

LUDViC: regras de transcrição

Maio 2021

Fernanda Pratas & Raïssa Gillier

1. Disposições Gerais

1.0. Introdução

Adotamos genericamente o Alfabeto Cabo-verdiano (antigo ALUPEC), mas com sugestões de adaptações necessárias para respeitar os traços morfofonológicos que nos pareçam relevantes em cada variedade, de acordo com os princípios específicos enunciados em “Regras de transcrição”, em 2. Isto implica alguns passos de regularização interna, recorrendo por vezes ao critério etimológico.

Depois de alguma investigação (Decker 2014; Sebba 2007), ficou entendido que rejeitar todas as normas ortográficas do português não pode ser um objetivo em si mesmo.

Sendo uma grande parte do léxico do caboverdiano herdado do português, é natural que algumas das regras da ortografia do português sejam transferidas para a ortografia do caboverdiano – desde que:

(i) cumpram aqui uma função importante e que, cumulativamente,

(ii) não sejam tão desagradáveis para os falantes que venham a causar a rejeição da ortografia por inteiro (ou seja, a eficácia linguística de um sistema de escrita é relevante, mas deve-se atender também a outros critérios que os falantes usam para adotar ou rejeitar uma grafia – como, precisamente, neste caso, ser demasiado próxima da língua do colonizador).

No que respeita à relação português-caboverdiano, há ainda a questão de os falantes de caboverdiano escolarizados em Cabo Verde ou em Portugal terem sido alfabetizados em português, pelo que as regras ortográficas desta língua já lhes são familiares. Assim, uma parte delas podem ser usadas onde são necessárias mas mantendo a ortografia no geral a uma confortável distância do português – desde que com critérios científicos claros. A este respeito temos por exemplo a adoção no Alfabeto Cabo-verdiano, que aqui também seguimos, da letra <k> ou dos dígrafos <dj> e <tx>, ambos os casos justificados pela escolha legítima de um sistema mais de tipo fonémico, que a ortografia do português não aplica.

Assim, em diversos casos podem intervir as regras ortográficas do português – sempre respeitando o princípio geral de um sistema fonémico – para resolver o problema das diferenças entre variedades (inserção de um <e> / *xevá*, por exemplo). Ou seja, trata-se igualmente de fazer corresponder um grafema a um fonema, mas submetendo esta regra a algumas outras (poucas e simples) que vêm trazer maior clareza / menor complexidade à ortografia.

1.1. Níveis de representação escrita

Os dados orais do LUDViC têm duas camadas de representação escrita:

A. transcrição ortográfica respeita a variação (morfo)fonológica entre variedades, e nalguns casos dentro de cada variedade, de forma a representar alguma da abundante variação fonético-fonológica interna da língua.

Isto passa por assumir na escrita algumas diferenças relevantes, tais como ter o mesmo item lexical a terminar em vogal no Sotavento e em consoante no Barlavento, entre outros.

(1) Exemplos:

- *sedu / sed* ('cedo')
- *frazi / fraz* ('frase')
- *désimu / désim* ('décimo')
- *bilheti / bilhet* ('bilhete')
- *dizenhu / dzenh* ('desenho')

No entanto, é necessário um certo grau de regularização interna, já que não é viável representar todo o tipo de variação. Ou seja, algumas variantes de um mesmo item lexical que ocorrem dentro da mesma variedade, ou até no discurso do mesmo falante, não são contempladas na transcrição.

B. lexema consiste num nível 'funcional' necessário à estrutura interna do corpus, de forma a poder relacionar entre si as variantes ortográficas de um mesmo item lexical.

A variedade linguística selecionada para representar o lexema foi a variedade de Santiago por ser a mais antiga, mas também por ser a mais bem descrita e com maior número de documentos de referência (gramáticas, dicionários, etc.). Assim, adotamos o *Dicionário do Crioulo de Santiago (Cabo Verde)* (Brüser et al 2002) como referência para a atribuição dos lexemas, fazendo os ajustes ortográficos necessários em função das regras descritas em 2.2.

Importa frisar que o lexema não deve ser interpretado como uma forma correspondente a qualquer variedade-padrão (que não existe na língua), mas sim como uma forma que permite agrupar todas as variantes equivalentes. Deste modo, uma pesquisa através do lexema *trabadja* permite obter todas as variantes: *trabaiá*, *trabalha* e *trabadja*.

Exemplos:

(2)

	São Vicente	Santiago
A	<i>trabaiá, trabalha</i>	<i>trabadja, trabalha</i>
B	<i>trabadja</i>	

2. Regras de transcrição

A regra genérica simples de respeitar ao máximo a morfofonologia das diferentes variedades tem de ser desdobrada em regras específicas para casos que, parecendo semelhantes num certo nível (fonológico, por exemplo), apresentam diferenças noutra nível (morfológico). Estes casos são explorados nas subsecções seguintes.

2.1. Fronteira de palavra

A fronteira entre palavras produzidas de forma contígua é marcada graficamente com espaço. Isto implica que certos fenómenos fonológicos nem sempre sejam respeitados na transcrição, como veremos a seguir.

2.1.1. Vogais elididas

Um dos fenómenos mais característicos e frequentes da oralidade consiste na elisão de vogais. Quando estabelecer a fronteira entre palavras contíguas implica acrescentar uma vogal no fim de uma palavra ou no início da seguinte, fazemos isto mesmo.

(3) Exemplos:

- diz-se *nhamigu* / *nhamig*, transcreve-se ***nha*** *amigu* / ***nha*** *amig*
- diz-se *pel*, transcreve-se ***pa*** *el*
- diz-se *djel*, transcreve-se ***djá*** *el*
- diz-se *as*, transcreve-se ***a*** *es*

E uma vez que os potenciais casos de ‘contrações’ existentes na língua ainda precisam de ser detalhadamente estudados, aplicamos esta mesma regra de transcrição aos casos em que temos claramente o significado de dois itens lexicais, como os apresentados em (4).

(4) Exemplos:

- diz-se *nun*, transcreve-se ***na*** *un*
- diz-se ***del***, transcreve-se ***dí*** *el* / ***d*** *el*
- diz-se *dundi* ‘de onde’, transcreve-se ***dí*** *undi*

Há outros casos, no entanto, que podem ser entendidos como novos itens lexicais, indivisíveis, e estes são aqui transcritos como apenas uma palavra, como por exemplo *dundi*, quando significa ‘onde’ e não ‘de onde’ (ver também 2.1.4).

Nota: Considera-se que há vogais elididas apenas quando o início ou o final da palavra dependem efetivamente do contexto, como acontece nos casos em (3) ou (4); ou seja, noutros contextos, as palavras a negrito ocorrem sem nenhum tipo de elisão: *nha*, *pa* e *na* (os contextos em que isto pode ocorrer e aqueles em que está proibido poderão vir a ser, aliás, um interessante objeto de estudo).

Note-se ainda que o clítico de sujeito parece aqui assumir o que se entende tradicionalmente como a forma de clítico de objeto – forma essa que, assim, não depende afinal da marcação morfológica de caso, mas apenas de encostar à esquerda.

LUDViC

2.1.2. Clíticos de objeto

Os clíticos pronominais de objeto são o único caso em que utilizamos o hífen para os demarcar da palavra a que se associam (o verbo). Desta forma deixamos claro que não só temos aqui duas unidades lexicais distintas (o que justifica a separação entre elas), como temos também uma nova palavra fonológica (com consequências, em certos contextos, para a mudança de acento de palavra) e *também* relações específicas ao nível da sintaxe.

(5) Exemplo: diz-se *dam*, ou *dã*, mas transcreve-se *da-m* ('deu-me').

2.1.3. Clíticos de sujeito

Para além do que ficou apontado acima (2.1.1.) sobre os casos em que a forma pronominal de sujeito se encosta à esquerda (a palavras como *djá*, *ma* ou *ki*) – ver exemplos em (3), com formas pronominais de terceira pessoa – temos também diferentes produções do pronome de primeira pessoa do singular noutras circunstâncias.

Em ambas as variedades, a forma base parece ser *N*, mas podemos ter o som /m/ em certos contextos:

- em casos em que o clítico encosta à esquerda, sejam eles os descritos em (5) ou os equivalentes a dois dos contemplados em (3): - diz-se *pã*, transcreve-se *pa N*; diz-se *djã*, transcreve-se *djá N*

- e em casos em que o clítico precede uma oclusiva bilabial (/b/ ou /p/): diz-se *m bai* e *m para*, transcreve-se *N bai* e *N para*

2.1.4. Expressões gramaticalizadas

Em ambas as variedades existem expressões provenientes de expressões multipalavra em português que aqui parecem não ter as diversas palavras intervenientes atuando individualmente de forma produtiva. Nestes casos, em que esteja bem claro que se trata de expressões gramaticalizadas nestas variedades, não respeitamos a fronteira de palavra original do português.

(6) Exemplos:

- *uki /uk* (pronome relativo em *uki é ki /uk é k* 'o que é que')
- *oki /ok* ('[na] hora que')
- *astanta* ('às tantas')
- *diazá* ('dias há')
- *sikalhar* ('se calhar')

2.2. As fricativas [s], [z], [ʃ] e [ʒ]

2.2.1.

A alternância entre as fricativas [s], [z], [ʃ] e [ʒ], que resulta frequentemente do contexto fonológico, não é contemplada. Assim, na transcrição optamos pela letra <s> para



LUDViC

assinalar todas estas ocorrências. Esta decisão deve-se ao facto de, como noutros casos, não ser sempre claro quando temos uma e quando temos outra.

Nota: a variante [s] e [z] parece ser mais frequente em Santiago, enquanto a alternância entre [ʃ] e [ʒ] é típica de São Vicente (também as razões diacrónicas para estas divergência parece ser um bom objeto de estudo; para isso, estes casos serão ‘facilmente’ recuperáveis no corpus através da combinação de <s> com outros sons que estejam tipicamente na sua vizinhança, considerando ainda a fronteira de palavra).

(7) Exemplos:

- *stá, skuru, sforsa, spanhol* (início de palavra)
- *diskoteka, nostalgja, transformod* (meio da palavra)
- *mas, kinhentus, ses, nhas, odja-s / oia-s* (final de palavra)

Nos casos em que a inexistência de qualquer vogal pode ter implicações na leitura, inserimos um <e> (xevá), de forma a favorecer o princípio de correspondência entre grafia e pronúncia da palavra. Assim, a inserção do xevá em *segund* (‘segundo’), por exemplo, garante que o <s> tem o valor de [s], ao contrário de *stá*, em que o <s> pode ser pronunciado como [s] ou [ʃ]. Utilizamos a mesma estratégia nos casos de <s> final.

(8) Exemplos:

- *sebí* (‘subir’)
- *semana* (‘semana’)
- *senhor* (‘senhor’)
- *setenta* (‘setenta’)

- *dose* (‘doce’)
- *asese* (‘acesso’)
- *xanse* (‘chance’)

2.2.2. Sequência -sC- interna

Um caso particular diz respeito à sequência interna -sc- do português, em palavras como *disciplina*, *nascer*, etc. Apesar de ser pouco frequente ou inexistente nas variedades do Sotavento, esta cadeia fonética parece ser comum nalgumas variedades do Barlavento (pelo menos São Vicente e Santo Antão). Este caso constitui uma exceção ao que foi descrito em 2.2.1. relativamente à não contemplação da alternância contextual entre fricativas. No entanto, por se tratar de um fenómeno fonético de potencial interesse, representamo-lo na transcrição como -xs-.

(9) Exemplos:

- *naxsê* (‘nascer’)
- *dixsiplina* (‘disciplina’)
- *adolexsent* (‘adolescente’)
- *konxsiênsia* (‘consciência’)
- *krexse* (‘crescer’)



LUDViC

2.3. Nasalidade

A nasalidade das vogais é marcada com <n> e a dos ditongos com til. Note-se que é sempre <n> a assinalar a nasalidade das vogais, e não <m>.

(10) Exemplos:

- *manhan* ('manhã')
- *nãu* ('não')
- *run* ('rum') (cf. com *rum* 'rumo')
- *asin* ('assim')
- *tanben* ('também')

Em São Vicente, há sílabas em final de palavra, ou em monossílabos, que têm a consoante nasal [n] como coda. Neste caso, para desambiguar, assinalamos este fenómeno com um <e> final, para ficar claro que não se trata de uma sílaba terminada em vogal nasal.

(11) Exemplo:

- *one* ('ano'), por oposição a *mon* ('mão')
- *alune* ('aluno'), por oposição a *algun*

Um caso específico diz respeito às consoantes 'pré-nasalizadas', em que uma consoante nasal parece constituir o núcleo de uma sílaba em início de palavra e tem a propriedade fonética de a sua nasalidade preceder a sua articulação oral (cf. definição diferente em Lang 2014: 151, segundo a qual o fenómeno das pré-nasalizadas diz respeito a "palavras que começam por nexos fonéticos constituídos por uma consoante oral precedida de uma consoante nasal homorgânica"). Quando se trata de palavras oriundas de outras línguas que não o português, seguimos a regra geral das línguas africanas em que este fenómeno também existe (e a regra de Lang em todos os seus trabalhos), representando a referida consoante em início de palavra, sem vogal inicial (12a). Quando se trata de palavras herdadas do português, não contemplamos este fenómeno, e por isso representamos uma vogal anterior, normalmente <i> ou <e> (12b). Ou seja, seguimos um critério etimológico para todos os casos.

(12) Exemplos:

- a. - *nbonji* ('tipo de fava')
- b. - *inkontru* ('encontro'); *enjinneru* ('engenheiro')

2.4. R duplo

As palavras cujo étimo do português contenham <r> duplo transcrevem-se também com dois <r>, independentemente de serem produzidas foneticamente como [r] ou [R]. Esta regra justifica-se pelo facto de se tratar de dois fonemas distintos, /r/ e /R/, cuja representação ortográfica é nalguns casos relevante para diferenciar palavras. Além disso, estes fonemas estão sujeitos a bastante variação intra-falante, e então optámos por este critério como forma de harmonizar as transcrições.



(13) Exemplos:

- *karu / karru* ('caro' / 'carro')
- *kor / korr* ('cor' / 'carro')

2.5. Acentuação

Para a nossa escolha do acento gráfico/'accent' - agudo ou circunflexo, conforme a abertura da vogal em questão - de uma palavra fonológica (incluindo-se aqui as palavras fonológicas formadas por forma verbal + clítico de objeto) consideramos as seguintes tendências observadas: o acento de palavra/'stress' da língua cai tendencialmente

- na penúltima sílaba se a palavra termina em vogal oral
- na última sílaba se a palavra termina em consoante, ditongo ou vogal nasal

Ou seja, considerando as codas consonânticas e as semivogais como segmentos moraicais, podemos dizer que o acento de palavra cai na penúltima mora (Pratas & Salanova 2005).

Decidimos por isto que colocamos acento gráfico:

- nas palavras que não estão de acordo com esta tendência
- nas palavras em que esta consideração é menos relevante por:
 - serem monossílabos
 - conterem um hiato, que desambiguamos face aos ditongos

Palavras lexicais com mais do que uma sílaba

1. a palavra termina em **vogal oral**:

- acento de palavra/'stress' na penúltima sílaba não leva acento gráfico, a não ser para assinalar os hiatos

(14) Exemplos:

- *rotxa* ('rocha')
- *padrinhu* ('padrinho')
- *sabi* ('bom')
- *saída* ('saída')

- acento de palavra/'stress' em qualquer outra sílaba leva o acento gráfico adequado

(15) Exemplos:

- *síviku* ('cívico')
- *klínika* ('clínica')
- *prósimu* ('próximo')
- *fazê* ('fazer') (a maioria dos verbos em São Vicente são acentuados na última sílaba)

2. a palavra termina em **consoante, ditongo (oral ou nasal) ou vogal nasal** (vogal+ <n>):

- acento de palavra/'stress' na última sílaba não leva acento gráfico, a não ser nos hiatos

LUDViC

(16) Exemplos:

- *papel* ('papel')
- *refrijerant* ('refrigerante')
- *kakau* ('cacau')
- *sabon* ('sabão')
- *saúd* ('saúde')
- *país* ('país')

- acento de palavra/'stress' em qualquer outra sílaba leva o acento gráfico adequado

(17) Exemplos:

- *konfortável* ('confortável')
- *paizájen* ('paisagem')
- *infânsia* ('infância')
- *rápid* ('rápido')
- *intérpreti* ('intérprete')

3. a palavra é um monossílabo terminado em:

(a) vogal oral – leva sempre o acento gráfico adequado;

(18) Exemplos:

- *djá* ('já')
- *kmê* ('comer')
- *pé* ('pé')
- *pô* ('pôr')
- *rí* ('rir')
- *gó* ('agora')

(b) consoante, vogal nasal (vogal+ <n>) ou **ditongo oral** – não leva acento gráfico

(19) Exemplos:

- *dos* ('dois')
- *rik* ('rico')
- *fan* ('fã')
- *bon* ('bom')
- *bai* ('ir')
- *poi* ('pôr')

(c) ditongo nasal – leva til

(20) Exemplos:

- *nãu* ('não')
- *mãu* ('mão')
- *pãu* ('pão')

Nota: os verbos que, na sua forma infinitiva, são acentuados graficamente não levam acento quando há um pronome clítico associado a eles.



(21) Exemplos:

- *lê / le-l* ('ler' / 'lê-lo')
- *ivítá / ivita-s* ('evitar' / 'evitá-los')

Palavras funcionais

As palavras funcionais (marcadores temporais, formas pronominais (exceto *bosê*), alguns conectores monossilábicos, etc.) não levam qualquer acento gráfico.

(22) Exemplos: *ki, na, ses, nos, kes, di, sata, ael, mi, ma ...*

Excetuam-se os casos em que o acento gráfico é relevante para desambiguar palavras.

(23) Exemplos:

- *tá* (forma reduzida de *tava*, em SV) vs *ta*
- *ká* (marcador aspetual, de 'acabar de') vs *ka* (negação frásica)

2.6. Estrutura silábica

Existe uma distinção entre a estrutura silábica de Santiago, tipicamente CVCV, e de São Vicente, em que temos o que parecem ser clusters de consoantes que na verdade soam de forma semelhante a muitas formas de produção das palavras correspondentes em português europeu (em que as vogais, apesar de contempladas na ortografia, não parecem ser pronunciadas). Note-se que isto é diferente do caso das 'vogais elididas' ilustrado em 2.1.1, já que aqui nunca existe uma versão em que essa potencial vogal seja pronunciada. Este fenómeno tem levado muitos trabalhos sobre esta variedade (Swolkien 2015, entre outros) a omitir essas vogais na escrita, e é isso que fazemos aqui também. Temos no entanto como referência o Princípio da Sonoridade relativamente a alguns ataques. Este Princípio baseia-se numa escala de sonoridade e estipula que "a sonoridade dos segmentos cresce da margem esquerda até ao núcleo e decresce do núcleo até à margem direita" (Selkirk 1984, apud Freitas 2016). Assim, de acordo com ele são admitidos clusters de consoantes quando estes são de sonoridade crescente nos ataques e de sonoridade decrescente nas codas.

Sonoridade crescente:

oclusiva (p, t, k, b, d, g) < **africada** (tʃ, dʒ) < **fricativa** (f, v, s, ʃ, z, ʒ) < **nasal** (m, n, ŋ)
< **líquida** (r, l, ʎ) < **semivogal** < **vogal**

Note-se que a universalidade deste princípio é controversa (Parker 2012), mas ainda assim temos aqui em consideração as restrições por ele impostas tanto no caso dos ataques como das codas – com uma exceção em ambos os casos (ver abaixo).

(24) Exemplos de ataques complexos permitidos pelo Princípio de Sonoridade:

Início de palavra	Meio de palavra
- <i>platea</i> ('plateia')	- <i>fevrer</i> ('fevereiro')
- <i>flis</i> ('feliz')	- <i>inkline</i> ('inquilino')
- <i>trá</i> ('tirar')	- <i>intresá</i> ('interessar')
- <i>tmá</i> ('tomar')	- <i>panflet</i> ('panfleto')
- <i>kmê</i> ('comer')	- <i>refrijerant</i> ('refrigerante')
- <i>txmá</i> ('chamar')	- <i>simplesment</i> ('simplesmente')
- <i>dzê</i> ('dizer')	
- <i>dzalojod</i> ('desalojado')	
- <i>bzot</i> ('você(s)')	
- <i>kzinha</i> ('cozinha')	
- <i>dxá</i> ('deixar')	
- <i>psoa</i> ('pessoa')	
- <i>dvagarinh</i> ('devagarinho')	

Quando certas sequências de consoantes em ataques silábicos não respeitam a escala de sonoridade, inserimos a vogal <e> (xevá) para 'desfazer' esse grupo consonântico.

(25) Exemplos:

- *deskonfiá* ('desconfiar')
- *beská* ('buscar')
- *fetografia* ('fotografia')
- *djegá* ('jogar')
- *xegá / txegá* ('chegar')
- *enxevalhá* ('enxovalhar')
- *medjor* ('melhor')
- *vezitá* ('visitar')

(26) Exemplos de clusters em coda final permitidas pelo Princípio de Sonoridade:

- <i>agost</i> ('agosto')
- <i>desd</i> ('desde')
- <i>mort</i> ('morte')
- <i>pasaport</i> ('passaporte')
- <i>pert</i> ('perto')
- <i>test</i> ('teste')

Quando certas sequências de consoantes em coda final não respeitam a escala de sonoridade, inserimos a vogal <e> (xevá) para 'desfazer' esse grupo consonântico.

(27) Exemplos:

- *mesme* ('mesmo')
- *moderne* ('moderno')
- *novembre* ('novembro')
- *nunbre* ('número')
- *kuatre* ('quatro')
- *kanpestre* ('campestre')
- *sentre* ('centro')
- *kontakte* ('contacto')

Consideram-se exceções as palavras que contêm um ataque silábico com a sibilante <s> seguida de qualquer consoante com um grau menor de sonoridade, ou uma coda formada por uma consoante com um grau menor de sonoridade seguida da sibilante <s>. Estas sequências consonânticas violam o Princípio da Sonoridade, já que não se verifica a sonoridade crescente no ataque ou a decrescente na coda. No entanto, por se tratar de fenómenos frequentes em caboverdiano, assim como em muitas outras línguas, talvez por envolver o comportamento idiossincrático das sibilantes, decidimos contemplar estes ataques e codas na ortografia, dispensando o xevá.

(28) Exemplos de ataques destes:

- *stájiu* ('estágio')
- *spadja* ('espalhar')
- *sforsa* ('esforçar')
- *skese* ('esquecer')

(29) Exemplos de codas destas – todas as que envolvem plurais:

- *miads* ('meados')
- *gats* ('gatos')
- *karrs* ('carros')

A não representação de vogais não pronunciadas aplica-se mesmo nos casos dos monossílabos, que assim são representados apenas por uma consoante (que pode contar como segmento moraicó, conforme o que foi defendido em Pratas & Salanova):

(30) Exemplos: *d, s, k, ...*

Referências:

Alfabeto Cabo-verdiano (ALUPEC), disponível em: <http://alupec.kauberdi.org/>

Brüser, Martina, André dos Reis Santos, Ekkehard Dengler, Andreas Blum (sob a direção de Jürgen Lang). (2002). Dicionário do crioulo da ilha de Santiago (Cabo Verde), sob a direção de Jürgen Lang. Tübingen: Narr.

Decker, K. (2014). Orthography Development for Creole Languages. [S.l.]: [S.n.].

Freitas, M. J. (2016). A sílaba na gramática do adulto e na aquisição de língua materna. In: Ana Maria Martins & Ernestina Carrilho (eds.), Manual de Linguística Portuguesa. Berlin/Boston: De Gruyter, p. 663-687.

Lang, J. (2014). A origem das consoantes pré-nasalizadas do crioulo de Santiago (Cabo Verde). *Papia* 24(1), p. 149-170.

Parker, S. (ed). (2012). *The Sonority Controversy*. Berlin: Mouton De Gruyter. <https://doi.org/10.1515/9783110261523>

Pratas, Fernanda & Andrés Pablo Salanova. 2005. The allomorphy of Capeverdean object pronominals: a case for late insertion. Paper presented at the 15^è Colloqui de Gramàtica Generativa, Universitat de Barcelona, April 2005.

Sebba, M. (2007). *Spelling and Society: The Culture and Politics of orthography around the world*. Cambridge: Cambridge University Press.

Swolkien, D. (2015). The Cape Verdean Creole of São Vicente: Its genesis and structure. Coimbra: Universidade de Coimbra.