

CONCORDÂNCIA NOMINAL E VERBAL: CONTRIBUIÇÕES PARA O DEBATE SOBRE O ESTATUTO DA VARIAÇÃO EM TRÊS VARIEDADES URBANAS DO PORTUGUÊS

Silvia Figueiredo BRANDÃO*

Silvia Rodrigues VIEIRA**

- **RESUMO:** Focalizam-se a concordância nominal de número e a concordância verbal de terceira pessoa do plural em variedades urbanas do Português Europeu, do Português do Brasil e do Português de São Tomé, com base na fala de indivíduos de níveis fundamental, médio e superior de instrução, distribuídos, ainda, por três faixas etárias e sexo. Com o objetivo de descrever as referidas variedades, o trabalho avalia as motivações de natureza estrutural e social segundo os pressupostos da Teoria da Variação e Mudança (WEINREICH; LABOV; HERZOG, 1968). Os resultados das diferentes análises, desenvolvidas no âmbito do Projeto *Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português*, sugerem que, embora as três variedades manifestem tendência à adoção da regra de concordância tanto no Sintagma Nominal quanto no Sintagma Verbal, a concordância apresenta estatuto diferenciado. No Português Europeu, os dados revelam um comportamento que não pode ser considerado efetivamente variável – trata-se, nos termos de Labov (2003), de uma regra categórica no âmbito da concordância nominal e semicategórica no âmbito da verbal; no Português do Brasil e no Português de São Tomé, ambas as regras são variáveis.
- **PALAVRAS-CHAVE:** Concordância nominal. Concordância verbal. Sociolinguística. Variedades do Português.

Introdução

Um dos temas mais focalizados no âmbito do Português, sobretudo no que se refere à variedade brasileira, a concordância nominal e verbal suscita, ainda, grande interesse em virtude não só de fatores linguísticos que se vêm mostrando recorrentes em diversos estudos, mas também de implicações sócio-histórico-culturais para a caracterização das variedades do Português. Em contraste com o que ocorre no Português Europeu (PE), em que a regra canônica de concordância parece ter *status* categórico ou semicategórico,

* Pesquisadora do CNPq. UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras. Rio de Janeiro – Rio de Janeiro – Brasil. 21941-590 - silvia.brandao@terra.com.br

** Pesquisadora da FAPERJ. UFRJ – Universidade Federal do Rio de Janeiro. Faculdade de Letras. Rio de Janeiro – Rio de Janeiro – Brasil. 21941-590 - silviavieira@hotmail.com

observam-se, em variedades não europeias, padrões de concordância variáveis, que, embora com maior ou menor índice de produtividade em certas áreas (rurais, urbanas), obedecem a determinadas restrições comuns do ponto de vista tanto estrutural quanto extralinguístico. Desde 2008, vem-se desenvolvendo o Projeto *Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias do Português*, inicialmente como projeto de cooperação internacional (Brasil e Portugal) apoiado pela CAPES¹ e, a partir de 2011, como um dos projetos da ALFAL. Ao idealizá-lo, pretendia-se, entre outros objetivos, testar hipóteses explicativas do conjunto dos fenômenos observados à luz dos princípios subjacentes a todas as gramáticas e dos parâmetros específicos a subconjuntos de gramáticas, bem como testar hipóteses prévias sobre a origem das diferenças/semelhanças observadas, tendo em conta a história do português.

As pesquisas empreendidas centraram-se, inicialmente, em variedades linguísticas urbanas. As cidades – onde hoje se concentra a grande maioria da população em virtude da crença de que nelas se encontram maiores oportunidades de trabalho e melhores condições de vida – constituem o espaço ideal para observar a interação entre indivíduos de diferentes estratos sociais. Sendo assim, organizou-se um *corpus*, disponível na internet², com entrevistas realizadas em diferentes comunidades das regiões metropolitanas do Rio de Janeiro e de Lisboa. No que se refere à variedade urbana de São Tomé e Príncipe, conta-se com entrevistas realizadas, em 2009, por Tjerk Hagemeijer e que compõem o *Corpus VARPOR*, do Centro de Linguística da Universidade de Lisboa.

Neste artigo, focalizam-se as variedades urbanas europeia (PE), brasileira (PB) e santomense (PST), no intuito de tecer considerações acerca das motivações de natureza estrutural e social à luz das análises empreendidas. Acredita-se que as análises contrastivas permitam observar até que ponto se pode avaliar como semelhante ou diferente a expressão gramatical de um fenômeno em termos quantitativos e qualitativos.

Perfil das comunidades

O confronto aqui realizado tem, necessariamente, de ser relativizado em termos não só das dimensões das áreas metropolitanas consideradas e do contingente populacional que as integra (a Região Metropolitana do Rio de Janeiro, por exemplo, tem população maior que a de Portugal como um todo), mas também de suas especificidades quanto a aspectos históricos, culturais e sócio-econômicos.

¹ Em sua versão inicial, o Projeto contou com a coordenação de Maria Antonia C. da Mota, da Universidade de Lisboa, e de Sílvia Rodrigues Vieira, da Universidade Federal do Rio de Janeiro. Como Projeto ALFAL, assumiram a coordenação das atividades as autoras do presente artigo.

² Disponível em: <www.lettras.ufrj.br/concordancia>

A Área Metropolitana de Lisboa constitui-se de 18 concelhos que recobrem 2.870km² e onde se concentram 2.819.433 habitantes (2008), “cerca de ¼ da população portuguesa”, bem como 30% das empresas nacionais.³ No concelho de Sintra, encontra-se Cacém, uma de suas 20 freguesias, com 21.289 habitantes.

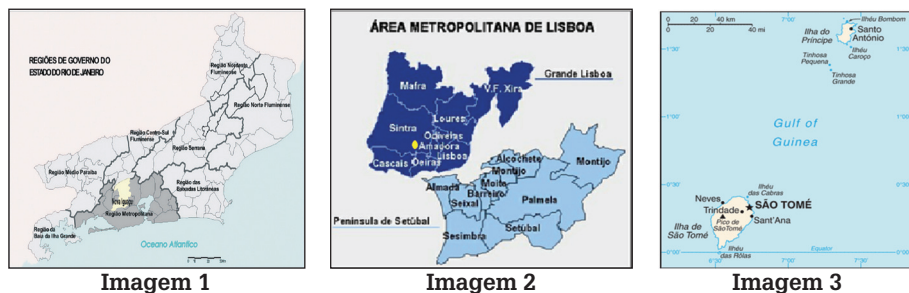
A Região Metropolitana do Rio de Janeiro compõe-se de 19 municípios, distribuídos por 5.645 km², e conta com 12.434.611 habitantes, o que corresponde a 76% da população do Estado. Dentre eles, destaca-se Nova Iguaçu, com 797.637 habitantes⁴, grande parte oriunda de diferentes áreas do Nordeste e de Minas Gerais.

Essas duas áreas metropolitanas têm em comum algumas características, tais como o crescimento desordenado da zona urbana, o significativo número de indústrias e a mobilidade populacional constante. Seus habitantes estão em permanente interação com os moradores das respectivas cidades-sede em função de a maioria deles nelas desenvolverem atividades profissionais.

Na área urbana de São Tomé, vive cerca de um terço da população de São Tomé e Príncipe (206.178 habitantes), arquipélago situado no Golfo da Guiné e onde, em virtude de uma série de contingências históricas, coexistem diversas línguas, entre as quais se destacam o Português e o Forro, que, de acordo com dados do censo realizado em 2001, são falados, respectivamente, por 98,9% e 72,4% dos indivíduos com mais de cinco anos (HAGEMELJER 2009, p. 18). A cidade, onde se desenrolam as principais atividades comerciais, é o principal porto do país e dispõe de estrutura precária.

As três áreas podem ser visualizadas nas imagens a seguir.

Figura 1 – Áreas metropolitanas (Rio de Janeiro, Lisboa e São Tomé)



Áreas Metropolitanas (1) do Rio de Janeiro, com a localização de Nova Iguaçu (mapa com alterações); (2) de Lisboa, com a localização de Cacém (cf. ponto amarelo); (3) de São Tomé (cf. estrela).

Fonte: METRODATA (2012); PLANO REGIONAL DE ORDENAMENTO DO TERRITÓRIO DA ÁREA METROPOLITANA DE LISBOA (2008); ILHA DE SÃO TOMÉ (2012).

³ Disponível em: <<http://www.aml.pt/index.php?&iLevel1=gaml&iLevel2=territorio&iContent=index.html>>. Acesso em: 10 jan. 2012.

⁴ INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. (2010).

Aspectos teórico-metodológicos e características dos corpora

A análise dos dados, tanto da concordância nominal quanto da verbal, parte dos pressupostos assumidos pela Sociolinguística Laboviana (WEINREICH; LABOV; HERZOG, 1968) no que diz respeito à concepção geral de regra variável com base no princípio da heterogeneidade ordenada, à postulação de restrições linguísticas e extralinguísticas que podem acelerar ou refrear processos de variação e mudança, bem como à tipologia proposta em Labov (2003), que diz haver três classes de regras linguísticas, conforme se demonstra a seguir, numa adaptação simplificada do quadro por ele delineado.

Quadro 1 - Tipos de regras linguísticas.

Tipo de regra	Frequência com que opera	Violações ¹
I - Categórica	100%	Nenhuma, na fala natural
II - Semicategórica	95-99%	Rara e relatável
III - Variável	5-95%	Nenhuma por definição e não relatável

Fonte: Com base em Labov (2003, p.243).⁵

Além dos conceitos sociolinguísticos ora assumidos, as investigações da concordância nominal e da verbal valem-se igualmente dos mesmos bancos de dados para a observação do fenômeno – o *corpus Concordância* e o *Corpus VARPOR* já referidos – e dos mesmos procedimentos analíticos, da coleta de dados ao tratamento quantitativo viabilizado pelo pacote de programas computacionais Goldvarb-X.

As análises foram desenvolvidas com base em dados coletados de entrevistas do tipo DID, realizadas com 53 indivíduos, 18 por variedade,⁶ distribuídos por sexo, três faixas etárias (18-35 anos, 36-55 anos e 56 anos em diante) e três níveis de escolaridade (fundamental: 5 a 8 anos; médio: 9 a 11 anos; superior). Os informantes brasileiros são naturais de Nova Iguaçu; os portugueses, de Cacém, e os de São Tomé vivem na capital do país. Além das três referidas variáveis extralinguísticas, controlaram-se (a) na análise da concordância nominal,

⁵ Labov assim descreve as violações, respectivamente, I, II e III: “*none in natural speech; rare and reportable; none by definition and unreportable.*”

⁶ Tendo em vista as variáveis extralinguísticas consideradas (sexo, três faixas etárias e três níveis de escolaridade), com um informante por célula, deveria haver um total de 54 informantes (18 por variedade). No entanto, no *Corpus VARPOR*, referente a São Tomé, não se encontrou uma mulher com mais de 56 anos com nível superior de instrução, o que se deve às próprias condições sócio-histórico-econômicas do país.

oito variáveis estruturais: (1) saliência fônica; (2) tonicidade e (3) número de sílabas do item singular; (4) classe do vocábulo; (5) posição linear e relativa do constituinte no SN; (6) marcas precedentes; (7) contexto fonológico subsequente; (8) animacidade do núcleo do SN; (b) na análise da concordância verbal, as variáveis (1) posição do sujeito em relação ao verbo; (2) distância entre o núcleo do SN e o SV; (3) presença de elementos intervenientes entre o SN sujeito e o verbo; (4) configuração morfossintática do SN sujeito; (5) paralelismo no nível oracional (entre sujeito e verbo); (6) animacidade do sujeito; (7) saliência fônica; (8) tempo verbal; (9) transitividade verbal.

Em relação a São Tomé, realizaram-se análises complementares sobre a concordância no SN, uma, com a ampliação do número de informantes de níveis fundamental e médio; outra, com base em amostra da fala de estudantes, em que foi possível controlar outras variáveis sociais. Quanto à concordância verbal, também foi realizada uma análise adicional com maior número de indivíduos de níveis fundamental e médio.

Sínteses das análises já realizadas

Aspectos gerais

Considerando os trabalhos já desenvolvidos com base no *corpus Concordância*,⁷ pode-se afirmar, tendo em vista exclusivamente os índices de concretização da marca de número,⁸ que, de um lado, os dados urbanos do PB e do PST em análise apresentam comportamento diferente do verificado no Português Europeu. Embora com expressões diferenciadas, brasileiros e são-tomenses apresentam variação entre concordância e não concordância, com preferência pela realização da marca de número, enquanto portugueses registram quase exclusivamente as marcas verbais e nominais de plural. A tabela e os gráficos a seguir permitem observar as tendências ora delineadas nas amostras constituídas com base na fala dos 53 informantes acima referidos.

⁷ Colaboraram na tarefa de coleta e tratamento dos dados de concordância verbal os então alunos de iniciação científica Fernanda Villares Vianna Barreto, Karen Cristina da Silva, Rodrigo Cunha da Silva e Daiane Rangel da Silva. Quanto à concordância nominal, colaboraram os alunos de graduação Aimée de Oliveira Silva e Vinicius Amado de Oliveira Ribeiro.

⁸ No âmbito da concordância verbal, foram coletadas todas as ocorrências de SN na 3ª pessoa do plural que inequivocamente desempenhasse o papel de sujeito do verbo da oração. Assim sendo, foram desprezadas as ocorrências em que tal SN viesse a funcionar como mero tópico discursivo, tendo sido abandonado em favor de uma concepção singular. Também foram desconsiderados os dados em que, por razões fonético-fonológicas, não foi possível determinar se houve ou não realização da marca de número.

Tabela 1 - Distribuição dos dados de concordância (marcação) e não concordância (não marcação) nominal de número e verbal de 3ª pessoa plural nas três variedades estudadas.

Variedade	Concordância nominal		Concordância verbal	
	Marcação	Não marcação	Marcação	Não marcação
PB (Nova Iguaçu)	3439/3777 (91,1%)	338/3777 (8,9%)	1036/1297 (78,1%)	261/1297 (21,9%)
PST (São Tomé)	2439/2612 (93,4%)	173/2612 (6,6%)	640/687 (93,1%)	47/687 (6,9%)
PE (Cacém)	2448/2449 (99,9%)	1/2449 (0,04%)	1499/1515 (98,9%)	17/1515 (1,1%)

Fonte: Elaboração própria.

Gráfico 1: Marcação/não marcação de pluralidade no âmbito do SN.

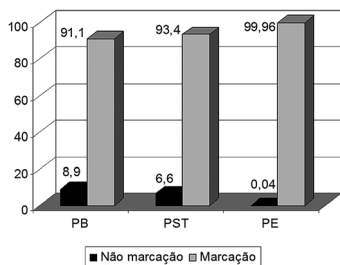
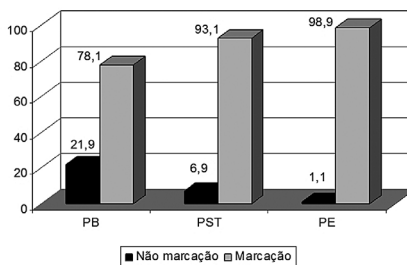


Gráfico 2: Marcação/não marcação de pluralidade no âmbito do SV.



Fonte: Elaboração própria.

Observando-se, apenas, o quantitativo empregado, chama à atenção, primeiramente, o que é comum às três variedades urbanas: a larga preferência pela concretização das marcas de número. Apesar dessa tendência geral, fica evidente, pelo que se verifica no gráfico, que do PB ao PE ocorre um crescimento dos índices da marca de plural, que, na última variedade, beira o valor categórico.

No PST, em posição intermediária, registra-se, em relação ao PE, maior índice de não marcação de pluralidade (6,6% e 6,9%, respectivamente, no âmbito do

SN e do SV). O menor índice de falta de concordância, nos dados são tomenses em relação aos do PB, talvez possa ser atribuído ao fato de a norma adotada oficialmente no PST ser a europeia e alguns membros da comunidade terem um contato mais estreito com os portugueses, tanto por conta de viagens a Portugal, quanto pela presença de professores portugueses nas poucas instituições de ensino, como é o caso do quadro docente do Liceu Português, escola privada que atende, em geral, aos que têm melhor situação sócio-econômica. Os altos índices gerais de concordância, todavia, mascaram, de certa forma, uma situação que se supõe semelhante à descrita por Gonçalves (2000, p. 3), que, ao referir-se à variedade moçambicana, diz haver “um continuum polilectal, ao longo do qual se distribuem com maior ou menor frequência” diversos traços gramaticais, *continuum* esse “composto por diversas subvariedades que oscilam entre um extremo muito próximo do padrão europeu, e um extremo ‘basilectal’.”

O presente trabalho confirma as generalizações apresentadas em Lucchesi, Baxter e Ribeiro (2009, p.331), ao propor que a “variação da concordância no PE é na melhor das hipóteses residual e periférica.” Segundo o autor, situação distinta se verifica no PB, em que os índices variáveis de concordância poderiam ser dispostos num *continuum* de polarização sociolinguística, que iria da norma urbana culta (com 94% de concordância, como registrado em Graciosa (1991)) até as comunidades rurais afro-brasileiras (com apenas 16% de concordância, índice verificado por Lucchesi, Baxter e Ribeiro (2009)), *continuum* que evidenciaria o efeito do contato entre línguas. Nos dados brasileiros da presente pesquisa, adotando a proposta ora comentada, pode-se atribuir a alta concretização do plural (91,1% para a concordância nominal e 78,1% para a verbal) ao fato de se tratar de amostra da fala metropolitana do RJ, com acentuado prestígio dos padrões idealizados para os meios escolarizados.

Levando-se em conta exclusivamente os totais de realização de concordância *versus* não concordância no presente estudo, pode-se avaliar o estatuto de cada variedade analisada quanto ao fenômeno. Consoante Labov (2003), acredita-se que a marca de apenas 1% de uma das supostas variantes põe em risco a concepção de regra variável. Segundo o autor, o intervalo de 95 a 99% de concretização de dada forma alternante constituiria parâmetro para o estabelecimento de uma regra linguística semicategórica, e não efetivamente variável, regra que implicaria alternância de formas em 5 a 95% das ocorrências.

A distinção dos tipos de regras propostos pelo autor, longe de constituir mera diferenciação de natureza terminológica, permite observar se as realizações de determinado fenômeno linguístico se configuram mais como uma espécie de parâmetro gramatical,⁹ naturalmente caracterizador de determinada variedade

⁹ Foge ao escopo deste artigo qualquer compromisso em interpretar os fenômenos variáveis à luz do referencial da Teoria de princípios e parâmetros, do quadro gerativista de análise linguística; entretanto, entende-se

linguística / língua, do que como uma forma alternante, que figura, por sua própria natureza, tão disponível quanto outra que lhe é equivalente.

No caso da concordância, tanto a nominal quanto a verbal, embora as três variedades em estudo exibam forte preferência pela realização da marca de número, são os dados do PE (Cacém) aqueles que assumem comportamento que se poderia denominar de semicategórico no que se refere à concordância verbal (98,9%) e de categórico no que toca à concordância nominal (99,6%), tendo em vista que o único registro de ausência de marca em um SN pode ser caracterizado, como se indicará, adiante, como resultante de uma falha de planejamento discursivo. Diferentemente dessa realidade, encontram-se os dados coletados no PST e no PB, cujos índices são inferiores à marca de 95% de realização do plural.

Não obstante o baixo número de dados de não marcação, entende-se ser de fundamental importância, para a determinação do estatuto de cada variedade, a observação detalhada das ocorrências, para que se verifique se algumas variáveis apontadas como relevantes em diversos estudos sobre a concordância de número nominal e verbal de 3ª pessoa operam também no material estudado. Em outras palavras, além dos índices quantitativos, faz-se necessária uma cuidadosa análise qualitativa dos dados. No caso das variedades brasileira e africana do Português, foi possível fazer essa verificação com base no tratamento sociolinguístico provido pelo Goldvarb-X. Quanto à variedade europeia, o procedimento mais produtivo foi, sem dúvida, a observação cuidadosa das poucas ocorrências de constituintes sem a expressa marca de número.

Aspectos específicos

Da concordância nominal

No âmbito da concordância nominal, foram realizadas diferentes análises: (a) confrontaram-se o PB e o PST, levando em conta, por variedade, a fala de 18 indivíduos de três níveis de escolaridade distribuídos, ainda, por sexo e três faixas etárias; (b) focalizou-se apenas o PST (i) considerando, com a ampliação de 12 para 22 do número de informantes, apenas os indivíduos com níveis fundamental e médio de escolaridade e controlando uma nova variável: **frequência de uso de um crioulo**; (ii) avaliando a fala de nove estudantes entre 10 e 19 anos,¹⁰ considerando a série que estavam cursando e, ainda, a

que a tipologia de regras propostas em Labov (2003) seja correlacionável e possa oferecer contribuições aos propósitos daqueles que se ocupam da investigação paramétrica de origem formalista.

¹⁰ O *corpus* Tjerk é constituído de 111 entrevistas. Destas, apenas as nove utilizadas no referido estudo contemplam a fala de indivíduos entre 10 e 19 anos ainda em fase de escolarização.

variável **desempenho individual**; (c) observou-se o PE, com base em amostra eliciada da fala de indivíduos distribuídos segundo os parâmetros explicitados em (a). Os estudos descritos em (a) e (b) foram desenvolvidos com o auxílio do pacote de programas Goldvarb-X, tendo como unidade de análise cada constituinte flexionável do SN; o mencionado em (c) consistiu, com apoio em índices percentuais, na análise qualitativa das ocorrências de SNs tomados como unidade de análise.

Na tabela 2, a seguir, apresentam-se as variáveis selecionadas nas análises empreendidas com amostras do PB e do PST, chamando-se a atenção para alguns aspectos: (a) o baixo *input* da regra em todas as análises empreendidas (.004 a .039); (b) a atuação das variáveis **posição linear e relativa do constituinte no SN e nível de escolaridade**, que aparecem como as mais salientes para a compreensão do fenômeno (em primeiro ou segundo lugar) e **animacidade do núcleo e saliência fônica**, que, embora de forma não tão sistemática, se mostram também relevantes.

Tabela 2 - Variáveis selecionadas em diferentes análises sobre a ausência de marca de pluralidade no SN no PB e no PST.

(a) Confronto entre PB e PST	
PB	PST
Posição linear e relativa do constituinte no SN	Nível de escolaridade
Nível de escolaridade	Posição linear e relativa do constituinte no SN
Faixa etária	Animacidade do núcleo
Saliência fônica	Gênero
Animacidade do núcleo	Contexto fonológico subsequente
Ausência de marca: 338/3777 (8,9%) Significância: .001 Input: .039	Ausência de marca: 173/2612 (6,6%) Significância: .023 Input: .016
(b) Análises sobre o PST	
(i) Indivíduos de níveis médio e fundamental	(ii) Estudantes de 10 a 19 anos
Posição linear e relativa do constituinte no SN	Desempenho individual
Nível de escolaridade	Posição linear e relativa do constituinte no SN
Frequência de uso de um crioulo	
Saliência fônica	
Animacidade do núcleo	
Ausência de marca: 305/2375 (12,8%) Significância: .000 Input: .004	Ausência de marca: 31/633 (4,9%) Significância: .000 Input: .005

Fonte: Elaboração própria.

No que toca à variável **Posição linear e relativa do constituinte no SN**, como se registra na tabela 3, não obstante os baixos percentuais de não marcação da pluralidade, pode-se depreender, também na fala urbana do PB e do PST, uma restrição indicada como muito atuante em variedades não urbanas: a propensão à ausência da marca de número a partir do núcleo em segunda posição. Pode-se observar que, em quase todas as análises feitas, é a partir dessa posição que se registram pesos relativos acima de .50, indicando o favorecimento da não marcação.

Tabela 3 - Atuação da variável *Posição linear e relativa no SN* para ausência de marca de número, em diferentes análises com amostras do PB e do PST.

Análise/ Amostra	Pré-nuclear		Nuclear			Pós-nuclear		
	1	2/3	1	2	3/4	2	3	4/5
PB	19/1450	6/155	6/155	238/1628	18/194	9/95	34/109	10/35
	1,3%	3,6%	3,6%	14%	5,1%	9,5%	31,2%	28,6
	.17	.47	.40	.74	.68	.72	.90	.89
(a) PST	7/1022	3/122	3/67	121/1120	14/144	10/58	10/60	5/19
	1%	2%	4%	10%	9,7%	17,2%	10/60%	26,3
	.15	.50	.57	.74	.76	.91	.88	.94
(b) (i)	9/909	6/14	9/909	228/1092	25/122	11/44	16/44	8/15
	1%	5,8%	4,4%	20,9%	20,5%	25%	36,4%	53,3%
P	.14	.55	.36	.75	.76	.88	.93	.95
S	4/262	2/26	0/12	16/268	3/35		6/30	
T (ii)	1,5%	7,7%	0%	6%	8,6%		20%	
	.25	.77	---	.62	.79		.90	

Fonte: Elaboração própria.

Invertendo-se os valores dos pesos relativos – de ausência para presença de marcas – registrados na tabela acima, com o objetivo de melhor caracterizar as áreas do SN em que se concentra a tendência à marcação da pluralidade, chega-se ao *continuum* proposto a seguir, cuja validade poderia ser testada em outras variedades urbanas do Português. Leva-se em conta o que aqui se denomina de SN prototípico, isto é, aquele composto por constituinte(s) pré-nuclear(es), núcleo e constituinte(s) pós-nuclear(es). Em cada posição, indicam-se os PRs mínimos e máximos de presença da marca de plural. Assim, a primeira posição linear, no pré-núcleo, tende a ser mais marcada, o que sugere seja esse o *locus* por excelência da marca. A partir do núcleo em segunda posição, vai decrescendo gradativamente a presença da marca, quer se considerem os menores ou os maiores índices obtidos (em destaque por meio de sublinhado). No esquema proposto, a linha vertical contínua que separa o pré-núcleo do núcleo representaria o limite do espaço mais saliente

para a marcação (o *locus* esquerdo), enquanto a linha vertical pontilhada entre o núcleo e o pós-núcleo indicaria que essas áreas não sofreriam, quanto à tendência à não marcação, solução de continuidade.

Tabela 4 - Continuum de marcação de pluralidade em SNs prototípicos no PB e no PST, com base em PRs obtidos nas diferentes análises realizadas.

[+ marcas]				[- marcas]		
Pré-núcleo		Núcleo		Pós-núcleo		
Pos. 1	Pos. 2/3	Pos. 2	Pos. 3/4	Pos. 2	Pos. 3	Pos. 4/5
.75 - <u>.86</u>	.23 - <u>.53</u>	.25 - <u>.38</u>	.21 - <u>.32</u>	.09 - <u>.28</u>	.07 - <u>.12</u>	.05 - <u>.10</u>
<i>locus esquerdo</i>						

Fonte: Elaboração própria.

SNs em que um constituinte que normalmente ocupa posição à esquerda ou à direita muda de lugar podem comportar-se, quanto à concordância, segundo o *continuum* acima proposto: mais marcas à esquerda (*locus* esquerdo), menos à direita, como se observa nos exemplos em (a) contrapostos aos transcritos em (b):

- 1- (a) todos os seus aluno
- (b) aquelas coisa toda
- 2- (a) ter condições financeira boa
- (b) ter boas ideia

A estrutura em 2(a) exemplifica, ainda, o que está patente na Tabela X: núcleos em primeira posição tendem a receber a marca (PRs. .64 - 1.0, com a inversão dos valores), diferentemente do que ocorre quando em segunda (PRs. .25 - .38) ou terceira/quarta posição (PRs .21 - .32).

Ainda do ponto de vista estrutural, mencionem-se as variáveis **animacidade do núcleo** e **saliência fônica**, também selecionadas para o PB e o PST (em algumas das análises) e consideradas em conjunto na tabela 5.

Constata-se que núcleos com traço semântico [+animado] (PRs variando de .39 a .43), embora com índices próximos do obtido para os que apresentam traço [-animado] (PR. .55 em todas as análises), inibem a não marcação. Quanto à saliência fônica são os constituintes com menor diferenciação entre a forma singular e plural (PRs. .52 e .54) os que mais desfavorecem a marca.

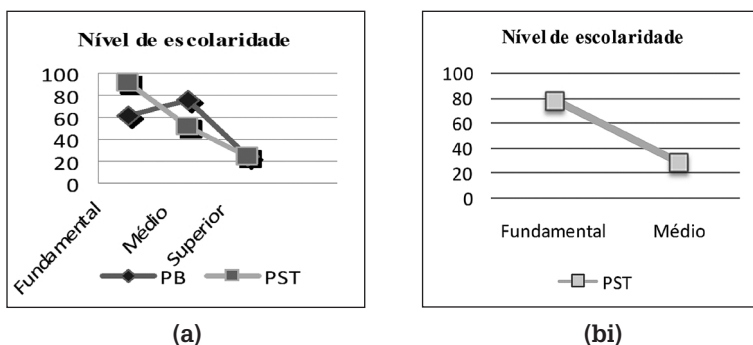
Tabela 5 - Atuação das variáveis animacidade do núcleo e saliência fônica para a não marcação de número no SN em diferentes análises com amostras do PB e do PST.

Análise/ Amostra	Animacidade do núcleo		Saliência fônica		
	[+ animado]	[-animado]	[+saliente]	[-saliente]	
(a)	PB	95/838	166/1095	25/374	313/3403
		11,4%	15,2%	6,7%	9,2%
	PST	.43	.55	.30	.52
		36/489	102/842	_____ ¹⁰	_____
(b)	PST (i)	7%	12%		
		.40	.55		
	PST (ii)	59/433	195/826	21/218	268/1.182
		13,6%	23,6%	9,6%	22,7%
	.39	.55	.25	.54	
	(ii)	_____	_____	_____	_____

Fonte: Elaboração própria.¹¹

Cabe, agora, discutir a variável **nível de escolaridade**, que parece desempenhar papel preponderante para a implementação e a produtividade dos padrões variáveis de concordância comuns ao PB e ao PST.

Gráfico 3 - Atuação da variável nível de escolaridade para a não marcação de número no SN em diferentes análises sobre o PB e o PST (pesos relativos).



Fonte: Elaboração própria.

Como se salientou inicialmente, os altos índices globais de concordância obtidos nas análises das duas variedades não europeias não deixam transparecer,

¹¹ O traço apresentado nas tabelas indica que a variável não foi selecionada na análise em questão.

de imediato, a existência de um *continuum* em cujos pólos se encontram os padrões usados pelos indivíduos menos escolarizados e os utilizados pelos de nível superior. Em estudo de natureza geolinguística, em que comentava os três *continua* – de urbanização, de oralidade-letramento, de monitoração estilística – segundo os quais Bortoni-Ricardo (2004, p.51-60) sugere que se estude a variação no PB, Brandão (2011) diz ser necessário ter em conta “[...] *un quatrième continuum - le niveau d’instruction - non seulement parce qu’il s’agit d’un paramètre fondamental pour représenter les distances sociales au Brésil, mais aussi pour se rattacher étroitement et parfois se superposer au continuum d’urbanisation [...]*”, que se apresenta mais adiante, adaptado ao que aqui se discute.

O gráfico 3(a), que retrata os resultados referentes ao confronto entre o PB e o PST realizado com base na fala de indivíduos de níveis fundamental, médio e superior de instrução, mostra que, tanto no PB quanto no PST, há, em relação à concordância nominal, uma situação sociolinguisticamente polarizada, reforçada pelo fato de o nível de escolaridade, em geral, se associar a outras variáveis, como nível socioeconômico, maior ou menor acesso a bens culturais, maior ou menor contato com indivíduos de diferentes estratos sociais.

Quadro 2 - Continuum de nível de instrução, segundo Brandão (2011), adaptado à ausência da marca de concordância nominal no PB e no PST.

Continuum de nível de instrução			
[- marcas] [+ marcas]	Fundamental	Médio	Superior
Não escolarizados/ Escolaridade: 1 a 4 anos		PB: (.62-.76)	PB: .24
	PST: .90 / .80	PST: .53 / .29	PST: .15 / --

Fonte: Elaboração própria.

Observe-se, no entanto, que, no âmbito da variedade brasileira, que já dispõe de uma norma consolidada, os indivíduos de níveis fundamental e médio se opõem aos de nível superior, enquanto na variedade santomense o caráter francamente escalar de adoção da norma canônica de concordância fica patente.

A análise do PST em que se confrontaram apenas indivíduos de nível médio e fundamental (bi) e aquela em que se focalizou a fala de estudantes de 10 a 19 anos (bii) dão conta da complexidade de fatores que concorrem para a variação.

Na primeira delas, a variável **frequência de uso de um crioulo** foi a terceira mais relevante para a não aplicação da marca, suplantada apenas pelas variáveis referentes à posição dos constituintes e à escolaridade. Embora todos os indivíduos que concorreram para a formação da amostra tenham se declarado falantes de

português como L1, tem-se de levar em consideração o caráter multilíngue da sociedade santomense, em que coexistem três crioulos, além de outras línguas minoritárias: há indivíduos que (a) só usam o português e (i) não dominam ou (ii) dominam um crioulo; (b) usam o português e um ou mais crioulos (i) com menor ou (ii) maior frequência (no mercado, em momentos de lazer com amigos), e assim por diante. O uso do crioulo tem caráter estigmatizante, haja vista que muitos dos informantes mencionaram o fato de seus pais obrigá-los a falar, mesmo em casa, apenas o português, para que, dominando as normas de prestígio, se tomassem aptos a aproveitar as poucas oportunidades de ascensão socioeconômica.

Diante desse quadro, buscou-se verificar se os indivíduos que fazem uso mais frequente de um crioulo (em geral, o forro) que do português teriam maior probabilidade de não usarem a marca de plural. A variável ficou assim definida, quanto à frequência de uso de um crioulo:

- **zero/baixa**, os indivíduos que se expressam fundamentalmente em português;
- **média**, os indivíduos que se expressam em português, mas dominam um crioulo e dele fazem uso eventualmente;
- **alta**, os indivíduos que, embora falem o português, se expressam, regularmente, num crioulo.

Na tabela 6, expõem-se os resultados referentes à atuação da variável.

Tabela 6 - Atuação da variável *frequência de uso de um crioulo para o cancelamento da marca de número no SN.*

Frequência de uso de um crioulo	Oco	Perc.	P.R.
Zero /Baixa (grau zero - G0)	75/1039	7.2%	.41
Média (grau 1 - G1)	140/1117	12.5%	.52
Alta (grau 2 - G2)	90/219	41.1%	.75

Fonte: Elaboração própria.

A classificação dos informantes quanto aos fatores da variável teve por base informações prestadas por eles durante as entrevistas. No primeiro fator, inscrevem-se nove informantes; no segundo, dez e, no terceiro, três. Independentemente dessa distribuição assimétrica, constata-se que são os indivíduos que se comunicam preferencialmente num crioulo os que mais tendem a não implementar a marca de número (41.1%; P.R. .75). Com esses indivíduos, contrastam os que nunca ou pouco se expressam em crioulo (P.R. .41) e os que o utilizam apenas eventualmente (p.r. 52).

A análise variacionista sobre a fala de estudantes santomenses, embora não se tenha realizado nos moldes clássicos – uma vez que se baseou na fala de apenas

nove informantes e não houve o preenchimento de todas as células sociais –, indicou a variável **desempenho individual** como a mais relevante. Na tabela a seguir, sintetizam-se os resultados.

Tabela 7 - Atuação da variável *desempenho individual* para a ausência de marca de plural em constituintes do SN, com base em amostra da fala de estudantes de São Tomé.

Informante	Variável <i>desempenho individual</i>			Informante	Oco	%	P.R.
	Oco	%	P.R.				
ST-E1-E6m	8/19	42.1%	.91	ST-E6-FDh	8/88	9.1%	.48
ST-E2E6m	0/28	0%	---	ST-E7-FDh	3/38	7.9%	.38
ST-E3-F8h	0/62	0%	---	ST-E8-FDh	3/83	3.6%	.26
ST-E4-F8m	0/261	0%	---	ST-E9-FDm	0/91	0%	---
ST-E5-F8m	9/26	34.6%	.91				
Input: .05					Significância: .000		

Fonte: Elaboração própria.

Os resultados sugerem que a concordância nominal tem, na sociedade santomense, um forte componente sócio-econômico-cultural. Independentemente do nível escolar em que se encontrem, enquanto, na fala de quatro estudantes, a regra tem caráter categórico, na de cinco outros, tem caráter variável, em maior ou menor grau. Tal quadro, certamente, vincula-se a aspectos que dizem respeito ao ambiente familiar, à maior ou menor exposição aos bens culturais, às línguas faladas na região, ao tipo de escola que frequentam. Vale referir as observações de dois dos estudantes que aplicam categoricamente a regra: um afirmou que o pai lhe propicia todos os meios para seu desenvolvimento intelectual; outro, que os pais preferem que estude na Escola Portuguesa por acharem que aquela escola, diferentemente das públicas, funciona de forma mais regular e, a princípio, propiciaria melhor qualidade de ensino.

Na fala dos estudantes que aplicam variavelmente a regra de concordância, as restrições que regulam a marcação ou não da pluralidade dizem respeito à **posição linear e relativa dos constituintes**, nos moldes aqui já discutidos.

Contrasta fortemente com o quadro descrito para o PB e o PST a variedade europeia, em que a regra de concordância nominal parece ter caráter categórico. A análise de 2.449 SNs, a grande maioria (79,40%) de dois constituintes, presentes na fala de 18 indivíduos de ambos os sexos e diferentes faixas etárias e níveis de escolaridade, revelou uma taxa de 99,96% de aplicação da marca de número. O índice de 0,04% de ausência de marca corresponde a um único SN, constituído

por um numeral e um substantivo, que foi produzido por um homem da faixa etária mais alta e com nível fundamental de escolaridade.

Como se poderá aquilatar pelo contexto que se transcreve abaixo, o cancelamento da marca parece ter ocorrido por uma falha de planejamento discursivo, um lapso, como indica Labov (2003) quando explicita o que seria uma regra categórica em oposição a regras semicategóricas e variáveis. O indivíduo organiza seu discurso no sentido de mencionar apenas uma praça, “**uma** pracetazinha”, como ele a caracteriza, continua e reitera que é **uma** e, ao se dar conta de que são **duas**, parece não ter organizado integralmente o enunciado, deixando, assim, de produzir a marca no substantivo.

[...] nada...ali nada tem **uma** pracetazinha...há **uma** ou **duas** **praceta** onde os miúdos se entretêm onde há menos carros pra jogar um bocadinho à bola ou andarem de bicicleta de resto não tem lá nada... aquela zona é mesmo mesmo mesmo sem nada. (Informante CACC1H).

Da concordância verbal

De forma semelhante ao que foi descrito no âmbito da concordância nominal, foram realizadas diferentes análises para o estudo da concordância verbal: (a) confrontaram-se o PB e o PST, levando em conta, em cada localidade, a fala de 18 indivíduos¹² de três níveis de escolaridade distribuídos, ainda, por sexo e três faixas etárias; (b) focalizou-se apenas o PST¹³ considerando a ampliação do número de informantes, de modo que houve maior número de indivíduos com níveis fundamental e médio de escolaridade (22 informantes), e controlando uma nova variável: **frequência de uso de um crioulo**; (c) observou-se o PE, também com base em amostra eliciada da fala de indivíduos distribuídos segundo os parâmetros explicitados em (a). As análises descritas em (a) e (b) também foram desenvolvidas com o auxílio do pacote de programas Goldvarb-X; por impossibilidade decorrente da falta de contextos variáveis, a mencionada em (c) baseou-se fundamentalmente na análise qualitativa dos dados encontrados.

Na tabela 8, a seguir, apresentam-se as variáveis selecionadas nas análises empreendidas com amostras do PB e do PST:

¹² Conforme já se esclareceu, apenas para o PST, o estudo só pôde contar com 17 informantes, o que resulta da dificuldade de localizar falantes de Português com curso superior.

¹³ Diferentemente da análise da concordância nominal, o estudo da concordância verbal também considerou nessa etapa da investigação (b) os informantes de curso superior; além disso, até o momento da pesquisa, não foi realizada a análise apenas com informantes de 10 a 19 anos.

Tabela 8 - Variáveis selecionadas em diferentes análises sobre a ausência de marca de pluralidade no verbo no PB e no PST.

(a) Confronto entre PB e PST	
PB	PST
Paralelismo no nível oracional	Nível de escolaridade
Nível de escolaridade	Sexo
Saliência fônica	Saliência fônica
Animacidade do sujeito	Animacidade do sujeito
Posição do sujeito em relação ao verbo	Faixa etária
Faixa etária	Posição do sujeito em relação ao verbo
Ausência de marca: 338/3777 (8,9%)	Ausência de marca: 47/687 (6,8%)
Significância: .004	Significância: 013
Input: .014	Input: .016

(b) Análise adicional sobre o PST	
Ampliação do número de indivíduos de níveis médio e fundamental + superior	
Nível de escolaridade	
Animacidade do sujeito	
Posição do sujeito em relação ao verbo	
Sexo	
Paralelismo no nível oracional	

Ausência de marca: 100/1053 (9,5%)
 Significância: .036
 Input: .049

Fonte: Elaboração própria.

No âmbito da concordância verbal, também chama a atenção o baixo *input* da regra em todas as análises empreendidas (.004 a .049), confirmando ser a concordância a estratégia preferencial nos dados coletados no Brasil e em São Tomé. Quanto às variáveis relevantes à compreensão do funcionamento dos dados, destaca-se a atuação da variável extralinguística **nível de escolaridade**, cuja descrição permitirá a comparação com os resultados relativos à concordância nominal. As variáveis estruturais **animacidade do núcleo** e **posição do sujeito em relação ao** verbo, em primeiro plano, e **saliência fônica**, em segundo, também se mostraram relevantes em relação ao fenômeno.

A tabela 9, a seguir, permite observar o comportamento das variáveis de cunho estrutural, considerando as análises desenvolvidas.

Tabela 9 - Atuação das variáveis *animacidade do núcleo* e *posição do sujeito em relação ao verbo* para a não marcação de número no verbo em diferentes análises com amostras do PB e do PST.

Análise/ Amostra	Animacidade do núcleo		Posição do sujeito em relação ao verbo ¹³		
	[+ animado]	[-animado]	[anteposto]	[posposto]	
(a)		206/1133	55/164	143/947	21/40
	PB	18,2%	33,5%	15,1%	52,5%
		.46	.70	.41	.81
		33/585	14/102	27/467	6/28
	PST	5,6%	13,7%	5,8%	21,4%
		.45	.71	.31	.77
(b)		63/879	37/174	55/688	15/53
	PST	7,1%	21,2%	7,9%	28,3%
		.43	.78	.46	.86

Fonte: Elaboração própria. ¹⁴

No que se refere ao traço semântico de animacidade, fica evidente que, nas três análises, a presença de referente [+animado] no SN sujeito (PRs variando de .43 a .45) desfavorece a não marcação, ao contrário do que ocorre com o traço [-animado], que nitidamente a favorece (PRs variando de .70 a .78).

Quanto à posição do sujeito em relação ao verbo, também fica comprovada a conhecida tendência: sujeitos pospostos, ao contrário dos antepostos, favorecem a ausência de marcação plural (PRs variando de .77 a .86).

A variável saliência fônica – que não se mostrou estatisticamente relevante apenas no caso da análise adicional realizada para o PST¹⁵ – foi controlada, para os dados de concordância verbal, consoante quatro graus de diferenciação das formas singular e plural dos verbos, conforme se expõe na tabela a seguir.

¹⁴ No controle desta variável, foram desconsideradas as ocorrências com sujeito não expresso.

¹⁵ Embora o programa Goldvarb-X não tenha selecionado a variável saliência fônica na amostra em questão, ficou claro, pela observação dos dados e pelo cruzamento entre grupos de fatores, que a atuação do referido grupo estava associado ao tipo de verbo em pauta quanto à transitividade verbal, o que está sendo apurado com mais detalhamento na presente etapa da investigação.

Tabela 10 - Atuação da variável *saliência fônica* para o cancelamento da marca de número no verbo na análise das amostras do PB e do PST.

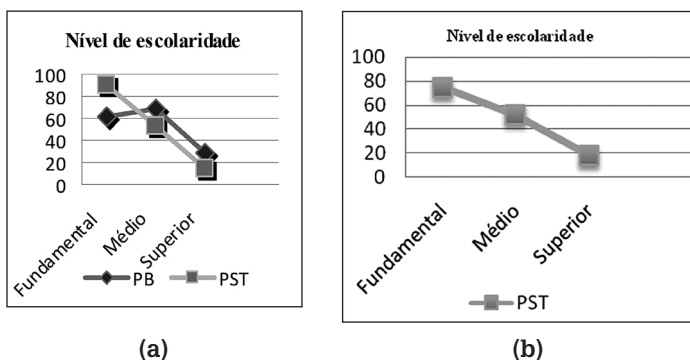
Análise/ Amostra	Saliência fônica			
	Grau 1 <i>come-comem; fala-falam</i>	Grau 2 <i>dá-dão; vai-vão</i>	Grau 3 <i>comeu-comeram; quis-quiseram</i>	Grau 4 <i>veio-vieram; é-são</i>
(a) PB	187/783 23,9%	46/226 20,4%	15/161 9,3%	13/127 10,2%
	.60	.46	.25	.26
PST	38/450 8,4%	2/79 2,5%	4/93 4,3%	3/65 4,6%
	.63	.17	.37	.26
(b) PST	—	—	—	—

Fonte: Elaboração própria.

Nas duas variedades do Português, constata-se maior tendência ao cancelamento da marca de número em verbos que apresentam pouca diferenciação entre as formas singular e plural (.60 no PB e .63 no PST) em relação aos demais (PRs variando entre .25 e .46 no PB e entre .17 e .37 no PST). Na amostra do PB, registram-se índices decrescentes de cancelamento da marca de número consoante o aumento da diferenciação fônica entre as formas verbais. Na amostra do PST, embora não se verifiquem tais índices progressivamente decrescentes, o efeito da **saliência fônica** pode ser aferido do contraste entre os índices obtidos para o grau 1 – que favorecem o cancelamento da marca de número – e os obtidos para os demais graus – que a desfavorecem.

Como se pode observar, fica patente a similaridade na atuação das variáveis estruturais. Resta, então, descrever o comportamento dos dados quanto ao grupo de fatores de cunho extralinguístico, que, como já se observou, desempenha papel preponderante para a implementação e a produtividade dos padrões variáveis de concordância comuns ao PB e ao PST, a variável **nível de escolaridade**. As figuras a seguir permitem observar, agora, o comportamento dos dados relativos à concordância verbal.

Gráfico 4 - Atuação da variável nível de escolaridade para a não marcação de número no verbo em diferentes análises sobre o PB e o PST (pesos relativos).



Fonte: Elaboração própria.

Os gráficos 4(a) e (b) demonstram que, também no âmbito da concordância verbal, tanto no PB quanto no PST, a concretização das marcas de número se associa fortemente ao nível de escolaridade do indivíduo, o que, observados os extremos, configura uma situação sociolinguisticamente polarizada. Na variedade brasileira, os indivíduos de níveis fundamental e médio (com índices acima de .50, indicativos de favorecimento da ausência de marca) opõem-se aos de nível superior (com nítido desfavorecimento da ausência de marca), enquanto na variedade santomense fica nítido que a adoção da norma canônica de concordância é progressiva. Assim, como demonstra o quadro 3, a seguir, também no eixo da concordância verbal, podem-se representar os *continua* de nível de instrução, tais como se configuram no PB e no PST.

Quadro 3 - Continuum de nível de instrução, segundo Brandão (2011), adaptado à ausência da marca de concordância verbal no PB e no PST.

Continuum de nível de instrução			
[- marcas]			[+ marcas]
Não escolarizados/ Escolaridade: 1 a 4 anos	Fundamental	Médio	Superior
	PB (.61-.69)		PB: .29
	PST: .90 / .75	PST: .53 / .52	PST: .15 / .18

Fonte: Elaboração própria.

A análise adicional do PST desenvolvida com os dados referentes à concordância verbal em que se utilizou maior número de indivíduos com nível médio e fundamental, além daqueles que cursaram o nível superior, também permitiu observar a atuação da variável **frequência de uso de um crioulo**.

Embora esse grupo de fatores não tenha sido selecionado pelo Programa na análise estatística dos dados, o cruzamento entre ele e a variável **nível de escolaridade** foi esclarecedor.

Tabela 11 - Cruzamento entre as variáveis *nível de escolaridade e frequência de uso de um crioulo para o cancelamento da marca de número no verbo.*

Frequência de uso de um crioulo	Ensino Fundamental	Ensino Médio	Ensino superior
Zero /Baixa (grau zero - G0)	14/105 13%	17/195 9%	4/142 3%
Média (grau 1 - G1)	26/154 17%	18/211 9%	3/157 2%
Alta (grau 2 - G2)	18/89 20%	-----	-----

Fonte: Elaboração própria.

Embora seja, de fato, a escolaridade o fator que faz mudar as tendências relativas ao fenômeno, a tabela 11 permite visualizar que, em paralelo ao acesso à escolarização, está a maior ou menor utilização da língua crioula. Só declararam que usam frequentemente o crioulo informantes que cursaram até o ensino fundamental, nível de escolaridade em que se registra maior cancelamento da marca de número. De fato, o prestígio que recebe o Português em terras são-tomenses e o maior ou menor propósito de dominar essa língua, o que pode se tornar possível graças ao acesso à escolaridade, trazem por consequência o maior ou menor domínio das regras canônicas de concordância.

Também no caso da concordância verbal, é absolutamente distinto o comportamento dos dados da variedade europeia, representada aqui pela comunidade de Cacém,¹⁶ em comparação ao descrito para o PB e o PST: no PE, a regra de concordância verbal assume um perfil de caráter semicategórico.

Das 1.515 ocorrências de construções de 3ª pessoa do plural, produzidas pelos dezoito informantes de ambos os sexos e diferentes faixas etárias e níveis de escolaridade, apenas dezessete não apresentaram marca de concordância verbal, o que corresponde a apenas 1,1%. Em termos extralinguísticos, cabe ainda informar que esses dados foram produzidos por apenas dez dos dezoito informantes considerados; em outras palavras, na fala de oito informantes, as marcas verbais plurais ocorreram em todos os dados, de forma categórica.

¹⁶ Já se encontram tratados os dados coletados em Oeiras/Lisboa, os quais não apresentaram comportamento substancialmente diferente do ora descrito, nem quantitativa nem qualitativamente. Para maiores detalhes acerca dos dados dessa localidade, consultar Vieira (2011). A fim de manter a comparabilidade dos resultados com os exibidos para a concordância nominal, consideram-se aqui exclusivamente as ocorrências coletadas na localidade de Cacém.

Tendo em vista que só se registraram dezessete casos de suposta não concordância na amostra do PE, passa-se a apresentar os tipos de estruturas que representam esses dados. A observação das dezessete ocorrências, uma a uma, permitiu detectar os padrões estruturais encontrados. Sistematizando os resultados dessa observação, procedeu-se à distribuição das ocorrências pelos seguintes contextos:

(i) sete dados de estruturas com o verbo “ser” em construções predicativas (exemplos 1 e 2):

(1) D: e: os seus amigos? Que interesse é que têm? Têm os mesmos interesses que o senhor o? L: a/a: os meus amigos meus são de engenharia digamos assim muitos deles **é** de engenharia há muitos de economia [...] (CACA1H)

(2) as melhoras **seria** eu em vez de ganhar quinhentos é assim não vamos ganhar quinhentos vamos ganhar só quatrocentos (CACC1H)

Cabe lembrar que construções do tipo SN sujeito + verbo *ser* + predicativo do sujeito costumam ser caracterizadas de forma particular, especialmente no que tange à concordância. Por vezes, admite-se, desde a tradição gramatical, que a relação de concordância se dê entre o verbo e o constituinte considerado predicativo. Além disso, essas construções podem ser consideradas estruturas “espelhadas”, de modo que as posições de sujeito e predicativo se invertem facilmente (como em “a mudança é a solução” ou “a solução é a mudança”).

(ii) cinco dados de posposição do sujeito (exemplos 3 e 4):

(3) D: há o comboio ali da linha de Sintra
L: sim sim sim passa mesmo no centro do Cacém e depois a partir da estação **existe**::: ... os autocarros para os diversos sítios (CACA3M)

(4) L: tal e qual eles passam a vida a ver aquelas anúncios...então nas épocas de Natal aquelas tretas todas aquilo é horrível mas é assim eles não têm :m não sei **escapa**-lhes as coisas não é? porque eu acho que com a idade deles apesar deles serem crianças eu já tinha outra maturidade pra ver que aquilo não era possível...(CACB3M)

As construções com sujeito posposto constituem, na realidade, um dos poucos contextos estruturais, amplamente debatido na literatura, em que as três variedades igualmente registram casos de cancelamento da marca de número plural. Há que se questionar o estatuto dessas construções quanto à identidade do sujeito. Sabe-se que, além da questão do processamento da informação, as construções com sujeitos pospostos são, em sua maioria, integrantes de categorias especiais de verbos, como os intransitivos e os inacusativos.

(iii) três dados em que o suposto referente sujeito não se encontra adjacente ao verbo (exemplos 5 e 6).

(5) D: não há o mínimo investimento?

L: nada prali não há nada nada e podiam fazer () que aquilo é uma zona onde esta zona aqui é onde eu moro pertence ao Cacém e esta que (é a rua) esta... os prédios daquele lado na mesma rua... já **pertence** a Rio de mouro...

(6) – oh minha menina eu onde eu trabalho já fomos assaltados duas vezes... duas vezes foram esses imigrantes que por aí entram aí não têm nada que fazer vêm para cá.. só **estraga**... e prontos foi uma experiência que não quero repetir espero que não volte a acontecer (CACB1H).

Nesses exemplos, cabe observar que a própria identificação do sujeito pode ser questionada: ser interpretado como nulo com um referente singular. No exemplo 6, é possível propor que “estragar” se relaciona a toda a situação descrita anteriormente (“isso tudo”, “essa situação”) e não especificamente aos “imigrantes” de que se falava.

(iv) dois dados em que o referente-sujeito foi retomado por pronome-cópia:

(7) L: muito alta é muito diferente de todas as outras brasileiras que eu conheço... é um estereótipo completamente diferente de todas as outras brasileiras que eu conheço... já as minhas vizinhas são todas brasileiras já são dentro do estereótipo de brasileiras que eu conheço... assim mais com um cabelo encaracolado uma pele mais morena tem umas feiçõeszinhas que **identifica** logo que elas são brasileiras...agora se ela fala a gente vê logo que ela não é de cá assim (CACA1M)

(8) é por quem lá tá dentro...não vão fazer leis a favorecer o mais pequeno... todas as leis que **é** feito a nível de isto ou de aquilo e de acoloutro é tudo leis (CACC1H)

A presença do pronome relativo *que* nesses exemplos constitui, sem dúvida, fator importante para a realização da forma verbal singular.

(v) um dado de infinitivo pessoal (exemplo 9):

(9) L: talvez pelo tal... por não haver tanta proximidade com os pais...no contacto...e pronto e se calhar o :o...a pô/a presença dos pais não ser tão efectiva e acaba por eles se **tornar** tão autónomos que quando alguém pede/ diz alguma coisa eles começam logo a disparatar (CACB3H)

A forma infinitiva constitui por si só um contexto absolutamente particular, motivo pelo qual muitos estudos variacionistas nem sequer a consideram na contagem dos dados.

Levando em conta o número de dados e os contextos em que se pôde comprovar efetivamente a ausência da marca plural no verbo, ao menos duas conclusões podem ser tiradas:

- (i) em termos quantitativos, a suposta falta de concordância no Português Europeu lisboeta configura-se um caso de regra semicategórica (LABOV, 2003);
- (ii) em termos qualitativos, a suposta falta de concordância no PE lisboeta está concentrada em contextos linguísticos específicos e não ocorre na diversidade de contextos já verificados no PB e no PST.

Reflexões sobre os resultados relativos à concordância nominal e à verbal: o estatuto das variedades do Português

Desenvolvidas as análises variacionistas sobre concordância nominal e verbal com dados das variedades brasileira, são-tomense e europeia, foi possível delinear o quadro a seguir, em que se observam aspectos divergentes e convergentes:

(i) parece haver, em termos percentuais gerais, tendência à adoção da regra de concordância tanto no SN quanto no SV, embora com estatuto diferenciado, nas diferentes áreas urbanas aqui consideradas:

- (a) na do Português Europeu, os dados revelam um comportamento que não pode ser considerado efetivamente variável – trata-se, nos termos de Labov (2003), de uma regra categórica no âmbito da concordância nominal e semicategórica no âmbito da verbal;
- (b) nas do Português do Brasil e do Português de São Tomé, embora se dê preferência à concordância – preferência compatível com meios em que o acesso à escola, em maior ou menor grau, é facultado aos indivíduos –, ambas as regras são variáveis, salientando-se que o PST, que apresenta maiores índices gerais de marcação de pluralidade do que o PB, é a variedade em que o *continuum* de nível de instrução tem caráter nitidamente escalar.

(ii) variáveis de cunho estrutural nos níveis sintático (**posição linear e relativa dos constituintes do SN; posição do sujeito em relação ao verbo**), semântico (**animacidade**) e fonético (**saliência fônica**) atuam como restrições na concretização das marcas de pluralidade, comprovadamente no PB e no PST, de acordo com as seguintes tendências:

- (a) nas estruturas sintáticas em que cabe indicar pluralidade por meio de flexão, a sua área esquerda é a que mais predispõe à presença de marca:
 - (i) nos constituintes pré-nucleares, quando se trata do SN considerado isoladamente, e no constituinte nuclear, quando ele se encontra em primeira posição;
 - (ii) no verbo, quando o SN sujeito está à sua esquerda.
- (b) o traço [+ animado] ou [-animado] do núcleo do SN repercute, respectivamente, na presença ou ausência da marca de número tanto no âmbito do SN, quanto no do SV;
- (c) menor ou maior grau de saliência fônica condiciona, inversamente, maior ou menor índice de marcação da pluralidade nas formas nominais e verbais.

(iii) o **nível de escolaridade** dos indivíduos atua como variável relevante no PB e no PST, motivação que, por ora, não pode ser atestada no PE.

Essas conclusões advindas das análises sugerem que, em dados contemporâneos, o estatuto da concordância no PE é de natureza absolutamente diferente do assumido nas variedades não europeias do Português ora estudadas. No que se refere à quantidade dos dados, não se detectou variação no âmbito da concordância nominal e registraram-se apenas 17 ocorrências de formas verbais no singular em construções em que se poderia propor a existência de SN sujeito plural.

Quanto à qualidade das construções analisadas, nas poucas estruturas relacionadas à concordância verbal, aventa-se a hipótese de que construções específicas em que não se dá a marca de plural também ocorrem no PE. Considerando apenas os resultados da presente investigação, pode-se supor que a posição do sujeito, a distância do constituinte sujeito ou, ao menos, do enunciado que evoca o referente sujeito, especialmente quando retomado pelo pronome relativo *que*, constituiriam evidências de que alguns parâmetros atuantes no PB e/ou no PST agiriam no PE.

Estudos científicos (NARO; SCHERRE, 2007; VAREJÃO, 2006; MONGUILHOT, 2009) declaram ter comprovado em dados europeus a atuação de outras variáveis, como, por exemplo, a saliência fônica e a animacidade do sujeito. Com base na observação de alguns exemplos e na atuação dessas variáveis, Naro e Scherre (2007, p.114-115) chegam a propor que “[...] o português brasileiro e o português europeu partilham estruturas semelhantes e que as estruturas variáveis de uso comum no Brasil também podem ser encontradas em Portugal, com menos intensidade, tanto hoje quanto antes da colonização.”

Embora não constitua objetivo do presente artigo, é fundamental que se comparem os dados da presente pesquisa com os de outras investigações sociolinguísticas com amostras do PE, sobretudo no âmbito do sintagma verbal.

Trabalhos como os de Varejão (2006) e Monguilhot (2010) registram percentuais gerais de não marcação mais altos do que os verificados nesta pesquisa (cerca de 10% em amostra de indivíduos não escolarizados do *corpus* Cordial-Sin, no primeiro caso, e cerca de 8%, no segundo, considerando amostra com indivíduos com pelo menos o nível fundamental de escolaridade), percentuais que permitiriam postular a existência de uma regra variável em determinadas variedades do PE. Por ora, uma observação preliminar desses trabalhos sugere que tenha havido diferença tanto na coleta de dados considerados variáveis, quanto na interpretação de determinadas estruturas como reveladoras de falta de concordância.

Nesse sentido, considera-se apropriada a hipótese de Naro e Scherre (2007, p.85) relativa à possibilidade de haver, mesmo no Português Europeu contemporâneo, padrões em diversas regiões de Portugal: “[...] são desconhecidas a verdadeira extensão e a intensidade da variação na concordância em terras lusitanas.” É preciso ter claro, por exemplo, que não se dispõe de trabalhos variacionistas baseados em amostras do PE falado por informantes não escolarizados.¹⁷

Considerando no debate ora apresentado apenas a realidade das comunidades urbanas representadas nos *corpora* estudados, pode-se afirmar que os dados europeus de não marcação de plural não ocorrem apenas com menos intensidade; antes, são periféricos quantitativa – como já se comprovou – e qualitativamente – visto que, no âmbito da concordância verbal, só foram encontrados dados de não marcação com verbo copulativo, com sujeitos pospostos ou com sujeitos antepostos cujos referentes são questionáveis ou expressos na forma de pronome cópia. Ainda assim, entende-se ser necessária uma reflexão adicional acerca da natureza das possíveis motivações estruturais para a não marcação de plural, mesmo que sugeridas por um pequeno número de ocorrências.

Em termos qualitativos, supõe-se que as restrições estruturais que atuam inequivocamente nos dados do PB e do PST, por princípio, poderiam atuar no âmbito de qualquer língua, visto que se referem, tanto no âmbito do SN quanto do SV, a componentes de ordem cognitivo-processual que se revelam nos níveis sonoro (saliência fônica), sintático (saliência posicional) e semântico (saliência do traço + animado). Fundamenta essa hipótese o que demonstra, por exemplo, Corbett (2000), ao atestar que, em línguas de perfis bem diferentes, a presença de sujeito posposto faz diminuir a ocorrência da marca de número.

¹⁷ A tese de Doutorado de Alexandre Monte, recentemente defendida na UNESP, contempla entrevistas sociolinguísticas de portugueses em Évora, escolarizados e não escolarizados.

Tabela 12 - Agreement with conjoined noun phrases.

	Animate		Inanimate	
	N	%PL	N	%PL
Subject-predicate				
Medieval Spanish	288	96	243	31
German	1.095	96	1.702	67
Russian	115	100	67	85
Serbo-Croat	21	100	35	91
Predicate-subject				
Medieval Spanish	318	69	239	6
German	379	93	925	40
Russian	89	84	114	28
Serbo-Croat	23	70	62	26

Fonte: Corbett (2000, p.201).

A tabela acima deixa claro que os percentuais de marcas, tanto com sujeito animado quanto com sujeito inanimado, os deste sempre mais baixos do que os daquele, são maiores na ordem sujeito-predicado do que na ordem predicado-sujeito, em línguas românicas ou não. Evidências como essas dão suporte à proposta de que a natureza dessas variáveis – que lidam com algum nível de saliência – está relacionada à atividade linguística geral e não se presta propriamente à indicação dos parâmetros gramaticais que caracterizam cada variedade ou cada língua.

Voltando aos dados da presente pesquisa, há indícios de que é a atuação de determinados fatores externos que acelera ou detém a implementação das restrições linguísticas. Nas realidades brasileira e são-tomense, ao que tudo indica, o valor social que se atribui à não concordância seria compatível com um estereótipo (nos termos de Labov (1972)) ou, na melhor das hipóteses, com um marcador. Por essa razão, o fenômeno variável funciona como nítido identificador do perfil social do usuário da língua. No PST – que configura um caso de variedade do português em formação, com normas objetivas ainda não totalmente definidas e assumidas pelos diferentes grupos de falantes –, sabe-se que dominar a concordância segundo a norma culta que lhe serve de modelo, a do Português Europeu, seria um dos índices de identificação de pleno domínio da Língua Portuguesa. No PB, ao contrário, o uso ou não da concordância canônica é uma opção que está claramente no nível de consciência dos indivíduos, de modo que é interpretado e funciona como forte traço caracterizador de classes sociais. Na realidade europeia, os dados desta pesquisa não permitem inferir diferença de comportamento vinculada a qualquer variável social, o que faz crer que a concordância pode constituir um parâmetro gramatical naturalmente assumido pelos falantes.

A confirmação dessas hipóteses depende, obviamente, de extensa pesquisa com pelo menos as seguintes diretrizes: (i) utilizar outras amostras de maior abrangência em termos diatópicos e diastráticos, (ii) levar em conta dados da aquisição, (iii) empregar metodologia apropriada à investigação de atitudes e, ainda, (iv) desenvolver comparação interlinguística. São necessários, sem dúvida, não só estudos de cunho sociolinguístico de outras variedades urbanas do português, mas também a observação dos mecanismos de concordância em outros sistemas linguísticos, de modo a aferir a universalidade das interpretações aqui propostas.

Por ora, ficam a reflexão e o convite à pesquisa, que, por instigante tanto do ponto de vista linguístico quanto social, segue seu rumo, sem limites no interesse de procurar respostas para as questões que inquietam os que buscam não só caracterizar cada variedade linguística, mas também compreender a gênese dessas características.

BRANDÃO, S. F.; VIEIRA, S. R. Nominal and verbal agreement: contributions to the debate on the status of variation in three urban varieties of Portuguese. *Alfa*, São Paulo, v.56, n.3, p.1025-1054, 2012.

- *ABSTRACT: We focus on nominal agreement and on verbal agreement in the third person plural in urban varieties of European Portuguese, Brazilian Portuguese and Portuguese of Sao Tome, based on the speech of individuals with primary, secondary and university education level, distributed, yet, in three age groups and sex. In order to describe these varieties, structural and social motivations have been evaluated in the light of the assumptions of the Theory of Variation and Change (WEINREICH; LABOV; HERZOG, 1968). Results from different analyzes, developed under the Project Comparative Study of Patterns of Agreement in African, Brazilian and European varieties of Portuguese, suggest that, although the three varieties tend to adopt plural marks in both the Noun Phrase and the Verbal Phrase, agreement presents distinct status in each case. In European Portuguese, linguistic data cannot be described as effectively variable – it is, according to Labov (2003), a categorical rule within the nominal agreement and semi-categorical rule within the verbal one. In Brazilian Portuguese and in Portuguese of Sao Tome, both rules are variable.*
- *KEYWORDS: Noun agreement. Verb agreement. Sociolinguistics. Varieties of Portuguese.*

REFERÊNCIAS

BORTONI-RICARDO, S. M. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2004.

BRANDÃO, S. F. Réalité sociolinguistique brésilienne et géolinguistique pluridimensionnelle. In: LE SYMPOSIUM INTERNATIONAL SUR LES LIMITES ET LES AIRES EN DIALECTOLOGIE, 2011, Lisboa. *Anais...* Lisboa: Centro de Linguística da Universidade de Lisboa, 2011. No prelo.

CORBETT, G. G. *Number*. Cambridge: Cambridge University Press, 2000.

GONÇALVES, P. *(Dados para a) história da língua portuguesa em Moçambique*. Maputo, 2000. Disponível em: <<http://cvc.instituto-camoes.pt/hlp/geografia/portuguesmocambique.pdf>>. Acesso em: 05 maio 2009.

GRACIOSA, D. *Concordância verbal na fala culta carioca*. 1991. 125f. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1991.

HAGEMEIJER, T. As línguas de São Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, Macau, n.1, p.1-27, 2009.

ILHA DE SÃO TOMÉ. In: WIKIPÉDIA a enciclopédia livre. Disponível em: <<http://pt.wikipedia.org/wiki/Ficheiro:Tp-map.png>>. Acesso em: 20 out. 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA [IBGE]. *Censo populacional 2010*. Brasília, 29 nov. 2010. Acesso em: 11 dez. 2010.

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

_____. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.

LUCCHESI, D.; BAXTER, A.; RIBEIRO, I. (Org.). *O português afro-brasileiro*. Salvador: Ed. da UFBA, 2009.

METRODATA. Região metropolitana de Rio de Janeiro (RJ): taxa de crescimento anual por municípios. *Observatório das Metrôpoles*, Rio de Janeiro. Disponível em: http://www.observatoriodasmetrosoles.ufrj.br/metrodata/ibrm/ibrm_rj_tca.htm. Acesso em: 20 out. 2011.

MONGUILHOT, I. O. S. *Estudo sincrônico e diacrônico da concordância verbal de terceira pessoa do plural no PB e no PE*. 2009. 228f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Santa Catarina, Florianópolis, 2009.

NARO, A.; SCHERRE, M. M. P. *Origens do português brasileiro*. São Paulo: Parábola, 2007.

PLANO REGIONAL DE ORDENAMENTO DO TERRITÓRIO DA ÁREA METROPOLITANA DE LISBOA. *Área Metropolitana de Lisboa*. 2008. Disponível em: <<http://protaml.inescporto.pt/>>. Acesso em: 20 out. 2012.

VAREJÃO, F. O. A. *Varição em estruturas de concordância verbal e em estratégias de relativização no português europeu popular*. 2006. 187f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

VIEIRA, S. R. O estatuto da regra variável e o fenômeno da concordância verbal em variedades do português. In: CONGRESO INTERNACIONAL DE LA ALFAL, 16., 2011, Alcalá de Henares. *Actas...* Alcalá de Henares: ALFAL, 2011.

WEINREICH, U.; LABOV, W.; HERZOG, M. Empirical foundations for theory of linguistic change. In: LEHMANN, W.; MALKIEL, Y. (Org.). *Directions for historical linguistics*. Austin: University of Texas Press, 1968. p.97-195.

Recebido em março de 2012

Aprovado em maio de 2012