



CIEA7 #42:

LÍNGUAS CRIOULAS DE BASE PORTUGUESA NA ÁFRICA.

Gabriel Antunes de Araujo<sup>©</sup>

g.antunes@usp.br

## Relações entre as fonologias das línguas crioulas de STP e a ‘proposta ortográfica’ ALUSTP<sup>1</sup>

*Neste trabalho, discutirei as relações entre as fonologias das línguas crioulas de São Tomé e Príncipe (stp) e o anteprojeto da proposta ortográfica unificada do Alfabeto Unificado para a Escrita das Línguas Nativas de S. Tomé e Príncipe (alustp): o São-tomense (Fôlo ou Forro), o Angolar (ou Ngola) e o Principense (Lung'îê). Assim, apresentarei, em primeiro lugar, as ‘ortografias autorais’ utilizadas anteriormente e, em seguida, mostrarei as principais características do alustp, destacando os seus pontos positivos e negativos. Tanto em stp como em outros países da África, o planejamento linguístico das línguas crioulas mostra-se problemático devido a questões relativas à oficialização, à manutenção, à codificação e à padronização destas línguas, que comumente convivem em ambientes multilíngues e/ou em situação de diglossia. Contudo, propostas ortográficas oficiais têm se mostrado de grande importância, posto que o uso de material linguístico (como gramáticas, dicionários etc.) auxilia o aprendizado das línguas crioulas nas escolas e as tornam línguas de prestígio.*

Línguas crioulas de base portuguesa, ortografia, fonologia,  
São Tomé e Príncipe, Angolar, Principense, São-tomense.

*In this paper, I will discuss the relationship between Portuguese-based Creole languages phonologies of São Tome and Principe (STP) and the draft of the ‘Unified Alphabet for the Languages of STP’ (Alfabeto Unificado para a Escrita das Línguas Nativas de S. Tomé e Príncipe, ALUSTP, for short): the Santomean (São-tomense or Fôlo), the Angolar (or Ngola) and Principense (or Lung'îê). In order to do so, first, I will present some orthographies previously used. Then, I will present phonological features of these languages and show the main features of ALUSTP, highlighting their strengths and weaknesses. In STP, language planning proves to be problematic due to issues relating to the codification, maintenance and standardization of local languages, which commonly exist in multilingual environments and/or in diglossia context. However, elsewhere, official orthography proposals have proved of great value, since the use of linguistic materials (such as grammars, dictionaries etc.) improves the use of these languages in society and make them more prestigious, increasing their chance of survival.*

---

<sup>©</sup> Universidade de São Paulo.

<sup>1</sup> A realização deste trabalho foi possível graças ao financiamento da Pró-reitoria de Pós-graduação da Universidade de São Paulo.

## INTRODUÇÃO

O objetivo deste trabalho é discutir o anteprojeto da proposta de alfabeto unificado para as línguas crioulas de base portuguesa da República de São Tomé e Príncipe (ALUSTP) (Pontífice *et al.* 2010) e, a partir das estruturas fonológicas das línguas crioulas, o São-tomense, o Principense e o Angolar, correlacionar as escolhas ortográficas da proposta com as características dessas línguas. Para tanto, em primeiro lugar, farei uma breve exposição sobre a situação linguística em São Tomé e Príncipe (STP), em seguida, abordarei as dificuldades no estabelecimento de um padrão ortográfico, as propostas idiossincráticas de vários autores do fim do século XIX ao fim do século XX e as propostas ortográficas 'científicas' dos últimos quarenta anos. Na seção seguinte, discutirei ponto a ponto o anteprojeto ALUSTP, no que diz respeito à relação entre a proposta e alguns pontos da fonologia das línguas São-tomense, Principense e Angolar. Ao final, apresentarei as conclusões.

## SITUAÇÃO LINGUÍSTICA DE STP

A República de São Tomé e Príncipe é um país multilíngue. Atualmente, são faladas quatro línguas crioulas (uma dessas línguas, o Kaboverdianu, é nativa da Alta Guiné) e o Português. Embora as relações genéticas dessas línguas sejam aceitas na literatura, há divergências sobre a gênese de todas as línguas crioulas de base portuguesa do Golfo da Guiné (cf. Ferraz 1979, Araujo 2010, Rougé e Schang 2010). Em número total de falantes, as línguas mais faladas são o São-tomense (também conhecida como *Forro*, *Fôlo* ou São-tomense), o Kaboverdianu (ou Crioulo de Cabo-Verde), o Angolar (ou Ngola) e o Principense (ou Lung'iê)<sup>2</sup>. Precisar o número de falantes é uma tarefa complexa, pois os dados sobre o número de habitantes do país são conflitantes e as estimativas variam muito (cf. INE 2001). O CIA The World Factbook<sup>3</sup> estimava, para julho de 2009, a população são-tomense em 212.679 pessoas, das quais cerca de 99,6 mil teriam entre 0 e 14 anos, 105,6 mil teriam entre 15 e 64 anos e as demais, mais de 65 anos. Além disso, cerca de 61% da população ocupa as aglomerações urbanas, enquanto 39% habitam o meio rural. Já os dados oficiais, como a do Instituto Nacional de Estatística (INE) do governo são-tomense que estimava o total de 151.912 habitantes em 2006<sup>4</sup> (cf. Cardoso 2007: 281) estão desatualizados. Embora as estatísticas sejam divergentes, pode-se calcular pela média das estimativas, que a população de São Tomé e Príncipe esteja na casa de 180 mil pessoas

<sup>2</sup> Holm e Madeira (2009) discutem os nomes das línguas crioulas de base portuguesa da África.

<sup>3</sup> <https://www.cia.gov/library/publications/the-world-factbook/geos/tp.html>

<sup>4</sup> [http://www.ine.st/files\\_pdf/STP\\_numeros\\_2006.pdf](http://www.ine.st/files_pdf/STP_numeros_2006.pdf), página 6.

em 2010. Da mesma forma, os dados publicados sobre a situação sociolinguística datam de 2003 e precisam ser renovados. Na ocasião, o INE apontava<sup>5</sup> que, basicamente, três línguas, o *São-tomense*, o *Angolar* e o *Principense* eram faladas, além do Português (o INE excluiu o cabo-verdiano do censo). Para o censo do INE, uma pessoa:

é considerada como [falante] de uma língua (estrangeira ou local) quando ela pode utilizá-la para comunicar oralmente e manter uma conversa com alguém. Quatro línguas foram explicitamente identificadas neste sentido (Português, Fôro, Lungwié, Francês). A população visada era aquela com 5 anos ou mais.<sup>6</sup>

O censo não oferece dados sobre o bilinguismo ou o multilinguismo, porém afirma que cerca de 98,9% da população são-tomense fala o Português (sem defini-lo como primeira língua, L1, ou como segunda língua, L2). Um total de 72,4% falaria o São-tomense e 2,4%, o Principense. O grupo formado por 'outras línguas', incluídos o Angolar e o Kaboverdianu, perfaz o total de 12,8%. O Kaboverdianu é falado por milhares de pessoas, residentes, sobretudo, nas roças ou em propriedade rurais isoladas. Já o Angolar, falado basicamente no Distrito de Caué, pode ser falado por cerca de 5 mil pessoas. É certo, todavia, apontar que a língua Principense, como língua materna, possui uso muito restrito e está limitada à população da Ilha do Príncipe. Maurer (2009: 3, nota 2) afirma que embora o censo de 2001 tenha apontado 16,3% da população ou aproximadamente mil pessoas da Ilha do Príncipe como falantes do Principense, trata-se de um dado superestimado, a não ser que se considere o conhecimento passivo da língua como critério. Isso entraria em contradição com a própria definição de falante dada pelo censo, pois é mencionada explicitamente a capacidade de 'manter uma conversa com alguém' como característica do falante. Agostinho (em preparação) aponta, corroborando Maurer, que há menos de cinquenta falantes da língua Principense. Sendo assim, a segunda língua mais falada em STP (depois do Português) é o São-tomense: o censo de 2001 mostrava que cerca de 72% da população é falante desta língua. Os níveis de domínio da língua variam, mas há claramente duas tendências em curso, observadas em trabalho de campo: o São-tomense tem se tornado a língua crioula mais falada (mesmo pelos outros grupos minoritários) em STP e, cada vez menos, as pessoas o aprendem como língua materna, papel desempenhado pelo Português. Igualmente, o São-tomense é a língua crioula que possui o maior número de falantes e que goza de maior prestígio na sociedade.

Atualmente, o Português é a língua oficial da República de São Tomé e Príncipe, empregada em todas as comunicações de Estado, na educação e na mídia. A norma portuguesa europeia é ensinada nas escolas e dominá-la é o objetivo do sistema escolar.

<sup>5</sup> [http://www.ine.st/files\\_pdf/Tema\\_Educa%E7%E3o.pdf](http://www.ine.st/files_pdf/Tema_Educa%E7%E3o.pdf), páginas 48-52.

<sup>6</sup> [http://www.ine.st/files\\_pdf/Tema\\_Educa%E7%E3o.pdf](http://www.ine.st/files_pdf/Tema_Educa%E7%E3o.pdf), página 14.

No entanto, há variação no país. Esta variação floresce no contexto de uma língua empregada pela população tanto como língua materna quanto segunda língua. O contexto multilíngue de STP, associado à forte migração campo/cidade, sobretudo nos últimos trinta anos, bem como ao sistema escolar não universalizado, favorece uma transmissão irregular do Português (para os falantes de Português L2) ou um afastamento da pressão da norma lusitana (para os falantes de Português L1). Duas consequências deste contexto podem ser observadas nas variantes regionais conhecidas como *Português dos tongas* (Rougé 1992, Lopes e Baxter 2010, *inter alia*) e no *Português vernacular são-tomense* (cf. Figueiredo 2009, Martiniuk 2010), isto é, exemplos de variantes reestruturadas do Português europeu. As variantes reestruturadas convivem com variantes próximas da ‘norma’ europeia, empregada pela pequena elite com acesso à língua culta portuguesa. Portanto, o uso generalizado do Português na mídia, como língua oficial, e na escolarização não pode ser descartado do contexto sociolinguístico de STP.

Neste ambiente multilíngue, contudo, não havia, até o momento, tentativas oficiais de normalização ortográfica das línguas nacionais. Portanto, as línguas nacionais ficaram limitadas ao mundo da oralidade, salvo raras exceções, comentadas na próxima seção.

## ORTOGRAFIA

Segundo Hosken (2003), descrever uma língua ortograficamente é importante por viabilizar a criação de documentos e o desenvolvimento de outras atividades ligadas à ortografia, tais como a literatura e o registro do Estado e de particulares (cf. para uma contextualização ampla dos processos de gramatização e padronização de línguas, ver Aurox 1992). Contudo, criar uma ortografia é um processo complexo, pois envolve aspectos técnicos, financeiros, políticos e culturais. Para se criar uma ortografia, é necessário uma boa descrição da língua, muito embora, haja casos em que a descrição parcial da língua é considerada melhor do que sua descrição total, devido às dificuldades que surgem no meio do processo. Basicamente, uma ortografia é considerada, entre outros fatores, como “*one of the ways that people write text in a particular language*” (Hosken 2003: 1). Além de gerar informações suficientes para se implementar um sistema de escrita, alguns aspectos devem ser levados em conta na criação da ortografia de uma língua (Hosken 2003):

- (1) Quantos falantes a língua tem? Os falantes funcionais da língua podem ser considerados potenciais usuários da ortografia?;

- (2) Qual a porcentagem de falantes alfabetizados (em outras línguas)? As comunidades linguísticas podem ter divisão social ou geográfica e seus falantes podem ser alfabetizados em diferentes ortografias, o que pode gerar vários tipos de uso ortográficos, sem levar em conta o nível de alfabetização na mesma;
- (3) A ortografia é atualmente usada, de que maneira e em qual medida? Saber se a ortografia é restrita apenas a alguns membros da comunidade linguística ou não, se há uma ortografia arcaica e outra mais moderna;
- (4) Qual porcentagem de falantes pode ser esperada para se tornar alfabetizada na ortografia? Os falantes da língua podem se tornar alfabetizados na ortografia baseados em diferentes divisões nacionais; e
- (5) Qual a atitude de cada comunidade em relação à ortografia? É importante no caso do grupo linguístico ter múltiplas comunidades para se saber a atitude de cada uma delas frente ao novo sistema.

Assim, é importante levar em consideração o contexto sociolinguístico no momento de elaboração de uma ortografia, pois através dele é que se terá o conhecimento necessário para se chegar à maneira ideal de criação desse sistema e também do potencial dos usuários dentro das comunidades linguísticas (Hosken 2003). Deve considerar, todavia, que a escrita não pode ser considerada uma representação da fala e não apresenta a capacidade de reproduzir vários fenômenos orais como a prosódia, os movimentos do corpo e dos olhos. A escrita, por sua vez, tem seus elementos próprios: o tamanho, o formato, a cor e o tipo da letra. Sendo assim, ambas as modalidades podem ser vistas não apenas como práticas, mas também como usos da linguagem. A escrita pode ser conservadora e, por ser registrada através de uma mídia temporal, pode atingir seu destinatário independentemente de sua época. Dessa maneira, o escritor é capaz de ordenar e expressar seus pensamentos com precisão e perenidade, ao mesmo tempo em que pode reduzir as situações de ambiguidade, em situações ideais.

De forma incontestável, na sociedade atual, a modalidade escrita é um bem social imperativo, que deve ser utilizada em todos os tipos de meio (rural ou urbano) e que se tornou essencial para o fortalecimento de línguas minoritárias, pois, muitas vezes, a manutenção de uma língua minoritária está associada ao seu prestígio frente às línguas majoritárias e a ortografia é um dos instrumentos que promove o prestígio das línguas.

A quem deve ser atribuída a tarefa de se propor uma ortografia? Ao longo da história, esta tarefa tem sido conduzida por indivíduos (sem conhecimento técnico), por linguistas (com conhecimento técnico) e/ou missionários ou sob encomenda de governos. No primeiro caso, um indivíduo ou indivíduos isoladamente, por iniciativa própria, cria(m) um sistema

ortográfico, normalmente baseado(s) em um sistema de uma língua majoritária disponível pelo método da cópia do plano detalhado (cf. Diamond 1997). Em geral, são criados poucos novos símbolos ou até mesmo nenhum. As especificidades fonético-fonológicas da língua podem ou não ser levadas em consideração. No segundo caso, linguistas (ou outros profissionais) ou missionários criam um sistema que pode ou não ser baseado em um sistema majoritário previamente existente. Comumente, estes profissionais levam em consideração as características da língua, pois receberam algum tipo de treinamento prévio. No terceiro caso, o Governo ou alguma outra instituição encomenda estudos a um grupo de especialistas ou adota um sistema proposto por um especialista ou por grupos de especialistas. Em todos os casos, o sistema ainda precisa ser, *de facto*, utilizado pela população-alvo.

Nos primórdios da ortografia da língua portuguesa, por exemplo, a iniciativa de vários indivíduos criou sistemas que, embora com múltiplas diferenças, compartilhavam muitas características (cf. Cagliari 1994). No entanto, a publicação de gramáticas no século XVI, sobretudo a de Fernão de Oliveira (1536) e a de João de Barros (1540), e do épico 'Os Lusíadas' (em 1572) lançou as bases para um modelo ortográfico. Ainda assim, várias reformas (cf. Aguiar 2007) ortográficas não foram capazes de limar as pequenas diferenças entre as variantes do Português, como pode ser observado nas diferenças entre a norma lusitana e a brasileira. Portanto, a construção social da ortografia é aberta a reformas, posto que a ortografia é um instrumento de representação da língua e não a própria língua. Além disso, a língua muda ao longo do tempo e a ortografia pode acompanhar esta mudança.

Basicamente, portanto, as ortografias se dividem em dois grupos: as ortografias científicas e as não-científicas. As ortografias científicas são aquelas nas quais se busca a relação de univocidade entre a representação do som (fonema) e o grafema, ou seja, para cada som, uma letra. Portanto, sons distintos (exceto alofones) devem possuir símbolos diferentes. Normalmente, ortografias científicas são propostas por linguistas ou outros especialistas. Já as ortografias não-científicas têm as mais variadas origens, porém, em geral, não conseguem manter a relação de univocidade e usam vários símbolos para um mesmo som ou o mesmo som é representado por vários símbolos. Contudo, muitas ortografias científicas são obrigadas a fazer uma ou outra concessão. No que diz respeito às línguas de São Tomé e Príncipe, mesmo para as ortografias científicas, há dificuldades em se representar as consoantes nasais, a oposição entre vogais médias altas e baixas, o sistema tonal e o sistema acentual.

## ORTOGRAFIAS PRÉVIAS DAS LÍNGUAS CRIOULAS

Este trabalho trata da proposta de um alfabeto unificado para as três línguas crioulas do Golfo da Guiné: o São-tomense, o Principense e o Angolar. Sobre esta última, não há evidências de ortografias prévias. Há, no que diz respeito ao São-tomense, quase uma dezena de ortografias autorais prévias e, para o Principense, três. Contudo, em primeiro lugar, serão abordadas as propostas ortográficas prévias para o São-tomense, seguidas pelas do Principense.

### São-tomense (Fôlo, Forro ou Santomé)

Negreiros (1895) é o primeiro trabalho a empregar uma ortografia para o São-tomense. Para representar os sons do São-tomense, sua ‘gramática’ emprega um sistema ortográfico de inspiração lusitana, com pequenas modificações. Negreiros, assim como as ortografias futuras, enfrenta dificuldades na representação das vogais médias ([↔, e, ↓, o]), da nasalidade, das consoantes palatais e ‘sibilantes’, e no acento prosódico. Assim, a vogal média-baixa posterior [e] é ora representada com o grafema <ê>, ora com o <e>, ao passo que a vogal média-alta posterior [↔] é representada com o grafema <é><sup>7</sup>. A consoante fricativa alveolar surda [s] é representada, como no sistema Português, como <c> diante de <e> e <i>, <ç>, <s> e <ss>. Já a consoante africada alveolar surda pode ser representada como <ch>, <gi> e <s>. A consoante alveo-palatal surda [tʃ] é representada por uma combinação de grafemas inexistente no Português <tch>. O acento prosódico pode ser marcado com os diacríticos grave e agudo, ou até mesmo não ser marcado. O sistema ortográfico de Negreiros foi usado para documentar a língua São-tomense no final do século XIX. Não há evidências que a gramática tenha sido utilizada com fins educacionais, nem a sua ortografia empregada por outros autores.

A ortografia de Negreiros pode ser observada nos exemplos a seguir:

(6) Exemplos ortográficos de Negreiros (1895):

<b>sálu</b>	‘sal’	[ᵛsa.lu]
<b>vêspla</b>	‘vespa’	[ᵛves.pla]
<b>fé mangaçón</b>	‘xingar’	[f↔.mã.ga.ᵛsõ]
<b>quentchi</b>	‘quente’	[ᵛkẽ.tʃi]
<b>méssé</b>	‘querer’	[m↔.ᵛs↔]
<b>méssê</b>	‘professor’	[ᵛm↔.se]
<b>fleminga</b>	‘formiga’	[fle.ᵛmĩ.ga]
<b>figiga</b>	‘fisga’	[ᵛfi.ʃ.ga]
<b>chilòla</b>	‘ceroulas’	[ʃi.ᵛlo.la]

<sup>7</sup> Neste texto, as representações fonológicas são representadas com barras inclinadas, / /, as fonéticas são representadas com colchetes, [ ], e as grafemáticas com diples < >. Os exemplos de representações grafemáticas em negrito foram retirados dos autores em questão, ou do anteprojeto da ALUSTP e os em fonte normal foram inferidos a partir do sistema da ALUSTP.

Depois de um hiato de quase cem anos, Luiz Ivens Ferraz, em sua tese de doutorado propõe uma nova ortografia para o São-tomense. Ferraz (1979) afirma que, já nesta data, havia várias notações em materiais já publicados sobre o São-tomense<sup>8</sup>, sendo que estes são largamente baseados na ortografia do Português, além do emprego de diacríticos e grafemas ‘mais do que o necessário’ (Cf. Ferraz 1979: 55). O autor propõe uma notação quase-científica, com pequenas concessões. A representação das vogais orais, por exemplo, mantém os grafemas <e> e <o> para as vogais médias [↔, e] e [↔, o], respectivamente. Ferraz assume que os pares /o/ e /ɔ/ e /e/ e /ɛ/ são minimamente contrastivos e, portanto, não há necessidade de diferenciá-los ortograficamente, seja através de diacríticos ou outros símbolos gráficos. As vogais nasais, por sua vez, são grafadas em final de palavra como vogal + <m> diante de <p> e <b>, ou seja, diante de consoantes bilabiais, e como vogal + <n>, alhures. Ferraz faz uma ressalva quanto à escrita da palavra <ũa> ‘um, uma’. Neste caso, a vogal nasalizada, marcada com o diacrítico til, é uma ocorrência única e a nasalidade da vogal não pode ser representada com nenhuma consoante nasal.

No que diz respeito à representação das consoantes, Ferraz utiliza os mesmos grafemas empregados no Português, com pequenas modificações: o grafema <c> representa o fonema /tʃ/. Sua contraparte sonora, a consoante africada alveopalatal sonora /dʒ/ é representada pelo grafema <dj>. Desta forma, a consoante fricativa alveolar sonora [ʒ] é representada por <j>. A consoante fricativa alveolar surda, por sua vez, é representada como <s> quando precedendo consoante na mesma sílaba e como <x> alhures. O grafema <nh>, usado em Português para representar o fonema /ɲ/, é mantido para representar o fonema nasal palatal /ɲ/. O fonema /k/, diferentemente das ortografias de inspiração lusitana, é grafado como <k>. Para os glides /w/ e /j/, Ferraz propõe o uso dos grafemas <w> e <y>. Segundo Ferraz (1979: 36), o fonema /ɰ/ não ocorre em São-tomense, sendo incorporado como /y/ ou /ɲ/ nos exemplos de origem portuguesa. Porém, em seguida, Ferraz afirma que o fonema é raro e deve ser representado como <lh>.

Um apóstrofo, segundo Ferraz (1979), deve ser usado para indicar a elisão de uma vogal em junção de palavra, muito embora possa ser dispensado em alguns contextos. Todavia, Ferraz não especifica estes contextos. Por fim, em seu trabalho, há poucos exemplos de implementação de sua própria proposta ortográfica, uma vez que os exemplos de sua gramática estão representados de forma fonêmica.

Já o trabalho de Espírito Santo (1998, 2001) não se apresenta formalmente como uma proposta ortográfica, mas, por conter centenas de exemplos na língua São-tomense. A ortografia adotada por Espírito Santo é baseada na ortografia do Português. E, por este

---

<sup>8</sup> Cf. Hagemeijer 2009, para informações sobre panfletos apócrifos escritos em São-tomense contra o Governo Colonial Português ao longo do século xx.



motivo, esta ortografia se assemelha a várias outras empregadas idiossincriticamente por muitos autores são-tomenses.

Espírito Santo aplica um sistema no qual a diferenciação das vogais orais médias altas e baixas é feita através do uso de diacríticos: /e/ → <ê>, /ɛ/ → <é>, /ɔ/ → <ó> e /o/ → <ô>. Contudo, esta opção nem sempre é obedecida, ou seja, as vogais médias podem também aparecer sem nenhum diacrítico. A marcação de tonicidade, quando na sílaba tônica aparecem as vogais orais /i/, /a/ e /u/ é também feita através do uso de diacrítico grave (´), como ocorre em **clincatá** [klĩkaŋta], por exemplo. Portanto, o diacrítico é empregado para diferenciar as vogais médias e para marcar o acento prosódico. Com relação às consoantes, o sistema preserva inconsistências como no sistema português. A consoante oclusiva velar surda /k/, por exemplo, pode ser representada como <c>, <qu> e <k>, assim como no Português:

(7) Uso dos grafemas <c>, <qu> e <k>

/k/					
<c>		<qu>		<k>	
Forma fonética	Forma ortográfica	Forma fonética	Forma ortográfica	Forma fonética	Forma ortográfica
[makeʧke]	<maquêquê>	[ʧblāku]	<blancu>	[ŋk↔]	<ké>
[ūa ʧkwatu]	<ūa quatu>	[ʧkalu]	<cálu>		
[kaʧki]	<caqui>	[tokaʧdo]	<tocado>		
[lekɛleʧkɛ ]	<lequé lequé>	[kaʧki]	<caqui>		
[liʧkeza]	<liqueza>	[kukuʧnu ]	<cucunú>		
[ʧkwātu]	<quantu>	[ʧkuma]	<cuma>		
[keʧbladū]	<quebladu>	[kleʧkle]	<clecle>		

A consoante fricativa alveolar surda, /s/, por sua vez, pode ser representada por quatro grafemas, <s>, <ss>, <ç> e <c>, como em (8). A consoante fricativa pós-alveolar surda, /ʃ/, é representada, na proposta de Espírito Santo, pelos grafemas <x>, como em **vivenxa** [viʧvɛ̃a], <ch>, em **tudachi** [tudaʃi] e <s>, como em **stlijón** [ʃtljõ]. Portanto, algumas vezes, sons diferentes são representados pelo mesmo símbolo.

(8) Uso dos grafemas <s>, <ss>, <ç> e <c>

/s/							
<s>		<ss>		<ç>		<c>	
Forma fonética	Forma ortográfica	Forma fonética	Forma ortográfica	Forma fonética	Forma ortográfica	Forma fonética	Forma ortográfica
[sãtoŋ↔]	<santopé>	[misuʧa]	<missuá>	[poʧsô]	<poçón>	[ŋnāse]	<nance>
[sapenlí]	<sapelin>	[baʧsola]	<bassola>	[ŋtasu]	<taçu>	[ŋs↔du]	<cédu>
[saʧta]	<satá>	[ŋbisu]	<bissu>	[ŋkoso]	<coço>	[fitʃinselu]	<fitxicelu>
[ŋsɛpk↔]	<sémplé>	[ʧvɛ̃su]	<véssu>	[ŋlɔsa]	<loça>	[flense]	<flece>
[ŋsebe]	<sebe>	[ŋusua]	<ussua>	[ŋmosu]	<moçu>		
[ŋsɔbe]	<sobe>	[decaʧsô]	<decassón>	[okonso]	<ocoço>		

A fidelidade ao sistema português pode ser observada na aplicação da ortografia de forma generalizada. A consoante oclusiva velar sonora, /g/, por exemplo, é representada como <g> antes de /a/, /o/ e /u/, enquanto diante de /e/ e /i/ este fonema é representado por <gu>. A consoante fricativa pós-alveolar sonora, /ŋ/, pode ser representada tanto como <g> como como <j>, porém o autor não explicita qualquer regra para clarificar a utilização de um grafema ou outro e, na própria obra, ora emprega um, ora emprega outro, como em [liiŋmɛtu] grafado **ligimentu** e **lijimentu**. A consoante africada alveo-palatal sonora, /dʒ/, é representado por <dj>, como em **djambi** [dʒãbi], enquanto a consoante africada alveo-palatal surda /tʃ/ é grafada <tx>, como em **txoco** [tʃok]. Por fim, na lista dos fonemas inexistentes no Português, Espírito Santo, emprega a combinação de grafemas <ng> para representar a consoante nasal velar sonora, /ŋ/, porém, ao mesmo tempo, como <ngue> [ŋge].

A escolha de uma ortografia lusitanizada tem a vantagem de ser conhecida da maioria do público-alvo, via escolarização em Português. Além disso, não é necessário memorizar um novo sistema. Por outro lado, há fonemas específicos do São-tomense não existentes em Português. Por isso, soluções *ad hoc* são sempre necessárias. Ademais, qualquer sistema baseado na ortografia portuguesa deve considerar as inconsistências e as soluções etimológicas incorporadas ao longo de mais de 500 anos. A ortografia portuguesa apresenta um alto grau de inconsistência. Assim, em contextos semelhantes, um grafema representa mais de um fonema e vice-versa. Isso gera grande dificuldade no aprendizado. A base de emprego da ortografia parece ser uma conexão etimológica com a ortografia portuguesa, muito embora esta conexão seja impossível nas palavras de origem não-portuguesa ou nas palavras cujo elo etimológico não são mais visíveis.

Fontes (2007), em sua dissertação de mestrado, organiza um dicionário bilíngue São-tomense-Português e, simultaneamente, propõe uma ortografia para o São-tomense. Em grande parte, Fontes segue o modelo proposto por Ferraz. Segundo Fontes, o sistema fonológico do São-tomense dispõe, com poucas diferenças, da quase totalidade dos fonemas do Português, a língua lexificadora, porém com alguns fonemas ou processos fonológicos diferentes. O autor, por exemplo, também citando Ferraz (1979: 25), oferece a seguinte definição para o processo fonológico de harmonia vocálica: “*Em são-tomense a harmonia vocálica consiste em uma tendência da mesma vogal ocorrer em duas sílabas consecutivas dentro de um morfema*”. Fontes (2007) postula que se o falante tomar ciência deste processo, ele poderá atribuir o mesmo grau de abertura às vogais de uma palavra, basta que este grau de abertura de uma dessas vogais seja identificada com um diacrítico, preferencialmente a tônica, e o falante atribuirá este grau para as demais vogais. Assim, a palavra **dotôlo**, tendo recebido a marca na sílaba tônica, representada pelo acento circunflexo e também definido o grau de abertura da vogal, seria realizada foneticamente

[doŋtolo], de forma transparente para o falante treinado, ao invés de ser realizada com vogais discordantes, do ponto de vista da altura. Portanto, para Fontes, as vogais médias, e apenas essas, são grafadas com um acento circunflexo. Adicionalmente, nas palavras com mais de uma sílaba, apenas a tônica é marcada, uma vez que, o processo de harmonia vocálica é altamente produtivo na língua e indica a qualidade das vogais átonas a partir da vogal tônica. Com relação aos pares de vogais contrastivos /o/ e /ɔ̃/ e /e/ e /ɛ̃/, Fontes propõe o emprego de um sinal diacrítico, diferentemente de Ferraz. As vogais orais médias seriam, portanto, grafadas /e/ → <ê>, /ɛ̃/ → <e>, /o/ → <ô> e /ɔ̃/ → <o>. Já em relação às vogais nasais, o autor diz adotar uma proposta anterior (Alcântara e Hagemeyer 2001), na qual a nasalidade destas vogais, em fim de palavra, é representada com ‘n’ após a vogal. Por isso, Fontes não emprega diacríticos para marcar a nasalidade dessas vogais, incluindo as que pertençam às sílabas tônicas. Portanto, em relação à nasalidade nas vogais, é proposto que seja representada com ‘n’ ou ‘m’, de acordo com o ponto de articulação da consoante seguinte. Mas assim como Ferraz (1979), Fontes afirma que a única exceção à esta regra é a palavra *ũa*, por se tratar de um arcaísmo do Português que permaneceu no São-tomense.

Os dois glides do São-tomense são representados pelos grafemas <w> e <y><sup>9</sup>. Com relação às consoantes, excluindo os grafemas comuns, são empregados os seguintes: /ʎ/ → <ch>, /tʎ/ → <tch>, /ç/ → <j>, /dç/ → <dj>, /ŋ/ → <ng>, /ɲ/ → <nɲ>, /r/ → <r>. Sobre a consoante fricativa pós-alveolar surda, /x/, Fontes discute o fato de este fonema poder ser representado na ortografia portuguesa de diversas formas diferentes: como <x>, na posição de onset, ‘xícara’ [ʁiçakã]; <ch>, também na posição de onset, como em ‘chave’ [çav]; <s>, na posição de coda, em ‘mais’ [mai], e também como <c>, como em ‘cinco’ [sẽkũ], como ocorre no dialeto de Lisboa. Reflete que este mesmo problema poderia ocorrer no São-tomense, optando por representar este fonema como <ch>, assim como sugere Hagemeyer (2000:16). O argumento para esta escolha é evitar uma possível confusão com a representação utilizada pelo IPA (International Phonetic Alphabet) para a consoante fricativa velar surda, [x] do Anobonense, língua crioula de base portuguesa falada na Guiné Equatorial. Esta confusão seria improvável nas línguas de STP, mas possível na comparação destas com o Anobonense. Por outro lado, esta justificativa não se aplica ao mundo são-tomense, no qual a utilização do alfabeto fonético pela população-alvo da ortografia praticamente é nula. Fontes também propõe que o fonema /ʎ/, quando ocorrer na posição de coda, seja representado por <s>. Segundo Fontes, a consoante vibrante alveolar /r/, antes realizada apenas em empréstimos do Português, vem sendo cada vez mais utilizada

<sup>9</sup> Na história dos trabalhos sobre o São-tomense, todos os autores que afirmam empregar o Alfabeto Fonético Internacional usam o símbolo ‘y’ para representar o glide aproximante palatal [j]. No entanto, a *International Phonetic Association* recomenda o símbolo /j/ para o glide aproximante palatal. O símbolo /y/ é empregado para a vogal posterior alta fechada arredondada.

como variante livre em relação à consoante //l/. Assim, a opção por /r/ ou //l/ depende do grau de domínio do São-tomense e/ou do Português por parte do falante, um fator correlacionado à idade, ao meio social e cultural predominante, e ao contexto de comunicação em que este falante está inserido.

Com relação às ligações entre palavras, Fontes (2007) segue novamente a sugestão dada por Alcântara e Hagemeijer (2001), em que o uso do apóstrofo é aconselhável nas contrações entre verbos e pronomes, entre as preposições *di* e *ni* e outras categorias sintáticas, quando efetivamente ocorrem, excetuando os pronomes pessoais, os demonstrativos e os advérbios de lugar, o que permite, segundo ele, evitar a confusão com os outros casos (Cf. Fontes 2007: 22). Por fim, o uso do hífen é proposto nos nomes compostos, assim como nos casos de reduplicação com origem no superstrato, cujas bases tenham autonomia morfológica e semântica. Os exemplos do substrato em que as bases possuem autonomia morfológica e semântica são raros, pois são constituídos, em sua maioria, por reduplicações, como os ideofones, em que as bases não funcionam autonomamente. Nesses casos, não é sugerido o uso do hífen, com exceção dos casos de hiato: **ũa-ũa** ‘cada qual’, e a variante **din-in-in** do ideofone **din**.

### Principense (Lung’le)

Há poucos documentos sobre a língua principense e o uso de ortografias idiossincráticas é limitado, não havendo documentos produzidos pela população local. Nesta seção, serão comentadas as ortografias das gramáticas de Ribeiro (1888), Günther (1973) e a de Maurer (2009)<sup>10</sup>.

A gramática de Ribeiro é o primeiro documento sobre a língua Principense. Ribeiro afirma que ‘se deve considerar o dialecto da ilha do Príncipe como uma corrupção da lingua portugueza’ (cf. Maurer 2009: 262). Desta forma, a escolha para um determinado grafema é fortemente influenciada pelo uso dos grafemas na ortografia da língua portuguesa que emprega múltiplos grafemas para o mesmo som e vice-versa. Se, por um lado, há uma clara conexão etimológica com a ortografia das palavras portuguesas, como é o caso de **cé** [ɲs↔] ‘céu’, por outro, Ribeiro, consciente das diferenças na pronúncia, provocadas pela assimetria dos quadros fonológicos do Principense e do Português, precisa criar um sistema ortográfico próprio, com o uso de combinações de grafemas do Português como as consoantes alveo-palatais /tʃ/ <tx> (segundo Ribeiro ‘t forte tch’), como em <txá> [ɲtʃ’a] ‘tia’, e <dg> para /dʒ/, como em **médge** [ɲmedʒe] ‘mês’. Embora, a palavra ‘tio’ seja também grafada <ti>, sem palatalização. Ainda para representar as consoantes, Ribeiro emprega ora <gh> para [g], como em **ghêza** ‘igreja’, ora <g> como em **migo** ‘amigo’. No que diz respeito à representação da nasalidade, Ribeiro utiliza, consistentemente, uma vogal marcada com o

<sup>10</sup> Conferir também Valkhoff (1966) e Rougé (2004).

diacrítico til na palavra **ũa**. Nos demais casos, ele emprega uma consoante nasal na coda silábica ou até mesmo uma vogal com til. Sua marcação da nasalidade é inconsistente, pois uma mesma palavra, como ‘pão’ aparece grafada como <upã>, <upam> e <upan>. Outra palavra, a forma do pronome pessoal primeira pessoal é grafada como <ũ> ou como <un>. Portanto, a inconstância na aplicação dos grafemas é uma característica notável na ortografia de Ribeiro.

A ortografia de Günther (1973) emprega, basicamente, o alfabeto IPA modificado. Günther marca os tons com diacríticos, sendo o tom baixo *default*, não marcado. O tom alto é marcado com o diacrítico agudo (´) e o tom decrescente é marcado com o diacrítico circunflexo (^). O autor usa os símbolos recomendados pela Associação Fonética Internacional, exceto o símbolo da consoante álveo-palatal surda /tʃ/ e do glide /j/, para os quais Günther emprega os grafemas <c> e <y>, respectivamente. Portanto, a sua ortografia é científica e, devido ao fato de o seu trabalho ter sido escrito em alemão, pouco influenciou os principenses.

A ortografia empregada por Maurer (2009) é de base científica, porém com uma série de adendos, sobretudo para lidar com a questão da representação das vogais médias, de algumas consoantes, do tom e da nasalidade. A vogal média /ɘ/ é representada pelo grafema <e>, enquanto a vogal média /ɤ/ é representada por <o>. As vogais médias-altas /e/ e /o/ são representadas, respectivamente, por <ê> e <ô>. A consoante fricativa palatal surda /ʃ/ é representada por <x>. Por isso, o grafema <tx> representa a consoante álveo-palatal surda /tʃ/. De forma semelhante, a consoante fricativa palatal sonora /ʒ/ é representada por <j>. Ao passo que, para representar o glide /j/, Maurer emprega o grafema <y>. A consoante álveo-palatal sonora /dʒ/ não ocorre como fonema no Principense, porém as sequências [d + j] são representadas com os grafemas <dy>. As consoantes lábio-velares /k/ e /g/ são representadas com os grafemas <kp> e <gb>, respectivamente. Quanto à representação da nasalidade, Maurer emprega as consoantes <n> e <m>, seguindo a vogal nasalizada. O grafema <m>, em coda, é sensível à presença da consoante homorgânica bilabial. O grafema <n> é usado alhures. Os tons e o acento lexical não são representados na ortografia de Maurer.

### **Angolar (Ngola)**

O Angolar é a única língua crioula de STP surgida em uma comunidade de escravos fugidos das roças ou quilombo (cf. Ferraz 1976, Seibert 2007) e, por isso, apresenta propriedades linguísticas que o distanciam do CST e do CP, sobretudo na realização das consoantes, nos processos fonológicos e em seu aporte lexical. Das línguas abordadas neste trabalho, o Angolar é aquela com o menos expressivo uso da escrita (não há, por exemplo, gramáticas publicadas na época da colonização portuguesa), porém, há duas

gramáticas descritivas científicas: Maurer (1995) e Lorenzino (2000)<sup>11</sup>, fazendo do Angolar uma das mais bem descritas línguas de STP. Contudo, apesar da disponibilidade deste material, a população angolar não possui um sistema ortográfico usado amplamente.

Na proposta ortográfica de Maurer (1995) para o Angolar, não há representação dos tons, as vogais médias /e/ e /o/ são representadas com grafemas com o diacrítico circunflexo, <ê> e <ô>, respectivamente, ao passo que /↔/ e /↵/ são representados por <e> e <o>. A consoante fricativa palatal surda /ʃ/ é representada por <x>, embora em começo de sílaba, em onsets complexos, Maurer também emprega o grafema <s>, como na variação entre **staka** e **xtaka**. Paralelamente, o grafema <tx> representa a consoante álveo-palatal surda /tʃ/. De forma semelhante, a consoante fricativa palatal sonora /ʒ/ é representada por <j> e o grafema <dj> é usado para representar a consoante álveo-palatal sonora /dʒ/. Para representar o glide /j/, Maurer emprega o grafema <y>. As consoantes /t/ e /d/ são representadas ortograficamente por <th> e <dh>, respectivamente. Há, no Angolar, uma ocorrência mais notável de consoantes pré-nasalizadas, representadas por Maurer, com o grafema <n> na posição inicial. Assim, tem-se <nd>, <ndj>, <nk> e <ng>. As consoantes pré-nasalizadas com segundo elemento bilabial são representada com 'm' inicial <mp> e <mb>, exceto <nf>, representada com 'n' inicial. Maurer (1995: 28) menciona a existência do fonema implosivo /ɓ/. No entanto, alhures (1995: 29), Maurer descreve este fonema como não-implosivo /b/.

## O ALUSTP

O anteprojeto do *Alfabeto Unificado para as Línguas Nativas de São Tomé e Príncipe* (ALUSTP), apresentado ao Governo de STP por uma Comissão formada por acadêmicos e intelectuais, é uma proposta para se representar alfabeticamente as línguas São-tomense, Angolar e Principense. A Comissão, no preâmbulo do anteprojeto do Decreto, declara que embora as línguas sejam mutuamente ininteligíveis, *partilham um número substancial de propriedades lexicais e gramaticais, justificando, assim, uma ortografia unificada*. Portanto, os cognatos lexicais juntamente com o compartilhamento de propriedades gramaticais justificam a adoção de um alfabeto unificado.

O alfabeto adotado é de base fonético-fonológica, em detrimento de um alfabeto lusitanizado, pois o léxico de origem portuguesa sofreu 'profundas alterações fonológicas', há grande porcentagem de cognatos compartilhados pelas três línguas e não há tradição escrita. O item 6 do Artigo 5º do anteprojeto do Decreto sugere que o princípio da

<sup>11</sup> Não tive acesso ao trabalho de Lorenzino em tempo hábil.

univocidade entre um fonema e sua representação gráfica (um grafema) esteja refletido na proposta, contudo este princípio não é seguido integralmente na proposta.

O anteprojeto propõe um alfabeto com trinta e um grafemas: cinco vocálicos <a, e, i, o, u>; e vinte e seis consonantais: <b, d, dh, dj, e, f, g, gb, j, k, kp, l, lh, m, n, nh, p, r, s, t, th, tx, v, w, x, y, z>.

Todavia, todas as três línguas possuem sete fonemas vocálicos e, desta forma, utilizando-se somente cinco grafemas, não se pode representar estes sete fonemas sem recorrer ao uso de diacríticos. Neste caso, o ALUSTP representa as vogais médias-altas /e, o/ com o diacrítico circunflexo <ê, ô> e as vogais médias-baixas /ɛ, ɔ/ sem diacrítico <e, o>, respectivamente. As vogais longas do Angolar e do Principense são representadas por uma sequência de vogais idênticas, como em <teela> ‘estrela’ do Principense, ou <dooba> ‘dobra (moeda nacional)’. Embora alguns autores (Ferraz 1979, Maurer 1995, Maurer 2009, entre outros) apontem a existência de vogais nasais nas línguas de STP, o estudo de fonologia experimental de Araujo e Agostinho (2010) mostrou que não há vogais fonologicamente nasais nessas línguas, sendo a nasalidade bifonêmica. Não obstante, o anteprojeto do ALUSTP defende a existência de vogais nasais e as representa como bifonêmicas. Assim, as ‘vogais nasalizadas são representadas no final de sílaba inicial ou medial com os grafemas vocálicos seguidos por <n>, porém, antes das consoantes <p> e <b>, usa-se o grafema vocálico, seguido por <m>’. Deste modo, tem-se, em São-tomense <tambôlô> ‘tambor’, em Principense <kompwe> ‘compadre’ e <kampu> ‘campo’ em Angolar, porém, em São-tomense <tamen> ‘grande’ e em Principense, <kôndê> ‘esconder’. Excepcionalmente, a representação do numeral e determinante <ũa> ‘um, uma’, em todas as línguas, e dos ideofones com repetição de vogal nasalizada, como no São-tomense <pya sũũũ> ‘olhar fixamente’, deve ser feita com o diacrítico nasal. Neste caso, a razão provém do fato de serem formas lexicalizadas.

A representação ortográfica das vogais nas três línguas viola o princípio de univocidade da ortografia, uma vez que um fonema pode ser representado por mais de um grafema. Além disso, a nasalidade fonética é idêntica ou similar em todos os ambientes, porém o ALUSTP mantém um elo etimológico com a ortografia lusitana ao preservar a notação do grafema <m> antes de <p> e <b>, e <n> em outras posições. Trata-se de uma escolha polêmica, pois atrela o alfabeto ao comportamento da ortografia de uma outra língua, o Português, muito embora em vários outros aspectos a ortografia proposta se distancie da ortografia portuguesa. Além disso, este alfabeto defende a existência de vogais intrinsecamente nasais, ao passo que sua representação ortográfica ‘bigrafêmica’ opõe-se diretamente à existência de vogais nasais.

A variação linguística é natural em todas as línguas vivas do mundo, contudo, normatizar uma língua e propor um sistema ortográfico necessariamente envolvem a eleição

de uma variante linguística como padrão. O ALUSTP propõe ortografias alternativas para casos de 'variação vocálica'. Isso pode trazer consequências indesejadas para o aprendiz. O anteprojeto cita três exemplos de variação ortográfica de cada língua: São-tomense <stluvisu> ~ <stlivisu> 'trabalho'; <bomu> ~ <bamu> 'vamos', <nanta> ~ <nantan> 'jamais'; Angolar <bi> ~ <bu> 'vir', <fibika> ~ <fubuka> 'beber com atenção', <lumo> ~ <lumon> 'irmão'; Principense <tem> ~ <ton> 'até', <ôzên> ~ <uzên> 'joelho' e <kumi> ~ <kumin> 'lugar'. Não fica estabelecido, no entanto, se a variação ortográfica poderá ocorrer somente nas palavras listadas no anteprojeto ou em todas as palavras com variação linguística ou casos similares. Caso a segunda hipótese seja válida, haveria muita variação na ortografia e somente a publicação de instrumentos linguísticos (vocabulários ortográficos ou dicionários) favoreceria a fixação de um padrão.

No que diz respeito ao sistema consonantal, as três línguas compartilham quase todos os elementos, graças à origem influenciada pelo superstrato português e à presença ostensiva dos colonizadores portugueses nos anos iniciais do século XVI, período de formação destas línguas. O fato de o Angolar ser uma língua de quilombo (Ferraz 1979, Maurer 1995, cf. Arends 1995) e de ter sofrido uma influência menor do superstrato português é visível em seu quadro consonantal, diferenciado em relação ao São-tomense e ao Principense. Portanto, as diferenças consonantais podem ser atribuídas às diversas influências das línguas do substrato e da dinâmica colonial de acesso à língua do superstrato.

A representação das consoantes no ALUSTP distancia as três línguas da ortografia portuguesa. Enquanto esta contém grafemas que representam vários fonemas e vice-versa, aquela pretende manter, ao máximo, o princípio de um fonema, um grafema, muito embora nem sempre consiga.

Cada uma das três línguas de STP possui vinte e uma consoantes na posição de início de sílaba (onset), sendo que as consoantes interdentais (*ɾ*, *ɳ*) ocorrem somente no Angolar (grafadas como <dh> e <th>, respectivamente) a oclusiva álveo-palatal sonora (*dʒ*) não ocorre no Principense (grafada como <dj> no São-tomense e no Angolar), a oclusiva velo-labial sonora (*ɡb*) ocorre no São-tomense e no Principense (grafada como <gb>, a oclusiva velo-labial surda (*kɸ*) ocorre somente no Principense (grafada como <kp>) e a oclusiva palatal sonora (*ʎ*) não ocorre no Angolar (grafada como <lh> no São-tomense e no Principense), como pode ser observado na tabela 1.

Na posição de coda, as ocorrências são limitadas nas três línguas: no Principense e no Angolar, somente o grafema nasal, representado por <m> ou <n>, ocupará esta posição; no São-tomense, tem-se o grafema nasal, como em <mpon> e a fricativa palatal surda <x>, como em <mlax.ka> 'máscara'.



Fonema	Grafema	São-tomense	Principense	Angolar
/b/ oclusiva bilabial sonora	<b>	<ô.bô> 'floresta'	<ba.la> 'bala'	<ku.bi> 'cobrir'
/p/ oclusiva bilabial surda	<p>	<o.pe> 'pé'	<pe.pe.lu> 'papel'	<ka.pa.ta.zi> 'capataz'
/m/ oclusiva nasal bilabial sonora	<m>	<mo.su> 'rapaz'	<u.mun.du> 'mundo'	<maa.ku> 'macaco'
/d/ oclusiva alveolar sonora	<d>	<du.mu> 'pisar'	<da.man> 'amigo'	<dê.lu> 'dinheiro'
/t/ oclusiva alveolar surda	<t>	<ta.son> 'sentar'	<tu.ban> 'tubarão'	<tam.ba> 'pescar'
/n/ oclusiva nasal alveolar	<n>	<no.vu> 'novo'	<u.nu> 'nu'	<ne.ni> 'anel'
/dʒ/ africada álveo-palatal sonora	<dj>	<djê.lu> 'dinheiro'		<pan.dji> 'padrinho'
/tʃ/ africada álveo-palatal surda	<tx>	<ki.txi.ba> 'banana prata'	<i.din.txi> 'dente'	<txi.ba> 'banana'
/ð/ oclusiva interdental sonora	<dh>			<dhu.ga> 'atirar'
/θ/ oclusiva interdental surda	<th>			<po.tho> 'cidade'
/v/ fricativa lábio-dental sonora	<v>	<vô.lô> 'zangar'	<vo.zu> 'voz'	<u.vwa> 'nove'
/f/ fricativa lábio-dental surda	<f>	<fla> 'falar'	<sô.fê> 'sofrer'	<fia> 'folha'
/g/ oclusiva velar sonora	<g>	<be.ga> 'barriga'	<gô> 'chorar'	<di.gô> 'fumo'
/k/ oclusiva velar surda	<k>	<fa.ka> 'faca'	<ku.di> 'responder'	<ka.thô> 'cão'
/gb/ oclusiva velo-labial sonora	<gb>	<gbê.gbê> 'espécie de caracol'	<i.gbê> 'corpo'	
/kp/ oclusiva velo-labial surda	<kp>		<i.kpê> 'grão, semente'	
/l/ oclusiva lateral sonora	<l>	<lem.bla> 'lembrar'	<u.la.du> 'lado'	<lon.ga> 'gamela'
/ʎ/ oclusiva palatal sonora	<lh>	<a.lha> 'areia'	<mi.lho> 'melhor'	
/ɲ/ oclusiva nasal palatal	<nh>	<fa.nha> 'farinha'	<gi.nha> 'galinha'	<nga.nha> 'galinha'
/r/ tepe vibrante alveolar	<r>	<ka.ru> 'carro'	<ra.nha> 'rainha'	<ku.ru> 'cru'
/z/ fricativa alveolar sonora	<z>	<o.ze> 'hoje'	<foo.za> 'ferrugem'	<za.zi> 'trovão'
/s/ fricativa alveolar surda	<s>	<se.la> 'cheirar'	<po.san> 'cidade'	<sin.gô> 'pescoço'
/ʒ/ fricativa palatal sonora	<j>	<ja> 'dia'	<ji.ne.la> 'janela'	<ji> 'fazer'
/ʃ/ fricativa palatal surda	<x>	<na.xi> 'ainda não'	<xin.ku> 'cinco'	<xi.ri.ga> 'mentira'

Tabela 1 – Fonemas e grafemas na posição do onset silábico.

No que diz respeito à variação, o ALUSTP define como 'grafias elegíveis' todas as variantes, certamente no intuito de não se posicionar sobre matéria tão sensível. Este tema é complexo, pois, uma vez que toda língua possui variantes (no Português, por exemplo, há muita variação fonética, como podemos ver nos pares lei[t.j]e e lei[t]e, porém, ambos são

grafados como **leite**) o papel da ortografia, é muitas vezes, neutralizar essa variação, sob pena de tornar a aplicação da ortografia inviável ou complexa demais para o aprendiz (cf. Cagliari 1994). Ainda não há como medir a influência que o não estabelecimento de um padrão para as variantes terá nas ortografias de cada língua, porém, esta é uma questão que deverá ser abordada no futuro. No anteprojeto do ALUSTP, são mencionadas algumas variantes fonéticas consonantais. No São-tomense (9), há uma variação previsível entre a consoante africada álveo-palatal e a fricativa palatal sonora, em começo de palavra. O fato de a variação ser previsível, torna a opcionalidade ortográfica desnecessária. Os demais casos de variação, ambos em começo de palavra, ocorrem em situações de consoantes complexas. Há variação entre a fricativa alveolar surda e a palatal sonora, como primeiro elemento de onset complexo, embora dados (ver Araujo 2007) sugiram uma preferência pela fricativa palatal surda, seria necessária uma investigação sociolinguística sobre este ponto. As consoantes pré-nasalizadas, em começo de palavra, variam com suas contrapartes orais. Portanto, em São-tomense, a variação está limitada ao início de palavra, uma posição na qual comumente consoantes complexas variam com elementos consonantais semelhantes.

- (9) São-tomense: a. Africada álveo-palatal sonora e fricativa palatal sonora  
 [dʒ] ~ [ʒ]  
 <dj> ~ <j>  
 <dja> ~ <ja> 'dia'
- b. Fricativa alveolar surda e fricativa palatal surda, como primeiro elemento de onset complexo  
 [s] ~ [ʃ]  
 <s> ~ <x>  
 <skola> ~ <xkola> 'escola'
- c. Consoantes pré-nasalizadas em início de palavra  
 [nɔ̃g] ~ [g], [mɔ̃p] ~ [p], [nɔ̃d] ~ [d] etc.  
 <ng> ~ <g>, <mp> ~ <p>, <nd> ~ <d> etc.  
 <ngembu> ~ <gembu> 'morcego'

No Principense, há uma variação entre consoantes complexas, pré-nasalizadas e consoantes velo-labiais, e os segundos elementos destas consoantes. Diferentemente do São-tomense, a variação não se limita ao início de palavra. Além disso, a variação nas consoantes pré-nasalizadas envolve a epêntese da vogal [i], ao invés da eliminação da parte nasal inicial do elemento complexo.

- (10) Principense a. Consoantes velo-labiais e labiais  
 [kɔ̃p] ~ [p], [gɔ̃b] ~ [b]  
 <kp> ~ <p>, <gb> ~ <b>  
 <ukpaku> ~ <upaku> 'pele'  
 <igbê> ~ <ibê> 'corpo'

- b. Consoantes pré-nasalizadas, em início de palavra, variam com vogal [i] nasalizada.  
 [n̩d] ~ [ind]; [m̩b] ~ [b]  
 <nd> ~ <ind>; <mb> ~ <imb>  
 <ndala> ~ <indala> ‘folha de palmeira’;  
 <mbasa> ~ <imbasa> ‘costela’

O quadro de variação mais complexo é encontrado no Angolar. Nesta língua, ocorre variação tanto no início como no meio da palavra. A variação ocorre com as consoantes alveolares sonoras orais, velares (sonora e surda), interdentalis (sonora e surda) e fricativas alveolares (sonora e surda). O Angolar também apresenta variação entre as consoantes pré-nasalizadas e as consoantes oclusivas (parte oral da pré-nasalizada) no começo de palavra. Contudo, o Angolar tem uma particularidade: nesta mesma posição, há exemplos de variação entre a pré-nasalizada e a fase nasal consonantal, ou seja, de forma imprevista, o elemento oclusivo oral é apagado.

- (11) Angolar
- a. Alveolares sonoras (orais)  
 [d] ~ [r] ~ [l]  
 <d> ~ <r> ~ <l>  
 <dêlu> ~ <rêlu> ~ <lêlu> ‘dinheiro’
  - b. Velares  
 [k] ~ [g]  
 <k> ~ <g>  
 <vutuka> ~ <vutuga>
  - c. Interdentais e fricativas alveolares  
 [ð] ~ [z], [θ] ~ [s]  
 <dh> ~ <z>, <th> ~ <s>  
 <dhuga> ~ <zuga> ‘atirar’  
 <thun> ~ <sun> ‘senhor’
  - d. Consoantes pré-nasalizadas em início de palavra  
 [m̩b] ~ [b], [m̩m] ~ [m], [m̩p] ~ [p] etc.  
 <mb> ~ <b>, <mp> ~ <p>, <mb> ~ <m> etc.  
 <mpêzi> ~ <mêzi> ‘lua, mês’  
 <mbambu> ~ <bambu> ‘bambu’

Como representar estes tipos de variação na escrita? Há, basicamente, quatro opções: eleger a variante mais complexa ou a mais comum (estatisticamente) como a básica, escolher aleatoriamente ou ainda representar todas as variantes. No primeiro caso, a consoante pré-nasalizada, por exemplo, é representada com os grafemas nasais e orais. Assim, a variação ocorre com o apagamento de um dos elementos, tornado o grafema menos complexo. Portanto, a ortografia torna-se intuitiva para o usuário. No segundo caso, a escolha do grafema se dá por um critério estatístico, porém é necessário avaliar se há tendências de mudança em curso. No terceiro caso, uma escolha aleatória (que muitas vezes envolve escolhas baseadas no prestígio) promove uma determinante variante considerada mais culta ou mais ‘correta’. Por fim, representar todas as variantes tem a

desvantagem de tornar a aplicação da ortografia inviável ou complexa demais para o usuário, pois, sem o auxílio da escolarização na língua ou de instrumentos linguísticos normativos, poderá decidir ora por um ora por outro grafema, gerando inconsistências no sistema.

Araujo e Agostinho (2010) afirmam que há três tipos de nasalidade em São-tomense e em Principense (não obstante, a nasalidade em Angolar não foi investigada a fundo). O estudo indicou que a nasalidade é bifonêmica, ou seja, a fonte da nasalidade vocálica é oriunda de espalhamento do traço nasal de uma consoante nasal na coda silábica. Posteriormente, processos fonológicos apagam a consoante e somente o traço de nasalidade permanece. Por isso, a duração das vogais nasalizadas em São-tomense e em Principense é maior do que a duração das vogais orais e menor do que a duração das vogais longas do Principense. O segundo tipo de nasalidade é fonético, resultante de um processo opcional de espalhamento de nasalidade da coda da sílaba tônica para a sílaba pré-tônica. Desta forma, o composto /ke + |glandzi/ é realizado foneticamente como [k↔~ŋ|glãdʒi] ‘casa grande’. O terceiro tipo de nasalidade é a existência de consoantes complexas pré-nasalizadas, com uma fase nasal e outra oral<sup>12</sup>. No começo de palavra, ocorre variação entre as consoantes complexas pré-nasalizadas e sua fase oral, ou seja, sua parte simples, isto é, a segunda parte da consoante pré-nasalizada. Destarte, trata-se de um processo fonético de apagamento ou ressilabificação da nasalidade: em São-tomense, ocorre frequentemente o apagamento. Portanto, /mpon/ pode ser realizado como [ᵐpõ] ou [ᵐpõ] ‘pão’. Já no Principense e no Angolar (Maurer 1995: 26), pode ocorrer a inserção de uma vogal epentética no começo da palavra, seguida do espalhamento da nasalidade e ressilabificação: /mbasa/ pode ser realizado como [ᵐbasa] ou como [ĩᵐbasa] ‘costela’. A inserção da vogal epentética neste contexto não ocorre no São-tomense. Além disso, Maurer (1995) e o anteprojeto do ALUSTP mencionam a existência de consoantes nasais silábicas, grafadas com um apóstrofe, quando seguidas por outra consoante nasal, como em <n’na> ‘criança’, <m’me> ‘comer’ e <m’ma> ‘eu como’. Contudo, Maurer não apresenta argumentos para qualificar a consoante nasal como silábica e, dados os exemplos, parece se tratar de uma interpretação para a parte nasal das consoantes complexas pré-nasalizadas. Contudo, uma investigação específica sobre este ponto deverá ser tema de um trabalho futuro.

Além da representação das vogais e das consoantes, o ALUSTP representa as vogais na margem do núcleo silábico ou ‘ditongos’ com os grafemas <w> e <y>, o primeiro para a vogal alta anterior arredondada [u] e o último para a vogal alta anterior não-arredondada [i]. Desta forma, as vogais na margem do núcleo diferem-se das vogais nucleares em núcleo distintos (hiatos) na ortografia. Assim, em São-tomense, <fya> [ᵐf̃≈a] ‘folha’ difere de <fi.a>

<sup>12</sup> O Angolar também possui consoantes complexas pré-nasalizadas.

[fi<sub>h</sub>a] ‘esfriar, fiar’ ou <bwê> [ɲb←e] ‘gado bovino’ distingue-se de <buê> [bu.ɲe] ‘querido(a)’.

A posição do acento ou as diferenças tonais não são marcadas com diacríticos no sistema ortográfico. Além das vogais médias <ê> e <ô>, somente a partícula <ka> (e sua variante de primeira pessoa <ga>) recebem diacrítico contrastivo: <ka> e <ga> para as formas tempo-aspectuais e <ká> e <gá> para os morfemas condicionais, como nos exemplos do São-tomense:

- (12) **Bô ka ba xkola, ê ka ba ke.**  
 ‘Tu vais para a escola e ele vai para casa.’  
**Bô ká ba xkola, ê ká ba ke.**  
 ‘Se fosses para a escola, ele iria para casa.’

O anteprojeto do ALUSTP propõe que os seguintes casos de sandhi sejam gravados com apóstrofe quando ocorrer:

- (13) (Exemplos do São-tomense)
- a. em verbos e pronomes:
- |                      |                |                               |
|----------------------|----------------|-------------------------------|
| <b>de + e</b>        | → <d'e>        | ‘dá-lhe’                      |
| <b>ê lige + inen</b> | → <ê lig'inen> | ‘ele os levantou/levantou-os’ |
- b. em preposições (<di> e <ni>) seguidas por palavra começada por vogal:
- |                         |                   |                                |
|-------------------------|-------------------|--------------------------------|
| <b>zamluku di + iya</b> | → <zamluku d'iya> | ‘inhame da ilha (do Príncipe)’ |
| <b>ni + ôtlo</b>        | → <n'ôtlo>        | ‘em outro/a/s, noutro/a/s’     |

Por fim, emprega-se o hífen quando (a) vocábulos reduplicados têm função semântica distinta da forma não-reduplicada, como em <kôlê> ‘correr’ e <kôlê-kôlê> ‘correr continuamente’; e (b) palavras compostas <bega-kolê> ‘diarréia’. Não se emprega o hífen para formas inerentemente duplicadas, como é o caso dos ideofones: <dana kotokoto> ‘estragar-se por completo’.

## CONSIDERAÇÕES FINAIS

Em resumo, o anteprojeto do ALUSTP propõe um sistema descolado da ortografia portuguesa, com exceção da representação da nasalidade vocálica. O ALUSP marca qualidade da vogal média, em detrimento de uma marcação no sistema prosódico e não marca as variações tonais. Contudo, o sistema se beneficia das sinergias dos quadros fonológicos convergentes das três línguas.

Do ponto de vista de sua coerência interna, a ORUSTP é um avanço em relação às ortografias lusitanizantes. Contudo, há ainda problemas pontuais que poderão ser resolvidos com o uso cotidiano do sistema ortográfico, por uma alteração antes da promulgação do decreto ou, futuramente, via reforma ortográfica. Os principais problemas são a possibilidade de grafias concorrentes para casos de variação fonética e ausência de estudos sobre áreas-chave como os sistemas tonais e acentuais, e a questão da nasalidade.

Desnecessário lembrar da complexidade de um sistema ortográfico e dos problemas de sua implementação. Todavia, após a aprovação do decreto, é fundamental que o governo de São Tomé e Príncipe, via Ministério da Educação e Cultura ou de alguma instituição criada para um fim específico, invista na produção de materiais didáticos para todas as fases da escolarização nacional, promoção de outros instrumentos linguísticos (dicionários, vocabulários ortográficos, gramáticas pedagógicas, literatura etc.) nas três línguas e manutenção do acervo cultural e linguístico do país.

## BIBLIOGRAFIA

- Agostinho, Ana Livia. Em preparação. Gramática pedagógica do principense. Universidade de São Paulo.
- Aguiar, Monalisa dos Reis. 2007. As reformas ortográficas da língua portuguesa: uma análise histórica, lingüística e ideológica. *Filologia e Linguística Portuguesa* 9: 11-26.
- Alcântara, Caustriño; Tjerk Hagemeyer. 2001. *Proposta Ortográfica para o Santome*. Colóquio Internacional sobre as Línguas Nacionais, S. Tomé
- Araujo, Gabriel. 2010. Onsets complexos e consoantes com dupla articulação nas línguas crioulas do Golfo da Guiné. Submetido à revista Estudos Linguísticos.
- Araujo, Gabriel. 2007. Empréstimos recentes do português, variação fonética e a sílaba na língua sãotomense da Ilha de São Tomé. *Pápiá* 17: 55-66.
- Araujo, Gabriel; Ana Livia dos Santos Agostinho. 2010. Fonte da nasalidade no saotomense e no principense. Universidade de São Paulo, inédito.
- Arends, Jacques. 1995. The socio-historical background of creoles. In: Arends, Jacques; Pieter Muysken; Norval Smith. *Pidgins and creoles: an introduction*, p. 15-24. Amsterdam: John Benjamins.
- Auroux, Sylvain. 1992. *A revolução tecnológica da gramatização*. Campinas: EDUNICAMP.
- Barros, João de. 1540 [1971]. *Gramática da língua portuguesa; Cartinha, Gramática, Diálogo em louvor da nossa linguagem e Diálogo da Viciosa Vergonha*. Reprodução facsimilar, leitura, introdução e anotações por Maria Leonor Carvalhão Buescu. Lisboa: Faculdade de Letras.
- Baxter, Alan. 2004. The development of variable NP plural agreement in a restructured African variety of Portuguese. In Geneviève Escure; Armin Schwegler (eds.). *Creoles, contact and language change: Linguistics and social implications*, 97- 126. Amsterdam: John Benjamins.
- Cardoso, Manuela. 2007. *Cabo Verde e S. Tomé e Príncipe: educação e infra-estruturas como factores de desenvolvimento*. Porto: Edições Afrontamento.
- Cagliari, Luiz Carlos. 1994. Algumas Reflexões sobre o Início da Ortografia da Língua Portuguesa. *Cadernos de Estudos Linguísticos* 27: 103-111.
- Camões, Luís de. 1572 [2002]. *Os Lusíadas*. Lisboa: Instituto Camões.
- Diamond, Jared. 1997. *Armas, Germes e Aço*. São Paulo: Record.
- Espírito Santo, Carlos. 1998. *A Coroa do Mar*. Lisboa: Cooperação.
- Espírito Santo, Carlos. 2001. *Enciclopédia Fundamental de São Tomé e Príncipe*. Lisboa: Cooperação.
- Ferraz, Luiz Ivens. 1976. A origem e o desenvolvimento de quatro crioulos portugueses do Golfo da Guiné. *Revista Brasileira de Linguística* 3 (2): 70-76.
- Ferraz, Luiz Ivens. 1979. *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press.

- Figueiredo, Carlos. 2009. A Configuração do SN do Português Reestruturado da Comunidade de Almojarife - São Tomé. *Revista de Crioulos de base lexical portuguesa e espanhola* 1(1): 28-55.
- Fontes, Carlos. 2007. Estudo do léxico do são-tomense com dicionário. Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Gunther, Wilfried. 1973. *Das portugiesische Kreolisch der Ilha do Príncipe*. Marburg am der Lahn: Im Selbstverlag.
- Hagemeijer, Tjerk. 2000. Serial Verb Constructions. Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Hagemeijer, Tjerk. 2009. As Línguas de S. Tomé e Príncipe. *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 1:1, 1-27.
- Holm, John; Sandra Madeira. 2009. À propos des noms de variétés du portugais restructuré em Afrique. In: *Le nom des langues III. Le nom des langues en Afrique sub-saharienne : pratiques, dénominations, catégorisations. Naming Languages in Sub-Saharan Africa: Practices, Names, Categorisations* (sous la direction de C. de Féral), BCILL 124, 2009, p. 109-118. Louvain-la-Neuve: Peeters,.
- Hosken, Martin. 2003. Creating an Orthographic Description. Disponível em [http://scripts.sil.org/cms/scripts/render\\_download.php?site\\_id=nrsi&format=file&media\\_id=OrthographyDescPaper&filename=OrthographyDescPaper.pdf](http://scripts.sil.org/cms/scripts/render_download.php?site_id=nrsi&format=file&media_id=OrthographyDescPaper&filename=OrthographyDescPaper.pdf)
- INE/Instituto Nacional de Estatísticas. 2001. *III Recenseamento Geral da População e da Habitação: Características Educacional da População*. São Tomé: INE.
- Lopes, Norma; Alan Baxter. 2010. A concordância verbal variável no português dos tongas. Submetido à revista *PAPIA*.
- Lorenzino, Gerardo. 2000. *The Angolar Creole Portuguese of São Tomé: its grammar and sociolinguistic history*. Munique: LINCOM Europa.
- Maurer, Philippe. 1995. *L'Angolar: Un créole afroportugais parlé à São Tomé; Notes de grammaire, textes, vocabulaires*. Hamburg: Helmut Buske Verlag.
- Maurer, Philippe. 2009. *Principense*. London: Battlebridge.
- Negreiros, António de Almada. 1895. *História Ethnográfica da Ilha de São Tomé*. Lisboa.
- Oliveira, Fernão de. 1536 [1975]. *A gramática da linguagem portuguesa*. Lisboa: Imprensa Nacional/Casa da Moeda.
- Pontífice, João; Castrino Alcântara; Beatriz de Castro Afonso; Tjerk Hagemeijer; Philippe Maurer. 2010. Alfabeto Unificado para a Escrita das Línguas Nativas de S. Tomé e Príncipe (ALUSTP). Inédito.
- Ribeiro, Manuel Ferreiro. 1888. Dialecto da ilha do Príncipe. MS número 11.23.12, Schuchardt Archive da Universidade de Graz, Áustria.
- Rougé, Jean-Louis. 1992. *Les langues des Tonga*. d'Andrade, Ernesto; Alain Kihm (org.) In *Actas do Colóquio sobre Crioulos de base lexical portuguesa*, 171-76. Lisboa: Colibri.
- Rougé, Jean-Louis. 2004. *Dictionnaire étymologique des créoles portugais d'Afrique*. Paris: Karthala.
- Rougé, Jean-Louis & Emmanuel Schang. (2010). The origin of the liquid consonant in Saotomense Creole. Inédito, Université d'Orléans/CORAL.
- Martiniuk, Nathaschka. 2010. Os clíticos pronominais no português vernacular são-tomense. In: 17 SIICUSP/Simpósio Internacional de Iniciação Científica da USP, 2009, São Paulo.
- Seibert, Gerhard. 2007. Angolares of São Tomé island. In Philip J. Havik; Malyn Newitt (eds.) *Creole societies in the Portuguese colonial empire*, 105-126. Bristol: Bristol University Press.
- Valkhoff, Marius F. 1966. *Studies in Portuguese and Creole*. Johannesburg: Witwatersrand University Press.