

Rita Gonçalves
Centro de Linguística
Universidade de Lisboa, Portugal

ritamgg@gmail.com

Mudança linguística e variação no português de São Tomé¹

São Tomé e Príncipe é palco de um processo de mudança linguística singular no quadro das ex-colónias portuguesas, tendo em conta a nativização do português, associada à perda da competência linguística das línguas crioulas. O fenómeno de transição do português L2 para L1 terá fomentado a transmissão irregular desta língua e, conseqüentemente, a emergência de estruturas gramaticais distintas do português padrão, entre as quais se evidencia a tendência para a transitivização direta. O presente trabalho analisa o efeito de fatores linguísticos, como a posição dos constituintes na frase, bem como a relevância de fatores sociolinguísticos, como o nível de escolaridade, o género e a idade, na variação e mudança linguística, em particular nas grelhas argumentais dos verbos.

Palavras-chave: português de São Tomé, mudança linguística, transmissão irregular, estrutura argumental, deslocação de constituintes, sociolinguística

Differently from other lusophone former colonies, in São Tomé and Príncipe a process of nativization of Portuguese is under way, which is related to the loss of proficiency in the locally spoken creole languages. This process of language shift resulted in irregular transmission of Portuguese, giving rise to the development of distinct properties from European Portuguese, such as a tendency toward direct transitivization. This paper assesses the role of linguistics effects, for example the position of constituents in the sentence, and sociolinguistics variables, such as the degree of instruction, age, and gender of the informants. These effects are essential to our understanding of language variation and change, especially in the domain of argument structure of verbs.

Keywords: Santomean Portuguese, language shift, irregular transmission, argument structure, constituent dislocation, sociolinguistics

¹ Agradeço a revisão e os comentários de Tjerk Hagemeijer. O texto apresentado faz parte de um trabalho de investigação em curso, razão pela qual algumas das questões não são mais desenvolvidas. A investigação é financiada pela Fundação para a Ciência e Tecnologia com a referência n.º SFRH/BD/73839/2010.

Introdução

Este trabalho tem como objetivos (i) caracterizar a situação linguística de São Tomé e Príncipe, discutindo o fenómeno de mudança linguística determinado pela hegemonia do português sobre os crioulos autóctones, de modo particular o forro; e (ii) analisar estruturas gramaticais que ilustram a variação existente no português falado em São Tomé, discutindo os possíveis fatores linguísticos e sociolinguísticos que a determinam.

Numa primeira parte, serão apresentados e discutidos dados estatísticos que ilustram o acelerado processo de nativização do português e a sua homogeneidade na sociedade são-tomense. Em seguida, tendo por base dados de *corpora* e os resultados de Brandão (2011), Alexandre, Gonçalves e Hagemeijer (2011) e R. Gonçalves (2010), debater-se-ão os fatores associados à variação nas áreas da concordância nominal, estratégias de relativização e estrutura argumental dos verbos. Numa terceira parte, dando continuidade ao trabalho encetado em R. Gonçalves (2010), serão confrontados os resultados da aplicação de testes de elicitación com dados de *corpora*, no sentido de avaliar o efeito de fatores linguísticos, como a posição dos constituintes na frase, e a relevância de fatores sociolinguísticos, como o nível de escolaridade, o género e a idade, na variação a nível da estrutura argumental dos verbos. Por fim, será apresentado um sumário dos resultados, apontando alguns dos princípios que estarão na base da mudança linguística e variação no português de São Tomé.

Mudança linguística

Apesar de constituir o mais pequeno dos PALOP em termos geográficos e populacionais, São Tomé e Príncipe caracteriza-se pela sua diversidade linguística motivada por fatores históricos e sociais. Nas ilhas são falados os crioulos autóctones – o forro, o lung'ie e o angular – decorrentes do contacto linguístico entre a língua dos colonizadores, o português, e as línguas dos escravos oriundos do antigo Reino de Benim (atual Nigéria), do Congo e de Angola, durante a primeira colonização (séc. XV e XVI)¹. Além destes crioulos, o arquipélago conserva ainda o português dos Tongas, a língua que surgiu do contacto entre o português falado nas roças (essencialmente, Água Ize, Monte Café e Agostinho Neto) e as línguas bantu faladas pelos descendentes de contratados provenientes de Angola e Moçambique durante a segunda colonização (séc. XIX e XX)². Fruto também desse segundo período, as ilhas acolhem ainda uma grande comunidade de falantes do caboverdiano, constituída por contratados não repatriados e pelos seus descendentes. O caboverdiano é usado um pouco por todo o arquipélago, chegando mesmo a constituir a língua dominante em

¹ Para um estudo mais aprofundado sobre a origem e a situação das línguas nacionais de São Tomé e Príncipe, veja-se Hagemeijer (2009) e Hagemeijer (no prelo).

² O português dos Tongas distingue-se do PST por apresentar traços do substrato Umbundu e por não ter recebido uma direta influência do forro, pelo menos nos primeiros estádios de aquisição. Como o crioulo era interdito nas roças, a língua de comunicação entre os escravos, e entre estes e os colonizadores, era o português que ali se desenvolvia sem possibilidade de escolarização. Segundo Baxter (2004, p. 104), “The L1 Tonga Portuguese (TP) of first generation Monte Café Tongas displays substantial and consistent morphosyntactic restructuring in the NP (in referencing, number, and gender representation), in the VP (tense-aspect marking), and in the representation of case relations and negation.” Sobre o português dos Tongas, veja-se, entre outros, Rougé (1992) e Baxter (2004).

algumas roças e no Príncipe, e a terceira língua em São Tomé, a seguir ao português e ao forro. Perante esta multiplicidade linguística, o português foi sendo reconhecido como fator unificador de grupos sociais e conquistou, sobretudo a partir do século XIX, uma hegemonia que veio a acentuar-se com a sua adopção como língua oficial após a independência.

Os dados sociolinguísticos reunidos em recenseamentos realizados durante o tempo colonial não nos permitem inferir o número de falantes das várias línguas neste período, uma vez que os crioulos nunca foram incluídos (cf. Lorenzino 1996, p. 442). Os dados relativos às línguas crioulas, número de falantes, distribuição etária e geográfica, só começam a ser recenseados a partir da independência de São Tomé e Príncipe, no primeiro Recenseamento Geral da População e Habitação (RGHP) de 1981 (publicado em 1987). Naturalmente que as políticas coloniais que depreciavam as línguas crioulas e consideravam o português como única língua de prestígio contribuiriam para a supremacia do português. Mesmo no período pós-colonial, esse estigma conduziu a população a usar e a transmitir às mais novas gerações o português adquirido como uma L2³, com todas as consequências que são inerentes à aquisição de uma língua não materna, continuando a considerá-lo a única língua de integração e ascensão social. O português foi conquistando o espaço onde primeiramente os crioulos tinham lugar e, conseqüentemente, estes foram sendo relegados para contextos cada vez mais informais e esporádicos.

O processo de nativização do português em São Tomé, segundo a aceção de Winford (2003, p. 244), de acordo com a qual uma anterior segunda língua é adoptada como a primeira língua (e língua nativa) de uma comunidade, é evidente se atentarmos nos dados estatísticos resultantes dos censos de 1981-2001. Não obstante todos os problemas associados à metodologia e à aplicação dos censos, desde logo por não ter sido indicado se as línguas são adquiridas como L1 ou L2, e por não terem sido consideradas outras línguas com relevância no arquipélago, como o angolano ou o caboverdiano⁴, é de destacar o aumento considerável do número de falantes do português, comparativamente às línguas crioulas e à(s) língua(s) não especificada(s).

	Português	Forro	Lung'ie	Outra
1981	60.519	54.387	1.533	8.180
1991	94.907	69.899	1.558	12.781
2001	136.085	99.621	3.302	17.612

Tabela 1 - Distribuição do número de falantes por língua e por ano de censo

³ O termo L2 ou LS (segunda língua) é usado para classificar toda a aprendizagem e uso de uma língua não materna, como ocorre, por exemplo, dentro de fronteiras territoriais em que ela é língua oficial. Por oposição, entende-se por L1, ou língua materna, a primeira língua à qual a criança está exposta e na qual estabelece a sua gramática.

⁴ Note-se que não só não foram contemplados o angolano e o caboverdiano, como foi incluído o francês que não é uma língua nacional. O facto de se privilegiar uma língua estrangeira em detrimento das línguas crioulas mostra o desprestígio que estas detêm na sociedade são-tomense. Consideramos, contudo, que, mesmo não tendo sido indicados, o angolano e o caboverdiano estão abrangidos pelo item de resposta “outra”.

A diferença no número de falantes do português e do forro aumenta significativamente de década para década. Enquanto em 1981, o intervalo era de apenas 5000 falantes, em 1991 e em 2001, este aumenta gradualmente para cerca de 25.000 e 35.000 falantes, respetivamente. Veja-se ainda que o número de falantes da(s) língua(s) não especificada(s) é superior ao de falantes do lung'ie, o que nos remete para a supremacia de outros crioulos, como o angolano, mas sobretudo o caboverdiano, sobre uma língua autóctone, o crioulo do Príncipe (cf. Hagemeijer, no prelo)⁵.

De acordo com o RGHP de 2001, o português é falado por 98,9% da população, contra 72,4% de falantes do forro. O gráfico 1 dá conta da relativa homogeneidade no uso do português, no que concerne ao género e à faixa etária.

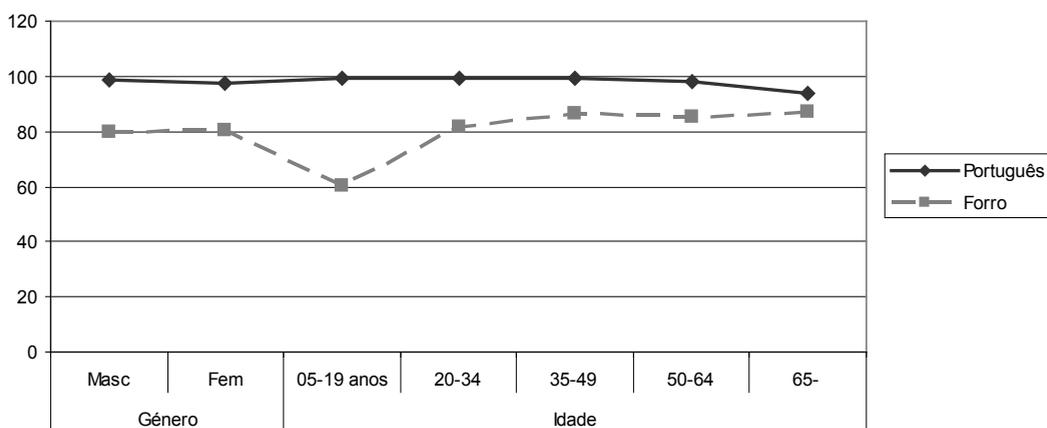


Gráfico 1 – Percentagem de falantes do Forro e do Português por género e idade (2001)

No primeiro caso, as percentagens apontam para um maior uso do português pelos informantes do género masculino, paralelamente a um maior uso do forro pelos informantes do género feminino. Esta situação pode ser corroborada pelo facto de as mulheres estarem mais circunscritas ao universo rural, ao passo que os homens estão mais integrados no contexto urbano, contactando com o mundo do comércio e com o exterior. Relativamente à faixa etária, verifica-se um ligeiro decréscimo no uso do português em informantes com idade igual ou superior a 65 anos, exatamente a que revela um aumento no número de falantes do forro, corroborando o que se verifica no terreno. A curva mais evidente dá conta da ausência de uso do forro pela população com idade igual ou inferior a 19 anos que tem, cada vez mais, apenas um conhecimento passivo das línguas crioulas e, neste caso particular, do forro.

Desta conjuntura, decorre uma situação em que o português primeiramente adquirido como L2 se começa a afirmar como a L1 da população, nomeadamente das gerações mais novas. Embora a língua-alvo seja, *a priori*, o português europeu, e o processo de mudança linguística não seja suficiente para o surgimento de uma língua

⁵ Esta diferença é ainda mais significativa se considerarmos que, de acordo com Maurer (2009, p. 3), no Príncipe há apenas cerca de 20 ou 30 pessoas, entre falantes com mais de 60 anos de idade e alguns jovens, que detêm um conhecimento relativamente ativo do lung'ie. Este facto corrobora o carácter pouco realista do censo apontado anteriormente.

totalmente nova, como ocorreu por exemplo com a crioulação, a verdade é que o resultado desta mudança permitirá a emergência de uma variedade linguística com propriedades distintas das do português europeu. No dia-a-dia é produzido um português local, com possíveis influências do crioulo, decorrente da transição histórica de segunda a primeira língua e das situações de contacto linguístico acima descritas. Esta variedade do português, resultado de uma transmissão linguística irregular⁶, desencadeada sobretudo através da oralidade, uma vez que o acesso à escola era bastante limitado, constitui o *input* a que as gerações mais jovens estão expostas, transformando-se, portanto, na língua-alvo no processo de aquisição da sua L1.

De acordo com Batibo (1992, *apud* Winford 2003, p. 258), o processo de mudança linguística envolve cinco estádios que vão desde o uso da L1 como único sistema linguístico de que o falante dispõe, passando por períodos intermédios de transição de monolingüismo a bilingüismo, até à completa “extinção” da L1 e consequente substituição pela L2⁷. A delimitação dos cinco estádios não é claramente visível, uma vez que a erosão linguística opera em tempos distintos nos vários falantes. No caso particular das ilhas de São Tomé e Príncipe, este processo situar-se-ia nos últimos estádios, os responsáveis pela mudança linguística, embora em algumas partes da sociedade, o forro ainda concorra com o português, numa situação de diglossia. Na verdade, mesmo que toda uma comunidade convirja para a adopção do português como L1, o processo não se desenvolve ao mesmo tempo e do mesmo modo em todos os falantes, mas depende de uma variedade de fatores sociolinguísticos que estão associados à idade, ao género e ao nível de escolaridade, como já apontámos e veremos com maior detalhe nas secções seguintes.

Os dados

O presente trabalho teve por base dois conjuntos de dados distintos: um *corpus* oral e testes de elicitación escritos. O *corpus* do português de São Tomé (PST) usado é constituído por trinta entrevistas à população da cidade São Tomé, recolhidas no âmbito do projeto VAPOR (Variedades Africanas do Português) do Centro de Linguística da Universidade de Lisboa. Apesar de, atualmente, o *corpus* ser constituído por cerca de 250.000 palavras transcritas, para este estudo foi considerado um *corpus*-amostragem de trinta entrevistas (correspondente a cerca de 120.000 palavras transcritas). O *corpus* é uniforme quanto ao género, dispondo de quinze informantes do género masculino e quinze do género feminino. As entrevistas foram realizadas a informantes com diferentes níveis de escolarização, desde o baixo (4.^a classe) até ao superior (licenciatura), e com idades compreendidas entre os 18 e os 45 anos. Os estudos sobre o PST aqui mencionados, nomeadamente Alexandre, Gonçalves e Hagemeyer (2011) e Brandão (2011), basearam-se no mesmo *corpus*, no primeiro caso, e num *subcorpus* deste, no segundo (correspondente a 17 entrevistas).

Os testes de elicitación foram aplicados no Instituto Superior Politécnico de São

⁶ Sobre o conceito de transmissão linguística irregular e a sua aplicação à formação de novas variedades de língua, veja-se Lucchesi & Baxter (2009).

⁷ Este fenómeno é conhecido e tratado na literatura sobre L2 por erosão linguística.

Tomé⁸ a alunos de licenciatura. O teste de juízos de gramaticalidade foi apresentado a 65 alunos (23 alunos do 4.º ano do curso de Língua Portuguesa (LP); 28 do 1.º ano do curso de Relações Públicas (RP); e 14 do 1.º ano do curso de Biologia (BL)), com idades compreendidas entre os 18 e os 45 anos de idade, sendo que 32 são do género feminino e 33 do género masculino. Dada a não comparência dos restantes elementos, o teste de produção foi apresentado nas mesmas turmas apenas a 58 alunos (22 do 4.º ano de LP; 23 de RP; e 13 de BL), 29 do género masculino e 29 do género feminino.

Variação linguística

Concordância nominal

Os estudos sobre concordância nominal em variedades não europeias do português, designadamente no português brasileiro (PB), são abundantes e têm mostrado que existe variação significativa a este nível. Brandão (2011) procurou estabelecer uma comparação entre os fatores associados à ausência de concordância nominal no PB e no PST, com o intuito de verificar se a mudança opera de forma convergente ou divergente.

Tendo por base o controlo de variáveis linguísticas e sociolinguísticas, a autora concluiu que, tanto no PB como no PST, essas variáveis se alternam, ainda que a posição dos constituintes dentro do SN se mostre relativamente mais relevante para a primeira e que o nível de escolaridade se destaque na segunda. Em suma, o núcleo do SN, bem como os elementos à sua direita, são mais afetados pelo apagamento da marca de número e a probabilidade de ausência de concordância nominal diminui conforme aumenta o nível de escolaridade do informante (cf. (1-2)).

(1) As dificuldade que nós temos (...). (PST, Brandão, 2011)

(2) As pessoas mais velha quando faziam calulu (...). (PST, Brandão, 2011)

Também Inverno (2009, p. 4), sobre a concordância nominal no português de Angola (PA)⁹, concluiu que “o núcleo do SN raramente recebe marcação de número. A pluralidade é indicada pela adição do sufixo –s apenas aos elementos não-nucleares mais à esquerda no SN, especialmente no discurso de falantes mais velhos ou menos instruídos ou no discurso informal daqueles que são mais jovens ou instruídos” (cf. (3)).

(3) Ele marca muitos golo. (PA, Inverno 2009)

Embora estes fatores remetam superficialmente para uma mudança convergente nas variedades do português, é de considerar que outros há que não são atuantes no PB e no PST, tais como o processo morfofonológico de formação de plural na-que-la, bem como o género e o contexto fonológico nesta. Brandão (2011) aponta

⁸ Agradece-se a participação dos alunos incluídos no estudo e das professoras Beatriz Afonso e Joana Castaño que colaboraram na aplicação dos testes.

⁹ Dados recolhidos no Dundo, província da Lunda Norte.

ainda importantes conclusões resultantes do cruzamento entre as variáveis nível de escolaridade e idade. No PST, os informantes entre os 36 e os 55 anos com escolarização apenas fundamental apresentam maiores índices de apagamento da marca de número. Estes dados vão ao encontro do defendido por Moreno e Tuzine (1997, p. 84) para o português de Moçambique (PM). Os autores mostram que, não obstante a pertinência do fator nível de escolaridade, os fenómenos de variação “não podem nem devem ser explicados recorrendo-se apenas a uma ou duas variáveis sociais isoladamente, parecendo, pelo contrário, decorrer de uma multiplicidade de factores, actuando em conjunto.”

Parece ser unânime, contudo, como vários estudos de teor sociolinguístico têm mostrado desde a década de sessenta, com os estudos labovianos, que o fator nível de escolaridade tem um papel relevante na variação linguística. No caso do PST, a par do nível de escolaridade, o género e a idade mostram-se também responsáveis pela variação a nível da concordância nominal. Vejamos, em seguida, de que modo estes fatores também atuam noutras áreas desta gramática em variação.

Estratégias de relativização

Apesar de existirem vários estudos sobre as estratégias de relativização nas variedades do português, estes centram-se essencialmente às variedades europeia e brasileira (e.g. Alexandre, 2000; Tarallo, 1985). Sobre as variedades emergentes do português em África, esses trabalhos limitam-se ao PM (e.g. Chimbutane, 1996) e ao estudo recente de Alexandre, Gonçalves e Hagemeyer (2011), sobre as estratégias disponíveis no PST e no português de Cabo Verde (PCV) e a sua relação com os crioulos com os quais estão em contato, o forro e o caboverdiano¹⁰.

De acordo com os autores, o PST dispõe de quatro estratégias de relativização: *pied-piping* (ou arrastamento da preposição) (cf. (4)), *chopping* (ou cortadora) (cf. (5)), resunção (cf. (6)) e cópia defetiva (cf. (7)), sendo que apenas a primeira é considerada canónica à luz da norma do PE. As três primeiras remetem, aparentemente, para uma mudança convergente, uma vez que estão disponíveis em todas as outras variedades do português, enquanto a última corrobora a hipótese de *transfer*, na medida em que constitui a única estratégia de relativização disponível no forro (cf. (8)).

(4) Uma das primeiras pessoas com quem eu trabalhei (...).

(5) É uma língua que nós devíamos dedicar.

(6) A mulher que eu vivo com ela também é católica.

(7) Os manuais que nós trabalhamos com ele (...).

(8) San tê inen mina nala san ku ome ka pô fla ku ê.

senhora ter PL filho lá POS REL homem IPFV poder falar com 3SG

‘Ela tem lá os seus filhos que os homens podem falar com ele.’

No quadro da variação a nível das estratégias de relativização, destaca-se o facto de o PST apresentar *pied-piping* argumental pouco expressivo (27%) e de a estratégia dominante ser a cortadora (66%). As estratégias de resunção e de cópia defetiva

¹⁰ Embora os autores já tenham alargado este estudo comparativo às restantes variedades africanas do português, designadamente o PA e PM, o trabalho não se encontra ainda publicado.

totalizam apenas 7% das ocorrências. Alexandre, Gonçalves e Hagemeyer (2011) defendem, contudo, que a primazia da estratégia cortadora no PST, contrariamente às restantes variedades do português, pode estar dependente da tendência para a perda substancial de preposições, em complementos verbais (cf. ((9 a. – 10 a.)) e em adjuntos temporais (cf. ((11))) (cf. Gonçalves, R. 2010).

- (9) a. Tem que apanhar carro para chegar Porto Alegre.
 b. Isto é que é a conclusão que se chegou.
- (10) a. Entrou pensão onde nós estávamos.
 b. Há casa que entra água.
- (11) a. Nasci sessenta e sete.
 b. Durante o tempo que estive de férias.

Uma vez que nesta variedade há contextos em que os verbos selecionam complementos que no PE seriam regidos por preposições, considera-se que há enunciados que podem não constituir verdadeiras cortadoras e tratar-se daquilo que os autores designam por pseudo-cortadoras (confrontem-se os exemplos (9-11 a. e b.)).

Apesar desse trabalho não ter como objetivo principal abordar a questão da variação sociolinguística, os autores admitem verificar-se “uma relação entre classe social/escolaridade e o uso das estratégias de *pied-piping*, de resumpção e de cópia defetiva, uma vez que estas estratégias são mais comuns em pessoas menos escolarizadas. A estratégia cortadora, por seu lado, ocorre de forma mais generalizada, tal como nas outras variedades” (Alexandre, Gonçalves e Hagemeyer, 2011, p. 27). Peres e Mória (1995) e Alexandre (2000) mostram como o uso desta estratégia é generalizada a falantes do português europeu mais escolarizados, inclusivamente na escrita.

Assim sendo, confirma-se a relação previamente estabelecida, a propósito da concordância nominal, entre a concorrência de fatores linguísticos e sociolinguísticos para a variação no PST. Vejamos agora em que medida esta é ou não extensível à variação na estrutura argumental dos verbos.

Estrutura argumental dos verbos

R. Gonçalves (2010) dá conta da variação existente ao nível da seleção categorial dos verbos, argumentando a favor de uma mudança no sentido da transitivação direta no PST. A tabela 2 dá conta da variação existente a esse nível, mostrando que a omissão de preposições é significativamente superior à sua substituição, o que permite afastar o PST de outras variedades não europeias do português.

		N.º Abs.	%
Omissão	OI	46	63%
	OBL	207	
Substituição	OI	23	37%
	OBL	125	

Tabela 2 – Distribuição da omissão/substituição de preposições no *corpus*

De acordo com o estudo de R. Gonçalves (2010), a tendência para a transitivização direta está dependente de fatores linguísticos como a natureza da preposição, uma vez que quanto menor é o grau de lexicalidade das preposições, maior é a probabilidade de estas serem afetadas. Na verdade, quando as preposições desempenham meramente funções sintáticas, como é o caso de *a* e *de*, é difícil captar o seu significado, contrariamente a outras como *até*, *desde* ou *sem*, semanticamente mais fortes, pelo que tipicamente as primeiras são as mais afetadas na aquisição da L2. No *corpus* do PST, as preposições *a* e *de* são tendencialmente apagadas ou substituídas, em contextos de objeto indireto e oblíquo, como mostram os exemplos seguintes.

- (12) Entrega senhor uma cerveja.
- (13) Cada um fala aquilo que ele imagina.
- (14) Com mais idade isso cria problema para homem.
- (15) Chego em casa, tenho coisas para fazer.

Nas variedades brasileira e angolana, a preposição *a*, é comumente substituída por *para* e *em*, respetivamente, em contextos de objeto indireto, e por *em*, em oblíquos selecionados por verbos de movimento, como *ir* ou *chegar* (cf. Berlinck e Torres Morais, 2007; e Cabral, 2005). O mesmo se verifica na variedade moçambicana, à exceção dos casos do objeto indireto, em que tendem a ocorrer construções de duplo objeto (e.g. P. Gonçalves, 2010).

Além do estatuto das preposições, R. Gonçalves (2010) estabelece uma associação entre a mudança nas grelhas argumentais dos verbos no PST e o efeito de *transfer* do forro, dado que nesta língua os verbos selecionam tipicamente complementos nominais.

- (16) Ê piji san lenha kopu d'awa
3SG pedir senhora rainha copo de água
'Ele pediu um copo de água à rainha.'
- (17) Ol'ô ka xiga kinte...
quando-2SG ASP chegar quintal
'Quando chegas ao quintal..'

Porém, se tivermos em conta que a preposição *a* é geralmente omitida, mesmo em complementos (preposicionados no PE) seleccionados por verbos que não existem no crioulo, como *assistir* ou *dedicar-se*, a hipótese de *transfer* é insuficiente para descrever todas as mudanças linguísticas. É nesse sentido que R. Gonçalves (2010) argumenta a favor de que, para compreender a variação existente a nível das grelhas argumentais, e em particular dos verbos de movimento no PST, é necessário considerar também fatores sintático-semânticos como, por exemplo, as propriedades

aspetuais dos verbos envolvidos¹¹. No entanto, consideramos que, embora as propriedades daquela que foi historicamente a L1 de grande parte dos falantes não sejam diretamente responsáveis pela estruturação da gramática da L2, terão, em grande parte, contribuído para esta.

Não obstante a análise dos vários fatores associados à mudança a nível das propriedades de subcategorização dos verbos no PST, apresentada por R. Gonçalves (2010), parece relevante verificar se há outros efeitos responsáveis pela tendência para o corte de preposições em complementos verbais. A secção seguinte ocupará-se da análise e discussão de efeitos linguísticos e sociolinguísticos na mudança na estrutura argumental dos verbos, recorrendo aos dados do *corpus* do PST e a resultados de testes de elicitación.

Efeitos linguísticos e sociolinguísticos na estrutura argumental dos verbos do PST

Corpus

Os dados PST foram organizados de acordo com seis variáveis sociolinguísticas que nos permitem verificar de que modo fatores como o género (masculino/feminino), a idade (18-34 anos/ 35-49 anos) e o nível de escolaridade (4.^a – 9.^a classe/ 10.^a ano – licenciatura) exercem influência na estrutura argumental dos verbos em variação, em particular na transitivização direta. A tabela 3 dá conta da percentagem geral de apagamento da preposição distribuída pelas variáveis em estudo.

Variáveis		%
Género	Masculino	40%
	Feminino	59%
Idade	18-34	61%
	35-49	38%
Nível de escolaridade	4. ^a – 9. ^a	76%
	10. ^a – licenciatura	37%

Tabela 3 – Percentagem geral de apagamento da preposição

Os resultados mostram que o fator nível de escolaridade é o mais significativo, sendo que o grupo de informantes com escolarização entre o 4.^o e o 9.^o ano é responsável por 76% das ocorrências de sintagmas nominais selecionados por verbos que no PE subcategorizam sintagmas preposicionados. Os fatores idade (18-34) e género feminino parecem ter uma responsabilidade secundária comparativamente ao primeiro, ainda que sejam responsáveis por mais de 50% das ocorrências. A relevância dos fatores género e idade já tinha sido apontada por Brandão (2011), a

¹¹ A diferença na seleção categorial dos verbos de movimento *chegar* e *ir* no PST (variação na seleção de um SN ou de um SP introduzido por *em*, no primeiro caso, e seleção de um SP introduzido por *a*, no segundo) pode estar associada ao facto de o verbo *chegar* implicar o atingir de uma finalidade (verbo télico), contrariamente ao verbo *ir* que, apesar de envolver deslocamento, não impõe que o percurso tenha sido cumprido (verbo atélico). Assim sendo, o recurso à preposição *em*, em complementos do verbo *chegar*, corrobora a noção de finitude subjacente à descrição da situação, contrariamente à preposição *a*, em complementos do verbo *ir*, que exprime dinamicidade e conserva a ideia de movimento.

propósito da concordância nominal, e é considerada igualmente noutros estudos com o mesmo cariz sociolinguístico. Baxter (2009, p. 285) constata que as mulheres lideraram a ausência de concordância em número na comunidade tonga de Monte Café, em São Tomé; e Lucchesi (2009, p. 313) apresenta conclusões semelhantes sobre a comunidade afro-brasileira de Helvécia, no sul da Bahia, sobre a concordância em género. Além disso, estes resultados vão ao encontro dos dados do RGHP (2003) (cf. gráfico 1), segundo os quais, ainda que de modo pouco expressivo, os homens fazem um maior uso do português e as mulheres do forro. Como já apontado, esta diferença pode estar associada ao facto de nas sociedades africanas a mulher estar mais limitada ao universo rural e doméstico, por contraste com os homens que têm um maior contato com o mundo exterior e dos negócios. Estes resultados contrastam com os obtidos em grandes centros urbanos, nos quais são as mulheres a liderar a mudança convergente com a norma (e.g. Romaine, 2000).

Dado que estruturas de deslocação de constituintes preposicionados são pouco frequentes no *corpus* de dados orais¹², para aferir o papel deste efeito linguístico nas grelhas argumentais dos verbos, foi necessário proceder à produção e aplicação de testes de elicitación. A secção seguinte dá conta desse processo e dos resultados obtidos.

Testes de elicitación

A aplicação de testes de elicitación vem no sentido de verificar se os informantes estabelecem diferenças na seleção categorial dos verbos (preferência pela seleção de um SN ou de um SP), tendo em conta a posição que o constituinte subcategorizado pelo verbo ocupa na frase. Foram aplicados dois testes de elicitación, um de juízos de gramaticalidade e um de produção. Com o primeiro, procurou-se testar o modo como os informantes classificam constituintes nominais e preposicionados realizados à direta do verbo, em declarativas (cf. (18)), por contraste com constituintes movidos em relativas (cf. (19)), interrogativas (cf. (20)), topicalizações (cf. (21)) e clivadas (cf. (22)):

- (18) Ninguém contribuiu (para) o almoço.
- (19) Esta é a viagem (em) que penso há anos.
- (20) (A) Que concerto é que eles assistiram?
- (21) (De) Esta cidade, gosto muito.
- (22) Foi (com) a decisão do patrão que eles concordaram.

Este tipo de contraste, em testes de juízos de gramaticalidade, permite observar em que contextos, com ou sem deslocação de constituintes, os informantes reconhecem a obrigatoriedade da presença da preposição para a aceitabilidade das frases. De acordo com Klein (1993, 2001), só se pode considerar que há efetivamente omissão de preposição¹³ se em relativas e/ou interrogativas a ausência de preposição

¹² Veja-se que num *corpus* de 120.000 palavras foram registadas apenas 24 ocorrências de topicalizações e 1 de interrogativa com 'é que'. Estes números são relativos a complementos oblíquos, sendo que, à exceção de relativas, não foram atestados contextos de deslocação de objetos indiretos.

¹³ Fenómeno de preposição nula, de acordo com a aceção da autora.

(resultante da estratégia cortadora) for aceite e nas declarativas correspondentes for considerada agramatical. No caso de a ausência de preposição ser aceite em ambos os contextos, o fenómeno deve ser perspectivado como uma alteração das propriedades de subcategorização dos verbos, o que no caso do PST corroboraria a hipótese da tendência para a transitivização direta (cf. Gonçalves, R. 2010).

No teste de produção, por sua vez, os informantes foram incitados a produzir as mesmas estruturas de deslocação de constituintes, no sentido de verificar se, nesses contextos, a preposição acompanha ou não o constituinte movido (cf. (23-26)). O confronto dos resultados deste teste com os de juízos de gramaticalidade permite constatar se há contextos em que os informantes, apesar de reconhecerem a obrigatoriedade da presença da preposição para a gramaticalidade das frases, não a produzem. Ao mesmo tempo, contribui também para observar quais são as preposições mais afetadas e os verbos cuja grelha argumental se encontra em variação e mudança¹⁴.

(23) O comerciante concordou com os clientes. Estes são os clientes _____.

(24) As pessoas assistiram à procissão. _____ é que as pessoas assistiram?

(25) As mulheres ontem foram à discoteca. _____, as mulheres foram ontem.

(26) Eles puseram os livros no saco. Foi _____ que eles puseram os livros.

Os testes de eliciação apresentados a alunos do Instituto Superior Politécnico de São Tomé foram construídos com base numa lista de verbos frequentes no *corpus* do PST¹⁵, a saber: *dar, dizer, ir, assistir, concordar, falar, gostar, precisar, pensar, pôr, voltar* e *contribuir* que, no PE, selecionam constituintes preposicionados, introduzidos pelas preposições *a, com, de, em* e *para*¹⁶.

O gráfico 2 dá conta dos resultados obtidos em ambos os testes, distribuídos por:

(i) variáveis linguísticas: COMP(lemento do verbo); REL(ativas); INT(errogativas), TOP(icalizações) e CLIV(adas);

(ii) e sociolinguísticas: nível de escolaridade – alunos do 1.º ano de licenciatura (Biologia – BL e Relações Públicas – RP) e alunos do 4.º ano de licenciatura (Língua Portuguesa – LP) e género (MASC(ulino) e FEM(imino)).

¹⁴ Fica fora do escopo deste trabalho a análise dos dois últimos aspetos, bem como a reflexão sobre uma possível diferença de desempenho na deslocação de objetos indiretos ou complementos oblíquos, apesar de ambos os argumentos terem sido considerados na produção dos testes. De acordo com Kato (2008, 2009), no PB, as preposições que atribuem caso inerente são opcionais, ao passo que as verdadeiras preposições não podem ser omitidas. A comprovar-se a hipótese de Kato, esperar-se-ia que a omissão de preposições fosse mais frequente (e aceite) em objetos indiretos, SN's introduzidos pelo marcador de caso dativo *a*, do que em oblíquos, SP's introduzidos por verdadeiras preposições.

¹⁵ A escolha dos verbos foi igualmente baseada numa lista de frequências do CRPC (*Corpus de Referência do Português Contemporâneo*) do Centro de Linguística da Universidade de Lisboa.

¹⁶ Note-se que no *corpus* analisado por R. Gonçalves (2010, p. 40), os verbos *assistir* (94%) e *precisar* (57%) apresentam uma percentagem de omissão de preposição superior à da sua realização. Nos argumentos dos verbos *dar* (29%) e *gostar* (13%), ainda que não ultrapasse os 50%, esta é igualmente significativa.

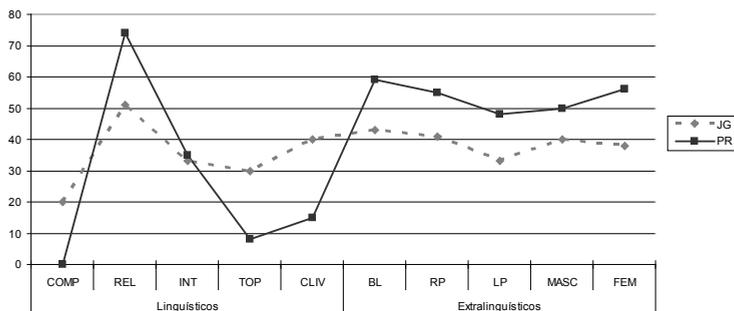


Gráfico 2 – Percentagem de apagamento da preposição por variáveis linguísticas e sociolinguísticas

Como o gráfico mostra, verifica-se em ambos os testes uma percentagem de corte da preposição mais elevada determinada por fatores linguísticos (relativas) e sociolinguísticos (nível de escolaridade mais baixo – alunos do 1.º ano da licenciatura em Biologia). Numa perspetiva comparativa entre os dois testes, verifica-se que a percentagem de corte é superior no teste de produção, o que mostra que a necessidade da ocorrência de preposição faz parte do conhecimento implícito do falante, no entanto não se manifesta na sua produção escrita. Repare-se que em dados de produção oral espontânea, o corte de preposição em relativas (*chopping*) é bastante mais significativo (66%) do que o de arrastamento da preposição (*pied-piping*) (27%) (cf. Alexandre, Gonçalves e Hagemeyer (2011))¹⁷.

Ao compararmos os dados de complementos realizados à direita do verbo com os de estruturas de deslocação, no teste de juízos de gramaticalidade, verifica-se que os constituintes não deslocados são claramente menos afetados pelo apagamento de preposição. Por sua vez, ao atentarmos nas percentagens dos constituintes deslocados, regista-se um menor índice de corte em topicalizações. Uma vez mais, se confrontarmos os resultados do teste de produção com os do *corpus* de dados orais, observa-se uma diferença significativa, na medida em que neste a preposição tipicamente não é realizada em contextos de topicalização (em 24 ocorrências de topicalizações, ocorre *pied-piping*¹⁸ da preposição em apenas 5).

Conclui-se, assim, que, além dos fatores linguísticos associados à mudança nas grelhas argumentais dos verbos descritos no ponto 3.3. deste trabalho, a deslocação de constituintes também exerce um papel relevante. Esta favorece o apagamento da preposição, destacando-se as relativas e as topicalizações por nos testes ocuparem, respetivamente, as fronteiras de maior e menor índice de apagamento da preposição¹⁹.

Relativamente aos dados sociolinguísticos, os resultados mostram uma clara progressão na aprendizagem do 1.º para o 4.º ano de licenciatura, em ambos os testes,

¹⁷ Registe-se que em estudo posterior (ms), com base num *corpus* alargado, Alexandre, Gonçalves e Hagemeyer (2011) verificaram que a diferença entre *chopping* e *pied-piping* no PST é ainda mais significativa (68% vs. 23%).

¹⁸ Por o processo que está na base da derivação de topicalizações e clivadas ser o mesmo das relativas e interrogativas – movimento-A', para dar conta do arrastamento ou corte de preposição nestas estruturas, serão igualmente usados os termos *pied-piping* e *chopping*.

¹⁹ O apagamento da preposição nestas estruturas já tinha sido apontado em Gonçalves (2010, pp. 35-39), embora o número reduzido de ocorrências no *corpus* não tivesse permitido formular esta generalização.

o que vai ao encontro dos dados fornecidos pelo *corpus* apresentados na secção anterior. Verifica-se, igualmente, uma diferença entre os alunos de 1.º ano, sendo que os de Biologia apresentam resultados mais divergentes com a norma do PE do que os de Relações Públicas. No entanto, a possibilidade de esta diferença estar associada à exposição a diferentes programas de português no ensino secundário é descartada, na medida em que em São Tomé essa distinção não tem lugar²⁰. O gráfico 2 permite verificar ainda que se regista uma maior percentagem de corte de preposições em informantes do género feminino no teste de produção, corroborando os dados do *corpus* oral, e em informantes do género masculino no teste de juízos de gramaticalidade.

No sentido de melhor aferir a variável nível de escolaridade, e de responder à diferença observada nas duas turmas do 1.º ano de licenciatura, procedeu-se ao seu cruzamento com a variável idade. O gráfico seguinte apresenta a percentagem de apagamento da preposição em ambos os testes distribuída pelas essas variáveis sociolinguísticas.

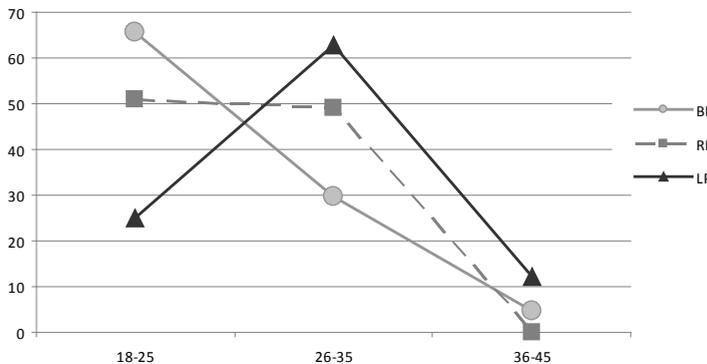


Gráfico 3 – Percentagem de apagamento da preposição, por idade e nível de escolaridade

O gráfico 3 regista uma diminuição significativa de corte de preposições nos três grupos de informantes a partir de 36 anos de idade, confirmando os dados do *corpus*. Na faixa etária entre os 26 e os 35 anos, é o grupo de informantes do nível de escolaridade mais alto (4.º ano – LP) que se destaca por apresentar um maior índice de apagamento de preposição (63%), concorrendo com os informantes do nível de escolaridade mais baixo (1.º ano – BL), na faixa etária entre os 18-25 anos (65%). O grupo que se mantém mais estável é o de 1.º ano (RP), com uma percentagem de apagamento na ordem dos 50%, nas duas primeiras faixas etárias. Contudo, se por um lado fica claro que a percentagem de desvio diminui conforme aumenta a faixa etária (média de 6% de desvio em informantes entre os 36 e os 45 anos de idade), por outro, não é possível responder à diferença de desempenho sentida nas turmas

²⁰ Agradeço este comentário à Beatriz Afonso e à Joana Castaño que contactam no terreno, quer com professores, quer com alunos do ensino secundário, observando a prática de ensino e analisando o programa de Língua Portuguesa adoptado. Este consiste numa adaptação dos programas de Português homologados pelo Ministério da Educação de Portugal em 2001 para o ensino secundário nos cursos científico-humanísticos e nos cursos tecnológicos.

de 1.º ano (cf. gráfico 2) com base no número de informantes por faixa etária. Em ambas, o número de alunos entre os 18 e os 25 anos de idade é superior ao da faixa seguinte (26-35 anos). Além disso, os dois grupos apresentam uma média de 47% de desvio.

Em suma, não obstante a existência de variação, estes dados corroboram os de Brandão (2011) sobre a pertinência de cruzar as variáveis nível de escolaridade e faixa etária e, a par dos resultados do *corpus* do PST, apontam para uma maior percentagem de convergência com a norma em informantes com idade superior a 36 anos de idade, com escolarização avançada. Ao confrontarmos este resultado com os dados do RGHP (2003) (cf. gráfico 1), verificamos que é exatamente a partir desta faixa etária que se começa a verificar um maior uso do forro paralelamente ao decréscimo do uso do português. Aparentemente, este facto fornece evidência negativa para o papel do *transfer* do crioulo e aponta para os efeitos e uma diglossia mais estável no passado.

Comprova-se, portanto, que os fatores sociolinguísticos testados têm um papel de relevo na variação e mudança a nível das grelhas argumentais dos verbos no PST. Além do nível de escolaridade que tem, desde há muito, um papel de destaque nos estudos sociolinguísticos, e que se comprova ser responsável pela variação entre a seleção de um SN ou de um SP no PST, constata-se a relevância do género e da idade na tendência para o apagamento das preposições e consequente transitivização direta no português falado em São Tomé. Do mesmo modo, constatou-se que o efeito linguístico de deslocação dos constituintes favorece largamente o apagamento da preposição. No entanto, fica por controlar se há regularidade nesse apagamento tanto por verbos como por tipo de argumentos subcategorizados e preposições associadas.

Considerações finais

O presente trabalho procurou mostrar que a variação existente no português falado em São Tomé é dependente de um processo histórico de mudança linguística caracterizado pela transição do português L2 para L1, motivada pela primazia do uso deste em detrimento das línguas crioulas. Verificámos que, tal como na concordância nominal em número e nas estratégias de relativização, os fatores associados à variação a nível das grelhas argumentais dos verbos no PST são de ordem sociolinguística e linguística.

Por um lado, os dados do *corpus* oral, bem como os resultados dos testes de elicitación, atestam a relevância do fator nível de escolaridade na tendência para a transitivização direta. A percentagem de seleção de um SN em detrimento de um SP aumenta em função de uma menor exposição à escolarização. Além disso, o género e a idade também se apresentam como fatores relevantes, na medida em que produções de informantes do género masculino ou na faixa etária entre os 36 e os 45 anos se destacam por apresentarem uma mais alta percentagem de convergência com a norma do português europeu. Por outro lado, o confronto entre a ocorrên-

cia de um SN/SP em posição de complemento do verbo com contextos em que o constituinte é movido, designadamente em relativas, interrogativas, topicalizações e clivadas, permite concluir que a deslocação de constituintes favorece o apagamento da preposição, de modo particular a relativização. Assim se conclui que a análise da gramática do PST requer a observação de um conjunto de fatores linguísticos e sociolinguísticos que concorrem no processo de variação e mudança linguística.

Referências

Alexandre, Nélia (2000). *A estratégia resumptiva em relativas restritivas do português Europeu*. Dissertação de mestrado, Universidade de Lisboa, Portugal.

Alexandre, Nélia, Gonçalves, Rita & Hagemeijer, Tjerk (2011). A formação de frases relativas de PP no português de Cabo Verde e de São Tomé. In *Actas do XXVI Encontro da Associação Portuguesa de Linguística* (pp. 17-34). Lisboa: APL.

Baxter, Alan (2009). A concordância de número. In Lucchesi, D., Baxter, Alan & Ribeiro, Ilza (eds.), *O português afro-brasileiro* (pp. 269-294). Salvador: EDUFBA.

Baxter, Alan N. (2004). The development of variable NP plural agreement in a restructured African variety of Portuguese. In G. Escure & A. Schwegler (eds.), *Creoles, contact and language change: linguistics and social implications* (pp. 97-126). Amsterdam: John Benjamins.

Brandão, Sílvia. (2011). Concordância nominal em duas variedades do português: convergências e divergências. *Veredas*, 15 (1), pp. 164-178.

Chimbutane, Feliciano (1996). A estratégia de pronome resumptivo na formação de orações relativas de OD e de OBL do português de Moçambique. In *Actas do XI Encontro Nacional da APL* (pp. 225-248), vol. III. Lisboa: Colibri.

Gonçalves, Perpétua (2010). *A génese do Português de Moçambique*. Lisboa: INCM.

Gonçalves, Rita (2010). *Propriedades de subcategorização verbal no português de S. Tomé*. Dissertação de Mestrado, Universidade de Lisboa, Portugal.

Inverno, Liliana (2009). A transição de Angola para o português vernáculo: estudo morfossintático do sintagma nominal. In Ana M. Carvalho (ed.), *Português em contato* (pp. 87-106). Madrid, Frankfurt: Iberoamericana, Editorial Vervuert

Lorenzino, Geraldo A. (1996). Uma avaliação sociolinguística sobre São Tomé e Príncipe. In Duarte, Inês & Leiria, Isabel (orgs), *Actas do Congresso Internacional sobre o Português* (pp. 435-449). Vol. II. Lisboa: APL e Edições Colibri.

Lucchesi, Dante. 2009. A concordância de género. In Lucchesi, Dante. Baxter, Alan. e Ribeiro, Ilza (eds.), *O português afro-brasileiro* (pp. 295-318). Salvador: EDUFBA.

Lucchesi, Dante & Baxter, Alan (2009). A transmissão linguística irregular. In Lucchesi, Dante. Baxter, Alan & Ribeiro, Ilza (eds.). *O português afro-brasileiro* (pp. 101-124). Salvador: EDUFBA.

- Hagemeyer, Tjerk (2009). As línguas de S. Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola*, 1 (1), pp. 1-27.
- Hagemeyer, Tjerk. (submetido). S. Tomé e Príncipe: labirinto e laboratório de línguas. In Gerhard Seibert (org.), *Arquipélagos crioulos. Cabo Verde e São Tomé e Príncipe numa perspectiva comparada*. Lisboa: Vega.
- Kato, Mary (2008). Optional prepositions in Brazilian Portuguese. In Arregi, K., Fagyal, Z., Montrul, Silvina A. & Tremblay, Annie (eds.), *Romance Linguistics 2008: interactions in romance*, (pp. 171-184). John Benjamins Publishing Company.
- Kato, Mary & Nunes, Jairo (2009). A uniform raising analysis for standard and nonstandard relative clauses in Brazilian Portuguese. In Jairo Nunes (ed.), *Minimalist Essays on Brazilian Portuguese Syntax* (pp. 93-120). Cambridge: University press.
- Klein, Elaine C. (1993). *Toward second language acquisition: a study of null-prep*. Dordrecht: Kluwer Academic Publishers.
- Klein, Elaine C. (2001). (Mis)construing null prepositions in L2 intergrammars: a commentary and proposal. *Second Language Research*, 17 (1), pp. 37-70.
- Maurer, Philippe (2009). *Principense (Lung'le). Grammar, texts, and vocabulary of the afro-Portuguese creole of the island of Príncipe, Gulf of Guinea*. London: Battlebridge.
- Moreno, Albertina & Tuzine, António (1997). Distribuição social de variáveis linguísticas no português oral de Maputo. In Christopher Stroud & Perpétua Gonçalves (orgs.), *Panorama do Português Oral de Maputo* (pp. 71-91), Vol II. Maputo: INDE.
- Peres, João & Mória, Telmo (1995). *Áreas críticas da língua portuguesa*. Lisboa: Caminho.
- RGPH – 1981. (1987). *Recenseamento geral da população e da habitação*. Lisboa: Instituto Nacional de Estatística.
- RGPH – 1991. *II Recenseamento geral da população e da habitação*. República Democrática de S. Tomé e Príncipe.
- RGPH – 2001. (2003). *Características educacionais da população – Instituto Nacional de Estatística*. S. Tomé e Príncipe.
- Romaine, Suzanne (2000). *Language in society – an introduction to sociolinguistics* (2.^a ed.). New York: Oxford.
- Rougé, Jean-Louis (1992). Les langues des Tonga. In Ernesto de Andrade, Maria Antónia Mota & Dulce Pereira (eds.), *Actas do colóquio sobre crioulos de base lexical portuguesa* (pp. 171-176). Lisboa: Colibri.
- Stroud, Christopher (1997). O *corpus*: antecedentes, quadro teórico e aspirações teóricas. In Christopher Stroud & Perpétua Gonçalves (orgs.), *Panorama do Português Oral de Maputo* (pp. 11-42). Vol I. Maputo: INDE.

Tarallo, Fernando (1985). The filling of the gap: pro-drop rules in Brazilian Portuguese. In L. King & C. Maley (eds.), *Selected Papers from the XIIIth Linguistic Symposium on Romance Languages* (pp. 355-375), Amsterdam: John Benjamins Publ..

Tuzine, António (1997). O papel da rede social na variação e mudança linguística. In Christopher Stroud e Perpétua Gonçalves (orgs.), *Panorama do Português Oral de Maputo* (pp. 75-94) Vol I. Maputo: INDE.

Winford, Donald (2003). *An Introduction to contact linguistics*. Oxford: Blackwell.