

A COESÃO DIALETAL NAS CONCORDÂNCIAS NOMINAL E VERBAL NO PORTUGUÊS PAULISTANO¹

LECTAL COHESION IN SÃO PAULO PORTUGUESE NOMINAL AND VERB AGREEMENT

LIVIA OUSHIRO
Universidade de São Paulo
livia.oushiro@usp.br

A questão do encaixamento simultâneo de múltiplas variáveis sociolinguísticas é raramente abordada em estudos sobre a variação e a mudança. Com base em um *corpus* de 118 gravações com falantes paulistanos, este artigo discute a covariação entre a concordância nominal, a concordância verbal de 3PP e a concordância de 1PP, com vistas a investigar se falantes que tendem a empregar a variante não padrão de uma variável também tendem a fazê-lo para as outras duas, ou se tais variáveis se encaixam independentemente no sistema linguístico e social. As tendências de emprego de cada falante para as três variáveis foram analisadas através do cálculo de coeficientes de Pearson entre pares de variáveis (CN-3PP, CN-1PP e 3PP-1PP). Os resultados mostram que a covariação entre CN e 3PP é mais forte do que aquela entre as concordâncias verbais de 3PP e de 1PP. Sugere-se que a covariação depende não apenas de semelhanças estruturais, mas também de coesão social e de fatores linguísticos que se correlacionam com múltiplas variáveis, tais como a saliência fônica.

Palavras-chave: concordância nominal, concordância verbal, Português Paulistano, covariáveis

The question of the simultaneous embedding of multiple sociolinguistic variables in a community is rarely addressed in studies of language variation and change. Based on a *corpus* of 118 recordings with Paulistano speakers, this article discusses co-variation among three variables of Brazilian Portuguese: nominal agreement, 3PP verb agreement, and 1PP verb agreement. The main goal is to analyze whether speakers who tend to use the nonstandard variant of a variable also tend to employ other nonstandard variants of the other two variables, or if they are independently embedded in the social and linguistic system. Each speaker's weight for each variable was analyzed through Pearson correlation coefficients between pairs of variables (NP-3PP, NP-1PP, and 3PP-1PP). The results show that co-variation between NP and 3PP is stronger than that between the verb agreement variables 3PP and 1PP. We suggest that co-variation depends not only on structural similarities, but also on social cohesion and on internal factors correlated with multiple variables, such as phonic salience.

Keywords: nominal agreement, verb agreement, Paulistano Portuguese, co-variables

¹ Este trabalho é fruto de pesquisa de doutorado financiada pela FAPESP (Processos 2011/09122-6 – Regular e 2012/01930-9 – BEPE).

0. INTRODUÇÃO

O mapeamento da concordância de número em sintagmas nominais e verbais no Português já é bastante extenso (ver, p.ex., Braga 1977, Scherre 1978, 1988, Guy 1981, 2000, Naro & Scherre 1991, Fernandes 1996, Lopes 2001, Antonino 2007, Salomão 2010, Brandão & Vieira 2012, Yacovenco *et al* 2012, Silva e Scherre 2013 sobre a concordância nominal; Naro 1981, Guy 1981, Bortoni-Ricardo 1985, Rodrigues 1987, Naro *et al* 1999, Zilles *et al* 2000, Pereira 2004, Coelho 2006, Scherre & Naro 2006, Scherre *et al* 2007, Monguilhott 2009, Brandão & Vieira 2012, Lucchesi 2012, Rubio 2012 sobre concordâncias verbais). Certos resultados são recorrentes nas mais variadas comunidades em que os fenômenos foram investigados, o que permite generalizações a respeito dessas variáveis. Por exemplo, a posição linear do vocábulo dentro do sintagma nominal, sua posição relativa ao núcleo, a saliência fônica (Scherre 1988) e o efeito serial do paralelismo (Scherre & Naro 1992, Scherre 1998, 2001) normalmente se mostram correlacionados à concordância nominal. A saliência fônica (Naro 1981, Naro *et al* 1999) e o paralelismo discursivo (Scherre & Naro 1992) também se correlacionam com as concordâncias de 1PP e de 3PP, bem como o tipo de sujeito, sua posição em relação ao verbo ou sua animacidade. Socialmente, as variantes não padrão tendem a ser desfavorecidas pelas mulheres e por falantes de faixa etária intermediária em centros urbanos (ver, p.ex., Scherre & Naro 1998), e a ser desfavorecidas pelos homens e pelos mais jovens em comunidades rurais ou rurbanas (p.ex. Bortoni-Ricardo 1985, Lucchesi 2012). O nível de escolaridade ou a classe social dos falantes se mostram invariavelmente relevantes, com a marca zero favorecida por falantes menos escolarizados ou de classes mais baixas.

No entanto, a maior parte desses trabalhos focaliza apenas uma variável sociolinguística. Quando mais de uma é estudada em uma mesma comunidade ou em um mesmo *corpus* (por exemplo, as concordâncias verbais de 1PP e de 3PP – Bortoni-Ricardo 1985, Rodrigues 1987, Pereira 2004, Rubio 2012), normalmente se realizam análises separadas, cujos resultados são posteriormente comparados. Entretanto, os falantes de uma comunidade lidam com múltiplas variáveis simultaneamente em sua fala cotidiana, mas pouco se sabe sobre como um mesmo falante se comporta em relação a mais de uma delas. Será que falantes que tendem a empregar a marca zero de concordância nominal (p.ex., “os menino”) também tendem a empregar as variantes não padrão de 1PP (p.ex. “nós foi”) e de 3PP (p.ex., “eles foi”)? Ou será que tais variáveis se encaixam de modo independente uma das outras no sistema linguístico e social? O estudo da covariação visa a verificar se múltiplas variáveis se correlacionam nos usos de falantes individuais e, se sim, quais fatores sociais e linguísticos promovem a covariabilidade.

Por um lado, os recorrentes padrões de estratificação social, variação estilística e mudanças em progresso descritos em estudos sociolinguísticos (ver, p.ex., Labov 2001) conduzem à expectativa de que certos pares de variáveis covariem; se, em determinada comunidade, formas não padrão de diversas variáveis tendem a ser favorecidas pelos falantes do sexo masculino e menos escolarizados, por exemplo, uma expectativa razoável é que tais variáveis também se correlacionem fortemente entre si. Por outro lado, também se pode esperar covariação devido a pressões internas à língua, como a simetria fonológica (Martinet 1955, Jespersen 1953 [1933]) ou regularização paradigmática (Duarte & Varejão 2013).

De fato, tanto leigos quanto linguistas tendem a reificar certas variedades sociais. Fala-se, sem muitas ressalvas, a respeito de um estilo “formal” ou de um estilo “informal”, sobre um “Português Popular” ou um “Português Padrão”, ou sobre a existência de um Português

Paulistano, um Português Caipira, um Português Carioca etc. Essa prática traz consigo a implicação de que as variedades nomeadas constituem objetos coerentes, que se caracterizariam por um conjunto de variantes que coocorrem. Contudo, tal expectativa foi poucas vezes colocada à prova empiricamente.

Este trabalho investiga possíveis correlações entre a concordância nominal (CN), a concordância verbal de 1PP (CV-1PP) e a concordância verbal de 3PP (CV-3PP), e discute o papel de fatores linguísticos e sociais na covariação. A próxima seção apresenta uma breve revisão da literatura sobre covariáveis, cujos desenvolvimentos são relativamente recentes. A seção 2 descreve os materiais e os métodos empregados na presente análise, bem como certas previsões a partir dos resultados de análises multivariadas separadas sobre cada variável. Em seguida, descrevem-se padrões gerais de covariação, assim como análises mais detalhadas quanto a fatores sociais e linguísticos. Demonstra-se que a correlação mais forte é entre o par CN-3PP, e não entre 3PP-1PP, diferentemente do que uma perspectiva puramente estrutural poderia prever; tal fato se deve a uma maior coesão da comunidade paulistana quanto ao uso de CN e 3PP do que em relação a 1PP. Sugere-se, simultaneamente, que a covariação é condicionada por variáveis internas que se correlacionam com múltiplas variáveis, tais como a Saliência Fônica: formas mais salientes tendem a estabelecer correlações mais fracas com outras formas linguísticas. Por fim, o artigo se conclui com uma síntese dos resultados e avalia novas perspectivas de análise que se abrem nos estudos sociolinguísticos com o exame de covariáveis.

1. O ESTUDO DE COVARIÁVEIS

A questão da covariação remonta ao trabalho seminal de Labov sobre a estratificação social do inglês falado na cidade de Nova Iorque (Labov 2006 [1966]). O autor fornece fortes evidências da coesão estrutural de socioletos ao demonstrar a covariação entre pares de variáveis fonéticas: o alçamento de (æh) e (oh) em palavras como *bad* ‘mau’ e *law* ‘lei’ respectivamente; entre (ay) e (aw) como em *ride* ‘andar, dirigir’ e *loud* ‘alto’; entre (ah) e (oh) como em *hot* ‘quente’ e *law*; e entre (æh), (oh) e (ah). Por exemplo, falantes que tendem a um baixo índice de alçamento de (æh) também tendem a um baixo índice de alçamento de (oh), e aqueles com um alto índice de alçamento de uma também tendem a ter um alto índice para a outra; além disso, os falantes tendem a se agrupar de acordo com seu gênero e etnia (Labov 2006 [1966]: 363; ver Figura 1). Com tais análises, Labov demonstra que a organização do espaço vocálico dentro do sistema fonológico do inglês de Nova Iorque segue o princípio de economia funcional proposto por André Martinet, uma tendência à simetria e ao espaçamento regular entre unidades fonêmicas.

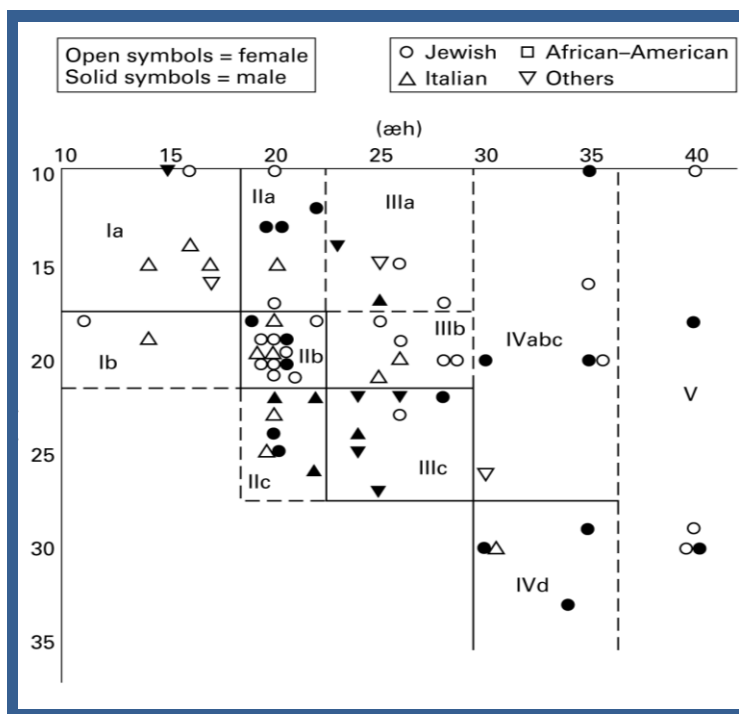
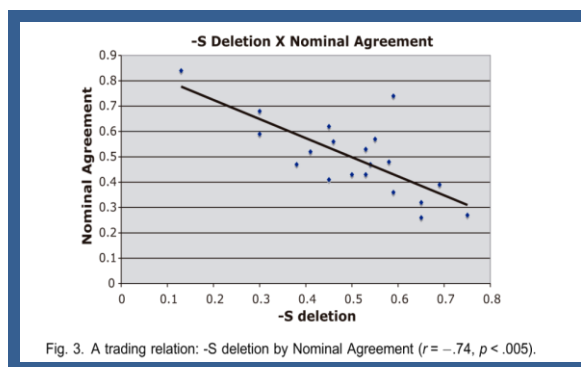


Figura 1: Covariação entre (æh) e (oh) no inglês de Nova Iorque
(Fonte: Labov 2006 [1966]: 363)

Mais recentemente, Guy (2013) demonstra que a covariação também ocorre entre variáveis de diferentes níveis linguísticos, de natureza fonológica e sintática. Em seu estudo sobre covariação no Português do Rio de Janeiro, Guy mostra que o apagamento de (-s) em coda (p.ex. *menos* vs. *meno*) covaria com a concordância nominal (quanto maior a tendência ao apagamento, menor a tendência à concordância), e que a desnasalização (p. ex. *vagem* [va.gĩ] vs. [va.gi]) covaria com a concordância verbal de 3PP (quanto maior a tendência à desnasalização, menor a tendência à concordância verbal). Essas correlações estão ilustradas na Figura 2. Em ambos os casos, também há motivações internas à língua para prever que esses pares de variáveis se correlacionem, uma vez que, em Português, a marca de plural em nomes frequentemente se realiza com -s, e a marca de 3PP nos verbos em muitos casos se realiza com a nasalização.



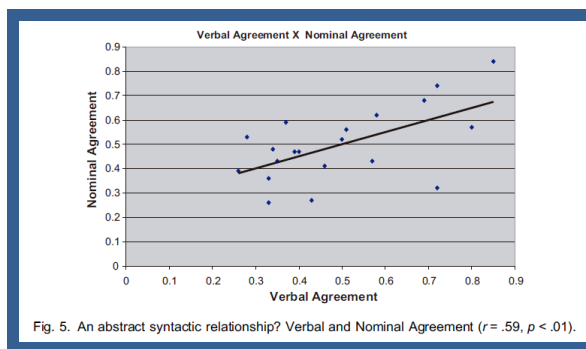
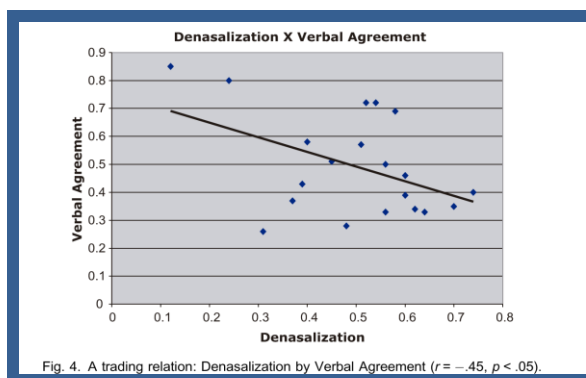


Figura 2: Covariação entre (i) apagamento de *-s* e concordância nominal (acima); (ii) desnasalização e concordância verbal de 3PP (meio); (iii) e concordância nominal e concordância verbal (abaixo).
(Fonte: Guy 2013:67-68)

A Figura 2 também ilustra a correlação entre a concordância nominal e a concordância verbal de 3PP no Português Carioca, um par de variáveis que também concerne ao presente estudo. As duas variáveis se correlacionam nos usos dos falantes, de acordo com o coeficiente $r = .59$ e o diagrama de dispersão: quanto maior a tendência à concordância verbal, maior também é a tendência à concordância nominal. De fato, Guy (2013) demonstra que os três pares de variáveis que potencialmente possuem motivações linguísticas para covariar se correlacionam significativamente (apagamento de *-s*/concordância nominal; desnasalização/concordância verbal; e concordância nominal/concordância verbal); por outro lado, de três pares estruturalmente não relacionados, somente um apresenta uma correlação significativa (desnasalização/concordância nominal).

Guy (2013) também investigou amplos padrões de agrupamento das variantes ao classificar os usos dos falantes em (A), médio (M) ou baixo (B). Pode-se inferir maior grau de coesão social se as quatro variáveis estiverem na mesma categoria (AAAA, MMMM, BBBB) para um grande número de indivíduos. Embora tenha constatado uma distribuição não aleatória, em que 50% dos falantes têm tendências semelhantes para três ou quatro das variáveis, os resultados não são fortemente conclusivos: os outros 50% de sua amostra têm tendências iguais apenas para duas das quatro variáveis – algo que forçosamente deve ocorrer em uma classificação ternária dos usos dos falantes, já que uma categoria deverá se repetir.

O presente trabalho busca avançar a investigação sobre covariáveis, sobretudo quanto à atuação de fatores sociais e linguísticos que promovem ou enfraquecem as correlações observadas entre pares de variáveis dependentes.

2. CORPUS, MÉTODO E HIPÓTESES

A presente análise se baseia em um *corpus* composto de 118 gravações de entrevistas sociolinguísticas com falantes nascidos na cidade de São Paulo, coletadas entre 2009 e 2013 pelo Grupo de Estudos e Pesquisa em Sociolinguística (GESOL-USP) e que fazem parte do Projeto SP2010 – Amostra da Fala Paulistana (Mendes & Oushiro 2013).² As entrevistas são estratificadas de acordo com quatro variáveis sociais: o sexo/gênero dos falantes (F: feminino; M: masculino); três faixas etárias (1: 20-34 anos; 2: 35-59 anos; 3: 60 anos ou mais); dois níveis de escolaridade (M: até Ensino Médio; S: Ensino Superior); e duas regiões de residência na cidade (C: bairros mais centrais; P: bairros mais periféricos). O cruzamento desses parâmetros resulta em 24 perfis sociolinguísticos (F1MC, F1MP, F1SC etc.), para cada um dos quais o *corpus* conta com 3-5 falantes. Trata-se, portanto, de uma amostra bastante representativa do Português Paulistano. As três variáveis aqui analisadas estão exemplificadas em (1)-(3).

- (1) Concordância nominal (CN)
(os) – *meninos* vs. (os) – *menino*-Ø
- (2) Concordância verbal de 3PP (CV-3PP)
eles falaram vs. *eles falou*
- (3) Concordância verbal de 1PP (CV-1PP)
nós falamos vs. *nós falou*

A concordância nominal foi tratada a partir da perspectiva atomística (Scherre 1988), em que cada palavra do sintagma nominal que é flexionável quanto ao número é considerada uma unidade de análise. Tais dados foram codificados de acordo com a presença ou não de marca explícita de número. Desses, a análise final excluiu os dados em primeira posição (p.ex. “os” em “os meninos”) devido à presença praticamente categórica de marca explícita de plural nesses casos (99,7%). No total, somaram-se 19.884 ocorrências, das quais 2.786 (14,0%) apresentaram a marca zero.

A análise sobre as concordâncias verbais considera os sintagmas de 3PP e de 1PP com sujeitos pronominais (*eles, elas, nós*) e não pronominais (p.ex., *meus filhos, minha irmã e eu*), explícitos na mesma sentença ou em sentença com estrutura paralela (p.ex. *eles foram lá e falaram com ela*). Excluem-se sujeitos de 3PP ditos indeterminados (p.ex., *já roubaram meu carro duas vezes*) e sentenças com os verbos *ter, vir e por* e formas derivadas (*manter, propor* etc.) nos casos em que as formas de singular e plural não se distinguem fonologicamente na 3PP (cf. *tem/têm*). Tampouco se consideram SVs com o pronome *a gente*; com os verbos *haver e ter* com sentido existencial (*tinha/tinham muitas pessoas lá*); sujeitos partitivos (*a maioria das pessoas foi/foram*); o chamado infinitivo flexionado opcional (Canever 2012; *nós temos tudo para crescer/crecermos como nação*); e construções impessoais como *faz/fazem dez anos*. Embora a variação seja possível nesses casos, a marca zero é aquela considerada “padrão” de acordo com as gramáticas normativas (ver, p.ex., Bechara 2005, Cunha & Cintra 2007). Os dados foram codificados de acordo com a presença ou não de morfema de plural no verbo.³

² Parte dessas gravações são disponibilizadas gratuitamente no portal do projeto: <http://projetosp2010.fflch.usp.br/>. Último acesso em 14.ago.2014.

³ Não se analisa a 2PP (*vocês vão* vs. *vocês vai*) pelo número bastante reduzido de dados: no *corpus* de 118 entrevistas, houve apenas 103 ocorrências de sintagmas verbais de 2PP.

Houve um total de 9.480 dados para 3PP, dos quais 1.191 (12,6%) com marca zero de número, e 1.074 ocorrências de 1PP, sendo a variante não padrão 101 (9,4%) dos casos.

Cada variável foi de início analisada separadamente em modelos de efeitos mistos no programa R (R Core Team 2013), através do pacote Rbrul (Johnson 2009)⁴. Em modelos de efeitos mistos, é possível incluir dois tipos de variáveis independentes no modelo estatístico: efeitos fixos e efeitos aleatórios. Os primeiros são variáveis cujos exemplares da amostra são representativos da população amostrada, como o sexo e a classe social do informante, ou a classe morfológica da palavra. Os efeitos aleatórios, por sua vez, referem-se a variáveis específicas da amostra analisada, como, por exemplo, os falantes individuais ou os itens lexicais que fazem parte dos dados coletados. A inclusão de efeitos aleatórios no modelo estatístico permite verificar se as correlações observadas se devem aos efeitos fixos – em geral, o principal foco de interesse do estudo, já que permitem chegar a generalizações – ou se possivelmente se devem à contribuição de certos indivíduos ou itens lexicais que eventualmente se comportam de modo distinto da população em geral (Baayen 2008, Johnson 2009). As análises de cada variável incluíram Falante e Item Lexical como efeitos aleatórios, e um conjunto de variáveis sociais e linguísticas como efeitos fixos.

As variáveis sociais (Quadro 1) incluem, além das quatro variáveis que estratificam a amostra (Sexo/Gênero, Faixa Etária, Nível de Escolaridade e Região de Residência), a Classe Social dos falantes (divididos em A-B1, B2, C1 e C2-D, a partir de índice que considera a renda, nível de escolaridade, ocupação, região de residência e condição dos pais – ver Oushiro 2015), Origem dos Pais (São Paulo-capital, interior, Norte/Nordeste, estrangeira e mista), e Mobilidade (baixa: sempre morou no mesmo bairro; média: mudou-se para outro bairro da mesma zona; alta: já morou em diferentes zonas da cidade).

Variável	Variantes
Sexo/Gênero	Feminino / Masculino
Faixa etária	1ª faixa etária: 20 a 34 anos 2ª faixa etária: 35 a 59 anos 3ª faixa etária: 60 anos ou mais
Nível de escolaridade	Até Ensino Médio / Ensino Superior
Região de residência	Bairro mais central / Bairro mais periférico
Classe social	Classe alta e média alta – A-B1 Classe média média – B2 Classe média baixa – C1 Classe baixa alta e baixa média – C2-D
Origem dos pais	São Paulo - capital Interior de SP e MG Norte/Nordeste Estrangeira Mista
Mobilidade geográfica	Baixa: sempre morou no mesmo bairro Média: mudou-se, mas sempre morou na mesma zona Alta: morou em diferentes zonas
Falante (efeito aleatório)	

Quadro 1: Variáveis sociais incluídas nas análises multivariadas de CN, CV-3PP e CV-1PP

⁴ Para uma análise detalhada sobre cada variável, ver Oushiro (2015).

O conjunto de variáveis linguísticas incluídas em cada análise difere, evidentemente, para cada variável dependente. A análise sobre CN (Quadro 2) incluiu a Saliência Fônica, a Tonicidade da Palavra, a Posição Relativa ao Núcleo, a Classe Morfológica, o Contexto Fônico Seguinte e o Paralelismo Intrassintagmático.

Variáveis	Variantes
Saliência Fônica	Vogal oral /s/, como em “casa, casas” Vogal nasal /s/, como em “homem, homens”, “mão, mãos” -r /is/, como em “professor, professores” -l /is/, como em “azul, azuis”, “pastel, pasteis” -z /is/, como em “mês, meses” -ão /õjs/, como em “profissão, profissões” -ão /ãjs/, como em “pão, pães” Metafônicos, como em “ovo, [ó]vos”
Tonicidade do Item Singular	Monossílabo tônico, como em “vez, vezes” Monossílabo átono, como em “o, os” Polissílabo oxítono, como em “morador, moradores” Paroxítono, como em “ano, anos” Proparoxítono, como em “quilômetro, quilômetros”
Posição Relativa	Pré-nuclear, como em “todas <i>as</i> pessoas mais velhas” Nuclear, como em “todas <i>as pessoas</i> mais velhas” Pós-nuclear, como em “todas <i>as pessoas</i> mais <i>velhas</i> ”
Classe Morfológica	Substantivo, como em “meus <i>amigos</i> , os <i>parques</i> ” Adjetivo, como em “placas <i>informativas</i> , países <i>européus</i> ” Quantificador indefinido, como em “essas <i>várias</i> pessoas” Demonstrativo, como em “todas <i>essas</i> coisas” Possessivo, como em “amigos <i>meus</i> , os <i>seus</i> direitos” Artigo definido, como em “todos <i>os</i> dias” Pronome, como em “todos <i>eles</i> ”
Contexto Fônico Seguinte	Consoante, como em “as <i>meninas</i> foram” Vogal, como em “as <i>meninas</i> então” Pausa, como em “as <i>meninas</i> ... então”
Paralelismo Intrassintagmático (somente para 3ª, 4ª e 5ª posições)	Precedida por palavra com marca zero, como em “nos <i>último-Ø</i> anos” Precedida por palavra com marca de plural ou invariável, como em “nos <i>últimos</i> anos, uns <i>dezesete</i> anos”
Item lexical (efeito aleatório)	

Quadro 2: Variáveis linguísticas incluídas na análise multivariada de CN

Para CV-3PP, analisaram-se a Saliência Fônica (de acordo com a escala proposta por Naro 1981), o Paralelismo Discursivo (Scherre & Naro 1992, Scherre 1998), o Tipo de Sujeito, a Posição do Sujeito e a Animacidade do Sujeito. A análise de CV-1PP incluiu as mesmas variáveis de CV-3PP, exceto Animacidade, uma vez que o sujeito de 1PP sempre tem o traço [+humano].

Variável	Variantes - 1PP	Variantes - 3PP
Saliência fônica (Naro 1981, Naro <i>et al</i> 1999)		
Grau 1 – oposição não acentuada		
1a – nasalização de [i]	NA ⁵	<i>come/comem</i>
1b – nasalização e mudança da qualidade vocálica	NA	<i>fala/falam</i>
1c – adição de segmento	<i>falava/falávamos</i>	<i>faz/fazem</i>
Grau 2 – oposição acentuada		
2a – oposição –V/-V-mos acentuada em uma das formas	<i>fala/falamos</i>	NA
2a' – 1PP: oposição –V/-V-mos acentuada em ambas as formas 3PP: nasalização e ditongação da vogal tônica	<i>está/estamos</i>	<i>dá/dão</i>
2b – 1PP: forma singular contém ditongo crescente, adição do segmento [-mus] 3PP: forma singular contém ditongo crescente, adição do segmento [rũ]	<i>comeu/comemos</i>	<i>comeu/comeram</i>
2c – mudança da vogal tônica	<i>falou/falamos</i>	<i>falou/falaram; é/são; veio/vieram</i>
Paralelismo discursivo		
precedido de SV com concordância	(de acordo com a presença de outros SVs plurais nas cinco sentenças precedentes)	
precedido de SV sem concordância		
isolado ou primeiro de uma série		
Tipo de sujeito		
composto	<i>minha esposa e eu casamos/casou</i>	<i>a minha mãe e o meu pai casaram/casou</i>
sintagma nominal simples	NA	<i>os meus pais casaram/casou</i>
pronomes	<i>nós casamos/casou</i>	<i>eles/elas casaram/casou</i>
Posição do sujeito		
imediatamente precedente	<i>nós fomos/foi</i>	<i>eles vieram/veio na minha direção</i>
precedente (1-4 sílabas)	<i>nós ainda não casamos/casou</i>	<i>os equipamentos que tem lá são/é importados</i>
precedente distante (5+ sílabas)	<i>nós com o estilingue com mamona atirávamos/atirava no pênis do jumento</i>	<i>as pessoas lá de Goiás mesmo que vieram/veio pra cá</i>

⁵ NA: variante não se aplica à variável.

posposto	<i>aí íamos/ia sempre <u>nós...</u> nós quatro pra lá</i>	<i>chegaram/chegou <u>dois caras</u> assim do nada</i>
pergunta do documentador	NA	D1: <i><u>seus pais</u> moravam no Bom Retiro?</i> S1: <i>moravam no Bom Retiro</i>
Animacidade do sujeito		
[+humano]	NA	<i><u>meus amigos</u> são/é todos daqui</i>
[-humano]	NA	<i><u>os carros</u> geralmente não param/para né?</i>
Item lexical – verbo (efeito aleatório)		

Quadro 3: Variáveis linguísticas incluídas nas análises multivariadas de CV-1PP e CV-3PP

Da perspectiva da coesão social, as três variáveis compartilham as mesmas correlações. A Tabela 1 apresenta um resumo dos resultados para os fatores sociais, que foram obtidos nas análises separadas de cada variável sociolinguística. Em todas as análises, o valor de aplicação foi a marca zero de número. Os fatores sociais são indicados como não selecionados pelo Rbrul “--”, ou pelo fator que favorece o valor de aplicação (“Masculino”, “Ens. Médio” etc.). Entre parênteses se encontra o valor de *range* (diferença entre maior e menor pesos relativos), que sinaliza a importância relativa da variável social para a respectiva variável dependente.

	CN-Ø	CV-3PP-Ø	CV-1PP-Ø
Sexo/Gênero	Masculino (18)	Masculino (13)	Masculino (62)
Faixa Etária	--	--	--
Escolaridade	Ens. Médio (53)	Ens. Médio (42)	Ens. Médio (70)
Região de Residência	--	--	--
Classe Social	C2-D (64)	C2-D (50)	C2-D (81)
Origem dos Pais	--	--	--
Mobilidade	Baixa (18)	--	--

Tabela 1: Correlações com variáveis sociais
(--: variável não selecionada pelo Rbrul; (00): *range* da variável)

Os resultados gerais mostram que, em São Paulo, a marca zero de número é favorecida pelos homens, falantes com menor nível de escolaridade e de classe social mais baixa em todos os casos. Para CN, há uma correlação adicional com a Mobilidade do falante: aqueles que sempre moraram no mesmo bairro favorecem a marca zero em relação aos falantes com maior mobilidade, que já viveram em bairros ou zonas diferentes. Faixa Etária, Região de Residência e Origem dos Pais não se correlacionam significativamente com nenhuma das três variáveis.

Ao comparar os valores de *range*, a ordem relativa de importância dos fatores sociais é bastante semelhante: Classe Social é sempre selecionada em primeiro lugar, seguida de Escolaridade e Sexo/Gênero (além de Mobilidade para CN). Assim, da perspectiva de semelhanças sociais, a seleção das mesmas variáveis e na mesma hierarquia de importância conduz à expectativa de que as três variáveis se correlacionem fortemente nos usos individuais

dos falantes – aqueles que tendem a empregar a marca zero em uma das variáveis também devem empregar a marca zero nas outras duas.

Estruturalmente, as três variáveis claramente se relacionam. CV-3PP e CV-1PP se referem à morfologia de número no paradigma verbal, que tende a formar um sistema rigidamente organizado, em que as regras se definem por analogia paradigmática. Duarte & Varejão (2013) propõem que as mudanças na morfologia flexional verbal no Português Brasileiro, de um sistema de seis para duas formas (oposição entre a primeira pessoa do singular e as demais pessoas gramaticais), conduziu a uma mudança paramétrica no Português Brasileiro, de língua de sujeito nulo para uma não pro-drop, em que pessoa e número são marcados pelo pronome em vez da flexão verbal. Se a marcação número-pessoal se relaciona fundamentalmente a duas gramáticas distintas, pode-se esperar grande pressão interna do sistema para a simetria entre 3PP e 1PP. Assim, de um ponto de vista puramente estrutural, espera-se que a concordância de 3PP e de 1PP estejam fortemente correlacionadas.

A Tabela 2 resume os resultados sobre as variáveis linguísticas nas análises separadas das concordâncias verbais. Verifica-se que Saliência Fônica, Posição e Tipo de Sujeito se correlacionam com ambas, na mesma hierarquia de importância (comparem-se os valores de *range* entre parênteses). Animacidade do Sujeito se correlaciona com 3PP mas, conforme se apontou acima, tal variável não foi incluída na análise de 1PP. A única diferença entre os condicionamentos internos das variáveis se refere a Paralelismo Discursivo, relevante para 3PP mas não selecionada para 1PP. Essa diferença, no entanto, possivelmente se deve a um número reduzido de ocorrências de SVs precedidos por SVs com marca zero para 1PP.

	CV-3PP-Ø	CV-1PP-Ø
Saliência Fônica	Grau 1a – come/comem (49)	Grau 1 – falava/falávamos (90)
Paralelismo Discursivo	SV precedido de SV com marca zero (40)	--
Posição do Sujeito	Posposto (38)	Posposto (59)
Animacidade do Sujeito	[-humano] (19)	NA
Tipo de sujeito	Sintagma nominal (18)	Sintagma nominal (43)

Tabela 2: Correlações de CV-3PP e CV-1PP com variáveis linguísticas
(--: variável não selecionada pelo Rbrul; (00): *range* da variável)

A concordância nominal, de modo semelhante a CV-3PP e CV-1PP, também envolve um processo em que a pluralidade pode ser marcada de maneira não redundante dentro do sintagma. Na Tabela 3, que apresenta os resultados para CN, podem-se destacar dois condicionamentos linguísticos semelhantes ao que ocorre com as concordâncias verbais: Saliência Fônica (analisada de acordo com os processos morfofonológicos de formação de plural) e Paralelismo (de acordo com a presença ou ausência de marca explícita de número no vocábulo precedente); formas menos salientes e ocorrências precedidas de marca zero favorecem a variante não padrão para CN. Desse modo, de um ponto de vista estrutural, também se pode esperar que CN covarie com CV-3PP e CV-1PP.

	CN-Ø
Saliência Fônica	Vogal oral + s (37)
Posição Relativa	Nuclear ou pós-nuclear (40)
Contexto Fônico Seguinte	Consoante ou pausa (5)
Paralelismo Intrassintagmático	SN precedido por SN com marca zero (53)
Classe Morfológica	Substantivo (42)

Tabela 3: Correlações de CN com variáveis linguísticas⁶
 (--: variável não selecionada pelo Rbrul; (00): range da variável)

Para examinar se tais previsões sociais e estruturais se observam empiricamente nos usos dos falantes, o presente estudo analisou os pesos relativos de cada falante para cada variável, que foram obtidos através da variável Falante da análise multivariada. Tais medidas foram primeiramente analisadas em termos de amplos padrões de agrupamento social. As tendências de cada indivíduo foram classificadas como alta (A), média (M) ou baixa (B) para cada uma das três variáveis, de modo semelhante ao estudo de Guy (2013); haverá um alto grau de coesão dialetal se todas as variantes se encaixarem na mesma categoria (p.ex., AAA ou BBB) para a maioria dos indivíduos da amostra.

A covariação também foi analisada através do cálculo de coeficientes de correlação de Pearson (ou *r* de Pearson), a partir dos mesmos pesos relativos. Os valores de *r* vão de -1 a +1 e indicam correlação negativa quando próximos de -1 (quanto mais *x*, menos *y*), correlação positiva quando próximos de +1 (quanto mais *x*, mais *y*) e não correlação quando próximos de zero. Tais medidas foram calculadas com auxílio da função `pairscor.fnc` do pacote `languageR` (Baayen 2008) no programa R.

Por fim, coeficientes de Pearson também foram calculados por grupos sociais (sexo feminino, masculino, classe A-B1, B2, C1, C2-D etc.) e por fatores linguísticos (especificamente, a Saliência Fônica), a fim de melhor determinar o efeito de condicionamentos sociais e estruturais na covariação. A comparação entre os valores de *r* para fatores distintos (p.ex., mulheres vs. homens) foi realizada no programa R através da função `r.test` do pacote `psych` (Revelle 2014).

3. ANÁLISES

3.1. Padrões gerais

As tendências de uso dos falantes foram classificadas em alta (A), média (M) ou baixa (B) para cada uma das três variáveis, de acordo com seus pesos relativos: abaixo de 0.40, entre 0.40 e 0.60 ou acima de 0.60, respectivamente. Ao atribuir A, M ou B, há 27 padrões possíveis em que um falante pode ser classificado ($3^3 = 27$): AAA, AAM, AAB etc. Em um cenário de completa aleatoriedade, a expectativa seria a de uma distribuição equilibrada do número de falantes por padrão – uma média de 4,4 (118 falantes / 27 padrões) para a presente amostra. No cenário oposto, os falantes seriam classificados nas mesmas poucas categorias. Como

⁶ Posição Relativa e Classe Morfológica foram analisadas em rodadas distintas, já que as variáveis não são independentes entre si (Guy & Zilles 2007): possessivos geralmente ocupam posições pré-nucleares, adjetivos tendem a ocorrer em posições pós-nucleares, e substantivos sempre ocupam a posição nuclear. Paralelismo Intrassintagmático, por sua vez, também foi analisado em rodada distinta, que incluiu apenas os dados em 3ª, 4ª e 5ª posições do SN; visto que, em 1ª posição, as ocorrências apresentam marcas semicategoriais de número, faz-se necessária a exclusão de dados de 2ª posição dessa variável.

normalmente é o caso quando se opera com dados empíricos, a real distribuição é um cenário intermediário: neste estudo, ocorreram 21 dos 27 padrões possíveis, e é pouquíssimo provável que a distribuição dos falantes nessas categorias tenha sido aleatória ($p < 0,001$). A Tabela 4 mostra os dez padrões mais frequentes.

	N	CN-Ø	3PP-Ø	1PP-Ø
(i)	22	B	B	B
(ii)	16	A	A	B
(iii)	15	A	A	A
(iv)	8	A	M	B
(v)	8	M	M	B
(vi)	7	M	B	B
(vii)	6	B	M	B
(viii)	4	A	M	A
(ix)	2	A	B	B
(x)	2	M	A	A

Tabela 4: Padrões mais frequentes de uso de CN-Ø, 3PP-Ø e 1PP-Ø

(A: P.R. > 0.60; M: 0.40 < P.R. < 0.60; B: P.R. < 0.40.)

O padrão que ocorreu mais vezes é aquele em que as tendências de emprego de marca zero são baixas para as três variáveis de concordância (BBB, N = 22 falantes). O padrão oposto AAA, de altas tendências de emprego da variante, também está entre os mais recorrentes (N = 15). Dos dez padrões mais frequentes, nove apresentam tendências iguais para pelo menos duas variáveis (AA ou BB); a exceção é o padrão (iv), AMB. É interessante notar que, quando há tendências diferentes, o uso de 1PP-Ø parece ser normalmente mais baixo (M ou B) do que para CN-Ø ou para 3PP-Ø (padrões ii, iv, v, vi, vii e ix).

Embora tal distribuição tenha baixa probabilidade de ser aleatória, ela não é fortemente persuasiva. Tendências idênticas de emprego para as três variáveis (AAA e BBB) ocorrem apenas para 37 falantes, o que equivale aproximadamente a um terço da amostra. Os demais dois terços se distribuem por diversos outros padrões, o que parece apontar para uma baixa coesão da comunidade nos usos linguísticos de variáveis distintas.

Os coeficientes de correlação de Pearson revelam um quadro mais detalhado da covariabilidade ao considerar não apenas uma tendência ternária de uso (A, M ou B), mas também a força da correlação entre cada par de variáveis. A Figura 3 é uma matriz de correlações que contém os três pareamentos possíveis entre as variáveis. Os quadros na linha diagonal mostram a distribuição dos pesos relativos dos falantes, dividida em decis (0 a 0.10, 0.11 a 0.20 etc.); os quadros do canto inferior esquerdo mostram os coeficientes de correlação (r) junto a seus respectivos valores de significância (p); por fim, os quadros do canto superior direito mostram os gráficos de dispersão para cada par de variáveis e suas respectivas linhas de regressão.

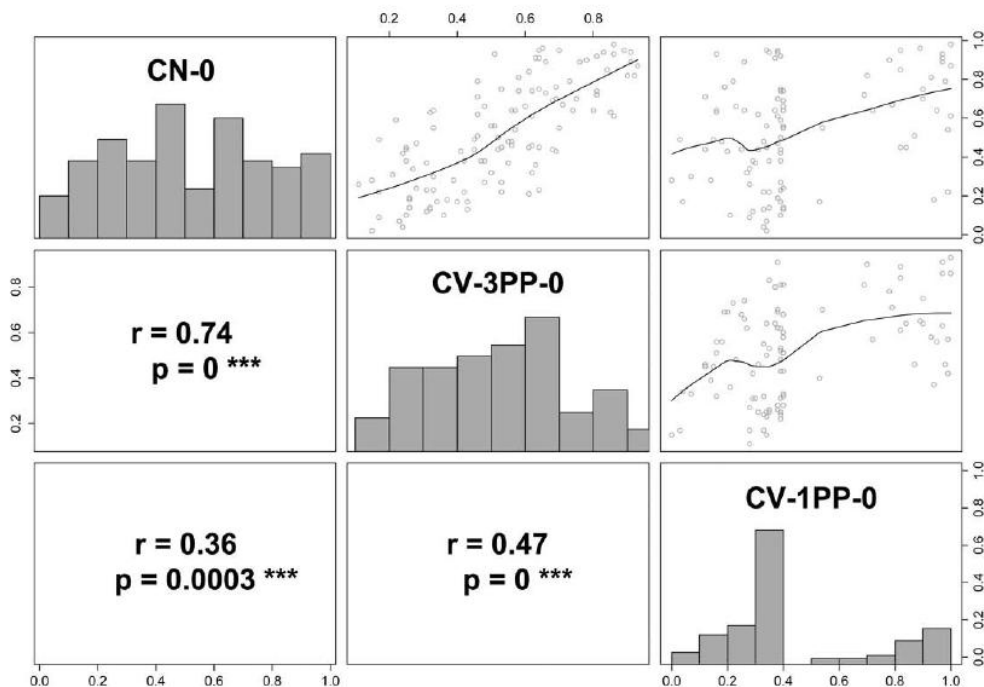


Figura 3: Matriz de correlações entre variáveis

Os resultados mostram que todos os pares de variáveis covariam significativamente, com valor de p abaixo de 0,001 (ou seja, é baixíssima a probabilidade, menor do que uma chance em mil, de que as distribuições observadas tenham ocorrido aleatoriamente). O emprego de CN-Ø se correlaciona com o emprego de CV-3PP-Ø ($r = 0.74$) e com CV-1PP-Ø ($r = 0.36$), e as duas concordâncias verbais também se correlacionam entre si ($r = 0.47$). Isso indica que falantes que tendem a empregar a forma não padrão de uma variável também tendem a fazê-lo para as outras duas. Isso não é muito surpreendente ao contrastar esse resultado com as correlações sociais verificadas, para cada uma dessas variáveis, com classe social, escolaridade e sexo/gênero dos falantes. Os mesmos grupos sociais tendem a apresentar correlações bastante semelhantes com cada uma dessas variáveis.

No entanto, a Figura 3 mostra que a correlação mais forte entre esses três pares de variáveis é aquela entre CN e CV-3PP, e não entre CV-1PP e CV-3PP, do modo como uma perspectiva puramente estrutural poderia prever. O gráfico de dispersão para a correlação entre CN e CV-3PP mostra uma linha de regressão bastante mais inclinada e regular, o que significa que o uso de uma variante prediz o uso da outra com alto grau de confiabilidade. Contudo, o que se verifica na correlação entre as concordâncias verbais é algo diferente. Cabe, portanto, examinar tais correlações mais detalhadamente.

Um olhar mais atento à correlação entre CV-1PP-Ø e CN-Ø, bem como a de CV-1PP-Ø e CV-3PP-Ø, mostra que as linhas de regressão são relativamente planas no lado esquerdo do gráfico de dispersão e se inclinam mais propriamente apenas no lado direito (Figura 4).

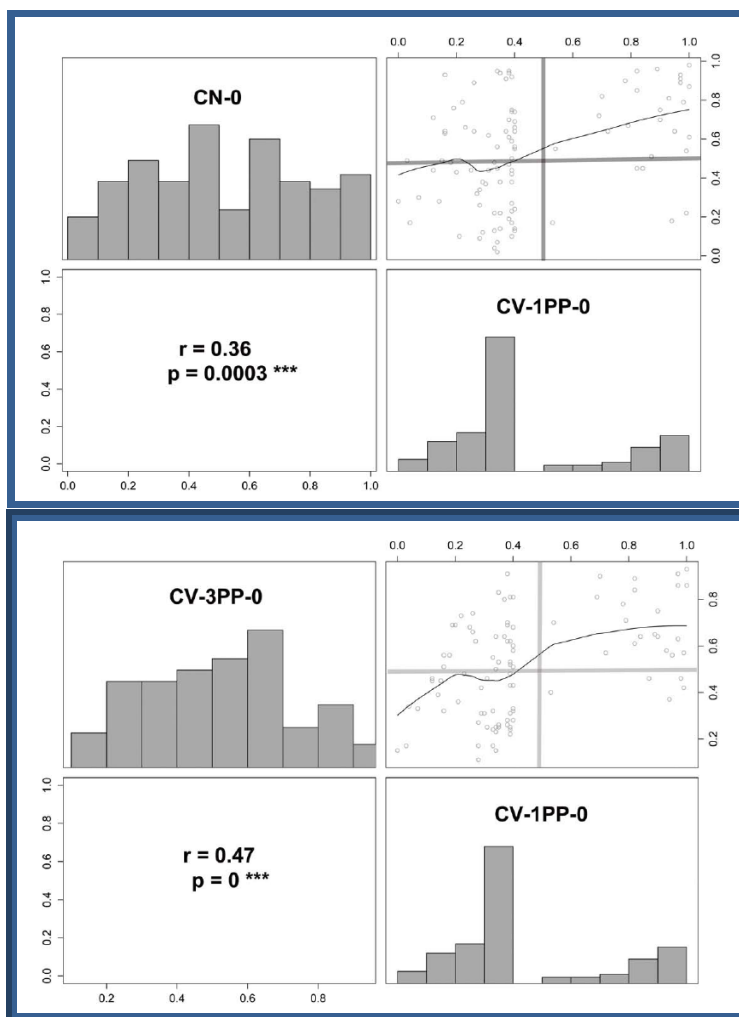


Figura 4: Correlações entre CN-Ø (acima)/CV-3PP-Ø (abaixo) e CV-1PP-Ø

Em cada um desses gráficos, os pontos no canto superior direito representam os falantes que tendem a empregar a marca zero de duas variáveis (pesos relativos acima de 0.50 para ambas), enquanto aqueles no canto inferior esquerdo são os falantes que tendem a empregar a forma padrão de ambas (pesos relativos abaixo de 0.50). Os cantos superior esquerdo e inferior direito, por sua vez, representam falantes que têm tendências opostas para cada variável (favorecimento de uma variante e desfavorecimento de outra). Note-se que, dentre os falantes que tendem a empregar a marca zero de 1PP (lado direito dos gráficos de dispersão), há muito mais falantes no quadrante superior do que no inferior. Diferentemente, os falantes que tendem a empregar a forma padrão de 1PP (lado esquerdo dos gráficos de dispersão) se distribuem mais equilibradamente entre os quadrantes superior e inferior, ou seja, entre aqueles que empregam a forma padrão ou não padrão de CN e de 3PP. Isso significa que, se um falante tende a empregar a forma *não padrão* de 1PP, há uma probabilidade bastante alta de que ele também empregue a forma *não padrão* de CN ou de 3PP; por outro lado, se um falante tende a empregar a forma

padrão de 1PP, tal fato não prevê como ele se comporta em relação a CN e CV-3PP. Decorrem daí as correlações mais fracas dos pares com 1PP-Ø.

Desse modo, os três pares de variáveis se correlacionam significativamente, de acordo com as previsões estruturais e sociais de que haveria coesão dialetal em seus usos. Contudo, tais correlações não são tão fortes quanto se previu, tampouco seguem a hierarquia esperada de correlação mais forte entre as concordâncias verbais do que entre essas e a concordância nominal. As correlações relativamente mais fracas nos pares que envolvem a 1PP devem-se ao fato de que o emprego da forma padrão dessa variável não necessariamente implica em uso da forma padrão para CN ou 3PP.

As próximas subseções investigam de modo mais aprofundado a atuação de fatores sociais e linguísticos, a fim de examinar possíveis motivações para as correlações verificadas.

3.2. Fatores sociais

Os coeficientes de correlação de Pearson também foram calculados para subgrupos de falantes, de acordo com as sete variáveis sociais incluídas nas análises multivariadas: Sexo/Gênero, Faixa Etária, Nível de Escolaridade, Região de Residência, Classe Social, Origem dos Pais e Mobilidade.

A Tabela 5 mostra os três pareamentos e seus respectivos coeficientes de correlação para a amostra geral nas primeiras duas linhas, e os coeficientes por subgrupos de falantes nas linhas subsequentes. Para facilitar a visualização, apresentam-se apenas os valores de significância das correlações, representados pelos asteriscos ($*p < 0,05$; $**p < 0,01$; $***p < 0,001$). A última coluna é a soma de asteriscos em cada linha, que indica quantas vezes houve correlações significativas em cada grupo social, bem como a força dessas correlações, como uma medida de quais grupos são mais coesos⁷. As células sombreadas indicam que a correlação para um determinado subgrupo de falantes (p.ex. sexo feminino) é significativamente mais forte do que aquela verificada em outro fator da mesma variável (p.ex. sexo masculino).

	r de Pearson:	CN-3PP	CN-1PP	3PP-1PP	N*
		0.74 ***	0.36 ***	0.47 ***	
Sexo/Gênero	Feminino	***	***	***	9
	Masculino	***		*	4
Faixa Etária	20-34 anos	***		*	4
	35-59 anos	***	*	**	6
	60+ anos	***	**	***	8
Escolaridade	Até Ensino Médio	***	***	***	9
	Ensino Superior	***			3
Região de Residência	Central	***	***	***	9
	Periférica	***		**	5

⁷ Trata-se de uma medida *ad hoc*, uma vez que *significância* não equivale a *tamanho de efeito*. Ver Oushiro (2015) para uma discussão mais detalhada.

Classe Social	A-B1	**			2
	B2	***			3
	C1	***		**	5
	C2-D	***	**	**	7
Origem dos Pais	Norte/Nordeste	*		**	3
	Interior SP/MG	***			3
	Estrangeiros	***			3
	São Paulo	***	**	***	8
Mobilidade	Mista	*			1
	Baixa	***	*	**	6
	Média	***	*	*	5
	Alta	***		*	4

Tabela 5: Correlações de Pearson por fatores sociais (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$. Células sombreadas indicam correlação significativamente mais forte do que em outro subgrupo da mesma variável social)

A forte correlação entre CN-Ø e 3PP-Ø ($r = 0.74$) resulta do fato de que todos os subgrupos da comunidade (mulheres, homens, todas as faixas etárias, os dois graus de escolaridade etc.) se comportam de maneira similar: quando CN-Ø é favorecido, 3PP-Ø também é favorecido, e quando CN-Ø é desfavorecido, 3PP-Ø também é desfavorecido. Tal comportamento gera as correlações significativas em todos os subgrupos de falantes. Não é demais enfatizar que tal correlação entre CN e 3PP *não* significa que todos os falantes de determinado grupo tendem a empregar a mesma variante, mas sim que eles se comportam de modo semelhante ao empregar ou evitar *pares* de variantes – CN-Ø e 3PP-Ø, ou CN padrão e 3PP padrão simultaneamente. As correlações entre CN e 3PP em diversos subgrupos, portanto, apontam para um consenso dentro da comunidade paulistana de que suas respectivas variantes padrão ou não padrão “andam juntas”, independentemente de se o falante favorece uma ou outra variante. Em comparação, as correlações com 1PP são mais fracas porque nem todos os grupos demonstram o mesmo nível de coesão dialetal, como se nota pela ausência de asteriscos em diversas células das colunas CN-1PP e 3PP-1PP.

Quando se examinam as correlações por grupos, parece haver maior coesão entre mulheres em relação aos homens, falantes menos escolarizados em relação aos mais escolarizados, moradores de bairros mais centrais em relação à região periférica, classes mais baixas (C1 e D2-D) em relação às mais altas, filhos de paulistanos em relação a filhos de pais não paulistanos. Isso é indicado pelas correlações significativamente mais fortes, nas células sombreadas. As diferenças por Faixa Etária e Mobilidade não são significativas dentre esses pares de variáveis. O fato de que falantes mais escolarizados, de classe social mais alta, que vivem em regiões mais periféricas e cujos pais não são paulistanos não apresentam o mesmo grau de coesão dialetal possivelmente se deve ao contato com maior variedade de normas sociais linguísticas. Falantes com pais provenientes de outros lugares e aqueles que moram na periferia provavelmente têm maior contato dialetal com outras variedades do Português. Ao mesmo tempo, embora Nível de Escolaridade e Classe Social sejam independentes nesta amostra, trata-se de duas variáveis que

interagem na experiência cotidiana dos membros da comunidade: falantes com maior nível de escolaridade geralmente pertencem a classes mais altas, cuja mobilidade social e demográfica possivelmente é maior do que o de classes mais baixas. Embora as diferenças para Mobilidade não tenham se revelado significativas, falantes com maior mobilidade também parecem tender a uma menor coesão em seus usos linguísticos. Tais considerações, no entanto, não explicariam o porquê de as mulheres terem, aparentemente, maior coesão dos que os homens. Essas são hipóteses que devem ser examinadas em estudos futuros, com um conjunto maior de variáveis sociolinguísticas e em outras comunidades, a fim de avaliar sua validade.

3.3. Fatores linguísticos: o papel da Saliência Fônica

Conforme se sugeriu, pode haver outros fatores linguísticos que condicionam a covariação para além da similaridade estrutural entre variáveis dependentes. A fim de separar o efeito de diferentes fatores linguísticos, rodaram-se novas análises multivariadas em subconjuntos de dados, subdividindo-os em fatores internos; aqui, discutem-se os resultados para a divisão das ocorrências quanto à Saliência Fônica: o fato de que CV-1PP estabelece correlações mais fracas com CN e CV-3PP levantou a hipótese de que a covariação possa ser influenciada pelo grau de saliência das variantes. Tais análises geraram novos valores de pesos relativos para cada falante, que foram então empregados no cálculo de novos coeficientes de correlação de Pearson. A questão aqui é se diferentes fatores geram correlações significativamente diferentes entre variáveis; se esse for o caso, a variável linguística em questão tem uma influência na covariação.

A Saliência Fônica (Naro 1981, Scherre 1988, Naro *et al* 1999) é uma variável linguística já bastante analisada em estudos sobre concordâncias. Embora os fatores sejam diferentes para CN, 3PP e 1PP, a hipótese que subjaz a essa variável é a mesma em todos os casos: formas menos salientes tendem a favorecer a marca zero, e formas mais salientes a desfavorecê-la. O modo como essa variável foi codificada para cada uma das três variáveis está exemplificado nos Quadros 2 e 3 acima. Nas novas análises de covariação, os fatores foram reorganizados em dois níveis nas três variáveis, como formas menos e mais salientes. Para CN, opuseram-se as formas regulares de plural (terminadas em vogal oral ou nasal) às formas irregulares (terminadas em $-r$, $-l$, $-s$, e plurais em $/\tilde{o}js/$, $/\tilde{a}js/$ e metafônicos). Para 3PP, contrastaram-se o Grau 1 (oposição não acentuada) ao Grau 2 (oposição acentuada); e para 1PP, também o Grau 1 (oposição não acentuada) ao Grau 2 (oposição acentuada).

As Tabelas 6 a 8 mostram, na segunda coluna, os coeficientes de correlação da análise geral como uma base de comparação. Os valores de r de Pearson para formas menos e mais salientes encontram-se na terceira e quarta colunas. Na última, apresenta-se um valor de significância que avalia a probabilidade de a diferença entre as correlações ser aleatória ou não.

Ao comparar os dados menos e mais salientes de CN (Tabela 6), embora os valores de r pareçam baixar para as correlações envolvendo as formas mais salientes (*cf.* $0.72 \rightarrow 0.65$ com 3PP, e $0.35 \rightarrow 0.31$ com 1PP), tais mudanças não são tão fortemente significativas e podem ter ocorrido aleatoriamente ($p > 0,05$).

	Todos os dados	-saliente	+saliente	-saliente vs. +saliente
	N = 19.884	N = 16.460	N = 3.424	<i>p</i>
CN-Ø	NA	NA	NA	NA
3PP-Ø	0.74***	0.72***	0.65***	0.08
1PP-Ø	0.36***	0.35***	0.31**	0.28

Tabela 6: Comparação de valores de *r* de Pearson entre formas menos e mais salientes de CN, em correlação com outras variáveis linguísticas (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Para 3PP-Ø (Tabela 7), o coeficiente de Pearson fica significativamente mais fraco na correlação entre as formas mais salientes de 3PP e CN-Ø (*cf.* 0.77 → 0.55, $p = 0$), mas não com 1PP-Ø (*cf.* 0.44 → 0.48, $p = 0.26$).

	Todos os dados	-saliente	+saliente	-saliente vs. +saliente
	N = 9.480	N = 5.309	N = 4.171	<i>p</i>
CN-Ø	0.74***	0.77***	0.55***	0.00***
3PP-Ø	NA	NA	NA	NA
1PP-Ø	0.47***	0.44***	0.48***	0.26

Tabela 7: Comparação de valores de *r* de Pearson entre formas menos e mais salientes de CV-3PP, em correlação com outras variáveis linguísticas (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Para 1PP-Ø (Tabela 8), os coeficientes de Pearson se tornam significativamente mais fracos tanto com CN-Ø (*cf.* 0.50 → 0.11, $p = 0,004$) quanto com 3PP-Ø (*cf.* 0.54 → 0.21, $p = 0,009$). Seria possível questionar se os valores de *r* diminuem com formas mais salientes devido a um menor número de dados, como um efeito estatístico colateral de uma amostra menor; contudo, vale notar que enquanto há menos dados de CN mais salientes (*cf.* 16.460 vs. 3.424 na Tabela 6) e 3PP mais salientes (*cf.* 5.309 vs. 4.171 na Tabela 7), há mais dados de 1PP mais salientes (*cf.* 297 vs. 777 na Tabela 8), mas os coeficientes de correlação ainda assim se enfraquecem com as formas mais salientes.

	Todos os dados	-saliente	+saliente	-saliente vs. +saliente
	N = 1.074	N = 297	N = 777	<i>p</i>
CN-Ø	0.36***	0.50***	0.11	0.004**
3PP-Ø	0.47***	0.54***	0.21*	0.009**
1PP-Ø	NA	NA	NA	NA

Tabela 8: Comparação de valores de *r* de Pearson entre formas menos e mais salientes de CV-1PP, em correlação com outras variáveis linguísticas (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$)

Em conjunto, as escalas de Saliência Fônica para CN, 3PP e 1PP mostram que as formas mais salientes tendem a se correlacionar de modo mais fraco com outras variáveis. Quando proposta por Naro (1981) para os morfemas de 3PP, a escala de Saliência Fônica tinha o objetivo de discutir o problema da implementação (Weinreich *et al* 2006 [1968]), o ponto de início de uma mudança sintática. Para a concordância de 3PP, a mudança começou em formas menos

salientes (*come, comem*) e se espalhou pelo sistema verbal seguindo a escala de saliência. A influência dessa variável na covariação demonstra que as escalas também têm um papel sincrônico. É mais provável que um falante empregue a marca zero em formas menos salientes de CN, 3PP e 1PP; alguns falantes, de fato, raramente ou nunca empregam a marca zero em ocorrências mais salientes, o que resulta em correlações mais fracas com outras variáveis nesse contexto. Colocado de outro modo, o uso da marca zero em formas mais salientes implica seu uso em formas menos salientes, o que gera correlações mais fortes entre estas; por outro lado, o uso de marca zero em formas menos salientes *não* implica seu uso em formas mais salientes, o que resulta em correlações mais fracas nesses casos.

Os resultados também mostram que a escala de saliência para 1PP afeta 3PP e CN; aquela para 3PP afeta CN; e aquela para CN não afeta as demais. Esses fatos podem ser explicados por uma escala mais geral de saliência considerando-se as três variáveis morfossintáticas. Os morfemas de plural nominal geralmente envolvem menos material fônico [-s, -is, -õjs, -ãjs] do que os morfemas de 3PP [-ĩ, -ũ, -ãw̃, -e.rũ, -i.rũ, -o.rũ] que, por sua vez, envolvem menos material fônico do que para morfemas de 1PP (*á.va.mos, -a.mos, -e.mos*). A Saliência Fônica, desse modo, afeta somente formas menos salientes de outras variáveis, mas não as mais salientes. Isso também pode explicar por que a 1PP não covaria tão fortemente com 3PP quanto o par 3PP-CN; o maior grau de saliência das ocorrências de 1PP-∅ enfraquece as correlações com outras variáveis.

4. CONSIDERAÇÕES FINAIS

Este artigo descreve padrões de covariação entre três variáveis do Português: a concordância nominal, a concordância verbal de 3PP e a concordância de 1PP. Verificou-se que os três pares de variáveis (CN-3PP, CN-1PP, 3PP-1PP) covariam significativamente, mas que tais correlações não são tão fortes quanto se poderia prever a partir da semelhança dos resultados separados de análises multivariadas, tampouco seguem uma hierarquia coerente de acordo com previsões estruturais, já que a correlação entre concordâncias verbais não é mais forte do que entre CN-3PP.

Na comparação desses padrões, destacaram-se fatores sociais e linguísticos que parecem promover e inibir a covariação, e que podem dar conta das correlações observadas. Socialmente, a coesão dialetal parece estar relacionada a grupos sociais que têm contato com menor variedade de normas linguísticas (moradores de bairros mais centrais, filhos de paulistanos, falantes de classes mais baixas). Linguisticamente, sugeriu-se que a covariação é promovida não apenas por similaridades estruturais entre as variáveis dependentes – por exemplo, as concordâncias e 3PP e de 1PP –, mas também por restrições linguísticas mais gerais que se correlacionam com múltiplas variáveis, como a Saliência Fônica.

Tais considerações estabelecem hipóteses mais específicas a serem testadas em estudos futuros. Será interessante observar o efeito de variáveis como Paralelismo, Posição do Sujeito e Animacidade, que já se revelaram correlacionadas a múltiplas variáveis sociolinguísticas em diversas comunidades. Outras variáveis como Contexto Fônico Precedente e Seguinte, Classe Morfológica ou Frequência do Item Lexical, que também se correlacionam com muitas variáveis, possivelmente também têm influência na covariação entre outros pares de variáveis. A análise de variáveis de diferentes níveis linguísticos (fonológico, morfológico, sintático, discursivo), mesmo que não sejam estruturalmente relacionadas entre si, também pode ser

relevante, com o potencial de iluminar o papel de fatores sociais e linguísticos na covariação; se variáveis dessemelhantes quanto à estrutura mas semelhantes quanto aos condicionamentos sociais covariarem, será possível separar quais efeitos sociais têm maior efeito na coesão dialetal.

Sendo uma área relativamente pouco explorada em estudos sobre a variação e a mudança linguística, a análise da covariação ainda levanta muitas questões e abre novos caminhos de análise nesse campo de estudos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Antonino, Vivian. 2007. *A concordância nominal em predicativos do sujeito e em estruturas passivas no Português Popular do interior do estado da Bahia*. Dissertação de Mestrado, UFBA, Salvador, 119f.
- Baayen, R. Harald. 2008. *Analysing linguistic data: a Practical Introduction to Statistics*. Cambridge, Cambridge University Press.
- Bechara, Evanildo. 2005. *Moderna Gramática Portuguesa*. Rio de Janeiro, Lucerna.
- Bortoni-Ricardo, Stella Maris. 1985. *The urbanization of rural dialect speakers*. Cambridge, Cambridge University Press.
- Braga, Maria Luiza. 1977. *A concordância de número no sintagma nominal no Triângulo Mineiro*. Dissertação de Mestrado, PUC, Rio de Janeiro.
- Brandão, Sílvia Figueiredo & Sílvia Rodrigues Vieira. 2012. Concordância nominal e verbal: contribuições para o debate sobre o estatuto da variação em três variedades urbanas do Português. *Alfa*, vol. 56(3): 1035–1064.
- Canever, Fernanda. 2012. *Evidências para um modelo de língua baseado no uso: o infinitivo flexionado em Português Brasileiro*. Dissertação (Mestrado em Linguística), Universidade de São Paulo/FFLCH, São Paulo, 165f.
- Coelho, Rafael F. 2006. *É nós na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana. O pronome de primeira pessoa do plural e a marcação de plural no verbo*. Dissertação de Mestrado, Universidade de São Paulo/FFLCH, São Paulo, 182f.
- Cunha, Celso & Lindley Cintra. 2007. *A nova gramática do Português contemporâneo*, 3ª ed., Rio de Janeiro, Lexikon Informática.
- Duarte, Maria Eugênia Lamoglia & Filomena Varejão. 2013. Null subjects and agreement marks in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, vol. 12(2): 101–124.
- Fernandes, Marisa. 1996. *Concordância Nominal na Região Sul*. Dissertação de Mestrado, UFSC, Florianópolis.
- Guy, Gregory R. 2013. The cognitive coherence of sociolects: how do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, vol. 52: 63–71.
- Guy, Gregory R. 1981. *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax and language history*. Tese de Doutorado, University of Pennsylvania, Philadelphia, 406f.
- Guy, Gregory R. 2000. A identidade linguística da comunidade de fala: paralelismo interdialeto nos padrões de variação linguística. *Organon*, vol. 14(28–29): 17–32.
- Guy, Gregory R. & Ana M. S. Zilles. 2007. *Sociolinguística quantitativa: instrumental de análise*. São Paulo, Parábola.
- Jespersen, Otto. [1933] 1953. *Essentials of English Grammar*. London, George Allen and Unwin Ltd..
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, vol. 3(1): 359–383.
- Labov, William. 2001. *Principles of Linguistic Change: external factors*. Oxford & Cambridge, Blackwell.
- Labov, William. [1966] 2006. *The social stratification of English in New York City*. São Paulo, Cambridge University Press.
- Lopes, Norma da Silva. 2001. *Concordância nominal, contexto linguístico e sociedade*. Tese de Doutorado, UFBA, Salvador. 408f.
- Lucchesi, Dante. 2012. A concordância verbal e a polarização sociolinguística do Brasil. Ms.
- Martinet, André. 1955. *Economie des changements phonétiques*. Berne, A. Francke.
- Mendes, Ronal Beline & Livia Oushiro, Livia. *Documentação do Projeto SP2010 – Construção de uma amostra da fala paulistana*. Disponível em <<http://projetsp2010.fflch.usp.br/producao-bibliografica>>. Último acesso em 21/03/2014.
- Monguillott, Isabel de Oliveira e Silva. 2009. *Estudo sincrônico e diacrônico da concordância verbal de terceira pessoa do plural no PB e no PE*. Tese de Doutorado, UFSC, Florianópolis. 229f.

- Naro, Anthony J. 1981. The social and structural dimensions of a syntactic change. *Language*, vol. 57: 63–98.
- Naro, Anthony J. & Maria Marta Pereira Scherre. 1991. Variação e mudança linguística: fluxos e contrafluxos na comunidade de fala. *Caderno de Estudos Linguísticos*, Campinas, vol. 20 (jan./jun/): 9–16.
- Naro, Anthony J., Edair Görski & Eulália Fernandes. 1999. Change without change. *Language Variation and Change*, vol. 11: 197–211.
- Oushiro, Livia. (2015) *Identidade na pluralidade: avaliação, produção e percepção linguística na cidade de São Paulo*. Tese de Doutorado, Universidade de São Paulo/FFLCH, São Paulo.
- Pereira, Deize Crespim. 2004. *Concordância verbal na língua falada nas trilhas das bandeiras paulistas*. Dissertação de Mestrado, Universidade de São Paulo/FFLCH, São Paulo, 116f.
- R CORE TEAM. 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Disponível em <<http://www.R-project.org/>>. Último acesso em 14/08/2014.
- Revelle, William. 2014. *Package 'psych'* v.1.4.5. Procedures for psychological, psychometric, and personality research. Disponível em <<http://personality-project.org/r/psych.manual.pdf>>. Último acesso em 18/10/2014.
- Rodrigues, Angela C. S. 1987. *A concordância verbal no Português Popular em São Paulo*. Tese de Doutorado, Universidade de São Paulo/FFLCH, São Paulo. Dois volumes. 380f.
- Rubio, Cassio Florêncio. 2012. *Padrões de concordância verbal e de alternância pronominal no Português Brasileiro e Europeu: estudo sociolinguístico comparativo*. Tese de Doutorado, Universidade Estadual Paulista/IBILCE, São José do Rio Preto, 393f.
- Salomão, Mircia Hermenegildo. 2010. *A variação de pluralidade nas estruturas predicativas da variedade falada na região de São José do Rio Preto*. Dissertação de Mestrado, Universidade Estadual Paulista/IBILCE, São José do Rio Preto. 162f.
- Scherre, Maria Marta Pereira. 1978. *A regra de concordância de número no sintagma nominal em Português*. Dissertação de Mestrado, UFRJ/ Faculdade de Letras, Rio de Janeiro. 158f.
- Scherre, Maria Marta Pereira. 1988. *Reanálise da concordância nominal em Português*. Tese de Doutorado, UFRJ, Rio de Janeiro. Dois volumes. 555f.
- Scherre, Maria Marta Pereira. 1998. Paralelismo linguístico. *Revista Estudos Linguísticos*, vol. 7(2): 29–59, Belo Horizonte.
- Scherre, Maria Marta Pereira. 2001. Phrase-level parallelism effect on noun phrase number agreement. *Language Variation and Change*, vol. 13: 91–107.
- Scherre, Maria Marta Pereira & Anthony J. Naro. 1992. The serial effect on internal and external variables. *Language Variation and Change*, vol. 4: 1–13.
- Scherre, Maria Marta Pereira & Anthony J. Naro. 1998. Sobre a concordância de número no Português falado do Brasil, em G. Ruffino (org.) *Dialettologia, geolinguística, sociolinguística*. (Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza) Centro di Studi Filologici e Linguistici Siciliani, Università di Palermo. Tübingen: Max Niemeyer Verlag, 5: 509–523.
- Scherre, Maria Marta Pereira & Anthony J. Naro. 2006. Mudança sem mudança: a concordância de número no Português Brasileiro. *SCRIPTA*, vol. 9(18): 107–129.
- Scherre, Maria Marta Pereira, Anthony J. Naro & Caroline Rodrigues Cardoso. 2007. O papel do tipo de verbo na concordância verbal no Português Brasileiro. *DELTA*, vol. 23(n.esp.): 283–317.
- Silva, Janaína Biancardi da & Maria Marta Pereira Scherre. 2013. A concordância nominal na fala cabixaba: fatores sociais, em C. R. Cardoso; M. M. P. Scherre; H. M. M. Lima-Salles & C. Pacheco (eds.), *Variação linguística: contato de línguas e educação*, Campinas, Pontes Editores: 129–143.
- Weinreich, Uriel, Labov, William & Herzog, Marvin I. [1968] 2006. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo, Parábola. Tradução de Marcos Bagno.
- Yacovenco, Lilian Coutinho, Maria Marta Pereira Scherre, Leila Maria Tesch, Marcela Langa L. Bragança, Elaine Meireles Evangelista, Alexandre Kronemberger de Mendonça, Elba Nusa Calmon, Heitor da Silva Campos Júnior, Astrid Franco Barbosa, Jucilene Oliveira Sousa Basílio, Carlos Eduardo Deoclécio, Janaína Biancardi da Silva, Aline Tomaz Fonseca Berbert & Samine de Almeida Benfica, 2012. Projeto Portvix: A fala de Vitória/ES em cena. *Alfa*, vol. 56(3): 771–806.
- Zilles, Ana Maria Stahl; Leonardo Zechlinski Maya & Karine Quadros da Silva. 2000. A concordância verbal com a primeira pessoa do plural na fala de Panambi e Porto Alegre, RS. *Organon - Estudos da língua falada*, vol. 28/29(14): 195–219. Porto Alegre, UFRGS – Instituto de Letras.