Towards a Morphological Analyzer for the Umbundu Language

Alberto Simões
2Ai, School of Technology, IPCA, Barcelos, Portugal
asimoes@ipca.pt

Bernardo Sacanene
Centro de Estudos Humanísticos da Universidade do Minho, Braga, Portugal
besacanene@gmail.com

Álvaro Iriarte
Centro de Estudos Humanísticos da Universidade do Minho, Braga, Portugal
alvaro@ilch.uminho.pt

José João Almeida
Algoritmi, Departamento de Informática, Universidade do Minho, Braga, Portugal
jj@di.uminho.pt

Joaquim Macedo
Algoritmi, Departamento de Informática, Universidade do Minho, Braga, Portugal
macedo@di.uminho.pt

Abstract

In this document we present the first developments on an Umbundu dictionary for a jSpell, a morphological analyzer. Initially some comments are performed regarding the Umbundu language morphology, followed by the discussion on jSpell dictionaries structure and its environment. Last, we describe the Umbundu dictionary bootstrap process and perform some final experiments on its coverage.

2012 ACM Subject Classification Computing methodologies → Language resources

Keywords and phrases Umbundu, Angolan Languages, Morphological Analysis, Spell Checking

Digital Object Identifier 10.4230/OASIcs.SLATE.2020.10

Funding This research was partially funded by Portuguese National funds (PIDDAC), through the FCT – Fundação para a Ciência e Tecnologia and FCT/MCTES under the scope of the projects UIDB/05549/2020 and UIDB/00319/2020. Bernardo Sacanene acknowledges from the Angolan govenment his PhD grant, through INAGBE (Instituto Nacional de Gestão de Bolsas de Estudos).

1 Introduction

A large part of African populations just as many other populations from other parts of the world, live in a situation of diglossia, in the most restricted sense of the term, presented by Ferguson [4], in which communities have two languages at their disposal: one language A (or prestigious language) and a language B (familiar or domestic language).

In these situations, it is normal to produce, sooner or later, a process of linguistic substitution (by direct imposition or as a result of a valuation that the new speaker makes, in terms of usefulness, of the language), starting from an original monolingual situation and going through a provisional bilingual state (in languages A and B) to finally reach a monolingual state (surviving solely the language A) [2].

But linguistic diversity is always a treasure. Each language has a type of relationship with reality and diversity, a guarantee of the system's durability and stability, since it does not destroy resources (in this case, linguistic, cultural, etc.). In the case of Africa, "*the*

9th Symposium on Languages, Applications and Technologies (SLATE 2020).
Editors: Alberto Simões, Pedro Rangel Henriques, and Ricardo Queirós; Article No. 10; pp. 10:1–10:11
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

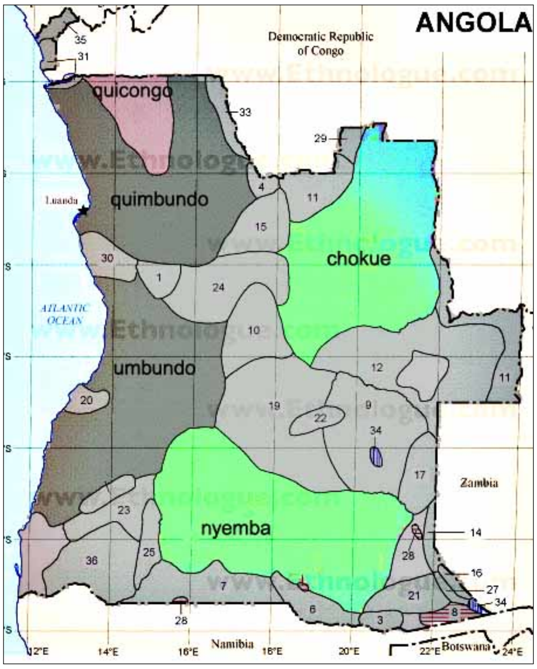

Figure 1 Map of languages in Angola[1].

languages of some peoples which have attained sovereignty are consequently immersed in a process of language substitution as a result of a policy which favours the language of former colonial or imperial powers" [12].

Only linguistic policies can change this process, together with the speaker's appreciation of the language, in terms of necessity or usefulness. That is the reason we must create conditions for people to enjoy their linguistic rights. These conditions include ensuring the use of languages B by different communities and creating conditions for interrelation and dialogue of these less favoured languages with both prestigious languages and formal languages.

Defending the linguistic rights of speakers involves not only linguistic policies that favor the use of community languages in as many contexts as possible, including formal contexts, but also by creating conditions for the interrelation of these languages with others, encouraging the practice of translation between both languages and the connection of these languages with formal languages, with the possibility of being used as an object of study in the Processing of Natural Languages, in corpus linguistics, automatic translation, data mining for the extraction of terms and lexicon for the elaboration of terminological and lexicographic products, etc.

The creation of Natural Language Processing tools for African languages is an urgent and imperative task to defend the wealth that current linguistic diversity signifies.

In this document we explore the development of a first Natural Language Processing (NLP) tool for Umbundu, one of Angola's languages (see Figure 1).

The construction of Umbundu corpora and other Bantu languages corpora (as well as other African languages) will allow access to a set of information (linguistic, cultural, etc.) that can be used for the elaboration of traditional dissemination products (newspapers,

[1] Map obtained from `http://seguindoadiante.blogspot.com/2008/08/torre-de-babel.html` [accessed on 2020-07-03].

school textbooks, dictionaries, grammars, medical records, restaurant menus, tourist guides, etc.). It could also be used for the development of Apps and other resources, which will be available for agents involved in the management and modernization of the cultural, linguistic, tourist, etc. of the African states and the different communities that make them up.

The next section explores the Umbundu language, namely some of the most basic rules of its morphology. Section 3 presents the environment of jSpell, the morphological analyzer used in our developments. Some of the rules included in the Umbundu morphological analyzer are depicted on Section 4. In the last section, we discuss the obtained results.

2 Contextualization of the Umbundu Language

Umbundu is the language of the ovimbundo ethnolinguistic group, belonging to zone R, group 10, R11 [6] and comes from the bantu group. It is spoken in central and Southern of Angola, specifically in Bié, Huambo, Benguela and extends to neighbouring province such as Namibe, northwest Cuando Cubango and northern Huíla.

According to the data from the last census [7] regarding the languages usually spoken at home, in the first place we have Portuguese (71%), and Umbundu as the second language (23%), followed by Kikongo and Kimbundu.

2.1 Orthographic Binormativism

The Umbundu language presents, for a specific linguistic reality, different written forms. As an example, “tch”, “ch”, “c” in “tchina”, “china”, “cina” (thing), “ng̃”, “ñ” in “Ng̃ala”, “Ñala” (God) or “dj” and “j” in “ondjo” and “onjo” (house). This is a common problem for languages before any attempt of standardization, and related to the fact that the first studies about the Umbundu languages where performed by religion persons that used an alphabet from another language with Latin origins, modifying it in order to simulate the sounds that are nonexistent in European languages [3].

The study of Umbundu by persons tied to religious institutions, together with UNESCO recommendations, lead to different ways to represent graphically this language. Therefore, there is a “catholic version and a protestant one,” [8, p. 479] as well as another, adopted by the Institute of National Languages [3], as an attempt for standardization. This orthographic duplicity or even triplicity is a problem, not just for the literacy of the population, but also for the creation of natural language processing tools for these languages.

2.2 A brief introduction to Umbundu

To build a morphological analyzer of the Umbundu language we are studying three grammatical categories: noun, adjective and verb. First and foremost we analyze how the formation of the number and the gender are done, and also the inflexion of verbs.

2.2.1 Nouns

As other bantu languages, Umbundu presents the following characteristics:

1. Nouns are organized in classes defined by prefixes. The classes are combined to distinguish the opposition of the number (singular/plural). Table 1 shows how number affix is done. According to Fernandes and Ntondo [5], the class changeover is done in two ways: replacement of prefixes (exchange of class prefix with that of the class in which it is inserted) and addition of prefixes (the noun inserted is close to the prefix of that class).

Table 1 Noun prefixes. Based on [5].

Umbundu language			**Examples**	
Class	Prefix	Class Information	Umbundu	English
1	omu-, u-, o-	human-related	omunu, ukombe	person, guest
2	oma-, ova-, a-		omanu, akombe	people, guests
3	u-	plants, animals, body parts and other	uti / uta	tree / gun
4	ovi-		oviti / ovita	trees / guns
5	e-	human realities, plants, animals	eka / epela, epito, etimba	hand / baldness, door, body
6	a-, ova-		ovaka / apela, apito, atimba	hands / baldness, doors, bodies
7	oci-	several concepts	ocimunu*, ocitangi	thief, problem
8	ovi -		ovimunu, ovitangi	thieves, problems
9	o-, ∅-	common for animals	ombweti, omoko	stick, knife
10	olo-		olombweti, olomoko	sticks, knives
11	olu-	delimitation	olumapo, olukalo, alupandu	model, opportunity, thanks
12	oke-	several concepts and diminutives	okalenda, okamõla, okambeyi	small tumor, little child, sicky
13	atu-		okatumola, otumbeyi	small tumors, little children, sickies
14	u-		ukolo	rope
15	oku-	verb-nominal	okulota, eloto	to dream, dream
Locative classes: Number variation does not occur for nouns [5, 10].				
16	pa-	surface	okulya ikasi po mesa wacipaka p'osi watumala k'omangu	the food is on the table put down sit in the chair
17	ko-	direction, route	eye wanda k'epya weya k'imbo kosindikile k'onembele eye okasi ko samwa	he went to the farm he came from the village walk him in to the church he it out
18	vu-	inside	eye okasi v'onjo vapiluka v'ocitali wakupukila v'ocikungu	he is at home they danced in the backyeard fell in the hole

*Umbundu allows three different orthographic forms for *ocimunu*. While this is the most common, the pronunciation and the attempt to make phonetic part of the word makes *ochimunu* and *otchimunu* acceptable orthographic forms.

Table 2 Umbundu regular paired noun classes. Based on [5].

Umbundu		**English**	
Singular	Plural	Singular	Plural
class 1: u- ukombe	class 2: a- akombe	guest	guests
class 3: u- upange	class 4: ovi- ovopange	work	works
class 3: u- ulume / ukãyi	class 6: a- alume / akãyi	man / woman	men / women
class 5: e- ekepa / etimba / etapalo / ekandu	class 4: a- akepa / atimba / atapalo / akandu	bone / body / road / sin	bones / bodies / roads / sins
class 5: e- ekomohiso, ewe	class 4: ovi ovikomohiso, ovawe	wonder, rock	wonders, rocks
class 7: oci- ocitungu, ocitangi	class 8: ovi- ovitungu, ovitangi	bundle, problem	bundles, problems
class 9: o-, ∅- ohumbo, omunda	class 10: olo- olohumbo, olomunda	needle, mountain	needless, mountains
class 11: olu- olunyi	class 10: olo- olonyi	fly	flies
class 12: oka okavisungo	class 13: otu okatuvisungo	little song	little songs
class 14: u- utima	class 6 ova- ovitima / ovilwa	heart / whistle	hearts / whistles
class 15: oku- okulama	class 4: ovi- ovilamo		greetings
class 5: e- eteke	class 10: olo- oloneke	day	days

Table 3 Umbundu irregular paired noun classes. Based on [5].

Class	**Prefix merge**	**English**
Class 11: olu olumbo	class 6 (a-) + class 11 (olu) alumbo	fence
Class 15: oku okwenye	class 6 (a) + class 15 (oku) akwenye	dry season

The pairing of the class system allows us to know how the plural is done in Umbundu. Although some slant is recognized, there is, as Katamba [9, p. 129] states, “distributional characteristics of noun classes showing a high degree of coherence.”

2. There is no specification of the gender, but there are words that gives us hints:
   - Common nouns of human kinship: *ulume* (man) and *ukãyi* (woman); *ukwenje* (boy) and *ufeko e ukano* (girl); *ise* (father) and *ina* mother; *mulume* (brother) and *mukãyi* (sister).
   - Kinship nouns more closely related to membership: *tate*, *so* and *ise* (father); *nyoho* and *ina* (mother).
   - Gender-related nouns for animals:
     onwi (bull), *ongombe* (ox), *onjindi* (cow), *onale* (calf), *ombelipa* (heifer), *ekondombolo* (rooster), *osanji* (hen), *ochitupi* (goat), *oselenge* (castrated goat), *ondume* (animal male) and *omange* (animal females)
3. The gender of the nouns is indicated by the postposition of the words for male or female (*ulume* or *ukãyi*), and the plural being made in the class leaving intact the radical of the word.

- For humans *ulume* (man), *omola ulume* (son), *ukãyi*: (woman), and *omola ukãyi*: (daughter).
- For irrational animals, the indication of gender occurs in two ways:
  a. Applying *ulume* (man) and *ukãyi* (woman) as *ombwa yulume* (dog), *ombwa yukãyi* (female dog), *ongulu yulume* (pig) and *ongulu yukãyi* (sow).
  b. Applying *ondume* or *ochilume* for male and *omange* for female: *onjamba yondume* (male elephant) and *ombwa yomange* (female dog).
- For birds and little animals, when in the diminutive, the concordant is the one corresponding to the class, but the determinant can be diminutive for females and augmentative for males: *okanjila k'okamange* (little female bird) or *okanjila k'ochilume* (little male bird).
- There are no articles: *Omoko* (knife), *Ositu* (meat) or *Oviti* (trees).

2.2.2 Adjectives

In Umbundu, the adjectives are divided into two groups: simple and verbal (derived from verbs), as shown in Table 4.

Table 4 Adjectives classes.

Simple adjectives		**Verbal adjectives**	
Umbundu	**English**	**Umbundu**	**English**
nene	big	pepa	tasty, delicious
tito	small, little	lula	bitter
lepa	tall	Lehã, neha	smell
Vi	evil, bad	yela	white
wa	good	tekãva	black
ewa	very, much	vela	sick
mbumbulu	short	pya	boiled
sumuluhwa	happy	vola	rotten
osuke	poor	Nyolehã	spoiled
umahele	young	kola	strong
umosi	uno	lile	weak
ukwangusu	forceful, strong	neta	fat
kavali	mutual	kachikapa	despicable
ukwafeka	Native, home	leluka	easy
		Letiwe, moleha	visible

Adjectives are neutral or common for gender, just the names [13]: *ukãyi walepa* (tall woman) and *ulume walepa* (tall man); *ukãyi una uvi* (that woman is evil) and *ulume una uvi* (that man is evil), *ongulo inene* (the pig is big) and *omange yongulo inene* (the sow is big), *ekondombolo inene* (big cock) and *onsanji inene* (big hen).

2.2.3 Verbs

According to Le Guennec and Valente [10], the verbs in Umbundu can be simple or derived. There are only three modes: indicative, conjunctive and imperative. Regarding the indicative mode, taking as reference *-linga*, the radical of the verb *to do* in the three main tenses (present, past and future) is marked by the particles *e-* (for the present: *ndi-linga* – I do]); *a-* (for the past tense: *nda-linga* – I did) and *ka-* (for the future: *ndi-ka-linga* – I will do).

There is no change in the radical of the verb for either the time or the number:

oku-linga	(to do)
ame ndi-linga	(I do)
etu tu-linga	(we do)
ame nda-linga	(I did)
etu twa-linga	(we did)
ame ndi-ka-linga	(I will do)
etu tu-ka-linga	(we will do)

Regarding concordance patterns [5, 13]:

1. Noun + Adjective (singular/plural)
   uti unene / oviti vinene (big tree / big trees)
   ukãyi wafina / akãyi vafina (pretty woman / pretty women)
2. Noun + Determinant Demonstrative
   Onjo ina / olonjo vina (that house / those houses)
3. Verb (verbal form) + Noun
   twayeva ondaka (we heard the message)
4. Pronoun + verbal form
   Ame ndilya (I eat)
   Ovo valya (they eat)
5. Noun + numeral
   omoko imosi (one knife)
   olomoko vitatu (three knives)

3 jSpell Environment

jSpell was developed as a fork of the well known Unix spell checker, ispell[2] with the main objective of allowing it to work as a morphological analyzer [1]. It was originally used to construct a morphological analyzer for the Portuguese language.

While jSpell might be used both as a command line application (with a similar interface as ispell) or as a C programming library, a Perl module was developed to help in the process of using it from within NLP tools developed in the Perl programming language [11].

jSpell dictionaries are comprised of two different files:

- A dictionary with a list of lemmas (word roots), the morphological properties for the lemma, and a list of flexion paradigm identifiers.
- An affix file, containing the flexion paradigm rules. Each flexion paradigm includes rewriting rules that specify how a word form can be generated from the lemma (what portions of the word are removed, added or rewritten) and how the morphological analyses changes with the flexion (for example, changing gender, number, or verb tense).

From this set, jSpell is not only able to check spelling, but also perform morphological analysis (without disambiguation) and to generate word forms accordingly with the morphological constrains needed.

[2] `https://www.gnu.org/software/ispell/` [accessed on 2020-07-03]

Another interesting property of jSpell is the ability to work in a guess mode, where unknown words are analyzed for possible derivation from any known lemma. This allows the tool to do morphological analysis on unknown words, while trying to apply any flexion paradigm to any lemma in the dictionary.

As the Portuguese dictionary for jSpell was the source of the first spell checking dictionary for the Portuguese language, that was shipped within all major Linux distributions together with ispell, the necessity to adapt the dictionary to other engines, like aspell[3], HunSpell[4] and MySpell[5] soon arrived. This lead to the development of an environment able to convert from dictionaries from jSpell format into any other dictionary format [14], automating the creation of packages for all major open source tools, like LibreOffice, OpenOffice, Firefox, Thunderbird and all Linux spell checking engines.

4 Dictionary Bootstrap

To help in the bootstrap process, the first task was to compile corpora from the Internet. Given we do not intend to release (yet) an Umbundu corpus, at the moment the approach did not take into account copyright issues.

Being a language with oral tradition, and with a relatively small number of written resources, some of the issues found with other languages years ago are still present in Umbundu: there is a lot of different ways to write words. While there are some dictionaries, the majority of the population is illiterate.

Therefore, our task cannot be completely guided from corpora, forcing us to query language users and to consult written references, as Le Guennec and Valente [10] dictionary of Portuguese/Umbundu.

We focused on nouns and regular verbs. The next sections present some of the rules currently in use, exemplifying the results.

4.1 Nouns

For the plural paradigm, the rules are written as:

```
flag p:
  > -E    , A     ; "N=p"     # door: epito / apito
  > -S    , OLOS  ; "N=p"     # monkey: sima / olosima
  > -OMU  , OMA   ; "N=p"     # person: omunu / omanu
  > -OMO  , OMA   ; "N=p"     # child: omola / omala
  > -OLU  , OLO   ; "N=p"     # grain: olumema / olomema
  > -OCI  , OVI   ; "N=p"     # song: ocisungo / ovisungo
  > -OTCHI, OVI   ; "N=p"     # white: otchindele / ovindele
  > -OTCHE, OVYE  ; "N=p"     # election: otchela / ovyela
  > -OKU  , OVI   ; "N=p"     # food: okulya / ovilya
  > -OKA  , OTU   ; "N=p"     # bread roll: okambolo / otumbolo
  > -M    , VAM   ; "N=p"     # brother: mange / vamange
  > -U    , A     ; "N=p"     # man: ulume / alume
```

This paradigm is comprised of different rules for different prefixes. Before the > sign, a pattern to match lemmas prefixes is specified. In this example, none of the rules have a pattern, meaning that the rules will be applied if their rewrite rule can be applied. After

[3] http://aspell.net/ [accessed on 2020-07-03]

[4] https://hunspell.github.io/ [accessed on 2020-07-03]

[5] https://www.openoffice.org/lingucomponent/dictionary.html [accessed on 2020-07-03]

the > sign, there is the rewrite part: first the characters to remove from the beginning of the word, and then the characters to be prefixed. After the semicolon, a set of properties to rewrite the morphological analysis metadata is presented. Everything after the sharp sign are comments, illustrating the rules usage.

Note that the rules, themselves, do not specify if the rewrite will take place in the beginning of the words or at the end. That information is provided by a separator in the rules file, specifying that rules following that separator will be applied as prefixes.

Unfortunately not all words follow the generic plural construction. For some of them there was the need to create another paradigm, regarding irregular plurals:

```
flag q:
  E [WKYP] > -E, OVA  ; "N=p"  # hand: eka / ovaka
  E [^WTKY]> -E, OVI  ; "N=p"  # wrinkle: enha / ovanha
  U [^TN]  > -U, OVO  ; "N=p"  # work: upange / ovopange
  U [TN]   > -U, OVI  ; "N=p"  # heart: utima / ovitima
           > -O, OLO  ; "N=p"  # window: ombana / olombana
```

4.2 Verbs

A first set of rules for the basic regular verb flexion was already added, including the present, perfect-past and future tenses, and the semantic negative form for the present tense. A subset of the rules follows.

```
flag v:
 # Examples for okulya (to eat)

 # TENSE: present
 > -OKU, NDI  ; "P=1,N=s,T=ip" # ndilya
 > -OKU,  TU  ; "P=1,N=p,T=ip" # tulya
 > -OKU,   U  ; "P=2,N=p,T=ip" # ulya
 > -OKU,  VA  ; "P=3,N=p,T=ip" # valya

 # TENSE: perfect-past
 > -OKU, NDA  ; "P=1,N=s,T=ipp" # ndalya
 > -OKU,   O  ; "P=2,N=s,T=ipp" # olya
 > -OKU,  WA  ; "P=3,N=s,T=ipp" # walya
 > -OKU, TWA  ; "P=1,N=p,T=ipp" # twalya

 # TENSE: future
 > -OKU, NDIKA ; "P=1,N=s,T=if" # ndikalya
 > -OKU,   UKA ; "P=2,N=s,T=if" # ukalya
 > -OKU,   OKA ; "P=3,N=s,T=if" # okalya
 > -OKU,  VAKA ; "P=3,N=p,T=if" # vakalya

 # Negated sense: to starve

 # TENSE: present
 > -OKU,  SI ; "P=1,N=s,T=ip,M=neg" # silya
 > -OKU,  KU ; "P=2,N=s,T=ip,M=neg" # kulya
 > -OKU,KAVU ; "P=2,N=p,T=ip,M=neg" # kavulya
 > -OKU,KAVA ; "P=3,N=p,T=ip,M=neg" # kavalya
```

Another different kind of rule is the modal abstraction, creating a noun from a verb:

```
flag u:
  > U ; "CAT=nc,MO=abs"  # to fear: sumba / fear: usumba
```

4.3 Dictionary

The dictionary was constructed from different sources in the Internet. As stated previously, at this moment we are not pretending to create a public corpus, and therefore we did not manage the rights for the obtained text. Nevertheless, these texts allowed us to create lists of terms, and understand and test the morphological rules.

The dictionary itself, is a list of lemmas, followed by a morphological information (in the examples presented, #v stands for a verb, while #nc stands a common name. For reference, and to allow further uses of the dictionary, we also included a Portuguese translation. At this point we are not concerned with polysemy. Our main goal is to register the information, to help us understand and work with the language. The Portuguese language was chosen given it is the main language used in Angola's written documents. At the end of each line, there is a flag, that represents the inflexion paradigm that is being applied to that word.

```
okulya/#v,pt=comer/v
okwiva/#v,pt=roubar/v
okupapala/#v,pt=brincar,#d/v
okutanga/#v,pt=estudar,#d/v

ekandu/#nc,pt=pecado/p
epito/#nc,pt=porta/p
etali/#nc,pt=pedra/p
sekulu/#nc,pt=ancião/p
ukãyi/#nc,pt=mulher/p
ukombe/#nc,pt=visita/p
ulume/#nc,pt=homem/p
olunyi/#nc,pt=formiga/p
okavisungo/#nc,pt=canção/p
okulama/#nc,pt=saudação/p

upange/#nc,pt=trabalho/q
ekomohiso/#nc,pt=maravilha/q
ewe/#nc,pt=pedra/q
ondjo/#nc,pt=casa/q
ovilwa/#nc,pt=assobio/q
eteke/#nc,pt=dia/q
```

5 Results Discussion

As previously referred, this is the kick-off for a project on the creation of tools and resources for the Angolan indigenous languages. Although we intend to collaborate with institutions and researchers from Angola, this first proof of concept was developed to understand these language characteristics.

The current morphological analyzer coverage is quite limited. The dictionary is comprised of two different sources:

- about 650 lemmas, encoded with their derivation paradigm, generating more than 6 500 different forms;
- a second dictionary with more than 1 700 forms, that are being manually validated, and classified with a derivation paradigm.

These two sources together cover more than 8 000 different word forms.

While in an early stage of development, we foresee to have a free and publicly available dictionary for the Umbundu language available in the near future.

References

1 José João Almeida and Ulisses Pinto. Jspell – um módulo para análise léxica genérica de linguagem natural. In *Actas do X Encontro da Associação Portuguesa de Linguística (APL'1994)*, pages 1–15, 1995.

2 Lluís V. Aracil. *Papers de sociolingüística.* Edicions La Magrana, Barcelona, 1982.

3 Boubacar Diarra. Choice and description of national languages with regard to their utility in literacy and education in Angola. In *A paper delivered at the UNESCO Expert Pool Meeting on Language issues in Literacy and Basic Education*, 1992.

4 Charles Ferguson. Diglossia. *Word*, 15:325–340, 1959.

5 João Fernandes and Zavoni Ntondo. *Angola: povos e línguas.* Editorial Nzila, Luanda, 2002.

6 Malcolm Guthrie. *The classification of the Bantu languages.* Oxford University Press, 1948.

7 INE. Resultados definitivos do recenseamento geral da população e da habitação de angola. Technical report, Instituto Nacional de Estatística, Gabinete Central do Censo, Subcomissão de Difusão de Resultados, Luanda, 2016.

8 Botelho Isalino Jimbi. A reflection on the umbundu corpus planning for the Angola education system: towards the harmonization of the catholic and the protestant orthographies. In *Actas Do XIII Congress Internacional de Linguistica Xeral*, page 475–482, 2018. Retrieved from `http://cilx2018.uvigo.gal/actas/pdf/661789.pdf`.

9 Francis Katamba. Bantu nominal morphology. In D. Nurse and G. Philippson, editors, *The Bantu Language.* Routledge Tayloer & Francis Group, London, 2014.

10 Gregoire Le Guenec and José Francisco Valente. *Dicionário Português-Umbundu.* Escolar Editora, Lobito, 2010.

11 Alberto Manuel Simões and José João Almeida. `jspell.pm` – um módulo de análise morfológica para uso em processamento de linguagem natural. In *Actas da Associação Portuguesa de Linguística (APL'2001)*, pages 485–495, 2002.

12 Universal Declaration of Linguistic Rights, 1996. (Barcelona Declaration) World Conference on Linguistic Rights, Barcelona, Espanha, Junho de 1996. Retrieved January 31, 2020, from `https://unesdoc.unesco.org/ark:/48223/pf0000104267`.

13 João Francisco Valente. *Gramática Umbundu. A língua do centro de Angola.* Junta de Investigação do Ultramar, Lisboa, 1964.

14 Rui Vilela. Geração de dicionários para correcção ortográfica do português. Master's thesis, Escola de Engenharia, Universidade do Minho, 2009.