

**Dois pastel e um chopes:  
a concordância nominal e identidade(s) paulistana(s)**

(Two pastel and a draught ‘beers’: nominal agreement and Paulistano identities)

**Resumo**

Este artigo analisa a realização variável da concordância nominal em uma amostra contemporânea e representativa do português paulistano, composta de 118 entrevistas sociolinguísticas, a partir dos pressupostos teórico-metodológicos da Sociolinguística Variacionista (LABOV, 1972). Além de contribuir para seu mapeamento em uma comunidade ainda pouco investigada de um ponto de vista sociolinguístico, este estudo discute as identidades sociais associadas à variante marca zero na comunidade paulistana, a fim de melhor compreender os mecanismos que regulam a manutenção de usos linguísticos considerados “não padrão”. A partir da análise qualitativa e quantitativa dos dados, sugere-se que a marca zero funciona como índice de masculinidade na fala de homens paulistanos e de identidade local com certos bairros e com a cidade.

**Palavras-chave:** Concordância nominal de número, São Paulo, identidades sociais.

**Abstract**

Based on Variationist Sociolinguistics theory and methods (LABOV, 1972), this paper analyzes variable nominal agreement in a contemporary and representative sample of Paulistano Portuguese, consisting of 118 sociolinguistic interviews. In addition to contributing to the variable’s description in an understudied community, this paper discusses the social identities associated to the zero morpheme variant, in order to shed light on the mechanisms regulating the continued use of the so-called “nonstandard variants”. The qualitative and quantitative analyses suggest that the zero morpheme variant indexes masculinity in male Paulistanos’ speech, as well as local identities.

**Keywords:** Nominal number agreement, São Paulo, social identities.

## **Introdução<sup>1</sup>**

A variação na concordância nominal de número (CN) dentro de sintagmas nominais (SNs) e entre o SN e o predicado – como em *a-s casa-s* vs. *as casa-Ø* e *as casas são bonita-s* vs. *as casas são bonita-Ø* – tem sido objeto de numerosos estudos sociolinguísticos no português brasileiro (ver, p.ex., SCHERRE, 1978, 1988; GUY, 1981, 2000; NARO; SCHERRE, 1991; LOPES, 2001; ANTONINO, 2007; SALOMÃO, 2010; BRANDÃO; VIEIRA 2012; YACOVENCO *et al.*, 2012; MENDES, no prelo, *inter alia*).

Certos resultados têm sido recorrentes nas mais variadas comunidades em que o fenômeno foi investigado. A Posição Linear do vocábulo dentro do SN, sua Posição Relativa ao núcleo ou o cruzamento entre essas duas variáveis sempre se mostram correlacionados com a variável: a primeira posição e as posições pré-nucleares tendem a favorecer a ocorrência da marca de plural (ver, p.ex., SCHERRE, 1988; BRANDÃO; VIEIRA, 2012). A saliência fônica (SCHERRE, 1988), analisada através dos Processos de Formação do Plural (-s, -is, -ões etc.), também se apresenta invariavelmente relevante; embora a hierarquia de fatores possa diferir de comunidade para comunidade, ou mesmo entre subgrupos da mesma comunidade (SCHERRE, 1988), o plural regular (com simples acréscimo de /-s/) sempre favorece a marca zero; a saliência fônica quanto à Tonicidade da Palavra nem sempre é analisada, mas tende a ser relevante quando considerada, com desfavorecimento das marcas de plural em monossílabos átonos, paroxítonas e proparoxítonas (ver, p.ex. SALOMÃO, 2010). O efeito serial do Paralelismo (SCHERRE; NARO, 1992; SCHERRE, 1998, 2001), seja entre as palavras dentro de um sintagma, seja entre sintagmas de uma sequência discursiva, costuma revelar a forte tendência ao emprego de formas semelhantes (marcas levam a mais marcas e zeros levam a mais zeros). O Contexto Fonológico Seguinte, quando se mostra relevante, indica o favorecimento da marca zero quando a palavra é seguida por consoantes (ver, p.ex. GUY, 1981). Para Animacidade do Núcleo, sintagmas com o traço [-humano] (*os gato-s, as coisa-s*) tendem a desfavorecer a marcação de plural (ver, p.ex. BRANDÃO; VIEIRA, 2012).

Quanto a variáveis sociais, o Nível de Escolaridade dos falantes apresenta-se invariavelmente como uma das mais fortemente correlacionadas com a marcação de plural: quanto maior o nível de escolaridade, maiores as taxas de emprego da variante prescrita pela norma culta, que normalmente se toma como “variante padrão”. A análise em tempo aparente

tende a quadros de variação estável, com o desfavorecimento da marca zero por parte dos falantes da faixa etária intermediária, embora subgrupos dentro de uma mesma comunidade possam apresentar movimentos divergentes entre si (NARO; SCHERRE, 1991; SCHERRE; NARO, 2006); o Sexo/Gênero, por fim, tende a revelar o favorecimento da marca zero pelos homens, embora a variável nem sempre seja selecionada como relevante pelo programa Varbrul (ver, p.ex. SALOMÃO, 2010 para a região de São José do Rio Preto e BRANDÃO; VIEIRA 2012 para a Região Metropolitana do Rio de Janeiro).

Entre paulistanos, a alternância tem sido analisada da perspectiva de sua correlação com identidades de gênero (MENDES, no prelo) e do contraste entre diferentes subcomunidades da cidade, como paulistanos e alagoanos (GOMES DA SILVA, 2014). No entanto, um levantamento abrangente a respeito dos padrões de variação entre paulistanos ainda não havia sido realizado. O presente trabalho tem como objetivo mais geral descrever padrões de variação quanto a variáveis sociais e linguísticas tradicionalmente analisadas, a fim de contribuir para o mapeamento da concordância nominal no português brasileiro. Além disso, o artigo discute as identidades sociais que se associam à marca zero de plural na cidade de São Paulo: em uma comunidade altamente urbanizada, em que é cada vez mais raro encontrar pessoas que não tenham concluído pelo menos o nível médio de escolaridade, quais fatores se correlacionam com a marca zero de plural e quais grupos se associam ao emprego da variante?

Para atender a essas questões, um *corpus* de 118 entrevistas sociolinguísticas foi analisado qualitativa e quantitativamente. Seu roteiro inclui tópicos de conversação sobre o bairro do informante, sua infância, seu trabalho, suas atividades de lazer e a cidade de São Paulo.<sup>2</sup> Na parte final das entrevistas, o documentador busca avaliações sociolinguísticas sobre certas variantes, entre as quais se inclui a marca zero de concordância nominal. A análise qualitativa examina as respostas dos informantes à pergunta “O que você acha desse modo de falar: ‘me vê dois pastel e um chopes?’”. O discurso metalingüístico dos falantes sugere uma associação estereotípica da marca zero com a cidade de São Paulo e, mais especificamente, com o tradicional bairro da Mooca, embora os falantes mais jovens pareçam não mais associar a variante com o português paulistano. A análise quantitativa, por sua vez, mostra que variáveis linguísticas tradicionalmente investigadas em estudos sobre concordância nominal, como Saliência Fônica e Paralelismo, também se apresentam como relevantes para (CN) no português paulistano; no entanto, um resultado que aqui difere em relação a outras comunidades é aquele

para a variável Animacidade, que revela tendência à marca zero em SNs com o traço [+humano]. Quanto às variáveis sociais, discute-se o papel da Classe Social (muitas vezes não analisada na sociolinguística brasileira), a forte correlação com Sexo/Gênero (diferentemente de outros centros urbanos), e a hipótese de que os moradores do bairro da Mooca tendem à marca zero. Argumenta-se que a variante funciona como índice de masculinidade e de uma identidade local, para além de rótulos como forma “padrão” ou “não padrão”.

Os 118 informantes são nascidos na cidade de São Paulo e estão estratificados de acordo com quatro variáveis sociais: (i) Sexo/Gênero; (ii) Faixa Etária (20 a 34 anos; 35 a 59 anos; 60 anos ou mais); (iii) Nível de Escolaridade (até Ensino Médio; Ensino Superior); e (iv) Região de Residência na cidade (bairro mais central; bairro mais periférico). As análises estatísticas de correlações foram realizadas no programa R (R CORE TEAM, 2013), em modelos de efeitos mistos, com auxílio do pacote Rbrul (JOHNSON, 2009). Tais modelos permitem a inclusão de variáveis aleatórias, específicas da amostra analisada – como, por exemplo, o Falante e o Item Lexical –, juntamente a efeitos fixos como Sexo/Gênero, Faixa Etária ou Saliência Fônica, mais tradicionalmente analisadas em programas como o GoldVarb X. A inclusão de variáveis aleatórias gera modelos estatísticos mais confiáveis ao levar em conta a influência idiosincrática de falantes ou itens lexicais que possivelmente se comportem diferentemente dos demais (para uma discussão mais detalhada, ver BAAYEN, 2008; JOHNSON, 2009; XXX, 2014).

### **A marca zero de concordância nominal e identidade(s) paulistana(s)**

Nacionalmente, os paulistanos são estereotipados por frases como “dois pastel e um chopes”, com marca zero de plural em “pastel” e um “s” adicional em “chopes”. Essa associação pode ser vista na mídia, na cultura pop e no discurso popular sobre o português paulistano. A Figura 1 apresenta o título de uma matéria da sessão de turismo do jornal carioca *O Globo*, na qual os jornalistas fazem uso dessa expressão popular: “Um chopes, dois pastel: aproveite nossa seleção dos melhores bares de São Paulo”. O exemplo (1) é excerto de uma letra de música da banda paulistana Velhas Virgens, que faz múltiplas referências a localidades e personagens da capital paulista: esquina da Ipiranga com a São João, o bar Brahma, a Catedral da Sé, Adoniran Barbosa, Noel Rosa.



Figura 1: Notícia do jornal *O Globo online* (28 abr./2010). Fonte: <<http://oglobo.globo.com/boa-viagem/um-chopes-dois-pastel-aproveite-nossa-selecao-dos-melhores-bares-de-sao-paulo-3017181>>. Último acesso em 24 abr. 2014.

(1) **Um chopp e dois pastel (Velhas Virgens)**  
(Paulão de Carvalho e Alexandre Cavallo Dias)

*Encontrei Caetano  
Na esquina da Ipiranga com a São João  
Sabe o que ele tinha na mão  
Um chopp do Bar Brahma e um disco do Adoniran  
Um chopp do Bar Brahma e um disco do Adoniran*

*Encontrei o professor Pasquale  
Atrás da Catedral da Sé  
Ele estava a pé  
Trazia dois pastel e um disco do Noel  
Trazia dois pastel e um disco do Noel*

*Pasquale, vejam vocês  
Deu pastel pro Caetano e acertou o português  
Caetano, que é baiano  
Pôs “s” no chopp e virou um paulistano  
(...)*

O excerto em (2) é de uma notícia do jornal Folha de São Paulo *online* (05 jun./2011) sobre um projeto de lei para preservar o “mooquês”, suposto sotaque do bairro da Mooca, na Zona Leste da cidade. Dessa notícia, destaca-se o comentário da presidente da associação dos moradores do bairro, Crescenza Giannoccaro de Souza, que faz questão de ressalvar que os moradores “sabem fazer o plural corretamente”. Embora negue a associação da marca zero de concordância com os mooquenses, seu comentário reforça a existência de um discurso corrente que afirma justamente o contrário.

(2) **Sotaque da Mooca pode virar patrimônio imaterial de SP**

*O “cantado” modo de falar da Mooca, bairro da zona leste de São Paulo, que inclui também expressões nacionalmente famosas como “orra meu” e “belo”, pode se tornar o primeiro bem imaterial protegido da cidade.*

*Está no Conpresp (Conselho Municipal do Patrimônio Histórico) um pedido para transformar o “mooquês” em patrimônio de São Paulo. A ideia é registrar e preservar esse jeito peculiar de falar de parte dos paulistanos.*

*“Fico maravilhada com a ideia”, afirma Crescenza Giannoccaro de Souza Neves, presidente da Associação Amo a Mooca. Ela faz, porém, uma ressalva sobre uma característica atribuída ao modo de falar dos moradores do bairro: a falta de “s” nos plurais. “Os imigrantes, quando chegaram, tinham dificuldade de dizer os plurais, pois era diferente da língua deles. Nós, descendentes, também falamos cantado, mas usamos bem os plurais.” (...)*

Fonte: Folha de São Paulo *online*. 05/06/2011. Disponível em <<http://www1.folha.uol.com.br/cotidiano/925287-sotaque-da-mooca-pode-virar-patrimonio-imaterial-de-sp.shtml>>. Último acesso em 24 abr. 2014.

Visto que se trata de uma variável difundida por todo o território brasileiro, é curioso notar sua associação estereotípica especificamente com a cidade de São Paulo. No roteiro das entrevistas sociolinguísticas aqui utilizadas, buscaram-se avaliações sobre a marca zero de concordância nominal através da pergunta: “O que você acha desse modo de falar: ‘me vê dois pastel e um chopes’?”. Foram pouquíssimos os informantes que não apresentaram algum tipo de comentário metalinguístico sobre a concordância nominal. Trata-se de uma variante com alto grau de saliência social e que se apresenta como sinônimo de “não saber falar português”.

A fim de avaliar quais noções são mais recorrentemente associadas à sentença-alvo, elaborou-se a Figura 2 no portal Wordle ([www.wordle.net](http://www.wordle.net)), uma ferramenta gratuita para criação de nuvens de palavras. O programa calcula a frequência relativa de cada vocábulo e dá maior proeminência àqueles que são mais frequentes; nas reações dos informantes à sentença-alvo, o termo “errado” ocorreu mais frequentemente do que o termo “Mooca”, que por sua vez ocorreu mais do que “classe baixa”. Trata-se de um modo de representar visualmente quais noções são mais recorrentemente associadas com a expressão “dois pastel e um chopes”.



Figura 2: Nuvem de palavras associadas à sentença-alvo “me vê dois pastel e um chopes”

Nota-se na Figura 2 a alta frequência de associação com o falar paulistano e, menos frequentemente, com a imigração italiana e o bairro da Mooca. O discurso popular atribui a presença da variante na cidade à vinda maciça de imigrantes italianos, em cuja língua nativa o plural nominal não se realiza com o morfema –s.

- (3) D1: [risos] *e o que o senhor acha de por-exemplo eh... “me vê dois pastel e um chopes”*  
S1: *bem paulista<sup>3</sup> isso aí... isso é bem paulista e tradicional daqui do Tatuapé do/ da Mooca dessa região* (Lucio A., M2MP)<sup>4</sup>
- (4) *então assim a gente tem inclusive essa coisa do... de comer o esse né no no do dos plural... né do ter tomado ‘duas cerveja’... eh [...] ‘um chopes e dois pastel’... essa coisa de comer o esse é vem/ acho acredito eu que venha também do italiano porque no italiano o plural não é feito com esse é feito com i né... e então a gente acabou herdando isso da convivência com os italianos aqui então* (Romulo S., M3MC)
- (5) D1: *e o que que você acha de... “me vê dois pastel e um chopes”?*  
S1: [risos] *ah sei lá... ah pessoal menos instruído que fala assim né... “dois pastel e um chopes”... é mais o povo né?*  
D1: *(vo)cê conhece pessoas que falam assim?*  
S1: *olha... na Mooca eu conheci muita gente assim... inclusive... eu tenho um irmão... ele não/ quer dizer ele até (es)tá morando na Mooca eu tinha me esquecido disso... fazem uns três meses que ele voltou pra Mooca... mas ele corta muitos esses finais das palavras... apesar de ser médico... mas ficou assim ‘os carro’ não sei por que que fala assim... acho esquisito não sei é um pouquinho da origem italiana ou se é influência do bairro né* (Iara S., F2SC)

A informante Iara S. (5) associa a marca zero de plural primeiramente com pessoas menos escolarizadas e de classes mais baixas; em seguida, no entanto, ela lembra de muitas pessoas do bairro da Mooca, em que a informante vivera até a adolescência e no qual ainda possui parentes. Lá, mesmo os falantes mais escolarizados, como seu irmão médico, tendem a “cortar os esses finais das palavras”, algo que é atribuído, pela informante, à influência italiana ou do bairro.

Outros informantes, no entanto, afirmam se tratar mais de um estereótipo criado pelos cariocas, ainda que admitam a possibilidade de haver algum fundo empírico para a gozação (6).

- (6) D1: *é... e quanto a/ a esse modo de falar “me vê dois pastel e um chopes”?* [...] (vo)cê acha que é característico de algum/ de algum grupo  
S1: *carioca tirando sarro de paulista*  
D1: [risos] (vo)cê acha que... que é mais uma caricatura então?  
S1: *n...*  
D1: *assim... um estereótipo... não é... não reflete tanto a realidade*  
S1: *“dois pastel e um”... pode refletir...*  
D1: *uhum*  
S1: *é que eu nunca prestei atenção nisso [...] não vou dizer que não aconteça não (xxx)*  
D1: *mas aí você acha que... que é algo que caracteriza o paulistano ou...*  
S1: *sim caracteriza*  
D1: *como um todo assim?*  
S1: *(todo) alguns você escuta isso de (ir em feira)*  
D1: *aham*  
S1: *se (vo)cê ficar prestando atenção você vai acabar escutando* (Pietro C., M2SC)

Certos falantes, por outro lado, não só assumem o estereótipo como defendem o emprego da variante como marca de identidade paulistana:

- (7) D1: *e o “me vê dois pastel e um chopes”?*  
S1: *é... é coisa nossa e não deve mudar (es)tá bom?*  
D1: *é*  
S1: *se (es)tá errado (xxx) falando se é errado em português (es)tá errado isso aqui mas não tem que deixar/ isso é uma coisa que tem que manter e deixar os carioca porque o carioca pra começo os cariocas não sabe vestir...* (Plinio C., M3MP)

Além da associação estereotípica com o falar paulistano, o segundo tipo de comentário mais recorrente acusa explicitamente o “erro” gramatical, associado a falantes menos escolarizados ou de classes mais baixas, e seguido geralmente de avaliações negativas como

“feio, horrível, brega, ignorante” (Figura 2). Andrea C. (8), professora de geografia do Ensino Fundamental e Médio, não disfarça seu desgosto de ver seus alunos “falando errado”. Cláudomiro S. acredita que a associação com paulistanos e imigrantes se refere a um fenômeno do passado, mas que o emprego da variante com marca zero hoje em dia se deve à “falta de conhecimento”, não mais à imigração (9).

- (8) D1: *e o que (vo)cê acha dessa daqui... “me vê dois dois pastel e um chopes”?*  
S1: *isso aí mostra uma ignorância (atroz) uma coisa que nunca/ “dois pastel”... quando os alunos falam tenho vontade de descer um livro na cara deles... é forma de falar e e passa como sendo uma coisa que todo paulista fala... e de novo todo brasileiro gosta de futebol... todo paulista fala errado “me dá dez pãozinho”...* (Andrea C., F2SP)
- (9) D1: *tá... e por exemplo essa daqui o que (vo)cê acha dessa?... “me vê dois pastel e um chopes”... “me vê dois pastel e um chopes”*  
S1: *essa daí é a clássica do falar errado né*  
D1: [risos] *por quê?*  
S1: *essa daí era... pelo menos nas ane/ anedotas que corriam... era o carioca tirando sarro de paulista*  
D1: *é né sei*  
S1: *essa daí... eu/ como eu procuro falar corretamente e fazer tudo corretamente isso daí... dói no ouvido quando eu escuto... eu acho feio*  
D1: *então como que seria o jeito correto?*  
S1: *é “um pastel”... não/ “dois pastéis... dois pastéis e um chope” [...] isso daí antigamente era... na época que (vo)cê tinha aquela coisa de... ter os grupos de migrantes ou de imigrantes... que praticamente não saía do bairro... ahn eles tiravam muito sarro dos italianos que os italianos que falavam desse jeito então a região ali Bexiga Brás Mooca... se falava desse jeito... e aí com o passar do tempo foi mudando um pouco mas... virou anedota... hoje em dia é possível que se fale desse jeito mas... aí não é porque é um imigrante... que que não sabe falar direito... a língua do lugar pra onde ele imigrou.... mas hoje em dia é por causa realmente da... falta de conhecimento... que (es)tá que (es)tá se popularizando...* (Cláudomiro S., M2SP)

Com efeito, as avaliações que associam a expressão “dois pastel e um chopes” com o falar paulistano, com a imigração italiana e com certos bairros da cidade parecem se restringir aos falantes mais velhos, a partir da segunda faixa etária (falantes acima de 35 anos). Entre os mais jovens, sobretudo os de classes mais altas e residentes de bairros mais centrais, são mais recorrentes avaliações negativas e a associação com baixa escolaridade (ex. 10-11). Alguns jovens chegam a reverter o estereótipo e atribuir a expressão aos cariocas (ex. 12).

- (10) D1: *da pessoa que fala assim “me vê dois pastel e um chopes”?*  
S1: *horrível*

D1: [risos]

S1: coisa mais ridícula do mundo... e é tão comum né a gente escuta não só isso mas outros erros... que é uma coisa assim surreal (Carolina A., F1MC)

- (11) D1: [...] e (vo)cê acha que tem bairros da cidade em que as pessoas falam mais desse jeito “dois pastel” e outros que falam menos?

S1: acho acho

D1: por bairro?

S1: acho que sim acho que principalmente os da periferia né porque

D1: ahn

S1: infelizmente o nível de instrução né de escolaridade é menor então acho que isso é uma tendência (Tatiana M., F1SP)

- (12) D1: e o... “dois pastel e um chopes?”

S1: nossa aí é... meio caiçara meio carioca né... eu já ouvi muito isso de os cara falando de zoeira né... por causa do lance do plural... coisa de feira né... meu vô zoava isso não sei da onde que é

D1: mas por que que ele zoava você sabe?

S1: não... não sei essas... não sei esse tipo de zoeira que você faz meio por fazer sabe... frase do tipo meio dito popular assim... (Fernando B., M1MC)

Os excertos (10)-(12) dos falantes mais jovens parecem indicar que a associação estereotípica da expressão “dois pastel e um chopes” com paulistanos pode estar em vias de desaparecimento da comunidade. Na análise quantitativa a seguir, a associação com paulistanos e com falantes de certos bairros é investigada através do mapeamento das tendências de emprego de acordo com a região de residência do falante. Além disso, investiga-se se haveria tendência de mudança em direção à marca explícita de número devido ao aumento geral dos níveis de escolarização na comunidade e à forte avaliação negativa demonstrada pelos falantes mais jovens.

## Análises Quantitativas

Das 118 entrevistas, extraíram-se todas as ocorrências de sintagmas nominais plurais com duas ou mais palavras em que pelo menos uma foi flexionada quanto ao número (ou seja, não se consideram dados invariáveis como *mais ônibus*, *dois óculos*, tampouco aqueles em que a noção de pluralidade é apenas semântica, como *muita criança* ou *uma série de coisa*). Os mais de 17 mil SNs foram analisados a partir de duas perspectivas: a atomística, em que cada palavra é considerada uma unidade de análise (p.ex. *o-s* vs. *o-Ø*; *menino-s* vs. *menino-Ø*), e a não-atomística, em que se analisa o SN como um todo (p.ex. *os meninos* vs. *os menino-Ø/ o-Ø*)

*menino-s*) (SCHERRE, 1988). Em ambas as análises, excluíram-se dados seguidos de [s, z, ſ, ʒ] (p.ex. *os meninos são...*, *os meninos já...*, *as – suas – chamadas...*), que constituem um contexto de neutralização em que não é possível determinar se o morfema de plural foi de fato realizado. Em sintagmas com três ou mais palavras, em que há múltiplas combinações possíveis quanto à presença ou ausência de morfema de plural, a análise de SNs contrapôs a concordância plena (p.ex. *os meus meninos* - SSS) à concordância parcial – ou seja, a marca zero em pelo menos uma das palavras flexionáveis (p.ex. *os meus menino* (SSØ), *os meu menino* (SØØ), *o meus meninos* (ØSS), *o meus menino* (ØSØ), *o meu meninos* (ØØS)). Ressalta-se que as análises são complementares entre si e objetivam iluminar diferentes processos que ocorrem seja no nível da palavra (como a posição do vocábulo em relação ao núcleo), seja no nível do sintagma (como a animacidade do SN). Trata-se, com efeito, de um mesmo conjunto de dados.

Os termos “marca zero” e “CN-Ø” são aqui empregados como sinônimos para fazer referência à variante sem concordância plena, tanto na análise atomística quanto na de SNs (ainda que, a rigor, só possa haver “concordância” entre palavras, ou seja, em SNs). Para ambas as perspectivas, as variáveis sociais incluídas nas análises foram as mesmas.

A distribuição geral dos dados se encontra na Tabela 1, que apresenta o número de ocorrências de CN-Ø de ambas as perspectivas, o número total de dados e a proporção de CN-Ø em toda a amostra. Percebe-se que a marca zero de plural é relativamente infrequente na comunidade, com taxa de 8,3% quando se consideram os itens lexicais separadamente e 15,2% do total de SNs.

Tabela 1: Distribuição de ocorrências de CN-Ø das perspectivas atomística e de SNs

	CN-Ø	N Total	% CN-Ø
Atomística	2.829	34.130	8,3
Sintagmas Nominais	2.709	17.866	15,2

Quanto à Posição Linear do vocábulo dentro do SN, verifica-se uma distribuição extrema quando a palavra ocupa a primeira posição: a taxa de marca zero nesse contexto é de apenas 0,3% (Tabela 2), o que reforça a constatação de estudos prévios de que essa é a posição privilegiada para a marcação de número em SNs no português brasileiro (GUY, 1981; SCHERRE, 1988; BRANDÃO; VIEIRA, 2012). Em análise multivariada de regressão logística, a disparidade entre a primeira posição e as demais fica clara nos pesos relativos (de .08 a .86) e

no amplo *range* (diferença entre maior e menor pesos relativos) de 78, o que torna essa variável a mais relevante para a variação em (CN).

Tabela 2: Tendências de emprego de CN-∅ quanto à Posição Linear (análise atomística)  
(N CN-∅ = 2.829 (8,3%); N Total = 34.130)

	P.R.	% CN-∅	N Total
<b>Posição Linear</b>			
1 <sup>a</sup> posição	.08	0,3	14.246
2 <sup>a</sup> posição	.85	14,1	16.934
3 <sup>a</sup> posição	.85	13,5	2.725
4 <sup>a</sup> posição	.86	14,3	210
5 <sup>a</sup> posição	.85	13,3	15
<i>Range</i>	78		
Input = 0.022.			

As demais posições lineares (2<sup>a</sup> a 5<sup>a</sup>) apresentam taxas de marca zero em torno de 14% no português paulistano e pesos relativos bastante próximos entre si. De fato, uma análise que exclui os dados de primeira posição deixa de selecionar Posição Linear como uma variável relevante (Tabela 3).

Tabela 3: Tendências de emprego de CN-∅ quanto à Posição Linear sem dados de 1<sup>a</sup> posição  
(análise atomística) – (N CN-∅ = 2.786 (14,0%); N Total = 19.884)

	P.R.	% CN-∅	N Total
<b>Posição Linear (sem 1<sup>a</sup> posição)</b>			
2 <sup>a</sup> posição	[.50]	14,1	16.934
3 <sup>a</sup> posição	[.49]	13,5	2.725
4 <sup>a</sup> posição	[.51]	14,3	210
5 <sup>a</sup> posição	[.49]	13,3	15

Input = 0.140. [ ] indicam a não seleção da variável. Pesos relativos provenientes de análise *one-level*.

Feitas essas considerações sobre a distribuição dos dados quanto à Posição Linear da palavra, as análises que seguem consideram somente as ocorrências a partir da segunda posição do SN (para a análise atomística), visto que a marcação do plural é semicategórica na primeira posição. Discutem-se inicialmente os resultados para as variáveis linguísticas nas análises atomística e de SNs; aqueles para as variáveis sociais, que também foram incluídas nos dois conjuntos de análises, vêm descritos na seção seguinte.

## Análise atomística

As variáveis linguísticas da análise atomística são discriminadas no Quadro 1.

Análise atomística	
Processos morfonológicos de formação do plural (terminação da palavra e morfema)	Vogal oral /s/: “casa, casas” Vogal nasal /s/: “homem, homens”, “mão, mãos” -r /is/: “professor, professores” -l /is/: “azul, azuis”, “pastel, pastéis” -S /is/: “mês, meses” -ão /ɔ̃js/: “profissão, profissões” -ão /ãjs/: “pão, pães” Metafônicos: “ovo, [ó]vos”
Tonicidade do item singular	Monossílabo tônico: “vez, vezes” Monossílabo átono: “o, os” Polissílabo oxítono: “morador, moradores” Paroxítono: “ano, anos” Proparoxítono: “quilômetro, quilômetros”
Posição relativa	Pré-nuclear: “todas <i>as</i> pessoas mais velhas” Nuclear: “todas as <i>pessoas</i> mais velhas” Pós-nuclear: “todas as pessoas mais <i>velhas</i> ”
Classe morfológica	Substantivo: “meus <i>amigos</i> , os <i>parques</i> ” Adjetivo: “placas <i>informativas</i> , países <i>europeus</i> ” Quantificador indefinido: “essas <i>várias</i> pessoas” Demonstrativo: “todas <i>essas</i> coisas” Possessivo: “amigos <i>meus</i> , os <i>seus</i> direitos” Artigo definido: “todos <i>os</i> dias” Pronome: “todos <i>eles</i> ”
Contexto fônico seguinte	Consoante: “as <i>meninas</i> <u>foram</u> ” Vogal: “as <i>meninas</i> <u>então</u> ” Pausa: “as <i>meninas...</i> então”
Paralelismo intrasintagmático (somente para 3 <sup>a</sup> , 4 <sup>a</sup> e 5 <sup>a</sup> posições)	Precedida por palavra com marca zero: “nos <u>último-</u> Ø <i>anos</i> ” Precedida por palavra com marca de plural ou invariável: “nos <u>últimos</u> <i>anos</i> , uns <u>dezessete</u> <i>anos</i> ”
Item lexical (efeito aleatório)	

Quadro 1: Variáveis linguísticas da análise atomística

Embora a codificação tenha sido bastante detalhada inicialmente, foram realizadas certas amalgamações no desenvolvimento das análises estatísticas, que serão destacadas oportunamente. Em todos os casos, a redução no número de fatores de uma variável levou em conta semelhanças estruturais cujas distribuições não se diferenciavam significativamente, de acordo com o valor de qui-quadrado, os graus de liberdade e valor de *p* (GUY; ZILLES,

2007:188-197). Ao mesmo tempo, certos pares de variáveis linguísticas não são verdadeiramente independentes entre si – por exemplo, Posição Relativa e Classe Morfológica –, de modo que foi necessário realizar análises separadas, ora com uma, ora com outra.

Dentre essas variáveis, Processos Morfonológicos de Formação de Plural e Tonicidade do Item Singular investigam, ambas, o papel da saliência fônica (SCHERRE, 1988). A segunda variável, no entanto, não é selecionada como significativa para a marcação variável de plural em um modelo de efeitos mistos, com a inclusão do item lexical como efeito aleatório. Embora já tenha se mostrado relevante em outros estudos (SCHERRE, 1988; DIAS; FERNANDES, 2000; SALOMÃO, 2010), é possível que o efeito de tonicidade se deva a alguns itens lexicais frequentes, como “os”, “coisas” ou “vezes”.

Processos Morfonológicos, por sua vez, é uma das variáveis mais relevantes entre as selecionadas como significativas para (CN) (Tabela 4). Nesta variável, foram amalgamados, por um lado, os plurais em -ães (*pães*), -ões (*profissões*), -r /is/ (*professores*) e metafônicos (*ovos*), e os plurais em -l /is/ (*azuis*) e -S /is/ (*meses*), por outro ( $\chi^2 = 5,74(4)$ ,  $p > 0,20$ ).<sup>5</sup> Palavras terminadas em vogal oral (*casa*) ou nasal (*homem, mão*) favorecem a marca zero de plural (P.R. .80 e .71). Os plurais irregulares, por outro lado, têm pesos relativos menores para a marca zero; os fatores que mais desfavorecem a marca de número são as palavras terminadas em -l e -S (P.R. .43).

Tabela 4: Tendências de emprego de CN-∅ quanto a variáveis linguísticas (análise atomística)  
(N CN-∅ = 2.786 (14,0%); N Total = 19.884)

	P.R.	% CN-∅	N Total
<b>Processos Morfonológicos de Formação de Plural</b>			
vogal oral /-s/	.80	16,0	16.066
vogal nasal /-s/	.71	10,7	394
-ães, -ões, -r /-is/, metafônicos	.66	8,1	1.386
-l /is/, -S /-is/	.43	2,9	2.038
<i>Range</i>	37		
<b>Posição Relativa</b>			
Pré-nuclear	.47	1,7	1.215
Nuclear ou pós-nuclear	.87	14,8	18.669
<i>Range</i>	40		
<b>Contexto Fônico Seguinte</b>			
Consoante ou pausa	.54	14,9	14.824
Vocal	.49	11,3	5.060
<i>Range</i>	5		
Input = 0.044.			

A Posição Relativa da palavra separou inicialmente os itens lexicais pré-nucleares, nucleares ou pós-nucleares, mas a diferença entre os dois últimos não se mostrou relevante ( $\chi^2 = 1,21(1)$ ,  $p > 0,20$ ). Essa variável também se apresenta como uma das mais importantes para a marcação de plural, com leve desfavorecimento da marca zero em posição pré-nuclear (P.R. .47), em contraste com a nuclear/pós-nuclear (P.R. .87).

O Contexto Fônico Seguinte, por sua vez, mostra uma diferença pequena – mas significativa – entre palavras seguidas por consoante ou pausa (P.R. .54,  $\chi^2 = 0,02(1)$ ,  $p > 0,80$ ), que favorecem a marca zero de plural, e aquelas seguidas de vogal (P.R. .49). Tal tendência, já demonstrada em outros estudos (GUY, 1981), deve-se à tendência de formação da sílaba canônica CV: quando seguido de vogal, o morfema –s de plural pode ser ressilabificado [ka.za.za.ma.re.las], enquanto que sua manutenção em palavras seguidas de consoante ou pausa gera sílabas CVC [ka.zaz.ver.dIs.].

As demais variáveis da análise atomística foram examinadas em rodadas que não incluem aquelas da Tabela 4, a fim de evitar interações. Paralelismo Intrasintagmático analisa a tendência de marcação de plural de acordo com a palavra imediatamente precedente dentro do sintagma. Considerando-se que a marcação de plural é semicategórica na 1<sup>a</sup> posição, excluíram-se dessa análise os dados de 2<sup>a</sup> posição. A Tabela 5 mostra que, quando a 3<sup>a</sup>, a 4<sup>a</sup> ou a 5<sup>a</sup> palavras do SN são precedidas por uma palavra com marca de plural ou invariável, a tendência de marca zero é desfavorecida (P.R. .45); por outro lado, quando precedidas por uma palavra sem marca de plural, a tendência a CN-Ø é bastante alta: ocorre em 81,8% desses dados, equivalente a um peso relativo de .98. Em outras palavras, uma vez que se emprega uma palavra com marca zero, a probabilidade de que a(s) palavra(s) seguinte(s) no mesmo sintagma apresente(m) o morfema de plural, embora possível, é muito pequena.

Tabela 5: Tendências de emprego de CN-Ø quanto ao Paralelismo Intrasintagmático (análise atomística) - (N CN-Ø = 400 (13,6%); N Total = 2.950)

	P.R.	% CN-Ø	N Total
<b>Paralelismo intrasintagmático</b>			
Precedida por palavra sem marca de plural	.98	81,8	143
Precedida por palavra com marca de plural	.45	10,1	2.807
<i>Range</i>		53	
Input = 0.057.			

A variável Classe Morfológica interage com Posição Relativa ao Núcleo (p.ex., substantivos sempre ocupam a posição nuclear), de modo que foram analisadas separadamente. Os resultados na Tabela 6 indicam que substantivos e adjetivos favorecem a marca zero de plural (P.R. .58 e .56); quantificadores indefinidos (*vários*, *muitos*) estão próximos do ponto neutro (PR. .49); demonstrativos (*esses*, *aquelas*) e possessivos (*meus*, *sus*) desfavorecem a variante (P.R. .16); e não houve ocorrências de marca zero em artigos definidos (*os*, *as*) ou pronomes (*eles*) nessa amostra. Tal hierarquia se assemelha àquela observada para a Posição Relativa, já que artigos definidos, demonstrativos, possessivos e quantificadores indefinidos tendem a ocupar posições pré-nucleares (desfavorecedoras da marca zero), e adjetivos e substantivos tendem a ocupar posições nucleares e pós-nucleares (favorecedoras da marca zero).

Tabela 6: Tendências de emprego de CN-Ø quanto à Classe Morfológica (análise atomística)  
(N CN-Ø = 2.786 (14,0%); N Total = 19.884)

	P.R.	% CN-Ø	N Total
<b>Classe Morfológica (sem Posição Relativa)</b>			
Substantivo	.58	14,8	16.950
Adjetivo	.56	13,2	1.953
Quantificador indefinido	.49	11,5	61
Demonstrativo + possessivo	.16	2,2	554
Artigo definido + pronome	--	0,0	366
<i>Range</i>	42		

Input = 0.057.

A fim de separar o efeito de Classe Morfológica e Posição Relativa, realizou-se uma análise alternativa com o cruzamento dessas variáveis. Excluíram-se substantivos (pelo fato de sempre ocuparem a posição nuclear) e artigos definidos e pronomes (por não apresentarem variação na presente amostra de dados); adjetivos e quantificadores definidos, por sua vez, foram amalgamados tanto em posição pré- quanto pós-nuclear ( $\chi^2 = 0,05(2)$ ,  $p > 0,95$ ). A Tabela 7 mostra que permanece a diferença entre as posições pré-nuclear, por um lado, e nuclear/pós-nuclear, por outro, já que as primeiras têm pesos relativos menores do que as segundas; além disso, há uma diferença significativa entre demonstrativos/possessivos e adjetivos/quantificadores indefinidos. Os primeiros relativamente desfavorecem a marca zero em relação aos segundos (Cf. .13 vs. .29; .35 vs. .69). Desse modo, o efeito de Classe Morfológica independe daquele de Posição Relativa: ambas as variáveis têm influência no uso da marca zero de (CN).

Tabela 7: Tendências de emprego de CN-Ø em cruzamento entre Classe Morfológica e Posição Relativa (análise atomística) - (N CN-Ø = 276 (10,7%); N Total = 2.568)

	P.R.	% CN-Ø	N Total
<b>Classe Morfológica + Posição Relativa</b>			
Adjetivo + quant. indefinido – nuclear e pós-nuclear	.69	15,3	1.640
Demonstrativo + possessivo – nuclear e pós-nuclear	.35	6,1	66
Adjetivo + quant. indefinido – pré-nuclear	.29	3,5	374
Demonstrativo + possessivo – pré-nuclear	.13	1,6	488
	<i>Range</i>	56	
Input = 0.036.			

### Análise de SNs

A análise da perspectiva não-atomística (SNs) incluiu as seguintes variáveis linguísticas:

Análise de SNs	
Classe Morfológica do Primeiro Elemento do SN	Artigo indefinido: “ <i>uns meninos</i> ” Artigo definido: “ <i>os meninos</i> ” Demonstrativo: “ <i>essas coisas todas</i> ” Numeral: “ <i>duas pessoas</i> ” Possessivo: “ <i>meus irmãos</i> ” Adjetivo: “ <i>melhores condições</i> ” Quantificador indefinido: “ <i>alguns erros</i> ” “ <i>todos/todas</i> ”: “ <i>todas essas pessoas</i> ” Substantivo: “ <i>pessoas bonitas</i> ”
Configuração Sintagmática Pós-Nuclear	Com posições pós-nucleares: “ <i>as pessoas mais velhas</i> ” Sem posições pós-nucleares: “ <i>as pessoas</i> ”
Animacidade do SN	[+humano]: “ <i>os professores</i> ” [-humano]: “ <i>as coisas</i> ”
Paralelismo Sintático (5 orações precedentes)	Primeiro de uma série Precedido por SN com marca zero Precedido por SN com marca explícita de número
Síntagma Nominal (efeito aleatório)	

Quadro 2: Variáveis linguísticas da análise de SNs

Quanto à Configuração Sintagmática, separaram-se os contextos pré- e pós-nuclear; o primeiro foi analisado de acordo com a Classe Morfológica da primeira palavra do SN; o segundo considerou a existência ou não de palavras após o núcleo. Animacidade levou em conta o traço [+humano], e Paralelismo Sintático atentou para a presença de outros SNs, com ou sem

marca explícita, nas cinco orações precedentes à ocorrência. O modelo também analisou o próprio SN como um efeito aleatório, a fim de avaliar se SNs frequentes como “os cara-s” e “as coisa-s” distorcem os resultados para os efeitos fixos. A correlação com essas variáveis é apresentada na Tabela 8.

Tabela 8: Tendências de emprego de CN-Ø quanto a variáveis linguísticas (análise de SNs)  
(N CN-Ø = 2.709 (15,2%); N Total = 17.866)

	P.R.	% CN-Ø	N Total
<b>Classe Morfológica do Primeiro Elemento do SN</b>			
Artigo indefinido / demonstrativo	.72	28,4	2.424
Artigo definido / numeral / possessivo	.52	14,5	12.375
Adjetivo	.33	8,9	1.060
Quantificador indefinido / “todos”	.28	6,6	1.407
Substantivo	.15	6,3	600
	<i>Range</i>	57	
<b>Paralelismo Sintático</b>			
Precedido por SN com marca zero	.70	48,4	1.393
Primeiro de uma série	.52	16,1	7.806
Precedido por SN com marca explícita	.45	9,0	8.667
	<i>Range</i>	25	
<b>Configuração Sintagmática Pós-nuclear</b>			
Com posições pós-nucleares	.69	16,6	1.456
Sem posições pós-nucleares	.48	15,0	16.410
	<i>Range</i>	21	
<b>Animacidade</b>			
[+humano]	.54	16,2	5.581
[-humano]	.48	14,7	12.285
	<i>Range</i>	6	
Input = 0.051.			

Classe Morfológica do Primeiro Elemento, inicialmente com nove fatores, foi simplificada para uma variável com cinco ( $\chi^2 = 5,52(4)$ ,  $p > 0,20$ ). SNs encabeçados por artigos indefinidos e demonstrativos são aqueles que mais favorecem a marca zero (P.R. .72); artigos definidos, possessivos e numerais revelam tendência neutra (P.R. 52); e adjetivos (P.R. 33), quantificadores indefinidos (P.R. .28) e substantivos (P.R. .15) desfavorecem a variante. É possível que o favorecimento de CN-Ø em SNs iniciados por artigos indefinidos e demonstrativos deva-se a um traço semântico [-definido]; não apenas os artigos indefinidos, mas também muitas das ocorrências de demonstrativos não possuem uma referência específica, referindo-se a categorias gerais de pessoas ou coisas:

- (7) *aquele pessoal que vai no funk tem aquelas calcinhas coladinha-* $\emptyset$ ... *tem uma roupa de funk baile funk* (Clara C., F2SC)
- (8) *dá vontade de rir né? porque elas parece aquelas velha-* $\emptyset$  *que reclama né?* (Mariana B., F2MP)

Quanto à Configuração Sintagmática Pós-nuclear, em consonância com o resultado verificado para Posição Relativa na análise atomística, a presença de palavras em posições pós-nucleares favorece a marca zero (P.R. .69), enquanto que sua ausência a desfavorece (P.R. .48).

O Paralelismo Sintático (SCHERRE; NARO, 1992; SCHERRE, 1998, 2001), investigado em diversos estudos prévios, manifesta resultados similares em São Paulo: SNs precedidos por outros SNs com marca zero favorecem CN- $\emptyset$  (P.R. .70), e SNs precedidos por SNs com marca explícita a desfavorecem (P.R. .45); quando se trata do primeiro SN de uma série, a tendência é próxima do ponto neutro (P.R. .52), semelhante à probabilidade geral da amostra.

Animacidade, por sua vez, revela tendência oposta àquela normalmente verificada em outros estudos (ver, p.ex. BRANDÃO; VIEIRA, 2012): são os SNs com o traço [+humano] que favorecem CN- $\emptyset$  (P.R. .54) em relação aos SNs com o traço [-humano] (P.R. .48). De fato, alguns SNs recorrentes que tendem a favorecer a marca zero são “os cara” e “os moleque”. Esses resultados indicam que a correlação com Animacidade nem sempre segue a mesma direção, e que o favorecimento de CN- $\emptyset$  por SNs [-humano] ou [-animado] não é generalizado no português brasileiro.

### Variáveis sociais

As análises atomística e não-atomística (SNs) incluíram ambas as mesmas sete variáveis, discriminadas no Quadro 3, além do Falante como efeito aleatório. Junto às variáveis estratificadoras da amostra (Sexo/Gênero, Faixa Etária, Nível de Escolaridade e Região de Residência), analisaram-se a Classe Social,<sup>6</sup> a Origem dos Pais e a Mobilidade Geográfica dos falantes. De modo semelhante à inclusão de Item Lexical e Síntagma Nominal como efeitos aleatórios para as variáveis linguísticas, a consideração do Falante no modelo estatístico tem o objetivo de se certificar de que a seleção de variáveis sociais deve-se a um efeito real das variáveis Sexo/Gênero, Faixa Etária etc. Nas palavras de Johnson (2009, p. 365), as variáveis são

selecionadas como estatisticamente significativas apenas quando as correlações “são fortes o suficiente para superar a variação inter-falantes”.

<b>Variáveis sociais (análises atomística e de SNs)</b>	
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 – A Classe média alta – B1 Classe média média – B2 Classe média baixa – C1 Classe baixa alta – C2 Classe baixa média – D
Origem dos pais	São Paulo-capital Norte/Nordeste Interior de SP/MG Estrangeiros 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 3: Variáveis sociais

Ainda que as análises atomística e de SNs tenham sido realizadas separadamente, os resultados são apresentados nas mesmas tabelas apenas para que possam ser mais diretamente comparados: ambas exibem resultados similares, quando não idênticos: coincidem o conjunto de variáveis selecionadas, sua hierarquia de seleção e os valores de *range*.

Não são selecionadas como significativas para (CN) a Região de Residência, a Origem dos Pais e a Faixa Etária. Este último indica que a concordância nominal parece ser um caso de variação estável em São Paulo. A despeito da avaliação de Tatiana M. que associa a marca zero com os moradores de periferia (ex. 11), não há correlação com a Região de Residência do falante. De modo semelhante, não obstante a constante associação da variante com um falar

“italianado”, não se pode afirmar, a partir dessa amostra – não estratificada de acordo com a origem dos pais –, que o emprego de CN-Ø tenha alguma influência da imigração italiana na cidade.

A frequente associação da variante com o bairro da Mooca, entretanto, foi investigada através do cruzamento entre região e zona de residência, que permite um mapeamento mais detalhado da cidade. A nova variável, com nove fatores (as zonas norte, sul, leste e oeste subdivididas entre parte central e periférica, mais a zona central), passou por amalgamações até se chegar ao modelo mais sintético da Tabela 9 ( $\chi^2 = 2,95(5)$ ,  $p > 0,70$  para análise atomística e  $\chi^2 = 3,29(5)$ ,  $p > 0,50$  para análise de SNs).

Tabela 9: Tendências de emprego de CN-Ø em cruzamento entre Região e Zona de Residência (análise atomística e de SNs)

Região + Zona de Residência	Atomística <sup>a</sup>			SNs <sup>b</sup>		
	N CN-Ø = 2.786 (14,0%); N Total = 19.884		N Total	N CN-Ø = 2.709 (15,2%); N Total = 17.866		N Total
	P.R.	% CN-Ø	P.R.	% CN-Ø	P.R.	P.R.
Leste Central	.85	35,1	.934	.85	37,7	851
Leste Periférica	.65	18,9	2.612	.66	19,7	2.374
Norte Periférica + Sul + Oeste	.46	12,7	13.681	.46	13,8	12.275
Central + Norte Central	.41	8,7	2.657	.40	9,6	2.366
<i>Range</i>	<i>44</i>			<i>45</i>		

<sup>a</sup>Input = 0.064. <sup>b</sup>Input = 0.053.

Os moradores da Zona Leste Central são aqueles que mais favorecem a marca zero (P.R. .85), seguidos dos moradores da Zona Leste Periférica (P.R. .65 e .66); os residentes das Zonas Central e Norte Central são aqueles que mais desfavorecem essa variante (P.R. .41 e .40), e os das demais regiões exibem tendência mais próxima ao ponto neutro (P.R. .46). Ora, a Zona Leste Central é justamente onde se localiza o bairro da Mooca, frequentemente associado à expressão “dois pastel e um chopes” pelos falantes paulistanos. Neste caso, a avaliação da comunidade sobre um subgrupo que tende a empregar a variante coincide com aquilo que se observa empiricamente. Essa correlação será examinada mais detalhadamente na próxima seção.

Em estudos prévios sobre (CN), o Nível de Escolaridade, quando considerado, invariavelmente se mostra relevante para a variação. A presente amostra foi estratificada entre falantes que atingiram o Ensino Médio e aqueles que prosseguiram para o Ensino Superior; no

entanto, a análise apresentada na Tabela 10 separa do primeiro grupo os falantes com nível de Ensino Fundamental. A categorização em três níveis permite entrever que a correlação é gradual e sistemática: falantes que estudaram até o Ensino Fundamental favorecem fortemente a marca zero (P.R. .85 e .81); aqueles que estudaram até o Ensino Médio também a favorecem, em menor grau (P.R. .67 e .64); e aqueles com nível superior tendem a desfavorecer-la (P.R. .32 e .35). Em outras palavras, quanto menor o nível de escolaridade, maior a tendência de emprego das variantes proscritas pela norma culta.

Tabela 10: Tendências de emprego de CN-Ø quanto ao Nível de Escolaridade  
(análise atomística e de SNs)

Nível de Escolaridade	Atomística <sup>a</sup>			SNs <sup>b</sup>		
	P.R.	% CN-Ø	N Total	P.R.	% CN-Ø	N Total
			N CN-Ø = 2.786 (14,0%); N Total = 19.884			N CN-Ø = 2.709 (15,2%); N Total = 17.866
Até Ensino Fundamental	.85	29,7	2.415	.81	31,8	2.192
Até Ensino Médio	.67	18,4	6.358	.64	19,8	5.775
Ensino Superior	.32	8,1	11.111	.35	8,8	9.899
Range	53			46		

<sup>a</sup>Input = 0.068. <sup>b</sup>Input = 0.074.

Pode-se questionar, no entanto, se esse resultado se deve à interferência direta da escolarização ou se trata-se de um efeito colateral da classe socioeconômica dos falantes, uma vez que diferentes níveis de escolaridade restringem as oportunidades de posicionamento no mercado de trabalho, de ascenção social e de estabelecimento de redes sociais dos falantes. O contraste com os resultados para a variável Classe Social pode iluminar essa questão.

A Classe Socioeconômica dos falantes se apresenta como a variável social que se correlaciona mais fortemente com CN-Ø, com *range* de 64-65. Da Tabela 11 depreende-se que, quanto mais baixo o nível socioeconômico, mais alta é a tendência ao emprego da marca zero, desde as classes mais altas A/B1 (P.R. .13 e .23) até as classes mais baixas C2/D (P.R. .77 e .88). A comparação dos valores de logaritmo de verossimilhança (GUY; ZILLES, 2007:191) das análises com nível de escolaridade e classe social, com grau de liberdade igual a um, indica que a segunda tem maior poder de previsão do que a primeira para a ocorrência da marca zero ( $\chi^2 = 93,3(1)$ ,  $p < 0,001$  para a análise atomística,  $\chi^2 = 19,64(1)$ ,  $p < 0,001$  para a análise de SNs). Desse modo, parece mais razoável interpretar que as tendências de emprego observadas em

diferentes níveis de escolaridade são decorrência do *status* social dos falantes e das redes sociais que eles estabelecem com falantes da mesma classe social.

Tabela 11: Tendências de emprego de CN-∅ quanto a variáveis sociais  
(análise atomística e de SNs)

	Atomística <sup>a</sup>			SNs <sup>b</sup>		
	N CN-∅ = 2.786 (14,0%); N Total = 19.884		N Total	N CN-∅ = 2.709 (15,2%); N Total = 17.866		N Total
	P.R.	% CN-∅	P.R.	% CN-∅		
<b>Classe Social</b>						
A-B1	.13	5,4	5.119	.23	5,9	4.528
B2	.38	8,9	5.981	.35	9,7	5.343
C1	.64	18,0	5.127	.63	19,0	4.656
C2-D	.77	28,9	3.657	.88	31,1	3.339
<i>Range</i>	64			65		
<b>Sexo/Gênero</b>						
Feminino	.38	9,8	9.327	.36	10,4	8.434
Masculino	.64	17,8	10.557	.63	19,4	9.432
<i>Range</i>	26			27		
<b>Mobilidade</b>						
Baixa	.54	21,9	4.122	.65	23,8	3.727
Média e alta	.36	12,0	15.762	.46	12,9	14.139
<i>Range</i>	18			19		

<sup>a</sup>Input = 0.044. <sup>b</sup>Input = 0.051.

O Sexo/Gênero dos falantes (Tabela 11) também se correlaciona com o emprego de CN-∅, com seu favorecimento por parte dos homens (P.R. .64 e .63) e desfavorecimento pelas mulheres (P.R. .36 e .46). Tendo em vista as correlações que geralmente se observam para variáveis cujas variantes se diferenciam em graus de estigma ou prestígio, os presentes resultados estão de acordo com a expectativa de que a variante não padrão fosse favorecida pelos falantes de sexo masculino. No entanto, os estudos sobre a concordância nominal, sobretudo em comunidades urbanas, nem sempre verificam a correlação com a variável Sexo/Gênero (ver, p.ex., SALOMÃO, 2010; BRANDÃO; VIEIRA, 2012) ou correlações tão fortes quanto a que se observa em São Paulo, com *range* de 26-27. Mendes (no prelo), em seu estudo sobre a concordância nominal e identidades de gênero no português paulistano, constatou correlação semelhante em sua amostra de 36 entrevistas sociolinguísticas: homens heterossexuais tendem a favorecer a marca zero em relação a mulheres, e homens gays efeminados tendem a desfavorecerla ainda mais. A partir desse resultado, o autor sugere que, na comunidade paulistana, os homens

tendem a empregar CN-Ø de modo a evidenciar sua masculinidade. Ainda que o presente estudo diferencie os falantes mais propriamente quanto ao *sexo*, e não quanto ao *gênero* ou quanto à marcação, no discurso, de um modo mais ou menos masculino de ser, uma explicação semelhante parece plausível para dar conta não apenas da direção observada para a correlação, mas para a grande diferença nas tendências entre homens e mulheres. Na cidade de São Paulo, a marca zero parece funcionar não só como forte estratificador de classes sociais e de níveis de escolaridade, mas também como um índice de masculinidade.

Quanto à Mobilidade Geográfica, os falantes com menor mobilidade (aqueles que sempre viveram no mesmo bairro) tendem a empregar a marca zero (P.R. .54 e .65), ao passo que os falantes com maior mobilidade tendem a evitá-la (P.R. .36 e .46). De acordo com Milroy (1987 [1980]), pessoas com maior mobilidade tendem a adotar valores de um grupo de referência externo, enquanto aquelas com menor mobilidade tendem a se orientar pelas práticas locais. Esse resultado sinaliza a importância de uma identidade local para o emprego da variante.

A próxima seção se volta para a questão das identidades locais, seja com o bairro, seja com a cidade: como se dá o encaixamento social de CN-Ø em diferentes partes da cidade? Em especial, a forte associação no discurso popular entre a marca zero e o bairro da Mooca ocorreria apenas pelo seu uso mais frequente, ou também por padrões divergentes de variação linguística em relação à comunidade? Lembre-se do comentário de Iara S. (ex. 5), que afirma que na Mooca a variante é comumente empregada mesmo na fala de pessoas mais escolarizadas, como seu irmão médico. Poderia a associação da marca zero com o bairro advir de uma maior semelhança nos usos linguísticos por parte de pessoas com diferentes graus de escolaridade ou pertencentes a classes sociais distintas?

### **Identidade local**

As análises acima mostraram que, embora não haja correlação significativa entre (CN) e a divisão da cidade entre bairros mais centrais e mais periféricos, há uma região específica da cidade em que a marca zero é fortemente favorecida – a Zona Leste Central, onde se localiza o bairro da Mooca. Realizaram-se então cruzamentos, utilizando-se as porcentagens, entre o mapeamento mais detalhado da cidade de acordo com Região e Zona de residência, e a Escolaridade e Classe Social dos falantes.

As Figuras 3 e 4 mostram os resultados desses cruzamentos. Em todas as áreas, as taxas de emprego de CN-∅ são maiores entre os falantes com menor nível de escolaridade (representados pela coluna mais escura) em relação aos de Ensino Superior (colunas mais claras), exceto para as Zonas Central e Norte Central, para as quais as taxas foram idênticas; no entanto, entre os falantes menos escolarizados, os habitantes da Zona Leste Central empregam-na com frequência muito maior (aproximadamente 55%) do que os das demais regiões da cidade, cujas taxas variam entre 9,5 a 23,3%.

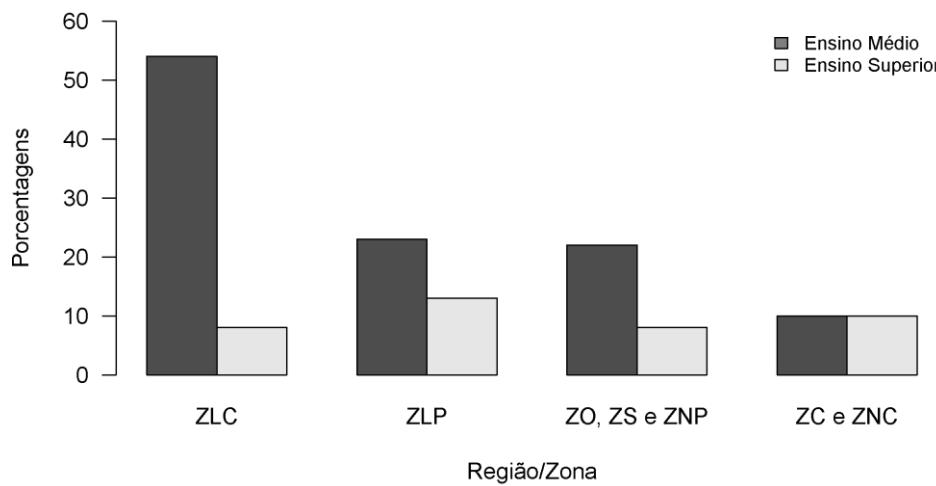


Figura 3: Cruzamento entre Região/Zona e Nível de escolaridade

Algo semelhante se verifica no cruzamento com Classe Social. Na Figura 4, há taxas gradualmente maiores de emprego de CN-∅ em todas as regiões quanto mais baixa for a classe social (com exceção, novamente, das Zonas Central e Norte Central, cujas taxas se aproximam de 10%). Nas Zonas Leste Central e Periférica, não há, nesta amostra, falantes das classes sociais mais altas A e B1, o que deve ter contribuído para as maiores tendências ao emprego da marca zero nessas regiões. Entretanto, entre os falantes das classes mais baixas C2 e D, a taxa de emprego de CN-∅ é muito superior na Zona Leste Central (73,1%), superior ao dobro daquelas das demais regiões para falantes da mesma classe social (29,4% para a Zona Leste Periférica; 32,8% para as Zonas Oeste, Sul e Norte Periférica; e 10% nas Zonas Central e Norte Central).

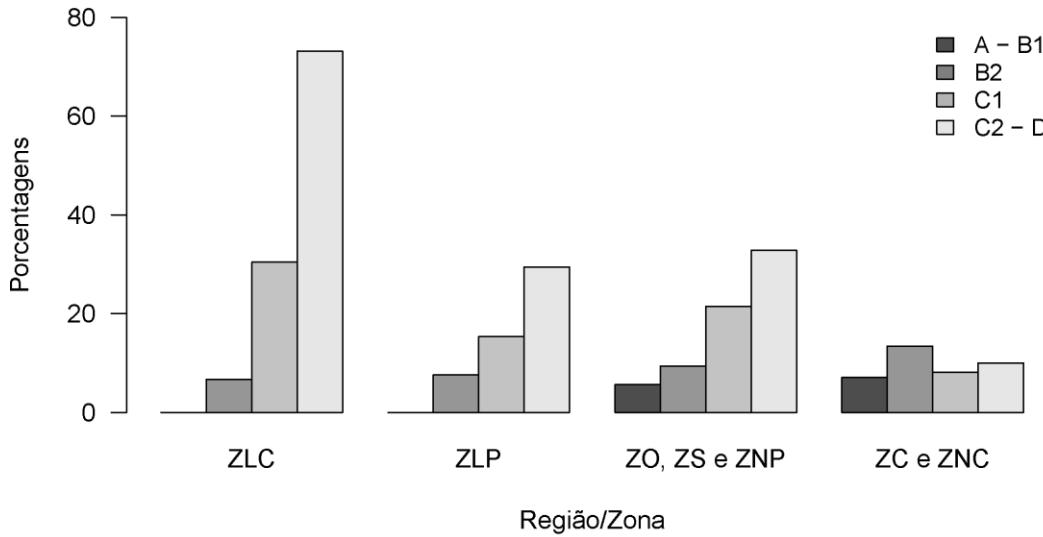


Figura 4: Cruzamento entre Região/Zona e Classe Social

A associação da variante com os falantes que residem especificamente no bairro da Mooca parece ser decorrência tão somente de taxas exageradas de emprego da marca zero por parte de certos residentes dessa área da cidade, que evidentemente destoam das médias em outras regiões. O estereótipo acerca do português paulistano provavelmente advém daí: alguns dos falantes mais prototípicos da cidade, cuja variedade é identificada, nomeada e comentada no discurso popular – o “mooquês” – apresentam taxas bastante superiores de emprego da marca zero.

Nesse aspecto, a identidade local com o bairro ou com a cidade parece reforçá-las, uma vez que “dois pastel”, antes de ser “errado”, é considerado “paulistano” ou “mooquense” para muitos dos falantes da amostra. Entretanto, é interessante notar que não se verificam tendências verdadeiramente divergentes entre os residentes da Zona Leste Central; nessa parte da cidade, as regras variáveis quanto à hierarquia de restrições se mantêm: falantes mais escolarizados ou de classes mais altas empregam a marca zero relativamente menos do que os falantes menos escolarizados e de classes mais baixas, assim como em outras partes da cidade.

Tais resultados contribuem para a discussão sobre o *locus* das identidades sociais que se manifestam através de usos linguísticos: até que ponto os falantes são capazes de manipular o emprego de certas variantes a fim de demonstrar filiação a determinados grupos sociais? Os resultados acima sugerem que a associação de traços linguísticos a grupos de indivíduos reside

em suas *frequências* de uso, e não em *tendências*: ainda que as taxas de emprego de uma variante possam diferir, as regras variáveis em todos os grupos de falantes são semelhantes.

## Considerações finais

Os diversos estudos sobre a concordância nominal no português brasileiro já permitem diversas generalizações quanto ao encaixamento linguístico das variantes, cujo emprego se correlaciona à Posição Linear e Relativa do vocábulo, à Saliência Fônica, ao Contexto Fônico Seguinte, ao Paralelismo Intra e Intersintagmático, à Configuração Sintagmática Pré- e Pós-nuclear e à Animacidade do SN. No presente estudo, o único resultado divergente daquele observado em outras comunidades refere-se a esta última variável, em que se verificou o favorecimento de CN-Ø em sintagmas com o traço [+humano], o que indica que a direção da correlação não é a mesma para todas as variedades do português brasileiro. Quanto a fatores sociais, a presente análise reitera a forte correlação com o Nível de Escolaridade dos falantes, mas permite defender que tal resultado é mais propriamente interpretado como efeito indireto da Classe Social e das redes sociais que os falantes estabelecem. Além disso, observou-se a influência do grau de Mobilidade Geográfica e do Sexo/Gênero dos falantes – este último, em um grau mais forte do que já se verificou em outras comunidades urbanas.

Tendo em vista que muitos dos padrões sociais e linguísticos acerca da concordância nominal de número são recorrentes em diferentes comunidades, é necessário voltar-se a novas questões. O presente trabalho enfoca os possíveis mecanismos que operam na associação dos usos linguísticos com certas identidades sociais, através da análise do discurso metalingüístico dos informantes, bem como dos padrões de variação na comunidade. Em especial, investigaram-se as motivações para a associação da marca zero especificamente com os paulistanos, quando a variável está presente em todo o território brasileiro. Sugere-se que tal associação se deve às altas taxas de emprego das variantes por parte de habitantes de bairros tradicionais de São Paulo, sobretudo a Mooca, mas sinaliza-se que as percepções da comunidade estão em vias de mudança, em direção a uma associação mais forte com o nível de escolaridade e condição socioeconômica dos falantes. Além disso, a marca zero parece funcionar como índice de identidade de gênero, dada a ampla diferença entre as tendências de homens e mulheres (maior do que em outras comunidades urbanas em que tal variável se revelou significativa); CN-Ø parece funcionar como estratégia de indexação de masculinidade (MENDES, no prelo). O estudo do significado social

da variação, para além dos amplos padrões observados, pode conduzir a novas interpretações da vitalidade de certas formas linguísticas em diferentes comunidades.

## Notas

<sup>1</sup> Esta pesquisa foi desenvolvida com auxílio financeiro da FAPESP (Proc. no. XXX).

<sup>2</sup> As entrevistas fazem parte do *corpus* do XXX (XXX, 2013), que disponibiliza parte das gravações gratuitamente no portal <XXX>. Último acesso em 12 set. 2014.

<sup>3</sup> A distinção entre os termos “paulista” e “paulistano”, como pessoas nascidas no *estado* e na *cidade* de São Paulo respectivamente, é raramente feita pelos informantes da amostra. Note-se que, em (3), Lucio A. emprega o termo “paulista” para se referir aos habitantes dos bairros do Tatuapé e da Mooca, presumivelmente nascidos na cidade de São Paulo.

<sup>4</sup> Nos exemplos extraídos das entrevistas, o documentador é identificado por “D1” e o informante por “S1”; o perfil social do informante é indicado entre parênteses ao final de cada excerto: pseudônimo, sexo (M = masculino; F = feminino), faixa etária (1 = 20 a 34 anos; 2 = 35 a 59 anos; 3 = 60 anos ou mais), nível de escolaridade (M = até Ensino Médio; S = Ensino Superior) e região de residência (P = bairro periférico; C = bairro central).

<sup>5</sup> Lê-se “qui-quadrado igual a 5,74, com quatro graus de liberdade e valor de *p* (ou significância) maior do que 0,20”. A notação indica que a diferença entre os resultados para a análise com oito fatores ou quatro ( $8-4 = 4$  graus de liberdade) tem uma alta probabilidade (maior do que 20%) de ser aleatória e que, portanto, a diferença não é significativa (um valor abaixo de 0,05 indicaria diferença significativa); seguindo a Navalha de Occam, a análise mais sintética (com quatro fatores a menos) deve ser preferida (ver GUY; ZILLES, 2007, p. 188-197).

<sup>6</sup> Os critérios de definição de classe social dos falantes se baseiam em um índice composto pelo nível de escolaridade do informante, sua ocupação, escolaridade e ocupação dos pais, e renda média do bairro de residência. Ver XXX (2014) para uma descrição mais detalhada.

## Referências

- ANTONINO, V. A concordância nominal em predicativos do sujeito e em estruturas passivas no português popular do interior do estado da Bahia. Dissertação (Mestre em Letras). Salvador, UFBA, 2007. 119f.

- BAAYEN, R. H. *Analysing Linguistic Data: a Practical Introduction to Statistics*. Cambridge: Cambridge University Press, 2008.
- BRANDÃO, S. F.; VIEIRA, S. R. Concordância nominal e verbal: contribuições para o debate sobre o estatuto da variação em três variedades urbanas do português. *Alfa*, São Paulo, 56(3), p. 1035-1064, 2012.
- DIAS, J. F. V.; FERNANDES, M. A inter-relação da concordância nominal e da concordância nos predicativos/particípios passivos, sob enfoque da teoria da variação e mudança linguística. *Organon* 14, p. 115-131, 2000.
- GOMES DA SILVA, F. Alagoanos em São Paulo e a concordância nominal de número. Dissertação (Mestre em Letras). São Paulo, Universidade de São Paulo, 2014. 103f.
- GUY, G. R. Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax and language history. Tese (Doutorado). Filadélfia, University of Pennsylvania, 1981. 406f.
- GUY, G. R. A identidade linguística da comunidade de fala: paralelismo interdialetal nos padrões de variação linguística. *Organon* 14, p. 17-32, 2000.
- GUY, G. R.; ZILLES, A. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola, 2007.
- JOHNSON, D. E. Getting off the GoldVarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistic Compass* 3(1), p. 359-383, 2009.
- LABOV, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- LOPES, N. S. Concordância nominal, contexto linguístico e sociedade. Tese (Doutorado), Salvador, UFBA, 2001. 408f.
- MENDES, R. B. Nonstandard Noun Phrase Agreement as an Index of Masculinity. In: LEVON, E.; MENDES, R. B. *Language, sexuality and stance: studies in distributional sociolinguistics*. Oxford: Oxford University Press, no prelo.
- MILROY, L. *Language and Social Networks*. Oxford: Basil Blackwell, 1987 [1980]. 2nd edition.
- NARO, A. J.; SCHERRE, M. M. P. Variação e mudança linguística: fluxos e contrafluxos na comunidade de fala. *Cadernos de Estudos Linguísticos*, Campinas, vol. 20, p. 9-16, 1991.
- R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. Disponível em <<http://www.R-project.org/>>. Último acesso em 21 mar. 2014.

- SALOMÃO, M. H. A variação de pluralidade nas estruturas predicativas da variedade falada da Região de São José do Rio Preto. Dissertação (Mestrado). São José do Rio Preto, UNESP, IBILCE, 2010. 162f.
- SCHERRE, M. M. P. A regra de concordância de número no sintagma nominal no português. Dissertação (Mestrado). Rio de Janeiro, UFRJ, 1978. 158f.
- SCHERRE, M. M. P. Reanálise da concordância nominal em português. Tese (Doutorado). Rio de Janeiro, UFRJ, 1988. 555f.
- SCHERRE, M. M. P. Paralelismo linguístico. *Revista Estudos Linguísticos* 7(2), p. 29-59, 1998.
- SCHERRE, M. M. P. Phrase-level parallelism effect on noun phrase number agreement. *Language Variation and Change* 13, p. 91-107, 2001.
- SCHERRE, M. M. P.; NARO, A. J. The serial effect on internal and external variables. *Language Variation and Change* 4, p. 1-13, 1992.
- SCHERRE, M. M. P.; NARO, A. J. Mudança sem mudança: a concordância de número no português brasileiro. *SCRIPTA*, vol. 9(18), p. 107-129, 2006
- YACOVENCO, L. *et al.* Projeto PortVix: a fala de Vitória/ES em cena. *Alfa* 56(3), São Paulo, p. 771-806, 2012.