



Proposal for the Creation of a Portuguese Dictionary: An Ongoing Project Based on Corpus Analysis of the Angolan Variety

MÁRCIO UNDOLO 

Translator:

LISA SHAW

*Author affiliations can be found in the back matter of this article

 MLO Modern Languages Open

SPECIAL COLLECTION:
LANGUAGE AND
EDUCATION IN
THE LUSOPHONE
COUNTRIES: THEORY
AND PRACTICE – LÍNGUA
E EDUCAÇÃO NOS
PAÍSES LUSÓFONOS:
TEORIA E PRÁTICA

ARTICLES –
LINGUISTICS


LIVERPOOL
UNIVERSITY PRESS

ABSTRACT

This article reflects on the question of the variety of Portuguese spoken in Angola where, on the one hand, a national educated-variety of Portuguese is emerging and, on the other, the use of Portuguese in accordance with the norms of standard European Portuguese, the written standard for Angola, is not correlated with speakers of high socio-economic status. These speakers use the emerging educated variety, which is based around the speech of the capital, Luanda. This variety acts as a centripetal force with respect to the other varieties of Portuguese spoken in Angola and contributes to the national and cultural cohesion of the state. This is contrasted with the standard variety from Portugal, which acts as a centrifugal force, since speakers are not usually exposed to this form of speech and do not connect with it. I draw attention to the lack of research on and resources for Angolan Portuguese and I present new and ongoing research from the Variedade do Português de Angola (VAPA) project on Angolan Portuguese and specifically the ongoing work towards the creation of the first ever monolingual dictionary for Angolan Portuguese. Methodologically, the work for the dictionary follows a linguistic approach based on the analysis of carefully selected corpus data. Specifically, data from a set of corpora from social media is being used as the empirical basis of the project since in official documents, publications, and formal communication contexts of official institutional bodies there is always a stringent observation of the European standard variety. The results of the research show that there are various linguistic processes and means by which the Angolan Portuguese lexicon is being enriched and expanded. Of particular interest is the finding that borrowing of words or structures from Bantu languages is not the main source of neologisms. Rather, other processes internal to the Angolan Portuguese linguistic system dominate, such as reductions or abbreviated forms and other types of semantically and morphologically motivated neologisms.

CORRESPONDING AUTHOR:

Márcio Undolo

Doutor em Linguística pela
Universidade de Évora;
Professor Associado do
Instituto Superior de Ciências
da Educação de Benguela
(Angola), AO

undolomarcio@gmail.com

TO CITE THIS ARTICLE:

Undolo, Márcio 2022
Proposal for the Creation of
a Portuguese Dictionary: An
Ongoing Project Based on
Corpus Analysis of the Angolan
Variety. *Modern Languages
Open*, 2022(1): 19, pp. 1–12.
DOI: [https://doi.org/10.3828/
mlo.v0i0.403](https://doi.org/10.3828/mlo.v0i0.403)

In this project, the aim is to create a monolingual dictionary of everyday language, which brings together habitual linguistic forms based on local social practices. John Robert Schmitz (qtd. by Henriques, 2011) reminds us that words create and shape our understanding of the world we live in, and that we base our decisions and judgements about others on them. The opposite is also true. It is based on this assumption that we, the project group, seek to demonstrate the importance of creating a dictionary of the Angolan variety of Portuguese.

The Angolan variety of Portuguese can be understood as a set of local linguistic phenomena that arise within a specific social-political-cultural matrix and with which speakers immediately identify. As is well known, languages of different origins coexist in the Angolan context, including: (i) Koi, San; (ii) Bantu varieties Umbundu, Kimbundu, Ngagela, Kikongo, Nyaneka, Helelo, Cokwe, Luba, Kyombe, Hanya, Nyemba, and so on; and (iii) Portuguese.

Angola's linguistic position is still uncertain and, therefore, controversial. In general, a language is defined from a political standpoint, which conflicts with studies that certain linguists have been conducting. Lusakalalu (2005, p. 9), for example, contends that "an analysis of this issue necessarily involves the fact that, when the question 'how many languages?' is asked, the next question has to be 'which ones?'".

Portuguese is the only official language, the main vehicular language, and, above all, the language of national unity in Angola. As an official language, a status conferred by the Constitution of the Republic (cf. article 19, no. 1), it is used by the various bodies and departments within the three branches of government, namely the executive, the legislature, and the judiciary, and in the public administration in general. Therefore, it is the language of promotion and social ascension *par excellence*. In addition, at the household level, it is a *lingua franca* used whenever people from different ethnolinguistic communities interact with each other.

The starting point for this article is the hypothesis that linguistic contact between Portuguese and other, particularly the Bantu, languages is key to explaining certain phenomena of variation and change in Angolan Portuguese. However, this hypothesis is not sufficient, given that different factors contribute to this effect. These include changes in the social structure, the governance system, and the national economic system; Martinet's (2014, p. 195) "principle of least effort"; and the prevalent pedagogical approach within the national education and teaching systems.

Structurally, this article is divided into four parts. The first concerns the technique of preparing lexicographic resources known as *practical lexicography*, together with a brief overview of the history of lexicography in Portugal and especially in Angola. The second part relates the concept of language to the concept of culture, focusing on the primary function of language—communication. The third analyses the properties of the lexicon. Finally, the fourth part presents the methodological assumptions of this research, namely the dictionary model, research criteria, and the specific object of linguistic analysis.

LEXICOGRAPHY

DICTIONARIES

For many authors, practical lexicography (also called "dictionary lexicography" by the French school) refers to the "technique of creating a dictionary". Although there are those who designate it an art or even a set of methodological assumptions necessary for the development of dictionaries, it is important to mention that, in addition to creating dictionaries, lexicography is also the technique used to prepare a glossary, among other linguistic resources such as lexicographic reference works. In Angola, the term *lexicography* is often confused with the term *lexicology* among arts and humanities students. The latter term is defined as the science of the lexicon of a given language. The term "lexicon", as we know, refers to the active and passive reservoir of a people's culture. We will return to this subject later. First, we must review the state of play regarding lexicographic work in Angola.

The history of Angolan lexicography is filled with lexicographic reference works, namely bilingual and plurilingual dictionaries: Olunyaneka-Portuguese, by Fr. and B. M. Bonnefoux (1941); Portuguese-Kimbundu-Kikongo, by Fr. António da Silva (1961); Portuguese-Kimbundu,

by Fr. António da Silva Maia (1964); Portuguese-Olunyaneka, by Fr. António Joaquim da Silva (1966); Portuguese-Umbundu, by the priests Gregório LeGuennec and José F. Valente (1972); and Portuguese-Umbundu, by Rev. Henrique Etaungo Daniel (2010).

Unlike Bantu languages, Angolan Portuguese has never had a monolingual dictionary. In its place is an unsystematic, limited lexical survey of the Angolan variety, seen as a vehicular language.

In Portugal, the second half of the eighteenth century is seen as the landmark period for the appearance of the first monolingual dictionaries. Noteworthy are the works of Melo Bacelar, published in 1783, and Morais Silva, published in 1789, as well as that of the Lisbon Academy of Sciences, published in 1793.

In the Americas, “the history of general lexicography in Brazil [...] begins, in fact, only in 1938, the year that *Civilização Brasileira* publishes the *Short Brazilian Dictionary of the Portuguese Language*” (Henriques, 2011, p. 19), in other words, at the end of the first half of the twentieth century.

In relation to Angolan Portuguese, the scant work of Portuguese lexicographers is well known, as can be seen in the *Dictionary of Contemporary Portuguese Language* produced by the Lisbon Academy of Sciences (2001). In terms of the history of *practical lexicography* in general, bilingual dictionaries appeared first. For example, in relation to English lexicography, bilingual dictionaries appeared at the beginning of the sixteenth century and monolingual dictionaries at the end.

In the lexicon of the Angolan variety of Portuguese, phenomena arising from various processes can be seen. One of these is the evolutionary or dynamic nature of languages, in time and space, as they come into contact with different mental or cultural frameworks. This explains the changes that already distinguish the national variety of Portuguese from other varieties of the same language around the world.

According to Raposo, Vicente, and Veloso (2013), “one of the ways to classify the different phenomena of linguistic change is to indicate [...] the changes in the inventory of words that speakers know—lexical substitution—and in the meanings associated with words—semanticization” (p. 23). Brocardo (2014) states that lexical change “is generally understood as either the appearance of new lexical items in a given language (innovation) or the disuse of words, which are no longer part of its lexicon (obsolescence)” (p. 83).

In the particular case of Angolan Portuguese, a set of changes have been noted with regard to the range of words which are common among (most) Angolans but not shared with Portuguese speakers of other nationalities. In the corpus we analysed, especially within sports vocabulary, we often find the lexical unit <girabola>:

O **Girabola** Zap, que sofreu uma interrupção neste fim-de-semana, prossegue no próximo dia 10 deste mês. (in Angonotícias, 5 June 2016)

[Girabola Zap, which suffered an interruption this weekend, resumes on the 10th of this month]

In Angola, the noun <girabola>, belonging to the domain of sport, refers to the national 1st division football championship; and <girabolaZap> refers to the national 1st division football championship sponsored by Zap. Zap is the name of a private law company operating in telecommunications that provides satellite television in southern Africa.

In terms of morphological structure, two lexical morphemes stand out—the inflected form of the verb <girar> (3rd person singular, indicative mood) and the feminine noun <bola>, giving rise to a lexical unit of morphosyntactic composition, with a <verb+noun> scheme, with masculine gender, and without grammatical number variation.

In addition, there are other lexical units such as: <kinguila>, <zungueira>, and <Kwanza>. It is noteworthy that, in written language, the noun <kwanza> with a “k” refers to the national currency, while <Cuanza> with a “c” can refer to either the largest river in Angola or two Angolan provinces, that is, <Cuanza Norte> and <Cuanza Sul>, as evidenced by extracts from our corpus. Also of note is the word <kinguila>, which means “travelling money changer”. This

word co-occurs in written discourse with the form <quingla>, an adaptation from the Kimbundu language to local Portuguese.

... na Mutamba, as **kinguilas** vendiam a nota de 100 dólares a 78 mil **kwanzas**. (in Angonoticias¹, 23 May 2008) [AQ: please provide full details for the references]

...in Mutamba, the **kinguilas** used to sold the 100 dollars for 78 thousand of **Kwanzas**.

Em declarações quarta-feira à imprensa [...], a directora provincial do Gabinete da Saúde no **Cuanza Norte**[...] frisou que a recuperação dos pacientes ... (in Angop,² 13 January 2021) [AQ: please provide full details for the references]

[In statements to the press on Wednesday (...), the provincial director of the Health Department in **Cuanza Norte** [...] stressed that the recovery of patients ...]

O livro [...] que homenageia a mulher **zungueira**, será apresentado sexta-feira (15), em Luanda, na sede da União dos Escritores Angolanos.

[The book (...) that pays tribute to the **zungueira** woman, will be launched on Friday (15), in Luanda, at the headquarters of the Angolan Writers Union]

Trata-se da segunda edição de uma obra em formato de revista que circulou a primeira vez há quinze anos, contendo histórias de vida de várias mulheres que asseguram a sobrevivência das suas famílias na “**zunga**”. (in Angop, 14 January 2021)

[This is the second edition of a work in a magazine format that was circulated for the first time fifteen years ago, containing the life stories of several women who provide for their families via the “**zunga**” trade]

There are, in the “local orthographic system”, lexical units with alternating written forms. This is the case, for example, with the noun Cuanza-Norte or Cuanza Norte, whose morphosyntactic composition may or may not contain a hyphen.

The noun <zungueira> is not only a case of suffix derivation but also of hybridity. In the internal structure of this lexical unit, the suffix <-eira>, typical of the Portuguese language, stands out. It is linked to the typically Bantu Kimbundu stem <zung-> from which the verb <zungar> is also formed, that is, stem <zung-> + thematic vowel <-a-> + infinitive morpheme <-r>.

Based on our corpus, *zungueira* refers to a woman who practises an itinerant trade, and *zunga* is the activity itself. Linguistic forms peculiar to the Angolan variety of Portuguese can often be found in formal and informal communication contexts.

LANGUAGE AND CULTURE

We know that the primary function of a language is communication. In this sense, according to Martinet (idem, p. 18), “every given language changes in the course of time [...] because of the need to adapt itself in the most economical way to the needs of communication of the community which speaks it”. Among linguists, there is a consensus that “languages, we know, are not necessarily identical over the whole territory in which they are spoken” (Martinet, idem, p. 38). In other words, if at the level of a national territory a given language varies and takes the form of different dialects, such variations are more marked from country to country, from continent to continent, and there are sometimes divergences that affect mutual intelligibility, both at the phonic and lexical levels.

Martinet (idem, p. 192) argues that “we can assume that the evolution of a language depends on changes in the needs of communication of the group which uses it”. Likewise, it can be argued that the evolution of communicative needs is a consequence of the social, cultural,

1 <https://angonoticias.com/>

2 <https://www.dev.angop.ao/>

political, and economic dynamics of a given community or society. This theoretical context explains the Angolan variety of Portuguese. Furthermore, Angolan Portuguese is a vehicle for the national way of thinking, being, and acting typical of the Angolan people.

Miranda (2013) states that “every group, every society, produces models for representing and interpreting the world, that are widely shared by its members, and that intervene in a decisive way in their behaviour and understanding of reality. The study of language could be, from this perspective, a privileged way of accessing the mental and cultural world of peoples” (p. 15). This clearly illustrates the consensus among linguists that all languages are at the service of their speakers and not the other way around. Farias (2016, pp. 25–26) adds that

language is loaded with the culture of the people who use it, both in its semantics and its lexicon, as well as in its syntax, systemic elements that reflect fundamental attributes of that culture. Because of this specificity, there are many theorists who argue that each language obeys an intricate ethnolinguistic apparatus, an association of customs, traditions, experiences, information, lifestyles and ways of apprehending reality, for which language presents a fitting and immediately available expressiveness. Consequently, the multiplicity of languages among peoples can be attributed to cultural multiplicity.

Furthermore, within the same language, to paraphrase Farias, the multiplicity of dialects and varieties among social groups or societies can be attributed to cultural diversity. Thus, the study of Angolan thought or culture, as a product of “Angolanness”, necessarily entails the study of the Angolan variety of Portuguese. Miranda (2013, pp. 16–17) continues,

Unlike Chomsky (1957), who advocated the study of an individual’s internal mental structures, for Hymes (1972), a speaker, to be communicatively competent or to be integrated [into] a community, must not only master linguistic structures, but also know how the language is used by the members of this speech community, that is, s/he must be able to produce utterances that are appropriate for the context or environment, and do not sound strange. This competence is integrated with their attitudes, values and motivations regarding the properties and use of language and with their attitudes and competence to understand the interrelation of language with another communicative code of conduct [...] The use of language is, therefore, culturally mediated, and is thus an aspect of cultural identification, as shown by Duranti (2000, p. 39) when stating that

If we are to understand the role of language in people’s lives, we need to go beyond the study of its grammar [and lexicon] and enter the world of social action, where words are embedded and constitutive of specific cultural activities such as telling stories, asking for a favour, showing respect, insulting.

This illustrates the relationship that is established between language and culture. Given this relationship, we aim to reflect on the cultural value that a dictionary holds in society. The reality of the Angolan case is that there is a lack of linguistic resources intrinsic to national culture such as, for example, dictionaries, glossaries, lexicographic databases, and descriptive grammars. It is therefore essential to begin developing activities that will lead to what we hope will be an exhaustive work on Angolan lexicography. This will permit the gathering, systematisation, harmonisation, and dissemination of collective linguistic knowledge, which forms part of the nation’s intangible cultural heritage.

LEXICON AND WORDS

A study conducted over more than thirty years by Rey-Debove (1984) established that there are three types of lexicon: (i) a language’s set of morphemes, (ii) a language’s set of words, and (iii) a language’s set of open class words. For the purposes of this article, we adopt the second definition which, although traditional, is the one that enjoys the greatest consensus. Rey-Debove (1984, p. 50) states that the lexicon, as a set of words in a language, includes most morphemes (free morphemes) and all coded units of various morphemes (derived and compound words). This is the vision of the lexicon that dictionaries give us. Azevedo and Silva

(2017, p. 39) believe that “the lexicon can be defined as the repertoire of words present in the language and which enables verbal interaction”.

Camara Jr. (2007, p. 232) defines a word as a free linguistic form, in other words, a lexeme, to use North American terminology. However, he admits that there are those who include dependent linguistic forms in this category. For this research, from the outset we chose three terms: (i) lexical unit, (ii) lexicographic unit, and (iii) grammatical categories. For Azevedo and Silva (2017), “when we take the word as a lexical unit, we also need theoretical investigation capable of conceptualizing what a lexical unit is [...] This theoretical investigation is a function of lexicology”. Likewise, a lexicographic unit must be conceptually differentiated from a lexical unit and a grammatical category. Lehmann and Martin (2000, p. 30) define a lexical unit as being “le mot. L’identité d’un mot est constituée de trois éléments: une forme, un sens et une catégorie grammaticale. Un mot résulte de l’association d’un sens donné à un ensemble de sons donnés susceptible d’un employ grammaticale donné” [Word. The identity of a word is made up of three elements: a form, a meaning and a grammatical category. A word results from the association of a given meaning to a group of given sounds likely to have a given grammatical use] (qtd. in Quivuna, 2014, p. 45; my translation). Quivuna (2014, pp. 45–46) adds that

in fact, after analysing this definition, we see the complexity of the nature of the “lexical unit”, a term in lexicology, which designates a unit of meaning (with 46 different categories) but which is not to be confused with a written word or a form (between two blank spaces). For example: wagon is a lexical unit; restaurant is another unit; but if we unite the two units through the composition process, we will have a unit of meaning, that is, a lexical unit: restaurant car (car with restaurant service inside a train).

According to Lino (2004, pp. 264–265, qtd. in Quivuna, 2014, p. 46), the lexical unit evolves into several different categories: “at the level of the linguistic system, the lexical component takes the name of *lexicon*; and the lexical unit, *lexeme*. At the level of discourse, in other words, when language becomes speech, the lexical component is called a *vocabulary item*; and the lexical unit, a *word*. In a dictionary, the lexical unit is called a *lexicographic unit*”. The term *grammatical category* refers to a set of morphological and/or morphosyntactic features inherent to lexicographic units in the context of usage.

RESEARCH METHODOLOGY AND RATIONALE

This article is the result of research conducted as part of the doctoral programme in Linguistics at the University of Évora, Portugal between 2011 and 2014. I completed the programme and defended my thesis on 23 February 2015, and this led on to the current research project, the working title of which is *Varietade do Português de Angola (VAPA)* [Angolan Variety of Portuguese]. Its theoretical basis is to pursue the following overall objective: to contribute to knowledge about the current status of the Portuguese language in Angola, characterising and establishing the educated language norms of Angolan Portuguese.

To this end, the project uses a methodology based on corpus analysis of texts extracted from local media sources, namely TV, radio, and the press. These sources were chosen as in formal contexts they strive to use the standard language, thus enabling us to assess local linguistic usages and make judgements about national educated language norms. For the collection of oral data, we employed audio recording and, for written data, a computer program that enabled us to assess the frequency, occurrence, and concordance of certain linguistic data. We analysed the data using contrastive and semasiological methods.

The first linguistic studies took place in post-independence Angola between 1975 and 1999. One of the criticisms levelled at the authors of those studies is the fact that they followed an intuitive data analysis methodology based on grammatical correction. However, the studies that followed in the period from 2000 to 2010 evidenced a degree of concern with linguistic phenomena, although there was still a tendency to focus on difficulties encountered in teaching, grammatical traditions, empirical methods, and comparative linguistics between Portuguese and the Bantu languages.

Most of the authors focused on contrasting the particularities of popular variants in Angola with the parameters of standard European Portuguese. However, as of 2011, it was understood that contrastive studies had to be based on a logic that respected principles of equality or equity. This approach would justify contrasting standard European Portuguese (also referred to as standard Portuguese of Portugal) with the variant of Portuguese that Angolans use to communicate in formal contexts, given that, in such situations, educated speakers seek to adapt their speech to the communication context. This is why the VAPA project favours media sources as a source of empirical evidence, spaces where people with secondary or higher education come into social contact.

After independence on 11 November 1975, Angola adopted Portuguese as its official language. However, the issue of the standard official norms of this language was not addressed in language policy. The European model is often referred to as the linguistic reference point and, therefore, enjoys greater prestige than local linguistic features. According to Raposo et al. (2013), the standard norm enjoys “privileged social and cultural status, which is demonstrated in the fact that it is the variety taught in schools, used by the media (newspapers, magazines, radio, television), especially in formal communication contexts, by the majority of Portuguese people who have completed secondary or higher education” (p. xxv). Castro (2006) has stated that this standard “consists [...] of the written or spoken variety [...] used by the educated population and recognized as the variety that lends itself more easily to being reproduced, without obvious errors, but also without forms that are too cultured, incomprehensible or erudite” (p. 3).

A study we have been carrying out with teachers, students, and heads of primary schools, both in urban schools and in the outskirts, within the scope of the LODcom project (an educational alternative for learning Portuguese in primary education), has generated empirical evidence of how the Portuguese adopted in the classroom context, especially in oral communication, does not correspond to the ideal of standard European Portuguese. The research findings evidence drift in the following domains:

- phonetic-phonological, with emphasis on vocalic and consonantal systems, articulatory and prosodic schemes;
- syntactic, focusing on nominal and verbal government, nominal and verbal agreement, conjunctions and structures of coordination and subordination, gradation constructions, selection and distribution of pronouns;
- vocabulary, with an emphasis on phenomena of lexical innovation and renewal.

Outside the school environment, it can be seen that greater command of the standard norm is not directly related to the socio-economic status of speakers or the institution they are part of. In fact, the Portuguese language is used within a social framework that conditions its structure and functioning, which is similar to what happens in Portugal. The VAPA project’s informants, in their everyday life, have virtually no exposure to the European standard and they are educated individuals who have completed secondary or higher education, namely public servants, entrepreneurs, promoters or agents of cultural events, and so on. Despite their almost non-existent exposure to the European standard in their day-to-day lives, we would not categorise them as being poor speakers of Portuguese. Mateus (2005) argues that “We will all agree that it cannot be said that the language spoken in the fifteenth or seventeenth century is better than the one spoken in the twenty-first century. Likewise, it cannot be said that the Portuguese spoken in Portugal is better than that used in Brazil or Mozambique” (p. 24).

The status of Portuguese as the only official language of the Angolan state, its promotion as the only language of social mobility, the main language used in commercial and cultural relations, in the army, and in religion, as well as the exodus of the population to urban centres to the detriment of rural areas as a result of the long period of civil war, has all contributed to the spread and exponential growth of this language in Angola. We must also consider the deficient teaching of this language, due to a shortage of qualified teaching staff, and the loss of qualified staff who moved abroad shortly before and after 11 November 1975. The situation is aggravated further by the fact that at the time of independence around 85 per cent of the population was illiterate; there was a precarious school infrastructure; and teaching policy was based on the ideas of John Locke, ignoring the linguistic profile of students and the sociolinguistic reality of the country.

It is worthwhile considering some of the factors, both external and internal, that lead to change in any language:

- external factors: linguistic contact resulting from the country being multilingual, being characterized by cultural diversity, systematic updating of communicative needs resulting from changes in the social structure, in the governance system, or in the national economic system, which contribute to the formation of the Angolan variety of Portuguese;
- internal factors: reanalysis and the “least effort principle” (Martinet, *idem*), tending towards a simplification of parameters in different grammatical domains. A case in point is monophthongisation, assimilation, the suppression and reduction of number agreement, changes to pronoun government, comparative sentences and prepositional government. (Undolo, 2016)

Freitas, Ramilo, and Arim (2005, p. 33) contend that the media, along with school and parliament, “are seen as the main vehicle for disseminating the standard language”, but in the particular case of Angola, it appears that these three institutions are not disseminators of the European standard, unlike the public service broadcaster (RTP), schools, and parliament in Portugal. The issue that arises is whether the Angolan state will, in the near future, accept revisions to its language policy, recognising the Angolan variety of Portuguese—a necessary step to accelerate publicly funded research in this field.

As we know, since independence, Angolan society has been struggling with the following problem: knowing what constitutes “correct” usage of the Portuguese language. In Portugal, the European standard variety fulfils the social role of bringing speakers closer together, in the face of the different dialects on the mainland and the islands (Madeira and the Azores). However, in Angola, as can be seen, this same standard variety is not proving to contribute to sociocultural cohesion and normative linguistic identification, due to the factors mentioned above.

Therefore, given that “the existence of a standard variety is necessary as a reference for linguistic production and as a guarantee of the acceptability of certain behaviour in the socio-cultural context where we are located” (Mateus, 2005, p. 26), the failed attempt in Angola to attain an ideal linguistic system stems from, among other factors, the emergence of a national norm that has been demarcating itself from the popular local variants of Portuguese.

Authors such as Mingas (1998) and Nzau (2011) recognise that Portuguese is the national language in Angola. According to Mateus (2005), “all national varieties have their standard norm that school is a special depository of”. Given the position of the Portuguese language in Angola, it is fitting to question the concept of a standard official language and its sociocultural purpose. The term “standard language” refers to a linguistic variant of a language that has, for various historical, political, social, or cultural reasons, acquired a privileged status and is the variant used in schools, newspapers, magazines, radio, and television, as well as by the various bodies within the executive, the legislature, and the judiciary (Raposo et al., 2013). This begs the question: what is, in fact, the standard Angolan variety of Portuguese? Angolan language policy is silent on this matter.

For now, we will not refer to a standard, but rather an educated variant/norm. Mateus (2005) also says that “several factors determine the definition and acceptance of the [standard] which cannot, today, model itself exclusively on normative grammars or on established authors” (p. 27). In Angola, the educated norm can be determined by the linguistic usage in formal communication contexts, whether in oral or written modes, of speakers who have had extensive schooling. Their language stands out as different, especially in the phonetic/phonological and lexical domains. Such an educated norm coexists with other popular norms of the Angolan variety of Portuguese and it appears that its epicentre is the urban space of the capital city, Luanda. This is because external factors that influence the general linguistic behaviour of speakers, namely political, economic, and cultural power, are concentrated in the capital. Consequently, the aforementioned educated norm of the Angolan variety of Portuguese is the variant that contributes, according to Emiliano (2009), to the national and cultural cohesion of the state and the autonomous nation—the “centripetal force” in the context of different uses of Portuguese throughout the national territory.

Therefore, the standard norm of the variety from Portugal, pretentiously held up as a linguistic reference in Angola for use in formal communication contexts, acts as a “centrifugal force”, in opposition to the emerging educated norm of the Angolan variety. Therefore, the planned dictionary will, depending on the extent of the material from the corpus under analysis, allow us to delineate lexical uses emanating from the educated norm.

According to Martinet (idem, p. 55), “language is described in terms of what we find in it”. To address the need to develop a work based on real usage, corpus analysis provides the ideal methodology. Halliday and Matthiessen (1989, p. 34) argue that “the corpus is fundamental to the enterprise of theorizing about language”. Effectively, in this sense, our aim is to show that this is a methodology that allows for in-depth study of the empirical linguistic knowledge of speakers, leading to new theoretical insights and acceptations in relation to natural languages. However, following the ideas of Oliveira (2009, p. 49), it is important to bear in mind that a corpus can ensure the existence of evidence but not information. Oliveira contends that it is linguists who produce (new) information, theoretical or applied, from the corpus. It is necessary to define here the concept of “corpus”, which we interpret as denoting a set of oral or written texts that have been carefully selected with a specific purpose in mind. For Oliveira (2009), a corpus is necessary to describe linguistic facts precisely since, in order to interpret the data drawn from the corpus, we often also have to take into account the contexts and/or sociocultural aspects involved.

PROPOSED DICTIONARY STRUCTURE

We intend to make public a linguistic resource useful not only for native Portuguese speakers, but also non-native national and foreign speakers. Therefore, in addition to the lexicographic unit (also known as a lexicographic entry) and the lexicographic definition, we are including the phonetic transcription, the grammatical categories, and the context, in addition to additional historical and linguistic information.

The following is an example of a lexicographic record. For each record (also called an entry) there is a lexicographic entry, organised overall in alphabetical order.

LEXICOGRAPHIC ENTRY
Phonetic transcription
Grammatical categories
Lexicographic definition
Context

The lexicographic entry field designates the *definiendum* [element to be defined]. The phonetic transcription field is the representation of speech sounds, in phonetic symbols, modelled on the works I published in 2016, 2017, and 2018, respectively (Undolo 2016, 2017, 2018). The grammatical category field gives minimal morphological and/or morpho-syntactic features for each lexicographic unit, in their contexts of use. The lexicographic definition field states the *definiendum* in question. The context field, extracted from our corpus, illustrates the *definiendum* indicated in its situational use, in order to facilitate understanding of it.

LEXICOGRAPHIC ENTRY	KWANZA
Phonetic transcription	['kwāza]
Grammatical categories	m.n
Lexicographic definition	1. national currency. 2. money
Context	Formal, official. _____ The kwanza fell to its lowest State value ...

NB: The abbreviation <m.n> means “masculine noun”.

DELINEATION OF THE OBJECT OF ANALYSIS

We plan to develop a monolingual dictionary of common language based on the corpus, using contrastive and semasiological methods, and seek to highlight a sample of the “lexical reservoir of Angolan culture”. Vilela (1995) argues that the lexicon of a language is not a patchwork, but

nor is it a homogeneous whole; it constitutes what we usually call the diasystem: everyday words coexist with specialist words, spoken language words live side by side with those of the written language, old words coexist with archaisms and neologisms. Indeed, as we can see from the systematic results of linguistic research carried out as part of the VAPA project, neologisms are one of the most common phenomena in this variety. In general, in linguistic studies focused on variation and change, neologisms are not only a phenomenon at the lexical level, but also at the phonetic, phonological, morphological, and syntactic levels.

Neologisms can be explained through neology, a field of lexicology that deals with (re)creativity and/or lexical innovation. According to Mendes (2000, p. 55), “we can find new lexical units, either in the “common language” or in the “speciality language”, which can be called “language neologisms” or “terminological neologisms”. Current language neology is characterised by spontaneity and is dependent on the “use” and “collective consciousness of the community”. In our work, we have noted the following processes that form part of (re)creativity and/or lexical innovation:

- formal neology (morphological neologisms):
 - *derivation* (involving simple lexical units):
 - prefixation: <desconseguir>;
 - suffixation: <zungueira>;
 - abbreviation:
 - acronyms: <UNTA> (União Nacional dos Trabalhadores Angolanos [National Union of Angolan Workers]), <EPAL> (Empresa Pública de Água [Public Water Company]), <Zap> (satellite TV operator);
 - initialisms: <BNA> (Banco Nacional de Angola [National Bank of Angola]), <BPC> (Banco de Poupança e Crédito [Bank of Savings and Credit]), <RNA> (Rádio Nacional de Angola [National Radio of Angola]);
 - truncation: <moto> (motorised).
 - *composition* (involving composite and complex units):
 - morphosyntactic: <girabola> (football league, first division), <soba-grande> (traditional authority), <kixi-crédito> (bank service or product), <janela-aberta > (“open-window”, an informal, home-based commercial establishment).
 - *lexical importation*:
 - external or interlinguistic loans (also known as foreignisms): <cupapata> (motorbike taxi driver), <alambamento> (traditional marriage), <quilape> (monetary loan of a certain amount), <dipanda> (independence).
- semantic neology:
 - polysemy: <saldo> (in the context of mobile phones, this means the “telecommunications rate unit”), <anexo> (residence, housing).
 - metaphor: <janela-aberta>.
 - metonymy: <força de trabalho> (workers).

This lexicographic project aims to help manage the Portuguese language, and to have an effect on both interpersonal and interorganisational communication. For Angola, as a member state of the Community of Portuguese Language Countries and other regional organisations, it is of paramount importance to have linguistic resources that convey elements of the country’s culture, economy, politics, and religion to be able to ensure effective intercultural communication.

FINAL CONSIDERATIONS

The importance of a dictionary in the life of a people lies in the fact that it serves as an organised reservoir of their culture and the natural world that surrounds them, encompassing the most diverse aspects of a society’s existence and providing systematic, consistent descriptions of language units.

The monolingual dictionary of common language that we are proposing already includes the phonetic transcription of each lexicographic unit, in order to allow national and, above all, foreign speakers, to deepen their understanding of phonetic-phonological elements. This is a resource that we believe will be useful not only for the study and understanding of the Angolan variety of Portuguese, but principally for language management work in Angola, bearing in mind that this language is an important tool for fully engaging with life in a wide variety of sectors and contexts.

Among the various processes of enrichment of the Angolan Portuguese lexicon, it is important to mention that the importation of Bantu into Portuguese is not the main source. For many years researchers believed that formal neologisms, amongst other features, were mostly the result of linguistic interference or language contact, ignoring other processes inherent to each language's system. On the one hand, the discredit attached to certain linguistic constructions, whether in the phonetic/phonological domain or in the lexical or even syntactic and morphological domains, gives rise to the formation of a prestige norm. On the other hand, on the practical level of the language, in the Angolan sociocultural context the prestige norm is not a faithful reproduction of the Portuguese of Lisbon (the Portuguese capital), because speakers do not perform the language in the same way in different geographic spaces (or geo-cultural) spaces. Rather, it tends to be the Portuguese of Luanda, which manifests itself as a vertical dialect, marked by the local ruling class, and by the concentration of political, economic, and cultural power in the capital city, to the detriment of other regions of the country. Therefore, this dialect has become the centripetal force in social relations, given the diversity of uses of Portuguese throughout the national territory. Indeed, radio, television, newspapers, and magazines serve as the main means of influencing the linguistic behaviour of language users.

Formal or informal communication context serves as a criterion for the selection of texts that constitute the corpus based on which we can also assess the <+ educated> or <+ popular> characteristics of given lexical units so that we can then determine the norm they belong to, whether educated or popular. Knowledge about a norm establishes the boundaries in terms of how it relates to other norms. We know that studying an educated norm paves the way for the relevant state bodies to determine the linguistic standard in the future. In this sense, any decision must, above all, take into account descriptive, systematic, and exhaustive studies.

In short, the shifts and changes experienced by the Portuguese language in Angola distinguish it from other varieties of this language which are spoken in different nations. In this sense, the immediate function of the Angolan variety of Portuguese as a communication tool in Angola is more effective/efficient than is the case in Portugal or Brazil, for example. Indeed, a commitment to researching and managing the language can contribute to the “empowerment” of its users, in its most diverse usage contexts. With this project, we will hopefully be able, at least on a lexical level, to create an increasingly comprehensive dictionary of Angolan Portuguese.

ACKNOWLEDGEMENTS

The translation of this article was made possible thanks to funding from the University Council of Modern Languages and the University of Liverpool.

AUTHOR AFFILIATIONS

Márcio Undolo  orcid.org/0000-0001-6735-5357

Doutor em Linguística pela Universidade de Évora; Professor Associado do Instituto Superior de Ciências da Educação de Benguela (Angola), AO

REFERENCES

- Azevedo, D. J. O. de, & Silva, F. M. da (2017). Colocações, estereótipos e clichês: definições e diferenças. *ReVEL*, 15(29). <http://www.revel.inf.br/files/b4f9129fed69407b4ec7c6e747aa1c5f.pdf>
- Brocardo, M. T. (2014). *Tópicos de história da língua Portuguesa*. Edições Colibri.
- Camara Jr., J. M. (2007). *Dicionário de linguística e gramática* (26th ed.). Editora Vozes.
- Castro, I. (2006). Norma linguística e ensino do Português. *Caderno Escolar, Pensar à Escola*, 3, 30–34.
- Dicionário da Língua Portuguesa. (1793). Lisboa: Academia Real das Ciências de Lisboa.

- Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa. (2001). Lisboa: Academia das Ciências de Lisboa e Editorial Verbo.
- Emiliano, A. (2009). *Fonética do Português Europeu: descrição e transcrição*. Guimarães Editores.
- Farias, M. A. R. de (2016). *Antropologia linguística & etnografia toponímica: vivências e narrativas em linguagens socioculturais de Murinin-Benevides-Pará* [Unpublished doctoral dissertation]. Universidade Federal do Pará.
- Freitas, T., Ramilo, M. C., & Arim, E. (2005). O Português Europeu e os meios de comunicação. In H. M. Mateus (Ed.), *A língua Portuguesa em mudança* (pp. 33–36). Caminho.
- Halliday, M. A. K., & Hasan, R. (1989). *Language, context and text: Aspects of language in a social-semiotic perspective*. Oxford University Press.
- Henriques, C. C. (2011). *Léxico e semântica*. Elsevier.
- Lusakalalu, P. (2005). *Língua e unidades glossonímicas*. Editorial Nzila.
- Martinet, A. (2014). *Elementos de linguística geral*. Clássica Editora.
- Mateus, H. M. (2005). A mudança da língua no tempo e no espaço. In H. M. Mateus (Ed.), *Língua Portuguesa em mudança* (pp. 14–30). Caminho.
- Melo Bacelar, B. (1783). *Diccionario da Língua Portuguesa*. Lisboa: Aquino Bulhoens.
- Mendes, I. (2000). *O Léxico no Português de Moçambique: aspectos neológicos e terminológicos*. Promédia.
- Mingas, A. A. (1998). *O Português em Angola: reflexões*. Paper presented at the VIII Encontro da Associação das Universidades de Língua Portuguesa, Macau.
- Miranda, V. M. R. de (2013). *Léxico e cultura: estudo linguístico na área rural de Sabinópolis-MG* [Unpublished doctoral dissertation]. Universidade Federal de Minas Gerais.
- Morais Silva, A. (1789). *Grande Diccionario da Língua Portuguesa, composto pelo Padre D. Rafael Bluteau, reformado e acrescentado*. Lisboa: Ferreira.
- Nzau, D. G. N. (2011). *A língua Portuguesa em Angola: um contributo para o estudo da sua nacionalização* [Unpublished doctoral dissertation]. Universidade da Beira Interior.
- Oliveira, L. P. de (2009). Linguística de corpus: teoria, interfaces e aplicações. *Matraga*, 16(24), 1–20.
- Quivuna, M. (2014). *Lexicologia aplicada ao ensino do léxico em Português língua não materna*. Edições Colibri.
- Raposo, E. P., Vicente, G., & Veloso, R. (2013). *Gramática do Português*. Fundação Calouste Gulbenkian.
- Rey-Debove, J. (1984). Léxico e dicionário. *Alfa: Revista de Linguística*, 28(supp.), 45–69.
- Undolo, M. (2016). Estabelecimento do “plural vazio” no sintagma nominal em Português Angolano contemporâneo. *Revista LUCERE da Universidade Católica de Angola*, 177–188.
- Undolo, M. (2017). Elementos da norma culta do Português Angolano: sistemas vocálico e consonântico. In A. Oliveira Figueiredo (Ed.), *A língua Portuguesa na Índia e em outras terras*. Droit Publisher.
- Undolo, M. (2018). *Língua Portuguesa: subsídios para o seu ensino*. Edições ECO7.
- Vilela, M. (1995). *Léxico e gramática*. Almedina.

TO CITE THIS ARTICLE:

Undolo, Márcio 2022
Proposal for the Creation of a Portuguese Dictionary: An Ongoing Project Based on Corpus Analysis of the Angolan Variety. *Modern Languages Open*, 2022(1): 19, pp. 1–12.
DOI: <https://doi.org/10.3828/mlo.v0i0.403>

Published: 14 November 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Modern Languages Open is a peer-reviewed open access journal published by Liverpool University Press.