média de ocorrências por segmento seguinte, e com um ponto de exclamação todos os segmentos seguintes com altas taxas de ocorrência, definidas aqui como mais de um terço acima da média. Os itens marcados com !, com muitos dados, tendem a mostrar resultados similares nos dialetos, enquanto os marcados com *, com uma ocorrência muito baixa, tendem a divergir muito mais.

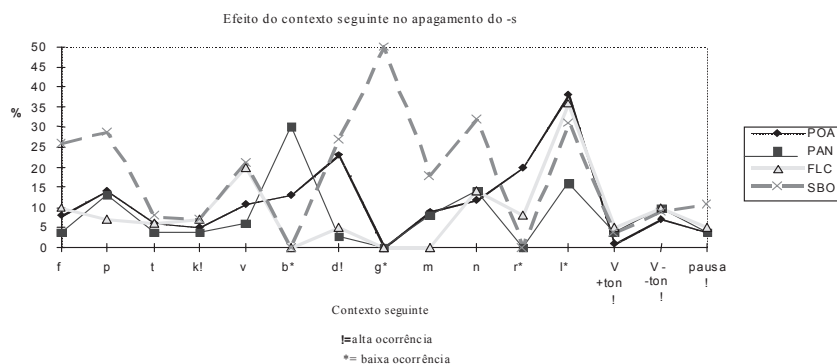


Fig.3

Passando a tratar agora da regra sintática de marcar o plural, podemos perguntar quão coerentes são as cidades nas restrições linguísticas desse processo? Nesse caso, com menos *knock-outs*, pudemos conduzir análises adequadas com o programa Varbrul em todas as cidades, e apresentaremos, conseqüentemente, os pesos relativos do Varbrul ao invés das frequências brutas.

r todo o Brasil, em inúmeros estudos sociolingüísticos, a começar com Braga (1977), a restrição mais forte nesse processo sempre foi a posição da palavra no sintagma nominal. Palavras mais à esquerda levam a marca de plural e palavras

mais à direita, não. Os dados do Varsul mostram esse efeito claramente, na Figura 4. Com impressionante coerência, todas as cidades têm essencialmente efeitos restritivos idênticos para posição: probabilidade muito alta de marcar o plural na primeira posição, e grandes reduções em posições subsequentes. Nesses dados, não há diferença significativa entre a segunda posição e as posições seguintes, a oposição é essencialmente binária entre primeira posição e todas as demais. Essa restrição, acredito, deve ser tratada como uma regra da gramática do português brasileiro coloquial, e em vista dos muitos estudos que foram feitos sobre esse fenômeno pelo país, é evidentemente uma regra válida em todo o Brasil.

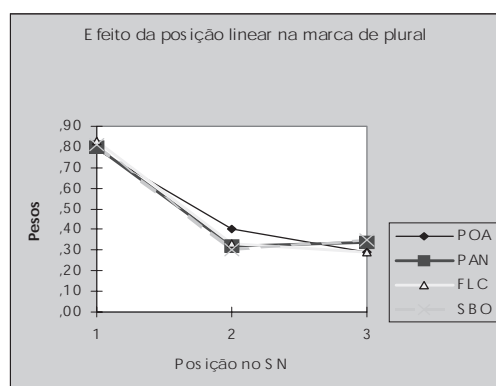


Fig.4

A próxima restrição que considerarei é um outro efeito de posição, a posição de uma palavra em relação ao núcleo do sintagma nominal, se ela o precede, o segue, ou é o núcleo. Isso é bastante independente da ordem linear: palavras de primeira posição podem ser o núcleo ou ser anterior a ele; palavras de segunda posição podem ser o núcleo, ou estar antes ou depois dele; e palavras da chamada terceira posição, uma vez que esta categoria inclui a quarta e as demais posições, podem ser também o núcleo, anteceder-lo ou segui-lo.

O que mostram os resultados? Na Figura 5 vemos evidências para talvez duas gramáticas diferentes nessa restrição. Duas cidades, PAN e SBO, marcam preferencialmente as posições anteriores ao núcleo, e desfavorecem a marca de plural do núcleo e das palavras que o sucedem. Uma cidade, POA, trata tanto o núcleo quanto palavras a sua esquerda como igualmente favoráveis a receber marca de plural, e apenas palavras pós-nucleares são desfavorecidas. A quarta cidade, FLC, só para tornar a vida mais interessante, fica entre as duas, com um declínio bem uniforme da esquerda para a direita no gráfico. Sugiro que a gramática de todos esses dialetos estabelece que a posição SPEC do sintagma nominal, geralmente a primeira palavra e à esquerda do núcleo, seja a posição

estrutural onde a marca de plural é exigida. Para PAN e SBO, essa é a única posição que favorece a marca de plural, enquanto para POA, tanto o SPEC quanto o núcleo favorecem a marca. FLC, eu esperaria que fosse mais como o padrão de PAN e SBO, por duas razões. Primeiro, esta cidade apresenta uma queda substancial em peso relativo entre a posição pré-nuclear e o núcleo, e segundo, ela apresenta uma diferença menor entre o núcleo e o que está à direita do que entre o núcleo e o que está à esquerda. Minha suposição é, então, que ela não representa uma gramática diferente, mas, ao contrário, não é significativamente diferente do tipo que marca preferivelmente a posição anterior ao núcleo, observado em PAN e SBO.

Efeito da posição em relação ao núcleo do SN

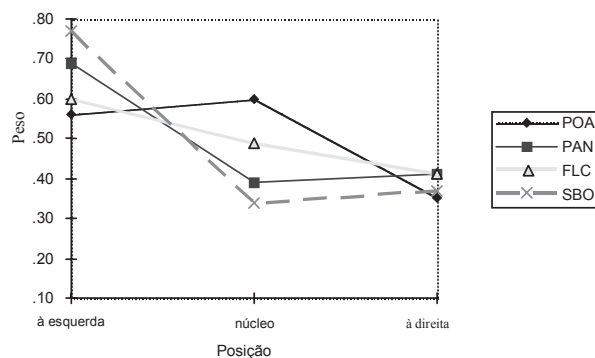


Fig.5

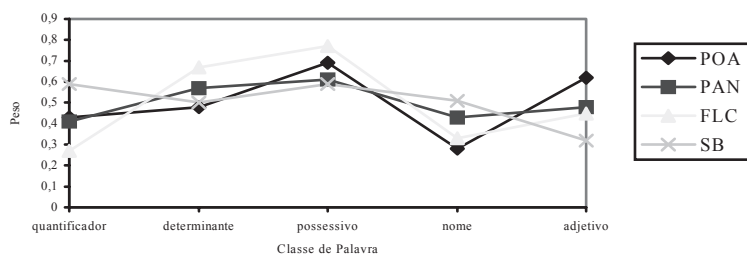


Fig.6

A terceira restrição morfossintática a ser considerada é classe de palavras, na qual distinguimos 5 categorias: quantificadores, determinantes, possessivos, substantivos e adjetivos, os quais listei nas suas posições mais comuns no sintagma

nominal, da esquerda para a direita, na Figura 6. (Um sintagma nominal raramente contém todos esses elementos, mas se um tivesse, como em *todos os meus livros pesados*, eles normalmente ocorreriam na ordem listada, embora alguns adjetivos ocorram antes do substantivo.) Então, os três primeiros elementos da lista normalmente são determinantes ou modificadores que antecedem o núcleo e quando apenas um deles está presente em um sintagma nominal (o que acontece na grande maioria dos sintagmas), este elemento geralmente estará na primeira posição. O substantivo é o núcleo do sintagma nominal, e ocorrerá mais freqüentemente na segunda posição, mas pode também aparecer na primeira, na terceira, ou em posições mais avançadas em certos tipos de sintagmas nominais. Finalmente, a maioria dos adjetivos aparecerá à direita do núcleo, na segunda, terceira ou em posições mais avançadas no sintagma nominal, mas alguns estarão à esquerda do núcleo, podendo mesmo ocorrer na primeira posição.

Com essas especificações em mente, a ordem da esquerda para a direita na Figura 6 reproduz parcialmente a ordem linear da esquerda para a direita mostrada na Figura 4 e a ordem da posição pré-nuclear até a posição pós-nuclear na Figura 5. Em geral, a Figura 6 mostra que as classes de palavras que antecedem o núcleo e podem aparecer na primeira posição do sintagma favorecem a marca, com uma redução de marca dessa posição para o núcleo e dali em diante. Há, porém, duas divergências nesse padrão. A primeira é que os quantificadores têm índices excepcionalmente baixos. Trata-se, provavelmente, de um efeito estrutural, já que muitos quantificadores na verdade aparecem fora do sintagma nominal, como SPEC de NP-barra². Esse seria o caso dos sintagmas nominais que possuem tanto um quantificador quanto um artigo – por exemplo, *todos os rapazes* –, uma estrutura que freqüentemente ocorre nos dados na forma *todo os rapazes*, sem marca de plural na posição inicial pré-nuclear, que se esperava, ao contrário, ser uma posição extremamente favorável à marca de plural. Esse resultado tenderia a apoiar a hipótese de que a posição preferida para a marca de plural é o SPEC do NP.

A segunda anomalia na Figura 6 é que os adjetivos têm várias vezes índices mais elevados do que se poderia esperar para uma classe de palavra que aparece geralmente mais para o final do sintagma e após o núcleo. Lembremos, porém, que há um conjunto substancial de adjetivos pré-nominais que apresentam uma freqüência relativamente alta. Essa, eu sugeriria, é a causa provável da diversidade entre as cidades nesses índices.

O último ponto sobre o fator classe de palavras é que ele não é selecionado como significativo na fase do *step-up* (na análise estatística com o Varbrul) para duas das cidades (PAN e SBO). Conseqüentemente, o *status* dessa variável como uma restrição independente na marcação de plural é duvidoso; talvez seja meramente derivado das duas variáveis anteriores.

² N.T. Apesar de optar pela tradução de *noun phrase* por *sintagma nominal*, a abreviatura NP foi mantida por ser convencional nos trabalhos de sintaxe.

CONCLUSÕES

O que podemos concluir, então, disso tudo? Vamos retomar o que descobrimos. Para três restrições do processo fonológico de apagamento do *-s*, geralmente encontramos forte coerência de efeitos dos fatores nas quatro comunidades. Onde há grande divergência de valores, foram identificadas duas explicações: primeiramente, muitos dos valores divergentes estão associados com baixo número de ocorrências e não têm significância estatística; provavelmente, eles são devidos a flutuações aleatórias da amostra. Em segundo lugar, um valor divergente com um número robusto de ocorrências foi atribuído a um efeito lexical: a cidade de SBO tem uma taxa divergentemente alta de ausência do *-s* em um morfema, o sufixo verbal *-mos*. Isso provavelmente reflete uma diferença de entradas lexicais, e não uma diferença estrutural na fonologia.

Em seguida, para as três restrições do processo sintático, encontramos alta coerência no efeito da posição linear, mas o efeito da posição relativa sugere a possibilidade de afirmações estruturais diferentes nas gramáticas sobre a localização favorecida para a marca de plural.

Dada a diversidade tipológica entre as línguas do mundo sobre as localizações estruturais da marca de plural nominal, isso talvez não seja surpreendente. Em todo caso, porém, o efeito prático de tal diferença entre POA e as outras comunidades que analisamos seria sutil: um pouco mais de marca de plural nos núcleos dos sintagmas nominais. E finalmente, a diversidade de efeitos para as classes de palavras, vista na figura 6, ocorre em uma restrição de significância marginal, e é maior para aquelas classes de palavras (quantificadores e adjetivos) que têm a maior área de ocorrência em diferentes posições estruturais. Consequentemente, também é improvável que isso represente diferenças gramaticais reais.

Geralmente, então, quando os dados são numerosos, as quatro cidades tendem a ser similares, se não idênticas, e a maioria das divergências que vemos entre as cidades é encontrada quando a quantidade de dados é pequena ou quando uma categoria tem uma ambigüidade maior quanto a sua posição estrutural canônica. As duas possibilidades que identificamos para diferenças gramaticais significativas incluem uma que é essencialmente lexical, e outra que envolve uma definição mais ampla contra uma mais restrita do alvo preferido da regra.

O que isso nos diz sobre a identidade lingüística da comunidade de fala? O português do Brasil, ou ao menos o português do estado do Rio Grande do Sul é uma só comunidade de fala ou são muitas? A evidência apresentada aqui sugere uma resposta mais sutil a essa questão. Em um nível, essas comunidades compartilham forte e sistematicamente a maior parte das mesmas restrições à variação. Mais ainda, as restrições que elas compartilham são elementos estruturais e, com base em outros tipos de evidências, poderíamos esperar que fossem altamente coerentes. Em dois pontos, contudo, identificamos restrições que

parecem receber um tratamento distinto por uma ou outra comunidade. Se pesquisa subsequente sobre os indivíduos dessas comunidades mostrar a mesma consistência interna que o efeito da pausa seguinte tem no apagamento do *-t,d* inglês em Nova Iorque e Filadélfia, eu concluiria que essas são distinções dialetais nascentes, definidas diferentemente por comunidades de fala locais. Essas comunidades, portanto, estão fazendo o que a teoria sociolinguística prediz: os muitos traços que são amplamente compartilhados refletem a unidade maior do português do Brasil como uma comunidade, distinto de outras línguas e de outros dialetos nacionais do português. No entanto, o número menor de traços que distingue certas cidades reflete o *status* delas como subcomunidades distintas dentro da comunidade maior.

Para a moral dessa história, tomarei a liberdade de fazer referência a Shakespeare: enquanto as comunidades de fala desenvolvem-se e mudam, a diferenciação que estabelecem sobrevive a elas, ao passo que similaridade é geralmente sepultada com seus ossos. Ao contrário de Marco Antônio, deveríamos louvar ambos os traços, e não enterrá-los.

BIBLIOGRAFIA

- BRAGA, Maria Luisa. *A concordância de número no sintagma nominal no Triângulo Mineiro*. Pontifícia Universidade Católica do Rio de Janeiro, 1977. Dissertação de Mestrado.
- CAMERON, Richard. Ambiguous agreement, functional compensation, and nonspecific tu in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change*, New York, n.6, p.305-333, 1994
- CEDERGREN, H. & Sankoff, D. Variable rules: performance as a statistical reflection of competence. *Language*, Baltimore, v.50, n.2, p.333-55, jun. 1974
- GUY, G.R. Variation in the group and the individual. *Locating language in time and space*, org. por W. Labov. New York, Academic Press, p.1-36, 1980.
- LABOV, W. The linguistic consequences of being a lame. *Language in the Inner City*. Philadelphia, University of Pennsylvania Press, capítulo 7, p. 255-92, 1972.