

LUDViC
ANNOTATION MANUAL
(lexematization and POS)

Fernanda Pratas

Raïssa Gillier

December 2021

Index

1. LEXEMATIZATION	3
2. MORPHOSYNTACTIC ANNOTATION – POS.....	5
ACRONYMS (ACR).....	5
ADJECTIVES (ADJ)	5
ADVERBS (ADV)	6
CONJUNCTIONS (CONJ).....	6
DETERMINERS (DET).....	6
INTERJECTIONS (INTJ).....	7
MORPH (MORPH).....	7
SENTENTIAL NEGATION (NEG)	7
NAMES (N and NPR)	8
NUMERALS (NUM)	8
PUNCTUATION (PNT) and GRAPHIC SIGNS (GR)	8
PROGRESSIVE (PROG).....	9
PREPOSITIONS (PREP)	9
PRONOUNS (PRON).....	10
PERSONAL PRONOUNS (PP1 / PP2).....	10
POSSESSIVE PRONOUNS (PPOSS).....	11
REFLEXIVITY (REFL)	12
TEMPORAL (TEMP).....	12
VERBS (V).....	12
References:.....	15

1. LEXEMATIZATION

Since the orthographic transcription of oral data reflects some aspects of the phonetic-phonological variation between linguistic varieties, the same lexical item can occur in different graphic forms. Thus, the lexeme consists of a 'functional' level that allows grouping and relating distinct variants corresponding to the same lexical item. Lexemes should not, however, be interpreted as forms corresponding to a standard variety, as this does not exist in the language.¹

The selection of the lexeme was based on the Santiago variety, as this is the oldest dialectal variety and also the one best described and with the greatest number of reference documents (grammars, dictionaries, etc.). Thus, the *Dicionário do Crioulo de Santiago (Cabo Verde)* (Brüser et al 2002) was adopted as a reference for the attribution of lexemes, making some orthographic adjustments according to the rules elaborated for data transcription (Cf. LUDViC Transcription Rules).

All words are lexematized in their 'neutral' form, that is, without plural, gender, degree or TMA morphemes. So, the words *kondisõis*, *boa*, *dretinh* and *flaba*, for example, receive the respective lexemes *kondison*, *bon*, *dretu* and *flá*.

GENERAL CONSIDERATIONS

➤ Multiple variants

In the case of words with two (or more) variants registered in the Dictionary (e.g. *li* and *ali*), the lexeme selection criterion is the frequency factor. Thus, the form with the highest number of occurrences in the corpus is the one corresponding to the lexeme.

➤ Verbs

Some verbs, especially those that occur in the leeward varieties, present the alternation of the final vowel -e / -i.

(1) Examples:

bebe / bebi

nase / nasi

rabende / rabendi

¹ This is the reason why the word lexeme was adopted, and not lemma. However, for reasons related to the international established practice in digital corpora and the functioning of the internal system of the TEITOK platform (where the materials are available), the word *lemma* is used in the XML files to name the corresponding attribute of this level of annotation.

In these cases, the lexeme always corresponds to the verbal form ending in *-e*. This decision is related to the fact that this is the form to which the TMA morphemes *-ba*, *-du* and *-da*, characteristic of the leeward varieties, are associated: *bebeba* (**bebiba*); *fazedu* (**fazidu*); *kumeda* (**kumida*).

➤ Suppletive verbal forms

The suppletive verbal forms receive the lexeme corresponding to the infinitive form of Portuguese.

(2) Examples:

tinha LEXEME: *ter* (not *tene*)
sabia LEXEME: *saber* (not *sabe*)
kizer LEXEME: *querer* (not *krê*)

➤ Past Participle versus Adjective

Past participle forms that appear in compound tenses or in passive structures are lexematized as the base form of the corresponding verb (and are tagged as V:PART).

(3) Examples:

<i>Nes últim sink one N ten bod.</i>	LEXEME: <i>bá</i>
<i>Foi undi ki nu foi djuntadu.</i>	LEXEME: <i>djunta</i>
<i>Asves N ta flá sata perdedu txeu tenpu.</i>	LEXEME: <i>perde</i>
<i>Es tá ta pensá k já N tinha perdid.</i>	LEXEME: <i>perde</i>

When a participial form has the syntactic function of an adjective, it is lexematized as an adjective (and receives the tag ADJ).

(4) Examples:

<i>Entãu, nu ten txeu kuzas asin ligadu a alunus (...)</i>	LEXEME: <i>ligadu</i>
<i>Santantãu ten uns son k é mut moda afransezod.</i>	LEXEME: <i>afransezadu</i>
<i>(...) já bu ten kel mastru bistidu na bo kaza lá.</i>	LEXEME: <i>bistidu</i>
<i>(...) k é kel papa o kel fijon o kel batata firvid k let.</i>	LEXEME: <i>ferbedu</i>

2. MORPHOSYNTACTIC ANNOTATION – POS

The tag set used for morphosyntactic annotation was designed considering data specificity. It consists of main morphological categories, to which we have added, when necessary, the subcategories relating to inflection in gender, number, degree or TMA. The two levels of information are separated by a colon (:). In the case of personal and possessive pronouns and suppletive verbal forms, the separation is done with a dot (.), since the 'parts' that form the tag do not correspond to distinct morphemes in the word.

(5) Examples:

N:PL = plural name (e.g. *fidjus*)

V:PART = past participle verb (e.g. *fitxadu*)

PP1.1SG = personal pronoun, 1^a person singular (e.g. *N*)

V.PAST = suppletive verbal form in past (e.g. *era*)

Inflection-related tags follow the order in which the morphemes occur in the word.

(5) Examples:

brazilera = ADJ:F (braziler + a)

kondisons = N:PL (kondison + s)

senhurinha = N:DIM:F (senhur + inh + a)

ACRONYMS (ACR)

Acronyms are tagged as ACR.

(6) Examples:

*Trabadju di skola, pamó mi é profesora di **EBI [ACR]**.*

*Enton el falá má kel senhora lá na institut, lá na **IP [ACR]** ond é k jent ta formá.*

ADJECTIVES (ADJ)

The ADJ tag applies to canonical adjectives, adjective participle forms, and ordinal numerals.

(7) Examples:

*Mininus era mas **pikinoti [ADJ:DIM]***

*(...) es ka ta fazi abuzu na otu, es é só **idukadu [ADJ]**.*

*Ah ántis e era mas **sabi [ADJ]** purkê N ta vendeba otrus venda (...)*

*El tava **kubrid [ADJ]** k un pone (...)*

Alguns part k tava **danifikod [ADJ]**, es metê pilar.
 Pur ise k N dzê-b, asves bo ka ta konsegí dá kel atensãu **individualizod [ADJ]**.
 Grinhasin, un ves bo tava bá eh **terser [ADJ]** one, **kuart [ADJ]** one, **kint [ADJ]**
 one, sest, sétim.
 N studá **kinta [ADJ:F]** i **sesta [ADJ:F]** klase na Krakinha.

ADVERBS (ADV)

The ADV tag applies to all adverbs (note that the *ka* and *ne* morphemes are annotated with the NEG tag as they are not adverbs, but just functional heads marking sentential negation).

(9) Examples:

*Djá N podi fla-u zona e muda **txeu [ADV]**.*
*N ta mora **li [ADV]** na Krus di Piku.*
*Dipos di ki N pasaba **mal [ADV]**, oji N ben pasa **mas [ADV]** razuável.*
*N ta dze-s: “**nãu [ADV]**, N ta bá **dvagarinh [ADV:DIM]**”.*

CONJUNCTIONS (CONJ)

The tag CONJ applies to some conjunctions, as well as to the complementizers *ki / k*, *ma* and *kmá*.

(10) Examples:

*Grinhasin d not, **s [CONJ]** N bá pa Sonsent, N ten med d saí d nos kaza.*
*Sáudi stá mas **o [CONJ]** menus.*
*N ka baba skola, **mas [CONJ]** N ka morri.*
***Nen [CONJ]** maskulinu, **nen [CONJ]** fiminina.*

*Kaminhu é un bokadinhu distanti, N ten **ki [CONJ]** anda txeu.*
*Au tud N ta kontá **k [CONJ]** N tiv des fidj.*
*Nãu, Kabu Verdi, ami, N sata atxa **ma [CONJ]** stá dizenvolvi bastanti.*
*Es ta enxiná ses fidj **kmá [CONJ]** spanhol é inportant.*

DETERMINERS (DET)

The DET tag applies to definite and indefinite articles, quantifiers and demonstrative adjectives.

(11) Examples:

*Fefon tanben **u [DET]** **mesmu [DET]** forma.*

*Grinhasin, bo ten menine k **uns** [DET:PL] konportament!*

*Nu ta uza **txeu** [DET] kartas, nu ta fazi **txeu** [DET] jogus, nu ta kanta **txeu**.*

*N ka ta sai pa anda **ninhun** [DET] kau.*

*Mas a vida a vida kontinua, jent tá na luta **tod** [DET] dia.*

*N ka ta skesê **ese** [DET] koza li.*

*Es txoma **kes** [DET:PL] alunu di ária siênsia i teknolojia i alunus di eletrisidadi.*

INTERJECTIONS (INTJ)

The INTJ tag applies to the words/expressions that express a value of admiration, confirmation, denial, pain, etc..

(12) Examples:

*N mora lá, **prontu** [INT], iá N ten nhas kumadri ku nhas konpadri lá.*

***Nah** [INT], es tá ku ses pai.*

*N ta dzê: “**uaba** [INT], portugés é un língua k N ta konprendê”.*

MORPH (MORPH)

The MORPH tag is used exclusively to annotate the morpheme **a**, which interpretation is still uncertain. The two main lines of analysis discussed in the literature consist in its interpretation as an allomorph of the adverb *djá* or as a grammaticalized morpheme (Swolkien 2015).

(13) Examples:

***A** [MORPH] m tiv na Praia.*

*Mi era só ku kes meninas, k es era jems, **a** [MORPH] bo oiá?*

*El ben ku se vint i tal ane, **a** [MORPH] l reformá li, **a** [MORPH] l bá pa Kab Verd.*

SENTENTIAL NEGATION (NEG)

The NEG tag applies only to the sentential negation markers **ka**, common to all dialectal varieties, and to **ne**, which occurs at least in the Santo Antão and São Vicente varieties. The negative item *nãu* / *nau* is tagged as ADV.

(14) Examples:

*En prinsípiu kabra nu **ka** [NEG] ten, nu ten só bodi.*

*Asin, mi, N **ka** [NEG] ten inimig.*

*Bo ka [NEG] podê ergí d bo lugar.
 N spiá pa un banda, N spiá pa ot, N dzê "ese ne [NEG] é Rotterdam".
 Tud menine el ne [NEG] é igual.
 N spiá kes jent, nen un d es e ne [NEG] era kriol.*

NAMES (N and NPR)

The N tag applies to all names. The proper names (person's names, places, institutions or others) receive the NPR tag.

(15) Examples:

*Mas é poti [N] di barru [N], bandeja [N] di barru [N].
 Kantu N pidiba noiba [N:F] pa N kaza, trinta dia pa N podê dá rasposta.
 N ben Merka k dznov anes [N:PL] d idad.
 Dipos dexa arma sfria un kuzinha [N:DIM], pô leti.*

*É lá mas pa riba, lá pa Água di Gatu [NPR].
 Agó bo ta fazê bo kaza é na Kab Verd [NPR] agó.
 No tá stod ta sperá desizãu d Ministériu d Edukasãu [NPR].*

NUMERALS (NUM)

The NUM tag applies to ordinal numerals, as well as to dates and to hours.

(16) Examples:

*N kaza di dizaseis [NUM] anu di idadi.
 Tres [NUM] irmon, dos [NUM] motx, un [NUM] fema.*

*N bá pa São Tomé na 1955 [NUM], N ben na 66 [NUM].
 Mi, N ta saí d ali 12h30 [NUM] kansod, stresod.*

PUNCTUATION (PNT) and GRAPHIC SIGNS (GR)

Punctuation marks (period, comma, exclamation point, question mark, colon, ellipsis, and semicolon) are tagged as PNT. Other graphic signs (dash, quotation marks, etc.) are tagged as GR.

(17) Examples:

*N ingravida-s li, [PNT] N fika prenha li, [PNT] N bá parí na Praia. [PNT]
 Kum é k N ta dzê? [PNT]
 Soment é k ta uzá "[GR] txuk"[GR].*

PROGRESSIVE (PROG)

The PROG tag applies to the morphemes **tita** (in the windward varieties), **sata**, **ata** and **sta** (in the leeward varieties), markers of the progressive aspect (Pratas 2018a, 2018b).

(18) Examples:

*Mod k só gó k bo **tita [PROG]** djgá na kaza?*

*Grinhasin pai **tita [PROG]** saí, mãi **tita [PROG]** saí, enton es **tita [PROG]** dxá kargu d edukasãu d ses fidj só pa skola.*

*Mas djá N monda, **sata [PROG]** ten un kuzinha ki nu ta vivê na el, aian.*

*Alguns zonas **sata [PROG]** fazedu txeu.*

*I daki a alguns anu N ka sabi mó ki nu **ata [PROG]** fazi.*

*Si N **ata [PROG]** meste algen pa N poi na trabadju, klaru ki N ta poi (...)*

*Es ka **sta [PROG]** paga-bu kel dinheru.*

*(...) e **sta [PROG]** mestê un psoa pa pinta un kuzinhas pra lá pa el ta pô dose.*

PREPOSITIONS (PREP)

The PREP tag applies to all prepositions, regardless of the contexts in which they occur. The same tag is used to annotate the lexical items **komu/kom**, **modi/mod/mó**, **sima/suma** and **moda** when these occur in 'comparative contexts'.

(19) Examples:

*Asves es ta teimá **má [PREP]** mi.*

*Agora no tá feká asin **k [PREP]** grasa d el bai.*

***Na [PREP]** mensájen é só kriol.*

*Mas, **apartir [PREP]** di momentu ki N entra, N ta gosta.*

*(...) pamodi é nos kultura ki nu atxa **desdi [PREP]** tenpus antigus.*

*Es ta vira un melting pot **modi [PREP]** United Nations, nãu? **Mó [PREP]** Nasõis Unidas.*

***Kom [PREP]** es ta dzê, li ten mas oportunidad pa traboi.*

*Maior parti di atividadi propi é agrikultura, **tantu di sekeiru komu [PREP]** di regadiu.*

*Santa Katarina i Sal é diferenti **sima [PREP]** séu ku terra.*

*Normalmenti, mi, **sima [PREP]** N ta pensa, N ta krê ser un pisikólogo (...)*

*Kada ilha ten se sutak, **moda [PREP]** vosê sabê, má nós tud (...)*

PRONOUNS (PRON)

The PRON tag applies to interrogative (*kuzé, kantu, kenha*, etc.), relative (*ki / k*), and indefinite (*isu, istu, nada, ninhun, ningen*, etc.) lexical items.

This label does not apply to personal, possessive and reflex pronouns.

(20) Examples:

N ten konsulta, N ten medikament, kuzé [PRON] mas k N krê?

Nhos é kantu [PRON]?

Bu ka sabi é kenha [PRON]?

Kel ki [PRON] é filha ki [PRON] manda buska mi di Purtugal ten tres.

Mi mas ken [PRON] k tá morá lá ka ten konhesiment.

Grinhasin, eses menine, es ta sin nada [PRON] i ese ajitasãu k es ten na vida...

Ninhun [PRON] ka stá ku mi.

Es ka ten ningen [PRON] pa konpanha-s.

PERSONAL PRONOUNS (PP1 / PP2)

The PP1 tag applies to free personal pronouns and to clitics, and the PP2 tag applies to strong pronouns. Both tags are followed, after a period (.), by the person and number to which the pronoun corresponds. Tables 1 and 2 bring together the free, clitic and strong pronominal forms that occur in the corpus.

POS	Free Forms	Clitic Forms		LEXEME
		Without a left host	With a left host	
PP1.1SG	<i>mi, min</i>	<i>N</i>	<i>m</i>	mi
PP1.2SG	<i>bo</i>	<i>bu</i>	<i>bu, b, u</i>	bo
PP1.2SG	<i>nho</i>	<i>nhu</i>		nho
PP1.2SG	<i>nha</i>	<i>nha</i>		nha
PP1.2SG	<i>bzot</i>			bzot
PP1.2SG	<i>vosê, bosê</i>			vosê
PP1.3SG	<i>el</i>	<i>e</i>	<i>l</i>	el
PP1.1PL	<i>nos</i>	<i>nu, no</i>	<i>nu</i>	nos
PP1.2PL	<i>nhos</i>			nhos
PP1.2PL	<i>vosês, bosês</i>			vosês
PP1.2PL	<i>bzot</i>			bzot
PP1.3PL	<i>es</i>		<i>s</i>	es

Table 1 – Free and Clitic Personal Pronouns (PP1)

POS	Strong Forms	LEXEME
PP2.1SG	<i>ami</i>	ami
PP2.2SG	<i>abo</i>	abo
PP2.2SG	<i>anho</i>	anho
PP2.2SG	<i>anha</i>	anha
PP2.3SG	<i>ael</i>	ael
PP2.1PL	<i>ano, anos</i>	anos
PP2.2PL	<i>anhos</i>	anhos
PP2.3PL	<i>aes</i>	aes

Table 2 – Strong Personal Pronouns (PP2)

POSSESSIVE PRONOUNS (PPOSS)

The PPOSS tag only applies to possessive pronouns. It is always followed, after a period (.), by the person and number to which the pronoun corresponds.

Table 3 gathers the possessive pronouns attested in the corpus.

POS	Pronominal Forms	LEXEME
PPOSS.1SG	<i>nha, mi, mia, minha</i>	nha
PPOSS.1SG	<i>(di) meu</i>	meu
PPOSS.2SG	<i>bo, bu, u</i>	bu
PPOSS.2SG	<i>(di) bosa</i>	bosa
PPOSS.2SG	<i>bzot</i>	bzot
PPOSS.3SG	<i>se, si</i>	se
PPOSS.3SG	<i>(di) sel, (di) seu</i>	seu
PPOSS.1PL	<i>nos</i>	nos
PPOSS.2PL	<i>nhas, mias, mis</i>	nhas
PPOSS.2PL	<i>bos, bus</i>	bus
PPOSS.3PL	<i>ses, sis</i>	ses

Table 3 – Possessive Pronouns (PPOSS)

REFLEXIVITY (REFL)

The REFL tag applies to the reflexive forms. In Caboverdean, reflexivity is conveyed through 'expressions' usually composed of a possessive pronoun followed by the lexical item *kabesa*. This tag, which includes the number and person corresponding to the possessive pronoun, applies to the entire expression.

All forms are lexematized as *si kabesa*.

(21) Examples:

*Bu stá ingana **bu kabesa [REFL.2SG]** i bu ka stá (...)*

*N ta konsidera **nha kabesa [REFL.1SG]** afrikanu, kabuverdianu afrikanu.*

*iá e prega **se kabesa [REFL.3SG]***

TEMPORAL (TEMP)

The TEMP tag applies, for now, to the morphemes *ta*, *tá* and *tava*, markers of the habitual aspect (Pratas 2018a, 2018b).

(22) Examples:

*Enton txeu bes e **ta [TEMP]** bai pa ladu di kuzas mariadu.*

*(...) pamodi iá bira riju, ka **ta [TEMP]** dá pa kumi.*

*Es **tá [TEMP]** da-m lisensa pa bá konplementá kel disciplina.*

*No **tá [TEMP]** betá koltxon na txon, no **tá [TEMP]** detá.*

*Mas, mi, N **tava [TEMP]** priokupod purkê (...)*

*Es **tava [TEMP]** feká li má mi, es **tava [TEMP]** ijda-m, es **tá [TEMP]** ta [TEMP] aprendê.*

VERBS (V)

The V tag applies to the base form of verbs. Most verbs receive this tag, since temporal interpretation in Caboverdean language is achieved by the combination of tense-mode-aspect morphemes (TEMP and PROG tags) with the bare form of the verb, together with the linguistic and extra-linguistic context (Pratas 2018b).

(23) Examples:

*N **ben [V]** pa nha terra.*

*(...) ta **kunsa [V]** **ben [V]** **laba [V]** pa **bisti [V]** pa **bá [V]** skola.*

*Ten kes ki ta **lavora [V]**, ten kes ki ta **simia [V]**, ten kes ki ta **monda [V]** (...)*

➤ V:PAST

Forms containing the past morpheme *-ba*, which occurs predominantly in leeward varieties, are tagged as V:PAST.

(24) Examples:

*Mi, N ka **baba** [V:PAST] skola, mas N ka morri.
Senpri N ta **flaba** [V:PAST] ma N **gostaba** [V:PAST] di **trabadjaba** [V:PAST] asin.
Riba kaza ta **manxeba** [V:PAST] kes areinha finu, pretu, bunitu.*

➤ **V:PART**

The V:PART tag applies to past participles that appear in compound tenses and in the passive voice. The past participle is formed with the postverbal morphemes **-du / -d**.

(25) Examples:

*Pamodi alves ta **fladu** [V:PART] asin.
Asin k no foi **kriod** [V:PART], nãu é?
Es ka ta **fazedu** [V:PART] nada di kes animasãu.*

➤ **V:PART.PAST**

The V:PART.PAST tag applies to participle forms that contain the postverbal morpheme **-da**, which occurs in passive structures that refer to a time previous to the 'simple' passive.

(26) Examples:

*Na kel altura ta **bada** [V:PART.PAST] pa Sãu Tomé era kontratadu (...)
Kel bes ta **flada** [V:PART.PAST] resebedoria.
É kes kuza asin mas ki ta **kunprada** [V:PART.PAST] na loja.*

➤ **V.PRES / V.PAST / V.FUT**

The suppletive verb forms, generally Portuguese forms with inflectional morphemes that are atypical in the Caboverdean verbal system, receive the main tag V, followed by the tag with generic specification of the verb tense: PRES (present), PAST (past) and FUT (future) .

As previously mentioned (point 1 - lexematization), in these cases, the period (.) is used to separate the main category from the subcategory and it is assigned the Portuguese lexeme.

(27) Examples:

*Vida na Sãu Tomé **era** [V.PAST] tantu difísil pamó é fetu skravu.
El **tivi** [V.PAST] lá oitu anu.
Grasas a Deus, N **foi** [V.PAST] d kes numiod.
A bzot imajiná s bo **tiver** [V.FUT] ta pasá na rua, trankuil, bo sintí vontad (...)
Ántis di N ben pa Sal, N ta **vivia** [V.PAST] na Sãu Nicolau ku un pai di fidju.
(...) nha pai, N **sabia** [V.PAST] manera k el tava asotá nos.*

List of POS tags

Meaning	POS tag	
Acronym	ACR	--
Adjective	ADJ	F
		PL
		DIM
Adverb	ADV	DIM
Conjunction	CONJ	--
Determiner	DET	F
		PL
Interjection	INTJ	--
Morph	MORPH	--
Phrasal negation	NEG	--
Name	N	F
		PL
		DIM
Proper name	NPR	--
Numeral	NUM	--
Punctuation	PNT	--
Preposition	PREP	--
Progressive	PROG	--
Pronoun	PRON	PL
Personal pronoun	PP1 / PP2	1SG / PL
		2SG / PL
		3SG / PL
Possessive pronoun	PPOSS	1SG / PL
		2SG / PL
		3SG / PL
Graphic sign	GR	--
Temporal	TEMP	--
Verb	V	PART
		PRES
		PAST
		FUT

References:

Brüser et al. (2002). Dicionário do crioulo da ilha de Santiago (Cabo Verde), sob a direção de Jürgen Lang. Tübingen: Narr.

Lang, Jürgen (ed.) (2014). A Variação geográfica do crioulo caboverdiano. Erlangen: Friedrich-Alexander-Universität University Press.

LUDViC Transcription Rules. Available at: <http://teitok.clul.ul.pt/ludvic/>

Pratas, Fernanda (2018a). Progressive forms and meanings: The curious case of Capeverdean. *Estudos de Lingüística Galega* 10. 103–128. <https://doi.org/10.15304/elg.10.5045>

Pratas, Fernanda (2018b). Tense results from mood and aspect: Temporal meanings in Capeverdean. Paper presented at GREAT, within Going Romance XXXII, Utrecht, December 10th-14th.

Pratas, Fernanda (2021). The expression of temporal meaning in Caboverdean. Berlin: De Gruyter. <https://doi.org/10.1515/9783110626629>

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