

## LUDViC transcription rules

May 2021

Fernanda Pratas & Raïssa Gillier

### 1. General Provisions

#### 1.0. Introduction

We generally adopt the Alfabetu caboverdiano (Cabo Verdean Alphabet, formerly ALUPEC), but with necessary adaptations suggested to respect the morphophonological traits that seem relevant to each variety, according to the specific principles set out in “Transcription rules”, in 2. This implies some steps of internal regularization, sometimes using the etymological criterion.

After some investigation (Decker 2014; Sebba 2007), we understood that rejecting all Portuguese spelling rules cannot be an objective in itself.

Being a large part of the Cabo Verdean lexicon inherited from Portuguese, it is natural that some of the Portuguese spelling rules are transferred to the Cabo Verdean spelling - provided that:

(i) they fulfil an important function here, and that, cumulatively,

(ii) they are not so uncomfortable for the speakers that they will cause the rejection of the spelling entirely (that is, the linguistic effectiveness of a writing system is relevant, but one must also meet other criteria that the speakers use to adopt or reject a spelling – as, precisely, in this case, the fact that it is too close to the colonizer language).

About the Portuguese-Cabo Verdean relationship, there is also the question that Cabo Verdean speakers with schooling in Cabo Verde or Portugal are literate in Portuguese, so the spelling rules of this language are already familiar to them. Thus, a part of these rules can be used where they are needed but keeping the spelling in general at a comfortable distance from Portuguese – as long as it is supported by clear scientific criteria. In this respect we have, for example, the adoption in the Cabo Verdean Alphabet, of the letter <k> or of the digraphs <dj> and <tx>, both justified by the legitimate choice of a more phonemic type system, which the Portuguese spelling does not apply.

Thus, in several cases, the Portuguese orthographic rules can intervene – always respecting the general principle of a phonemic system – to solve the problem of differences between varieties (insertion of an <e> / schwa, for example). In other words, it is also a matter of matching a grapheme to a phoneme, but subjecting this rule to some others (few and simple) that bring greater clarity / less complexity to the spelling.

# LUDViC

## 1.1. Levels of written representation

The oral data of LUDViC has two layers of written representation:

**A. orthographic transcription** respects the (morpho)phonological variation between varieties, and in some cases within each variety, in order to represent some of the abundant internal phonetic-phonological variation of the language.

This involves assuming some relevant differences in writing, such as having the same lexical item ending in a vowel in the Sotavento varieties and in a consonant in the Barlavento varieties, among others.

(1) Examples:

- *sedu* / *sed* ('early')
- *frazi* / *fraz* ('phrase')
- *désimu* / *désim* ('tenth')
- *bilheti* / *bilhet* ('ticket')
- *dizenhu* / *dzenh* ('drawing')

However, a certain degree of internal regularization is necessary, given that it is not feasible to represent all types of variation. That is, some variants of the same lexical item that occur within the same variety, or even in the speech of the same speaker, are not included in the transcript.

**B. lexeme** consists of a 'functional' level necessary for the internal structure of the corpus, in order to relate the orthographic variants of the same lexical item to each other.

The linguistic variety selected to represent the lexeme was the Santiago variety because it is the oldest, but also because it is the one best described and with the greatest number of reference documents (grammars, dictionaries, etc.). Thus, we adopted the *Dicionário do Crioulo de Santiago (Cabo Verde)* (Brüser et al 2002) as a reference for the attribution of lexemes, making the necessary spelling adjustments according to the rules described in 2.2.

It should be noted that the lexeme must not be interpreted as a form corresponding to any standard variety (which does not exist in the language), but as a form that allows grouping all the equivalent variants. In this way, a search using the lexeme *trabadja* allows to obtain all the variants: *trabaiá*, *trabalha* and *trabadja*.

Examples:

(2)

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

# LUDViC

## 2. Transcription rules

The simple general rule of respecting the morphophonology of the different varieties as much as possible must be broken down into specific rules for cases that, looking similar at a certain level (phonological, for example), present differences at another level (morphological). These cases are explored in the following subsections.

### 2.1. Word boundary

The boundary between words produced in a row is marked graphically with a space. This implies that certain phonological phenomena are not always respected in the transcription, as we will see below.

#### 2.1.1. Deleted vowels

One of the most characteristic and frequent phenomena of orality concerns vowel deletion. In cases where establishing the boundary between contiguous words implies adding a vowel at the end of one word or at the beginning of the next, we do this.

(3) Examples:

- we say *nhamigu* / *nhamig*, but transcribe ***nha*** *amigu* / ***nha*** *amig*
- we say *pel*, but transcribe ***pa*** *el*
- we say *djel*, but transcribe ***djá*** *el*
- we say *as*, but transcribe *a* ***es***

Because the potential ‘contractions’ in the language still need to be thoroughly studied, we also apply this same transcription rule to the clear cases where we have the meaning of two lexical items, like the ones exemplified in (4).

(4) Examples:

- we say *nun*, but transcribe ***na*** *un*
- we say *del*, but transcribe ***di*** *el*
- we say *dundi* ‘from where’, but transcribe ***di*** *undi*

There are some cases, however, which can be understood as a new, indivisible, lexical item, and these cases we transcribe them as an only word, like when *dundi* means ‘where’ and not ‘from where’ (see also 2.1.4).

Note: we consider that there are deleted vowels only when the beginning or the end of the word effectively depends on the context, as it happens in the cases in (3) and (4); that is, in other contexts, the words in bold occur without any type of elision: *nha*, *pa* and *na* (the contexts in which this can occur and those in which it is prohibited may, in fact, become an interesting object of study).

It should also be noted that the subject clitic here seems to assume what is traditionally understood as the object clitic form – a form that, therefore, does not depend, after all, on any morphological case marking, but only on its leaning to the left.



# LUDViC

## 2.1.2. Object Clitics

The pronominal object clitics are the only case in which we use the hyphen to demarcate them from the word to which they associate to (the verb). With this we make it clear that not only do we have two distinct lexical units here (which justifies the separation between them), but we also have a new phonological word (with consequences, in certain contexts, for changing the word stress) and also specific relationships at the level of syntax.

(5) Example: we say *dam*, or *dã*, but transcribe *da-m* ('gave me').

## 2.1.3. Subject Clitics

In addition to what was pointed out above (2.1.1.) about the cases in which the subject pronominal form leans to the left (onto words like *djá*, *ma* or *ki*) – see examples in (3), with third-person pronominal forms – we also have different productions of the first-person singular pronoun in other circumstances.

In both varieties, the base form appears to be *N*, but we can have the sound /m/ in certain contexts:

- in cases where the clitic leans to the left, whether they are the ones described in (5) or equivalent to two of those contemplated in (3): we say *pã*, but transcribe *pa N*; we say *djã*, but transcribe *djá N*

- and in cases where the clitic precedes a bilabial stop (/b/ or /p/): we say *m bai* and *m para*, but transcribe *N bai* and *N para*.

## 2.1.4. Grammatical expressions

Both varieties have forms from multi-word expressions in Portuguese that in Caboverdean do not seem to have the various intervening words acting individually in a productive way. In these cases, in which it is quite clear that these are grammatical expressions in these varieties, we do not respect the original word boundary in Portuguese.

(6) Examples:

- *uki* / *uk* (relative pronoun in *uki é ki* / *uk é k* 'what is that')
- *oki* / *ok* ('[at] the time that')
- *astanta* ('late')
- *diazá* ('days ago')
- *sikalhar* ('maybe')

## 2.2. The fricatives [s], [z], [ʃ] and [ʒ]

### 2.2.1.

The alternation between the fricatives [s], [z], [ʃ] and [ʒ], which often results from the phonological context, is not considered. Thus, in the transcription, we choose the letter



# LUDViC

<s> to mark all these occurrences. This decision is due to the fact that, as in other cases, it is not always clear when we have one and when we have the other.

Note: the variant [s] and [z] seems to be more frequent in Santiago, while the alternation between [j] and [ʒ] is typical of São Vicente (also the diachronic reasons for these divergences seem to be a good object of study; for this, these cases will be ‘easily’ recoverable in the corpus by combining <s> with other sounds that are typically in its neighbourhood, also considering the word boundary).

(7) Examples:

- *stá, skuru, sforsa, spanhol* (beginning of word)
- *diskoteka, nostalgja, transformod* (middle of the word)
- *mas, kinhentus, ses, nhas, odja-s / oia-s* (end of word)

Whenever the absence of any vowel may have implications for the reading of the word, we insert a <e> (schwa), in order to favor the principle of correspondence between spelling and pronunciation of the word. Thus, inserting the schwa in *segund* (‘second’), for example, guarantees that the <s> has the value of [s], as opposed to *stá*, where the <s> can be pronounced as [s] or [ʃ]. We use the same strategy in cases of final <s>.

(8) Examples:

- *sebí* (‘going up’)
- *semana* (‘week’)
- *senhor* (‘sir’)
- *setenta* (‘seventy’)
  
- *dose* (‘sweet’)
- *asese* (‘access’)
- *xanse* (‘chance’)

## 2.2.2. Internal -sC- sequence

A particular case concerns the internal sequence -sC- of Portuguese, in words such as *disciplina, nascer*, etc. Despite being infrequent or non-existent in the Sotavento varieties, this phonetic chain seems to be common in some varieties of the Barlavento (at least São Vicente and Santo Antão). This case is an exception to what was described in 2.2.1. regarding the non-contemplation of the contextual alternation between fricatives. However, because this is a phonetic phenomenon of potential interest, we represent it in the transcription as -xs-.

(9) Examples:

- *naxsê* (‘be born’)
- *dixsiplina* (‘discipline’)
- *adolexsent* (‘teenager’)
- *konxsiênsia* (‘conscience’)
- *krexsê* (‘grow’)



# LUDViC

## 2.3. Nasality

The nasality of vowels is marked with <n> and that of diphthongs with tilde. Note that it is always <n> to indicate the nasality of vowels, not <m>.

(10) Examples:

- *manhan* ('morning')
- *nãu* ('no')
- *run* ('rum') (cf. with *rum* 'course/route')
- *asin* ('so')
- *tanben* ('too')

In São Vicente, there are syllables at the end of the word, or in monosyllables, which have the nasal consonant [n] as a coda. In this case, to disambiguate, we mark this phenomenon with a final <e>, to make it clear that it is not a syllable ending in a nasal vowel.

(11) Example:

- *one* ('year'), as opposed to *mon* ('hand')
- *alune* ('student'), as opposed to *algun* ('some')

A specific case concerns 'prenasalized' consonants, in which a nasal consonant seems to constitute the nucleus of a syllable at the beginning of a word and has the phonetic property that its nasality precedes its oral articulation (cf. Lang different definition 2014: 151, according to whom the phenomenon of prenasalization relates to "words that begin with phonetic nexuses consisting of an oral consonant preceded by a homorganic nasal consonant"; our translation). When it comes to words from languages other than Portuguese, we follow the general rule of African languages in which this phenomenon also exists (and Lang's rule in all his works), representing the referred consonant at the beginning of a word, without an initial vowel (12a). When it comes to words inherited from Portuguese, we do not consider this phenomenon, and therefore we represent a previous vowel, usually <i> or <e> (12b). In other words, we follow an etymological criterion in all cases.

(12) Examples:

- a. - *nbonji* ('type of broad bean')
- b. - *inkontru* ('meeting'); *enjinheru* ('engineer')

## 2.4. Double R

The words whose etymology from Portuguese contains double <r> are also transcribed in Caboverdean with two <r>, regardless of whether they are phonetically produced as [r] or [R]. This rule is justified by the fact that these are two distinct phonemes, /r/ and /R/, whose orthographic representation is often relevant to differentiate the words. Moreover, these phonemes are subject to a lot of intraspeaker variation, and so we resorted to this criterion to harmonize the transcriptions.



(13) Examples:

- *karu / karru* ('expensive' / 'car')
- *kor / korr* ('color' / 'car')

## 2.5. Accents

For our choice of the graphic accent – acute or circumflex, depending on the opening of the vowel in question – of a phonological word (including here the phonological words formed by verbal + object clitic), we consider the following trends observed: the word stress of the language tends to fall

- a. on the penultimate syllable if the word ends in an oral vowel
- b. on the last syllable if the word ends in a consonant, a diphthong or a nasal vowel

In other words, considering consonant codas and semivowels as moraic segments, we can say that the word accent falls on the penultimate mora (Pratas & Salanova 2005).

For this reason we decided that we put a graphic accent:

- i. in words that are not in line with this trend
- ii. in words where this consideration is less relevant because:
  - a. they are monosyllables
  - b. they contain a hiatus, which we distinguish from diphthongs

### Lexical words with more than one syllable

#### 1. the word ends in an **oral vowel**:

- word stress on the penultimate syllable takes no graphic accent, except for hiatuses

(14) Examples:

- *rotxa* ('rock')
- *padrinhu* ('godfather')
- *sabi* ('good')
- *saída* ('exit')

- word stress on any other syllable takes the appropriate graphic accent

(15) Examples:

- *síviku* ('civic')
- *klínika* ('clinic')
- *prósimu* ('next')
- *fazê* ('do') (most verbs in São Vicente are stressed on the last syllable)

#### 2. the word ends in **consonant, diphthong (oral or nasal) or nasal vowel** (vowel+ <n>):

- word stress on the last syllable takes no graphic accent, except for hiatuses

# LUDViC

(16) Examples:

- *papel* ('paper')
- *refrijerant* ('soft drink')
- *kakau* ('cocoa')
- *sabon* ('soap')
- *saúd* ('health')
- *país* ('country')

- word stress on any other syllable takes the appropriate graphic accent

(17) Examples:

- *konfortável* ('comfortable')
- *paizájen* ('landscape')
- *infânsia* ('childhood')
- *rápid* ('fast')
- *intérpreti* ('interpreter')

**3. the word is a monosyllable ending in:**

**(a) oral vowel** - always takes the appropriate graphic accent

(18) Examples:

- *djá* ('already')
- *kmê* ('eat')
- *pé* ('foot')
- *pô* ('to put')
- *rí* ('laugh')
- *gó* ('now')

**(b) consonant, nasal vowel** (vowel + <n>) or **oral diphthong** – takes no graphic accent

(19) Examples:

- *dos* ('two')
- *rik* ('rich')
- *fan* ('fan')
- *bon* ('good')
- *bai* ('to go')
- *poi* ('to put')

**(c) nasal diphthong** – takes a tilde

(20) Examples:

- *nãu* ('no')
- *mãu* ('hand')
- *pãu* ('bread')

Note: verbs which, in their infinitive form, take a graphic accent do not take the accent whenever a clitic pronoun is associated to them.





(21) Examples:

- *lê* / *le-l* ('read' / 'read it')
- *ivitá* / *ivita-s* ('avoid' / 'avoid them')

## Functional words

Functional words (temporal morphemes, pronominal forms (except *bosê*), some monosyllabic connectives, etc.) take no graphic accent.

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

The exception to this is when the graphic accent is relevant to disambiguate words.

(23) Examples:

- *tá* (reduced form of *tava*, in São Vicente) vs *ta*
- *ká* (aspectual marker, 'finish') vs *ka* (sentential negation)

## 2.6. Syllabic structure

There is a distinction between the syllabic structure of Santiago, typically CVCV, and the one of São Vicente, in which we have what appear to be clusters of consonants that actually sound similarly to many forms of the corresponding words in European Portuguese (in which the vowels, despite being registered in the spelling, do not seem to be pronounced). Note that this is different from the case of the 'deleted vowels' illustrated in 2.1.1, since in the case here there is never a version in which this potential vowel is pronounced.

This phenomenon has led many works about this variety (Swolkien 2015, among others) to omit these vowels in writing, and that is what we do here as well. However, we have as a reference the Sonority Sequencing Principle regarding some onsets. This principle is based on a sonority scale and stipulates that "the sonority of the segments increases from the left margin to the nucleus and decreases from the nucleus to the right margin" (Selkirk 1984, apud Freitas 2016, among others; our translation). Thus, according to this principle, clusters of consonants are allowed when they are of increasing loudness in attacks and of decreasing loudness in codas.

### Increasing sonority:

**plosive** (p, t, k, b, d, g) < **affricate** (tʃ, dʒ) < **fricative** (f, v, s, ʃ, z, ʒ) < **nasal** (m, n, ɲ) < **liquid** (r, l, ʎ) < **semivowel** < **vowel**

It should be noted that the universality of this principle is controversial (Parker 2012), but even so we here respect the restrictions imposed by it in both the onsets and codas – with one exception in both cases (see below).

(24) Examples of complex attacks allowed by the Sonority Sequencing Principle:

Begginging of word	Midword
- <i>platea</i> ('audience')	- <i>inkline</i> ('tenant')
- <i>flis</i> ('happy')	- <i>intresá</i> ('interest')
- <i>trá</i> ('take out')	- <i>sinplesment</i> ('simply')
- <i>tmá</i> ('take')	- <i>fevrer</i> ('February')
- <i>kmê</i> ('eat')	- <i>panflet</i> ('pamphlet')
- <i>txmá</i> ('call')	- <i>refrijerant</i> ('soft drink')
- <i>dzê</i> ('say/tell')	
- <i>dzalojod</i> ('homeless')	
- <i>bzot</i> ('you')	
- <i>kzinha</i> ('kitchen')	
- <i>dxá</i> ('let/leave')	
- <i>psoa</i> ('person')	
- <i>dvagarinh</i> ('slowly')	

When certain sequences of consonants in syllabic onsets are not in accordance with the scale of sonority, we insert the vowel <e> (schwa) to 'undo' this consonant cluster.

(25) Examples:

- *deskonfiá* ('suspicious')
- *beská* ('search')
- *fetografia* ('photography')
- *djegá* ('play')
- *xegá* / *txegá* ('arrive')
- *enxevalhá* ('soil')
- *medjor* ('best')
- *vezitá* ('visit')

(26) Examples of clusters in final codas allowed by the Sonority Sequencing Principle:

- <i>agost</i> ('August')
- <i>desd</i> ('since')
- <i>mort</i> ('death')
- <i>pasaport</i> ('passport')
- <i>pert</i> ('close')
- <i>test</i> ('test')

# LUDViC

When certain sequences of consonants in final codas are not in accordance with that scale of sonority, we insert the vowel <e> (schwa) to ‘undo’ this consonant cluster.

(27) Examples:

- *mesme* (‘same’)
- *nunbre* (‘number’)
- *kuatre* (‘four’)
- *kanpestre* (‘rural’)
- *sentre* (‘center’)
- *kontakte* (‘contact’)

An exception to this are words that contain a syllable onset formed by the sibilant <s> followed by another consonant with a lesser degree of sonority, or a coda formed by a consonant with a lesser degree of sonority followed by the sibilant <s>. These consonant strings violate the Sonority Sequencing Principle, given that there is no rising sonority in the onset or falling sonority in the coda. However, as these are frequent phenomena in Caboverdean, as well as in many other languages, maybe because it involves the idiosyncratic behaviour of sibilants, we decided to contemplate them in the written form, dispensing the schwa.

(28) Examples of onsets:

- *stájiu* (‘stage’)
- *spadja* (‘spread’)
- *sforsa* (‘make an effort’)
- *skese* (‘forget’)

(29) Examples of codas – all that involve plurals:

- *miads* (‘middles’)
- *gats* (‘cats’)
- *karrs* (‘cars’)

The non-representation of non-pronounced vowels applies even in the case of monosyllables, which are thus represented only by a consonant (which can be assumed to count as a moraic segment, given what has been defended in Pratas & Salanova 2005):

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



## References:

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

Brüser, Martina, André dos Reis Santos, Ekkehard Dengler, Andreas Blum (under supervision of Jürgen Lang). (2002). *Dicionário do crioulo da ilha de Santiago (Cabo Verde)*. 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<sup>è</sup> 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.