



Veredas atemática Volume 18 nº 2 – 2014

O apagamento de (-r) em coda nos limites da variação

Livia Oushiro (USP)
Ronald Beline Mendes (USP)

RESUMO: A partir das considerações de Weinreich et al. (2006) sobre a *questão do encaixamento e da implementação* da mudança linguística, este artigo discute os condicionamentos sociais e linguísticos do apagamento variável de (-r) em coda silábica, em São Paulo. Comparam-se três estágios da mudança linguística: seu final (enquanto morfema de infinitivo), um intermediário (em outras classes de palavras) e um inicial (quando se consideram apenas substantivos e adjetivos). Os resultados indicam que, nos estágios iniciais, a variação é condicionada apenas por fatores internos; nos estágios finais, prevalece o *estilo* (definido como grau de atenção à fala – Labov, 2001), o que aponta para a natureza supravernacular da manutenção de (-r) em certos contextos.

Palavras-chave: (-r) em coda; variação; mudança linguística; questão do encaixamento; português paulistano.

Introdução¹

Este artigo discute o encaixamento social e linguístico da pronúncia variável de (-r)² em coda silábica, do ponto de vista de sua realização *vs.* seu apagamento, em contextos

¹ Este trabalho foi previamente apresentado em forma de comunicação oral no II Jornadas Internacionales (Buenos Aires, ago.2013) graças ao auxílio financeiro da FAPESP (Proc. no. 2011/09122-6), a quem agradecemos. Estendemos também os agradecimentos aos dois pareceristas anônimos, ao público presente no encontro científico e, em especial, às perguntas e sugestões apresentadas por Marta Scherre – que, naturalmente, eximem-se de qualquer falha que aqui possa haver permanecido.

variáveis definidos pela classe morfológica e *estilo* de fala (LABOV, 2001, 2006 [1966]). Na comunidade paulistana, quando (-r) desempenha a função de morfema do infinitivo, seu apagamento é praticamente categórico na fala espontânea, o que denota os estágios mais avançados da mudança; por outro lado, quando em outras classes de palavras, sobretudo em substantivos e adjetivos, seu apagamento é bem menos frequente. O objetivo é comparar as correlações verificadas em diferentes estágios de mudança. A partir de análises multivariadas, demonstra-se que, nos estágios iniciais, a variação é condicionada apenas por fatores internos; em estágios intermediários, há a atuação de variáveis sociais e linguísticas; e nos estágios finais, prevalecem pressões supravernaculares, que previnem a erradicação do segmento.

Tratando-se de uma variável sociolinguística já bastante estudada em diversas comunidades brasileiras (ver, p.ex., OLIVEIRA, 1983 para Belo Horizonte-MG; SKEETE, 1997 para João Pessoa-PB; OLIVEIRA, 1999 para Salvador-BA; CALLOU et al., 1996, 2002 para as capitais documentadas pelo Projeto NURC; BRESCANCINI; MONARETTO, 2008 para as capitais do sul do país; LEITE, 2010 para Campinas-SP; *inter alia*), a realização variável de (-r) permite certas generalizações: (i) em muitas comunidades, parece estar em processo de mudança, em direção a realizações com o traço [+posterior]; (ii) o apagamento de (-r) é muito mais frequente em verbos do que em nomes, o que parece indicar que se trata de processos variáveis distintos: no primeiro caso, refere-se ao apagamento de um morfema redundante, visto que o infinitivo também é marcado pelo acento lexical; no segundo caso, trata-se de um caso de variação fonética propriamente; (iii) o apagamento é muito mais frequente em final de palavra do que em posição medial, o que dá pistas do primeiro contexto em que a alternância passou a ocorrer.

Na cidade de São Paulo, as taxas de apagamento de (-r) variam entre cerca de 3% em substantivos até cerca de 97% em verbos infinitivos, em situação de entrevista sociolinguística. Tais distribuições extremas raramente são analisadas em estudos sociolinguísticos, uma vez que podem ser consideradas casos categóricos ou semicategóricos (LABOV, 2003), ou seja, casos de não variação, que dificilmente permitem a análise quantitativa de regressão logística em programas como o GoldVarb X, devido a *knockouts*, interações entre grupos de fatores – que devem ser independentes entre si – ou células com um número muito pequeno de dados (GUY; ZILLES, 2007). Desse modo, na grande maioria de estudos sociolinguísticos, analisam-se variáveis em estágios intermediários (tanto casos de mudança como de variação estável), em que a distribuição dos dados permite visualizar claramente o encaixamento social e linguístico das variáveis.

Contudo, a observação dos estágios extremos da variação também deve ser capaz de lançar luz a fatos caros à Teoria da Variação e da Mudança. Weinreich et al. (2006 [1968]), em seu texto seminal e programático, formulam como uma das questões centrais deste campo de estudos o problema do *encaixamento*, que se refere à natureza e à extensão das correlações com variáveis independentes no decorrer da mudança linguística (p. 36). Acerca disso, afirmam os autores:

No desenvolvimento da mudança linguística, encontramos estruturas linguísticas encaixadas desigualmente na estrutura social; e *nos estágios iniciais e finais de uma mudança, pode haver pouca correlação com fatores sociais.* (WEINREICH et al., 2006 [1968], p. 123. Grifo acrescido.)

² Seguindo convenção em estudos variacionistas (LABOV, 1969), utilizam-se parênteses para fazer referência à variável, em contraste com a notação fonológica entre barras // e fonética entre colchetes [].

Essa questão se relaciona diretamente com o problema da *implementação*, que diz respeito às razões para que se observem as mudanças em certa língua e em certa época, e não em outras línguas e outras épocas com condições semelhantes. Para alguns autores, a motivação para a mudança linguística é sempre de natureza social (MEILLET, 1906, *apud* WEINREICH et al., 2006 [1968]; MUFWENE, 2008), uma consequência direta ou indireta de mudanças na estrutura da comunidade. Para outros, a mudança se dá em virtude de pressões internas da própria língua e por reanálise das estruturas gramaticais pelas novas gerações (LIGHTFOOT, 2010). Sobre essa questão, Weinreich et al. assumem uma postura mais prática ao afirmar que não é tarefa do linguista demonstrar a motivação social de uma mudança, mas sim descrever o grau de correlação com fatores sociais a fim de avaliar seu peso sobre o sistema linguístico. Os autores sugerem então que a mudança linguística começa quando um traço de variação se difunde através de um subgrupo da comunidade; ao longo do percurso, o traço se generaliza a outros elementos do sistema e pode adquirir certa significação social, possivelmente tornando-se um estereótipo; ao se aproximar do fim da mudança, no entanto, o traço perde qualquer significação social na passagem de variável para constante.

Nesse sentido, o presente trabalho se volta à questão do encaixamento da realização variável de (-r) em São Paulo a fim de testar a hipótese de Weinreich et al. sobre os estágios iniciais e finais da mudança: existem correlações entre o apagamento variável de (-r) e variáveis sociais nos contextos de substantivos/adjetivos e de verbos no infinitivo? Tanto em caso negativo quanto positivo, há diferenças entre um processo de mudança incipiente e outro quase completo?

Tais questões são analisadas através de análises multivariadas no programa Rbrul, em modelos de efeitos mistos (BAAYEN, 2008; JOHNSON, 2009), com a inclusão do falante e do item lexical como efeitos aleatórios. Os dados foram extraídos de 118 entrevistas sociolinguísticas que fazem parte do *corpus* do Projeto SP2010 (MENDES; OUSHIRO, 2013).³ Os falantes, todos paulistanos, estão estratificados de acordo com o seu sexo/gênero, três faixas etárias (20 a 34 anos, 35 a 59 anos, 60 anos ou mais), dois níveis de escolaridade (até Ensino Médio, Ensino Superior) e duas regiões de residência na cidade (bairros mais centrais e bairros mais periféricos).

Ainda que se opere com distribuições extremas, a amostra de 118 gravações permite a análise de um grande número de dados e a geração de modelos estatísticos viáveis. Entretanto, certos cuidados quanto à definição dos contextos variáveis ainda se fazem necessários, conforme se descreve na próxima seção. Em seguida, analisam-se três subconjuntos de dados, que se definem pela morfologia do segmento ou do vocábulo: em morfema de infinitivo (subseção 1.1); em outras classes de palavras (subseção 1.2); e somente em adjetivos/substantivos (subseção 1.3). Comparam-se, então, os resultados das análises quanto à natureza, social e linguística, das variáveis selecionadas como significativas (seção 2), e conclui-se com as implicações dos resultados para processos de variação e mudança linguística.

Cabe ainda ressaltar que se assume, dessarte, um quadro geral de mudança linguística no português brasileiro, de longa duração, e que tal premissa não necessariamente implica mudança em progresso em tempo aparente. Veja-se, por exemplo, a constatação de Callou et

³ Ver página do Projeto SP2010: <http://projetosp2010.fflch.usp.br/>. Último acesso em 08 abr. 2014.

al. (1998) de que (-r) se encontra atualmente em variação estável no Rio de Janeiro, ou a observação de Oushiro e Mendes (2013) sobre movimentos divergentes na comunidade quanto à realização de (-r) como tepe ou retroflexo em São Paulo. Especialmente para o caso de substantivos e adjetivos, que se definem aqui como contextos de “mudança incipiente”, não se pretende dar a entender que o apagamento de (-r) nessas classes de palavras deverá inevitavelmente seguir seu curso até a supressão total do segmento em algum momento futuro. Assume-se aqui o princípio geral de que já se observa mudança quando uma alternância adquire caráter de *heterogeneidade ordenada* (WEINREICH et al., 2006 [1968], p. 125); como se verá, tal é o caso dos três subconjuntos de dados que aqui se analisam.

1. Análise

A distribuição geral de dados de (-r) em coda silábica nas 118 entrevistas sociolinguísticas encontra-se na Tabela 1. Verifica-se que, de um total de mais de 68 mil dados, o apagamento é a variante mais frequente (56,3%), seguida das variantes tepe/vibrante múltipla (29,2%) e retroflexa (14,1%).⁴ Embora ocorram na fala de paulistanos nativos, as realizações fricativas são pouquíssimo expressivas na comunidade (0,4%) e foram desconsideradas na presente análise.

Codificação	%
Tepe/vibrante	29,2
Aproximante (retroflexa/alveolar)	14,1
Fricativa (velar/glotal)	0,4
Apagamento	56,3

Tabela 1: Distribuição geral de ocorrências na amostra (N = 68.764)

Desse total de ocorrências, foram codificados de 30 a 50 dados por falante. Tal decisão norteou-se pelo objetivo de obter um número razoavelmente equilibrado de dados de leituras e da conversação.⁵ Os dados de leituras advêm da parte final da entrevista, em que o documentador pede ao informante que leia (i) uma lista de palavras; (ii) uma notícia de jornal; e (iii) um texto com características da língua oral, chamado “Depoimento”.⁶ Em cada um desses estilos, espera-se que o informante apresente diferentes graus de monitoramento de sua fala, desde o mais cuidadoso (na lista de palavras) até o menos cuidadoso (na leitura de depoimento). A conversação no restante da gravação, por outro lado, pode ser considerada como relativamente mais espontânea do que os três estilos de leitura.

De fato, ao verificar a distribuição das variantes de (-r) de acordo com esses quatro estilos de fala, percebe-se que há taxas gradualmente maiores de sua realização quanto maior

⁴ Para uma análise sobre a alternância entre tepe e retroflexo no português paulistano, ver Oushiro e Mendes (2013).

⁵ Na Tabela 2 adiante, a taxa geral de 32% de apagamento, menor do que a taxa geral de 56,3% da Tabela 1, deve-se a esse balanceamento.

⁶ Ver Roteiro da entrevista sociolinguística, disponível no site do Projeto SP2010 (MENDES; OUSHIRO, 2013): <http://projetosp2010.fllch.usp.br/sites/projetosp2010.fllch.usp.br/files/SP2010-Roteiro.pdf>. Último acesso em 09 abr.2014.

é o grau de monitoramento por parte do falante. Na Tabela 2, da esquerda para a direita, apresentam-se o número de dados (colunas não sombreadas) e as respectivas taxas de apagamento e de realização de (-r) (colunas sombreadas) para cada um dos estilos. A última coluna apresenta os totais parciais, por classes de palavras, e o total geral: 8.531 dados.

		Conversaço		Depoimento		Jornal		Palavras		Total	
		N	%	N	%	N	%	N	%	N	%
Morfema de infinitivo (fazer)	Ø	1842	97	148	33	14	10	9	5	2013	76
	R	50	3	298	67	124	90	174	95	646	24
		1892		446		138		183		2659	
Conj./Prep. (conforme, porque)	Ø	409	52	0	--	0	0	0	--	409	49
	R	385	48	0	--	48	100	0	--	433	51
		794		0		48		0		842	
Verbos Finitos/raiz (conversar)	Ø	160	33	39	19	0	0	0	0	199	22
	R	318	67	164	81	164	100	69	100	715	78
		478		203		164		69		914	
Advérbios (normalmente)	Ø	2	12	0	--	0	--	0	--	2	12
	R	15	88	0	--	0	--	0	--	15	88
		17		0		0		0		17	
Adjetivos (moderna)	Ø	45	10	0	0	0	--	1	0	46	5
	R	422	90	107	100	0	--	447	100	976	95
		467		107		0		448		1022	
Substantivos (lugar)	Ø	44	3	2	1	0	0	3	0	49	2
	R	1300	97	177	99	256	100	1295	100	3028	98
		1344		179		256		1298		3077	
Total	Ø	2502	50	189	20	14	2	13	1	2718	32
	R	2490	50	746	80	592	98	1985	99	5813	68
		4992		935		606		1998		8531	

Tabela 2: Distribuição geral de (-r) de acordo com classe morfológica e estilo de fala

Observe-se que, para os morfemas de infinitivo, as taxas de apagamento são de 97% na conversaço, 33% na leitura do depoimento, 10% na leitura da notícia de jornal e 5% na lista de palavras. Para substantivos, embora a taxa de apagamento de (-r) seja bastante baixa na conversaço (3%), observa-se a sua diminuição no estilo de depoimento (1%) e na notícia de jornal e lista de palavras (0%).⁷

De cima para baixo na coluna “Conversaço”, também se observam taxas gradualmente menores para o apagamento de (-r) a depender do estatuto morfológico do segmento ou da palavra com (-r). A maior taxa, como se viu, encontra-se entre os morfemas de infinitivo (97%), seguida de conjunções/preposições (52%), segmentos de (-r) em verbos finitos ou na raiz do verbo (33%), advérbios (12%), adjetivos (10%) e substantivos (3%). O

⁷ A diferença entre as taxas de apagamento de (-r) em substantivos, em um contraste entre conversaço vs. leituras, é significativa de acordo com um teste de qui-quadrado ($\chi^2 = 43,6(1)$, $p < 0,001$) e uma análise em modelo de efeitos mistos ($p < 0,001$).

apagamento de (-r), portanto, é mais frequente em morfemas e palavras mais gramaticais, e relativamente menos frequente em palavras “de conteúdo”.⁸

Tomem-se, para fins de comparação, três estudos em diferentes regiões do país. Em Salvador, em dados de Diálogo entre Documentador e Informante (DID) do Projeto NURC, Oliveira (1999) constatou taxas de 18% de apagamento de (-r) em substantivos/pronomes e em adjetivos/advérbios, acima daquelas observadas na coluna “Conversação” da Tabela 2; considerando-se que se trata da fala de informantes cultos na década de 1970 – enquanto a presente amostra também inclui falantes menos escolarizados – e considerando-se que ali se observou favorecimento do apagamento pelas faixas etárias mais jovens, pode-se supor que as taxas atuais de apagamento de (-r) na fala soteropolitana sejam ainda maiores.

No Rio de Janeiro da década de 1990, na fala de informantes cultos, Callou et al. (1998) diferenciaram verbos (correspondentes às linhas “morfema de infinitivo” e “verbos finitos/raiz” da Tabela 2) de não-verbos. Para os primeiros, constataram uma taxa de 82% de apagamento, semelhante aos 84,5% para a presente amostra paulistana (2.002 casos em um total de 2.370 ocorrências de (-r) em morfemas de infinitivo e em verbos finitos); por outro lado, o apagamento em não-verbos correspondia a 32% no Rio de Janeiro, enquanto que a taxa comparável dessas categorias na São Paulo atual é de 19,1% (500 casos em um total de 2.622 em conjunções/preposições, advérbios, adjetivos e substantivos). Tanto no Rio de Janeiro quanto em Salvador, comunidades em que também ocorrem as variantes fricativas de (-r), parece haver maiores taxas de apagamento em não-verbos do que em São Paulo.

Em contraste, o levantamento de Monaretto (2000) nas capitais do sul do país parece revelar taxas semelhantes à capital paulista:⁹ o apagamento de (-r) em verbos é cerca de 81%, e em não-verbos (excluindo-se palavras funcionais) é de 5%. Na Tabela 2, a taxa geral de apagamento em advérbios, adjetivos e substantivos em dados de Conversação é igualmente 5% (91 casos de 1.828).

Nos presentes dados, exceto para o morfema de infinitivo, quase não se observa variação nos três estilos de leitura, seja porque não se incluíram certas classes de palavras no *design* da entrevista (por exemplo, não há conjunções e preposições no depoimento e na lista de palavras), seja porque o comportamento dos informantes tenha sido categórico (por exemplo, dos 448 dados de adjetivos na lista de palavras, houve apenas uma ocorrência de apagamento). Tais contextos, portanto, devem ser descartados da análise quantitativa. A partir das distribuições observadas, definem-se três conjuntos de dados: (i) morfemas de infinitivo (que foram analisados com a inclusão e a exclusão dos dados de leituras); (ii) todas as demais classes de palavras, em dados de Conversação; e (iii) somente substantivos e adjetivos, em dados de Conversação.

Em todas as análises no Rbrul, incluíram-se o falante e o item lexical como efeitos aleatórios, e sete variáveis sociais como efeitos fixos (BAAYEN, 2008; JOHNSON, 2009): (i) Sexo/Gênero; (ii) Faixa Etária; (iii) Nível de Escolaridade; (iv) Região de Residência (conforme descritas na estratificação da amostra); (v) Classe Social (em três estratos: classe alta/média alta; classe média média; classe média baixa e baixa); (vi) Origem dos Pais (paulistanos e não-paulistanos); e (vii) Mobilidade (o informante sempre morou no mesmo bairro; sempre morou na mesma zona; viveu em diferentes zonas da cidade). Faixa Etária e Classe Social também foram analisadas como variáveis contínuas – denominadas “Idade” (de

⁸ Tal gradação nas taxas de apagamento coaduna-se com uma visão de língua que conceitualiza o léxico, a morfologia e a sintaxe em um *continuum* (p.ex., LANGACKER, 1991).

⁹ Embora a autora assinala diferenças entre Curitiba, Florianópolis e Porto Alegre.

20 a 89 anos) e “Índice Socioeconômico” (de 1.0 a 5.0),¹⁰ respectivamente – em modelos que excluíram as variáveis qualitativas correspondentes. As variáveis linguísticas, por sua vez, compreendem: (i) o Contexto Fônico Precedente, reorganizado nas análises quanto aos traços [±alto] e [±anterior]; (ii) o Contexto Fônico Seguinte, analisado quanto ao Ponto de C (CLEMENTS; HUME, 1995), os traços [±sonoro] e [±contínuo]; (iii) a Posição de (-r) no vocábulo (final ou medial); (iv) a Tonicidade da sílaba com (-r) (tônica ou átona); e (v) Classe Morfológica (de acordo com as categorias da Tabela 2). Para as duas variáveis referentes ao contexto fonológico, a análise de cada traço foi realizada, evidentemente, em rodadas distintas: quando se incluiu o traço [±alto] da vogal precedente, não se incluiu [±anterior], e vice-versa. A variável Estilo foi incluída apenas em uma das análises de morfemas de infinitivo, visto que todas as demais incluem apenas os dados de Conversação.

1.1 Apagamento do morfema de infinitivo

O apagamento de (-r) na função de morfema de infinitivo é o caso mais avançado da mudança, conforme se visualiza na Tabela 2. Nessa análise, não foram incluídas as variáveis Posição de (-r), Tonicidade e Classe Morfológica, uma vez que os fatores são aqui constantes: todas as instâncias de (-r) ocorrem na sílaba tônica final de verbos no infinitivo. O resultado da análise multivariada que inclui todas as demais variáveis encontra-se na Tabela 3: são selecionadas como significativas apenas Estilo e o Contexto Fônico Precedente [±alto].

	P.R.	% R-Ø	Total
Estilo			
Conversação	.889	97,4	1.892
Depoimento	.013	33,2	446
Jornal	.005	10,1	138
Lista de Palavras	.001	4,9	183
<i>Range</i>		89	
Contexto Fônico Precedente [±alto]			
[-alto]	.70	85,8	1.236
[+alto]	.32	67,0	1.423
<i>Range</i>		38	

Tabela 3: Tendências de apagamento de (-r) em dados de morfemas de infinitivo (N = 2.659; % R-Ø = 75,7%). Input: 0,965.

Como já se poderia prever pelas drásticas diferenças de distribuição dos dados da Tabela 2, o apagamento de (-r) é altamente favorecido na fala mais espontânea (Conversação – P.R. 0.889) e, inversamente, muitíssimo desfavorecido nos estilos de leitura, com pesos relativos gradualmente menores quanto maior a atenção do informante à própria fala: P.R. 0.013 na leitura do depoimento, P.R. 0.005 na notícia de jornal, e P.R. 0.001 na lista de palavras. O Contexto Fônico Precedente quanto ao traço de altura também revela correlação:

¹⁰ O Índice Socioeconômico foi composto a partir de quatro parâmetros, a cada qual se atribuíram valores entre 1.0 a 5.0: (i) escolaridade do informante; (ii) ocupação do informante; (iii) escolaridade + ocupação do pai e da mãe do informante; (iv) renda média familiar do bairro de residência. O índice de cada falante é a média dos valores (i)-(iv). Ver Oushiro (2014) para uma descrição mais detalhada.

o apagamento de (-r) é favorecido quando precedido de vogais com o traço [-alto] (a, ó, é; P.R. 0.70) e desfavorecido por vogais com o traço [+alto] (i, e, o, u; P.R. 0.32).¹¹

O fato de que as demais variáveis não foram selecionadas poderia ter sido um efeito do forte grau de correlação com Estilo, que pode ter “obscurecido” outras correlações. Além disso, a reversão das tendências nos estilos de leitura em relação à fala mais espontânea sugere que a realização de (-r) se deve tão somente a pressões normativas – “deve-se falar do modo como se escreve” –, mas que já não se encontra no inventário vernacular dos falantes. Uma análise que inclui somente os dados de Conversação pode revelar o verdadeiro encaixamento da variável na comunidade paulistana. Os resultados dessa análise se encontram na Tabela 4.

	P.R.	% R-Ø	Total
Contexto Fônico Precedente [±alto]			
[-alto]	.74	98,8	1.012
[+alto]	.24	95,7	880
<i>Range</i>		50	
Nível de Escolaridade			
Até Ensino Médio	.63	98,5	926
Ensino Superior	.37	96,3	966
<i>Range</i>		26	

Tabela 4: Tendências de apagamento de (-r) em dados de morfemas de infinitivo (somente dados de Conversação. N = 1.892; % R-Ø = 97,3%) Input: 0,998.

Aqui, verifica-se que o Contexto Fônico Precedente quanto à altura da vogal se mantém como relevante para o apagamento de (-r), com *range*¹² maior do que na rodada anterior (50 vs. 38).¹³ No entanto, de todas as variáveis, apenas Nível de Escolaridade passa a ser selecionada pelo Rbrul. O apagamento de (-r) é favorecido pelos falantes com menor nível de escolarização (P.R. 0.63) e desfavorecido entre falantes com nível superior de escolaridade (P.R. 0.37). Ainda que as taxas de apagamento sejam bastante altas na Tabela 4, note-se que a realização de (-r) entre os falantes menos escolarizados (1,5%) é cerca de 2,5 vezes menor do que entre falantes mais escolarizados (3,7%). Tal diferença se reflete na seleção da variável como relevante para a variação.

Embora Faixa Etária como variável discreta (1ª, 2ª e 3ª faixas etárias) não se correlacione significativamente com (-r), a sua análise de uma perspectiva contínua revela uma leve tendência de decréscimo da probabilidade de apagamento quanto mais velho é o falante (*logodds* -0,027, $p < 0,05$; ver JOHNSON, 2009). Em outras palavras, quanto mais jovem o falante, maior a sua tendência de apagar (-r) em morfemas de infinitivo, o que está de

¹¹ Quando reorganizado pelo traço [±anterior], em rodada distinta, o Contexto Fônico Precedente também é selecionado: vogais [-anterior] (a, ó, o, u) favorecem o apagamento (P.R. 0.65), e vogais [+anterior] (i, e, é) o desfavorecem (P.R. 0.37). O valor de *log-likelihood* da rodada com o traço de altura (-511,0764), mais próximo de zero do que a rodada com o traço de anterioridade (-514,1088), revela que o primeiro tem maior poder explicativo, embora ambos sejam relevantes (GUY; ZILLES, 2007).

¹² Diferença entre maior e menor pesos relativos.

¹³ O Contexto Fônico Precedente [±anterior] também é novamente selecionado, com favorecimento do apagamento em vogais com o traço [-anterior] (P.R. 0.68) e desfavorecimento em vogais [+anterior] (P.R. 0.29).

acordo com a hipótese de que se trata de uma mudança na comunidade e no português brasileiro.

1.2 Apagamento em outras classes de palavras

A análise que exclui os dados de morfemas de infinitivo revela um encaixamento muito mais complexo de (-r), com múltiplas correlações de natureza linguística e social.

	P.R.	% R-Ø	Total
Classe morfológica			
Conjunção/Preposição	.96	51,5	794
Verbo finito	.76	33,5	478
Adjetivo/Advérbio	.27	9,7	484
Substantivo	.12	3,3	1.344
<i>Range</i>	84		
Posição da sílaba com R			
Final	.88	36,9	667
Medial	.37	17,0	2.433
<i>Range</i>	51		
Contexto Fônico Precedente [±alto]			
[-alto]	.64	22,1	1.099
[+alto]	.42	19,8	2.001
<i>Range</i>	22		
Contexto Fônico Seguinte [±contínuo]			
[-contínuo]	.54	26,3	2.146
[+contínuo]	.41	10,0	954
<i>Range</i>	13		
Região de Residência			
Bairro mais periférico	.55	23,1	1.588
Bairro mais central	.45	19,4	1.512
<i>Range</i>	10		
Sexo/Gênero			
Masculino	.55	22,1	1.553
Feminino	.45	20,5	1.547
<i>Range</i>	10		

Tabela 5: Tendências de apagamento de (-r) em dados de outras classes de palavras (N = 3.100; % R-Ø = 21,3%) Input: 0,071.

A Tabela 5 mostra que Classe Morfológica é a variável mais importante para o apagamento de (-r), com *range* de 84 e pesos relativos que vão de 0.96 (para conjunção/preposição) até 0.12 (para substantivo). A hierarquia observada entre esses fatores é a mesma da tabela de distribuição geral mais acima: há maior tendência de apagamento de (-r) em palavras mais gramaticais, e menor tendência relativa em palavras “de conteúdo”. A Posição da sílaba com (-r) também se mostra relevante, à semelhança de outras comunidades

brasileiras: a posição final (p.ex. *mulher, melhor*) é bastante favorecedora do apagamento (P.R. 0.88), em contraste com a posição medial (p.ex. *sorte, perto*, P.R. 0.37).

Assim como na análise para morfemas de infinitivo, a altura da vogal precedente se correlaciona com o apagamento, na mesma direção previamente observada: favorecimento para vogais com o traço [-alto] (P.R. 0.64) e desfavorecimento para vogais [+alto] (P.R. 0.42).¹⁴ O Contexto Fônico Seguinte não é selecionado quando organizado por Ponto de C nem por sonoridade, mas é relevante para o apagamento de (-r) quanto ao traço [±contínuo]; o resultado, no entanto, é contrário àquele observado em outras comunidades (Cf. OLIVEIRA, 1999; SKEETE, 1997): é o traço [-contínuo], como em africadas e oclusivas (p.ex. *porque*) que favorece o apagamento (P.R. 0.54) em relação ao traço [+contínuo], como em líquidas e fricativas (p.ex. *versão*, P.R. 0.41).

As variáveis sociais Região de Residência e Sexo/Gênero também passam a se mostrar relevantes, em contraste com a análise prévia de morfemas de infinitivo. O apagamento de (-r) é favorecido por habitantes de bairros mais periféricos e do sexo masculino, ambos com peso relativo de 0.55, o que parece ser reflexo de certo grau de estigma da variante quando se trata de outras classes morfológicas que não o morfema de infinitivo. A seleção da variável contínua Índice Socioeconômico reforça essa interpretação: quanto mais alta a classe social do falante, menor a probabilidade de apagamento de (-r) nesse contexto (*logodds* -0,341, *p* < 0,03). É interessante notar que nem Faixa Etária nem Idade são selecionadas como significativas nessas rodadas, o que indica ser este um caso de variação estável.

1.3 Apagamento em substantivos e adjetivos

Embora substantivos e adjetivos já tenham sido incluídos na análise anterior, o exame desse subconjunto de dados pode revelar o encaixamento da variável em seus estágios iniciais da mudança. Conforme já se notou anteriormente, essas classes de palavras são aquelas que menos favorecem o apagamento de (-r), tanto em termos de frequência (3% e 10%; ver Tabela 2) quanto em termos de tendência (P.R. 0.12 e 0.27; ver Tabela 5).

A Tabela 6 mostra que, nesses casos, há apenas correlações de natureza linguística. De modo semelhante à análise que inclui todos os dados exceto morfemas de infinitivos, desfavorecem o apagamento de (-r) a posição medial da sílaba (P.R. 0.28), a vogal precedente com traço [+alto] (P.R. 0.28) e substantivos (P.R. 0.44); o contexto fônico seguinte [±contínuo], no entanto, revela direção contrária àquela observada na seção 2.2: para substantivos e adjetivos, é o traço [+contínuo] (p.ex. *cerveja*) que favorece o apagamento (P.R. 0.62), em consonância com resultados verificados em outras comunidades. Em outros estudos que se utilizaram da categorização “verbos vs. não-verbos”, a maior tendência de apagamento quando (-r) é seguido por consoantes com o traço [+contínuo] pode ter sido motivada simplesmente pela tendência de haver mais dados de substantivos/adjetivos do que outras categorias de “não-verbos”.

O Contexto Fônico Precedente quanto ao traço [±anterior] também revela correlação distinta: neste caso, são as vogais [+anterior] que favorecem o apagamento (P.R. 0.80) em relação às vogais [-anterior] (P.R. 0.31). Tais diferenças na direção da correlação reforçam a necessidade de separar os dados em diferentes classes morfológicas, para além da categorização “verbos vs. não verbos”.

¹⁴ O traço [±anterior] também é novamente selecionado: [-anterior] 0.64; [+anterior] 0.43.

	P.R.	% R-Ø	Total
Posição da sílaba com R			
Final	.95	19,0	436
Medial	.28	0,4	1.375
	<i>Range</i>	67	
Contexto Fônico Precedente [±alto]			
[-alto]	.74	9,3	853
[+alto]	.28	1,0	958
	<i>Range</i>	46	
Classe Morfológica			
Adjetivo	.66	9,6	467
Substantivo	.44	3,3	1.344
	<i>Range</i>	22	
Contexto Fônico Seguinte [±contínuo]			
[+contínuo]	.62	5,1	669
[-contínuo]	.43	4,8	1.142
	<i>Range</i>	19	

Tabela 6: Tendências de apagamento de (-r) em dados de adjetivos e substantivos (somente dados de Conversação. N = 1.811; % R-Ø = 4,9%) Input: 0,006.

Nota-se, por fim, que Faixa Etária e Idade não são selecionadas como significativas para a variação de (-r) neste contexto, à semelhança do conjunto “outras classes de palavras” e diferentemente do conjunto “morfemas de infinitivo”.

2. Discussão

O Quadro 1 resume as variáveis selecionadas como significativas para o apagamento de (-r) na análise de cada conjunto de dados. Para substantivos e adjetivos, casos em que (-r) é menos frequentemente apagado, observam-se correlações apenas com fatores internos. Em dados que excluem somente os morfemas de infinitivo, com distribuições intermediárias, verificam-se correlações tanto com fatores internos quanto com fatores externos. Para morfemas de infinitivo, há correlação com apenas uma variável interna, e com as variáveis Idade e Nível de Escolaridade (quando se consideram apenas dados de Conversação) ou ainda apenas Estilo (quando se considera o conjunto Conversação e Leituras).

Subconjunto de Dados	Fatores externos	Fatores internos
Verbos infinitivos (Conversação + Leituras)	Estilo	Contexto Fônico Precedente
Verbos infinitivos (Conversação)	Nível de Escolaridade Idade	Contexto Fônico Precedente
Outras classes morfológicas (Conversação)	Região de Residência Sexo/Gênero Índice socioeconômico	Classe morfológica Posição da sílaba com R Contexto Fônico Precedente Contexto Fônico Seguinte
Substantivos e adjetivos (Conversação)	--	Posição da sílaba com R Contexto Fônico Precedente Classe Morfológica Contexto Fônico Seguinte

Quadro 1: Resumo dos resultados das análises

Retorne-se, então, às questões colocadas inicialmente neste trabalho: existem correlações entre o apagamento de (-r) e variáveis sociais nos contextos de verbos no infinitivo e de substantivos/adjetivos? E há diferenças entre um processo de mudança incipiente e um quase completo?

O Quadro 1 indica que não há correlações com variáveis sociais nas classes morfológicas que estão nos estágios iniciais da mudança (substantivos e adjetivos), e que o encaixamento social se dá mais propriamente em classes que se encontram em estágios intermediários (outras classes morfológicas). Embora Estilo não seja uma variável linguística, tampouco pode ser considerada exatamente como uma variável social, do mesmo modo que o sexo ou a classe socioeconômica do falante; trata-se, de fato, de uma variável intra-falante, e não de uma macro-categoria social que pode ser atribuída a indivíduos.¹⁵ A distribuição geral dos dados (Tabela 2) e a análise multivariada (Tabela 3) mostram claramente o efeito do monitoramento da fala; na situação de conversação na entrevista sociolinguística, mesmo considerando-se o Paradoxo do Observador (LABOV, 2006 [1966]), os falantes apresentam taxas altíssimas de apagamento de (-r) infinitivo; a sua realização só se dá quando os informantes se veem na tarefa de ler textos e listas de palavras, em que os vocábulos com (-r) em coda foram todos grafados de acordo com a ortografia padrão – *ver* e não *vê*, *costrar* e não *costrá*, *erguer* e não *erguê*. Trata-se da pressão normativa e da ideologia escolar, imposta sobre o vernáculo dos falantes, que prescreve que o modo “correto” de falar é tal qual se escreve. A seleção de Nível de Escolaridade (em vez de classe social, sexo/gênero, região de residência, origem dos pais ou mobilidade), quando não se inclui a variável Estilo, pode ser entendida da mesma forma: antes de estratificação social, trata-se da ativação de normas supravernaculares mais ou menos salientes de acordo com o nível de escolarização. Desse

¹⁵ É digno de nota que, dentro do próprio arcabouço de estudos sociolinguísticos, há diferentes concepções da variável “Estilo”. Para Eckert (2001), por exemplo, estilo não se limita ao grau de atenção à fala, mas refere-se ao *locus* privilegiado da criação de significados sociais, através de um processo contínuo de *bricolage*, em que os falantes combinam e recombina variantes na construção de *personae* e identidades sociais. Para Irvine (2001), estilo diz respeito a “distintividade”, características que distinguem indivíduos ou grupos e que se formam a partir da “ideologização da diferenciação”. Para Bell (2001), ainda, a variação estilística na fala individual pode ser atribuída aos interlocutores (reais ou imaginados), em seu modelo do *design* da audiência. Ver referências em Eckert e Rickford (2001).

modo, pode-se afirmar que, ao menos quanto ao apagamento de (-r), a hipótese de Weinreich et al. (2006 [1968]) sobre os estágios limítrofes da mudança encontra confirmação empírica: não há correlações sociais no início e no fim da mudança linguística.

Quanto à *questão da implementação*, o encaixamento inicial, apenas de natureza linguística, poderia ser interpretado como evidência de que a mudança teve início por pressões puramente internas à língua. No entanto, embora aqui se analisem apenas dados de falantes paulistanos, há que se lembrar que a cidade de São Paulo é cenário de contato entre diversas variedades do português brasileiro, devido à intensa migração de habitantes de outras cidades e regiões: segundo levantamento do IPEA (2011), 46% da população adulta na Região Metropolitana não nasceu do estado de São Paulo (portanto, os 54% restantes incluem tantos os paulistanos nativos quanto os migrantes do interior do estado). Dos migrantes, a grande maioria advém de estados do Norte e do Nordeste do país (sobretudo Bahia 11%, Pernambuco 7% e Ceará 7%) e se estabelece nas regiões periféricas da cidade. Vale notar que Região de Residência torna-se relevante para a variação de (-r) em estágios intermediários, e que são os residentes de bairros mais periféricos que favorecem o apagamento. É possível, desse modo, que o apagamento de (-r) em outras classes de palavras, incluindo-se substantivos e adjetivos, também tenha sofrido a influência do contato dialetal com variedades cujas taxas de emprego da variante são maiores. O fato de que a variação em substantivos e adjetivos tem caráter de heterogeneidade ordenada somente no sistema linguístico, e não no sistema social, não exclui a possibilidade de motivações externas à língua para o início da mudança.

Considerações finais

Esta análise contrasta o encaixamento linguístico e social do apagamento variável de (-r) na comunidade paulistana em três contextos que podem ser considerados representativos dos estágios inicial, intermediário e final de uma mudança linguística. A análise de um *corpus* robusto, composto de 118 entrevistas sociolinguísticas, permitiu tal exame mesmo com distribuições extremas em alguns contextos, como em substantivos e em morfemas de infinitivo. Os resultados de análises multivariadas mostram que, nos estágios inicial e final, há apenas correlações com variáveis internas; a manutenção de (-r) nos estágios finais pode ser atribuída tão somente a pressões supravernaculares. O pleno encaixamento social e linguístico, desse modo, dá-se apenas nos estágios intermediários, o que fornece subsídios empíricos para a hipótese de Weinreich et al. (2006 [1968]) sobre a questão do encaixamento. A generalização sobre os tipos de correlações encontradas em diferentes estágios de mudança linguística certamente dependerá de novos estudos sobre outras variáveis e em outras comunidades. A presente análise, entretanto, já aponta para possíveis respostas para as questões do encaixamento e da implementação no quadro geral da Teoria da Variação e da Mudança Linguística.

Coda (-r) deletion at the limits of variation

ABSTRACT: Based on Weinreich et al.'s (2006) remarks about the embedding and the actuation problems in language change, this article analyzes the social and linguistic constraints on the variable realization of coda (-r) in the city of São Paulo. We compare three phases of language change: its end (when (-r) is the infinitival morpheme), its intermediate stage (in other word classes) and its beginning (in nouns and adjectives). The results suggest that, in the initial phase, the variation is conditioned only by internal factors and, in the final phase, only by *style* (defined as attention paid to speech – Labov 2001), which points to the supravernacular nature of (-r)'s continued occurrence in certain contexts.

Keywords: coda (-r); language variation; language change; the embedding problem; São Paulo Portuguese.

Referências

BAAYEN, R. H. *Analysing Linguistic Data: a Practical Introduction to Statistics*. Cambridge: Cambridge University Press, 2008.

BRESCANCINI, C.; MONARETTO, V. N. de O. Os róticos no sul do Brasil: panorama e generalizações. *Signum: Estudos Linguísticos*, vol. 11, n.2, p. 51-66, 2008.

CALLOU, D.; LEITE, Y.; MORAES, J. Processo(s) de enfraquecimento consonantal no português do Brasil. In: ABAURRE, B.; RODRIGUES, A. (Eds.). *Gramática do português falado: novos estudos descritivos*, vol. 8. Campinas: Editora da Unicamp, 2002. p. 537-555.

CALLOU, D.; MORAES, J.; LEITE, Y. Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil. In: KOCH, I.; MORAES, J.; LEITE, Y. (Eds.). *Gramática do português falado*, vol. 6. Campinas: Editora da Unicamp, 1996. p. 465-494.

CALLOU, D.; MORAES, J.; LEITE, Y. Apagamento do R final no dialeto carioca: um estudo em tempo aparente e tempo real. *DELTA* [online], vol. 14, 1998. Disponível em <<http://dx.doi.org/10.1590/S0102-44501998000300006>>. Último acesso em 10 abr. 2014.

CLEMENTS, G. N.; HUME, E. V. The internal organization of speech sounds. In: GOLDSMITH, J. A. (Ed.). *The Handbook of Phonological Theory*. Oxford: Blackwell, 1995. p. 245-306.

ECKERT, P.; RICKFORD, J. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001.

GUY, G. R.; ZILLES, A. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola, 2007.

IPEA. *Comunicados do IPEA no. 115: Perfil dos migrantes em São Paulo*, 2011. Disponível em

<<http://www.ipea.gov.br/portal/images/stories/PDFs/comunicado/111006comunicadoipea115.pdf>>. Último acesso em 16 nov. 2011.

JOHNSON, D. E. Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, vol. 3, n. 1, p. 359-383, 2009.

LABOV, W. Contraction, deletion, and inherent variability of the English copula. *Language* vol. 45, n. 4, p. 715-762, 1969.

LABOV, W. The anatomy of style-shifting. In: ECKERT, P.; RICKFORD, J. R. (Ed.) *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001. p. 85-108.

LABOV, W. Some sociolinguistic principles. In: PAULSTON, C. B.; TUCKER, G. R. (org.). *Sociolinguistics: the essential readings*. Oxford: Blackwell, 2003. p. 235-250.

LABOV, W. *The social stratification of English in New York City*. 2. ed. New York: Cambridge University Press, 2006 [1966].

LANGACKER, R. *Foundations of Cognitive Grammar*, vol II: descriptive application. Stanford: Stanford University Press, 1991.

LEITE, C. M. B. *O /R/ em posição de coda silábica no falar campineiro*. 2010. 227f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Unicamp. Campinas, 2010.

LIGHTFOOT, D. Language acquisition and language change. *Cognitive Science*, vol. 1, n. 5, p. 677-684, set./out. 2010.

MENDES, R. B.; OUSHIRO, L. Documentação do Projeto SP2010 – Construção de uma amostra da fala paulistana. 2013. Disponível em <<http://projetosp2010.fflch.usp.br/producao-bibliografica>>. Último acesso em 21 mar. 2014.

MONARETTO, V. N. de O. O apagamento da vibrante posvocálica nas capitais do sul do Brasil. *Letras de Hoje*, vol. 35, n. 1, p. 275-284, mar. 2000.

MUFWENE, S. *Language evolution: Contact, competition, and change*. London: Continuum Press, 2008.

OLIVEIRA, J. M. *O apagamento do /R/ implosivo na norma culta de Salvador*. 1999. 80f. Dissertação (Mestrado em Letras) – Universidade Federal da Bahia. Salvador, 1999.

OLIVEIRA, M. A. *Phonological variation and change in Brazilian Portuguese: the case of the liquids*. 1983. 286f. Tese (Doutorado em Linguística) – University of Pennsylvania. Philadelphia, 1983.

OUSHIRO, L. *Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São Paulo*. Tese (Doutorado em Linguística). São Paulo, Faculdade de Filosofia, Letras e Ciências Humanas, USP, 2014.

OUSHIRO, L.; MENDES, R. B. A pronúncia do (-r) em coda silábica no português paulistano. *Revista do GEL*, vol. 8, n. 2, p. 66-95, 2013.

SKEETE, N. A. O uso variável de vibrante na cidade de João Pessoa. *Graphos*, vol. 2, n. 1, p. 77-96, 1997.

WEINREICH, U.; LABOV, W.; HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola, 2006 [1968]. Tradução: Marcos Bagno.

Data de envio: 26/05/2014

Data de aceite: 21/01/2015

Data de publicação: 23/04/2015