

3º CICLO DE ESTUDO EM CIÊNCIAS DA LINGUAGEM
ÁREA DE ESPECIALIZAÇÃO EM TECNOLOGIAS DA LINGUAGEM HUMANA

Estudo Comparativo de Formas Lexicais e Estruturas Passivas em Português de Angola e Europeu com o Auxílio de Técnicas de *Text Mining* e Processamento de Linguagem Natural

TIMÓTEO SUMBULA MUHONGO

D

2021



Timóteo Sumbula Muhongo

**Estudo Comparativo de Formas Lexicais e Estruturas Passivas em
Português de Angola e Europeu com o Auxílio de Técnicas de
Text Mining e Processamento de Linguagem Natural**

Tese realizada no âmbito do Doutoramento em *Ciências da Linguagem*,
orientada pela Professora Doutora Maria de Fátima Henriques da Silva
e pelo Professor Doutor Pavel Bernard Brazdil

Faculdade de Letras da Universidade do Porto

abril de 2021

Estudo Comparativo de Formas Lexicais e Estruturas Passivas em
Português de Angola e Europeu com o Auxílio de Técnicas de
Text Mining e Processamento de Linguagem Natural

Timóteo Sumbula Muhongo

Tese realizada no âmbito do Doutoramento em Ciências da Linguagem,
orientada pela Professora Doutora Maria de Fátima Henriques da Silva
e pelo Professor Doutor Pavel Bernard Brazdil

Membros do Júri

Aos meus pais, esposa e filho.

Índice

Declaração de honra	12
Agradecimentos.....	14
Resumo	17
Abstract	19
Índice de Ilustrações.....	21
Índice de Tabelas	24
Índice de Quadros	27
Lista de Abreviaturas e Siglas	29
Parte I – Enquadramento para o Estudo Comparativo na Área de Léxico e Sintaxe.....	31
Capítulo 1 – Introdução	33
1.1. Motivação e objetivos	33
1.2. Contextualização e enquadramento metodológico	35
1.3. Objetivos da investigação e metodologia	36
1.4. Organização da dissertação.....	37
1.5. Resultados atingidos.....	38
1.6. Publicações	38
Capítulo 2 – Situação Linguística de Angola: o português e as outras línguas de Angola ...	40
2.1. Localização geográfica	40
2.2. Línguas e surgimento do Português em Angola.....	44
2.3. Síntese do capítulo	52
Capítulo 3 – Fundamentos Teóricos em Lexicologia e Lexicografia	53
3.1. As Ciências do léxico: Lexicologia e Lexicografia.....	54
3.2. Língua geral e língua de especialidade.....	57
3.3. Transformação conceptual da dimensão neológica	59
3.3.1. Empréstimos lexicais	66
3.4. Síntese do capítulo	68
Capítulo 4 – Estrutura do Léxico e Processos de Formação de Palavras	69
4.1. Estrutura do léxico.....	69
4.2. Estrutura morfológica do Português.....	72
4.3. Processos de inovação lexical	75
4.4. Alomorfia	76
4.5. Processos de derivação em Português.....	76
4.5.1. Derivação afixal e subtipos semânticos	77

4.6. Conversão	82
4.7. Processos de construção não concatenativa.....	83
4.7.1. Cruzamento vocabular.....	84
4.7.2. Truncação	85
4.7.3. Reduplicação	86
4.7.4. Siglação e acronímia	87
4.8. Processos de composição em Português	88
4.8.1. Composição morfológica.....	89
4.8.2. Composição morfossintática	90
4.8.3. Composição sintagmática	91
4.9. Síntese do capítulo	92
Capítulo 5 – Estruturas Passivas e Estruturas com <i>Se</i>	93
5.1. Estruturas passivas	93
5.1.1. Passiva eventiva	94
5.1.2. Passiva resultativa	97
5.1.3. Passiva estativa	99
5.2. Análise dos valores de <i>se</i>	100
5.2.1. <i>Se</i> impessoal.....	101
5.2.3. <i>Se</i> anafórico	105
5.2.4. Alternância causativa: <i>se</i> decausativo.....	109
5.3. Síntese do capítulo	110
Capítulo 6 – Trabalhos Relacionados com o Estudo Comparativo	112
6.1. Estudos comparativos na área do Léxico.....	112
6.2. Estudos comparativos na área da Sintaxe	113
6.3. Estudos comparativos em Linguística Computacional.....	114
6.4. Síntese do capítulo	115
Parte II – Enquadramento na Área de Linguística Computacional	117
Capítulo 7 – Técnicas de <i>Text Mining</i> e Processamento de Linguagem Natural	119
7.1. <i>Text Mining</i> e Processamento de Linguagem Natural	119
7.1.1. Tipos de documentos textuais	120
7.2. Técnicas básicas de <i>Text Mining</i>	121
7.2.1. <i>Tokenization</i>	122
7.2.2. Representação de textos	123
7.2.3. Lematização e <i>stemming</i>	124
7.2.4. Normalização de texto: conversão para letra minúscula	125
7.3. Técnicas avançadas de <i>Text Mining</i>	126
7.3.1. <i>Part-of-speech tagging</i>	126
7.3.2. Identificação de entidades nomeadas	127
7.3.3. <i>Dependency parsing</i>	129
7.3.4. Expressões regulares	131
7.4. Medidas de desempenho do método automático	132
7.4.1. Medidas para problemas de classificação	132

7.4.2. Medidas para problemas de regressão.....	134
7.4.3. Testes estatísticos	134
7.5. Síntese do capítulo	136
Parte III – Técnicas de <i>Text Mining</i> e Processamento de Linguagem Natural em Lexicologia Contrastiva.....	138
Capítulo 8 – Metodologia de Processamento de Dados para a Detecção de Empréstimos... 140	
8.1. Metodologia adotada – visão geral.....	140
8.2. Constituição e caracterização de corpora.....	142
8.2.1. Constituição do corpus de Telejornal de Angola	145
8.2.2. Sumário de corpora usados.....	146
8.3. Importação e pré-processamento de dados	148
8.3.1. Importação de dados e constituição de corpus em R.....	149
8.3.2. Pré-processamento de dados	149
8.4. Matriz Documento-Termo e Termo-Documento.....	152
8.4.1. Frequência de alguns termos	154
8.4.2. Visualização de termos: <i>barplot</i> e <i>word cloud</i>	155
8.5. Detecção de candidatos a empréstimos	156
8.6. Síntese do capítulo	158
Capítulo 9 – Extração de Empréstimos Lexicais e Estudo Comparativo	159
9.1. Detecção de empréstimos verbais.....	159
9.1.1. Identificação de elementos lexicais em textos do Português de Angola.....	160
9.1.2. Extração de <i>lema</i> e classe lexical de elementos lexicais (<i>tokens</i>).....	161
9.1.3. Uso do léxico do Português Europeu.....	164
9.1.4. Extração de candidatos a empréstimo	164
9.1.5. Redução de candidatos a empréstimo usando empréstimos existentes	165
9.2. Identificação de empréstimos no subconjunto de candidatos	166
9.3. Extração de empréstimos de outras classes lexicais	168
9.3.1. Extração de empréstimos nominais	169
9.3.2. Extração de empréstimos adjetivais	171
9.3.3. Extração de empréstimos de nomes próprios	172
9.4. Sumário de resultados sobre os empréstimos	174
9.5. Processamento incremental de textos.....	174
9.6. Vantagens do processo automático	175
9.7. Extração de cotextos de empréstimos	176
9.8. Síntese do capítulo	182
Capítulo 10 – Organização do Subconjunto de Empréstimos Lexicais	184
10.1. Categorização lexical e etimologia dos empréstimos	184
10.2. Estruturação dos empréstimos em campos lexicais.....	188
10.3. Processos de formação dos empréstimos.....	190
10.3.1. Empréstimos formados por derivação.....	190

10.3.2. Empréstimos formados por composição	193
10.3.3. Empréstimos por construção não concatenativa	194
10.4. Protótipo de Dicionário de Regionalismos Angolanos	195
10.4.1. Proposta de modelo de dicionário	195
10.4.2. Ficha lexicográfica de empréstimos	195
10.5. Síntese do capítulo	197
Parte IV – Análise de Estruturas Passivas com Técnicas de <i>Text Mining</i> e Processamento de Linguagem Natural	199
Capítulo 11 – Estruturas Passivas em Português de Angola e Europeu	201
11.1. Metodologia para o estudo de estruturas predicativas	201
11.1.1. Identificação de frases no corpus Telejornal	201
11.2. Extração de passivas	204
11.2.1. Extração de sintagmas com verbo auxiliar e principal	207
11.2.2. Extração de sintagmas com três elementos	209
11.2.3. Análise de estruturas passivas eventivas e estativas	210
11.2.4. Avaliação do processo de extração	213
11.2.5. Estruturas passivas resultativas	214
11.3. Análise comparativa de ocorrências de passivas	215
11.3.1. Aplicação do teste estatístico de proporções	216
11.4. Análise comparativa de ocorrências de tipos de passiva	216
11.5. Síntese do capítulo	218
Capítulo 12 – Análise das Estruturas Passivas	219
12.1. Relação de dependência entre os constituintes	219
12.2. Comparação das estruturas passivas	226
12.3. Síntese do capítulo	231
Capítulo 13 – Conclusões e Trabalho Futuro	232
Referências	236
Anexos da Tese	247
Anexo I – Cotexto de empréstimos nominais, verbais e adjetivais	249
Anexo II – Lista de empréstimos nominais próprios	293
Anexo III – Estruturas passivas do Telejornal de Angola	333
Anexo IV – Estruturas passivas do Telejornal de Portugal	342

Declaração de honra

Declaro que a presente Tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição e encontram-se devidamente indicadas no texto e nas referências bibliográficas de acordo com as normas de referência. Tenho naturalmente consciência de que a prática de plágio e auto plágio constituem um ilícito académico.

Porto, abril de 2021

Timóteo Sumbula Muhongo

Agradecimentos

Gostaríamos de inevitavelmente começar por expressar os nossos mais profundos agradecimentos ao Todo e a todas as entidades pelo adminículo e empenhamento pessoal adjutórios que tornaram possível a exequibilidade desta reflexão.

Primeiramente, um muito especial à Fundação Calouste Gulbenkian pela concessão da bolsa de estudo e formações que nos deu, as quais sisudamente propiciaram esta culminação. É neste contexto que aproveitamos também o ensejo para remerrear à Doutora Margarida Abecasis por todas as pronunciações sapienciais, com as quais pudemos de facto chegar a resoluções maduramente tomadas. Estamos muito gratos pela proficiência e qualidade de trabalho a que a FCG nos habituou e nos incentivou a ter.

Gratulamos imenso aos nossos orientadores, Professora Doutora Maria de Fátima Henriques da Silva e Professor Doutor Pavel Bernard Brazdil, pela tão atenta, intensa, empenhada e exultante orientação, pela compreensão, fiúza e perspicácia demonstradas com quem pudemos, por um lado, sentir o demiurgo e a dualidade primordial, e depreender que, tal como não se faz um médico sem o auxílio de um médico, não se faz um linguista nem um informático sem o auxílio de outro linguista e informático e, sobretudo, pelo incentivo à melhora contínua, de maneira que houve de facto ação local com visão global. Por outro, verificamos que pode haver uma harmonia, uma correlação entre as áreas do conhecimento. Reconhecemos o quão interventivos foram e estamos certos de que esta perscrutação é, decerto, uma colheita e uma sementeira. Com a trimúrta de amor, vontade e inteligência é possível notar o *solve et coagula*, enunciado no livro mudo.

Cabe-nos o dever e a responsabilidade de agradecermos ao LIAAD – INESC TEC por todo apoio em congresso demonstrado, particularmente à Professora Doutora Conceição Rocha e ao Professor Doutor Alípio Jorge pelos exímios materiais e sugestões das quais jamais nos olvidaremos. A nossa profunda gratidão à Universidade de Santiago de Compostela, particularmente ao Professor Doutor Pablo Gamallo, a quem devemos muito respeito e consideração pelo fornecimento do Léxico do Português Europeu – elemento fulcral desta investigação. Ao CLUP por todo apoio demonstrado e consentir, por unanimidade, a nossa integração na sua equipa de investigação. Aos Professores Doutores Ana Maria Brito, Fátima Oliveira e João Veloso pelo apoio científico que nos coadjuvam para a egrégora académica, que permite que se efetive a metanoia e a dianoia. O nosso agradecimento é extensivo à AOTP, APL e à Câmara Municipal do Porto.

À Domingas Sumbula, Manuel Lopes Muhongo, à Luísa Muhongo e ao Reuel Carlos

Muhongo, pois, se aquilo que o homem faz pode ecoar pela eternidade, aquilo que é deixará frutos no futuro e tem sementes no passado. Além disso, aquilo que é não pode deixar de ser nem no tempo nem no espaço. Aos nossos irmãos e aos Quindombes, pois pudemos sentir o Tau e constatar que as coisas são o Todo temporariamente condensado e que nada que é, realmente, nosso pode ser tirado de nós, certificando-nos de que somos fator de soma na vida daqueles em que participamos.

Resumo

O objetivo deste trabalho é comparar formas lexicais e estruturas passivas eventivas, estativas e resultativas em Português de Angola e Português Europeu. Esta investigação situa-se na interface entre a Linguística e as Ciências da Computação e explora técnicas de Lexicologia Contrastiva, Sintaxe, *Text Mining* e Processamento de Linguagem Natural.

Este estudo tem dois objetivos. O primeiro insere-se na área de neologia e foca em empréstimos que promanam de línguas de Angola. Constituímos os corpora de extração de modalidade escrita e oral que foram transcritos com o auxílio de métodos computacionais para poder avançar com o processamento seguinte. O nosso objetivo era identificar, mediante os corpora à nossa disposição, os candidatos a empréstimos dimanantes de Angola por classe lexical (verbos, nomes, adjetivos e nomes próprios).

Quanto aos métodos computacionais, usamos diversos métodos, incluindo *part-of-speech tagging*, *named entity recognition*, *dependency parsing* e *regular expressions*, disponíveis em linguagem R e alguns pacotes específicos orientados para *Text Mining* e Processamento de Linguagem Natural. Esta análise revelou que alguns empréstimos não constam dos dicionários de referência. Além disso, apresentamos o protótipo de dicionário de regionalismos angolanos. Os termos são acompanhados pelos respetivos cotextos, que poderão ser úteis para trabalho futuro.

O nosso segundo objetivo era analisar as estruturas passivas em Português de Angola e Português Europeu e, em particular, as estruturas passivas eventivas, estativas e resultativas, que foram extraídas automaticamente dos nossos corpora. Analisamos as proporções de estruturas passivas e frases em Português de Angola e Português Europeu. O resultado principal deste estudo é a constatação de que as estruturas passivas são mais frequentes em Português de Angola do que em Português Europeu, particularmente no corpus oral analisado. O teste estatístico de proporções indicou que as diferenças são estatisticamente significativas.

Acreditamos ter contribuído para o estado da arte na área da neologia e da descrição das estruturas passivas em Português de Angola, visto que a metodologia proposta pode ser reaproveitada noutros contextos, com outros corpora de Português de Angola e Europeu. Como mostramos, isso pode ser feito de forma incremental e, dessa forma, poupar muito esforço manual na elaboração de novas versões de um dicionário de empréstimos e estudo de estruturas passivas.

Palavras-chave: *Text Mining*, Processamento de Linguagem Natural, Neologia, Empréstimos, Métodos Computacionais em Lexicologia Contrastiva, Estruturas Passivas.

Abstract

Our aim in this dissertation was to compare the lexical forms, as well as eventive, stative and resultative passive structures in Angolan and European Portuguese. This study is at the interface between Linguistics and Computer Science and exploits techniques from the areas of Contrastive Lexicology, Syntax, Text Mining and Natural Language Processing. This research allow us to highlight some substantial differences between Angolan and European Portuguese.

The study has two aims. The first one is concerned with neology and focuses on loanwords originating from Angolan languages. We use different Angolan and European Portuguese corpora, some of which were originally oral form. These were transcribed into text format to permit further processing. Our objective was to identify loanwords originating from Angolan languages and categorize them by lexical class (verbs, nouns, adjectives, and proper nouns). As for computational methods, we have used a set of methods, including *part-of-speech tagging*, *named entity recognition*, *dependency parsing* and *regular expressions* in R programming language and some specific packages oriented towards Text Mining and Natural Language Processing. This analysis revealed that a significant proportion of loanwords does not actually appear in existing reference dictionaries. In addition, we present a prototype dictionary of Angolan regionalisms. These are accompanied by examples of cotexts, which can be useful for follow-up studies.

Our second aim was to analyse the passive structures in Angolan and European Portuguese with the focus on eventive, stative and resultative structures. These were extracted automatically from the available corpora. We have analysed the proportions of passive structures and phrases in Angolan and European Portuguese. Our main finding was that the passive structures were much more common in Angolan Portuguese than in the European one, particularly when considering the oral corpora. A statistical test of proportions indicated that the difference is statistically significant.

We believe that we have contributed to the state-of-the art in the area of neology and syntax of passive structures, as the methodology proposed can be reused in other settings, with other corpora Angolan and European Portuguese. As we have shown this can be done in an incremental way and this way save a great deal of manual effort when preparing new versions of the dictionary of loanwords and study of passive structures.

Keywords: Text Mining, Natural Language Processing, Neology, Loanwords, Computational Methods in Contrastive Lexicology, Passive Structures.

Índice de Ilustrações

Fig. 2.1 – Mapa da divisão político-administrativa de Angola, United Nations (2020).....	41
Fig. 2.2 – Mapa sobre a distribuição <i>bantu</i> em Angola, Pires (2020).....	42
Fig. 2.3 – Mapa etnográfico de Angola, Dias (2003).....	43
Fig. 2.4 – Situação linguística de Angola, Ethnologue (2020).....	45
Fig. 2.5 – Distribuição das línguas <i>bantu</i> por zonas geográficas, Wikimedia (2021).....	46
Fig. 2.6 – África antes e após a Conferência de Berlim, Lynch (2020).....	50
Fig. 3.1 – Configuração da dimensão neológica.....	65
Fig. 3.2 – Relação de dependência sintática de <i>Humbes</i>	67
Fig. 4.1 – Esquema da estrutura do léxico.....	71
Fig. 5.1 – <i>Tokenisation, POS Tagging e Dependency Relation</i> de estrutura recíproca.....	108
Fig. 7.1 – Procedimentos para Ciência de Dados, Wickham & Grolemund (2017).....	121
Fig. 7.2 – <i>Part-of-speech tagging</i> de estrutura ativa.....	127
Fig. 7.3 – Esquema de abordagem de <i>Machine Learning</i> para NER, Konkol (2015).....	128
Fig. 7.4 – Extração de entidades nomeadas em <i>A Conjura</i>	129
Fig. 7.5 – Um exemplo de relações de dependência sintática.....	130
Fig. 7.6 – Exemplo de relações de dependência sintática com <i>spaCy</i>	131
Fig. 7.7 – Marcação de verdadeiro e falso positivo, Konkol (2015).....	134
Fig. 8.1 – Representação da tríade de Processamento de Linguagem Natural.....	140
Fig. 8.2 – Gravação do corpus Telejornal usando o <i>Audacity</i>	145
Fig. 8.3 – Transcrição automática do corpus Telejornal usando o <i>Dictate</i>	146
Fig. 8.4 – Procedimentos de <i>Text Mining</i> aplicados ao corpus, adaptado de Edureka (2019).....	148
Fig. 8.5 – Importação de dados e criação de corpus de <i>A Conjura</i>	149
Fig. 8.6 – Matriz Documento-Termo de <i>A Conjura</i>	153
Fig. 8.7 – Matriz de Termos do corpus <i>Jornal de Angola</i> com remoção de <i>stop word</i>	154
Fig. 8.8 – Frequência de termos do corpus <i>A Conjura</i>	154
Fig. 8.9 – Frequência de termos do corpus <i>Jornal de Angola</i>	155
Fig. 8.10 – <i>Barplot</i> dos termos mais frequentes em <i>A Conjura</i>	155
Fig. 8.11 – Tokenização do corpus <i>A Conjura</i>	156
Fig. 8.12 – Candidatos a empréstimo lexical do corpus <i>A Conjura</i>	156
Fig. 8.13 – Frequência falsa dos candidatos a empréstimo do corpus <i>A Conjura</i>	157
Fig. 8.14 – Frequência dos candidatos a empréstimo do corpus <i>A Conjura</i>	157
Fig. 8.15 – <i>Word cloud</i> de candidatos a empréstimo do corpus <i>A Conjura</i>	157
Fig. 9.1 – Representação de extração de empréstimo	161
Fig. 9.2 – Tokenização e <i>Part-of-speech Tagging</i> de <i>A Conjura</i>	161
Fig. 9.3 – Seleção das variáveis <i>lemma</i> e <i>upos</i>	161
Fig. 9.4 – Extração de alguns lemas verbais de <i>A Conjura</i>	162
Fig. 9.5 – Supressão de <i>NA</i> e dimensão do <i>data frame</i> de verbos.....	162
Fig. 9.6 – Uso do Léxico do Português Europeu.....	164
Fig. 9.7 – Extração de candidatos a empréstimos nominais.....	169
Fig. 9.8 – Extração de adjetivos candidatos a empréstimos.....	171
Fig. 9.9 – Extração de candidatos a empréstimos.....	172
Fig. 9.10 – Análise <i>duma</i> das frases com o empréstimo <i>maka</i>	177
Fig. 9.11 – Extração de cotexto simplificado do empréstimo <i>maka</i>	178

Fig. 9.12 – Extração de frase com empréstimo verbal.....	181
Fig. 9.13 – Extração de cotextos de empréstimos verbais.....	181
Fig. 10.1 – Algumas línguas de Angola (Edmundo, 2020).....	186
Fig. 10.2 – Ficha lexicográfica da entrada <i>Maka</i>	196
Fig. 10.3 – Ficha lexicográfica da entrada <i>Machila</i>	197
Fig. 11.1 – Tokenização e <i>Part-of-speech Tagging</i> de Telejornal de Angola.....	203
Fig. 11.2 – Estrutura com <i>NA</i> do corpus Telejornal de Angola.....	204
Fig. 11.3 – Estruturas passivas do corpus Telejornal de Angola.....	205
Fig. 11.4 – Ciclo de extração de verbo auxiliar e principal.....	207
Fig. 11.5 – Ciclo de extração de verbo auxiliar e principal melhorado.....	208
Fig. 11.6 – Ciclo de extração de estruturas passivas estativas.....	208
Fig. 11.7 – Ciclo de extração de sintagmas com três elementos.....	209
Fig. 11.8 – Relação de dependência de estrutura passiva eventiva.....	211
Fig. 11.9 – Relação de dependência de estrutura passiva estativa.....	212
Fig. 11.10 – Relação de dependência de estrutura mal classificada.....	212
Fig. 11.11 – Extração do verbo ficar como auxiliar.....	214
Fig. 11.12 – Relação de dependência de estrutura passiva resultativa.....	215
Fig. 11.13 – Teste estatístico de proporções de sintagmas passivos.....	216
Fig. 12.1 – <i>Part-of-speech tagging</i> de estrutura passiva.....	219
Fig. 12.2 – <i>Dependency parsing</i> de passiva eventiva curta no PE	220
Fig. 12.3 – <i>Dependency parsing</i> de passiva eventiva curta no PA.....	220
Fig. 12.4 – <i>Dependency parsing</i> de passiva eventiva longa no PA.....	220
Fig. 12.5 – <i>Dependency parsing</i> de passiva eventiva longa no PE.....	220
Fig. 12.6 – <i>Dependency parsing</i> de passiva estativa no PA.....	220
Fig. 12.7 – <i>Dependency parsing</i> de passiva estativa no PE.....	220
Fig. 12.8 – Contextos do verbo remeter (Leipzig, 2020).....	230

Índice de Tabelas

Tabela 2.1 – Formação do plural em <i>kimbundu</i> , António e Osório (2018).....	47
Tabela 4.1 – Comparação da sufixação e flexão, adaptado de Arquiola (2009).....	80
Tabela 4.2 – Posição dos constituintes nos compostos morfológicos.....	89
Tabela 7.1 – Relação entre lematização e <i>stemming</i>	125
Tabela 7.2 – Sistematização para classes lexicais em <i>udpipe</i>	126
Tabela 7.3 – Principais mecanismos de <i>regular expression</i>	132
Tabela 7.4 – Matriz de confusão, adaptado de Gama, Faceli, Lorena e Oliveira (2017).....	133
Tabela 7.5 – Testes estatísticos não paramétricos, Teixeira, Fuccio e Oliveira (2020).....	135
Tabela 8.1 – Sumário do corpus do Português de Angola.....	146
Tabela 8.2 – Sumário do corpus do Português Europeu.....	147
Tabela 8.3 – Parâmetros de classificação do corpus.....	147
Tabela 8.4 – Frequência de ocorrência do corpus do PA por classe lexical.....	148
Tabela 9.1 – Quantificação dos lemas verbais nos textos de PA.....	163
Tabela 9.2 – Quantificação dos lemas sem repetição nos textos de PA.....	163
Tabela 9.3 – Quantificação dos candidatos a empréstimo por classe lexical.....	165
Tabela 9.4 – Empréstimos verbais em <i>A Conjura</i>	166
Tabela 9.5 – Empréstimos verbais falsos em <i>A Conjura</i>	167
Tabela 9.6 – Quantificação dos empréstimos verbais.....	167
Tabela 9.7 – Quantificação dos lemas sem repetição nos textos de PA.....	168
Tabela 9.8 – Quantificação dos candidatos a empréstimo por classe lexical.....	168
Tabela 9.9 – Empréstimos nominais em <i>A Conjura</i>	169
Tabela 9.10 – Empréstimos nominais falsos em <i>A Conjura</i>	170
Tabela 9.11 – Quantificação dos empréstimos nominais.....	170
Tabela 9.12 – Empréstimos adjetivais em <i>A Conjura</i>	171
Tabela 9.13 – Empréstimos adjetivais falsos em <i>A Conjura</i>	172
Tabela 9.14 – Quantificação dos empréstimos adjetivais.....	172
Tabela 9.15 – Empréstimos de nomes próprios em <i>A Conjura</i>	173
Tabela 9.16 – Empréstimos de nomes próprios falsos em <i>A Conjura</i>	173
Tabela 9.17 – Quantificação dos empréstimos de nomes próprios.....	174
Tabela 9.18 – Quantificação dos empréstimos por classe lexical.....	174
Tabela 9.19 – Taxa de redução de elementos lexicais por classe lexical.....	176
Tabela 9.20 – Cotextos de empréstimos nominais.....	179
Tabela 9.21 – Cotextos de empréstimos adjetivais.....	180
Tabela 9.22 – Cotextos de empréstimos verbais.....	182
Tabela 10.1 – Extrato da lista de empréstimos dimanantes de Angola.....	185
Tabela 10.2 – Quantificação das entradas dos empréstimos por etimologia.....	187
Tabela 10.3 – Empréstimos formados por derivação.....	191
Tabela 10.4 – Empréstimos formados por composição.....	193
Tabela 10.5 – Empréstimos por formação não concatenativa.....	194
Tabela 11.1 – Estruturas passivas do corpus Telejornal de Angola.....	205
Tabela 11.2 – Estruturas passivas do corpus Telejornal de Portugal.....	206
Tabela 11.3 – Matriz de confusão do Telejornal de Angola.....	214
Tabela 11.4 – Matriz de confusão do Telejornal de Portugal.....	214

Tabela 11.5 – Medidas de desempenho do método automático.....	214
Tabela 11.6 – Sumário de ocorrências de passivas.....	215
Tabela 11.7 – Sumário de ocorrências de sintagmas passivos por tipologia.....	216

Índice de Quadros

Quadro 5.1 – Passiva verbal: estrutura temático-argumental, Ribeiro (2011).....	95
Quadro 5.2 – Caraterização das componentes nas estruturas passivas.....	100
Quadro 5.3 – Estrutura temático-argumental da passivas de se, Ribeiro (2011).....	101
Quadro 5.4 – Diferenças entre a passiva eventiva, resultativa, estativa e pronominal.....	102
Quadro 5.5 – Estrutura temático argumental dos predicadores transitivos em estruturas transitivas plenas e estruturas de se nominativo, adaptado de Ribeiro (2011).....	103
Quadro 5.6 – Organização temático-argumental das estruturas reflexas, Ribeiro (2011).....	106
Quadro 5.7 – Organização temático-argumental das estruturas recíprocas, Ribeiro (2011).	107
Quadro 5.8 – Caraterização do SE anafórico, SE indeterminador e SE apassivador quanto à sua capacidade referencial, Ribeiro (2011).....	108

Lista de Abreviaturas e Siglas

A/ADJ – Adjetivo

COMP – Composto

DER – Derivado

DERIV – Derivacional

DP – Determiner Phrase

DTM – Document Term Matrix

EL-EMP – Elementos Lexicais Empréstimos

IT – Índice Temático

PREF – Prefixo

RA – Radical Adjetival

RN – Radical Nominal

RV – Radical Verbal

SN – Sintagma Nominal

SP – Sintagma Preposicional

SUF – Sufixo

SV – Sintagma Verbal

TA – Tema Adjetival

TDM – Term Document Matrix

TMA – Marcador de Flexão em Tempo-Modo-Aspetto

TN – Tema Nominal

TV – Tema Verbal

V – Verbo

VT – Vogal Temática

Parte I

Enquadramento para o Estudo Comparativo na Área de Léxico e Sintaxe

Capítulo 1 – Introdução

Este trabalho situa-se na interface entre a linguística e as ciências da computação. Pretendemos comparar formas lexicais e estruturas passivas em Português de Angola e Português Europeu com o auxílio de técnicas de *Text Mining* e Processamento de Linguagem Natural. Qualquer lucubração preambular sobre comparação das variedades do português e a sua existência enfrenta, inicialmente, a questão de indagar se é possível estabelecer fronteiras que as delimitam sob o ponto de vista lexical e sintático. Este estudo permite identificar as formas lexicais, mais concretamente palavras simples e algumas estruturas passivas que são mais frequentes na variedade do Português de Angola em comparação com as formas lexicais do corpus do Português Europeu, usado para comparação.

A nossa investigação advém da colaboração entre a Faculdade de Letras da Universidade do Porto, Centro de Linguística da Universidade do Porto, e da Faculdade de Economia do Porto, Laboratório de Inteligência Artificial e Apoio à Decisão (LIAAD) do Instituto de Engenharia de Sistemas e Computadores Tecnologia e Ciência (INESC TEC).

Este capítulo tem cinco objetivos. No primeiro momento, explicita as motivações da investigação desenvolvida. Visto que este estudo comparativo teve início com a pesquisa desenvolvida durante a dissertação de mestrado, discorre sobre os avanços feitos nesta investigação em relação à dissertação do mestrado. No segundo momento, apresenta os objetivos da presente investigação. No terceiro momento, descreve a contextualização da investigação e o enquadramento metodológico. No domínio da linguística, optamos pela metodologia semasiológica, tendo como fundamento a Teoria Comunicativa da Terminologia, que consideramos apropriada para o estudo de variantes de uma língua. No domínio da computação, cingimo-nos a procedimentos de técnicas de *Text Mining* e Processamento de Linguagem Natural. No quarto momento, descreve a organização da dissertação. Por fim, menciona os resultados alcançados.

1.1. Motivação e objetivos

Na sequência de frequentes ponderações sobre a formação, o funcionamento e as motivações subjacentes aos processos de formação de formas lexicais, surgiu o presente estudo comparativo de formas lexicais e estruturas passivas em Português de Angola e Português Europeu com o auxílio de técnicas de *Text Mining* e Processamento de Linguagem Natural. Pelo que referimos, podemos notar que este trabalho é fundamentalmente um cruzamento quadrífido de áreas de conhecimento: Lexicologia Contrastiva, Sintaxe, *Text Mining* e Processamento de Linguagem Natural.

A investigação tem como objeto uma comparação de formas lexicais e estruturas passivas nas variedades do Português de Angola (PA) e Português Europeu (PE), com o objetivo de prospear as especificidades do Português de Angola, incluindo alguns empréstimos lexicais dimanantes de sistemas linguísticos angolanos que podem ser encontrados nos escritos que retratam hodiernamente a realidade linguística angolana; é a esta que damos proeminência. Para isso, cingimo-nos a critérios como fiabilidade e representatividade. Constituímos e descrevemos as medidas de estatística descritiva do corpus de extração composto por vários documentos, incluindo um livro, transcrição de telejornais e léxico de Português Europeu, do qual, com o auxílio de métodos computacionais, se extraíram mais de um milhar de empréstimos dimanantes de línguas de Angola. Estes dividem-se em verbos, nomes, adjetivos e nomes próprios. Quantificamos os empréstimos por etimologia, étimo, e organizamo-los em campos lexicais. Extraímos o contexto de empréstimos nominais, verbais e adjetivais. Como produto final da investigação, além desta análise, propomos um protótipo de dicionário de regionalismos angolanos que possa desempenhar uma função didática.

Como se poderá notar, verificamos ainda os processos de formação subjacentes aos empréstimos, nomeadamente a derivação, composição e formação não concatenativa. Deste modo, detetamos os contextos de empréstimos verbais, nominais e adjetivais do corpus e verificamos as coocorrências, o que nos permitiu perceber as combinatórias mais frequentes no corpus com o auxílio de métodos computacionais.

Por fim, engendramos um programa que deteta estruturas passivas. Com efeito, mensuramos o desempenho do método computacional. Procedemos à verificação das proporções de passivas no Português de Angola e Português Europeu. Verificamos que as proporções de passivas em PA eram mais frequentes do que em PE. A questão é se este fenómeno foi influenciado por alguma das línguas autóctones.

A motivação da nossa investigação é dupla. Por um lado, consideramos que há uma lacuna no conhecimento sobre o léxico do Português de Angola, nomeadamente no que se refere à sua descrição e à constituição de dicionários que beneficiem de estudos de corpora sistemáticos. Além disso, constatamos que é um estudo antessignano, em Angola e Portugal, que conglobera um estudo de léxico e estruturas passivas em Português de Angola e Português Europeu com o auxílio de técnicas de *Text Mining* e Processamento de Linguagem Natural e aplica teste estatístico de proporções de modo que possamos mensurá-lo.

Associado a esta falta de estudos, consideramos também a necessidade de engendrar precisamente um protótipo de um dicionário de regionalismos angolanos que, com um pendor didático, possa auxiliar quer os estudantes e professores em Angola quer investigadores na

compreensão das especificidades do Português de Angola. Advogamos que as formas lexicais são o reflexo de realidades histórico-culturais e, como património de valor inestimável, merecem, sob este ponto de vista, alguma reflexão. Com efeito, partindo do princípio de que a língua, cultura e história formam de facto um trinómio especial em lexicática, podemos relacioná-lo na descrição e compreensão do funcionamento de formas lexicais do Português de Angola e Português Europeu.

1.2. Contextualização e enquadramento metodológico

Para a resolução da lucubração tratada, privilegiamos o quadro teórico da Teoria Comunicativa da Terminologia (TCT) proposta por Cabré (1993), em detrimento da Teoria Geral da Terminologia (TGT) proposta por Wüster (1998). Apesar de dimanar do Círculo de Viena¹, primar por uma língua universal e uniformidade da comunicação, ser internamente coerente, baseada num sistema logicista inquestionável na documentação e Inteligência Artificial, a opção pela TGT tem implicações no tipo de resultado que pretendemos obter. Tratando-se de situação de comunicação natural, de base social, com pretensões identitárias, a TGT revela-se insuficiente dado o seu carácter prescritivo.

Deste ponto de vista, em virtude da realidade sociolinguística de Angola, o estudo de formas lexicais do Português de Angola, numa vertente sincrónica, ou mais precisamente, na análise das formas lexicais e os seus elementos constitutivos no seu desenvolvimento histórico, tomando em consideração a forma e o conteúdo, é recomendável a metodologia semasiológica, a qual parte do termo para o conceito e permite-nos explicar a variação tanto denominativa como conceptual das formas lexicais.

Tendo como fundamento matricial interdisciplinar a Teoria do Conhecimento, que permite explicar como se conceptualiza a realidade, os tipos de conceptualização, a relação dos conceitos entre si e as suas possíveis denominações, a Teoria da Comunicação, que permite notar a correlação entre o tipo de situação e o tipo de comunicação em toda a sua amplitude e diversidade, e a Teoria da Linguagem, que permite notar as unidades terminológicas dentro da linguagem natural, singularizando o seu carácter terminológico e explicando como este se ativa na comunicação, a opção pela TCT, tanto do ponto de vista

¹ Tendo como principais representantes Carnap, Neurath e Wittgenstein, o Círculo de Viena é uma comunidade de investigadores reunidos em 1929, em torno do manifesto intitulado *A Conceção científica no mundo*, que anuncia uma ciência unitária fundada numa metodologia comum. Ela reivindica, assim, a tradição positivista, advogando o positivismo lógico, e recusa, precisamente, a metafísica e a tese segundo a qual o conhecimento se constrói com base em experiências sensoriais, de que a linguagem matemática traduz os dados (Besnier, 1996).

teórico como metodológico, permite prospetar a variação linguística em toda a sua dimensão, a adequação dos termos às condições de produção do discurso e ainda integrar aspetos psicolinguísticos, sociolinguísticos e pragmáticos.

1.3. Objetivos da investigação e metodologia

São os seguintes os objetivos centrais deste estudo.

Em primeiro lugar, pretendemos reunir textos que espelham a realidade linguística angolana e organizá-los, tendo em conta o género do discurso e a datação. No passo seguinte, a finalidade é identificar os elementos lexicais no Português de Angola, com o foco, em primeiro lugar, nas palavras simples, e compará-los com as entradas no léxico disponível (dicionário de Português Europeu). Todos os elementos que aparecem no primeiro conjunto e que não fazem parte do Português Europeu são candidatos a empréstimos lexicais. A análise manual do subconjunto permite identificar os empréstimos lexicais. Esses são analisados com o objetivo de identificar certas subcategorias lexicais, etimologia e outros aspetos. Estes elementos servem para elaborar o novo protótipo de dicionário de regionalismos angolanos. Com efeito, os dados coletados são apresentados em fichas lexicográficas, cujo modelo contém variados campos.

Um outro objetivo consiste em construir um sistema que permita extrair estruturas passivas em Português de Angola e Português Europeu para auxiliar comparações, em particular em relação às suas frequências. Pretendemos testar a nossa hipótese inicial de que as passivas são mais frequentes em Português de Angola do que em Português Europeu.

Subjacente à formulação dos objetivos gerais enunciados, foram delimitadas as seguintes questões de pesquisa:

- Quais são as especificidades das formas lexicais e estruturas passivas em Português de Angola e Português Europeu?
- Como é que o Português de Angola integra os empréstimos de outras línguas de Angola?
- Que critério se pode usar para a determinação do carácter neológico de unidades lexicais?
- Que diferenças estruturais há entre estruturas passivas no Português de Angola e Português Europeu?

1.4. Organização da dissertação

De modo a cumprir com o seu escopo, este trabalho é estruturado em treze capítulos divididos em quatro partes. Na parte I, *Enquadramento para o estudo comparativo na área de Léxico e Sintaxe*, descrevemos os fundamentos teóricos em léxico e sintaxe numa interface com a morfologia, que sustentam o desenvolvimento da descrição dos empréstimos e das estruturas passivas eventivas, estativas, resultativas e pronominais e os valores de *se*. Com efeito, apresentamos a situação linguística de Angola, os fundamentos teóricos em lexicologia e lexicografia, alguns aspetos de semântica lexical e estrutura do léxico, as estruturas predicativas e, por fim, os trabalhos nas áreas de léxico, sintaxe e linguística computacional relacionados com o nosso objeto de investigação.

Na parte II, *Enquadramento na área de Linguística Computacional*, descrevemos as técnicas de *Text Mining* e Processamento de Linguagem Natural usadas no decurso da nossa investigação. Debruçamo-nos sobre os tipos de documentos textuais, a estrutura de dados, medidas de desempenho de método automático e teste estatístico de proporções.

Na parte III, *Técnicas de Text Mining e Processamento de Linguagem Natural em Lexicologia Contrastiva*, descrevemos a metodologia de processamento de dados. Entretanto, descrevemos também a constituição e classificação do corpus, o pré-processamento de dados e as medidas de estatística descritiva dos dados. Explicitamos os processamentos para a extração de empréstimos lexicais e estudo contrastivo, comparamos os resultados obtidos do método computacional proposto. Por fim, engendramos um protótipo de dicionário de regionalismos angolanos.

Na parte IV, *Análise de estruturas passivas com técnicas de Text Mining e Processamento de Linguagem Natural*, descrevemos a metodologia para o estudo de extração de estruturas predicativas, ou mais precisamente, estruturas passivas eventivas, estativas e resultativas. Mensuramos o desempenho do método automático proposto e implementamos o teste estatístico de proporções. Para finalizar, apresentamos as considerações finais e trabalho futuro.

Dos anexos, constam as listas de empréstimos detetados na parte III e estruturas passivas extraídas na parte IV da presente investigação.

1.5. Resultados atingidos

1. Elaboração de corpora na base de áudio de telejornais, *Telejornal de Angola (2020)* e *Telejornal de Portugal (2020)*, úteis não só para o nosso trabalho, mas que podem ser disponibilizados a outros investigadores;
2. Processo semi-automático de extração de empréstimos lexicais do Português de Angola, que pode ser reutilizado noutros contextos;
3. Identificação de um conjunto de 1784 empréstimos lexicais do PA, dos quais 923 não aparecem no dicionário existente de regionalismos angolanos nem no dicionário da língua portuguesa contemporânea;
4. Para todos os empréstimos lexicais do PA, verificação da etimologia, étimo, campo lexical e significação;
5. Extração automática de cotextos de empréstimos nominais, verbais e adjetivais;
6. Extração automática de estruturas passivas eventivas, estativas, resultativas em PA e PE;
7. Descoberta de que as estruturas passivas são bem mais frequentes em PA do que em PE;
8. Detecção automática da relação da estrutura de constituinte, para analisar as especificidades de estruturas predicativas;
9. Proposta para fichas lexicográficas, que possibilitam uma melhor compreensão do protótipo de dicionário de regionalismos angolanos a ser desenvolvido para uma aplicação.

1.6. Publicações

- Muhongo, T. About Dialogism and the Construction of the Discursive Force in Solémnia Verba. 13th Edition of the *Portuguese Language Journal. American Organization of Teachers of Portuguese. University of California* (2019).
- Muhongo, T. Sobre Valores de *Se* no Português Europeu e Português de Angola. *Diseminaciones. Revista de Investigación y Crítica en Humanidades y Ciencias Sociales*, nº 3 (jan-junho 2019).
- Muhongo, T., Brazdil, P. & Silva, F. Comparison of Lexemes in Angolan and

European Portuguese with the Help of Text Mining Techniques. 15th Edition of the *Portuguese Language Journal*. American Organization of Teachers of Portuguese. University of California (2021).

- Muhongo, T. Sobre Avaliação de Modelo Preditivo em Lexicologia Contrastiva (no prelo, revista *Metalinguagens*).
- Muhongo, T. Sobre Evolução Conceptual da Dimensão Neológica (no prelo, revista *Polissema*).
- Muhongo, T., Brazdil, P. & Silva, F. Comparative Study of Passive Structures in Angolan and European Portuguese: A Text Mining Approach. [a ser submetido à revista *Inteligência Artificial*].
- Muhongo, T., Brazdil, P. & Silva, F. Detection of Loanwords in Angolan Portuguese: A Text Mining Approach. [a ser submetido à revista *International Journal of Language and Linguistics*].

Capítulo 2 – Situação Linguística de Angola: o português e as outras línguas de Angola

Da nossa língua, vemos a história de povos: suas viagens, cultura, contactos internacionais e momentos bélicos. É esta indelével história que podemos verificar através das formas lexicais no português em Angola. Este capítulo, à guisa de reflexões preliminares, tem três objetivos. No primeiro momento, pretende descrever a localização geográfica de Angola. Em um segundo momento, apresenta o resultado do contacto linguístico do português e as trinta e quatro línguas de Angola, as condições históricas e sociolinguísticas em que surge e se desenvolve a variedade do Português de Angola. Constatamos que o estudo da variedade do Português de Angola implica uma reflexão e um exame prévio das condicionantes históricas, sociais e linguísticas que caracterizam este país. Com efeito, depreendemos que a variedade do Português de Angola surge mediante um sistema educativo-religioso e desenvolve-se ao longo do português médio. Em seguida, apresentamos os grupos etnolinguísticos em Angola.

2.1. Localização geográfica

Angola é um país da África Austral, composto por dezoito províncias e com uma extensão territorial de 1.246.700 km², é o sétimo maior país de África e vigésimo terceiro maior do mundo. Tem ao norte a República Democrática do Congo, ao sul a República da Namíbia, a este a República da Zâmbia e a oeste o Oceano Atlântico, sem nos olvidarmos do seu papel na região dos Grandes Lagos, conforme podemos verificar na Fig. 2.1, que apresenta o mapa n^o 3727, versão 4, de agosto de 2008 das United Nations (UN) sobre a divisão político-administrativa de Angola. Revelam os dados do Instituto Nacional de Estatística (INE) sobre o censo populacional de 2014 que Angola possui um total de 25.789.024 habitantes, dentre os quais 71,1% se comunica em português, mas apresenta uma insuficiência pelo facto de não se detalhar quantos têm o português como língua materna, como língua segunda e, sobretudo, quantos o têm como língua estrangeira. Se se fizesse este detalhe, o número de falantes seria reduzido. Convém aludir que isto alteraria inclusive os dados do Instituto Internacional da Língua Portuguesa (IILP) e das UN, segundo os quais há mais de 250 milhões de falantes de português como língua materna ou língua segunda, quatro quintos dos quais são brasileiros e, ainda neste século, Angola e Moçambique serão os países com mais falantes de português (United Nations, 2020). Recorde-se que a variedade do Português de Angola surge num contexto político-religioso em confluência com as línguas autóctones, como veremos em 2.2, e

um sistema de educação, como sublinha Foucault (1997), pode ser considerado como uma maneira política de manter ou modificar a apropriação dos discursos da classe dominante, com os saberes e poderes que eles transportam consigo.



Fig. 2.1 – Mapa da divisão político-administrativa de Angola, United Nations (2020)

Recorde-se que os primeiros habitantes de Angola não são decerto os povos *bantu*. Visto que o povo *khoisan* é normalmente nómada, os povos *bantu*, provenientes da República Democrática do Congo, da Zâmbia e da Namíbia, instalaram-se e delimitaram o seu espaço dentro do território nacional. Em virtude disso, apresentamos, na Fig. 2.2 (Pires, 2020), a sua inserção e distribuição no território que hodiernamente é Angola.

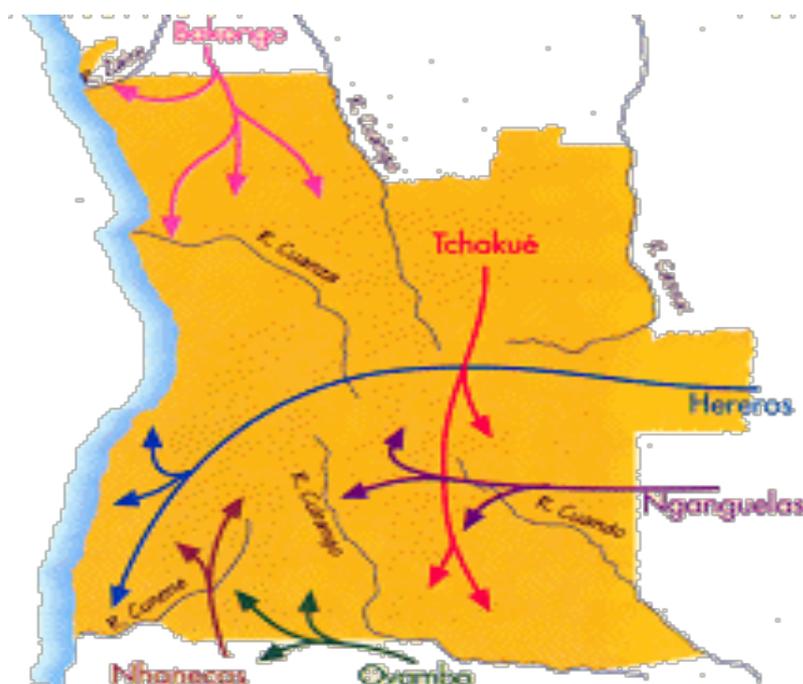


Fig. 2.2 – Mapa sobre a distribuição bantu em Angola, Pires (2020)

Por um lado, é preciso ter em conta que, em Angola, a língua nacional – falada pelos habitantes de uma nação – e a língua oficial – a que os cidadãos usam, podendo ser obrigatória, no contacto com a administração do país onde residem – não coincidem, pois o português, conquanto tenha o estatuto de língua oficial, da instrução, do poder político, da cultura e do meio de comunicação social, não é falado por toda a população angolana. Por outro lado, não há uma língua *bantu* ou *khoisan* que seja falada em toda a dimensão do território nacional, uma vez que Angola tem vários grupos etnolinguísticos.

Altuna (2006) advoga a existência dos seguintes nove grupos etnolinguísticos: Bakongos, Kimbundos, Ovimbundos, Lunda-Tchokwé, Ngangelas, Nhanecas-Humbes, Ambós, Hereros, Xindongas. Entretanto, ao refletirmos sobre a realidade etnolinguística de Angola, temos de partir da noção de conjunto, pois dentro destes grupos etnolinguísticos há subgrupos, conforme podemos verificar na Fig. 2.3 (Dias, 2003), que retrata a etnografia de

Angola. Para elucidar, os povos Gingas, Kissamas, N'golas, etc. são do grupo etnolinguístico *Ambundu*, cuja língua é o *kimbundu*. Os Mussorongos, Sossos, Muzombos, etc. pertencem ao grupo *Bakongo*, cuja língua é o *kikongo*. Os Amboins, Bailundos, Bienes, etc. pertencem ao grupo *Ovimbundu*, cuja língua é o *umbundu*.

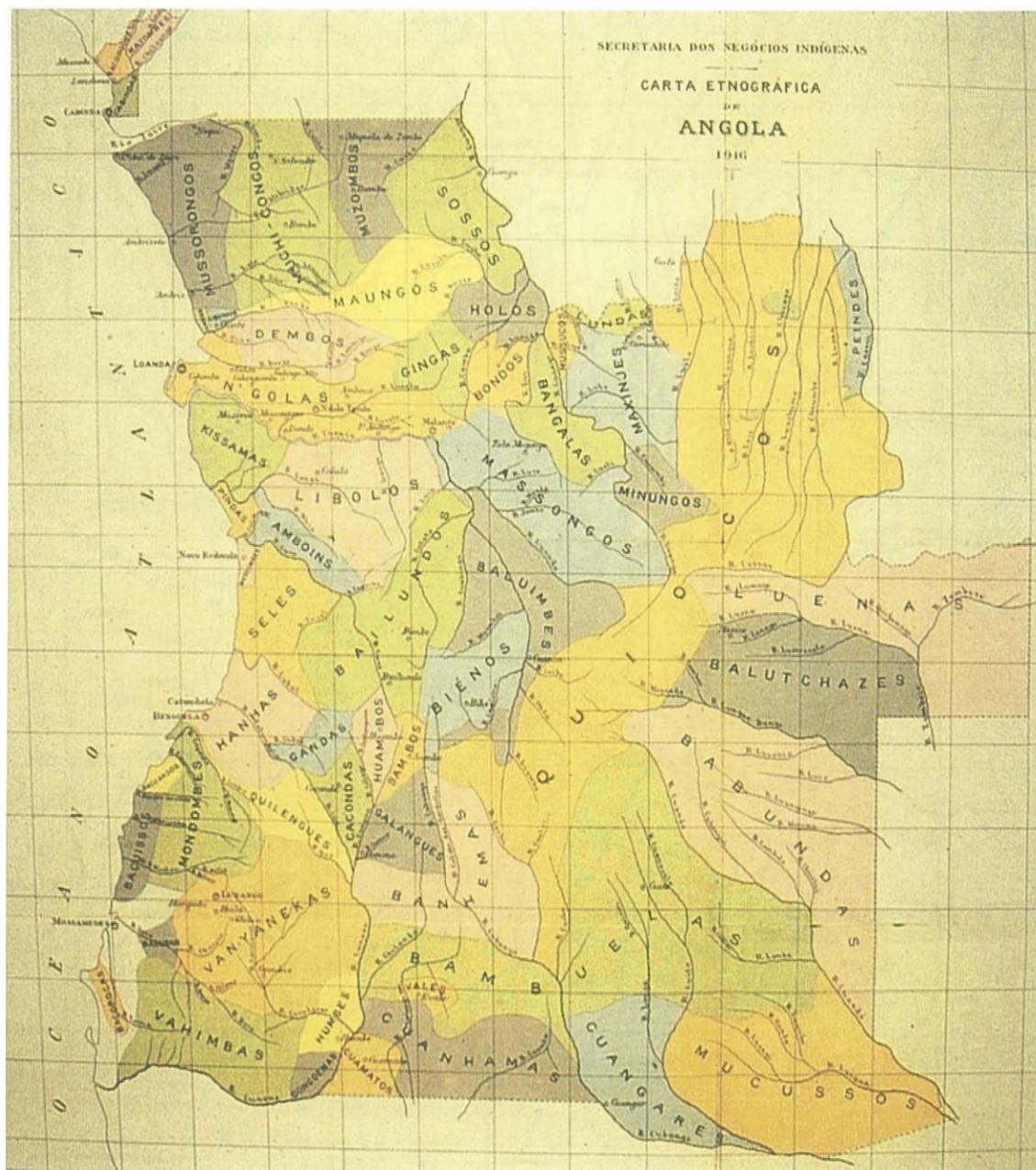


Fig. 2.3 – Mapa etnográfico de Angola, Dias (2003)

2.2. Línguas e surgimento do Português em Angola

Sem de facto retomar o que acima foi dito, queríamos assinalar que foi indubitavelmente a 23 de outubro de 1574 que, com o Decreto de D. Sebastião para a ocupação efetiva de Angola, se pronunciou a metamorfose do panorama geopolítico angolano. Dos primeiros contactos entre portugueses e angolanos, em 1482, correspondente ao período do português antigo, resultou posteriormente a colonização do segundo povo que se prorrogou até ao século XX, período do português moderno (Mateus & Cardeira, 2007). Os empréstimos lexicais, por definição, confirmam sobretudo a história deste contacto, pois cada lexema pode revelar a realidade de um povo.

Como sublinha o Tenente Coronel Francisco de Salles Ferreira (1854), que escreveu uma *Memória sobre Cassange*, cuja publicação foi feita no *Boletim do Conselho Ultramarino* n.º 2 de março de 1854, “enviado pelo Rei D. Sebastião, a frota de Paulo Dias de Novaes largou Lisboa a 23 de outubro de 1574. Avistou a Barra do *Kwanza* em fevereiro de 1575, depois de três meses e meio de viagem.” Isto seguramente riposta que a possessão em Angola não teve quinhentos anos. Quando os portugueses chegaram a Angola, encontraram os reinos e povos que vimos na Fig. 2.3. Estes compõem um mosaico de trinta e quatro línguas estruturadas em três grupos, as quais enumeraremos na Fig. 2.4, que apresenta a situação linguística de Angola.

O português era, como afirma Fonseca (2012), a língua mais falada pela população de assimilados que estava situada próxima à costa atlântica, principalmente em Luanda. Era a língua materna e nacional de muitos que estavam à frente do processo independentista e instala-se nos altos escalões do governo e na burocracia. Como advoga Muhongo (2019, p. 137), em virtude da diversidade linguística de Angola, não se podia ter, de facto, predileção por alguma língua nacional para que esta fosse oficial, pois isto desencadearia um conflito sem precedentes. O português só se generalizou entre a população quando foi apropriado pelos líderes independentistas angolanos como língua oficial em 1975, facto que também é notável em Moçambique, Cabo Verde, Guiné-Bissau e São Tomé e Príncipe. Parece evidente que, em Angola, o português é a língua oficial para que se mantivesse a unidade nacional. Porquanto, não privilegiaria nenhum grupo etnolinguístico. Constatamos que as trinta e quatro línguas se dividem em três grandes grupos *bantu*, *khoisan* e outro não identificado. Embora sejam mais usadas na comunicação, verificamos que as línguas pertencentes ao grupo *bantu* são as mais faladas oralmente (Inverno, 2018, p. 12). Enganam-se, portanto, os que denominam estas línguas de Angola de dialetos, uma vez que estas certamente se diferenciam estruturalmente e as comunidades que as falam definiram uma gramática e

espaço próprios (Bonvini, 1994). O português convive então com as demais trinta e quatro línguas de Angola, as quais podem ser vistas na Fig. 2.4 (Ethnologue, 2020).

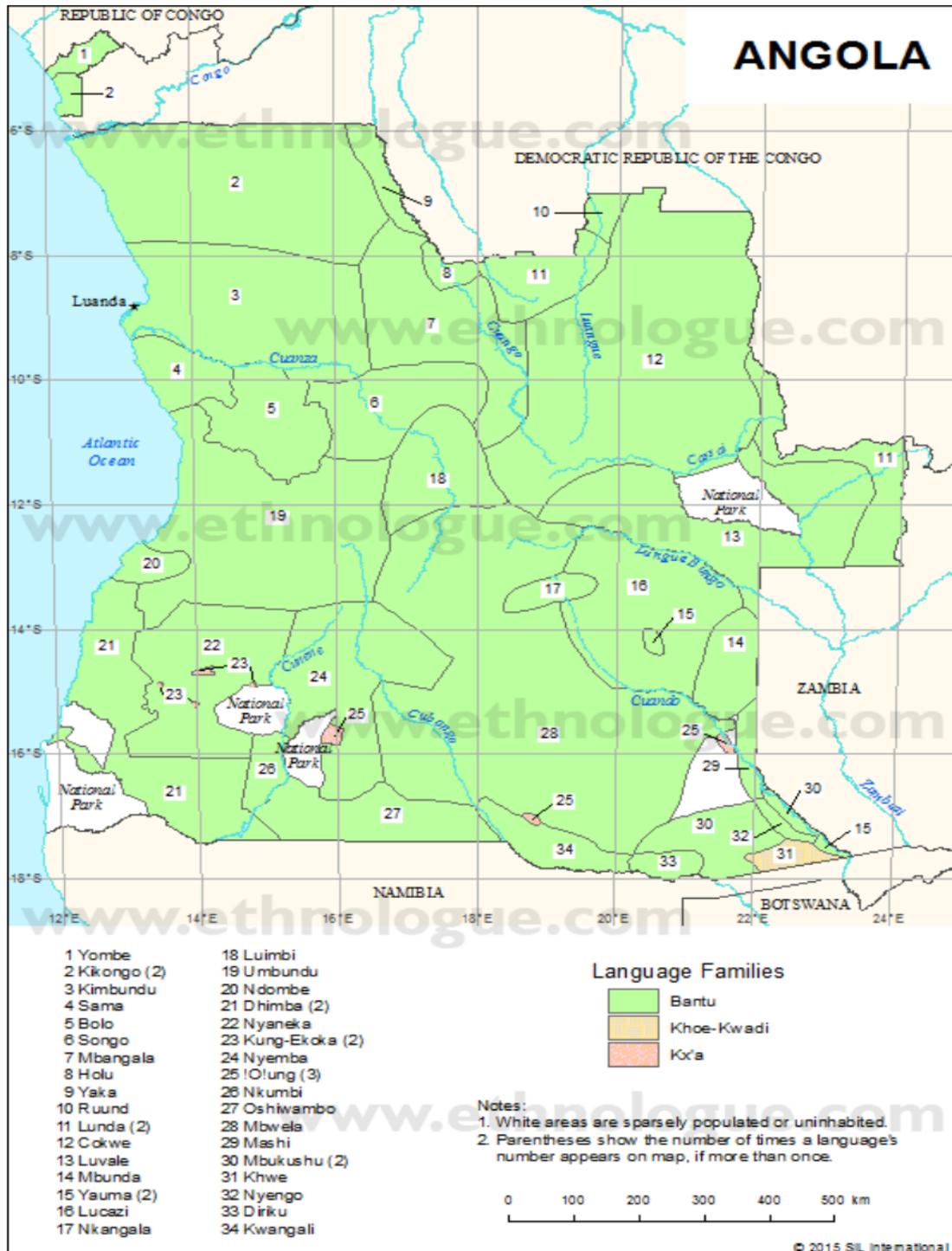


Fig. 2.4 – Situação linguística de Angola, Ethnologue (2020)

Para Guthrie (1971), as línguas *bantu* estão geograficamente divididas em zonas, segundo a localização natural de cada grupo etnolinguístico. Ao noroeste estão as zonas A, B e C. Ao sudoeste estão as zonas K e R. Ao nordeste estão as zonas D, E, F. No centro este estão as zonas G, M e N. Por fim, ao sudeste estão as zonas P e S, conforme podemos observar na Fig. 2.5.

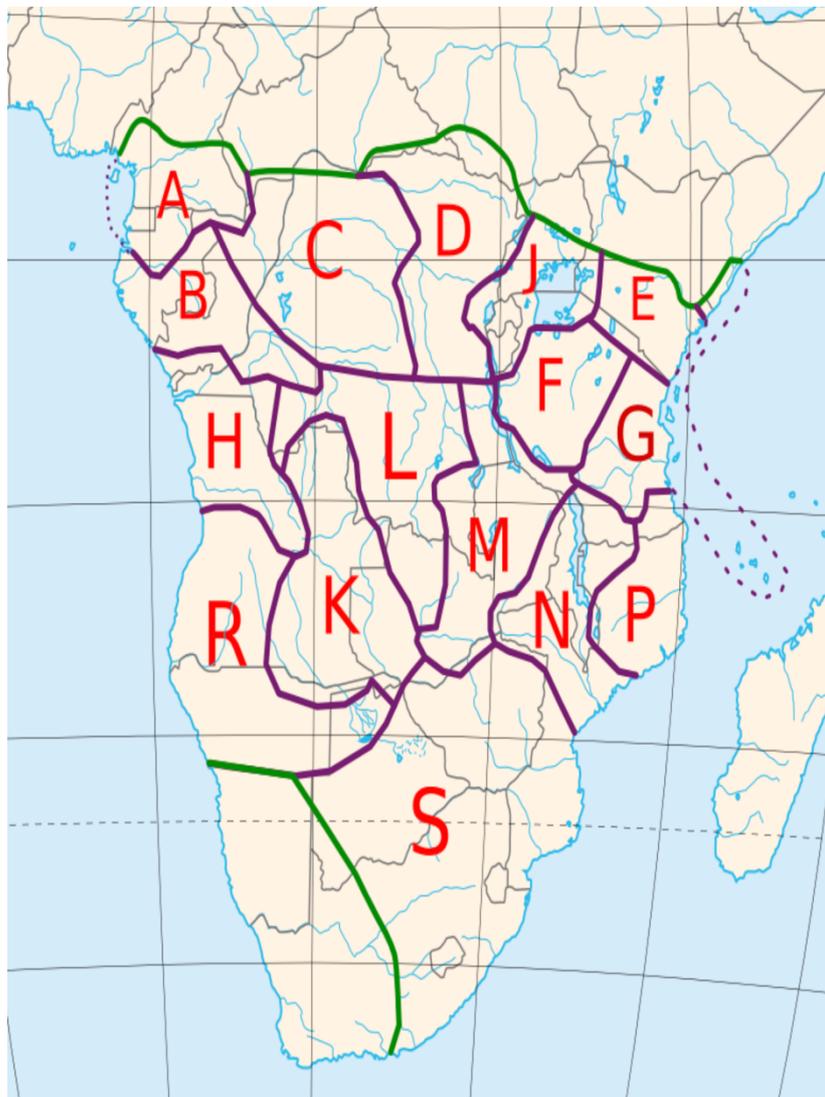


Fig. 2.5 – Distribuição das línguas bantu por zonas geográficas, Wikimedia (2021)

Em 1970, seis destas línguas foram adotadas como línguas nacionais (Inverno, 2008, p. 170), o que nos parece ser heresia da separatividade, pois as outras também são nacionais. À luz do princípio do *Congresso de Solvay* de 1927, segundo o qual todas as coisas que existem são na realidade uma mesma coisa, este estatuto de língua nacional mudou quando se implementou a Constituição da República de Angola de 2010, a qual as denomina línguas

de Angola, remetendo-nos de novo para a noção de unidade. A título ilustrativo da unidade linguística, nestas línguas, os indicadores de género são prefixos por meio dos quais os nomes se organizam em classes. Parece que os prefixos nominais são os catalisadores de todas as línguas *bantu* e regem a concordância da frase. No *kimbundu*, estas encontram-se associadas por pares que são singular e plural, conforme podemos verificar na Tabela 2.1 (António & Osório, 2018).

Classes e marcas do singular		Classes e marcas do plural		Exemplos		
				Singular	Plural	Português
1	mu	2	a	mulongexi	alongexi	professor(es)
3	mu	4	mi	mulonde	milonde	ponte(s)
5	di	6	ma	dikanu	makanu	boca(s)
7	ki	8	i	kyezo	isoneku	vassoura(s)
9	f,h,i,k, m,n,o,ph,x,z	10	ji	fundango	jifundango	pólvora(s)
11	lu	10 e 6	ji/ma	lukwako	malukwako	braço(s)
12	ka	13	tu	kanjila	tunjila	passarinho(s)
14	u	10 e 6	ji/ma	ufunu	mafunu	profissão(ões)

Tabela 2.1 – Formação do plural em kimbundu, adaptado de António e Osório (2018)

Qualquer cognoscente que escolhe a situação linguística angolana na fase da posseção percebe que o português, sob o pressuposto do postulado que se pode notar no Decreto n.º 77 emitido por Norton de Matos em 1921, era apoditicamente considerado a língua superior em relação às aborígenes e só se permitia o uso das línguas de Angola em impetras na catequese, pois decretou-se o que podemos encontrar nas seguintes inscrições:

“Artigo 1º: É obrigatório, em qualquer missão, o ensino da língua portuguesa. Artigo 2º: Não é permitido ensinar, nas escolas de missão, línguas indígenas. Artigo 3º: O uso de língua indígena só é permitido em linguagem falada na catequese e, como auxiliar, no período do ensino elementar da língua portuguesa. Artigo 4º: É vedado o ensino de qualquer língua estrangeira.” (Matos, 1921)

Apesar destas medidas político-administrativas, as línguas preexistentes resistiram ao processo de glotofagia. Na realidade, isso teve repercussões na produção escrita das

línguas autóctones.² O português convive, assim, com as demais línguas de Angola (Inverno, 2009). Por essa razão, constatamos que muitas são as dificuldades que os decisores de políticas culturais e educativas vêm enfrentando no que se refere à política linguística e notamos, na literatura, a forçosa adaptação de formas lexicais que emanam destas línguas.

Como advoga Foucault (1997, p. 18), “não há sociedade onde não existam narrativas maiores que se contam, se repetem e se fazem variar; fórmulas, textos, conjuntos ritualizados de discursos que se narram, conforme circunstâncias bem determinadas.” Ao debruçarmo-nos sobre a imprensa e literatura africana, mais precisamente a angolana, escrita numa língua de origem europeia, temos de partir de umas noções de choque de culturas como um fator determinante na consciência daqueles que produziam essa literatura aculturada. Notamos que muitos escritores e ensaístas angolanos procuraram produzir uma literatura que, apesar de não ser de fio a pavio numa língua africana, fosse autenticamente africana, tal como se pode notar na forçosa inserção de lexemas de línguas de Angola nos seus escritos, para criar maior impacto ou como forma de afirmação cultural, enquanto há uma forma lexical própria em português. No primeiro contacto com a literatura angolana, muitos leitores, por não entenderem o contexto sociopolítico de suas produções, podem deduzir que tais construções constituem precisamente uma anomalia semântica.³

Para além de o português em Angola ter um funcionamento hegemónico, é a língua de empoderamento cultural, político e cívico. Para alguns, o português é língua materna, para outros é, genericamente, língua segunda e para muitos ainda é língua estrangeira. Compreendemos que a língua é um instrumento de poder. O ensino de um padrão linguístico é, como sublinham Mateus e Cardeira (2007), não só um importante fator de estabilidade social, mas também uma poderosa forma de manutenção das ideologias dominantes. Inocência Mata *apud* Neto (2012, p. 14) sublinha “que se pode afirmar, sem receio de quaisquer melindres ideológicos, que em três dezenas de anos o poder pós-colonial fez mais pela difusão da língua portuguesa no território angolano do que o poder colonial na sua centena de dominação direta”.

² À semelhança do que observamos na Idade Média, a Igreja, aparelho ideológico de Estado religioso, acumulava muitas das funções hoje atribuídas a vários aparelhos de Estado distintos. Os manuais de ensino religioso estavam escritos em português, sendo as missões religiosas católicas e protestantes responsáveis por tal ensino. As primeiras eram pagas pelo Governo português e as segundas não recebiam nada. Assim, podemos inferir que, para o primeiro grupo, a “educação” era oferecida à classe dominante. Na ideologia protestante, não tem como a pessoa carregar o livro sagrado e não conseguir lê-la.

³ Entende-se por anomalia semântica a impossibilidade ou dificuldade de atribuir significado a uma expressão linguística (Móia, 2016, pp. 270-297).

Constata-se, em Neto (2006, p. 15), que “na região de Icolo e Bengo os colonos não construíram escolas, igrejas, hospitais nem campos de futebol”. Repare-se de facto que, ao longo das guerras de dominação coloniais holandesas e portuguesas, Icolo e Bengo – conquistada em 1595/96, no início da administração do Governador Geral João Furtado Mendonça – serviu, num curto espaço temporal, como abrigo e capital do Governo da Colónia (Mascarenhas, 2008, pp. 14-34).

Em virtude disso, compreendemos quando Brito (2011, pp. 32-33) advoga que tanto em Angola como em Moçambique a língua portuguesa não teve divulgação maciça antes dos finais do século XIX, porque a colonização portuguesa foi tardia, e sublinha que, nas regiões do interior, a língua teve uma presença relativamente pobre e que, quer num país quer noutra, a colonização só se verifica verdadeiramente a partir do final do século XIX. Isso é visível na Fig. 2.6, que apresenta o mapa de África em 1880 e 1914. Essa contextualização acerca da Partilha de África permite-nos deduzir que a colonização só se verifica após a Conferência de Berlim 1884/85, cuja resolução consistiu na divisão do território africano (Lynch, 2020).

O domínio de uma língua é, indubitavelmente, o resultado de práticas efetivas, significativas e contextualizadas. Parece evidente que, desde a época da posse, o ensino do português em Angola se tornou não só elemento sociocultural, mas também um fator de discriminação e exclusão social, pois uma língua não é apenas um sistema linguístico, mas uma instituição indissociável de todo um conjunto de crenças e valores que refletem uma determinada estrutura social (Mateus & Cardeira, 2007). Sob esta perspetiva do ponto de vista histórico, a norma pode ser considerada como elemento unificador entre as variedades diatópicas. Depreendemos, assim, que a variedade do Português de Angola surge mediante um sistema educativo-religioso, desenvolve-se ao longo do português médio. Tendo em conta os quatro períodos da transição do português enunciados por Mateus e Cardeira (2007), a fase da penetração, que compreende os primeiros contactos, data do século XV, período do português antigo.

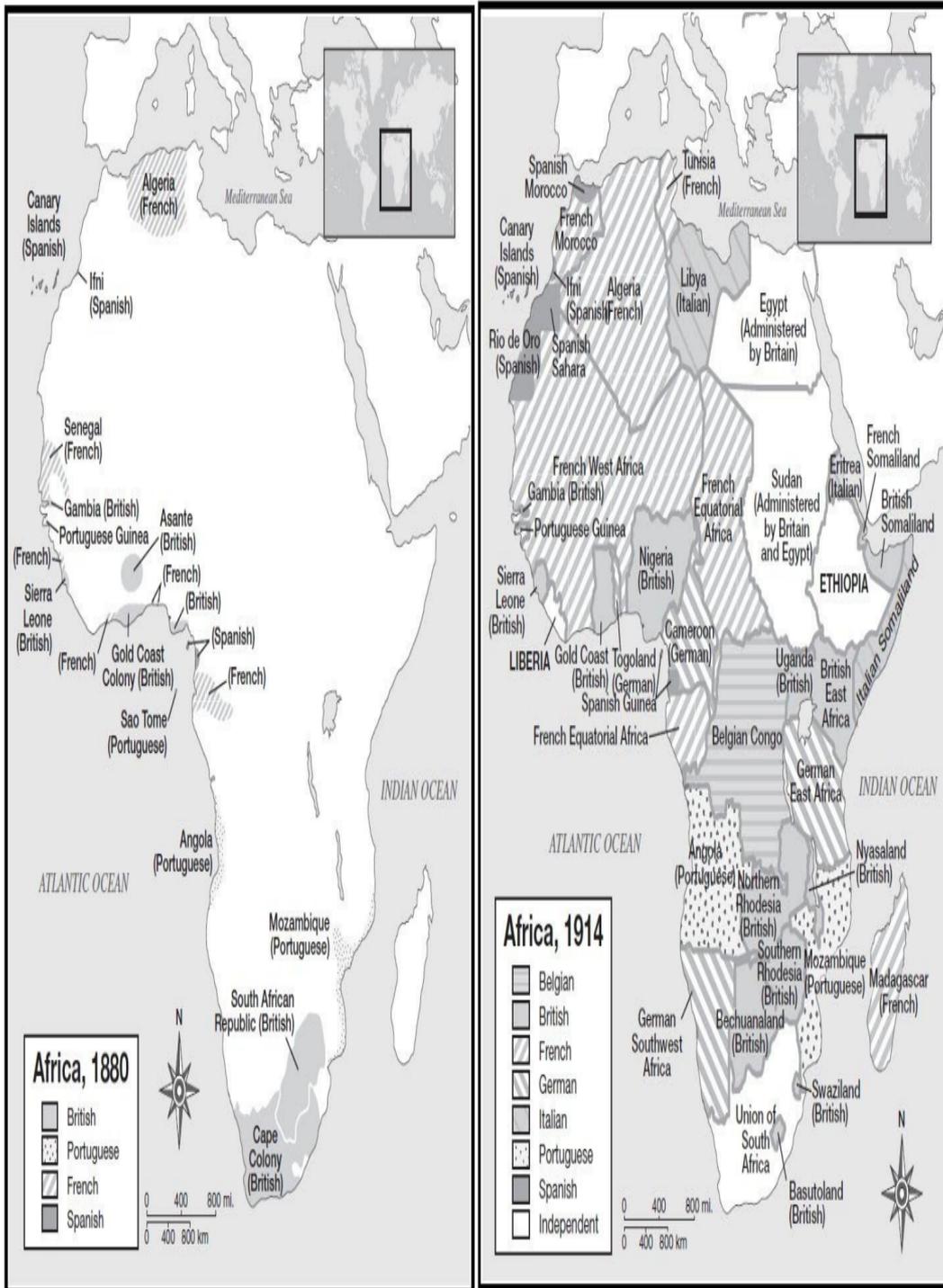


Fig. 2.6 – África antes e após a Conferência de Berlim, Lynch (2020)

Qualquer língua é um sistema heterogêneo, aberto e dinâmico, em constante mudança e caracterizado pela variação. Não sendo exatamente igual em todos que a falam, não existe língua sem variação, ela acompanha, por conseguinte, a dinâmica social (Mateus

& Cardeira, 2007). Qualquer variedade de uma língua é, assim, um sistema que varia e se transforma. Considerando que não nos parece haver, sob o ponto de vista linguístico, nenhuma megalomania entre as variedades, o magníloquo geralmente recorre a estes sistemas complexos em diferentes momentos dependentemente das condições de produção.

Embora a variedade padrão atue como força centrípeta de coesão social enquanto modelo de comunicação, como ideal linguístico até mesmo em Angola, cuja variedade do português é caracterizada pelo contato com as línguas autóctones, adquirindo certas formas lexicais que não podem, seguramente, ser consideradas como simulacro da sua sociedade, são significativamente, além de traços linguísticos próprios, laivos perenes da realidade histórico-cultural desta nação.⁴ Como resultado da situação linguística, a variedade do Português de Angola, criada em situação de contacto de línguas, apresenta mudanças em relação ao Português Europeu, sobretudo em mecanismos da concordância, a expressão da posse, a ordem de palavras e o sistema pronominal (Brito, 2011, pp. 32-33). Pensamos que isto tende a variar dependendo da coabitação com línguas de Angola, o nível de literacia e não o nível de escolaridade dos falantes, os usos sociais e individuais da língua e as condições de produção, etc.

Dentre os mecanismos de mudança linguística, como sublinham Hock e Joseph (1996) e Lightfoot (2006), e não temos como ripostar, podemos notar um subgrupo originário, a expansão do grupo originário restrito; gerações seguintes continuam o processo e acentuam o modelo; a expansão da mudança e generalização a toda a comunidade – geralmente por um valor estilístico; os reajustamentos estruturais fonológicos; os grupos que assumiram os novos traços podem levar a evolução mais longe do que os grupos iniciais – hipercorreção inconsciente; as mudanças tornam-se conscientes em maior ou menor grau; as correções mais ou menos conscientes da variável tendem a aproximá-la de modelos de prestígio; nota-se uma hipercorreção mais ou menos consciente; as formas mais estigmatizadas desaparecem – estereótipos; se o grupo mais bem situado na hierarquia da comunidade linguística estiver na origem da mudança, esta funciona como modelo de prestígio para todos os membros da comunidade. Os outros grupos integram-na no seu estilo cuidado e, embora menos e em graus variáveis, até no seu discurso familiar. No entanto, podemos afirmar que dentre as causas de mudança linguística estão o contacto com outras línguas e com outras realidades sociais, culturais e políticas (Mateus & Cardeira, 2007).

⁴ Mateus e Cardeira (2007) elucidam que é só em 1983 que, através da Lei de Normalização Linguística, as sociedades políticas oficializaram uma norma que passaria a ser promovida pelo ensino e comunicação social.

2.3. Síntese do capítulo

Neste capítulo, descrevemos a localização geográfica de Angola. Apresentamos o resultado do contacto linguístico do português e as trinta e quatro línguas de Angola, a descrição histórica e sociolinguística do surgimento da variedade do Português de Angola. Constatamos que esta data do século XV, período do português antigo e, mediante um sistema educativo-religioso, desenvolve-se ao longo do português médio. Explicitamos os grupos etnolinguísticos em Angola e distribuição das línguas *bantu* por zonas geográficas. Além disso, vimos que o estudo da variedade do Português de Angola implica uma reflexão e um exame prévio das condicionantes históricas, sociais e linguísticas que caracterizam este país.

Capítulo 3 – Fundamentos Teóricos em Lexicologia e Lexicografia

Este capítulo tem como respaldo o arquétipo teórico de Cabré (1993), Hall (2000), Mel'čuk (2000), Beard (2001), Haspelmath (2002), Pruvost & Sablayrolle (2003), Miranda (2009), Rio-Torto (2013), Rodrigues (2013), Caldas (2016) e Desmet (2016). Pretendemos apresentar os fundamentos teóricos da Lexicologia e da Lexicografia e um quadro teórico sobre a transformação conceptual da dimensão neológica.

No que concerne à estrutura do presente capítulo, a reflexão cinde-se em quatro pontos. Na Secção 3.1, explicitamos um conjunto de noções pertinentes sobre as duas áreas acima referidas, sua génese e datação como campo de conhecimento, seu liame com a sintaxe, etimologia, pragmática e semântica estrutural, o estruturalismo como modelo de análise do léxico, a tipologia de dicionários de língua corrente, as implicações da alomorfia na constituição do lexema, os processos fonéticos como resultado da combinação de morfemas, a recategorização das unidades léxicas como resultado das operações sintáticas e suas implicações na construção frásica. Na Secção 3.2, efetuamos uma apresentação dos conceitos de língua geral e língua de especialidade, as suas implicações na descrição do léxico, a dissimilitude entre dicionário geral e dicionário de especialidade, a ligação entre termo e conceito dentro de uma predicação e da estrutura conceptual. Na Secção 3.3, descrevemos o desenvolvimento da conceptualização das dimensões da neologia desde 1970 até ao presente, a morfologia derivacional, os fundamentos da neologia quer em língua geral, quer em língua de especialidade, a independência da morfologia em relação à sintaxe, as implicações do princípio de signo linguístico isomórfico no domínio da ciência e tecnologia, os parâmetros de análise da neologia. Em virtude de a composição também ser encarada como do domínio da sintaxe, debruçamo-nos sobre o contributo da Teoria X-Barra na amplitude da dimensão neológica. Na Secção 3.4, centramo-nos sobre a relação entre empréstimo lexical interlinguístico e intralinguístico e empréstimo semântico intralinguístico e interlinguístico, as implicações subjacentes ao aportuguesamento de unidades léxicas dimanantes de línguas de Angola.

3.1. As Ciências do léxico: Lexicologia e Lexicografia

A Lexicografia e a Lexicologia são duas áreas afins das ciências da linguagem. A primeira é a disciplina que se ocupa de compilar, analisar a forma e o significado das palavras de uma dada língua e propõe métodos de descrição do léxico, repercutindo os resultados de investigação de disciplinas contíguas como a Lexicologia, Sintaxe, Linguística de Corpus e a Linguística Histórica. Além disso, contempla também uma dimensão histórica, na medida em que trabalha com o corpus textual acumulado nos dicionários anteriores, refletindo sobre as descrições lexicais neles registadas e interpretando os dados de acordo com as especificidades da técnica lexicográfica usada em cada um desses dicionários (Pavel & Nolet, 2002, p. 17; Silvestre, 2016, p. 200).

Considerando as funções do dicionário, tais como a resolução de dificuldades linguísticas, o acesso a conhecimentos gerais, a síntese do conhecimento em domínios específicos, etc., Svensén (2009, p. 22) propõe uma tipologia dos dicionários, afirmando que estes podem classificar-se como monofuncionais ou polifuncionais, consoante ofereçam uma ou mais valências.

Recorde-se que, relativamente à lexicologia, Halliday (2004, p. 3) advoga que é o estudo do conteúdo das palavras ou itens lexicais. Consequentemente, a lexicologia tem como objeto de estudo a semântica lexical e a morfologia lexical. O primeiro aspeto compreende o estudo do conteúdo dos lexemas, enquanto o segundo abrange o estudo dos processos de formação dos lexemas (Rey, 1970, p. 10).

Compreendemos o léxico como conjunto virtual de lexemas que pertencem a um dado sistema linguístico. Os lexemas, quando atualizados no discurso, designam-se por vocábulos. O conjunto de vocábulos, as unidades do discurso, constitui efetivamente o vocabulário. Pensamos que, no domínio que aqui nos ocupa, compreender a lexicologia simplesmente como uma das disciplinas centrais da linguística que tem por objeto de estudo o léxico parece-nos restringi-la demasiado, pois ela estuda as unidades léxicas de uma língua e as relações sistemáticas que se estabelecem entre elas, em todos os seus aspetos: pode incluir a etimologia, a formação das unidades léxicas, a morfologia, a fonologia, a sintaxe e a pragmática, mas também tem uma ligação especial com a semântica estrutural. Atualmente, o estudo do léxico entende-se muito além da classificação cronológica dos significados. Assim, o cruzamento da lexicologia e da linguística computacional é uma área fundamental no domínio do processamento das línguas naturais (Amaro & Mendes, 2016, p. 178).

Pelos motivos acima expostos, consideramos que a lexicologia existe enquanto disciplina independente e incide sobretudo na análise da estrutura interna do léxico, suas relações e inter-relações, mas não é a única área de estudo linguístico que se interessa pelo

léxico. Isso remete-nos para a ponderação do seu conteúdo e seus limites. Repare-se que as unidades léxicas e suas relações sistemáticas são inclusive objeto de interesse de outras disciplinas da linguística, particularmente a morfologia, a semântica, a sintaxe e a pragmática, e de outras áreas relacionadas com o estudo da linguagem, designadamente a psicolinguística, neurolinguística e *sentiment analysis*. Além disso, a discriminação da sua unidade de estudo constitui uma unidade de natureza difusa para a teoria linguística, que tem recorrido a critérios ortográficos, fonéticos e fonológicos, morfológicos, sintáticos e semânticos. O conceito de palavra parece ainda um aspeto discutível em línguas isolantes como o chinês clássico, aglutinantes como o turco e flexivas como o português. Em síntese, partimos do pressuposto de que é importante perceber que a lexicologia compartilha a sua área de estudo com outras áreas da linguística e que durante anos o termo lexicologia estava desprovido de conteúdo e com efeito ausente dos dicionários, dos manuais de linguística e das gramáticas do inglês. O seu desenvolvimento deve-se, em grande parte, a investigações feitas de certas correntes linguísticas tais como a gramática generativa, a semântica cognitiva e a lexicologia funcional, e a estudos levados a cabo em determinadas áreas e enfoques como experimentação com corpus e bases de dados, estudos de terminologia e tradução, o desenvolvimento de fraseologias como ponto de interseção entre lexicógrafos, lexicólogos, semanticistas e pragmáticos, a elaboração de dicionários apoiados por novos recursos tecnológicos, a composição e decomposicionalidade do significado. A lexicologia pode denominar-se descritiva, aplicada, histórica e social dependendo das diferentes teorias linguísticas e métodos utilizados, centrando-se na possibilidade de encontrar no interior das palavras componentes sub-léxicos com repercussão semântica, sintática e com aplicações computacionais e lexicográficas. A sua função é, portanto, fornecer os pressupostos teóricos e delinear a coordenação do conteúdo e forma do léxico (Piera, 2009, p. 25; Miguel, 2009).

Se é verdade que é precisamente ao léxico como repositório das entidades léxicas que o processo formativo seguramente vai buscar os seus elementos, então podemos afirmar que, quanto ao aspeto morfológico, a combinação de elementos pode alterar a estrutura interna das palavras. Para ilustrar, o morfema derivacional – *ifíc* – seleciona muitas vezes alomorfes eruditos *petrificar*/**pedrificar*; *crucificar*/**cruzificar*. Portanto, notamos uma alomorfia na base. Os alomorfes eruditos são formas presas que só se manifestam em contexto derivacional *petr/cruc*. Dissimétrico destes, os alomorfes nativos ou autóctones podem ocorrer como palavras *pedra/cruz*. A combinação de morfemas pode, em muitos casos, resultar em transformações fonéticas. Assim acontece com *Japão* < *japonizar* (*desnasalização*), *lágrima* < *lacrimejar* (*dessoronização*) (Pereira, 2013, pp. 272-273). Quanto ao aspeto sintático, a recategorização das formas léxicas é o resultado das operações sintáticas e tem implicações

na construção frásica, tal como se nota na recategorização das bases açúcar [Nome de Tema Ø] > açúcar [Verbo] (Villalva, 2003a, p. 922).

Acrescentemos que as propriedades sintáticas do empréstimo lexical e a sua relação com as propriedades da forma vernácula são do interesse da lexicologia. Se considerarmos, por exemplo, os nomes deverbais – nomes derivados de verbos e que podem designar a ação do verbo, o resultado da ação do verbo, o portador da ação ou de instrumento – percebemos que, se o predicador verbal do qual promanam seleciona um argumento interno com a relação gramatical de objeto direto, o nome deverbal, por sua vez, também seleciona um sintagma preposicional, originando uma composição sintagmática. Considere-se, por exemplo, que o verbo *obter* é um evento, mais precisamente uma culminação que tem uma estrutura argumental binária, sendo um deles argumento interno. O nome deverbal *obtenção* também seleciona um argumento interno, podendo-se dizer *obtenção de visto* (Brito & Oliveira, 1997). Quanto ao aspeto enunciativo-pragmático, as condições de uso das palavras alteram-se, como podemos notar nos exemplos *cinema/cinematógrafo; felicitações/parabéns; prenome/nome de batismo*. Quanto à semântica estrutural, parece evidente que a formação de lexemas é o resultado de transformações semânticas (Vilela, 1994).

Sem pretendermos fazer uma gênese da lexicologia, importa salientar que a etimologia tem o seu lugar dentro da estrutura dos estudos lexicológicos e que muitas das reflexões sobre o lexema remontam à Antiguidade Grega e aos gramáticos indianos (a. C). O conceito de lexema é, por conseguinte, um dos problemas centrais quer na filosofia da linguagem, quer na filosofia do sentido. Torna-se, desse prisma, fundamental destacar alguns trabalhos de alguns estudiosos, particularmente Panini, pelo seu contributo em elaborar precisamente a Gramática do Sânscrito. Aristóteles, pelo seu contributo, a partir de artifícios retóricos, na observação da unidade lexical e os seus semas, aspeto que, mais tarde, foi muito analisado por Ullmann (1964), Coseriu (1977) e Lyons (1977).

Do que precede, é possível inferirmos que são justamente nesses mecanismos de análise das formas léxicas que assenta a ponderação de que o problema do lexema em termos morfossemânticos e socioculturais está presente tanto na história da linguística, como na história dos sistemas de escrita ou, mais precisamente, na evolução da técnica lexicográfica.

A divisão tripartida da linguística postulada por Benveniste (1966, p. 20) refere a época filosófica (pensadores gregos), época histórica (século XIX) e a época estruturalista. Procurando datar o surgimento da lexicologia como campo do conhecimento, Lino (1979, p. 12) afirma que a lexicologia se constitui como disciplina durante o período da linguística estrutural, a partir de 1920-1930. A linguística estrutural, podendo ser considerada um modelo de análise do léxico, eclode no século XX posterior à linguística histórica, toma como

ponto de partida a natureza dupla do signo linguístico (significado e significante), tem a França como epicentro, a figura central dela é Ferdinand de Saussure. Baseia-se no princípio de que a língua é um sistema cujas unidades são determinadas pelo seu lugar no sistema e não por um ponto de referência exterior como a realidade (Beedham, 2005, p. 3).

Algumas personagens emblemáticas do estruturalismo são Claude Lévy-Strauss (pai da antropologia moderna), Jacques Lacan (iniciador da corrente mais conhecida), Roland Barthes (figura chave da semiótica). Parece-nos evidente que muitas teorias linguísticas do século XX tomam como ponto de partida o conceito de estrutura linguística, partindo do postulado de que as unidades linguísticas não existem de forma individual e separada, mas sim em interdependência com outras unidades ou estruturas dentro do sistema. As ciências humanas (antropologia, psicanálise, filosofia e semiótica) partilham, na sua vertente estrutural, uma série de pressupostos provenientes da linguística (Díaz, 2009, pp. 219-222).

3.2. Língua geral e língua de especialidade

Se considerarmos a língua como um sistema de signos articulados, utilizados pelos membros de uma comunidade humana para comunicarem entre si, é notório que a linha de divisão teórica entre língua geral e línguas de especialidade é ténue. Um dicionário pode, portanto, ser geral ou especializado.

Silvestre (2016, p. 204) sublinha que o dicionário geral, sendo seguramente sincrético, refere-se a um repertório de lemas estruturados alfabeticamente, contendo na sua macroestrutura informações extralinguísticas e linguísticas de natureza ortográfica (estrutura silábica, distribuição dos diacríticos), fonológica, etimológica (identificação do étimo), dados diacrónicos (datação a partir de primeiras atestações, identificação de arcaísmos e neologismos), gramatical (morfologia, morfossintaxe, classes e categorias gramaticais), semântica (significado, definição, sinonímia, antonímia, paronímia e hiperonímia), sintática (valências sintáticas, estrutura e construção da frase), diatópica, diastrática, diafásica e integração de lexias complexas (estruturas combinatórias fixas, expressões idiomáticas, expressões formulares e provérbios).

Tendo em conta o número de línguas que são objeto de descrição, o dicionário geral pode ser monolíngue, bilingue ou multilingue. Os monolíngues são concebidos para utilizadores nativos da língua objeto ou pelo menos com uma elevada proficiência. Os dicionários bilingues distinguem-se entre a língua de partida, que constitui a nomenclatura, e a língua de chegada, que está presente em definições ou equivalentes. Assim, estes classificam-se em unidirecionais, quando na descrição linguística se supõe a distinção entre língua materna (L1) e língua estrangeira (L2). Se explorar a direção L1 > L2, é um dicionário

de codificação, usado como auxílio à produção de enunciados escritos e orais. Se o caso for inverso $L2 > L1$, é um dicionário de descodificação. Classifica-se como dicionário bidirecional, quando se pressupõe que o consulente tem uma competência de nativo quando acede à informação sobre cada uma das línguas em confronto.

Dissimétrico deste, o dicionário especializado apresenta uma descrição restrita de dados terminológicos relativos a um ou vários domínios ou subdomínios, usando linguagem de especialidade. Como observam Lehman e Martin-Berthet (2000, p. 3), “le lexique général est commun à tous les locuteurs ; les lexiques de spécialité sont liés à un domaine : science (chimie, astronomie), science et technique (informatique), métier (menuiserie), activité (jardinage). L'étude des lexiques de spécialité est la terminologie”. A terminologia, ao referir-se à descrição formal, semântica e funcional das unidades que podem adquirir um valor terminológico, estudar como se ativam e explicar as suas relações com outros tipos de signos do mesmo ou de diferente sistema linguístico, pode ser considerada como Terminologia Teórica.

Por sua vez, ao cingir-se à recompilação das unidades com valor terminológico num tema e situação determinados e estabelecendo as suas características de acordo com essa situação, pode ser considerada como Terminologia Aplicada (Cabré, 2002, p. 57). Tal compilação de termos pode ser feita a partir de um corpus de divulgação ou em corpus especializado. O primeiro é um conjunto de textos produzidos por especialistas de uma área de especialidade e dirigidos a um público não especializado nessa área e a utilizadores em geral. O segundo refere-se a um conjunto de textos produzidos por especialistas e também destinado a especialistas (Alves, 2002, p. 139). Quanto mais especializado for o registo, maior pode ser a sua sistematização e menor é a possibilidade de haver variação denominativa, pois a comunicação especializada admite níveis de especialização diferentes, graus de opacidade variados, índices diversos de densidade cognitiva, terminológica e propósitos distintos. Numa abordagem sob o ponto de vista terminológico, Desmet (2016, p. 125) sublinha que “nous parlons de concept et non de signifié, de dénomination et non de désignation”.

Se atentarmos na reflexão de Desmet (2016, p. 125), depreendemos que, ao falar em denominação, a autora se refere ao termo, o qual, surgindo posteriormente ao conceito, é uma unidade semiótica composta de conceito e denominação, cuja identidade só se justifica dentro de um campo de especialidade. As relações estabelecidas entre os conceitos formam uma estrutura conceptual. Assim, os conceitos estruturam-se lógica e ontologicamente de maneira hierárquica e um conceito pode participar em mais de uma estrutura com o mesmo ou diferente valor (Cabré, 2002, pp. 44, 57).

O valor do termo depende da posição que ocupa na estrutura conceptual de uma matéria. Portanto, termo e conceito são certamente signo e significado, uma vez que um conceito é uma unidade de conhecimento que contém os atributos de um dado referente, geralmente denominado termo, o qual, entendemos nós, é entidade variante e item do léxico especializado, porque faz parte de situações comunicativas distintas e passa por evolução (Faulstich, 2002, pp. 62-63).

Partindo deste pressuposto, está subjacente a percepção de que um conceito possui características específicas que se organizam em traços sémicos, os quais agrupam os objetos no mundo real e permitem fazer árvore de classificação. Se, como advoga Larat (1989, p. 56), todo o termo é um saber acerca de um objeto, todo o objeto é conceptualizado por um conceito, todo o conceito se exprime por um signo, todo o signo significa um conceito, todo o signo denomina um objeto e todo o objeto tem por nome um signo, então a relação entre o conceito e o termo dá-se por meio de predicções que particularizam a intensão e a extensão do referente, pois as predicções apresentam-se efetivamente sob a forma de características essenciais, acidentais e individualizadas do referente em questão (Faulstich, 2002, p. 63).

Adotando essa mesma perspetiva, é possível verificar que existem diferenças entre o tratamento lexicográfico e o tratamento terminográfico das unidades linguísticas. Em alguns casos, todavia, não será assim tão evidente a fronteira entre o léxico comum e o léxico terminológico (Sanromán, 2001, p. 223). Este esclarecimento notório tem como principal reflexo o facto de até um dicionário de língua corrente poder conter, como entrada, uma unidade terminológica, pois uma unidade terminológica é uma unidade léxica, contudo nem todas as unidades léxicas são unidades terminológicas. Sob esta égide, notamos que estamos, por conseguinte, perante unidades terminológicas vulgarizadas.

3.3. Transformação conceptual da dimensão neológica

Entende-se por neologia o processo de formação de novas unidades léxicas. Este processo caracteriza-se por dois grandes tipos de criação lexical, que englobam vários outros subtipos, nomeadamente a neologia formal, que consiste na criação de novas unidades léxicas, e a neologia semântica que consiste na utilização de um significante já existente na língua com uma nova significação. Por neologia entende-se ainda a adoção de uma unidade léxica proveniente de um outro sistema linguístico. Em princípio, a dimensão ou tipologia neológica era dicotómica e distinguia (Caldas, 2016, pp. 102-106):

- Matriz interna
 - *matriz morfossemântica* (construção (afixação ou composição) ou imitação e deformação);
 - *matriz sintático-semântica* (mudança de função ou de sentido);
 - *matriz morfológica* (redução da forma);
 - *matriz pragmático-semântica*
- Matriz externa: empréstimos

Atualmente, esta dimensão ou tipologia dicotômica passou para pentacótoma, a que nos dedicaremos nesta parte.

Como nenhum sistema linguístico é isolado, a comunicação intercultural pode implicar a inserção de unidades léxicas de uma língua na outra. Confrontamo-nos, assim, não apenas com a história da língua, com os processos neológicos (processos que participam na história da língua), mas também com a história, em geral, e com a história das ciências, em particular. A neologia pode ocorrer em língua corrente ou em língua de especialidade. Os seus elementos resultantes são denominados neologismos e neónimos respetivamente.

Como corrobora Sager (1990, p. 125), no domínio da ciência e tecnologia, o neónimo pode surgir como resultado da necessidade de encontrar uma denominação única para novos conceitos. Considerando que as bases têm autonomia semântica, este domínio evita alomorfia e aquiesce ao princípio de que um signo linguístico isomórfico é o ideal linguístico e que quanto mais um morfema se desvia desse ideal mais difícil é para as línguas sustentarem-no. Assim, os neónimos podem ser lexemas totalmente novos ou importados de outro sistema linguístico, integrando-se sob a forma de transformações imediatas ou progressivas no léxico.

Notando tal ocorrência dual, podemos pressupor que a neologia é, como advoga Guilbert (1975), o processo de criação de novas unidades léxicas, a atribuição de um novo sentido a uma unidade léxica já existente ou a aceitação de empréstimo interlinguístico dependendo das virtualidades do sistema lexical da língua. O elemento resultante do processo de criação de novos lexemas, dos novos sentidos ao lexema que já existe ou do lexema proveniente de um outro sistema linguístico é, conseqüentemente, denominado neologismo. Sob o domínio da lexicologia e da terminologia, a neologia é analisada sob os critérios linguístico, cultural e político (Desmet, 2016, p. 120).

Evoque-se que, em virtude da dinâmica social, os loquentes geralmente criam neologismo por necessidade, prazer ou por notarem que é o lexema ideal. Assim, aquilo que atualmente é denominado neologia por empréstimo lexical certamente surge para denominar

novas realidades, para substituir outras palavras e para concorrer com outras palavras já existentes (Pruvost & Sablayrolle, 2003, p. 8). Quanto ao primeiro motivo, podemos realçar sobretudo as necessidades culturais, científicas e da comunicação de um modo geral. Quanto ao segundo motivo, como se nota particularmente nos média, usa-se, às vezes, o empréstimo lexical para a mensagem dispor de maior repercussão. Quanto ao terceiro motivo, parece que este é o que mais suscetibiliza muitos loquentes, pois muitas vezes é um empréstimo interlinguístico sem necessidade, uma vez que já existe um lexema específico com que coocorre; como notamos constantemente em *mostrador* (unidade léxica do português) e *montra*, galicismo dimanante de *montre*.

É importante referir que o exercício meticoloso de explicitação da natureza fundamental do conceito de neologia cinde-se em cinco dimensões: formal, semântica, sintagmática, funcional e empréstimo. A proporção neológica não tem sido abordada pelos tratadistas da mesma maneira em todas as épocas. Na década de setenta, distinguem-se dois tipos de neologia: a neologia formal ou de forma e a neologia semântica ou de sentido.

A neologia formal, sendo hodiernamente a mais estudada, consiste na criação de novas unidades léxicas a partir de processos morfológicos e morfossintáticos da língua para denominar uma nova realidade (Desmet, 2016, p. 121). Por sua vez, a neologia semântica consiste na utilização de um significante em atribuir-lhe um conteúdo que não possuía anteriormente, quer esse conteúdo seja conceptualmente nupérrimo quer tenha sido expresso presentemente por outro significante; posição que ainda se mantém aceite pela comunidade científica (Miranda, 2009, p. 135).

Posteriormente, como linha divisória entre os clássicos e os modernos tendo como critério a lexicogénese, defende-se a necessidade de distinguir a neologia sintagmática que dimana da formal e da semântica. A composição sintagmática implica precisamente um composto sintagmático em que, à luz da Teoria X-Barra e do Léxico Generativo, um lexema funciona como núcleo e determina a categoria lexical e semântica do produto (Miguel, 2009, pp. 337-366; Pustejovsky, 1995, pp.57-58). Assim, a neologia por composição sintagmática e o seu fragmento tornam-se indissociáveis, como notamos em *casa de banho* e *pena capital* (Gross, 1998, pp. 52-72).

Para elucidar, consideremos o composto sintagmático *gestão de informação*. No eixo sintagmático, este composto sintagmático funciona como um sintagma nominal, cujo núcleo é o morfema *gestão* e os restantes morfemas funcionam, à luz do quadro teórico generativo, como complementadores. Para a compreensão do seu sentido, considera-se o princípio da composicionalidade, pois um composto sintagmático é, naturalmente, um constituinte cujo

valor semântico resulta da junção dos valores semânticos das suas partes e do modo como estes estão combinados.

Recorde-se que, sob o ponto de vista terminológico, Desmet (2016, p. 124) salienta a predileção pelo critério semântico no estudo da composição sintagmática. Como consideram Martinet (1967), Darmesteter (1967) e Benveniste (1966), tal processo de formação lexical funciona como uma unidade léxica simples. Partindo das reflexões destes autores, Picoche (1977, pp. 15-17) elenca os critérios sintáticos e semânticos subjacentes nos compostos sintagmáticos, ao evidenciar a inseparabilidade, a comutação e a unidade conceptual.

Contornando hodiernamente essa questão, Desmet (2016, p. 125) acopla aos critérios sintáticos a inseparabilidade e comutação, aos semânticos a identificação semântica de elementos compostos e ao critério extralinguístico as relações entre a designação e o objeto designado, o carácter constante da designação e o uso.

Como que regressando de certo modo à génese, podemos notar que Cabré (1993, p. 39) frisa uma dimensão neológica quadrífida: neologia formal (*atotolar*⁵; *casario*; *envelhecer*; *bocarra*), funcional (*voa*_{Nome}/*voa*_{Verbo}), semântica (*rato* em informática/zoologia; *vírus* em informática/medicina) e por empréstimo (*carimbo*; *cachimbo*; *muxima*⁶). A primeira inclui os neologismos formados por derivação afixal (prefixação, sufixal, circunfixação e infixação), formados por composição (culto, híbrida e tradicional), neologismos formados por composição sintagmática (*código de estrada*; *massa ambala*; *campo de Higgs*) e os neologismos formados por redução (sigla (*IMT*; *GNR*; *BPI*), acrónimo (*INESC TEC*; *EDEL*; *BAI*)). A segunda compreende os casos de lexicalização de uma forma e neologismo formal por conversão sintática. A terceira inclui três tipos de procedimentos: amplificação, restrição e mudança de sentido da forma de base. A quarta, por fim, compreende os empréstimos propriamente ditos e os decalques.

Podemos, assim, inferir que a tipologia dicotómica se baseia numa matriz interna que compreende aspetos morfológicos, morfossemânticos e sintático-semânticos. Mas, a tricotómica, além dos aspetos morfológicos, morfossemânticos, sintático-semânticos e pragmático-semânticos, baseia-se numa matriz externa que se refere aos empréstimos. Estas dimensões fundamentais formam, por conseguinte, um todo e são inseparáveis entre si, como de resto indicam as abordagens, pois, embora haja diferença entre elas, todas elas apontam para os mesmos procedimentos de formação neológica: o morfológico, o morfossintático e o

⁵ Na variedade do Português de Angola, esta unidade léxica refere-se a pisar o dedo de alguém.

⁶ Esta unidade léxica provém do *kimbundu* e significa coração. Na variedade do Português de Angola, o morfema *muxima* passou a ser utilizado como topónimo quando se estabeleceu o templo católico na região próxima de Massangano. *Muxima*, como topónimo, é equivalente a Nossa Senhora de Fátima em Portugal.

morfossemântico. Parece-nos que, quanto ao processo de formação morfológica, a derivação demonstra ser o mais produtivo, provavelmente pelo facto de ser o mais perceptível de todos os processos de formação morfológica. Quanto aos processos de formação morfossintática, parece-nos evidente que o mais importante é a composição sintagmática. Quanto ao processo de formação morfossemântico, destacam-se os decalques e os empréstimos (Desmet, 2016, p. 123).

Rodrigues (2013, p. 73) considera que, compreendendo a produtividade como um mecanismo inconsciente e a criatividade como um modo consciente de gerar lexemas, se verifica que a produtividade, sobretudo em morfologia derivacional, é um parâmetro gradativo. Os motivos que levam à produção de neologia prendem-se com fatores referenciais, sintáticos e avaliativos, por analogia com os padrões vigentes de formação de unidades léxicas.

É importante referir que as palavras podem passar por um processo de evolução fonética, morfológica e semântica, tais como a velarização (*tchilengue* > *quilengue*); desnasalização (*Ndembo* > *Dembo*; *Ndandji* > *Dandé*; *mbombo* > *bombô*); prótese (*Mbaka* > *Ambaca*; *Ngola* > *Angola*; *nguba* > *jinguba*). Com efeito, o exercício meticoloso de explicitação da natureza fundamental do conceito de neologia cinde-se em cinco dimensões: formal, semântica, sintagmática, funcional e empréstimo.

Se partirmos do pressuposto de que os morfemas podem apresentar uma carga funcional e semântica (Rodrigues, 2013, pp. 36-38), a neologia formal consiste na possibilidade de ordem sincrónica e diacrónica de criação de novas unidades léxicas, dependendo das virtualidades do sistema lexical de cada língua, a partir de processos morfológicos e morfossintáticos geralmente para denominar uma nova realidade, tais como *drone*, *nanotecnologia* (Caldas, 2016, p. 122). Considerando que os interfixos não têm efetivamente capacidade derivacional, contudo são uma sequência fonológica da afixação e da composição, este tipo de neologia pode dar-se por sufixação (*amplitude*), prefixação (*desamor*), circunfixação (*entardecer*), infixação (*zangarrão*), composição (*guarda-redes*), amálgama (*mucama*), reduplicação (*reco-reco*), lexicalização de uma forma gramatical ou ainda por variação denominativa (*amendoim/jinguba*) (Hall, 2000, p. 535; Beard, 2001, pp. 55-63).

Tendo como respaldo o arquétipo teórico de Hall (2000), Mel'čuk (2000), Beard (2001) Haspelmath (2002), Rio-Torto (2013) e Rodrigues (2013) é evidente que, na derivação, os processos de formação de palavras correspondem a mecanismos formais de criação de unidades léxicas que podem ser pela adjunção de afixo a uma base lexical, duas bases lexicais, mutação da categoria da base lexical sem adjunção de constituinte derivacional ou pela

alteração na estrutura fonológica/prosódica da base, apresentando uma representação fonológica, gramatical e semântica.

Considerando a posição lexicalista, as unidades léxicas são derivadas no léxico e emergem com uma estrutura interna à qual a sintaxe não tem acesso e asseveram a independência da morfologia em relação à sintaxe. Parece-nos evidente que os verbos conversos (*açúcar/açucarar*), cuja recategorização é feita na base sem o auxílio de operador derivacional entre base e produto, confirmam que a derivação pode ser vista como lexical. O que caracteriza a formação no léxico parece ser, de facto, a idiosincrasia que semanticamente os produtos adquirem (Rodrigues, 2013, pp. 102-107).

Em alguns casos, não se verifica a posição lexicalista. Por exemplo, os nomes conversos de adjetivos graduáveis dinâmicos (*os trabalhadores*), estativos (*os preguiçosos*) e valorativos (*os famosos*) remetem-nos para a dependência da morfologia em relação à sintaxe (Rio-Torto, 2013, pp. 153-155).

Assim, entendemos por afixação a formação de palavras através da adjunção de um afixo a base. Quanto à posição do afixo em relação à base, esta pode ser prefixação, sufixação, circunfixação e infixação. A primeira ocorre quando o afixo se adjunge à esquerda da base, não alterando a acentuação da palavra base, contanto que mantenham no produto a mesma categoria lexical da base. Podemos notar que os prefixos não denotam uma seletividade categorial tão notória, prototipicamente não alteram as categorias morfossintáticas das bases e, conseqüentemente, mantêm o mesmo género e possibilidade de flexão em número. Na segunda, a adjunção é feita à direita da base, podendo alterar a categoria lexical, a natureza semântica que inscreve o denotado pelo derivado numa classe ontológico-referencial diferente da da base. Convém sublinhar que alguns sufixos, como é o caso dos verbalizadores (*-iz*) e alguns nominalizadores (*-ção; -ment-*), têm evidentemente a capacidade de introduzir nos seus produtos uma estrutura argumental (Rodrigues, 2013, pp. 90-91).

Repare-se que, não obstante os prefixos e sufixos poderem apresentar alomorfia (*incomum/irrepreensível; amável/amabilidade*), um signo linguístico isomórfico é o ideal linguístico e quanto mais um morfema se desvia desse ideal mais difícil pode ser para as línguas sustentarem-no. As bases têm autonomia semântica. Há prefixos que têm autonomia fonológica, porém não têm autonomia semântica nem lexical e, com efeito, não funcionam como base. Em português, o afixo não admite alteração na sua posição em relação à base, não admite também a inclusão de outros constituintes tais como determinantes e quantificadores entre si e a base e o resultado semântico do produto não é, de facto, linear relativamente ao do sintagma (Rodrigues, 2013, p. 38).

A terceira corresponde à formação de palavras pela adjunção de um afixo descontínuo que se anexa à esquerda e à direita da base simultaneamente (Hall, 2000, pp. 535-536). Em português, a circunfixação ocorre apenas na formação de verbos denominais e deadjetivais (*anoitecer, envelhecer*). Dissimétrico dos interfixos, geralmente designados por vogais ou consoantes epentéticas que correspondem a constituintes morfológicos que detêm um papel funcional, mas não semântico na formação dos lexemas e, porquanto sem capacidade derivacional (*tecelagem*), recorde-se que a infixação possui funcionalidade semântica para o produto, sendo constituintes que ocupam uma posição medial dentro do produto na formação dos avaliativos (*sabichão*) (Mel'čuk, 2000, p. 28; Rodrigues, 2013, p. 94).

A neologia semântica ou neologia de sentido refere-se à utilização de um significante já existente na língua com um conteúdo conceptualmente novo. O neologismo semântico pode dar-se pela ampliação, restrição ou mudança do significado da unidade léxica (Sablayrolles, 2016, p. 78). A configuração das dimensões da neologia pode ser vista na (Fig. 3.1).

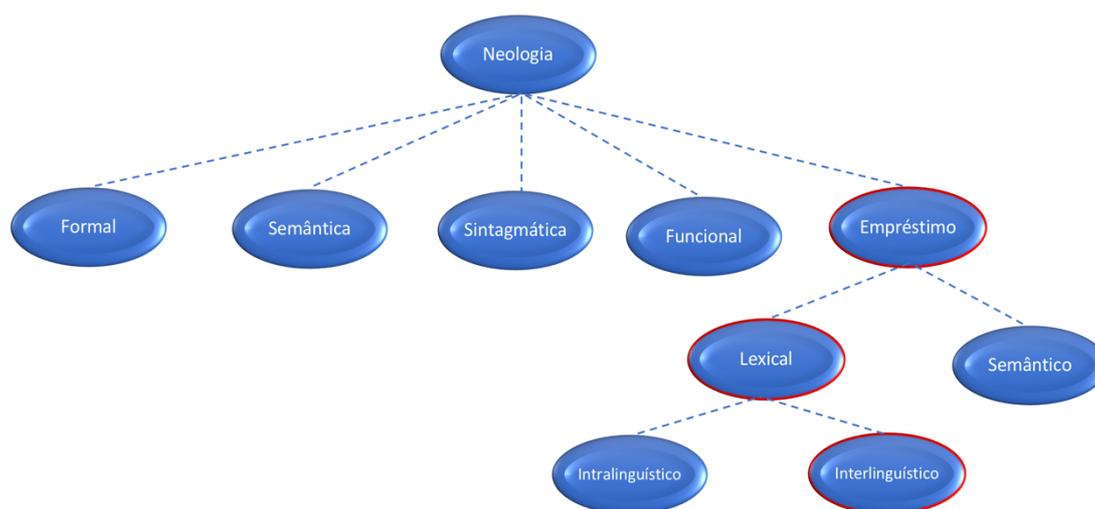


Fig. 3.1 – Configuração das dimensões da neologia

De entre os vários fatores que concorrem para a mudança do significado das unidades léxicas, Ullmann (1964, p. 411) destaca as causas linguísticas, históricas, sociais e psicológicas. A neologia sintagmática supõe a adjunção da neologia formal e a neologia semântica (Desmet, 2016, p. 122). Isso está intrinsecamente relacionado com o conceito de virtuemá, o qual se compreende aqui como o subconjunto do semema que é constituído pelos semas virtuais ou conotativos. A neologia funcional compreende os casos de lexicalização de uma forma fixa e neologia formal por conversão sintática. O empréstimo lexical pode ser compreendido como, além de elemento resultante, o processo de transferência de unidade

léxica de um sistema linguístico para o outro (interlinguístico) ou de um registo linguístico para o outro dentro do mesmo sistema linguístico (intralinguístico) (Sablayrolles, 2016, p. 245).

3.3.1. Empréstimos lexicais

Tentando explicitar a ocorrência de empréstimos lexicais nas línguas românicas, Caldas (2016) defende que “l’importation lexicale est un phénomène qui consiste à aller chercher une lexie dans une autre langue, plutôt que d’initier une fabrication de tout pièce.”

Uma mudança social pode implicar uma inovação lexical. Entende-se por empréstimo lexical, por um lado, o processo de transferência de uma unidade léxica de um registo linguístico para o outro dentro da mesma língua ou de uma língua para outra. Por outro lado, o empréstimo é a unidade léxica resultante do processo desta transferência (Correia & Lemos, 2009, pp. 13-18).

Parece evidente que alguns empréstimos podem ser considerados como uma visível consequência do contacto linguístico e podem constituir um sinal de mudanças linguísticas e sociais de uma dada época, pois um idioma tende a incorporar mais empréstimos de línguas culturalmente mais próximas de si e que geralmente entram na língua preferencialmente por via oral, embora a forma escrita pareça atingir, com o tempo, uma importância crescente (Miranda, 2009, pp. 142-144). Compreendemos que um empréstimo pode revelar uma realidade social, passar por adaptação fonológica, gráfica, morfológica e pode ocorrer tanto em língua corrente como em língua de especialidade. Nesta última, é denominado neónimo. Em virtude de os empréstimos interlinguísticos poderem ser estudados sob uma perspectiva prescritiva ou descritiva, ao longo da presente perquirição, considerá-lo-emos, numa perspectiva descritiva, de empréstimos quer como processo, quer como unidade resultante do processo.

Muitas unidades léxicas dimanantes de várias línguas de Angola foram adaptadas ao sistema linguístico português. Este processo aconteceu ao longo de séculos. Ao explicar o contacto de línguas africanas e não só com o português, Endruschat & Schmidt-Radefeldt (2015, p. 142) enfatizam que, na sequência da história da posseção em África (a partir de 1415), na Ásia (1498) e na América do Sul (1500), se juntaram ainda palavras e conceitos xenoglóssicos das línguas *bantu* africanas (por exemplo do quimbundo, de Angola, e de changana de Moçambique). Podemos observar o aportuguesamento de *Humbi* [Nome de Tema Ø] para *Humbes* [HUMB_[radical preso] + e_[IT] + s_[MNP]] no exemplo de *dependency parsing* na Fig. 3.2.

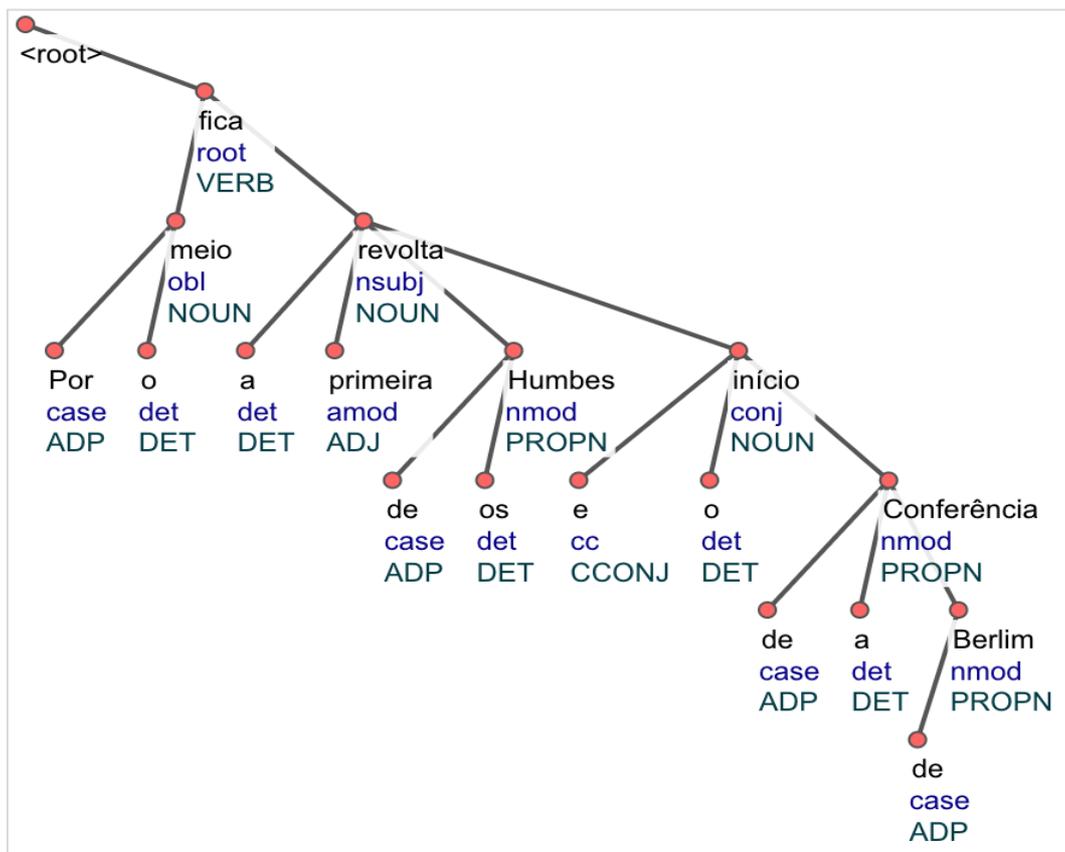


Fig. 3.2 – Relação de dependência sintática de Humbes

Por mais que, à primeira vista, se pense em empréstimos apenas como processos interlinguísticos, isso não é tão linear, pois estes também podem ocorrer dentro do mesmo sistema linguístico, por metassemia. Estes são geralmente denominados empréstimos intralinguísticos (Sablayrolles, 2000, p.232). Para ilustrar, veja-se as unidades léxicas *vírus* – termo médico e informático – e *rato*, que nos remete para a informática ou para zoologia, dependentemente do contexto (Miranda, 2009, p. 44).

Parece-nos evidente que alguns loquentes se enleiam com os empréstimos lexicais interlinguísticos. Tal impacto deve-se a dois fatores. Primeiramente, porque as unidades léxicas importadas são produzidas em sistemas linguísticos distintos e, como tal, apresentam características formais que, além de as tornarem opacas ou amórficas, infringem, em muitos casos, o sistema linguístico importador. Em seguida, e a outro nível, a importação em massa de unidades léxicas que muitas vezes se sobrepõem a formas vernáculas pode descaracterizar o idioma recetor *e-mail/correio eletrônico; slide/diapositivo; shopping center/centro comercial*.

A propósito da necessidade de distinção metódica entre empréstimo lexical e empréstimo semântico, Vilela (1994) destaca que o empréstimo semântico pode advir do acréscimo de um novo sentido proveniente de outro sistema linguístico a uma unidade léxica existente, embora conserve o antigo. Para ilustrar, podemos verificar que a unidade léxica *investir* que, inicialmente, significava *dar investidura, empossar*, recebe, por influência do francês, depois do século XVI, o sentido de “atacar” e, atualmente inclui, por influência do inglês *to invest*, o valor de aplicar capitais em.

Em virtude de ser ao léxico como repositório das entidades léxicas que o processo formativo vai buscar os seus elementos, deduzimos que, quanto ao aspeto morfológico, a combinação de elementos pode alterar a estrutura interna dos lexemas. Tal combinação pode resultar em transformações fonéticas. Quanto ao aspeto sintático, a recategorização das palavras é o resultado das operações sintáticas e tem implicações na construção frásica. Os verbos conversos podem ser indicados para esclarecer a independência da morfologia em relação à sintaxe, mas não se pode dizer o mesmo em relação aos nomes conversos.

Uma vez que os empréstimos são um tipo de neologismo, constatamos que alguns lexemas não chegam a generalizar-se como é, principalmente, o caso de muitas unidades terminológicas (*sinoblocos, dead wood*) que nos remetem para o automobilismo e a economia, respetivamente, e que ficaram apenas em textos de especialidade. Outros permanecem na língua por pouco tempo. Estes podem englobar lexemas com uma vida efémera, pois designam realidades sociais ou políticas temporárias (Freitas, Ramilho, & Arim, 2010). Com efeito, verificamos que a integração de um empréstimo consiste numa adaptação gradual e feita a vários níveis que podem ser transformações imediatas, progressivas e integração no léxico, obedecendo à critérios morfossintáticos, fonológicos, gráficos e semânticos (Freitas, Ramilho, & Arim, 2010a).

3.4. Síntese do capítulo

Neste capítulo, descrevemos o conceito de Lexicologia e Lexicografia. Não obstante os morfemas possam apresentar uma carga funcional e semântica, a capacidade derivacional do morfema advém da sua carga semântica. Efetuamos uma apresentação dos conceitos de língua geral e língua de especialidade e as suas implicações na descrição do léxico. Explicitamos o desenvolvimento das tipologias de neologia desde a década de setenta até a data. Além disso, apresentamos as dimensões da neologia e cingimo-nos aos empréstimos.

Capítulo 4 – Estrutura do Léxico e Processos de Formação de Palavras

Este capítulo tem três objetivos. No primeiro momento, pretende descrever a estrutura do léxico. Em um segundo momento, apresenta a estrutura interna de constituintes morfológicos em português, partindo da relação forma-significado. Tratando-se de um conceito relacional, constatamos que esta é regida por regras de formação ou restrições morfossemânticas. Compreenda-se, neste estudo, a semântica como a área da gramática que se ocupa do significado das expressões linguísticas. Como área do conhecimento que se dedica a estudar a forma como a linguagem representa o mundo, as análises semânticas centram-se, tradicionalmente, nos aspetos da interpretação exclusivamente dependentes do sistema linguístico, pois os elementos extralinguísticos constituem objeto de estudo da pragmática (Móia, 2016, p. 308). Embora a semântica estude os signos linguísticos a qualquer nível, cingimo-nos aos lexemas. Assim, a semântica lexical é uma disciplina confrontada com o objetivo de traçar padrões regulares de comportamento semântico das expressões linguísticas (Geeraerts, 2010, pp. 7-8). Em seguida, explicita os processos de inovação lexical e formação de palavras. A concepção de que o léxico é um componente estruturado, dinâmico e criativo permite verificar que as unidades léxicas estabelecem entre si relações formais e semânticas que constituem parte essencial do conhecimento que um falante tem da estrutura da língua.

4.1. Estrutura do léxico

A sistematicidade do léxico provém das suas relações paradigmáticas e sintagmáticas. Uma língua compõe-se de dicionário, que pode ser um repertório de palavras, e de uma gramática, a qual de facto pode equivaler a um conjunto de dispositivos que agrupando, distribuindo e modificando as palavras do dicionário, dão lugar a frases de uma língua. Para o português, a ideia de palavra está muito condicionada por dois fatores principais, nomeadamente a grafia e a tradição greco-latina. É comum dividir-se em palavra prosódica e palavra gramatical, em que, considerando critérios como a coerência, ordem e significado arbitrário, conclui-se que, para ser palavra, os elementos gramaticais aparecem sempre juntos, numa ordem fixa e têm uma coerência e um significado convencional (Piera, 2009, pp. 25-29).

Encontram-se, certamente, várias tentativas de determinar o conceito de palavra que hodiernamente ainda não é consensual tanto do ponto de vista interlinguístico como da perspectiva de uma língua concreta. A noção de palavra não é, portanto, a mesma em todas as

áreas (Villalva & Silvestre, 2014, p. 77). Parece evidente que, como sublinha Piera (2009, p. 30), a noção de palavra não é de um conceito primitivo da teoria linguística nem promete sê-lo. Villalva (2003a, p. 938) define palavra como a etiqueta da projeção máxima do radical, ou seja, da unidade morfológica que domina o tema e o seu especificador que é a flexão morfológica (= FM). Sob este prisma, uma palavra pode ter a seguinte representação:

(1) [[[X] _{Radical} [Y] _{Especificador temático}] _{Tema} [Y] _{Especificador morfossintático}] _{Palavra}	
[[[<i>lev</i>] _{RA} [<i>e</i>] _{IT}] _{TA} [] _{FM}] _{A singular}	<i>cf. leve</i>
[[[<i>lev</i>] _{RA} [<i>e</i>] _{IT}] _{TA} [<i>s</i>] _{FM}] _{A plural}	<i>cf. leves</i>
[[[<i>filh</i>] _{RN} [<i>o</i>] _{IT} [] _{FM}] _{N singular}	<i>cf. filho</i>
[[[<i>filh</i>] _{RN} [<i>o</i>] _{IT} [<i>s</i>] _{FM}] _{N plural}	<i>cf. filhos</i>
[[[<i>valid</i>] _{RV} [<i>a</i>] _{VT}] _{TV} [[<i>r</i>] _{TMA} [] _{PN}] _{FM}] _{V infinitivo}	<i>cf. validar</i>
[[[<i>valid</i>] _{RV} [<i>a</i>] _{VT}] _{TV} [[<i>va</i>] _{TMA} [<i>mos</i>] _{PN}] _{FM}] _{V infinitivo}	<i>cf. validávamos</i>

Visto que o núcleo é o constituinte que determina as propriedades da palavra inscritas na sua assinatura categorial, podemos notar que se trata de um predicador que pode seleccionar um complemento – predicador transitivo – como os afixos derivacionais (*livraria*) ou não seleccionar qualquer argumento – predicador intransitivo – como os radicais das palavras simples (*leve*, *poema*). Recorde-se que os radicais (*barbearia*), temas (*continua*) ou palavra (*incrivelmente*) podem funcionar como complemento, o qual é um constituinte seleccionado por um predicador transitivo. Os prefixos (*reescrever*, *prever*) e os sufixos avaliativos e z-avaliativos (*tesourinha*, *espiralzinha*) são modificadores, pois são adjuntos que apenas operam semanticamente. O principal contraste entre ambos é que os sufixos avaliativos se associam a radicais enquanto os sufixos z-avaliativos se associam a palavras. Os constituintes temáticos (*ponderar*, *elidir*), os morfemas flexionais (*camas*, *amáveis*) e os interfixos (*antropomórfico*, *estereofonia*) são especificadores (Villalva & Silvestre, 2014, pp. 87-88).

O morfema constitui a unidade significativa mínima de análise morfológica. Do ponto de vista estruturalista, o morfema tem sido caracterizado como a unidade significativa mínima ou signo linguístico mínimo, ou mais precisamente, como a combinação mais pequena possível de significante e significado. É comum caracterizar-se o morfema como “unidade gramatical mínima”, eliminando-se toda referência ao significado. Os morfemas podem ser livres ou presos. Entretanto, as palavras podem ser constituídas por um morfema livre, unidade linguística mínima com conteúdo e expressão, como *sapato*, *ontem*, *darma*. Podem ser

constituídas por morfemas livres e presos como *comprar*, *livraria*, *reescrever*, por vários lexemas, como *guarda-redes*, *chave inglesa* ou por vários morfemas presos, como *hierocracia*, *ignívoro* (Vilela, 1994, p. 11; Arquiola, 2009).

Quando um morfema é livre, corresponde a uma palavra. Quando é preso, não corresponde a uma palavra, uma vez que não pode ser integrado em frase pelo facto de obrigatoriamente precisar de se combinar com pelo menos mais um morfema preso para darem origem a um morfema livre, não podendo formar por si só frases isoladamente (Rodrigues, 2016, p. 49; Veloso, 2016). Destarte, consideramos, por conseguinte, que as unidades do léxico estão estruturadas em categorias funcionais e categorias léxicas. As categorias léxicas constituem lexemas e são os nomes, verbos, adjetivos e advérbios em *mente*. As restantes classes léxicas compreendem-se como categorias funcionais, pois atuam sobre os lexemas como operadores gramaticais (Villalva & Silvestre, 2014, p. 77). As propriedades das categorias léxicas podem ser canceladas ou substituídas por outras, se forem inseridas no campo gramatical (sintático ou morfológico) apropriado (Borer, 2005).

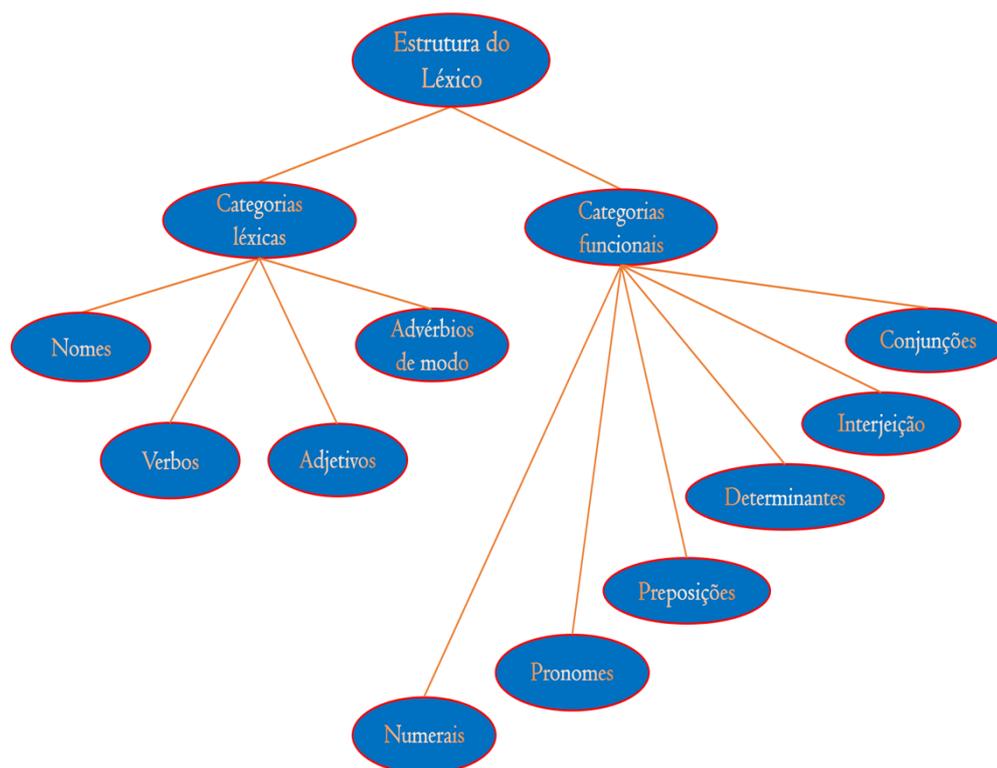


Fig. 4.1 – Esquema da estrutura do léxico

Tendo como respaldo a Hipótese da Integridade Léxica, a sintaxe não tem acesso à estrutura interna das unidades léxicas para operar com unidades que fazem parte dela, pois a sintaxe forma constituintes, frases, mas não palavras. A estrutura interna da palavra não é da alçada da sintaxe, mas da morfologia (Rosa, 2006, p. 80). Com efeito, as categorias léxicas, sendo x^0 ou potencial núcleo da frase à luz da Teoria X-Barra, podem receber categorias funcionais para formar palavras, dando lugar a uma interface entre a sintaxe e o léxico. Por essa razão, do ponto de vista morfológico, a palavra morfossintática corresponde à forma ou a uma das formas associadas a um mesmo lexema, subsumidas por ele – é uma forma flexionada em número, se a classe lexical a que pertence o lexema e os traços semânticos do mesmo admitirem flexão. Compreendemos, portanto, que as unidades multilexicais inscritas no léxico não são palavras, pois são frases que se lexicalizaram adquirindo significação própria, como *dar corda aos sapatinhos*, *roi-e-sopra* e *maria-vai-com-as-outras* (Mota, 2013a; Aronoff & Fudeman, 2011).

4.2. Estrutura morfológica do Português

Entende-se por estrutura morfológica um esquema de organização interna de constituintes morfológicos que, ordenados linearmente de determinada maneira e segundo uma hierarquia com regras estritas, correspondem à configuração de uma palavra. Essa palavra corresponde a um lexema se, na sua descrição, não se incluem elementos flexionais. Para determinar os constituintes internos, usamos o método da comutação, o qual consiste na substituição de um constituinte por outro do mesmo tipo que possa ocorrer na mesma posição e dê origem a uma palavra diferente, existente ou possível, em português (Mota, 2013a, p. 2801). Recorde-se, assim, que radical, tema e constituinte temático, que pode ser compreendido como o índice temático nas classes não verbais ou vogal temática nos verbos, são as etiquetas que identificam os constituintes morfológicos que ocupam os três vértices nucleares da estrutura básica das palavras.⁷ Portanto, a identificação de uma estrutura morfológica passa, assim, pela identificação do radical e do tema. Sendo o portador da significação básica da palavra, o radical de palavras simples é uma forma inanalísável, como por exemplo *lastim*, *progrede*, *persuade*, etc. O radical de palavras complexas é uma forma complexa, concretamente um radical derivado, como [lev]_{Radical Simples} [ez]_{Afixo}Radical Complexo. Se o radical for complexo, o tema também o é, podendo ser representado como [RAD_{Complexo} + IT]_{Tema Complexo}. Com efeito, os radicais simples são unidades léxicas portadoras de informação idiosincrática de natureza

⁷ Compreendemos que os afixos também são constituintes morfológicos, porém não os entendemos como constituintes morfológicos nucleares (Mota, 2013a).

morfológica, sintática e semântica. Funcionando como base das formas flexivas de uma palavra, entende-se por tema a unidade morfológica que domina o radical e o constituinte temático, o qual especifica a classe temática do radical, como [dign]_{Radical Simples} [o]_{Índice Temático} | Tema Adjetival = TA.

Convém notar que, de acordo com a natureza do índice temático, os nomes, adjetivos e advérbios podem ser tema em *-a, -e, -o*, como *logorreia, cuneiforme, crisântemo, agora, longe, decerto*. São formas de tema *-Ø~e* as terminadas em *-r, -s, -z* e *-l, mar, pedrês, xadrez, exequível*.⁸ Os constituintes temáticos nos nomes, adjetivos, advérbios, pronomes e determinantes são átonos. Assim, é evidente que o índice temático não é flexional nem derivacional, mas sim morfolexical e participa na morfologia de concordância (Mota, 2013b, p. 2875). Embora o radical veicule a informação básica, é no domínio do tema que os nomes têm a sua significação específica, uma vez que é no domínio do tema que o género tem expressão (Mota, 2013a, pp. 1804-1827). São atemáticas as que não integram índice temático, como *céu, coração, afinal, antes* mas também se encontram formas terminadas em vogal átona (*táxi*) e em consoante (*cais, lápis*). A vogal temática dos verbos é tónicas e pode ser *-a, -e, -i*, sendo primeira, segunda e terceira conjugação respetivamente⁹ (Villalva, 2003a, pp. 920-927). Recorde-se que as vogais temáticas são hierarquicamente superiores aos constituintes flexionais e representam as categorias de tempo-modo-aspeto (TMA) e de pessoa-número (PN) típicas dos verbos. Estas cinco categorias são, na morfologia do verbo, a contraparte de requisitos da sintaxe e da semântica da frase, com as quais podemos constatar sincretismo paradigmático (*cantava-Ø, cantasse-Ø, cantaria-Ø*). Podemos notar que a seleção do índice temático ou da vogal temática é, como sublinha Mota (2013a, p. 2819), determinada no nível lexical. Portanto, a vogal temática tem a função de identificação plena do lexema verbal, como o índice temático relativamente ao lexema nominal e adjetival.

Convém sublinhar que o conceito de raiz também confundido com o radical, a qual se trata de um segmento formal compartilhado por todos os cognatos e que contém o significado léxico fundamental comum, como *trist* nos lexemas *tristonho* e *entristecer*, cuja representação pode ser [[[trist]_{Raiz}[onh]]_{Radical Complexo}[o]_{Índice Temático}Adjetivo e [[en[trist]_{Raiz}ec]_{Radical Complexo}[e]_{Vogal Temática} [r]_{Desinência}Verbo. As unidades léxicas *competição, competir* e *competente* partilham também a mesma raiz *compet*; assim como *sorriso, sorrir* e *sorridente* contêm a mesma raiz *sorr*. A noção tradicional de raiz verbal, nominal, etc. carece de sentido claro. A raiz é somente raiz e passa a nominal ou verbal, quando recebe as categorias funcionais. A raiz e o radical

⁸ Como os advérbios não têm expressão num paradigma, não é possível verificar a sua inclusão nessa subclasse. Deverão ser considerados atemáticos (Mota, 2013a, p. 2827).

⁹ Verbos com vogal temática *-o* pertencem à segunda conjugação, tendo em conta a sua forma latina.

confundem-se quando se trata de radicais simples, como [[lev]Radical Adjetival [e]Tema]Adjetivo e [[lev]Radical Verbal [a]Vogal Temática [r]Desinência]Verbo. Portanto, a categoria do radical simples ou raiz é determinada em função da categoria sintática das palavras simples em que cada radical pode ocorrer (Arquiola, 2009, p. 55; Villalva, 2003a, p. 920). Do ponto de vista do léxico, o radical assegura a informação básica, permitindo que se estabeleça um parentesco não só formal como semântico entre, por exemplo, digno, digníssimo e dignatário, podendo-se dizer que o radical é inanalísável morfológicamente (Mota, 2013a, pp. 2789-2810). Recorde-se que o sufixo temático e o afixo derivacional são lexicais, porém o sufixo flexional (número e pessoa) não é lexical, uma vez que não está presente no lexema, apenas é requerido em sintaxe, respondendo a morfologia com material que permite realizá-lo e, conseqüentemente, não é derivacional porque a sua presença não dá origem a um novo lexema.

O termo base designa o elemento sobre o qual se aplica um processo morfológico. A base pode ser nominal, verbal, adjetival, adverbial ou pronominal. Trata-se de um conceito relacional regido por regras de formação ou restrições morfossemânticas. Para ilustrar a restrição morfológica, o sufixo *-mento* não se combina com verbos em *-ificar* como *calcificar* e *classificar* (**calcificamento / calcificação; *classificamento / classificação*); o sufixo *-vel* combina com verbos (*amar* *v* > *amável* *A*; *suportar* *v* > *suportável* *A*); o sufixo *-al* formador de adjetivo relacional combina-se com bases nominais (*mão* *N* > *manual* *A*; *cultura* *N* > *cultural* *A*); por sua vez, o sufixo *-ez* combina com adjetivos (*rígido* *A* > *rigidez* *N*; *plácido* *A* > *placidez* *N*). O sufixo *-ano* pode combinar com base adverbial, resultando em adjetivo deadverbial (*cerca* *Adv* > *cercano* *A*). A base pode ser pronominal (*vós* *Pron* > *vosear* *v*). Concernente às restrições semânticas, estas podem estar relacionadas com a estrutura argumental e o caráter aspetual. Com efeito, podemos considerar que as bases podem ser radicais (*mesa* > *mesinha*; *laranja* > *laranjal*; *plácido* > *placidez*) ou temáticas (*estuda* > *estudante*; *dura* > *duradouro*; *comercializa* < *comercializável*). Quando a base e o resultado da derivação são da mesma classe, as palavras são isocategoriais (MES-, radical nominal, e *mesinha*, nome). Se são de classe lexical distinta, são heterocategoriais (PLACID-, radical adjetival, e *placidez*, nome; ESTUDA-, tema verbal, e *estudante*, nome).

Recorde-se que, na formação de verbos com o prefixo *-auto*, a base deve possuir um argumento agentivo e pelo menos um argumento interno, ou mais precisamente, o verbo deve ser transitivo ou ditransitivo e cujo sujeito seja um agente (*autocriticar-se*; **autocaminhar-se*; **autotemer-se*). Quanto ao caráter aspetual, a formação de verbos com o prefixo *-re* com valor iterativo combina, assim, com predicados verbais télicos, isto é, aspetualmente delimitados (*reescrever*; *reenviar*), não combina com predicados que expressam estado (**ressaber a resposta*) nem com predicados de atividade durativa e não

delimitada (**recaminhar*). Com efeito, a classificação de palavra simples e complexa (derivada ou composta) baseia-se na constituição de seus respectivos temas: palavra simples – tema simples (*maduro*), palavra derivada – tema derivado (*imatur*), palavra composta – tema composto (*guarda-chuva*). A formação de palavras aquiesce, por conseguinte, a critérios categoriais, morfológicos e semânticos (Arquiola, 2009, pp. 56-57). Há, no entanto, uma restrição. Se a base da sufixação fosse uma palavra morfossintática previamente existente tal daria resultados agramaticais (**cabelo*]_N eira]_{SUF} / *cabel*]_{RN} eira] _{SUF}]_N), realçando que a derivação se faz com uma base que não apresenta IT, uma vez que uma base que permite a ocorrência de processos morfológicos sobre ela pode ser, decerto, um elemento lexical, um lexema, podendo ser denominada base-lexema (Mota, 2013a, p. 2797).

4.3. Processos de inovação lexical

O léxico tem três possibilidades para se adaptar a situações novas, nomeadamente: câmbio semântico, como acontece em *engarramento* (metáfora), *sesta* (metonímia), *ténis* (elipse: *sapato de ténis*), empréstimo como *shopping*, *carimbo* e pela formação de palavras a partir de elementos existentes na língua, aspeto muito estudado em morfologia lexical (Vilela, 1994, pp. 14-25). A morfologia léxica compreende o conjunto de procedimentos formais ampliados de uma língua para criar novas unidades léxicas a partir de unidades léxicas já existentes. Por sua vez, a morfologia flexiva dá lugar a distintas formas gramaticais de uma mesma unidade lexical, remetendo-nos a dois conceitos fundamentais: produtividade e bloqueio. O primeiro refere-se às regras disponíveis para a formação de palavras novas (*glorioso/inglório*). O segundo refere-se ao fenómeno pelo qual uma palavra já existente na língua impede a formação de uma palavra complexa com igual significado (*glorioso/*gloriosidade; harmonioso/*harmoniosidade*), devido ao desajuste da relação forma-significado. É evidente que a morfologia conta com dois tipos de elementos: unidades e regras que combinam essas unidades. Considera-se a palavra a unidade máxima da morfologia; por sua vez, é a unidade básica da sintaxe, tornando-se em ponto de contacto entre ambas as disciplinas (Arquiola, 2009, pp. 53-61). O léxico português enriquece-se mais frequente e criativamente, usando processos de formação de palavras. Dentre os processos de formação de palavras, podemos notar a derivação e a composição, os quais são elucidados mais adiante.

4.4. Alomorfia

Estamos diante de alomorfia, quando o morfema surge realizado sob formatos diferentes. A alomorfia confirma que é possível encarar o morfema como a ligação entre uma forma e significado (Rodrigues, 2016, pp. 45-46). Assim, o mesmo significado pode ser veiculado por formas fonológicas diferentes que se encontram correlacionadas no léxico mental. A alomorfia pode dar-se na base, remetendo-nos para a sua origem latina que pode ser marcada pelo traço [- autónomo] e [+ erudito], como em *cabelo/capilar*, *carvão/carbonizar* ou a sua configuração em fases pretéritas da língua, consideradas formas [- eruditas] e podem atuar como radicais de palavras livres, como *endiabrar/diabro/diabo*, *emperlar/perla/pérola* (Pereira, 2016, pp. 304-306). Recorde-se que a alomorfia pode dar-se no prefixo, como *imoral/ injusto*; ou no sufixo, como *queridinha/ mãezinha*, *contornável/ contornabilidade*; na desinência verbal *falo/estou*, etc. Além disso, o morfema -s de plural pode ser realizado por vários morfemas, como *amores/amáveis* (Rodrigues, 2016, p. 46).

4.5. Processos de derivação em Português

Tendo como respaldo o arquétipo teórico de Hall (2000), Mel'čuk (2000) Beard (2001), Haspelmath (2002), Arquiola (2009) e Rio-Torto (2016a), na derivação, os processos de formação de palavras correspondem a mecanismos formais de criação de unidades léxicas que podem ser pela adjunção de afixo a uma base lexical, duas bases lexicais, mutação da categoria da base lexical sem adjunção de constituinte derivacional ou pela alteração na estrutura fonológica/prosódica da base, apresentando uma representação fonológica, gramatical e semântica. Entende-se por derivação a formação de novas palavras ou de novos temas de palavra mediante a adição de um afixo a uma base (*separar/ inseparável*) ou mediante a modificação da base (*perdoar/perdão*). No primeiro caso, temos derivação afixal e no segundo caso temos derivação não afixal. Além disso, a derivação pode ser heterocategorial, quando há uma mudança categorial entre a base e o derivado; ou pode ser homogénea, quando não há mudança de categoria entre a base e o derivado.

4.5.1. Derivação afixal e subtipos semânticos

Importa sublinhar que a criação lexical em português é predominantemente concatenativa (Pereira, 2016). Entende-se por afixação a formação de palavras através da adjunção de um afixo à base. Quanto à posição do afixo em relação à base, esta pode ser prefixação, sufixação, circunfixação e infixação. Compreendemos que, tendo em conta a categoria da base, as formações derivadas podem classificar-se em denominais (*esperança* > *esperançoso*), deadjetivais (*tranquilo* > *tranquilizar*), deverbais (*doer* > *doente*), deadverbais (*cerca* > *cercano*) e depronominais (*vós* > *vosear*), se têm como base um nome, adjetivo, verbo, advérbio e pronome respetivamente.

Se o elemento resultante for um nome, adjetivo, verbo ou advérbio, a derivação é respetivamente nominal (*fingir* > *fingimento*), adjetival (*andar* > *andante*), verbal (*frágil* > *fragilizar*) e adverbial (*final* > *finalmente*). Arquiola (2009, p. 64) sublinha alguns subtipos semânticos da derivação nominal, adjetival e verbal, cujos exemplos assinalados com asterisco são nossos:

Derivação denominal

- nomes de ação: *explicar* > *explicação*, *peregrinar* > *peregrinação*, etc.
- nomes de agente/instrumento: *ganhar* > *ganhador*; *acalmar* > *calmante*, etc.
- nomes de qualidade: *leal* > *lealdade*; *sensato* > *sensatez**, etc.
- nomes coletivos: *corpo* > *corporação**, *pássaro* > *passarada**, *parte* > *partido**, etc.

Derivação adjetival

- adjetivos relacionais: *estudante* > *estudantil*, *banco* > *bancário*, etc.
- adjetivos possessivos: *barba* > *barbudo*, *dente* > *desdentado**, etc.
- adjetivos de semelhança: *serpente* > *serpiginoso**, *seda* > *sedoso*, etc.
- adjetivos ativos: *fugir* > *fugitivo*, *entender* > *entendido*, etc.
- adjetivos passivos: *suportar* > *suportável*, *arremessar* > *arremessado*, etc.

Derivação verbal

- verbos acusativos: *tranquilo* > *tranquilizar*, *mama* > *amamentar**, etc.
- verbos iterativos: *saltar* > *saltitar**, *martelo* > *martelar*, *sílaba* > *silabar*, etc.
- verbos incoativos: *lânguido* > *elanguescer*, *pálido* > *empalidecer*, etc.

Derivação prefixal

A prefixação, cuja estrutura é [prefixo + RADICAL]_{Radical Complexo} e tem essencialmente a função de modificar uma base não tendo a possibilidade de originar uma palavra de outra classe lexical que não a da base atribuindo-lhe informação semântica complementar (Mota, 2013a, pp. 2811,2822), ocorre quando o afixo se adjunge à esquerda da base, não alterando a acentuação da palavra base, contanto que mantenham no produto a mesma categoria lexical da base (*ter* *v*/ *reter* *v*; *empate* *N*/ *desempate* *N*). Podemos notar que os prefixos não denotam uma seletividade categorial tão notória como os sufixos. Para ilustrar, o prefixo *ante-* agrega-se a verbo (*ver*>*antever*), nome (*câmara*>*antecâmara*), adjetivo (*penúltimo*> *antepenúltimo*). O prefixo *semi-* requer que as suas bases sejam quantificadas ou graduadas de maneira que se combine com base verbal (*enterrar*> *semienterrar*), base nominal (*círculo*> *semicírculo*) e base adjetival (*transparente*> *semitransparente*). Alguns prefixos podem aparecer em estruturas coordenativas (*becas pré- e pós-doutorais*) (Rio-Torto, 2016b; Arquiola, 2009).

Prototipicamente, os prefixos não alteram as categorias morfossintáticas das bases, se mantiverem no produto a mesma categoria lexical da base (*movível* > *removível*)¹⁰. Do ponto de vista fonológico, os prefixos diferenciam-se dos sufixos não avaliativos pelo seu caráter átono, pois a prefixação tipicamente não altera a posição do acento da base (*maduro*> *imatur*). Caso seja um produto prefixado, o acento é alterado de modo a conciliar-se com o padrão prosódico da categoria resultante, como *torto* *A*> *entort* *v* (Rodrigues, 2016, p. 103). Contudo, os prefixos isocategoriais mantêm o mesmo género e possibilidade de flexão em número (*mortal* *A*> *imortal* *A*). Dissimilmente destes, alguns prefixos – denominados heterocategoriais – tais como *a-*, *en-* e *es-*, que formam verbos deadjetivais e denominais, podem promover alteração da categoria da base (*torto* *A* > *entortar* *v*; *terra* *N*> *aterrar* *v*) (Rodrigues, 2013, p. 89).

Tendo em conta a história do português, estabelece-se uma distinção entre prefixos preposicionais como *ante-*, *sub-* e *sobre-* (*antediluviano*, *antebraço*, *submarino*, *sobrenatural*), e prefixos adverbiais *des-* (*desobedecer*, *desmontar*) – se provêm de preposições ou advérbios latinos e como afetam semanticamente o nome assim como fazem as preposições e advérbios (Arquiola, 2009, pp. 66-67; Hall, 2000). Consideram-se os verdadeiros prefixos, elementos modificadores, os seguintes *a-*, *de(s)-*, *in-*, e *re-*, visto que alguns prefixos suscitam dúvidas quanto a serem realmente prefixos ou elementos de composição (Mota, 2013a, pp. 2822-2823).

¹⁰ O prefixo *re-* certamente pode ser agregado a qualquer base verbal, desde que esta designe um evento reversível (*reler*, *reprogramar*, *reabrir*, mas **remorrer*, **rematar* no sentido de tornar a matar) (Rodrigues, 2016, p. 104).

Derivação sufixal e sufixal avaliativa

Na sufixação, a adjunção é feita à direita da base, podendo alterar a categoria lexical, a natureza semântica que inscreve o denotado pelo derivado numa classe ontológico-referencial diferente da da base (*real* *A*/ *realidade* *N*; *cómoda* *A*/ *comodamente* *Adv*). Em alguns casos, acontece o fenómeno de cancelamento vocálico, o qual consiste na perda da vogal átona final da base (*fábula/fabuloso*; *cabeça/cabeçudo*). Noutros, pode acontecer alomorfa no sufixo, como *-dor*, *-tor* e *-or* em derivação heterocategorial (*jogar>jogador*; *infringir>infrator*; *desertar>desertor*). Convém sublinhar que alguns sufixos, como é o caso dos verbalizadores e alguns nominalizadores como os derivados em *-ção*, *-mente*, etc., têm a capacidade de introduzir nos seus produtos uma estrutura argumental (*criar/criação de...*) (Rodrigues, 2013, p. 91; Arquiola, 2009).

Além de combinarem com base nominal, verbal, adjetival, adverbial e pronominal, verificamos que os sufixos selecionam categorialmente as suas bases, os não apreciativos estão especificados com a sua própria categoria e, em muitos casos, os não apreciativos nominalizadores *-a* (*contenda*), *-ção* (*condução*), *-ez* (*altivez*), *-idad* (*sagacidade*), *-ado* (*iluminado*), etc. estão especificados com a sua própria categoria. Em virtude destes aspetos, considera-se que os sufixos não apreciativos constituem o núcleo morfológico da palavra derivada. Do ponto de vista semântico, um conteúdo pode ser expresso por diferentes sufixos (Arquiola, 2009, pp. 62-63).

Convém notar que, não obstante os prefixos e sufixos poderem apresentar alomorfa, um signo linguístico isomórfico é o ideal linguístico e que quanto mais um morfema se desvia desse ideal mais difícil é para as línguas sustentarem-no. As bases têm autonomia semântica. Há prefixos que têm autonomia fonológica (*pós-*, *ante-*, *sob-*), porém não têm autonomia semântica nem lexical e, com efeito, não funcionam como base. Em português, o afixo não admite alteração na sua posição em relação à base, não admite também a inclusão de outros constituintes tais como determinantes e quantificadores entre si e a base e o resultado semântico do produto não é de facto linear relativamente ao do sintagma. Alguns sufixos são homogéneos, isto é, não alteram a categoria do produto em relação à base (*mulher/mulherio*; *bonito/bonitão*) (Rodrigues, 2013).

A expressão sufixal da avaliação pode realizar-se com sufixos avaliativos, que combinam com radicais (*livro>livrinho*, *amigo>amigaço*), e com sufixos z-avaliativos que se combinam com palavras (*artista>artistazito*, *boleia>boleiazinha*). Estes podem operar como diminutivos, aumentativos (Rio-Torto, 2016a, p. 358). A Tabela 4.1 mostra a diferença entre a sufixação e a sufixação avaliativa.

Sufixação	Sufixação Avaliativa	Flexão
Significado denotativo Ex.: relógio ‘aparelho que mede o tempo’ > relojoeiro ‘pessoa que faz, arranja ou vende relógios’	Significado conotativo (tamanho ou apreço/desprezo) Ex.: relógio > relógiozito ‘relógio pequeno/relógio que apreço muito’	Significado gramatical Ex.: grande>grandes ‘grande + traço plural’
Muda a categoria da base Ex.: <i>alto</i> _A > <i>altura</i> _N	Não muda a categoria da base Ex.: <i>alto</i> _A > <i>altinho</i> _A	Não muda a categoria da base Ex.: <i>alto</i> _A / <i>alta</i> _A / <i>altos</i> _A / <i>altas</i> _A
Cria novas palavras Ex.: <i>combater</i> _V > <i>combatente</i> _N	Não cria palavras novas (exceto em lexicalização como <i>casinha</i>) Ex.: <i>casinha</i> aparece como entrada no dicionário.	Não cria palavras novas Ex.: <i>casas</i> não aparece como entrada no dicionário
Mais interno na palavra complexa Ex.: <i>moedeiro</i>	Intermédia entre a sufixação não avaliativa e flexão Ex.: <i>moedeiros</i>	Mais externa que em palavras complexas Ex.: <i>moedeiros</i>

Tabela 4.1 – Comparação da sufixação e flexão, adaptado de Arquiola (2009)

Derivação parassintética

A derivação parassintética, cuja estrutura é [x + RADICAL + x]_{Radical Complexo} (Mota, 2013a, p. 2811), é particularmente frequente na formação de verbos deadjetivais ou denominais, embora também se verifiquem alguns casos de adjetivalização. A parassíntese pode ser percebida como um procedimento de formação de palavras que consiste na aplicação simultânea de um prefixo e um sufixo a uma base dado que tanto a prefixação da forma de base quanto a sua sufixação geram formas não atestadas na língua (Villalva, 2003b, p. 952), cuja estrutura ternária pode ser representada por [[pref [X]]_x suf]_{x1} ou [[pref [[X]_x suf]_{x1}, dependentemente da base, como *trono* > *entronizar*.

Este processo é pouco produtivo em espanhol, mas muito produtivo em português (Arquiola, 2009, p. 69). Parece que a parassíntese mais comum em português pode ocorrer com a adjunção dos prefixos *es-*, *a-*, *em-*, e os sufixos *-ear*, *-ejar*, *-ecer*, *-izar*, como *esverdear*, *esclarecer*, *anoitecer*, *empobrecer* (Bechara, 2006, p. 503). Apesar de ser um problema provecto, Arim & Freitas (2010, p. 179) asseveram que os produtos parassintéticos, com

feito, devem ser considerados como palavras derivadas por sufixação, o que não nos parece crível dada a sua estrutura de constituintes.

Derivação circumfixal

A circumfixação corresponde, na realidade, à formação de palavra pela adunção de um constituinte descontínuo que se anexa à esquerda e à direita da base simultaneamente (Hall, 2000, p. 535). Este processo de formação é muito visível no neerlandês, como é o caso de *been* > *gebeente* (Lieber, 1992). Em português, a circumfixação ocorre com verbos apenas na formação de verbos denominais e deadjetivais (*tonto* > *entontecer*; *velho* > *envelhecer*; *verde* > *esverdear*; *claro* > *esclarecer*) em que o elemento da esquerda não se junta à base sem o elemento à direita **entontar*, nem o elemento à direita se junta à base sem o elemento à esquerda **tontecer*. É evidente que este processo carece de um esclarecimento, pois é diferente da parassíntese, processo em que ocorrem sucessivas afixações (Rodrigues, 2016, p. 106; Mota, 2013a).

Ao esclarecer o conceito de parassíntese, Arquiola (2009, p. 69) sublinha que esta também se denomina circumfixação, equiparando os dois processos. Compreendemos que pode ser que haja diferença entre a parassíntese e a circumfixação. Na parassíntese, não estamos diante de afixo descontínuo, pois temos um prefixo e um sufixo. Significa isto que, na realidade, a parassíntese tem uma estrutura ternária. No entanto, na circumfixação não temos prefixo e sufixo, mas sim um único afixo descontínuo. A estrutura da circumfixação é, por conseguinte, binária. Parece evidente que este processo de formação de palavras precisa de maior esclarecimento pelo facto de ainda não se definir os constituintes descontínuos do português. Como esclarecimento desse aspeto teórico, Mota (2013a, p. 2822) tece algumas considerações que ainda podem ser discutíveis, dada a estrutura dos dois processos e o facto de não se definir os afixos descontínuos em português.

“O processo de parassíntese dá origem apenas a novos lexemas verbais. As bases a que se ligam os afixos são maioritariamente nominais e adjetivais e os afixos têm uma configuração particular, formando um único afixo descontínuo. Este processo é denominado circumfixação, na tradição gramatical, e os afixos, circumfixos.” (Mota, 2013a, p. 2822)

Derivação infixal

Dissimétrica dos interfixos, geralmente designados por vogais ou consoantes epentéticas, que correspondem a constituintes morfológicos que detêm um papel funcional, mas não semântico

na formação dos lexemas e, porquanto, não têm capacidade derivacional, como (*fumívoro; tecelão*), repare-se que a infixação, possuindo funcionalidade semântica para o produto, corresponde a constituintes que ocupam uma posição medial dentro do produto e são usados na formação de avaliativos, como *zangarrão, sabichão* (Rodrigues, 2013). Uma vez que os interfixos, em alguns casos, evitam hiatos e preservam a identidade mórfica da base, a sua presença deve-se à necessidade mórfica, pois estes são semanticamente vazios. Por essa razão, estamos de acordo que os interfixos não têm, na realidade, capacidade derivacional (Rodrigues, 2016, p. 108).

4.6. Conversão

A conversão refere-se à transposição de uma unidade léxica de uma categoria para a outra sem afixação, sendo que a morfologia derivacional opera com alterações categoriais no léxico sem que haja vestígio formal direto dessa alteração, ou mais precisamente, sem que haja um morfema responsável por essa mudança (Beard, 2001, p. 62). A derivação dá-se através da conversão elaborada mental e paradigmaticamente sem auxílio de operador derivacional entre base e produto. Consistindo, realmente, na recategorização de radicais ou de palavras morfossintáticas preexistentes em radicais e palavras de outra classe lexical, compreendemos que se trata de um processo eminentemente lexical, pelo facto de não haver a intervenção de processos morfológicos e sintáticos para mudança de classe lexical, como em LIVR]_{RA} → livr]_{RV} [a]_{VT} [r]_{TMA}]_V; LAÇ]_{RN} → laç]_{RV} [a]_{VT} [r]_{TMA}; jantar]_V → jantar]_N; rápido]_A → rápido]_{ADV} (Mota, 2013a, pp. 2824-2826).

Sob o ponto de vista sintático, qualquer verbo pode ser convertido em nome, mas nem todos os nomes podem ser convertidos em verbo. O facto de a ocorrência em sintaxe do infinitivo como nominal admitir flexão em número e pessoa mostra que a alteração categorial não se deu no léxico, ao contrário do que acontece na conversão. Assim, a conversão ocorre no léxico, não na sintaxe. Os derivados conversos comprovam que há, decerto, restrições relacionadas com parâmetros estruturais morfológicos e léxico-conceptuais-argumentais bem definidos que impedem que determinados verbos sejam convertidos em nome. Se fosse sintático, o semantismo dos conversos seria possível e linear em relação ao semantismo da base. Em *rogo* temos o semantismo de prece; em *caça* temos o semantismo de conjunto de animais caçados, além da designação de evento. A conversão é, por conseguinte, um processo não sintático, mas que dá origem a novos lexemas. Trata-se de processo morfológico com consequências notórias em morfologia, pois acarreta alterações a nível da formatação formal.

O tipo de índice temático que se anexa ao radical é diferente consoante a categoria lexical do radical, nem todo verbo dá origem a um nome e nem todo o nome dá origem a um verbo e as mudanças ao nível das categorias e subcategorias indicadas morfológicamente estão dependentes da mudança na categoria lexical.

Concernente aos verbos conversos, considera-se que o radical da base sofreu essa recategorização no léxico, como *escorrega* v/ *escorrega* n; *mente* v/ *mente* n; *canto* v/ *canto* n. Os nomes deverbais conversos apresentam semantismo muito variados. Por exemplo, *passeio* designa o evento de passear e o local onde se passeia; *lixa* designa evento e instrumento. Portanto, seguindo a análise de Rodrigues (2016), o nome é derivado e o verbo, derivante, se estiverem presentes os prefixos *a-*, *en-*, *es-*, como em *ruga* n > *enrugar* v > *enruga* n; se o nome tiver apenas semantismo de carácter concreto, o nome é derivante e o verbo é derivado, como em *muro* n > *murar* v; se para além de semantismo concreto, o nome apresentar significação abstrata de evento, o nome é derivado, como em *colher* v > *colha* n; se o nome tiver acentuação esdrúxula não coincidente com a acentuação geral dos nomes do português, o nome é derivante, como em *âncora* n > *ancorar* v. Se o nome tiver estrutura argumental, o nome é derivado, como *gestão de dados*. Se o verbo for de tema *-e* ou *-i*, o verbo é derivante, porque a formação de novos verbos faz-se com a VT *-a*. Mas, se o verbo for de tema em *-a*, não se determina a direccionalidade da derivação através desse critério (Rodrigues, 2016, pp. 115-119).

4.7. Processos de construção não concatenativa

Considerando a posição lexicalista, as unidades léxicas são derivadas no léxico e emergem com uma estrutura interna à qual a sintaxe não tem acesso, asseverando a independência da morfologia em relação à sintaxe, pois, se a flexão é tão relevante para a sintaxe, parece evidente que os verbos e nomes conversos, cuja recategorização é feita na base sem o auxílio de operador derivacional entre base e produto, confirmam que a derivação é lexical. O que caracteriza a formação no léxico é, de facto, a idiosincrasia que semanticamente os produtos adquirem. A saída de uma regra de derivação é, portanto, uma nova unidade léxica que está sujeita à lista léxica (Rodrigues, 2013; Beard, 2001).

Trata-se de operações que geram produtos através de mecanismos que não assentam em princípios de natureza eminentemente morfológica, mas antes de natureza fonológica/prosódica (cruzamento vocabular, truncação, reduplicação) e/ou gráfica (siglação/acronímia), cujos padrões envolvidos não são lineares, pois nestas operações não são identificáveis

constituintes morfológicos encadeados linearmente. Não se trata de processos de formação de palavras com estatuto morfológico, pois raramente as bases mantêm integralmente o seu material segmental, exceto nos casos de reduplicação total (Pereira, 2016, p. 521).

Distinguimos, de seguida, os principais processos de construção não concatenativa. Neste contexto, seguimos essencialmente o trabalho de Pereira (2016), embora convocando também outros autores.

4.7.1. Cruzamento vocabular

Por cruzamento vocabular entende-se a adjunção de duas palavras existentes para formar uma palavra nova com supressão de material segmental de pelo menos uma delas e, em certos casos, sobreposição de segmentos, como *portunhol*_N (português_N + espanhol_N), *ofimática*_N (oficina_N + informática_N), *informática*_N (informação_N + automática_A). O cruzamento vocabular é muito frequente em textos literários e publicitários. A maioria dos produtos de cruzamento vocabular é nominal e adjetival pertencendo em geral a base à mesma categoria do produto. Este processo de formação de palavras é criado para denominar novas realidades, quer sejam entidades ou conceitos e exprimir uma avaliação. Assim, em *portunhol*, o constituinte hospedeiro é *português* e o qualificador é *espanhol*, pois o qualificador é, por certo, a forma estranha que se incorpora disfarçadamente na base hospedeira. O cruzamento vocabular é difundido e partilhado numa comunidade e depende de informação contextual, exceto quando é produzido em obra literária cuja significação é independente do contexto (Pereira, 2016, pp. 522-532).

Tem-se verificado que os produtos de cruzamento vocabular atestados resultam da junção de apenas duas bases. Visto que o conteúdo segmental das bases não é integralmente preservado, não são reconhecíveis constituintes morfológicos, há rutura da sequencialidade linear por meio de sobreposição, o produto do cruzamento vocabular constitui uma única palavra fonológica, perdendo-se a estrutura prosódica dos seus componentes, o cruzamento vocabular obedece a certas condições prosódicas, pelo que é de facto um processo que se situa na interseção da morfologia com a fonologia/prosódia (Pereira, 2016, p. 523).

Para que haja cruzamento vocabular é necessário obedecer a dois padrões basilares: Primeiramente, formas em que não existe semelhança fónica entre bases. Segundo, formas em que existe semelhança fónica entre as duas bases e em que, conseqüentemente, se verifica sobreposição. Isso determinará a localização de cada uma das bases e o ponto de fusão entre as duas. Do ponto de vista semântico, o cruzamento vocabular gera formas compostas com uma significação única resultante da combinação das bases (Pereira, 2016, pp. 524-530).

4.7.2. Truncação

Entende-se por truncação o processo de criação vocabular através do apagamento de material segmental de uma palavra-base, dando origem a uma forma diferente, que mantém o mesmo valor referencial, como *moto* < *motocicleta*, *cinema* < *cinematógrafo*. A truncação ocorre muito em registo familiar ou coloquial. Estes truncamentos podem ser feitos de três maneiras. A primeira é feita através de mecanismos redutivos dependentes de princípios fonológicos sem hipocorísticos, pois o material segmental apagado e o material segmental que se mantém não constituem unidades de nível morfológico, como *cusco* < *coscuvilheiro*, *tuga* < *português*. A segunda obedece a princípios morfológicos, ou seja, a sequência preservada constitui uma unidade morfológica da palavra-base e a segmentação é feita com base nessa identificação morfológica, mantendo a integridade segmental do constituinte, uma vez que essas palavras são, geralmente, criadas a partir de formas morfológicamente complexas, como *foto* < *fotografia*, *radio* < *radioterapia*. Na terceira que é a mais utilizada, a redução assenta em princípios de natureza fonológica e trata-se de um tipo de área lexical em que operam hipocorísticos pelo facto de permitirem maior variação de padrões de redução, como *Zé* < *José*, *Nela* < *Manuela* (Pereira, 2016, p. 532). Do ponto de vista categorial, a truncação pode afetar normalmente nomes (*prof* < *professor/a*) e adjetivos (*depre* < *depressivo*).

Sendo um processo de utilização frequente em português, a truncação obedece a alguns padrões de redução, de estrutura silábica e acentual do produto da redução. Encontram-se formas reduzidas bissilábicas (*níver* < *aniversário*), que são consideradas o padrão mais regular do truncamento, as formas trissilábicas (*eletro* < *eletrencefalograma*) e as formas monossilábicas, que ocorrem junto com a bissilábica em truncação com hipocorização (*Lu* < *Luísa*, *Bete* < *Elizabete*). Os hipocorísticos monossilábicos são muitas vezes gerados por seleção da sílaba tónica e apagamento de todo material segmental à sua esquerda e à sua direita, quando exista (*Gu* < *Augusto*, *Tó* < *Antónia*), cuja estrutura silábica é CV. No português europeu, as formas truncadas bissilábicas são paroxítonas. As formas trissilábicas obedecem ao mesmo padrão de segmentação: truncação à direita e manutenção da sequência segmental da esquerda da palavra matriz, por isso as formas resultantes do processo são paroxítonas. A truncação a esquerda (*noia* < *paranoia*) é muito escassa e comprova que não é um padrão regular na formação destes produtos. As formas bissilábicas ou trissilábicas mais frequentes são morfológicamente constituídas por um radical, que neste caso é a sequência segmental reduzida da palavra-base e por um constituinte temático que geralmente é *-a* ou *-o*, podendo variar em função do género da palavra. Em alguns casos, a forma resultante pode ser atemática ou mesmo com formatos fonológicos irregulares, como *biju* < *bijuteria*, *prof* < *professor(a)*. Considera-se que todas as formas com índice temático são

paroxítonas e as formas sem índice temático são oxítonas e não estão em conformidade com o mecanismo de truncação produtivo em português, que talvez tenham entrado na língua por empréstimo do francês. Os produtos de truncação têm o mesmo valor referencial da palavra que lhes dá origem, a diferença consiste em aspetos de natureza discursivo-pragmática e sociolinguística (Pereira, 2016, pp. 534-536). Apesar de se atribuir a formas truncadas um sentido pejorativo aproximando-as dos avaliativos, concedendo uma perspetiva diferente, Pereira (2016, p. 535) considera que o objetivo principal da truncação é ter uma carga expressiva superior à da palavra de que provém. Verificamos que a truncação também pode ser por elipse quando, em virtude da situação comunicacional que pressupõe a compreensão da elipse, a totalidade de um item lexical composto ou uma expressão sintagmática composta é representada por apenas um dos componentes, como *cachorro* < *cachorro-quente*, *ponta* < *ponta de lança direita ou esquerda* (Henriques, 2007, p. 126).

4.7.3. Reduplicação

A reduplicação consiste na repetição de uma sequência segmental para criar uma palavra nova, podendo essa transformação acarretar diferentes tipos de significação, nomeadamente gramatical ou lexical. Consequentemente, a reduplicação pode ser total, quando a palavra-base é repetida (*assim-assim* [Adv + Adv = N], *bombom* [A + A = N], *pula-pula* [V + V = N]), parcial, quando se nota apenas a repetição de uma parte da palavra-base (*vovó*, *papá*, *mamã*), reduplicação hipocorística (*Dadão*, *Zeze*, *Lulu*), reduplicação onomatopaica, quando não são criadas a partir de uma base léxica podendo ser considerada um recurso de natureza fonológica (*tiquetaque*, *zunzum*, *tautau*), reduplicação de sílaba ou sequência de sílabas sem ligação a uma base lexical identificável (*Jengalenga*, *gagá*, *blabláblá*), resultando numa interação entre a fonologia e a morfologia, e que alguns linguistas concebem como um tipo de afixação. Na reduplicação total, o constituinte prosódico copiado é a palavra prosódica, de tal maneira que a estrutura acentual do reduplicante é idêntica à da base. Daí resulta que as formas reduplicadas tenham dois acentos constituindo um sintagma fonológico. Tendo em conta as suas características estruturais, tanto morfológicas quanto fonológicas da reduplicação total, podemos aproximá-los dos compostos. A diferença entre elas consiste na regularidade semântica que só se verifica nos produtos de reduplicação e o facto de a relação semântica entre a base e o produto reduplicado não poder ser enquadrada no quadro tipológico da composição, pois o produto da reduplicação total é certamente diferente da das bases. Na reduplicação parcial, verifica-se uma simplificação da sílaba tónica da base no limite esquerdo da palavra e a sílaba da reduplicação parcial é, geralmente, de estrutura CV.

Na reduplicação hipocorística, as sílabas tónicas podem perder a coda (*Nonô* < *Leonor*, *Bebé* < *Isabel*) e não têm valor semântico referencial diferente da palavra original de que são sinónimos, usados em contextos específicos (Pereira, 2016, pp. 539-543).

4.7.4. Siglação e acronímia

É notável uma falta de concordância e explicitação entre os vários autores em relação à terminologia, definição e classificações, mesmo depois de se estabelecer o termo acrónimo (López Rúa, 2010, p. 336). Notando a dificuldade em estabelecer uma denominação, definição e classificação das abreviaturas, Geraldo Ortiz (2010, p. 73) defende que uma sigla pode ser percebida como uma unidade de redução formada por caracteres alfanuméricos procedentes de uma unidade léxica de estrutura sintagmática. De acordo com este autor, uma sigla forma uma sequência, cuja pronúncia pode ser alfabética, silábica ou ambas, e ocorre muito no âmbito técnico-científico e no léxico da política e economia.

A sigla não é um processo com estatuto morfológico, pois, além de ser intencional, não é um processo universal, sendo que nem todas as línguas têm siglas, como é o caso do *kimbundu*. Além disso, as novas palavras resultantes não mostram semelhanças sistemáticas entre o significado e o som com que os falantes de facto a reconheceriam (Haspelmath, 2002, p. 25). As siglas, sendo hiperónimos, e os acrónimos, sendo hipónimos constituem uma parte da extensão vocabular por serem um processo não regular e opaco em que não se reconhece a estrutura interna, sendo criados a partir de uma regra que estipula a extração da primeira letra de cada uma das palavras que constituem a expressão substituída ou simplificada. A forma como se faz a extração admite variações, sendo frequente a seleção de vários segmentos das bases para a constituição da sigla. A falta de consenso em torno da definição de siglação e da dificuldade em estabelecer uma terminologia está na origem do elevado grau de imprevisibilidade nestes produtos. As siglas obedecem a regras fonológicas no que respeita a estrutura acentual, a estrutura silábica e regras que afetam estruturas em contextos definidos. Tendo como critérios classificatórios a forma como são criadas e pronunciadas as unidades, a sigla pode ser própria ou mista. As siglas próprias resultam exclusivamente da extração das iniciais das unidades léxicas da estrutura sintagmática de base, cada constituinte mantém a sua proeminência acentual própria e a sigla funciona como um sintagma fonológico constituído pelas designações das letras cujas sílabas têm estruturas não marcadas (V ou CV), como CNE ‘Comissão Nacional Eleitoral’, BNA ‘Banco Nacional de Angola’. As siglas mistas são produtos em que se utilizam, além dos iniciais, caracteres secundários da estrutura de base ou em que se omitem partes fundamentais dessa estrutura.

As siglas mistas podem classificar-se em típicas, acrónimos e cruzamentos. As mistas típicas são aquelas em que se empregam ou omitem partes fundamentais da forma de base e cuja pronúncia pode ser silábica ou não, como HUC. As acronímicas são formadas por vários grupos de letras, e não apenas as iniciais, das unidades da forma de base e têm uma pronúncia exclusivamente silábica e a sequência é analisada como uma palavra morfológica cuja silabificação pode causar palavras regulares mais simples ou mais complexas (V, CV, CVC, CVV, CCVC, VCVCC), como INESC TEC ‘Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência’. As siglas mistas por cruzamento são formadas pela combinação de dois segmentos da estrutura sintagmática da base e que têm também pronúncia silábica (Pereira, 2016, pp. 543-545).

Há que distinguir acrónimos de abreviatura como *exc^a*, *apto*, etc. que possuem unicamente carácter gráfico de maneira que não se pronunciam, mas se interpretam. Além disso, as abreviaturas formadas por uma só letra refletem o plural da forma abreviada mediante a reduplicação da sua grafia tanto em abreviaturas simples (pp. ‘páginas’) como em abreviaturas compostas (AA.VV ‘vários autores’) (Arquiola, 2009, p. 80).

4.8. Processos de composição em Português

Se partirmos dos pressupostos teóricos apresentados por Fabb (2001, p. 66), Arquiola (2009, p. 71), Villalva (2003c, p. 971) e Mota (2013a, pp. 2823-2824), compreende-se que, não obstante os autores a associem ao domínio da sintaxe, a composição corresponde ao processo de formação de unidades léxicas em que se juntam dois ou mais constituintes – radicais ou palavras – que, na realidade, podem funcionar como bases, têm um acento primário comum, podendo remeter-nos para processos fonológicos, morfológicos, semânticos, como a metonímia, pois em muitos casos o resultado semântico seguramente não se subsume na simples adição dos semantemas das bases, as formas são dominadas por um acento primário comum, não podem ser separadas por outros constituintes e mantêm geralmente um significado composicional, um significado semelhante ao seu significado como bases isoladas, contudo com certas restrições. Se atentarmos nos compostos sintagmáticos, como *chave inglesa* e *olho de baleia*, podemos perceber que um composto pode carecer de composicionalidade semântica, de tal maneira que o seu significado não equivale literalmente à soma do significado de suas partes (Arquiola, 2009, p. 71).

Há várias tipologias de composição (Villalva, 2003c; Arquiola, 2009; Villalva, 2013; Mota, 2013a; Rio-Torto & Ribeiro, 2016; Caldas, 2016) tais como compostos morfológicos,

morfossintáticos e sintáticos. Há ainda a proposta de compostos greco-latinos/ compostos patrimoniais, compostos patrimoniais/ compostos léxicos e compostos sintagmáticos, compostos coordenantes e compostos subordinantes, e compostos endocêntricos e compostos exocêntricos (Arquiola, 2009, pp. 71-72). Compreendemos que a composição pode ser morfológica, morfossintática e sintagmática, pois esta proposta parece mais consensual (Rio-Torto & Ribeiro, 2016; Caldas, 2016). A composição é, por conseguinte, uma transformação sintática em expressão nominal (Bechara, 2006, p. 507).

4.8.1. Composição morfológica

Os compostos morfológicos, cuja estrutura é [RADICAL1 + VL + RADICAL2]_{Radical Complexo} (Mota, 2013a, p. 2811), incluem pelo menos um radical não autónomo, frequentemente de origem grega ou latina e caracterizam-se pela presença de uma vogal de ligação entre os respetivos elementos compositivos. São possíveis vários esquemas compositivos no âmbito dos compostos morfológicos (Rio-Torto & Ribeiro, 2016, pp. 476-477). Convém notar que a não autonomia funcional dos radicais em composição morfológica pode aproximá-los dos afixos derivacionais que, por definição, são formas presas. A diferença entre ambos consiste no facto de ser comum que os elementos compositivos possam ocorrer à direita ou à esquerda do composto. Rio-Torto e Ribeiro (2016, p. 477) apresentam três radicais que podem aparecer à esquerda ou à direita, nomeadamente *-antropo-*, *-fil-* e *-log-*. Nos nossos estudos, verificamos que, além destes, os radicais *-osteo-* e *-cardio-* também têm poder de mobilidade, ou seja, podem aparecer tanto à esquerda como à direita, conforme podemos verificar na Tabela 4.2.

Radicais	Inserção à esquerda	Inserção à direita
<i>-osteo</i>	<i>osteoporose</i>	<i>periósteo</i>
<i>-cardio</i>	<i>cardiologia</i>	<i>taquicardia</i>
<i>-logo</i>	<i>logorreia</i>	<i>biólogo</i>
<i>-antropo</i>	<i>antropofagia</i>	<i>misanthropo</i>
<i>-fil</i>	<i>filatelista</i>	<i>anglófilo</i>

Tabela 4.2 – Posição dos constituintes nos compostos morfológicos

Os radicais eruditos dos compostos caracterizam-se também por serem marcados categorialmente. São de facto muito frequentes em terminologias científicas e técnicas, como *hemograma*, *ignívomo*, *ornitologia*, *fusiforme* e precisam em alguns casos de uma vogal de ligação. A vogal de ligação é um resíduo de um marcador casual na estrutura dos compostos do latim e do grego antigo. É por esta razão que, no português, se encontram duas vogais de ligação, a saber *-o* e *-i*. Quando se combinam radicais greco-latinos, a origem do segundo elemento determina em parte o interfixo ou vogal de ligação. Se o radical neoclássico da direita é de origem grega, a vogal de ligação é *-o*, como *mareógrafa*, *musicologia*, *meritocracia*, *morfossintaxe*. Se o radical neoclássico da direita for de origem latina, a vogal de ligação será *-i*, como *vinícola*, *fumívomo*, *carbonífero*, *febrífugo*. Recorde-se que esta generalização não se aplica aos casos em que o radical da direita começa por vogal independentemente da sua etimologia, como *obtusângulo*, *geriatria*. Em muitos casos, acontece de facto a adjunção de dois radicais, como *quimioterapia*, *monoplegia*, *logomaquia*. Noutros, a generalização da vogal de ligação pode ser invertida, como *taxímetro*, *hidrófugo*. Há casos em que as duas formas existem, como *amperímetro*, *amperómetro*. Com efeito, a vogal de ligação na estrutura dos compostos morfológicos tem, efetivamente, a função de delimitador dos radicais. Os radicais greco-latinos agrupam-se geralmente por pares, um designa uma atividade e outro designa o agente ou tema, como *geologia/geólogo*, *parricídio/parricida*, *plutocracia/plutocrata*, *egolatria/ególatra* (Villalva, 2003c, pp. 972-978).

4.8.2. Composição morfossintática

Os compostos morfossintáticos são percecionados como estruturas que resultam da reanálise de uma estrutura sintática numa palavra, envolvem a combinação de duas palavras e caracterizam-se por algum grau de atipicidade relativamente aos padrões sintagmáticos do português ativos nas estruturas sintagmáticas correspondentes. Assim, os compostos morfossintáticos têm diferentes padrões sintáticos: [N + N]_N: *abóbora-menina*, *governos-sombra*, [A + A]_A: *trabalhador-estudante*, *morto-vivo*, [V + V]_N: *vaivém*, *pára-arranca*, [V + N]_N: *conta-quilómetros*, *guarda-joias*, [V + Pron]_N: *faz-tudo*, *sabe-tudo*. É notório que as estruturas [V + N]_N se afastam dos padrões sintáticos típicos, uma vez que geralmente o seu funcionamento como sintagma canónico exigiria a presença de um determinante a preceder o nome. Além disso, são caracterizados pela presença à direita de um nome singular, não massivo, que funciona como complemento da forma verbal usada à esquerda. Em [V + Pron]_N, notamos que há diferença entre compostos e singular homólogos através da impossibilidade de apagamento do núcleo em estruturas coordenadas, através da contradição semântica entre

o sintagma e o composto, como em *esse faz tudo e mais algumas coisas vs esse faz-tudo e mais algumas coisas*. Essas diferenças confirmam que os compostos morfossintáticos exibem uma configuração e significação fixas assentes na impossibilidade de inserção lexical no seu interior (Rio-Torto & Ribeiro, 2016, pp. 484-487).

À luz do que foi dito, se perscrutarmos os compostos morfossintáticos, verificamos como corolário que *escola-modelo* é uma única palavra na sintaxe, mas contém duas palavras morfológicas. A noção de palavra não é, portanto, a mesma em todas as áreas. Para o léxico, todas as noções de palavras são igualmente relevantes. Consideramos que as palavras são, decerto, conjunto de formas portadoras de informações fonológicas, morfológicas, sintáticas, semânticas e etimológicas, etc. (Villalva & Silvestre, 2014, pp. 76-77).

4.8.3. Composição sintagmática

Recorde-se que os compostos sintagmáticos se caracterizam por possuírem um significado unitário e são denominados por compostos sintagmáticos preposicionais (*chave de estrela*), compostos de nome e adjetivo (*guerra fria*) e compostos justapostos (*político-económico*) (Arquiola, 2009, p. 75). Os compostos sintagmáticos são unidades multilexicais, cuja estrutura segue os padrões próprios das estruturas sintáticas do português. Os compostos sintagmáticos têm, segundo Rio-Torto e Ribeiro (2016), os seguintes padrões estruturais: [N + Prep + N]_N: *rosa dos ventos, jardim de infância*; [N + A + Prep + N]_N: *processamento inteligente de texto*; [N + A]_N: *água benta, fita magnética*; [A + N]_N: *grande área, puro-sangue*; [N + Prep + V]_N: *máquina de lavar, ferro de engomar*; [Num + N]_N: *primeira-dama, segunda-comunhão*; [Num + Prep + N]_N: *quartos de final, oitavos de final* (Rio-Torto & Ribeiro, 2016, pp. 487-489). Constatamos que os compostos sintagmáticos, como sublinha Mendevil Giró (2009, p. 84), são um sintagma que funciona semântica e sintaticamente como uma palavra única.

Em alguns casos, os compostos sintagmáticos são denominados lexias (Bechara, 2006, pp. 506-507). Verifica-se que os compostos sintagmáticos podem conservar o seu próprio acento, funcionam como sintagmas livres e apresentam restrição do seu potencial combinatório. Tal restrição combinatória advém do nosso conhecimento do mundo e não do conhecimento da língua (Mendevil Giró, 2009, pp. 86-87). Depreendemos que, dentre muitos fatores, estes são caracterizados por cada constituinte conservar o seu acento (*puro-sangue*), pela formação do plural (*fitas magnéticas, jardins de infância*), poder apresentar modificações ou complementações (**olho de boi enfermo*), poder igualmente tornar impossível coordenar parte de um composto com outro elemento (**chave inglesa e estrela*) algo que na realidade é

possível com os sintagmas (*latas de conserva e de bebida*), tampouco podem ser alterados os seus elementos sem destruir o significado (*grande área/ área grande* ‘não se referem à mesma coisa’).

Os compostos sintagmáticos têm significado unitário, que não corresponde à soma do significado normal de seus componentes, e os seus elementos podem funcionar como sintagmas livres. Em todas estas características, compreende-se, assim, que os compostos sintagmáticos se comportam como palavras plenas, cujo significado é relativamente arbitrário, já dado, que se aprende de memória (Arquiola, 2009, pp. 75-76). Por isso, os compostos sintagmáticos, como sublinha Mel'čuk (2006, p. 14) e Mendevil Giró (2009, pp. 83-84), também são considerados como palavras com estrutura externa.

4.9. Síntese do capítulo

Neste capítulo, pudemos verificar que a sistematicidade do léxico provém das suas relações paradigmáticas e sintagmáticas. Uma língua compõe-se de dicionário, que pode ser um repertório de palavras, e de uma gramática, a qual pode, de facto, equivaler a um conjunto de dispositivos que agrupando, distribuindo e modificando as palavras do dicionário, dão lugar a frases de uma língua. Embora o conceito de palavra hodiernamente ainda não seja consensual tanto do ponto de vista interlinguístico como da perspectiva de uma língua concreta, debruçamo-nos sobre a estrutura do léxico, a conceptualização de palavra, os processos de inovação lexical tais como os tipos de derivação, conversão, processos de construção não concatenativa e processos de composição em português.

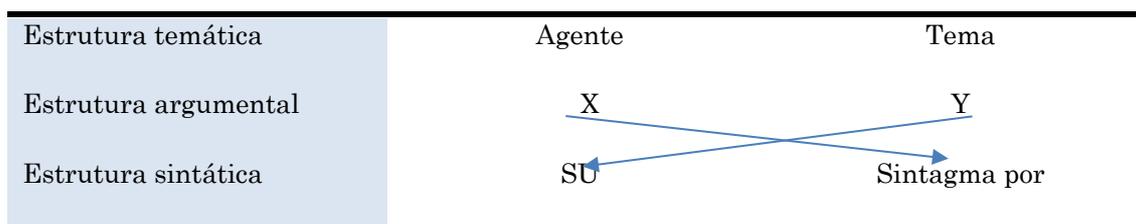
Capítulo 5 – Estruturas Passivas e Estruturas com *Se*

Este capítulo é uma continuação do trabalho anterior na área do léxico e tem cinco objetivos. Explicitamos, num primeiro momento, a passiva eventiva, sua estrutura temática, estrutura argumental e a estrutura sintática. Apresentamos, de seguida, uma breve caracterização das estruturas passivas resultativas e estativas. Num segundo momento, descrevemos a estrutura temático-argumental das estruturas passivas de *se* e apresentamos o *se* como clítico argumental de referência arbitrária. Num terceiro momento, descrevemos a organização e materialização da estrutura temático-argumental das estruturas reflexas e recíprocas. Além disso, debruçamo-nos sobre o *se* anafórico, o *se* indeterminador e o *se* apassivador quanto à sua capacidade referencial. De seguida, apresentamos o *se* como partícula de transitivização do verbo e como clítico sem conteúdo semântico.

5.1. Estruturas passivas

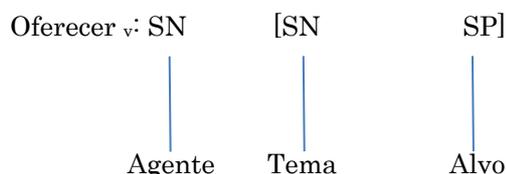
Pretendemos fazer uma caracterização das estruturas passivas eventivas, resultativas, estativas e passivas de *se* e outros valores de *se* numa perspetiva contrastiva – Português Europeu e Português de Angola, com o auxílio de métodos computacionais, que serão demonstrados no capítulo onze. No entanto, não se trata de um problema de fácil resolução, pois há imensas formas que podiam ser estudadas. Com efeito, decidimos analisar a literatura sobre o assunto e detetamos que alguns autores como Peres e Mória (1995), Mendikoetxea (1999), Mingas (2000), Duarte e Brito (2003), Duarte (2003a), Embick (2004), Mendes e Estrela (2008), Brito (2011), Ribeiro (2011), Duarte (2013), Estrela (2013), Hagemeyer (2016) estudaram as estruturas passivas, os valores de *se*, esquemas relacionais e ordem das palavras. Em virtude disso, decidimos estudar as estruturas passivas numa perspetiva contrastiva, tendo em conta os esquemas relacionais e ordem das palavras, com o auxílio de métodos computacionais.

As orações com diátese ativa ou passiva descrevem uma situação, que pode ser, sob o ponto de vista semântico, evento ou estado. A diferença entre as duas diáteses consiste no facto de a situação na ativa ser perspectivada a partir da entidade com o papel temático externo; enquanto na passiva se perspectiva a situação descrita pela frase a partir da entidade com o papel temático interno. As gramáticas gregas já falavam sobre a diátese ativa e passiva. A estas acrescentavam a diátese média como uma categoria intermédia, pois reunia características da diátese ativa e da passiva. As estratégias de uso do *se* como clítico, em Angola, revelam algumas estruturas estruturais.



Quadro 5.1 – Passiva verbal: estrutura temático-argumental, Ribeiro (2011)

No que se refere à estrutura sintática, convém lembrar que, como mostra Jackendoff (1991), há relação entre a seleção categorial e a seleção semântica e obedece-se a uma hierarquia da estrutura argumental. A título ilustrativo, notamos que a frase (1.a) tem um predicado ternário e obedece à hierarquia temática que pode ser esquematizada do seguinte modo:



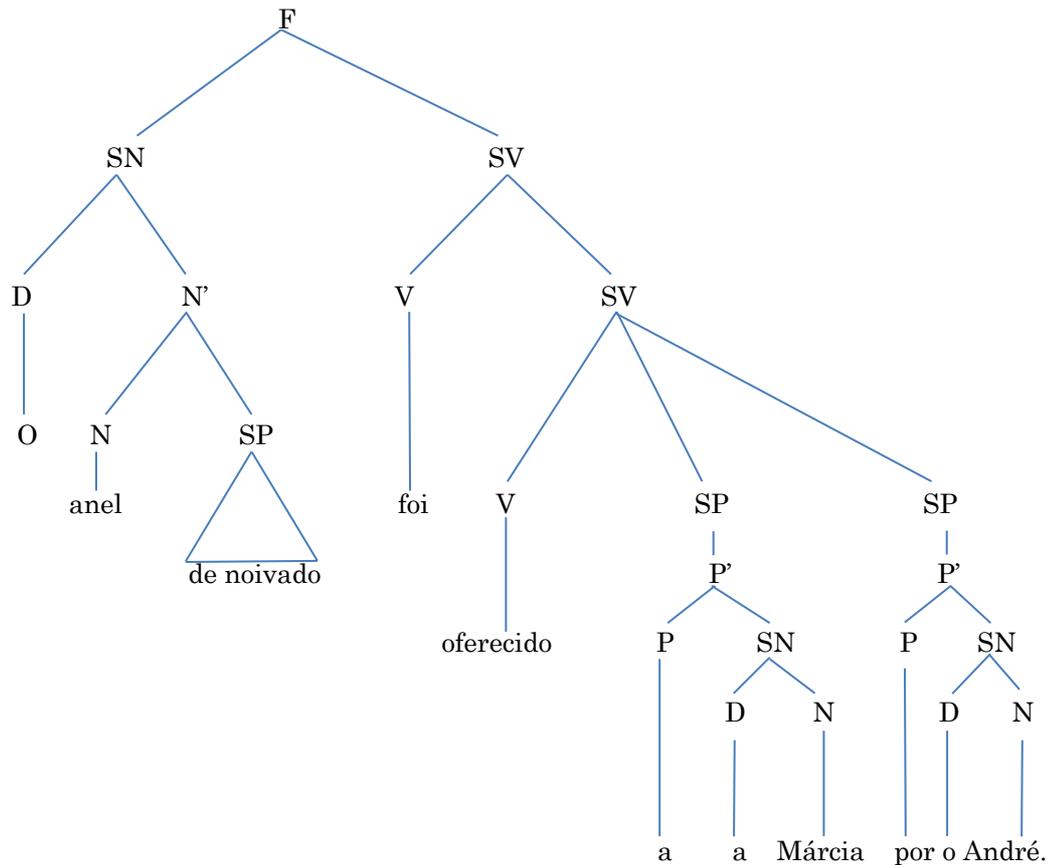
1. Agente
2. Tema
3. Alvo/Locativo¹³

Nas passivas, além de ocorrer uma forma auxiliar de *ser* no mesmo tempo e modo do verbo pleno da ativa correspondente, o verbo pleno da oração ativa assume na passiva correspondente a forma de particípio passado e concorda em número e género com o sujeito (Duarte, 2013). É igualmente notável a constância de papéis temáticos entre o sujeito da passiva e o objeto direto da ativa correspondente e entre o complemento agente da passiva e o sujeito da ativa correspondente (Duarte, 2003b). Portanto, a estrutura argumental do verbo é, sob o ponto de vista sintático, determinante, quando se trata de restrições passivas. Numa perspectiva semântica, fatores como a estatividade e agentividade do verbo também são determinantes. Julgamos importante sublinhar que o complemento agente da passiva é de

¹³ Certos verbos admitem que o argumento externo possa ter mais do que um papel temático. A título ilustrativo, isso acontece com o verbo matar – que pode selecionar um argumento externo com papel temático de Agente ou Fonte – e o verbo partir, que pode selecionar um argumento externo com o papel temático de Agente ou Experienciador (Duarte & Brito, 2003).

caráter opcional. As orações passivas em que ocorre o complemento agente da passiva são, por esta razão, denominadas passivas longas. As orações passivas em que o agente da passiva não está expresso são denominadas passivas curtas, nas quais o papel temático externo está implícito e pode ser comprovado através da existência de advérbios orientados para o agente, tal como é o caso de *atentamente* (Duarte, 2013).

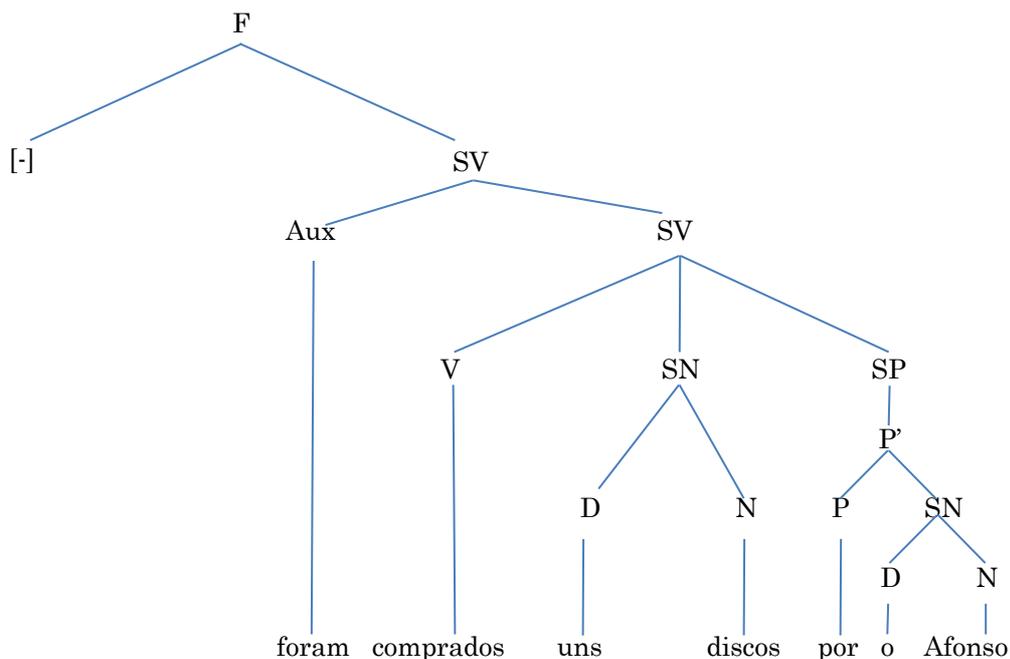
Na oração passiva eventiva, a estrutura temática mantém-se. Há, porém, uma estrutura sintática e informacional diferente. Este tipo de alinhamento faz das orações passivas um caso particular de orações intransitivas e de orações inacusativas. Só podem ocorrer em orações passivas eventivas os verbos de predicado binário ou ternário em que o argumento que se realiza como sujeito da frase tem o estatuto de argumento direto na entrada lexical do verbo. Partindo do modelo apresentado por Baker (2001), as categorias que funcionam como complemento ficam à direita do núcleo e os especificadores à esquerda. Podemos, segundo a Sintaxe Generativa, notar a seguinte representação arbórea da frase (2).



A frase com a representação arbórea anterior é, como advoga Duarte (2013), uma passiva pessoal, pois o sujeito dela é realizado na posição canónica pré-verbal. O sujeito da

oração passiva eventiva pode estar em posição pós-verbal, em especial se for uma expressão indefinida ou um sintagma nominal reduzido. Quando isso acontece, a oração é denominada passiva impessoal. Para ilustrar, atentemos na frase, cuja descrição é feita a partir das propostas de análise de Raposo (1992) e Wasow (2001).

(3) «Foram comprados uns discos pelo Afonso.»



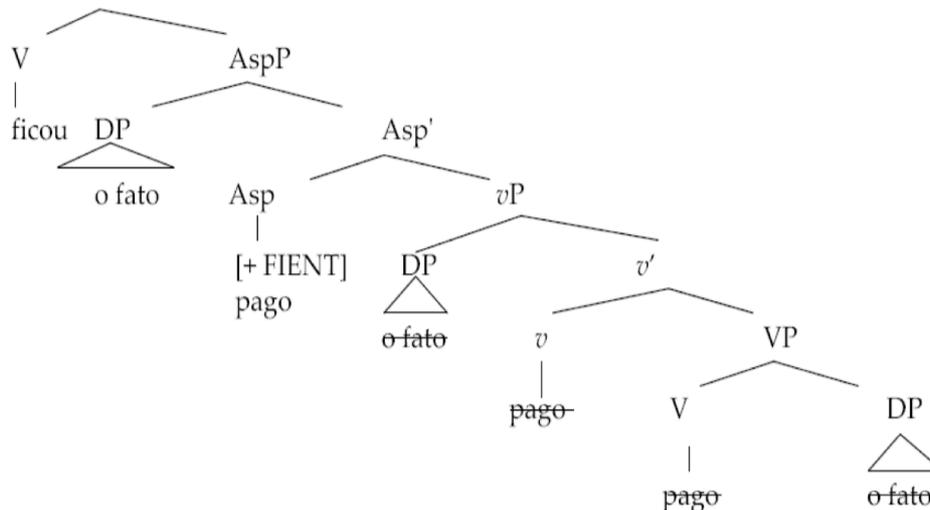
Nas passivas impessoais, a posição do sujeito frásico não está preenchida e o argumento interno ocupa a posição básica, característica do argumento interno nominal do verbo (Estrela, 2013).

5.1.2. Passiva resultativa

As passivas resultativas são, como sublinha Duarte (2013, p. 440), aquelas em que ocorre um grupo verbal complexo iniciado pelo verbo auxiliar *ficar* seguido de um particípio com o estatuto verbal ou adjetival correspondente ao verbo pleno da oração ativa. As estruturas passivas resultativas descrevem uma situação que é o resultado de uma mudança de estado, lugar ou posse (Embick, 2004). Marques (1998) defende-as como passivas de estado resultativo. Estas são caracterizadas pelo facto de terem uma componente eventiva e ausência de componente agentiva. Com efeito, estas estruturas não admitem geralmente a realização do argumento externo do verbo através de um agente da passiva, conforme os exemplos retirados do *CETEMPúblico*.

- (4a) «O contrato-programa que as criou agora **ficou desfalcado**.» (*par=ext552909-clt-93a-1*)
- (4b) «Por outro lado, os investidores **ficaram animados** com o corte que Bill Clinton propôs para o orçamento das forças armadas.» (*par=ext347322-eco-93a-1*)
- (4c) «O fato **ficou pago**.» (Duarte & Oliveira, 2010)
- (5a) «Os originais, por decisão de Camila Miguéis, **ficaram depositados** na John Hay Library, da Universidade de Brown.» (*par=ext552909-clt-93a-1*)
- (5b) «A encomenda **ficou entregue**.» (Duarte, 2013)
- (6a) «O pequeno **fica entregue** à avó, que, por vezes, «não tem mão» no rapaz.» (*par=ext1017605-soc-98a-2*)
- (6b) «O polémico Ministério da Economia e Finanças **fica entregue** a Carlos Quaresma, um jovem economista que se irá estrear na esfera governativa.» (*par=ext1077427-pol-94b-1*)

Repare-se que, nestas construções, o argumento externo não está implícito e descrevem-se situações de mudança de estado, mudança de lugar e mudança de posse nos exemplos 4, 5 e 6 respetivamente. Segundo o modelo de Embick (2004), a passiva resultativa em (4c) pode ter a seguinte representação sintática (Duarte & Oliveira, 2010).



Podemos constatar que outras particularidades das passivas resultativas é que, em virtude da ausência de uma componente agentiva, não admitem advérbios que pressupõem o agente, não coocorrem com orações subordinadas finais sem sujeito expresso e não admitem sintagmas preposicionais de valor instrumental.

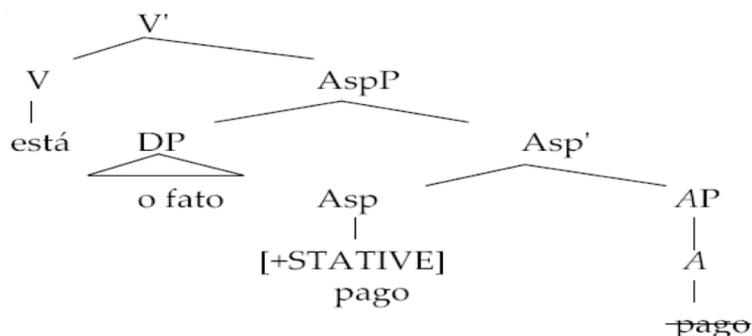
Recorde-se que as estruturas passivas eventivas e as passivas resultativas descrevem diferentes fases de situações dinâmicas. Enquanto as passivas eventivas focalizam, como sublinha Duarte (2013, pp. 440-443), a fase do processo de mudança de estado, lugar ou posse, as passivas resultativas focalizam o estado que resulta dessa mudança, perspetivando-o como uma consequência da mudança. Com efeito, as passivas eventivas e passivas resultativas podem ocorrer quer com predicados estáveis quer com predicados episódicos.

5.1.3. Passiva estativa

As passivas estativas são, como sublinha Duarte (2013, p. 440), aquelas em que ocorre um grupo verbal complexo iniciado pelo verbo auxiliar *estar* seguido de um particípio com o estatuto adjetival correspondente ao verbo pleno da oração ativa, conforme se pode verificar nos exemplos retirados do *CETEMPúblico*.

- (7a) «A Estrada Nacional 368, Tapada-Alpiarça, **estava submersa** devido à falta de drenagem da água, enquanto o rio Sorraia se apresentava nos seus limites.»
(*par=ext899746-soc-97b-2*)
- (7b) «Em princípio, as grandes empresas **estão organizadas** para fornecer este tipo de informações.» (*par=ext546639-soc-91b-1*)
- (7c) «O fato **está pago**.» (Duarte & Oliveira, 2010)

Partindo do modelo de representação de participios estativos proposto por Embick (2004), notamos que a possibilidade de prefixação com *i*(*inato, incompleto, incorreto, inculto*), é apenas produtiva com particípio estativo. A passiva estativa em (7c) pode ter a seguinte representação sintática (Duarte & Oliveira, 2010; Alexiadou & Anagnostopoulou, 2008).



Repare-se que a impossibilidade de ocorrência de expressões *em x tempo* revela que as estruturas passivas estativas têm uma natureza não eventiva ou atélica. Quando a

expressão *em x tempo* coocorre em passivas estativas, a interpretação primeira da expressão *em x tempo* é *daqui a x tempo* (Duarte, 2013).

(7d) «Se tudo correr bem, o filme está pronto em sete meses.» (*par=ext1013450-nd-95a-1*)

Pelo exposto acima, é evidente que as passivas eventivas são caracterizadas por uma componente agentiva e por uma componente eventiva. As passivas resultativas não têm componente agentiva, mas têm componente eventiva. As estruturas passivas estativas são na realidade caracterizadas por não terem componente agentiva nem componente eventiva (Duarte, 2013, p. 444). Estas propriedades são sintetizadas no Quadro 5.2.

	Passiva eventiva	Passiva resultativa	Passiva estativa
Componente eventiva	+	+	–
Componente agentiva	+	–	–

Quadro 5.2 – Caraterização das componentes nas estruturas passivas

5.2. Análise dos valores de *se*

Entende-se por predicar a atribuição de propriedades ou o estabelecimento de relações entre entidades (Duarte & Brito, 2003). Nesta secção, descrevemos os valores de *se* como impessoal, anafórico, decausativo e inerente. Reconhecemos que o morfema *se* pode, de facto, ter o valor de conjunção completiva e condicional autêntica ou potencial, condicional real ou factual, condicional irreal ou contra factual e condicional ilocutória (Maingueneau, 1997). Todavia, estes valores conjuncionais não são contemplados neste estudo. O português, não sendo realizado da mesma forma ao longo da sua extensão territorial, é de facto objeto de alguma reflexão. Os acontecimentos históricos, os contactos com falantes de outras línguas ou o tempo determinam que a língua, progressivamente, se diferencie e que cada região em que é falada desenvolva traços próprios.

5.2.1. *Se* impessoal

O *se* é usado com um valor impessoal quando não se pretende ou não se consegue identificar com precisão a entidade subjacente à situação descrita. Assim, este *se* ocorre seguramente nas estruturas passivas e nas estruturas de sujeito indeterminado (Mendes & Estrela, 2008).

(8a) «Vendeu-se muitas casas naquele bairro.»

↓
nominativo

(8b) «Estas casas venderam-se ontem.»

↓
Passivo

Estruturas passivas com o “se”

A passiva de *se* ou passiva pronominal é, como advoga Duarte (2013), aquela cuja diátese passiva é expressa através do pronome átono de 3.^a pessoa *se*, sem qualquer verbo auxiliar ou morfologia verbal especial no verbo pleno. Este *se* tem um estatuto quase-argumental e funcional (Brito, Duarte, & Matos, 2003). Notemos a estrutura temático-argumental e respetiva materialização da passiva de *se*.

Estrutura semântica	Agente	Tema
Estrutura argumental	X	Y
Estrutura sintática	SU	SE

Quadro 5.3 – Estrutura temático-argumental das passivas de *se*, Ribeiro (2011)

Além da constância do papel temático na transformação de uma oração ativa e uma passiva, uma das semelhanças entre as passivas eventivas e as passivas pronominais é, por um lado, o facto de admitirem expressões adverbiais que pressupõem um agente, incluindo ainda advérbios como *intencionalmente*, *propositadamente*, *voluntariamente* e orações subordinadas finais (Duarte, 2013). Podemos notar isso nos seguintes exemplos extraídos de Duarte (2013, pp. 444-445).

(9a) «O raciocínio algébrico é ensinado propositadamente na escola.» (Duarte, 2013)

(9b) «O raciocínio algébrico ensina-se propositadamente na escola.» (Duarte, 2013)

(9c) «As obras mais recentes foram publicadas numa editora espanhola para garantir maiores tiragens.» (Duarte, 2013)

(9d) «As obras mais recentes publicaram-se numa editora espanhola para garantir maiores tiragens.» (Duarte, 2013)

É, por outro lado, importante sublinhar que as duas passivas têm ainda semelhanças importantes, como assevera Mendikoetxea (1999), «las pasivas de *se* comparten con las pasivas perifrásticas el hecho de que tienen como sujeto gramatical al objeto de la oración activa.»

Dissemelhantemente da passiva eventiva, nas passivas pronominais, o agente não pode normalmente aparecer especificado num sintagma preposicional (Mendikoetxea, 1999).

(10a) «Compraram-se todos os discos de Il Divo.»



(10b) «*Compraram-se todos os discos de Il Divo pelos estudantes.»

Como se pode notar na frase (10a), é importante realçar que o *se* com valor passivo tem por referente uma entidade arbitrária identificada com o agente da passiva. O SN pós-verbal é o SU da frase (Brito, Duarte, & Matos, 2003).

O verbo tem de ser transitivo, como na ativa correspondente. Nesta estrutura, o clítico *se* suspende a atribuição de relação temática de agente à posição de argumento externo e de caso acusativo ao argumento interno do verbo, pois o *se* absorve o acusativo. Notemos o seguinte quadro em que se apresenta uma síntese das estruturas passivas.

		TIPOS DE PASSIVA			
		Eventiva	Resultativa	Estativa	Pronominal
PROPRIEDADES	Expressão da diátese passiva	Verbo «ser»	Verbo «ficar»	Verbo «estar»	Clítico «se»
	Forma do verbo pleno	Participial	Participial	Participial	Sempre na 3.ª pessoa
	Posição do sujeito	Pré ou pós-verbal	Pré ou pós-verbal	Pré ou pós-verbal	Em geral, pós-verbal
	Sintagma preposicional agentivo	Admite-o	Dispensa-o obrigatoriamente	Dispensa-o obrigatoriamente	Dispensa-o obrigatoriamente

Quadro 5.4 – Diferenças entre a passiva eventiva, resultativa, estativa e pronominal

Estruturas de sujeito indeterminado

As estruturas de sujeito indeterminado constroem-se a partir do recurso a formas verbais de 3ª pessoa do singular; com efeito, o uso de formas verbais na 3ª pessoa do singular é obrigatório, quando o verbo é acompanhado do clítico *se*, pois é um clítico argumental de referência arbitrária; o recurso a infinitivos não flexionados; por *a gente* ou por expressões lexicais nominais ou pronominais impessoais – como *eles, as pessoas, alguém* (Ribeiro, 2011). Notemos a diferença entre a estrutura transitiva lexicalmente plena e a estrutura transitiva de sujeito indeterminado.

	ESTRUTURA TRANSITIVA LEXICALMENTE PLENA		ESTRUTURA TRANSITIVA DE SUJEITO INDETERMINADO	
	“O andré dança Kizomba.”		“Dança-se Kizomba.”	
EST. TEMÁTICA	Agente	Tema	Agente	Tema
EST. ARGUMENTAL	X	Y	X	Y
EST. SINTÁTICA	SU	OD	SU	OD
	O André	Kizomba	SE	Kizomba

Quadro 5.5 – Estrutura temático-argumental dos predicadores transitivos em estruturas transitivas plenas e estruturas de *se* nominativo, adaptado de Ribeiro (2011)

O constituinte com a relação gramatical de sujeito pode ser um argumento externo de predicadores verbais transitivos (11a) e inergativos (11b); pode ser o argumento interno direto de predicadores verbais inacusativos (11c); e o argumento externo do predicador secundário em estruturas copulativas (11d).

(11a) «[A condução prolongada]_{SU} provoca fadiga.»

(11b) «[A Joana]_{SU} espirrou.»

(11c) «Chegado [o João]_{SU}. vs [O João]_{SU} chegou.» (Duarte, 2003a)

(11d) «[O André]_{SU} é engenheiro.»

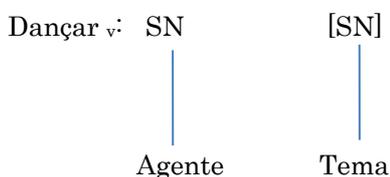
Quando usamos o clítico *se* como sujeito indeterminado – também denominado sujeito impessoal, indeterminado ou sujeito com interpretação arbitrária – podemos parafraseá-lo por expressões nominais como *alguém* ou *uma pessoa* (Brito, Duarte, & Matos, 2003).¹⁴ Em

¹⁴ Usa-se o clítico nominativo *se* acompanhado da 3.ª pessoa no singular de um verbo para exprimir o sujeito com interpretação arbitrária pois, no português, não existe de facto um pronome tónico para o exprimir (Duarte, 2003a).

alguns casos, este *se* absorve o caso nominativo. Para exemplificar, notemos a seguinte frase:

(12a) «[Aqui]_{Locativo} [o André]_{Agente} dança [kizomba]_{Tema} intencionalmente.»

(12b) «[Aqui]_{Locativo} dança-[se]_{Agente} [kizomba]_{Tema} intencionalmente.»



Trata-se precisamente de uma frase cujo verbo *dançar* é inergativo de atividade física e o argumento sujeito (12b) tem uma interpretação arbitrária. Além disso, a presença do advérbio orientado para o agente – *intencionalmente*, também substituível por *voluntariamente* e *propositadamente* – não nos dá margem de dúvida para deduzir que o *se* absorve o papel temático do argumento externo que o verbo tem para atribuir. Tanto pode ser agente como outro (Duarte & Brito, 2003).

Assim, o clítico *se* aparece na frase acima como símbolo do constituinte com a relação gramatical de sujeito indeterminado, ao qual se atribuiu, portanto, o caso nominativo (Duarte & Brito, 2005). Temos um predicado unário que, nesse caso, denota um processo, pois descreve uma situação dinâmica e atética, tem duração, é homogénea, não tem estado consequente e admite a expressão «durante x tempo» (Oliveira, 2003). Teremos a seguinte frase, caso verifiquemos a classe aspetual: «Aqui dança-se intencionalmente durante três horas.»

Usando o teste de substituição, podemos substituir o *se* por *alguém*, caso a frase seja afirmativa, ou *ninguém*, caso seja negativa. Teríamos assim a seguinte construção: «Aqui alguém dança intencionalmente.» Temos, mesmo assim, o sujeito indeterminado *alguém* (Mendikoetxea, 1999). Caso usássemos uma oração negativa, teríamos a seguinte construção: «Aqui ninguém dança intencionalmente.»

O tipo de sujeito não foi alterado. Poderíamos continuar a frase com uma adverbial final, a fim de identificar o agente. Teríamos, assim, a seguinte frase:

(13) «Aqui dança-se intencionalmente para causar uma boa impressão.»

Em alguns casos o *se* nominativo não tem o valor de agente. A título ilustrativo, notemos as seguintes frases:

(14) «Aqui, morre-[se]_{Tema} muito.»

(15) «Aqui, nasce-[se]_{Tema} pouco.»

(16) «Aqui, ama-[se]_{Experienciador} muito.»

Nestas frases de predicado unário, o *se* nominativo não tem o papel de agente. Aparece apenas como símbolo do constituinte com a relação gramatical de sujeito indeterminado, ao qual se atribuiu, portanto, o caso nominativo.

(17) «Aqui, morre-se muito às 5h da manhã.»

(18) «Aqui, nasce-se pouco às 2h da tarde.»

Convém relembrar que o *se* nominativo não admite construções de redobro de clítico. É obrigatoriamente referencial, e, por esta razão, não ocorre associado a uma posição de pronome expletivo, tal como se nota na agramaticalidade da seguinte frase: *«Aqui, chove-se.» (Brito, Duarte, & Matos, 2003).¹⁵

5.2.3. *Se* anafórico

Trata-se de construções em que o clítico *se* está sob c-comando do respetivo antecedente. Partindo deste pressuposto, uma anáfora tem de ter o seu antecedente dentro da oração a que pertence; não pode ela própria ocupar a posição de sujeito; entre o antecedente e a anáfora não pode interpor-se uma expressão nominal que seja, ela própria, um sujeito.

O *se* anafórico é um argumento interno do predicador verbal em estruturas reflexas e recíprocas. Estabelece, assim, relações gramaticais de objeto direto ou objeto indireto dentro do domínio sintático de predicação. O *se* reflexo e o *se* recíproco ocorrem em estruturas que descrevem situações com características diferentes, nomeadamente no que se refere ao número de intervenientes nelas envolvidos, às relações que entre si mantêm e ao tipo de papéis temáticos a que estão associados.

Dependendo da estrutura temática do predicador verbal, este *se*, tendo a relação gramatical de acusativo ou dativo, tem o papel temático de tema ou alvo. Este, apesar de ter estatuto argumental, carece de autonomia a vários níveis e revela uma configuração lexical compactada. Este *se* é destituído de acentuação própria, nunca tem capacidade referencial autónoma e, por isso, é sempre dependente de um antecedente, necessariamente pertencente à mesma frase com capacidade para fixar o seu valor referencial.

¹⁵ Importa sublinhar que, segundo Mendikoetxea (1999), «las impersonales con *se* se asemejan más a las oraciones activas asociadas que a las pasivas perifrásticas en cuanto a la realización sintáctica del objeto nocional como objeto gramatical.»

Estruturas reflexas

Trata-se de estruturas que descrevem uma situação em que uma entidade age sobre si própria, fazendo com que a ação que inicia se reflita em si mesma. As estruturas reflexas agentivas volitivas são as reflexivas prototípicas. O constituinte com a relação gramatical de sujeito é, em estruturas reflexas, agente e paciente (Ribeiro, 2011). Podemos notar isso na seguinte frase.

(19) «A Maria lavou[-se] _{Tema}.»

O *se* reflexo ocorre em estruturas transitivas. Nestas estruturas, o *se* pode ter a relação gramatical de objeto indireto, quando está com verbo bitransitivo (Bechara, 1999).

Quando se usa o *se* com o valor de clítico dativo, o mesmo não ocorre na sua posição canônica (Brito, Duarte, & Matos, 2003). Para uma compreensão disso, notemos a seguinte frase:

(20) «[A Maria] _{Agente Paciente} deu[-se] _{Alvo} a si própria [um presente] _{Tema}»



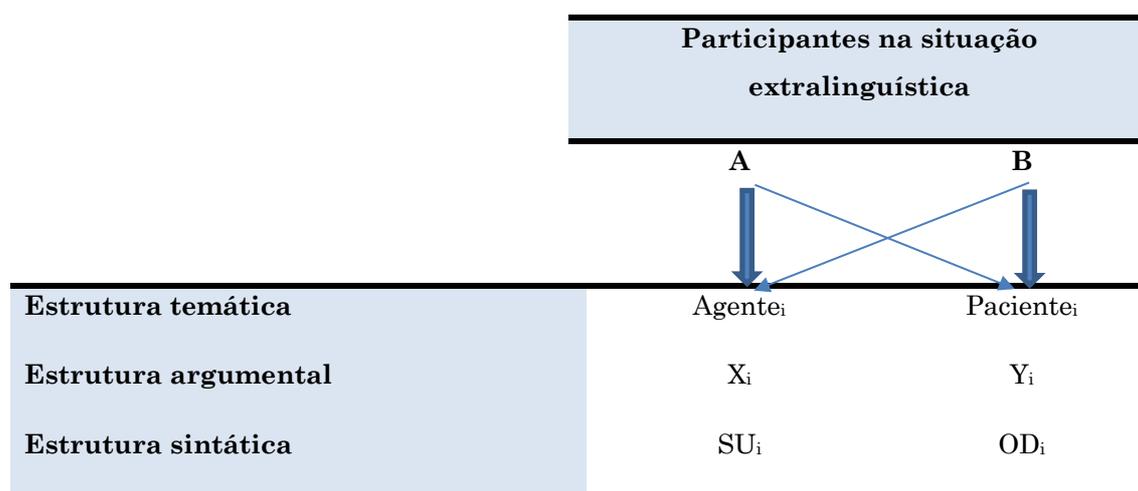
Temos, por certo, um predicado ternário, pois, tendo em conta a seleção categorial e semântica, o verbo é bitransitivo e seleciona, portanto, três argumentos: o argumento externo tem a relação gramatical de sujeito, um argumento interno tem a relação de objeto direto e outro, sendo o *se*, é um argumento interno com a relação gramatical de objeto indireto. Caso o predicador verbal fosse transitivo, como na frase (19), teríamos o esquema do Quadro 5.6.

Constituintes		
Estrutura temática	Agente _i	Paciente _i
Estrutura argumental	X _i	Y _i
Estrutura sintática	SU _i	OD _i

Quadro 5.6 – Organização temático-argumental das estruturas reflexas, Ribeiro (2011)

Estruturas recíprocas

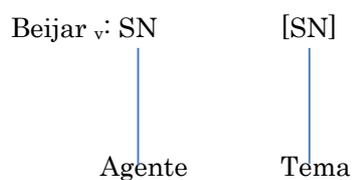
As estruturas recíprocas são construções verbais que codificam situações em que participam pelo menos dois intervenientes que realizam a mesma ação um sobre o outro, ou seja, A age sobre B e B age sobre A. Notemos a sua representação.



Quadro 5.7 – Organização temático-argumental das estruturas recíprocas, Ribeiro (2011)

O *se*, além de reflexo verdadeiro, tendo em conta a perspetiva de Bechara (1999), tem a relação gramatical de acusativo, quando está com verbo transitivo direto tanto na voz reflexiva como na recíproca. Atentemos na frase que se segue e, posteriormente, analisemo-la para se esclarecer o valor da estrutura em estudo como tema e clítico acusativo:

(21) «[O Reuel e a Acsa] _{Agente e Paciente} beijaram·[se] _{Tema} durante dois minutos.»



Pelo seu conteúdo, percebemos que o processo descrito no domínio de predicação é efetuado pelas duas entidades denotadas pelas expressões referenciais. O predicador verbal é um verbo transitivo, pois seleciona um argumento externo com a relação gramatical de sujeito e um argumento interno com a relação gramatical de objeto direto (Reinhart & Siloni, 2005). O argumento externo é ao mesmo tempo agente e paciente. O *se*, sendo argumento interno, tem a relação gramatical de objeto direto. Ilustramos a visualização da estrutura de constituinte de (21).

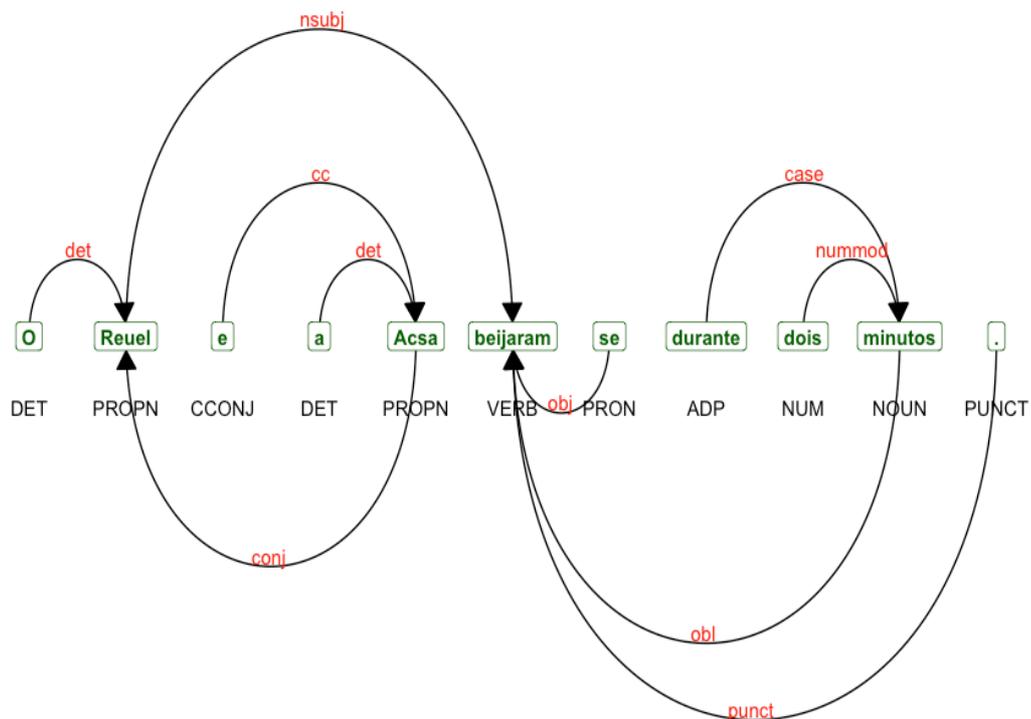


Fig. 5.1 – Tokenisation, POS Tagging e Dependency Relation de estrutura recíproca

Assim, o *se* acusativo para aparecer em estruturas recíprocas, tem de satisfazer algumas condições, tais como: o sujeito da oração, por um lado, deve ser plural; por outro, o sujeito e o verbo devem necessariamente exibir as mesmas marcas de concordância (Brito, Duarte, & Matos, 2003). Além disso, este *se* pode coocorrer em estruturas reflexas. A título ilustrativo, podemos ver na seguinte frase. «O Paulo elogiou-se em frente de toda a turma.» Julgamos importante apresentar o quadro dos três valores de *se* até agora abordados.

	SE anafórico	SE indeterminador	SE apassivador
Capacidade referencial autónoma	–	+	+
Referência delimitada/identificável	+	–	–

Quadro 5.8 – Caracterização do SE anafórico, SE indeterminador e SE apassivador quanto à sua capacidade referencial, Ribeiro (2011)

5.2.4. Alternância causativa: *se* decausativo

Este *se* também é denominado clítico ergativo, anticausativo ou inacusativo. Esta denominação advém, por um lado, do facto de a sua ocorrência inibir a presença do argumento externo (sujeito) do verbo a que se associa, o qual deteria as relações temáticas de Causador ou de Agente (Brito, Duarte, & Matos, 2003). Por outro lado, ocorrendo precisamente com verbos transitivos, a sua função fundamental é, acima de tudo, destransitivizar o verbo. Visto que o verbo é o seu hospedeiro, o *se* como clítico ergativo comporta-se como um sufixo derivacional destransitivizador. Este *se* pode ser omitido. A título ilustrativo, atentemos na seguinte frase.

(22a) «[A cadeira] Tema partiu(-se).»

(22b) «[A Maria] Agente partiu [a cadeira] Tema.»

O verbo *partir* apresenta, geralmente, uma estrutura argumental binária (22b), cujo argumento externo é um Agente e o argumento interno é um Tema. No exemplo (22a), o verbo *partir* apresenta-se como intransitivo, cuja partícula destransitivizadora e sem qualquer valor argumental nem conteúdo referencial é o clítico *se* (Mendes & Estrela, 2008). Sob o ponto de vista semântico, a frase expressa uma culminação, pois descreve uma situação dinâmica, télica, que tem estado consequente, não tem duração nem é homogénea, e admite a expressão «a x tempo» (Duarte & Brito, 2003). A título ilustrativo, é gramatical a seguinte estrutura:

(23) «A cadeira partiu às 9h da manhã.»

O SN que ocupa a posição de SU (pré ou pós-verbal) corresponde ao argumento interno do predicador verbal (daí a aproximação com a passiva); mas, diferentemente da passiva, é incompatível com a presença de agente, mas não com a causa.

(24) «A cadeira partiu-se por causa do peso do André.»

Este *se* tem tendência a ser omitido, inclusive no Português de Angola como mostramos em Muhongo (2019a); ao contrário da passiva, o uso deste *se* é limitado a certos verbos transitivos, parecendo ser um mecanismo lexical e não transformacional, idiossincrático de certos verbos. Tendo o valor de sufixo derivacional destransitivizador, o clítico *se* como decausativo, normalmente, não ocorre em construções de redobro de clítico, tal como se pode notar na agramaticalidade de (25).

(25) *«A cadeira partiu-se a si mesma.»

5.2.5. Estruturas de *se* inerente

Este *se* não tem conteúdo semântico. Há certos verbos que só *se* conjugam na forma pronominal. Designam-se como casos de clítico inerente as formas do pronome reflexo que não estão associadas a qualquer posição argumental ou de adjunto e em que o clítico não pode ser interpretado como uma partícula deansitivizadora. Este clítico não pode coocorrer com as expressões *a si próprio* ou *a si mesmo* (Brito, Duarte, & Matos, 2003). A título ilustrativo, notemos a seguinte frase.

(26) «Licenciou-se em ensino de Português[...].»

Notamos, em (26), uma informação sobre alguém que *se* formou em ensino de um idioma. Reparamos que o clítico é exigido pelo próprio verbo. O *se*, por conseguinte, faz parte integrante do verbo.

Neste caso em que o clítico *se* tem um valor de reflexo inerente, além de revelar uma incompatibilidade de coocorrência em construções de redobro com as expressões *a si próprio* ou *a si mesmo*, não tem conteúdo semântico. Esta impossibilidade de paráfrase com a expressão *a si próprio* e o facto de não receber nenhum papel temático mostra que não estamos perante verdadeiras anáforas reflexas. O clítico *se* com valor inerente não afeta a estrutura argumental do predicador verbal (Brito, Duarte, & Matos, 2003).

5.3. Síntese do capítulo

A compreensão da variedade do Português de Angola é essencial para o trabalho numa perspectiva comparativa com o Português Europeu. Visto que o domínio de uma língua é, por conseguinte, o resultado de práticas efetivas, significativas e contextualizadas, esta reflexão revestiu-se de seis pontos principais – a passiva eventiva, resultativa, estativa e valores de *se* tais como o *se* impessoal, o *se* anafórico, *se* decausativo e *se* inerente.

Explicitamos que a passiva eventiva ou perifrástica é expressa através do verbo *ser* como verbo auxiliar. Tem uma componente eventiva e uma componente agentiva. A estrutura temática desta mantém-se; tem, porém, uma estrutura informacional e sintática diferente. A passiva resultativa é expressa através do verbo *ficar* como verbo auxiliar. Estas são caracterizadas pelo facto de terem uma componente eventiva e ausência de componente agentiva. A passiva estativa é expressa através do verbo *estar* como verbo auxiliar. Estas são

caracterizadas pelo facto de não terem componente agentiva nem componente eventiva. A passiva pronominal ou passiva com o *se* é expressa através do pronome átono de terceira pessoa *se*, sem qualquer verbo auxiliar ou morfologia verbal especial no verbo pleno. Verificamos que, relativamente aos outros valores de *se* em português, há o *se* nominativo que ocorre como sujeito indeterminado com verbos na terceira pessoa do singular. O *se* anafórico ocorre em estruturas reflexas e recíprocas, o qual, dependendo da estrutura temático-argumental do predicador verbal, pode ter a relação gramatical de acusativo ou dativo. O *se* decausativo ocorre com verbos transitivos. A sua função é de transitivizar o predicador verbal. Há, por fim, o *se* inerente que não tem conteúdo semântico.

Capítulo 6 – Trabalhos Relacionados com o Estudo Comparativo

Nesta secção, apresentamos efetivamente os trabalhos desenvolvidos no âmbito do léxico, sintaxe e linguística computacional que têm relação com o nosso objeto de estudo. Procedemos à explanação da dissimilitude entre os estudos desenvolvidos e a nossa perscrutação. Com efeito, objetivamos apresentar sete estudos na área do léxico que se debruçam sobre empréstimos dimanantes de línguas de Angola, cinco estudos na área de sintaxe sobre estruturas passivas e valores de *se* em Português de Angola e Português Europeu. Por fim, debruçamo-nos sobre quatro estudos na área da linguística computacional.

6.1. Estudos comparativos na área do Léxico

Não temos, de facto, gnose de que existam muitas perscrutações nesta área, em particular no que diz respeito às abordagens computacionais. Tem havido uma predileção mais recente por esta área, o que se reflete na emergência de perquisições como Costa (2015), Silva (2015), Sacanene (2019) e Miguel (2019). O primeiro restringe-se ao *umbundu* e o segundo ao português e *kimbundu*, num desvelo de encontrar isomorfismos na área da medicina. Utilizando as etiquetas para etimologia das entradas da versão digital do Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa, o terceiro fez uma congérie de 709 lexemas, que promanam de Angola, nessa base de dados, dentre os quais 335 empréstimos são dimanantes de línguas de Angola. Recorde-se que nem Costa (2015), nem Silva (2015), nem Miguel (2019) utilizaram abordagens computacionais na análise e, precisamente por isso, se abstraem de técnicas de *Text Mining* e Processamento de Linguagem Natural, tais como reconhecimento de entidade nomeada, que foi utilizado por Maurice Gross (1997), mesmo o estudo tendo muitas entidades nomeadas.

Sem pretender ser exaustivo, é preciso não nos olvidarmos de que Teresa Lino e Christina Dechamps (2016) realizaram um estudo comparativo da terminologia jurídica do Português Europeu e do francês. Para tal, utilizaram o *Hyperbase*, programa de tratamento semiautomático de corpora. O estudo de Muhongo (2017) também se debruçou sobre empréstimos dimanantes de sistemas linguísticos angolanos e utilizou o *AntConc* e o *Concapp*. Miguel (2019) realizou um estudo sobre a integração morfológica e fonológica dos empréstimos lexicais provenientes de Angola. Usou uma abordagem sociolinguística e realizou 36 entrevistas a loquentes de diversos níveis de escolaridade em Luanda, cuja transcrição e deteção de empréstimos lexicais foi manual, nas quais identificou 255 EL-EMP.

Dissimétrico destes trabalhos, cingimo-nos a uma abordagem computacional, baseada em técnicas de *Text Mining* e Processamento de Linguagem Natural para a identificação de empréstimos, independentemente do sistema linguístico angolano de que promanam.

6.2. Estudos comparativos na área da Sintaxe

Quanto ao estudo de estruturas predicativas, notamos os estudos de Mendes e Estrela (2008), Ribeiro (2011), Estrela (2013), Muhongo (2019a) e Muhongo (2019b). Mendes e Estrela (2008) estudaram os valores de *se* no Português Europeu e compararam-no com outras variedades africanas do português. Na verificação do estudo, constatamos que o corpus usado não é atualizado e não foram usados métodos computacionais.

Ribeiro (2011) estudou os valores de *se* no Português Europeu em estruturas com *se* anafórico, impessoal e decausativo. Contudo, não usou métodos computacionais para detetar os valores de *se* nem comparou com o Português de Angola. Estrela (2013) verificou como se processa a aquisição de estruturas passivas eventivas, estativas e resultativas no Português Europeu, mais especificamente no que diz respeito à sua compreensão. Tal como o estudo anterior, constatamos que não usou métodos computacionais para detetar estruturas passivas. Além disso, não as comparou com outra variedade do português.

O presente estudo é uma continuidade das reflexões em Muhongo (2019a) e Muhongo (2019b) que, numa perspetiva incipiente, desenvolve um estudo da passiva eventiva e dos valores de *se* no Português de Angola e no Português Europeu, com a finalidade de prospetar algumas dissimilaridades. Com efeito, procedemos à extração de estruturas passivas eventivas, estativas, resultativas e pronominais e a deteção dos valores de *se* numa perspetiva contrastiva, com o auxílio de métodos computacionais. Além disso, explicitamos a estatística descritiva dos dados, calculamos as medidas de desempenho do método automático e usamos teste estatísticos de proporções para a estrutura passiva eventiva e estativa.

6.3. Estudos comparativos em Linguística Computacional

Nesta secção destacam-se as investigações desenvolvidas Canosa, *et al.* (2019), Pinto *et al.* (2016), Gamallo e Garcia (2011) e Iriguti e Feltrim (2019), os quais revelam um liame com o nosso objeto e procedimentos de análise.

Canosa, *et al.* (2019) descrevem um método para construir uma ferramenta destinada a reconhecer entidades geográficas nomeadas em textos medievais. Entretanto, a nova ferramenta foi desenvolvida a partir dos módulos de língua contemporânea do *LinguaKit*, um conjunto de ferramentas de processamento de linguagem natural, com a qual se elaborou o *gazetteers*, uma lista de topónimos medievais. Observaram padrões para a melhora e implementação de novas regras de reconhecimento dos topónimos. Depois da lista de entidades geográficas, os *triggers* – ativadores contextuais – foram o recurso determinante na melhora da *recall*.

Tendo como ponto de partida a avaliação do desempenho de modelos preditivos, Pinto *et al.* (2016) avaliaram uma série de conjuntos de ferramentas de processamento de linguagem natural com sua configuração padrão, enquanto executavam um conjunto de tarefas padrão (por exemplo, tokenização, marcação em POS, *chunking* e NER) em conjuntos de dados populares que cobrem jornais e redes sociais, usando a programação em *Java* e *Python*. Como medidas de desempenho do método automático foram usadas a precisão, a *recall*, medida-F, micro e macro *averaging*. Os resultados obtidos foram úteis para restringir sua escolha. 1998 ACM Subject Classification I.2.7 Natural Language Processing.

Gamallo e Garcia (2011) propõem um sistema de Classificação de Entidades Nomeadas (NEC) baseado em recursos, que combina a extração de entidades nomeadas com heurísticas simples independentes de linguagem. Extraíram-se automaticamente *gazetteers* de entidades nomeadas, fazendo uso de informações semiestruturadas da Wikipédia, como *infoboxes* e árvores de classificação. As heurísticas independentes do sistema linguístico foram usadas para desambiguar e classificar entidades que já foram reconhecidas no texto. Além disso, compararam o desempenho do sistema baseado em recursos com o de um módulo NEC supervisionado implementado para o *FreeLing*. Realizaram-se experiências sobre corpora de texto português tendo em conta vários domínios e géneros.

Partindo da análise de resumo de teses e dissertações, Iriguti e Feltrim (2019) buscaram identificar automaticamente categorias em nível de sentenças que compõem estruturas retóricas de resumo científico. O objetivo foi avaliar o impacto de diferentes conjuntos de atributos na implementação de classificadores retóricos para resumos científicos escritos em português. Avaliam um conjunto de atributos e algoritmos de classificação

aplicados à construção de classificadores retóricos sentenciais para resumos científicos escritos em português. Para isso, foram utilizados atributos superficiais extraídos como valores TF-IDF e selecionados com o teste χ^2 , atributos morfosintáticos implementados pelo classificador AZPort, e atributos extraídos a partir de modelos de word embedding (Word2Vec, Wang2Vec e GloVe). Estes conjuntos de atributos e as suas combinações foram usados para o treinamento de classificadores usando os algoritmos de aprendizagem supervisionada, tais como Support Vector Machines, Naive Bayes, K-Nearest Neighbors, decision Trees e Conditional Random Fields (CRF), cujos classificadores foram avaliados por meio de validação cruzada sobre três corpora compostos por resumos de teses e dissertações. Dentre estes, o melhor resultado, 94% de F1, foi obtido pelo classificador CRF com as seguintes combinações de atributos: (i) Wang2Vec–Skip-gram de dimensões 100 com os atributos provenientes do AZPort; (ii) Wang2Vec–Skip-gram e GloVe de dimensão 300 com os atributos do AZPort; (iii) TF-IDF, AZPort e *embeddings* extraídos com os modelos Wang2Vec–Skip-gram de dimensões 100 e 300 e GloVe de dimensão 300. A partir dos resultados obtidos, conclui-se que os atributos provenientes do classificador AZPort foram fundamentais para o bom desempenho do classificador CRF, enquanto que a combinação com *word embeddings* se mostrou válida para a melhoria dos resultados.

6.4. Síntese do capítulo

Constatamos que há alguns estudos relacionados com o nosso objeto de análise. Assim, procedemos a apresentação dos trabalhos desenvolvidos no âmbito do léxico, sintaxe e linguística computacional que têm relação com o nosso objeto de estudo. Procedemos à explanação da dissimilitude entre os estudos desenvolvidos e a nossa perscrutação. Com efeito, verificamos a existência de sete estudos na área do léxico que se debruçam sobre empréstimos dimanantes de línguas de Angola, cinco estudos na área de sintaxe sobre estruturas passivas eventivas, estativas, resultativas e pronominais e valores de *se* em Português de Angola e Português Europeu. Por fim, debruçamo-nos sobre os quatro estudos na área de linguística computacional.

Parte II

Enquadramento na Área de Linguística Computacional

Capítulo 7 – Técnicas de *Text Mining* e Processamento de Linguagem Natural

Este capítulo tem quatro objetivos. Em um primeiro momento, pretendemos explicitar uma conceptualização de *Text Mining* e Processamento de Linguagem Natural. Em um segundo momento, apresentamos as estruturas de dados, isto é, dados não estruturados, dados estruturados e semiestruturados. Além disso, descrevemos os procedimentos para o seu processamento. Em seguida, debruçamo-nos sobre técnicas de *Text Mining* e Processamento de Linguagem Natural. Assim, dividimos as técnicas em básicas e avançadas, as quais servirão de estrutura basilar para os capítulos subsequentes. Por fim, descrevemos a matriz da confusão, as medidas de desempenho de método automático, ou seja, precisão, *recall* e medida-F. Procedemos também à apresentação de testes estatísticos. Debruçamo-nos sobre estes aspetos, pois aplicamos, nos capítulos subsequentes, estes conceitos.

7.1. *Text Mining* e Processamento de Linguagem Natural

Se compreendermos o texto como uma coleção de documentos não estruturados, e sendo usada para analisar textos e transformá-los em uma forma mais estruturada a fim de obter *insights* a partir dela, notaremos que *Text Mining* é, como advogam Cielen, Meysman e Ali (2016), uma área da Inteligência Artificial que conglobera Ciências da Linguagem e Ciências da Computação com técnicas estatísticas e *Machine Learning*, a qual, sendo outra área da Inteligência Artificial e intimamente relacionada com a Matemática Aplicada e a Estatística, pode ser percebida como processo pelo qual um computador pode trabalhar com mais precisão à medida que recolhe e aprende com os dados que lhe são inseridos. Por Processamento de Linguagem Natural pode entender-se a parte das Ciências da Computação e da Inteligência Artificial que lida com a linguagem humana.

Estas áreas do conhecimento, *Text Mining* e Processamento de Linguagem Natural, têm sido muito usadas para várias aplicações como identificação de entidades, detecção de plágio, *topic identification*, *text clustering*, tradução, *automatic text summarization*, detecção de fraude, *spam filtering* e *sentiment analysis*. Para ilustrar, a Google usa-as para muito mais do que responder a consultas. Além de proteger os utilizadores do *Gmail* contra *spam*, também divide os e-mails em diferentes categorias, tais como social, atualizações e fóruns (Cielen, Meysman, & Ali, 2016; Weiss, Indurkha, & Zhang, 2015).

A comunidade de *Text Mining* e Processamento de Linguagem Natural tem tido sucesso no reconhecimento de entidades nomeadas, reconhecimento de tópicos, resumo, preenchimento de textos e análise de sentimento, mas modelos treinados num domínio não generalizam bem para outros domínios. Parece que mesmo as técnicas mais avançadas não são capazes de decifrar o significado de cada parte do texto. Isso é visível inclusive em humanos, pois eles também lidam com a linguagem natural que parece, naturalmente, ambígua. Até o conceito de significado também pode parecê-lo. Por exemplo, constatamos que, quando duas pessoas ouvem a mesma conversa, a conversa pode não ter o mesmo significado, portanto o significado das palavras pode variar, quando vêm de alguém irritado ou alegre, aspecto que é muito estudado em Pragmática e Análise do Discurso (Cielen, Meysman , & Ali, 2016, p. 6).

7.1.1. Tipos de documentos textuais

Os dados, i.e., documentos com texto, podem ter formas diferentes. Partindo do tripartite arquétipo de Oliveira, Guerra e McDonnell (2018), partimos do princípio de que os dados podem ser classificados em dados não estruturados, estruturados e semiestruturados. Por dados *não estruturados* pode entender-se os dados que não têm efetivamente uma estrutura previsível, cada conjunto de informações possui, assim, uma forma única. Estes dados geralmente são arquivos com um imenso teor textual, organizados para cada conjunto de informações, dentre os quais estão e-mail, *twitter*, *pdf*, imagens, vídeos, etc. Consideramos que os nossos dados são deste tipo. Portanto, a linguagem natural é, como asseveram Cielen, Meysman e Ali (2016, pp. 5-6), um tipo especial de dados não estruturados, cujo processamento requer conhecimento de técnicas específicas de Ciência de Dados e Linguística.

Entende-se por dados *estruturados* o conjunto de informações organizadas em forma tabelar. São dados comumente encontrados diretamente em bancos de dados, arquivos com algum tipo de organização/separação entre as linhas e colunas, como por exemplo, no formato de *Excel*. Consideramos que os dados estruturados são precisos, pois muitos métodos computacionais requerem esse tipo de dados, como podemos observar em matriz documento-termo no próximo capítulo.

Os dados *semiestruturados* são dados que também possuem uma organização fixa, todavia não seguem o padrão de estrutura linha ou coluna. Seguem umas estruturas mais complexas e flexíveis, geralmente hierárquicas, estruturadas em *tags* ou marcadores de campos. São exemplos de arquivos semiestruturados: JSON, XML, HTML, YAML, etc. são o

formato mais usado em troca de dados pela internet e consumo de *Application Programming Interface*. Reconhecemos de facto que não existe somente uma forma de estruturar e aplicar os conhecimentos de Ciências de Dados, que concilia as áreas das Ciências da Computação e Estatística. A forma de aplicação pode variar conforme a necessidade do projeto ou do objetivo que se pretende alcançar (Oliveira, Guerra, & McDonnell, 2018). Os dados semiestruturados, em geral, são facilmente transformados em dados estruturados, para se proceder ao processamento com recurso a técnicas de *Text Mining* que aquiesce ao arquétipo que pode ser visto na Fig. 7.1. Nesta investigação, usamos o modelo de procedimentos em Ciência de Dados apresentado por Wickham e Grolemond (2017).

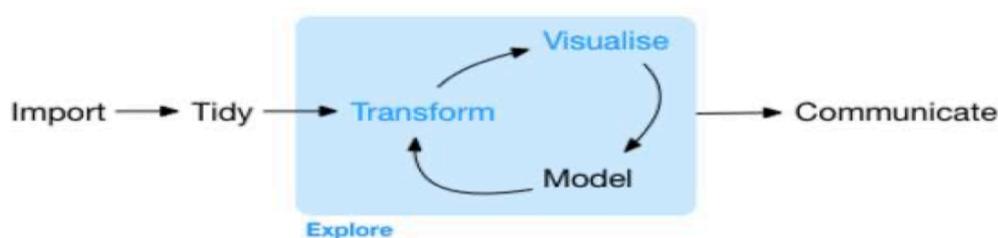


Fig. 7.1 – Procedimentos para Ciência de Dados, Wickham & Grolemond (2017)

7.2. Técnicas básicas de *Text Mining*

Na análise linguística de um texto digital em linguagem natural, é necessário que em qualquer documento sejam claramente definidos os caracteres, as palavras e as frases. Definir estas unidades apresenta desafios diferentes dependendo da língua processada e da fonte dos documentos. Destarte, o cruzamento da Lexicologia e Linguística Computacional torna-se, efetivamente, de grande relevância no processamento de linguagem natural (Amaro & Mendes, 2016).

Recorde-se, com efeito, que Palmer (2010) observa que o pré-processamento de texto se cinde em triagem documental e segmentação textual. Em geral, o primeiro corresponde ao processo de conversão de um conjunto de documentos digitais em documentos de texto bem definido, como por exemplo uma matriz documento termo. Para os primeiros corpora, este era evidentemente um processo prolongado, manual e raramente eram mais do que alguns milhões de palavras. Nesta secção, descrevemos a *tokenization*, lematização, *stemming* e *lower case*, que são usadas nos capítulos subsequentes.

7.2.1. *Tokenization*

A noção de *token* pode remeter-nos para a Teoria do Signo (Villalva & Silvestre, 2014, p. 78). Com a crescente importância da análise computacional de texto, requer-se um processo de triagem de documentos totalmente automatizado (Coniam, 1993). A *tokenization*, podendo ser o primeiro passo na manipulação de texto para fins de Linguística Computacional, pode ser entendida como o processo de dividir um texto em *tokens*. Isso é crucial para a análise computacional de texto, pois os textos completos são muito específicos para realizar quaisquer análises mais aprofundadas (Palmer, 2010; Welbers, Atteveldt, & Benoit, 2017). Este processo é conhecido pelas linguagens artificiais, como as linguagens de programação, de maneira que a conversão permita que se elidam ambiguidades lexicais e estruturais.

Os *tokens*, geralmente, são formas lexicais, porquanto componentes semanticamente mais comuns dos textos. Destarte, o espaço de caracteres, a tabulação e nova linha que assumimos são delimitadores e não são contados como *tokens*, pois são denominados coletivamente de espaço em branco. Os caracteres “<>!?” são, normalmente, delimitadores e podem ser *tokens* (Weiss, Indurkha, & Zhang, 2015). Esses espécimes salientam a importância da normalização do texto durante o processo de *tokenization*, de modo que datas, horas, expressões monetárias e todas as outras frases numéricas possam converter-se em um formulário consistente com o processamento exigido pelo pacote de Processamento de Linguagem Natural que se estiver a usar.

Além do pacote *tokenizers*, o R tem o *stringr*, o qual é frequentemente usado para a desambiguação de frase e palavra. Para isso, utiliza dicionários da biblioteca *International Components for Unicode (ICU)* (Welbers, Atteveldt, & Benoit, 2017). Partindo do princípio de que a escrita de expressões numéricas depende efetivamente de cada idioma, a sua *tokenization* requer uma gnose da sintaxe de tais expressões. Se isso implica conhecimentos específicos de técnicas de Ciências da Computação e Linguística, a segmentação de *tokens*, sendo a unidade básica de informação do modelo que se usará, permite verificar a sequência de caracteres de um texto.

7.2.2. Representação de textos

Nesta subsecção, debruçamo-nos sobre conceitos fundamentais para a área de Processamento de Linguagem Natural: TF, TF-IDF e *Bag of Words*. Recorde-se que um documento de texto pode ser representado como:

- o vetor de *tokens*,¹⁶
- *bag-of-words*,

Na representação de *bag-of-words*, o documento é representado em forma de um *bag* (*multiset*) que ignora a ordem das palavras (*tokens*). Em R, podemos notar que o *bag* pode ser representado, usando um vetor de *tokens*, que é normalmente acompanhado com o seu número de ocorrências (frequência) (Cielen, Meysman, & Ali, 2016).

Em relação à representação de vários documentos, esses podem ser representados em forma de uma lista, em que cada elemento representa um documento. No entanto, uma representação habitualmente usada é uma *matriz* ou *data frame*, em que cada documento é representado numa linha. As colunas incluem, como já foi adiantado, informação sobre cada documento. As células da matriz (ou *data frame*) relativas a token t (i.e., a coluna relativa a token t) representam informação sobre as suas ocorrências no respetivo documento. Essa informação pode indicar

- existência de termo usando representação binária (0/1);
- frequência de termos em certo documento (TF);
- representação TF-IDF.

A representação binária pode ser utilizada para a representação de documentos. Uma consulta (documento) é representada da mesma forma que outros documentos. A maioria dos documentos semelhantes são identificados utilizando alguma medida de proximidade, como a distância Euclidiana.

Convém notar que a representação TF-IDF se baseia nos seguintes pressupostos.

- O termo que aparece frequentemente no documento pode ser mais importante para a identificação do que o termo que aparece raramente. Este aspeto é indicado pelo TF, frequência de ocorrência do termo num documento.
- Se um termo aparecer em muitos documentos, será provavelmente irrelevante. Este aspeto é indicado pelo IDF, a frequência inversa do documento. Podemos

¹⁶ O vetor compreende uma sequência de valores numéricos ou de caracteres (letras, palavras). A *matriz* compreende uma coleção de vetores em linhas e colunas; todos os vetores têm de ser do mesmo tipo (numérico ou caracteres). Entende-se por *data frame* o mesmo que *matriz*, mas aceita vetores de tipos diferentes (numéricos e caracteres) em colunas diferentes.

calcular IDF usando a seguinte equação

$$IDF_t = \log_2 \left(\frac{N}{DF_t} \right)$$

Sendo:

- N o número de documentos no corpus;
- t um token do corpus;
- $DF(t)$ o número de documentos em que o token aparece.

Com efeito, o valor de TF-IDF do termo t_j no documento d_i pode ser calculado usando a seguinte equação

$$W_{tf-idf}(t_j, d_i) = TF(t_j, d_i) * IDF(t_j)$$

onde TF é, como já foi referido, a frequência de termo em certo documento e IDF a frequência inversa do termo nos documentos.

7.2.3. Lematização e *stemming*

Partindo de procedimentos da morfologia, cujas tarefas consistem em analisar os constituintes morfológicos das palavras e o modo como estes se organizam entre si, analisar as suas propriedades e compreender a organização que pode existir entre os constituintes para a estruturação de uma palavra ou para a variação da mesma palavra (Rodrigues, 2013), a lematização cumpre um propósito símil ao do *stemming*, mas não corta as extremidades das formas para normalizá-las.

Compreendemos que a importância da normalização consiste precisamente no facto de uma palavra base poder ter variações morfológicas diferentes, tais como os morfemas de tempo-modo-aspeto ou morfema de número-pessoa. Convém notar que uma técnica é evidentemente o *stemming*, que pode ser considerado essencialmente como um algoritmo baseado em regras que converte as palavras na sua forma base ou radical. A sua utilização requer também uso do pacote *SnowballC* e a instrução *stemDocument*. Na Tabela 7.1, podemos notar que a lematização substitui, efetivamente, as flexões ou formas variantes pela sua forma base ou lema, permitindo verificar os paradigmas do lexema (Welbers, Atteveldt, & Benoit, 2017; Haspelmath, 2002).

Nº	Palavras	Lematização	Stemming
1	carros	carro	carr
2	carrão		
3	carrinha		
4	cantasse	cantar	cant
5	cante		
6	cantaria		

Tabela 7.1 – Relação entre lematização e *stemming*

Se compreendermos o morfema como uma unidade mínima com função semântico-funcional, conquanto nem todos os morfemas tenham carga semântica, parece-nos crível depreender que uma característica diferencial das unidades morfológicas das unidades da sintaxe consiste no facto de as primeiras denotarem um carácter fixo em que ocorre a ordem dos constituintes léxicos, tornando seguramente perceptível casos de alomorfia tanto na raiz como nos afixos (Luschützky, 2000; Plag, 2003). Esta informação de lematização é dada pela coluna *lemma* na Fig. 7.2, que veremos mais adiante.

7.2.4. Normalização de texto: conversão para letra minúscula

O processo de normalização refere-se amplamente à transformação das palavras numa forma mais uniforme. Para converter as unidades lexicais para letra minúscula, podemos usar a instrução *tolower*. Em virtude da sensibilidade dos sistemas computacionais, isto pode ser importante se, para uma determinada análise, um computador tem de reconhecer quando dois lexemas têm aproximadamente o mesmo significado, mesmo que sejam de forma ligeiramente diferente.

Compreendemos que a importância da normalização consiste precisamente no facto de uma palavra base poder ter variações morfológicas diferentes, tais como os morfemas de tempo-modo-aspeto ou morfema de número-pessoa. Convém notar que este procedimento é fundamental para o nosso estudo comparativo, devido à sensibilidade do R à maiúscula.

7.3. Técnicas avançadas de *Text Mining*

Nesta secção, descrevemos algumas técnicas avançadas de *Text Mining*, fundamentais nos capítulos subsequentes. Com efeito, explicitamos os conceitos de *part-of-speech tagging*, *named entity recognition* e *dependency parsing*.

7.3.1. *Part-of-speech tagging*

Permitindo verificar a simbiose entre a Lexicologia e a Linguística Computacional, áreas fundamentais do Processamento de Linguagem Natural, compreenda-se *part-of-speech tagging* como identificação de categoria lexical/sintática tais como verbo, nome, adjetivo, preposição, advérbio, etc. Havendo, no português, unidades léxicas isocategoriais e heterocategoriais, a etiquetagem é precisamente expressa pela divisão das palavras em classes abertas (verbo, nome, adjetivo, etc.), fechadas (determinantes, pronomes, conjunções, etc.) e outras (Wijffels, 2019).

Para a concretização desta subsecção, usamos o pacote *udpipe*, que é orientado para Processamento de Linguagem Natural. Na Tabela 7.2, podemos observar a sistematização para classes lexicais em *udpipe* (Wijffels, 2019). Podemos depreender que este processo conglomerava categorias morfossintáticas como género, número, tempo-modo-aspeto e pessoa-número (Villalva, 2000; Borer, 2005).

Nº	Categoria	Abreviação	Significação
1	Lexical	VERB	Verbo
2		AUX	Verbo auxiliar
3		NOUN	Nome
4		PROPN	Nome próprio
5		ADJ	Adjetivo
6		ADV	Advérbio
7	Funcional	INTJ	Interjeição
8		ADP	Preposição
9		CCONJ	Conjunção coordenativa
10		DET	Determinante
11		NUM	Numeral
12		PART	Partícula
13		PRON	Pronome
14		SCONJ	Conjunção subordinativa
15		PUNCT	Pontuação
16		SYM	Símbolo
17		X	Outras

Tabela 7.2 – Sistematização para classes lexicais em *udpipe*

Podemos, sob este prisma, considerar que os critérios de identificação de palavras ajudam a perceber que, apesar de ser perceptível, a partir do contexto, que a morfologia tem uma grande interação com a sintaxe, principalmente a morfologia flexional, evoque-se que a morfologia é um domínio autónomo em relação à sintaxe, pois a determinação da categoria surge no léxico (Jackendoff, 2002). Depreendemos que *part-of-speech tagging* pode ser, por conseguinte, uma forma de filtrar classes sintáticas das estruturas léxicas tendo como base as informações sobre o modo como a palavra é usada na sintaxe (Welbers, Atteveldt, & Benoit, 2017; Rodrigues, 2013). A Fig. 7.2 ilustra uma representação da utilização do pacote *udpipe* para *part-of-speech tagging* de uma estrutura ativa, onde a classe lexical da palavra é indicada na coluna *upos*.

```
> sentence <- c("o meu pai construiu uma casa em tempo recorde.")
> install.packages("udpipe")
> library(udpipe)
> udmodel <- udpipe_download_model(language = "portuguese")
> udmodel <- udpipe_load_model(file = udmodel$file_model)
> sentence.annot <- udpipe_annotate(udmodel, x = sentence)
> sentence.annot <- as.data.frame(frase.annot, detailed = TRUE)
> sentence.annot[,c("token_id", "token", "lemma", "feats", "head_token_id",
"upos", "dep_rel")]
```

	token_id	token	lemma	feats	head_token_id	upos	dep_rel
1	1	o	o	Definite=Def Gender=Masc Number=Sing PronType=Art	3	DET	det
2	2	meu	meu	Gender=Masc Number=Sing PronType=Prs	3	DET	det
3	3	pai	pai	Gender=Masc Number=Sing	4	NOUN	nsubj
4	4	construiu	construir	Mood=Ind Number=Sing Person=3 Tense=Past VerbForm=Fin	0	VERB	root
5	5	uma	um	Definite=Ind Gender=Fem Number=Sing PronType=Art	6	DET	det
6	6	casa	casa	Gender=Fem Number=Sing	4	NOUN	obj
7	7	em	em	<NA>	8	ADP	case
8	8	tempo	tempo	Gender=Masc Number=Sing	6	NOUN	nmod
9	9	recorde	recorde	Gender=Masc Number=Sing	8	ADJ	amod
10	10	.	.	<NA>	4	PUNCT	punct

Fig. 7.2 – *Part-of-speech tagging* de estrutura ativa

7.3.2. Identificação de entidades nomeadas

O reconhecimento de entidades nomeadas (NER) foi definido como subárea da extração de informação em 1995 (Konkol, 2015, p. 12). Assim, por reconhecimento de entidades nomeadas pode entender-se uma técnica para identificar uma palavra ou sequência de palavras que representa uma entidade e que tipo de entidade, como uma cidade, região, data, pessoa, organização, etc. Geralmente, este processo é emparelhado com a resolução de correferência.

Konkol (2015, p. 24) descreve a abordagem de *Machine Learning* para deteção de NER que é hodiernamente a abordagem mais utilizada pelos investigadores. Este esquema,

ilustrado na Fig. 7.3, apresenta duas fases: treinamento e teste. A entrada da fase de treinamento é simplesmente *text* e *labels*. Esta entrada pode ser processada de várias formas, normalmente com *tokenization*, *part-of-speech tagging*, *stemming* ou lematização. O passo seguinte é a transformação da entrada para um formato compreensível para a máquina. O *token* (\approx palavra) é tomado como uma unidade básica para o processamento seguinte, que envolve a extração de certas características (*features*) que são usadas na construção de classificador com o auxílio de técnicas de *Machine Learning*.

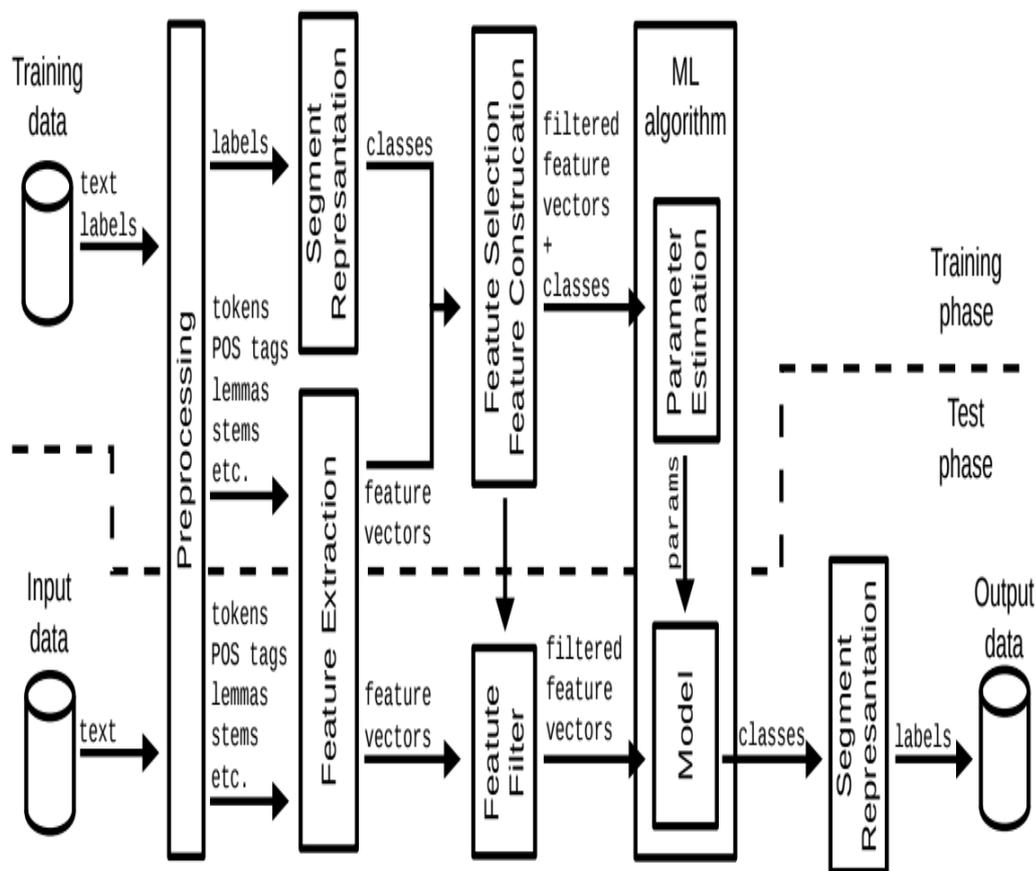


Fig. 7.3 – Esquema de abordagem de *Machine Learning* para NER, Konkol (2015)

Reconhecemos que o programa PAMPOPlus detecta entidades nomeadas, indicando as categorias (pessoa, local, organização, miscelânea, data, etc) (Rocha, et al., 2016). Pode-se detetar também as entidades nomeadas em outras linguagens, como auxílio a outros pacotes como, por exemplo, *spaCy* em *Python*. Na Fig. 7.4, podemos observar a deteção de entidades nomeadas com o pacote *spaCy* em *Python*, usando a instrução

```

In [25]:!pip install spacy
In [26]:!python -m spacy download pt
In [27]:import spacy
In [28]:nlp = spacy.load("pt")
In [29]:doc = nlp('Aqui se conta da chegada de Jerónimo Caninguili, moço
benguelense, à velha cidade de São Paulo da Assunção de Luanda.[...]' )
In [30]:print([(token.text, token.label_) for token in doc.ents])
[('Jerónimo Caninguili', 'PER'), ('São Paulo da Assunção de Luanda',
'LOC'), ('Caninguili', 'PER'), ('Loja de Barbeiro', 'PER'), ('Pomadas',
'LOC'), ('Fraternidade', 'PER'), ('Alice', 'PER'), ('Malange', 'PER'),
('Humbes', 'PER'), ('Conferência de Berlim', 'MISC'), ('Arantes Braga',
'PER'), ('dezassete de Novembro', 'LOC'), ('mucama Josephine', 'PER'),
('Luanda', 'LOC')]

```

Aqui se conta da chegada de **Jerónimo Caninguili PER**, moço benguelense, à velha cidade de **São Paulo da Assunção de Luanda LOC**. E de como, enquanto **Caninguili PER** dava os seus últimos retoques à sua **Loja de Barbeiro PER** e **Pomadas LOC**, dita ainda **Fraternidade PER**, a menina **Alice PER** soltava os pássaros do falecido pai. Conta-se também da confusa rixa que pelos finais de 1881 teve por pretexto as eleições para a câmara municipal e dos sucessos que levaram um rico agricultor de **Malange PER** a mandar assar uma escrava para a servir aos cães. Pelo meio fica a primeira revolta dos **Humbes PER** e o início da **Conferência de Berlim MISC**. Finalmente, dá-se conta do passamento de **Arantes Braga PER**, jornalista de incendiado verbo, socialista e independentista, o qual abandona esta estória no

Fig. 7.4 – Extração de entidades nomeadas em *A Conjura*

Constatamos que, na Fig. 7.4, algumas entidades nomeadas estão bem classificadas, como por exemplo *Jerónimo Caninguili*, *São Paulo da Assunção de Luanda*, *Humbes*, *Conferência de Berlim*, etc. Podemos observar que a unidade lexical *Malange* foi classificada como pessoa, mas é, na realidade, local. Problema de classificação também ocorre com *Loja de Barbeiro*. Notamos que isso ocorre em vários sistemas computacionais.

Neste trabalho, seguimos outra abordagem, pois recorremos ao *udpipe*, que reconhece entidades nomeadas. Estas são identificadas com a etiqueta (tag) *PROPN*. Vimos que o sistema não distingue as diferentes categorias (pessoa, local, organização, etc.).

7.3.3. *Dependency parsing*

Considerando que a morfologia, como frisa Rodrigues (2013), se dedica à análise da estrutura interna das palavras, pelo que não cabe à morfologia estudar a organização das palavras na frase, assim como não cabe à sintaxe estudar a organização interna das palavras, *dependency parsing* remete-nos precisamente para o domínio da sintaxe.

A análise de dependência permite-nos lobrigar a estrutura e relação dos constituintes e as relações sintáticas entre os *tokens*, que podem ser usadas para analisar textos no nível

de classes sintáticas, sendo possível notar a dependência entre a estrutura de constituintes. Esta informação é dada pelas colunas *token_id*, *head_token_id* e *dep_rel* na Fig. 7.2, que indica a relação entre um par de *tokens*.

Considere-se, por exemplo, os *tokens* nº 3 e 4, na Fig. 7.2. O *token* nº 4 foi classificado como verbo (*verb* na coluna *upos*). A etiqueta desse *token* na coluna *dep_rel* indica que esta palavra é considerada como a *raiz* (núcleo) e, conseqüentemente, não depende de nenhum outro *token*. Isso é representado pelo valor 0 na coluna *head_token_id*. Vamos agora analisar o *token* nº 3, i.e. o *token* *pai*, que é classificado como um nome (*noun* na coluna *upos*). Reparamos que esse *token* tem a ligação à *raiz*, pois o *head_token_id* é igual a 4, o número desse *token*. O *pai* tem a relação gramatical de sujeito, como se pode ver na coluna *dep_rel* que inclui o valor *nsubj*.

A mesma informação pode ser encontrada na Fig. 7.5, que mostra a informação da Fig. 7.2 em forma de gráfico. Esta figura mostra a sequência de *tokens*, tal como ocorreram na frase (ex. *pai* e *construiu* nas posições 3 e 4, classe lexical (ex. *noun* e *verb*) e a ligação entre tokens (ex. *nsubj*). Este gráfico pode ser obtido com as seguintes instruções:

```
> library(igraph)
> library(gggraph)
> library(ggplot2)
> plot_annotation(sentence.annot, size = 4)
```

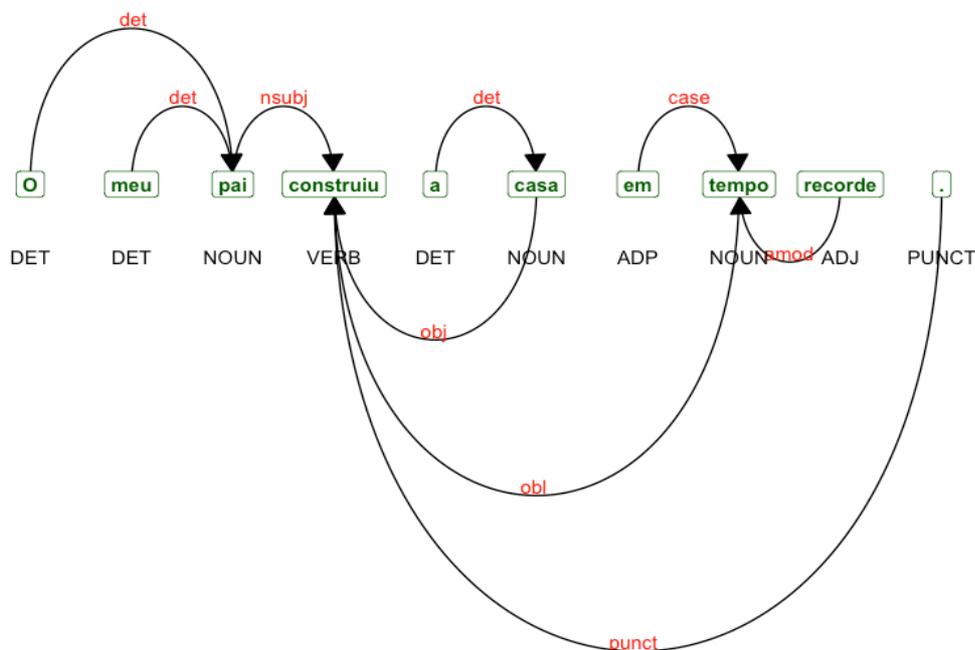


Fig. 7.5 – Um exemplo de relações de dependência sintática

Reconhecemos que se pode fazer *dependency parsing* em outras linguagens, com o auxílio a outros pacotes como, por exemplo, *spaCy* em *Python*. Entretanto, optamos pelo de *udpipe* em *R*. Na Fig. 7.4, podemos observar a detecção de relações de dependência com o pacote *spaCy* em *Python*. Constatamos que, quer o *udpipe*, conforme notamos na Fig. 7.5, quer o *spaCy*, como se verifica na Fig. 7.6, que apresenta a mesma frase, partem do princípio de que o verbo é o núcleo da estrutura sintática.

```
In [31]: doc = nlp("O meu pai construiu uma casa em tempo recorde.")
In [32]: from spacy import displacy
In [33]: displacy.serve(doc, style="dep")
```

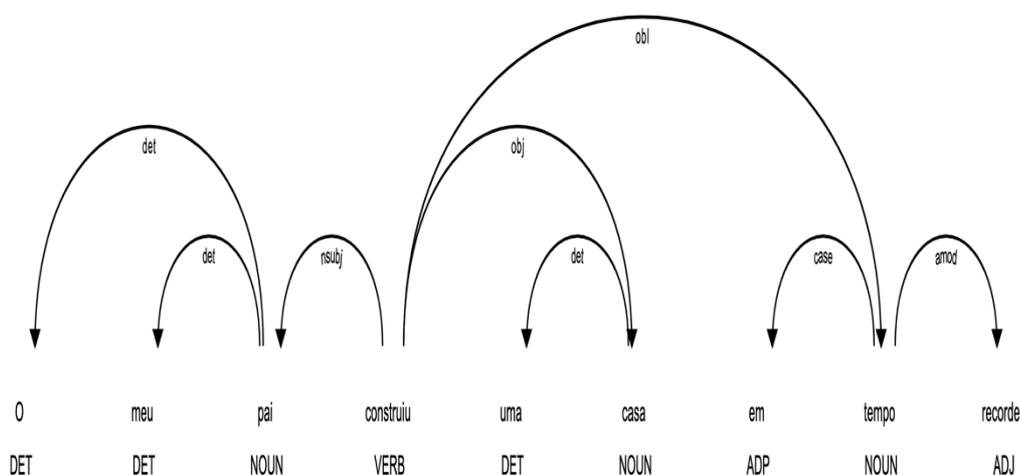


Fig. 7.6 – Exemplo de relações de dependência sintática com *spaCy*

7.3.4. Expressões regulares

Trata-se de um assunto bastante complexo e avançado. *Regular expression* ou *Regex* refere-se a símbolos especiais utilizados em funções textuais para reconhecimento de padrão (Oliveira, Guerra, & McDonnell, 2018; Wickham & Grolemond, 2017). Na Tabela 7.3, apresentamos uma lista dos principais mecanismos de *regex*.

Nº	Regex	Correspondência
1	^	começo de <i>string</i> (ou uma negação)
2	.	qualquer caracter
3	\$	fim da linha
4	[abc]	procura os caracteres a, b, c
5	[^abc]	procura qualquer caracter exceto a, b, c
6	[0-9]	números
7	\\d	qualquer número
8	[A-Z]	qualquer letra maiúscula
9	\\b[A-z]	qualquer letra maiúscula ou minúscula
10	\\w	uma palavra
11	\\b{1}	palavra com uma letra
12	\\W	não é palavra (pontuação, espaço, etc.)
13	\\s	um espaço (<i>tab</i> , <i>newline</i> , <i>space</i>)
14	\\s+	espaço extra

Tabela 7.3 – Principais mecanismos de *regular expression*

7.4. Medidas de desempenho do método automático

A utilização de um modelo para realizar predições de valores pode, naturalmente, implicar erros em relação aos valores reais. Consequentemente, a avaliação do método de análise preditiva requer uma definição de medidas de erro ou desempenho do método automático que, geralmente, variam consoante o problema seja de classificação ou de regressão. Podemos tomar como modelo o caso da classificação, como poderemos notar no capítulo 11, *Estruturas passivas em Português de Angola e Europeu*.

7.4.1. Medidas para problemas de classificação

Podemos verificar as medidas de desempenho do método automático com base na matriz de confusão de problema de duas classes. Convém recordar que uma classe é denotada positiva e outra é denominada negativa, conforme observamos na Tabela 7.4, em que:

- *VP* – *verdadeiro positivo*, que representa o número de casos cujo valor previsto era *positivo* e o valor real foi também *positivo*;
- *FP* – *falsos positivos*, que representa o número de casos cujo valor previsto era *positivo* e o valor real era *negativo*;
- *FN* – *falsos negativos*, que representa o número de casos cujo valor previsto era

negativo e o valor real era *positivo*;

- *VN* – *verdadeiro negativo*, que representa o número de casos cujo valor previsto era *negativo* e o valor real era *negativo*.

		Classe preditiva	
		+	-
Classe verdadeira	+	VP	FN
	-	FP	VN

Tabela 7.4 – Matriz de confusão, adaptado de Gama, Faceli, Lorena e Oliveira (2017)

Partindo da matriz apresentada na Tabela 7.4, podemos verificar as métricas definidas para a extração de informação, a fim de medir o desempenho de um método computacional. As métricas estabelecidas podem ser precisão, *recall* e a medida-F (Gama, Faceli, Lorena, & Oliveira, 2017; Pinto, Alves, & Oliveira, 2016):

A *precisão* é definida como proporção de instâncias corretamente classificadas (verdadeiros positivos) entre todas as instâncias classificadas sob uma determinada categoria (verdadeiros positivos e falsos positivos), conforme podemos observar na equação 1.

$$(1) P_i = \frac{VP_i}{VP_i + FP_i}$$

A medida *Recall* – é definida como a correspondência da taxa de acerto na classe positiva, ou seja, a razão das respostas corretas (verdadeiros positivos) e as respostas corretas totais possíveis (verdadeiros positivos e falsos negativos), como observamos na equação 2.

$$(2) R_i = \frac{VP_i}{VP_i + FN_i}$$

F-measure – corresponde à média harmônica da precisão e *recall*, como na equação 3.

$$(3) F_1 = \frac{2 \times P_i \times R_i}{P_i + R_i}$$

A delimitação entre os objetos positivos e negativos, como advoga Konkol (2015, p. 7), pode ser vista na Fig. 7.7 onde as curvas mostram a distribuição de objetos positivos e negativos e a linha pontilhada mostra o limiar da decisão do classificador. Nas áreas indicadas como FN e FP estão alguns objetos marcados incorretamente.

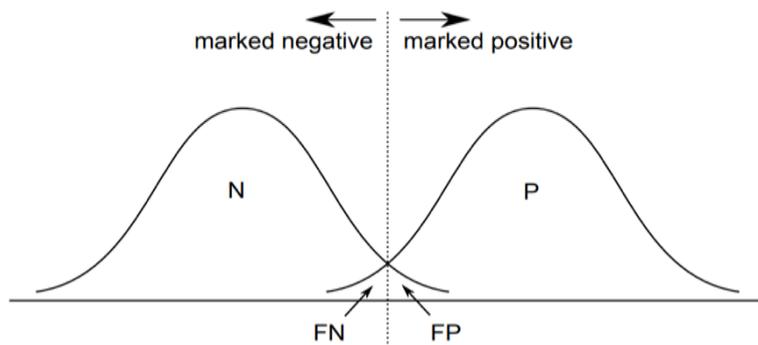


Fig. 7.7 – Marcação de verdadeiro e falso positivo, Konkol (2015)

7.4.2. Medidas para problemas de regressão

Relativamente à regressão, as medidas de desempenho baseiam-se em cálculos da diferença numérica entre os valores previstos e os valores reais, podendo ser utilizadas várias medidas, tais como soma do quadrado dos erros (SQE), raiz quadrada da média do quadrado dos erros (RMQE) e média dos desvios absolutos (MDA).

7.4.3. Testes estatísticos

Partindo desta perspetiva, consideramos que os testes estatísticos podem ser usados para verificarmos uma hipótese relacionada com a distribuição dos dados de uma amostra. Neste caso, gera-se uma hipótese nula (H_0) sobre a distribuição e calcula-se uma estatística de teste que permite aferir a probabilidade de rejeitar esta hipótese, cujo valor principal é o *p – value*.

Os testes estatísticos podem ser paramétricos e não paramétricos. Os testes paramétricos fazem suposições sobre a distribuição dos dados, assumindo a distribuição normal de ambos. Os testes não paramétricos não impõem qualquer distribuição nos dados, não assumindo o cálculo de qualquer parâmetro (Rocha & Ferreira, 2017). Além disso, os testes não paramétricos, apresentados na Tabela 7.5, podem ser úteis tanto para teste de hipóteses e análise de amostras grandes que não podem ser classificadas como dados paramétricos como para análise de amostras pequenas, não têm normalidade nos seus dados.

Nível de mensuração	TESTES ESTATÍSTICOS NÃO-PARAMÉTRICOS					Medidas de correlação não-paramétricas
	Caso de uma amostra	Caso de duas amostras		Caso de k amostras		
		Amostras relacionadas	Amostras independentes	Amostras relacionadas	Amostras independentes	
Nominal	Binomial e X ²	McNemar	Fisher e X ²	Q de Cochran	X ²	De contingência
Ordinal	Kolmogorov-Smirnov Iterações	Sinais Wilcoxon	Mediana U de Mann-Whitney Kolmogorov-Smirnov Iterações de Wald-Wolfowitz Moses	Friedman	Extensão da mediana Kruskal-Wallis	Por postos de Spearman Por postos de Kendall Parcial de postos de Kendall Concordância de Kendall
Intervalar		Walsh Aleatoriedade	Aleatoriedade			

Tabela 7.5 – Testes estatísticos não paramétricos, Teixeira, Fuccio e Oliveira (2020)

Teste estatístico McNemar

Nesta investigação, usaremos o teste estatístico McNemar, que se baseia na estatística de Q^2 , cuja equação é $Q_{obs}^2 = \frac{(a-d)^2}{a+d}$. É um teste de mensuração ou escala nominal¹⁷ e pode ser feito com duas amostras correlacionadas (Portal Action, 2020). Sendo um teste de escala nominal, as categorias diferenciam-se apenas pelo nome, é menos restritiva onde a equivalência é caracterizada por propriedades como a reflexibilidade – cada unidade de uma classe é igual a ela própria; a simetria – para cada duas unidades em uma mesma classe, sejam A e B, A = B implica B = A; e a transitividade – para quaisquer três unidades em uma classe, sejam A, B e C, A = B e B = C implica A = C (Piana, Machado, & Selau, 2009).

1. Cada indivíduo das amostras possui duas respostas, uma antes e outra depois de certo tratamento ser realizado ou cada indivíduo recebe um par de outro grupo que seja parecido com ele em relação às variáveis de interesse.
2. É analisada a possível mudança de resposta no teste McNemar, verificando se ela é significativa e deve ser considerada ou não.
3. As hipóteses do teste McNemar configuram-se como:
 - 3.1. H_0 = hipótese nula, quando não há diferença nas amostras em relação ao atributo observado. Portanto, a mudança de resposta é mínima.
 - 3.2. H_1 = Hipótese de que houve diferença em relação ao atributo observado nas amostras, portanto a mudança de resposta das amostras é significativa.

¹⁷ Recorde-se que as escalas de medida podem ser classificadas em quatro categorias: escala nominal, escala ordinal, escala intervalar e escala de razão ou racional (Piana, Machado, & Selau, 2009).

Neste trabalho, aplicamos o teste McNemar ao estudo de estruturas passivas eventivas, estativas e resultativas no Português de Angola e Português Europeu. Assim, estabelecemos as seguintes hipóteses:

H_0 : Considera-se que não existe diferença no uso das estruturas passivas no Português de Angola e Português Europeu.

H_1 : Considera-se que existe diferença no uso das estruturas passivas no Português de Angola e Português Europeu.

Para isso, engendramos um classificador. Os dados obtidos da criação do classificador é que serão usados para a aplicação do teste McNemar, o qual nos remete para probabilidade condicional e cuja discussão é feita no capítulo 11 desta tese.

7.5. Síntese do capítulo

Neste capítulo, explicitamos algumas técnicas básicas e avançadas de *Text Mining*. A secção *Named entity recognition* (NER) permite analisar os termos específicos em Português de Angola e ver quais são os termos gerais ou entidades nomeadas. Debruçamo-nos sobre *dependency parsing*, pois é fundamental para a comparação de estruturas passivas eventivas, resultativas e estativas em Português de Angola e Europeu que abordaremos no capítulo 11. Além disso, descrevemos as medidas de desempenho do método computacional e teste estatístico. No próximo capítulo, abordamos a aplicação de técnicas enunciadas e descrição da metodologia de extração de empréstimos lexicais.

Parte III

Técnicas de Text Mining e Processamento de Linguagem Natural em Lexicologia Contrastiva

Capítulo 8 – Metodologia de Processamento de Dados para a Detecção de Empréstimos

Este capítulo, à guisa de estudo preliminar dos passos mencionados no Capítulo 7, que descreve Técnicas de *Text Mining* e Processamento de Linguagem Natural, é dedicado ao procedimento metodológico para o estudo comparativo e tem quatro objetivos fundamentais. Em primeiro lugar, pretendemos descrever a constituição e caracterização do corpus escrito e a transcrição de corpus oral do Português de Angola e do Português Europeu. A seguir, descrevemos a sumarização dos dados estatísticos dos corpora. Por fim, apresentamos os procedimentos computacionais para o pré-processamento de dados e estudo comparativo prévio com o objetivo de detetar os candidatos a empréstimo de alguns textos do Português de Angola. A próxima secção apresenta mais informações sobre este processo, mas sem entrar em muitos pormenores.

8.1. Metodologia adotada – visão geral

Esta secção incide fundamentalmente na triagem documental para a preparação dos dados. Com efeito, aplicamos procedimentos de *Text Mining* e Processamento de Linguagem Natural, a qual, sendo uma vertente da inteligência artificial que ajuda os computadores a entender, interpretar e manipular a linguagem humana, resulta de diversas disciplinas, incluindo as ciências da computação e a linguística computacional, que buscam preencher a lacuna entre a comunicação humana e o entendimento dos computadores.

Em casos de volumes de dados textuais, o Processamento de Linguagem Natural coadjuva os computadores a comunicar com os seres humanos na sua própria linguagem e escala outras tarefas relacionadas com a linguagem. Assim, possibilita que os computadores leiam textos, ouçam, interpretem fala e indiquem sentimento. Na Fig. 8.1, podemos observar a tríade de Processamento de Linguagem Natural.

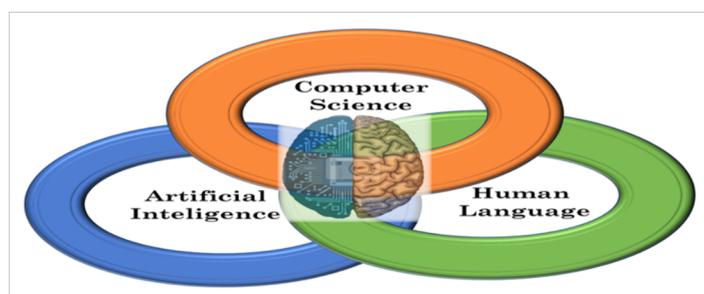


Fig. 8.1 – Representação da tríade de Processamento de Linguagem Natural

Retomamos alguns dos objetivos já enunciados no primeiro capítulo, *Introdução*, na medida em que são os objetivos centrais para a descrição do estudo descrito nesta parte da tese. Objetivamos comparar as formas lexicais do Português de Angola (PA) e o Português Europeu (PE), a fim de encontrar empréstimos lexicais dimanantes de Angola. Assim, o método adotado envolve os seguintes passos:

1. Constituição de corpora a serem usados nos nossos estudos.
2. Pré-processamento de corpora.
3. Identificação dos elementos lexicais em PA, EL-PA, com o foco, em primeiro lugar, nas palavras simples.
4. Identificação de candidatos a empréstimos lexicais.

Os seguintes passos fazem parte da metodologia adotada, mas são descritos no próximo capítulo (Capítulo 9):

5. Análise do subconjunto de candidatos a empréstimo manualmente, para identificar um subconjunto de empréstimos lexicais verdadeiros e distingui-los dos restantes casos, que foram erradamente incluídos na categoria de empréstimos.
6. Análise e organização da informação relativa aos empréstimos.

A seguir, apresentamos mais pormenores sobre cada um dos pontos acima.

Constituição de corpora

Para a concretização do intitulado desta secção, constituímos o corpus do Português de Angola e Português Europeu. Quanto ao primeiro corpus, este é composto por um livro, transcrição de vários telejornais e jornais, cuja descrição detalhada está na Secção 8.2.

Pré-processamento de dados

Este passo envolve a remoção de pontuação e números, conversão para letras minúsculas, remoção de *stop words*, *stemming*/lematização entre outras operações. Mais pormenores sobre estas operações estão na Secção 8.3. Na Secção 8.4, apresentamos mais detalhes sobre os dados pré-processados com o recurso à matriz documento-termo.

Identificação dos candidatos a empréstimos lexicais

O objetivo é elaborar um estudo no qual os elementos lexicais identificados de EL-PA são comparados com os elementos EL-PE do léxico disponível. Este passo é conseguido pela

identificação do subconjunto de unidades léxicas em texto de EL-PA que não aparecem no léxico de PE. Desse modo, foram extraídas, por exemplo, as seguintes palavras:

"ndunduma" "coxilar" "quindumbo" "xuaxualhar" "muxima"
"dicanza" "monangamba" "malange" "kissangua" "molumbar"

Depois, elaboramos um estudo que mostra quão frequente é cada ocorrência. Isso é útil, pois os elementos pouco frequentes poderão ter como origem erros na escrita ou números, que têm que ser eliminados. Mais detalhes sobre este processo são apresentados na Secção 8.5.

8.2. Constituição e caracterização de corpora

Entende-se por corpus a coleção de textos em linguagem natural que é armazenada eletronicamente e processada (Bijeikienė & Tamošiūnaitė, 2013; Sager, 1990). Em virtude de nos cingirmos à metodologia semasiológica e tratar-se de estudo comparativo, constituímos um corpus da variedade do Português de Angola composto por:

- *A Conjura (2008)*
- *Jornal de Angola (2019-2020)*
- *Telejornal de Angola (2020)*¹⁸

Quanto ao corpus do Português Europeu, este é composto por:

- *Corpus de Extratos de Textos Eletrónicos MCT/Público (CETEMPúblico)* disponível na *Linguateca*
- *Léxico do Português Europeu* (Gamallo, Léxico para Português Europeu, s.d)¹⁹
- *Telejornal de Portugal (2020)*²⁰

Relativamente a *A Conjura*, é uma obra literária angolana. Digitalizamos o texto e transformamo-lo em formato *word* e depois em *txt*. Visto que cada edição tem 32 páginas, o

¹⁸ Usamos o telejornal de julho e agosto de 2020 da TPA 1 das vinte horas, cuja duração de cada emissão é cerca de quarenta e sete minutos a uma hora. Este está disponível no *Youtube*.

¹⁹ Trata-se do Léxico do Português Europeu estruturado pela Universidade de Santiago de Compostela.

²⁰ Usamos o telejornal das vinte horas da TVI de julho e agosto de 2020, o qual está disponível na *TVIPlay*.

Jornal de Angola, em virtude da sua proporção, foi processado em formato *pdf*. Em ambos os casos, trata-se de dados não estruturados.

Ao longo da nossa abordagem sobre a caracterização dos corpora, optamos pela tipologia de Torruella e Llisterri (1999), Sinclair (2005) e Llamazares (2008), os quais sublinham que a tipologia de corpus é estabelecida segundo os sete parâmetros seguintes: a modalidade da língua, o número de línguas, os limites do corpus, a especificidade do texto, o período temporal que abarcam os textos, o tamanho dos textos e, por fim, o tratamento aplicado ao corpus.

Modalidade da língua

Concernente à modalidade da língua, o corpus pode ser escrito, oral ou misto. O corpus escrito contém unicamente a modalidade escrita da língua. Os orais contém unicamente amostras orais da língua. O corpus misto combina ambas as modalidades da língua (escrita e oral) (Llamazares, 2008, pp. 343-344). Podemos notar que, quanto à modalidade da língua, o nosso corpus de análise é misto.

Número de línguas

Quanto ao número de línguas, consideramos que o corpus pode ser monolíngue, bilingue ou multilingue (Llamazares, 2008, p. 343). Em virtude de o nosso corpus ser somente escrito em português, deduzimos, portanto, que se trata de um corpus monolíngue. Os corpora bilingues ou multilingues são, geralmente, usados para análise contrastiva de sistemas linguísticos e para equivalências em tradução.

Limites do corpus

No que se refere aos limites do corpus, o mesmo pode ser classificado em corpus fechado ou aberto. Consideramos que o nosso corpus de análise é fechado. O corpus fechado consiste num número finito de palavras, que é estabelecido antes da compilação do corpus. O corpus aberto ou de monitorização é dinâmico, está em constante crescimento, geralmente através da introdução periódica de novas quantidades de textos de acordo com proporções previamente definidas (Llamazares, 2008, p. 345).

Especificidade dos textos

Quanto à especificidade dos textos, o corpus pode ser geral ou especializado (Llamazares, 2008, p. 345). O corpus geral pretende refletir sobre a língua comum utilizada pelos falantes em situações comunicativas normais (Torruella & Llisterri, 1999, p. 10). O corpus

especializado contém textos que podem ter dados para a descrição de um tipo de língua de especialidade, como por exemplo textos jurídicos (Llamazares, 2008, p. 345). Neste trabalho, constituímos um corpus geral.

Período temporal dos textos

Ao debruçarmo-nos sobre o período temporal dos textos, referimo-nos ao ano em que o texto é produzido. Concernente ao período temporal que abarcam os textos, um corpus pode ser diacrónico ou sincrónico (Llamazares, 2008, p. 345). O corpus diacrónico inclui texto de diferentes épocas com a finalidade de se observar a evolução da língua. Dissímil deste, o corpus sincrónico tem como finalidade permitir o estudo de uma ou mais variedades linguísticas em um determinado tempo (ano, período, etc), sem prestar atenção à evolução. Assim, o nosso corpus de análise é sincrónico.

Tamanho do corpus

Quanto ao tamanho, um corpus pode ser de referência ou textual. O primeiro é formado por fragmento de textos dos documentos que o constituem, como é o caso do *CETEMPúblico*. O segundo inclui textos completos (Llamazares, 2008; Torruela & Llisterri, 1999). Estes são mais comuns, quando o objeto é uma língua de especialidade. Em virtude disso, notamos que o nosso corpus de análise é, nitidamente, de referência e textual.

Tratamento aplicado ao corpus

Quanto ao tratamento aplicado ao corpus, este pode ser não codificado ou codificado. O corpus não anotado ou não codificado consiste em textos armazenados sem qualquer formatação e sem adição de qualquer informação adicional, tais como códigos ou anotações. O corpus codificado ou anotado consiste em textos aos quais certas informações foram acrescentadas manual ou automaticamente (Llamazares, 2008, p. 346). O nosso corpus é não codificado, exceto o Léxico do Português Europeu e o *CETEMPúblico*.

8.2.1. Constituição do corpus de Telejornal de Angola

Quanto ao *Telejornal*, a fim de o processar, gravamos o áudio, utilizando o programa *Audacity*, com o qual, depois da gravação, procedemos à normalização das ondas sonoras e à remoção de ruídos, conforme podemos notar na Fig. 8.2. Usamos o programa *Audacity*, em virtude das múltiplas funcionalidades acústicas que possui. É evidente que este procedimento facilita a transcrição automática.

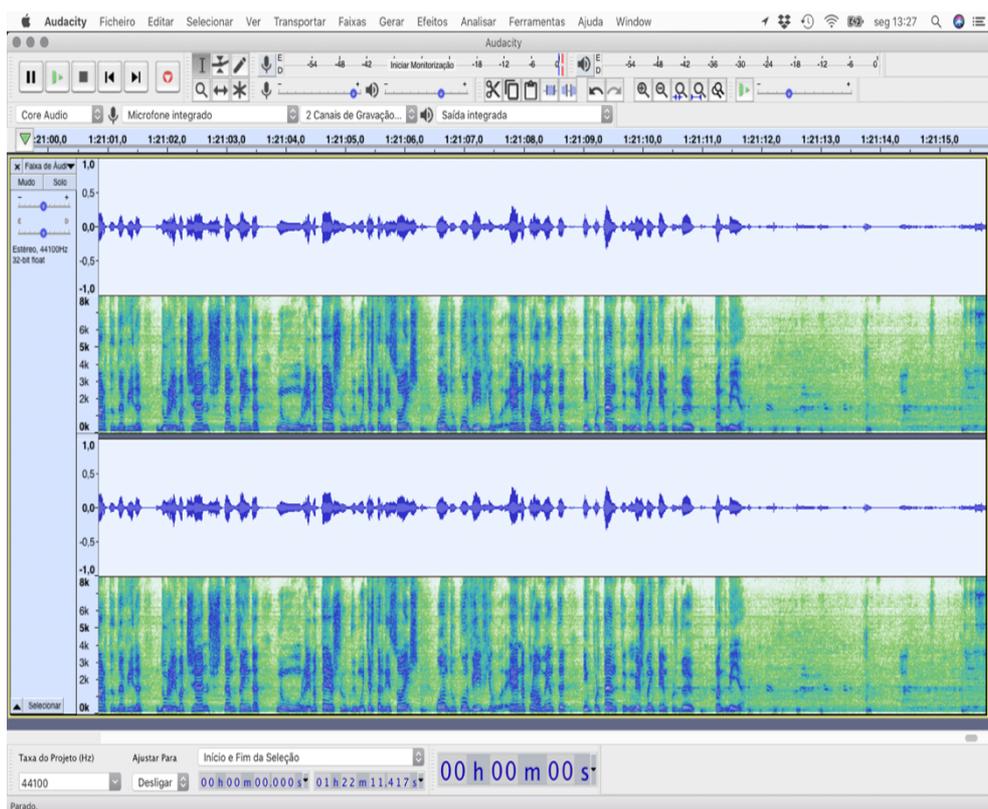


Fig. 8.2 – Gravação do corpus Telejornal usando o *Audacity*

Posteriormente, usamos como recurso o sistema de transcrição automática *Dictate*, o qual transcreve para 51 idiomas incluindo o português europeu, inglês e espanhol, conforme ilustrado na Fig. 8.3. Visto que o sistema de transcrição de voz não insere a pontuação, procedemos à verificação da transcrição com exatidão e inserimos a pontuação, tendo em conta a fluidez do discurso. Reconhecemos que o sistema *word* também faz transcrição automática, porém constatamos que transcreve para a variedade do Português Brasileiro, o que, em virtude da natureza do nosso estudo, não seria efetivamente o mais apropriado. Tendo em conta a subsecção 7.1.1, *Estrutura de dados*, consideramos que o corpus do *Telejornal* também é dado não estruturado.

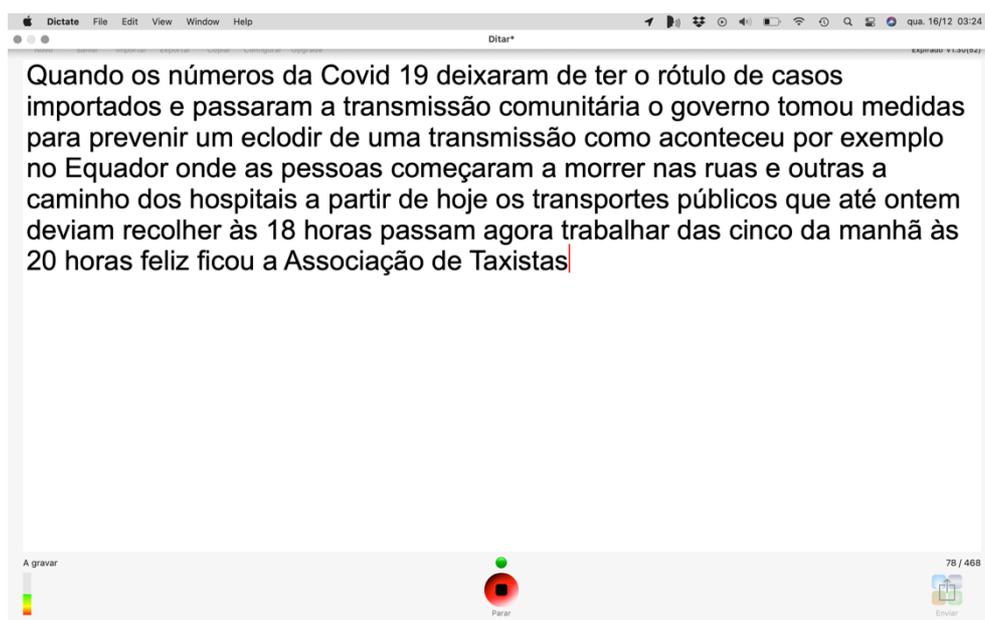


Fig. 8.3 – Transcrição automática do corpus Telejornal usando o *Dictate*

8.2.2. Sumário de corpora usados

Procedemos à lexicometria, a qual pode ser entendida como um conjunto de métodos que permitem operar, a partir de análise estatística, reorganização do vocabulário de um corpus (Bonnafous & Tournier, 1995). Na Tabela 8.1, podemos observar a constituição e estatística descritiva do corpus do Português de Angola (Wickham, 2015 ; Kennedy, 1998). As 23 edições do *Jornal de Angola* foram separadas em 744 documentos.

	Documentos	Frases	Palavras
A Conjura	1	2.026	10.392
Jornal de Angola	23	35.861	82.572
Telejornal TPA	5	1.707	30.524
Total	29	39.594	123.488

Tabela 8.1 – Sumário do corpus do Português de Angola

Escolhemos este corpus, pois precisamos de textos de diversos géneros discursivos dentro do mesmo curto espaço temporal que expressam com fiabilidade a atual variedade do Português de Angola, quer no código escrito quer no oral. Em virtude de ser um estudo comparativo, aplicamos o mesmo critério à constituição do corpus do Português Europeu, podendo observar a sua estatística descritiva na Tabela 8.2.

	Documentos	Frases	Palavras
CETEMPúblico	1	200	3.440
Telejornal TVI	3	2.271	29.524
Léxico do Português Europeu	1	–	1.110.724
Total	5	2.471	1.143.688

Tabela 8.2 – Sumário do corpus do Português Europeu

Tendo em conta os parâmetros de classificação de corpus provenientes da Linguística de Corpus propostos por Torruella e Llisterri (1999), Adam (2005), Sinclair (2005), Llamazares (2008) e Bijeikienė e Tamošiūnaitė (2013), consideramos que o corpus analisado é conforme a descrição da Tabela 8.3.

	Parâmetros	Classificação
Tipologia de Corpus	Modalidade da língua	Misto
	Número de línguas	Monolingue
	Limites	Fechado
	Especificidades dos textos	<i>Corpus</i> geral
	Período temporal	Sincrónico
	Tamanho	Referência e Textual
	Tratamento aplicado ao <i>corpus</i>	Codificado

Tabela 8.3 – Parâmetros de classificação do corpus

O nosso objetivo é determinar o número de palavras por classe lexical. No próximo capítulo (Cap. 9), apresentamos o procedimento que envolve o pacote *udpipe*, que pode ser usado para esse fim, o que permite extrair a quantificação de palavras por classe lexical. Aplicamos este procedimento ao texto do *Jornal de Angola* e *Telejornal*. Podemos observar a quantificação na Tabela 8.4, a qual apresenta os dados estatísticos por classe lexical do corpus do Português de Angola. Recorde-se que a explicação dos termos usados na Tabela 8.4 está no Capítulo 7, Subsecção 7.3.1 *Part-of-speech tagging*. As classes sublinhadas são as que decidimos analisar, pois estas têm maior produtividade de empréstimo.

	ADJ	ADP	ADV	AUX	CCONJ	DET	INTJ	NOUN	NUM	PART	PRON	PROPN	PUNCT	SCONJ	SYM	VERB	X
A Conjura	2.321	6.884	2.578	1.251	1.788	7.286	6	8.442	475	1	2.708	2.456	5.205	664	-	5.072	1
Jornal de Angola	43.534	158.990	23.375	18.058	23.979	129.603	108	175.811	35.862	1.006	24.727	160.104	158.510	6.921	1.222	70.425	1.265
Telejornal	1.890	6.130	1.961	1.585	955	5.789	7	7.523	649	-	1.386	1.434	2.754	517	22	3.456	-
Total	47.745	172.004	27.914	20.894	26.722	142.678	121	191.776	36.986	1.007	28.821	163.994	166.469	8.102	1.244	78.953	1.266

Tabela 8.4 – Frequência de ocorrência do corpus do PA por classe lexical

8.3. Importação e pré-processamento de dados

Tendo como modelo de procedimento de *Text Mining* (Feinerer, 2019), aplicamos algumas técnicas de *regular expressions* que é um conteúdo assaz complexo e avançado descrito no Capítulo 7, Subsecção 7.3.4, *Expressões regulares*. O processo envolve importação de dados e pré-processamento que estão descritos nas subsecções seguintes. (Wickham & Grolemund, 2017, pp. 200-202). Nesta secção, seguimos os parâmetros indicados no diagrama da Fig. 8.4.

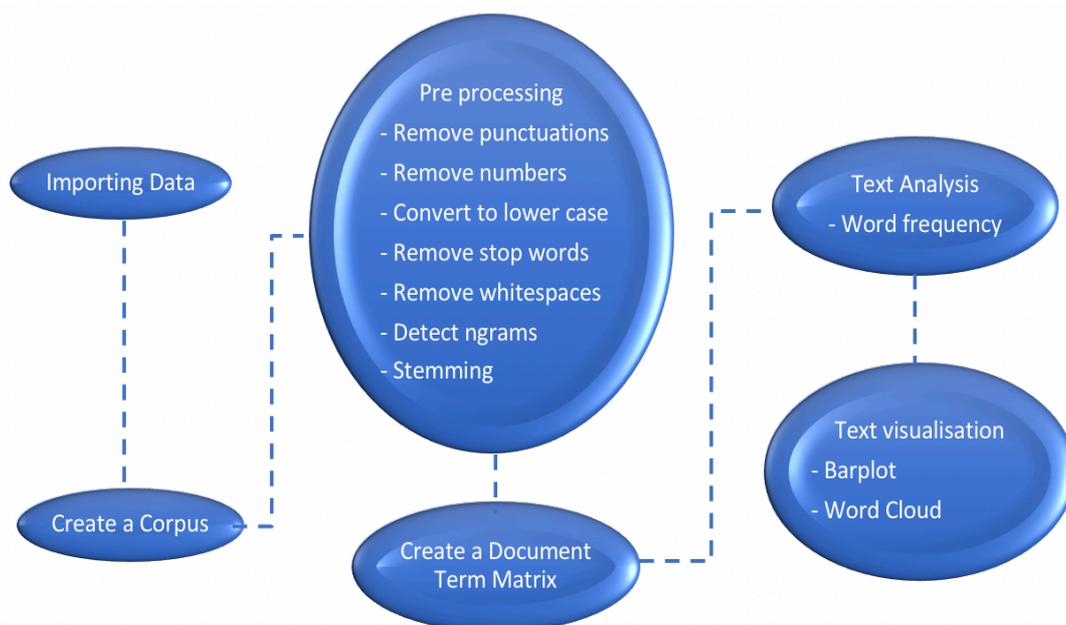


Fig. 8.4 – Procedimentos de *Text Mining* aplicados ao corpus, adaptados de Edureka (2019)

8.3.1. Importação de dados e constituição de corpus em R

Nesta subsecção, demonstramos os procedimentos para a constituição do corpus. Começamos por instalar os pacotes *NLP*, *tm* e *stringr*, conforme a Fig. 8.5. A primeira linha indica a demonstração correta de palavras com acentos e sinais diacríticos. A segunda, terceira e quarta linhas indicam os pacotes a serem usados para o pré-processamento dos dados. Por fim, a instrução *readLines* permite proceder à leitura do texto.

```
> Sys.setlocale("LC_ALL","pt_BR.UTF-8")
> library("NLP")
> library("tm")
> library("stringr")
> text.conjura <- readLines("A Conjura.txt")
> text.conjura
[1] "A CONJURA. Em memória de Pedro da Paixão Franco. Capítulo primeiro.
Aqui se conta da chegada de Jerónimo Caninguili, moço benguelense, à velha
cidade de São Paulo da Assunção de Luanda[...]"
```

Fig. 8.5 – Importação de dados e criação de corpus de *A Conjura*

8.3.2. Pré-processamento de dados

Nesta subsecção, apresentamos algumas técnicas de *Text Mining* e Processamento de Linguagem Natural de pré-processamento de dados, ou mais precisamente remoção de pontuação, números, conversão das unidades léxicas em letra minúscula, remoção de unidades léxicas com uma letra, remoção de *stop words* e de espaços extra em brancos. Descrevemos os procedimentos enunciados.

Remoção de pontuação

Em virtude da sensibilidade do R e tendo em conta os procedimentos posteriores, para a concretização deste passo, começamos por normalizar o texto seguindo o esquema de *regular expressions* de Wickham e Grolemond (2017, pp. 200-202), usando a instrução

```
> text.conjura <- gsub(pattern="\\W",replace=" ",text.conjura)
> text.conjura
[1] "A CONJURA Em memória de Pedro da Paixão Franco Capítulo primeiro
Aqui se conta da chegada de Jerónimo Caninguili moço benguelense à
velha cidade de São Paulo da Assunção de Luanda[...]"
```

para a remoção da pontuação do texto de *A Conjura*, pois objetivamos fazer a *word cloud*.

Remoção de números

Recorde-se que a primeira distinção entre *Data e Text Mining* consiste em: números *versus* texto. Isso não significa que estes sejam dois conceitos muito distintos. A composição dos exemplos é muito diferente (Weiss, Indurkha, & Zhang, 2015). Depois da remoção da pontuação, usamos a instrução

```
> text.conjura <- gsub(pattern="\\d",replace=" ",text.conjura)
> text.conjura
[1] "A CONJURA Em memória de Pedro da Paixão Franco Capítulo primeiro
Aqui se conta da chegada de Jerónimo Caninguili moço benguelense à velha
cidade de São Paulo da Assunção de Luanda[...]"
```

para elidirmos todos os números do texto. Foi necessário suprimi-los para termos as palavras.

Conversão para letras minúsculas

Repare-se que, no início do nosso vetor ilustrado na Fig. 8.5, a palavra *conjura* aparece com todas as letras maiúsculas, o que de facto não dá nenhum valor acrescido à nossa posterior matriz de termos. Usamos a instrução

```
> text.conjura <- tolower(text.conjura)
> text.conjura
[1] "a conjura em memória de pedro da paixão franco capítulo primeiro
aqui se conta da chegada de jerónimo caninguili moço benguelense à velha
cidade de são paulo da assunção de luanda[...]"
```

para normalizar o texto, transformando as palavras em minúscula. Este procedimento era necessário, para evitarmos que na *word cloud* apareçam adjetivos como *difícil* e nomes de meses como *junho* em maiúscula.

Remoção de unidades lexicais com uma letra

A remoção dos números resultou em algumas unidades lexicais com apenas uma letra, o que dificultaria os posteriores pré-processamentos, o *stemming*, *matriz documento-termo* e *word cloud*. Contudo, eliminamos as unidades lexicais com uma letra, utilizando a instrução

```
> text.conjura <- gsub(pattern="\\b[A-z]\\b{1}",replace=" ",text.conjura)
> text.conjura
[1] " conjura em memória de pedro da paixão franco capítulo primeiro
aqui se conta da chegada de jerónimo caninguili moço benguelense à velha
cidade de são paulo da assunção de luanda[...]"
```

Remoção de “stop words”

Palavras comuns como *a, de, para, etc.* podem não ser informativas sobre o conteúdo de um texto. Filtrar palavras como estas permite-nos reduzir o tamanho dos dados, reduzir a carga computacional e, em muitos casos, melhora também a precisão. Usamos as instruções

```
> head(stopwords("pt"))
[1] "de" "a" "o" "que" "e" "do"
> text.conjura <-tm::removeWords(text.conjura,tm::stopwords(kind="pt"))
> text.conjura
[1] " conjura memória pedro paixão franco capítulo primeiro aqui
conta chegada jerónimo caninguili moço benguelense velha cidade
paulo assunção luanda[...]
```

para remover *stop words*. Vários pacotes de análise de texto, tais como *NLP* e *TM*, fornecem listas de *stop words* para várias línguas que podem ser usadas para filtrá-las. (Welbers, Atteveldt, & Benoit, 2017, p. 251).

Remoção de extras espaços brancos (“white spaces”)

A partir dos procedimentos anteriores, notamos que a remoção de pontuação, números, unidades lexicais com uma letra e de *stop words* resultou em alguns espaços vazios que tivemos de suprimir. A instrução

```
> text.conjura <-trimws(gsub(pattern="\\s+", " ", text.conjura))
> text.conjura
[1] "conjura memória pedro paixão franco capítulo primeiro aqui conta
chegada jerónimo caninguili moço benguelense velha cidade paulo assunção
luanda[...]
```

remove os espaços vazios do conjunto de dados resultantes da elisão destes elementos.

Stemming

Por *stemming* pode entender-se o processo de normalização de texto que consiste num algoritmo baseado em regras da língua em causa que converte as palavras na sua forma base, retirando o constituinte temático, a flexão morfológica e o morfema de tempo-modo-aspeto no caso dos verbos (Welbers, Atteveldt, & Benoit, 2017, pp. 250-251).

Para esta parte, instalamos o pacote *SnowballC*. Para a concretização deste passo, usamos a instrução

```
> install.packages("SnowballC")
> library("SnowballC")
> text.conjura.stem <-stemDocument(text.conjura,language="portuguese")
> text.conjura.stem
[1] "conjur memór pedr paixã franc capítul primeir aqu cont cheg jerónim
caninguil moc benguelens velh cidad paul assunçã luand [...]
```

Além de possibilitar a verificação do radical de cada palavra, o *stemming* permite determinar o número de formas base de palavras do texto, a partir da sua frequência e uso da instrução

```
> text.conjura.stem <- str_split(text.conjura.stem, pattern = "\\s+")
> text.conjura.stem <- unlist(text.conjura.stem)
> freq.stem.conj <- sort(table(text.conjura.stem), decreasing=TRUE)
> freq.stem.conj[1:16]
  tod  severin   é   hav   outr  velh  aind  cas
180   152  133  133  116   98   97   89
ferreir   diz   mã  carm  luand  dias  adolf  anos
  84    81   80   76   75   71   70   68

> length(freq.stem.conj)
[1] 4668
```

onde podemos notar que a quantificação pode ser alterada, pois contabiliza-se não as palavras em si, mas a frequência de radicais unindo as frequências de todos os *cognatas* com a mesma base, ou mais precisamente, palavras com a mesma base são contadas como uma. Verificamos que *A Conjura* tem 4.668 formas base.

Recorde-se que a lematização cumpre um propósito símil ao do *stemming*, mas não corta as extremidades das formas para normalizá-las. Substitui efetivamente as flexões pela sua forma base ou lema, permitindo verificar os paradigmas do lexema (Welbers, Atteveldt, & Benoit, 2017; Haspelmath, 2002). No próximo capítulo, usamos a lematização.

8.4. Matriz Documento-Termo e Termo-Documento

Esta secção é a aplicação dos procedimentos enunciados no Capítulo 7, Subsecção 7.2.2, *Representação de texto*. Importa referir que os *data frames* e as matrizes, ambos com uma estrutura matricial, são as estruturas de dados centrais na representação de dados. Os primeiros apresentam algumas vantagens, nomeadamente pela sua capacidade de guardar numa forma matricial variáveis numéricas e nominais, ao passo que as matrizes devem ser homogéneas, permitindo simplesmente guardar dados de um determinado tipo, geralmente dados numéricos (Rocha & Ferreira, 2017).

Consideramos que os dados textuais podem ser resumidos numa matriz documento-termo e termo-documento. Com efeito, resumimos os dados de *A Conjura* numa matriz documento-termo, cujas linhas e colunas correspondem, respetivamente, a documento e termo. Convém notar que, para procedermos à matriz documento-termo, não foi aplicado o *stemming*, pois pretendíamos verificar as palavras e não os radicais das palavras. As entradas da matriz são as frequências observadas. Na Fig. 8.6, podemos observar, na primeira linha de instrução, a constituição do corpus. Na segunda, transformamos os dados

textuais numa matriz documento-termo. Na terceira, visualizamos uma parte da matriz, nomeadamente as colunas 6175 a 6184.

```

> corpus.conjura <- Corpus(VectorSource(text.conjura))
> dtm <- DocumentTermMatrix(corpus.conjura)
> dtm
<<DocumentTermMatrix (documents: 1, terms: 7760)>>
Non-/sparse entries: 7760/0
Sparsity           : 0%
Maximal term length: 18
Weighting          : term frequency (tf)

> m <- as.matrix(dtm[,6175:6184])
> m
      Terms
Docs   quiocos quis quisera quisesse quisessem quissama quissanguela quissondes quissongo quitandeira
A Conjura  1  6    2    1    1    3    1    1    4    4

```

Fig. 8.6 – Matriz Documento-Termo de *A Conjura*

Procedemos também ao resumo dos dados textuais do *Jornal de Angola*. Assim, fizemos a matriz de termos por documento. O utilizador pode decidir que representação usar. O algoritmo de contagem dos termos permite fazer a contagem do número de vezes que os termos aparecem em cada documento de um corpus. A instrução

```

> corpus.jang <- Corpus(VectorSource(text.jang))
> tdm.jang <- TermDocumentMatrix(corpus.jang)
<<TermDocumentMatrix (terms: 82572, documents: 744)>>
Non-/sparse entries: 388742/61044826
Sparsity           : 99%
Maximal term length: 144
Weighting: term frequency(tf)

```

permite observar o número de termos sem remoção de *stop word* e o número de documentos.

Recorde-se que o *Jornal de Angola* foi processado em formato *pdf*. Assim, utilizamos os pacotes *pdftools*, *NLP*, *tm* e *stringr*. Depois do pré-processamento que consistiu na remoção de números, pontuação, *stop words*, unidades lexicais com apenas uma letra, espaços vazios e transformação em letra minúscula, constatamos que o número de termos diminuiu para 44.463. Repare-se que, durante o procedimento computacional dos dados textuais do *Jornal de Angola*, constatamos que estes dados foram gravados em vários subdocumentos. Com efeito, as edições usadas foram fragmentadas em 744 documentos, conforme podemos observar na Fig. 8.7. Na variável *tdm*, notamos parte da matriz de termos,

onde podemos verificar termos relativos ao contexto angolano como *Mbandi*, *Luanda* e a sua frequência por documentos. Na variável *m_1*, visualizamos uma parte da matriz, nomeadamente as linhas 137 a 144 e as colunas 480 a 500.

```
> tdm <- TermDocumentMatrix(corpus.jang.red)
<<TermDocumentMatrix (terms: 44463, documents: 744)>>
Non-/sparse entries: 316082/32764390
Sparsity           : 99%
Maximal term length: 144
Weighting: term frequency(tf)

> m_1 <- as.matrix(tdm[137:144,480:500])
> m_1
```

Terms	Docs																				
	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500
luanda	1	2	0	10	0	12	3	5	7	1	1	4	0	1	2	6	0	0	0	4	0
mais	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
manuel	6	5	0	0	0	1	2	1	1	0	3	3	0	0	5	1	0	0	0	0	2
marinha	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
markel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mbandi	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
meios	1	1	0	0	1	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0
memória	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

Fig. 8.7 – Matriz de Termos do corpus *Jornal de Angola* com remoção de *stop words*

8.4.1. Frequência de alguns termos

Nesta subsecção, debruçamo-nos sobre a frequência de termos em *A Conjura* e *Jornal de Angola*. No que se refere à *Conjura*, usamos as instruções *str_split* e *unlist* para separar os termos do corpus e transformá-los numa lista, respetivamente. Com o auxílio da instrução *as.data.frame* da Fig. 8.8, a primeira linha indica a transformação do vetor em *data frame*. Na segunda linha, com a instrução *order*, notamos a ordenação dos termos por frequência.

```
> term.conj <- as.data.frame(text.freq.conjura)
> freq.conj[order(term.conj$Freq, decreasing = TRUE), ]
```

	Words	Freq
1	severino	152
2	é	133
3	ainda	97
4	ferreira	84
5	velho	79

Fig. 8.8 – Frequência de termos do corpus *A Conjura*

Aplicamos o mesmo procedimento aos dados textuais do corpus *Jornal de Angola*. Após a verificação da matriz de termos, procedemos à visualização das formas lexicais mais frequentes no corpus *Jornal de Angola*, conforme podemos observar na Fig. 8.9.

```
> freq.jang[order(term.jang$Freq,decreasing = TRUE),]
  Words  Freq
1     é    98
2    ser    82
3  luanda   58
4  angola   56
5     dia   54
```

Fig. 8.9 – Frequência de termos do corpus *Jornal de Angola*

8.4.2. Visualização de termos : *barplot* e *word cloud*

Para a visualização dos dados de *A Conjura* a partir do *data frame* ilustrado na Fig. 8.9, fizemos o gráfico de barras, usando a instrução indicada na Fig. 8.10, em cujo resultado podemos verificar os 7 termos mais frequentes.

Em virtude de a presente investigação ser um estudo comparativo e o *data frame* apresentar as frequências das unidades léxicas com mais ocorrência no conjunto de dados textuais de *A Conjura*, decidimos que a *word cloud* seria, não dos elementos que mais ocorrem no corpus, mas dos elementos detetados como candidatos a empréstimos dimanantes de Angola, conforme podemos verificar na secção subsequente. Procedemos ao *barplot* com o auxílio da instrução

```
> barplot(freq.conj,names.arg=terms_1,xlab="Text_Conjura",ylab =
"Frequency", col="lightgray")
```

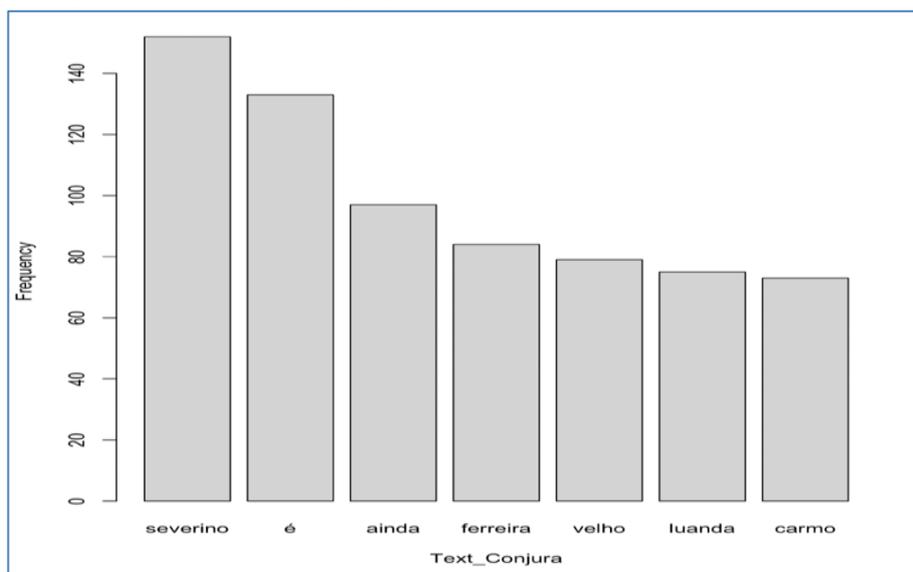


Fig. 8.10 – *Barplot* dos termos mais frequentes em *A Conjura*

8.5. Detecção de candidatos a empréstimos

Para procedermos ao estudo comparativo, começamos pela separação das unidades lexicais do texto e transformamo-las num vetor, cujo resultado podemos ver na Fig. 8.11.

```
> text.conjura <- str_split(text.conjura, pattern = "\\s+")
> text.conjura <- unlist(text.conjura)
> text.conjura[1:20]
[1] "conjura"      "memória"      "pedro"        "paixão"       "franco"
[6] "capítulo"    "primeiro"     "aqui"         "conta"        "chegada"
[11] "jerónimo"    "caninguiili"  "moço"         "benguelense"  "velha"
[16] "cidade"      "paulo"        "assunção"     "luanda"       "enquanto"
```

Fig. 8.11 – Tokenização do corpus A Conjura

Tendo em conta este vetor ilustrado na Fig. 8.11, detetamos os candidatos a empréstimo, verificando os elementos que estão em `text.conjura` e não constam do Léxico do Português Europeu (LPE-USC) com a instrução `setdiff`. Na Fig. 8.12, podemos observar o uso da instrução `setdiff` e os vinte candidatos a empréstimo selecionados.

```
> text.conjura.cand <- setdiff(text.conjura, léxico[,1])
> text.conjura.cand[1:20]
[1] "caninguiili"  "paulo"        "luanda"       "alice"        "malange"
[6] "humbes"      "berlim"       "arantes"      "dezassete"    "mil"
[11] "josephine"   "andua"        "carmo"        "dezasseis"   "vavó"
[16] "uála"        "ingombotas"  "ezequiel"    "feiura"       "alfredo"
> length(text.conjura.cand)
[1] 442
```

Fig. 8.12 – Candidatos a empréstimo lexical do corpus A Conjura

Podemos constatar que, embora esta abordagem permita identificar os candidatos a empréstimo, não permite separar as palavras, de maneira automática, em classes lexicais nem apresenta as categorias sintáticas. Por esse motivo, decidimos seguir uma abordagem que assenta no uso do pacote `udpipe`, descrita no próximo capítulo (Cap. 9).

Constatamos que a abordagem acima descreve 442 candidatos a empréstimos, enquanto o uso do `udpipe` descrito na Tabela 9.3 revela 621 candidatos. Depreendemos que isso se deve ao facto de unidades léxicas serem classificadas como pertencentes a mais de uma classe lexical. Para ilustrar, o lexema *muxima* foi classificado como nome próprio, adjetivo e verbo.

Recorde-se que a instrução `setdiff` retorna os elementos sem repetição, portanto todas as unidades léxicas são consideradas como tendo a mesma frequência, que é um. Além disso, indica os lexemas por ordem alfabética. Na Fig. 8.13, podemos observá-las.

```

> freq.cand.conjura <- sort(table(text.conjura.cand), decreasing=TRUE)
> freq.cand.conjura[1:12]
      abranches  acácio  adolfo  afonso  agualusa  ambaquense
           1         1         1         1         1         1
      amboim  ambriz  américa  andembo  andongo  andua
           1         1         1         1         1         1

```

Fig. 8.13 – Frequência falsa dos candidatos a empréstimo do corpus *A Conjura*

Procedemos à extração da frequência destas unidades léxicas. Na Fig. 8.14, podemos observar os 12 candidatos a empréstimos mais frequentes.

```

> text.conj.cand <- text.conjura[! text.conjura %in% léxico[,1]]
> freq.cand.conj <- sort(table(text.conj.cand[! text.conjura %in%
léxico[,1]]),decreasing=TRUE)
> freq.cand.conj[1:12]
      carmo      mil      adolfo      luanda      marimont      souza
           9         9         7         6         5         5
 caninguili     cesar      dois      josephine      judite      galeano
           4         4         4         4         4         3

```

Fig. 8.14 – Frequência dos candidatos a empréstimo do corpus *A Conjura*

Visto que o conjunto de dados de *A Conjura* é constituído por 7.944 unidades lexicais, extraímos a *wordcloud* do conjunto de candidatos a empréstimo, usando o pacote *wordcloud2*. Na Fig. 8.15, podemos observá-la. Para isso, usamos a instrução

```

> wordcloud2(freq.cand.conj,minRotation = -pi/6,maxRotation = -pi/6, minSize =
5,rotateRatio = 0)

```



Fig. 8.15 – *Word cloud* de candidatos a empréstimo do corpus *A Conjura*

Nota-se que muitas unidades léxicas da Fig. 8.15 são nomes próprios. Alguns são empréstimos dimanantes de Angola e outros não. Tendo em conta a classe lexical, constatamos, por exemplo, os seguintes empréstimos:

- Verbos: *kuribeka, uandi, xuaxualhar*;
- Nomes: *buxila, dicanzas, ndunduma, nga, quinzári, etc*;
- Adjetivos: *humbe, pembe, uanga*;
- Nomes próprios: *Andembo, Malanje, Mandume, Songos, Zambi, etc*.

A análise mais detalhada de todos os empréstimos é apresentada no próximo capítulo.

8.6. Síntese do capítulo

Neste capítulo, explicitamos a constituição e caracterização dos corpora. Descrevemos o pré-processamento de dados com o auxílio de algumas técnicas de expressões regulares (*regular expressions*). De seguida, representamos os dados em forma de matrizes documento-termo e termo-documento. Por fim, elaboramos os procedimentos para a deteção de candidatos a empréstimos.

Embora a abordagem descrita permita identificar os candidatos a empréstimo, verificamos que não permite separar as palavras, de maneira automática, em classes lexicais nem apresenta as categorias sintáticas. Em virtude disso, consideramos fundamental seguir uma abordagem que assenta no uso de pacote *udpipe*, descrita no próximo capítulo, de modo a extrairmos os empréstimos por classe lexical e verificarmos as categorias sintáticas.

Capítulo 9 – Extração de Empréstimos Lexicais e Estudo Comparativo

Este capítulo, continua com a temática do capítulo anterior, nomeadamente a descrição dos procedimentos computacionais para a extração de empréstimos dimanantes de Angola em textos a nossa disposição. Procedemos ao estudo comparativo para a extração de empréstimos por classe lexical com o foco em quatro classes lexicais (verbos, nomes, adjetivos e nomes próprios). A Secção 9.1 descreve o método de deteção de empréstimos verbais. Como o conjunto identificado pelo método automático inclui alguns elementos lexicais que não são empréstimos, é preciso eliminá-los manualmente. Mais pormenores sobre este assunto estão apresentados na Secção 9.2. A Secção 9.3 apresenta o estudo, junto com os resultados, para as outras classes lexicais que consideramos, nomeadamente nomes, adjetivos e nomes próprios. Para cada classe lexical apresentamos o total de empréstimos identificados e quantificamos a parte de empréstimos que não aparece no dicionário de regionalismos. A Secção 9.4 apresenta o sumário de todos os empréstimos identificados. Por fim, explicitamos o processamento incremental de texto, que melhora a eficiência elidindo a sobrecarga do sistema computacional (ver Secção 9.5). Por último, a Secção 9.6. descreve a metodologia e as vantagens do processo automático utilizado aqui.

9.1. Deteção de empréstimos verbais

A deteção de empréstimos verbais envolve os seguintes passos:

- Identificação de palavras em textos em Português de Angola;
- Extração de lema e classe lexical de elementos lexicais nesses textos;
- Uso do Léxico do Português Europeu;
- Extração de candidatos a empréstimo;
- Redução desses candidatos usando empréstimos existentes;
- Identificação de empréstimos no conjunto de candidatos a empréstimo.

A descrição de todos estes passos está apresentada a seguir.

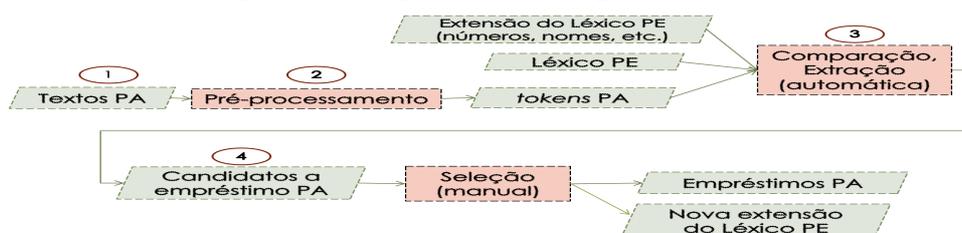


Fig. 9.1 – Representação de extração de empréstimos

9.1.1. Identificação de elementos lexicais em textos do Português de Angola

Este passo consiste na tokenização. Tal como os outros, é atingido com o auxílio à programação em linguagem R. Este passo pode ser feito com algumas instruções básicas de *Text Mining* para a segmentação textual, usando instruções dos pacotes *tm* e *stringr*. Primeiro, pode-se usar a instrução *readLines* para proceder à leitura do texto. A seguir, a instrução *strsplit*, para separar os *tokens*. Por fim, usa-se a instrução *unlist* que transforma os *tokens* num vetor, conforme se nota a seguir.

```
> library(tm)
> library(stringr)
> text <- readLines('conjura.txt')
> text <- strsplit(text, " ")
> text <- unlist(text)
```

Apesar de atingirmos o objetivo de maneira simples, decidimos usar um método mais complexo para obter tokenização, pois esse tem a vantagem de retornar também categorias sintáticas para a nossa análise. Com efeito, para a concretização deste primeiro passo, começamos com *part-of-speech (POS) tagging* do conjunto de dados ilustrado na Fig. 9.1, que, além do *token* demonstrado na quarta coluna (*token*), indica o lema (*lemma*) e a classe lexical (*upos*), além de outras informações fundamentais para o nosso trabalho descrito no Capítulo 11, que trata do estudo de estruturas passivas. Em virtude de o corpus ser de diferente tipologia textual, o processamento foi feito com o auxílio do pacote *udpipe*. Atente-se nos procedimentos computacionais. Na linha 2, é lido o texto de *A Conjura* linha a linha. Mais adiante, é criado um modelo para *part-of-speech tagging* em português, com o auxílio das instruções *udpipe_load_model* e *udpipe_annotate*. Transformamos os dados em *data frame*, usando a instrução *as.data.frame*, conforme a Fig. 9.2, que é uma parte da estrutura de dados que contém 50.445 observações e 17 variáveis. A lista das variáveis (colunas) pode ser obtida através da instrução *str(text.annot.conjura)*.

```
> Sys.setlocale("LC_ALL", "pt_BR.UTF-8")
> text.conjura <- readLines("CONJURA.txt")
> install.packages("udpipe")
> library("udpipe")
> udmodel <- udpipe_download_model(language = "portuguese")
> udmodel <- udpipe_load_model(file = udmodel$file_model)
> text.annot.conjura <- udpipe_annotate(udmodel, x = text.conjura)
> text.annot.conjura <- as.data.frame(text.annot.conjura, detailed = TRUE)
```

```
> text.annot.conjura <- text.annot.conjura[,c("term_id", "token_id", "token",
"lemma", "upos", "xpos", "feats", "head_token_id", "dep_rel", "deps", "misc")]
```

term_id	token_id	token	lemma	upos	xpos	feats	head_token_id	dep_rel	deps	misc
1	1	A	o	DET	<NA>	Definite=Def Gender=Fem Number=Sing PronType=Art	2	det	<NA>	<NA>
2	2	CONJURA	CONJURA	PROPN	<NA>	Gender=Fem Number=Sing	0	root	<NA>	SpaceAfter=No
3	3	.	.	PUNCT	<NA>	<NA>	2	punct	<NA>	<NA>
4	4	Em	em	ADP	<NA>	<NA>	2	case	<NA>	<NA>
5	5	memória	memória	NOUN	<NA>	Gender=Fem Number=Sing	0	root	<NA>	<NA>
6	6	de	de	ADP	<NA>	<NA>	4	case	<NA>	<NA>
7	7	Pedro	Pedro	PROPN	<NA>	Gender=Masc Number=Sing	2	nmod	<NA>	<NA>
8	8	5-6	da	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
9	9	de	de	ADP	<NA>	<NA>	7	case	<NA>	<NA>
10	10	a	o	DET	<NA>	Definite=Def Gender=Fem Number=Sing PronType=Art	7	det	<NA>	<NA>
11	11	Paixão	paixão	PROPN	<NA>	Gender=Fem Number=Sing	4	nmod	<NA>	<NA>
12	12	Franco	Franco	PROPN	<NA>	Number=Sing	7	flat:name	<NA>	SpaceAfter=No
13	13	.	.	PUNCT	<NA>	<NA>	2	punct	<NA>	<NA>
14	14	Capítulo	capítulo	NOUN	<NA>	Gender=Masc Number=Sing	0	root	<NA>	<NA>
15	15	primeiro	primeiro	ADJ	<NA>	Gender=Masc Number=Sing NumType=Ord	1	amod	<NA>	SpaceAfter=No
16	16	.	.	PUNCT	<NA>	<NA>	1	punct	<NA>	<NA>
17	17	Aquí	aquí	ADV	<NA>	<NA>	3	advmod	<NA>	<NA>
18	18	se	se	PRON	<NA>	Case=Acc Gender=Masc Number=Sing Person=3 PronType=Prs	3	expl	<NA>	<NA>
19	19	conta	contar	VERB	<NA>	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	0	root	<NA>	<NA>
20	20	4-5	da	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>	<NA>
21	21	de	de	ADP	<NA>	<NA>	6	case	<NA>	<NA>
22	22	a	o	DET	<NA>	Definite=Def Gender=Fem Number=Sing PronType=Art	6	det	<NA>	<NA>
23	23	chegada	chegada	NOUN	<NA>	Gender=Fem Number=Sing	3	obl	<NA>	<NA>
24	24	de	de	ADP	<NA>	<NA>	8	case	<NA>	<NA>
25	25	Jerónimo	Jerónimo	PROPN	<NA>	Gender=Masc Number=Sing	6	nmod	<NA>	<NA>
26	26	9	Caninquili	Caninquili	PROPN	<NA>	8	flat:name	<NA>	SpaceAfter=No
27	27	10	,	,	PUNCT	<NA>	11	punct	<NA>	<NA>
28	28	11	moço	moço	NOUN	<NA>	8	appos	<NA>	<NA>

Fig. 9.2 – Tokenização e *Part-of-speech Tagging* de A Conjura

9.1.2. Extração de *lema* e classe lexical de elementos lexicais (*tokens*)

A segmentação textual por classes lexicais é fundamental, pois extrairemos *verbos*, *nomes*, *adjetivos* e *nomes próprios* que são empréstimos dimanantes de Angola. Decidimos focar-nos nestas quatro classes lexicais, porque o nosso estudo preliminar mostrou que é nessas classes que mais incidem os empréstimos. Neste contexto, precisamos dos lemas e da classificação lexical para os procedimentos de extração de empréstimos. Para a extração do lema e classe lexical, selecionamos as colunas que correspondem a *lemma* e *upos* com o auxílio da instrução

```
> text.conjura.lemma <- text.annot.conjura[,c("lemma", "upos")]
```

e selecionamos posteriormente todas as classes mencionadas acima. Podemos ver as primeiras quatro linhas na Fig. 9.3.

```
> lemma.text.annot <- text.annot.conjura[,c("lemma", "upos")]
> lemma.text.annot[1:4,]
      lemma      upos
1         o      DET
2  CONJURA  PROPN
3          .  PUNCT
4         em      ADP
```

Fig. 9.3 – Seleção das variáveis *lemma* e *upos*

Extração de lemas verbais

A partir da seleção destas duas variáveis *lemma* e *upos*, procedemos à extração do subconjunto de todos os lemas verbais do conjunto de dados que foram classificados como verbo, com a instrução

```
> lemma.verb <- subset(lemma.text.annot[lemma.text.annot$upos=="VERB",])
```

cujos procedimentos servem de arquétipo para o processamento de outras classes lexicais. Assim, obtivemos uma lista de todos os verbos e outros elementos de *A Conjura*, da qual se pode ver um recorte na Fig. 9.4.

```
> lemma.verb <- subset(lemma.text.annot[lemma.text.annot$upos=="VERB",])
> lemma.verb[1:9,]
      lemma      upos
NA      <NA>      <NA>
19      contar    VERB
NA.1    <NA>      <NA>
NA.2    <NA>      <NA>
NA.3    <NA>      <NA>
52      dar       VERB
NA.4    <NA>      <NA>
67      dizer     VERB
74      soltar    VERB
```

Fig. 9.4 – Extração de alguns lemas verbais de *A Conjura*

Quanto à extração dos verbos, notamos que o conjunto de dados do *data frame* tem ocorrência de NA (*Not Available*) em algumas observações. Posteriormente, eliminamos as observações onde os NA se encontram, tendo como recurso a instrução

```
> lemma.verb <- na.omit(lemma.verb)
```

de modo a que pudéssemos ter apenas os verbos. Tal eliminação pode ser vista na Fig. 9.5. Depois de suprimirmos os NA, tivemos um *data frame* só com todos os verbos do conjunto de dados e notamos que contém 5.071 observações e 2 variáveis (i.e. *lemma*, *upos*).

```
> lemma.verb <- na.omit(lemma.verb)
> lemma.verb
      lemma      upos
19      contar    VERB
52      dar       VERB
67      dizer     VERB
74      soltar    VERB
```

Fig. 9.5 – Supressão de NA e dimensão do *data frame* de verbos

Em virtude de um dos objetivos da investigação ser comparar formas lexicais do corpus da variedade do português em Angola com o léxico disponível para a variedade do português europeu, usamos a instrução `lemma.verb <-c(lemma.verb[,1])`, que extraiu a coluna de *lemma*.

Tendo em conta a sensibilidade do R à maiúscula e minúscula, transformamos todos os *tokens* em minúscula, usando a instrução `lemma.verb <- tolower(lemma.verb)`. Este procedimento também foi aplicado à classe nominal, adjetival e de nomes próprios, cuja quantificação dos lemas se vê na Tabela 9.1, na primeira linha relativa ao texto *A Conjura*. Aplicamos também este procedimento no *Jornal de Angola* e no *Telejornal*, os quais constituem o nosso corpus do português de Angola.

	A Conjura	Jornal de Angola	Telejornal	Total
Verbo	5.072	70.425	3.456	78.953

Tabela 9.1 – Quantificação dos lemas verbais nos textos de PA

Depois da obtenção dos verbos de *A Conjura* (5.072 no total), pudemos reduzi-los, eliminando repetições. A instrução

```
> verb.red.conj <- table(lemma.verb)
> length(verb.red.conj)
[1] 1235
```

gerou a lista de verbos sem repetição. Constatamos que reduziu para 1.235 unidades léxicas. Repetimos este procedimento para as outras 3 classes lexicais. Aplicamos também este procedimento no *Jornal de Angola* e no *Telejornal*. Na Tabela 9.2, podemos observar o resultado da quantificação de elementos sem repetição.

	A Conjura	Jornal de Angola	Telejornal	Total
Verbo	1.235	2.029	540	3.804

Tabela 9.2 – Quantificação dos lemas sem repetição nos textos de PA

9.1.3. Uso do léxico do Português Europeu

Debruçamo-nos sobre o Léxico do Português Europeu (LPE-USC) (Gamallo, Léxico para Português Europeu, s.d), pois precisamos dele para o estudo comparativo das formas lexicais em português angolano e europeu. Com efeito, selecionamos a nova versão do LPE-USC *dicc.src1.txt*, uma vez que a antiga versão tinha simplesmente três colunas e, por causa disso, quando efetuávamos a comparação, o R baralhava os dados. A coluna *w1* corresponde à palavra, *w2* ao lema e *w3* a *part-of-speech tagging*. Assim, por etiqueta *NCMS* entende-se *nome comum, masculino singular*. Introduzimos mais nomes de colunas de modo a ser feita uma melhor leitura dos dados. Procuramos o diretório e procedemos à leitura do ficheiro com a instrução *read.csv*.

Verificamos a dimensão do Léxico do Português Europeu, com o auxílio da instrução *dim(léxico)*, a qual determinou que este tem 1.110.724 observações e vinte colunas, conforme pode ser observado na Fig. 9.6.

```
> léxico <- read.csv("dicc.src1.txt", sep = "")
> head(léxico)
      w1      w2      w3 w4 w5 w6 w7 w8 w9 w10 w11... w20
1  Abril  abril NCMS000
2  Abris  abril NCMP000
3  Agosto agosto NCMS000
4  Agostos agosto NCMP000
5     Ar   ar   NCMS000
6     Au   au   NCMS000

> dim(léxico)
[1] 1110724      20
```

Fig. 9.6 – Uso do Léxico do Português Europeu

9.1.4. Extração de candidatos a empréstimo

Para procedermos à identificação de formas lexicais específicas em português angolano, usamos a estratégia de comparação com o Português Europeu. A instrução

```
> c.emp.verb.conj <- setdiff(names(verb.red.conj),léxico[,1])
```

identifica o subconjunto de termos em *verb.red.conj* que não aparecem na primeira coluna do *léxico[,1]*, da qual obtivemos 92 candidatos a empréstimo lexical, conforme a Tabela 9.3. Na primeira linha, estão os resultados relativos aos dados de *A Conjura*. Aplicamos também este procedimento no *Jornal de Angola* e no *Telejornal*, os quais constituem o corpus do português de Angola, cujos resultados estão, respetivamente, na segunda e terceira linha.

	A Conjura	Jornal de Angola	Telejornal	Total
Verbo	92	1.407	53	1.552

Tabela 9.3 – Quantificação dos candidatos a empréstimo por classe lexical

9.1.5. Redução de candidatos a empréstimo usando empréstimos existentes

Depois da obtenção dos verbos de *Jornal de Angola* (1.407 no total), pudemos reduzi-los, eliminando os verbos extraídos de *A Conjura*. A instrução

```
> c.emp.verb.red.jang <- setdiff(c.emp.verb.jang,c.emp.verb.conj)
> length(c.emp.verb.red.jang)
[1] 1372
```

gerou a lista reduzida de 1.372 candidatos a empréstimo. Verificamos que eliminou 35 unidades léxicas.

Aplicamos o mesmo procedimento para a redução de candidatos a empréstimo no *Telejornal*. Em primeiro lugar, usamos a instrução

```
> c.emp.verb.conj.jang <- union(c.emp.verb.jang,c.emp.verb.conj)
```

que junta os candidatos a empréstimo de *A Conjura* e *Jornal de Angola*. Posteriormente, com o auxílio da instrução

```
> c.emp.verb.red.telj <- setdiff(c.emp.verb.telj,c.emp.verb.conj.jang)
> length(c.emp.verb.red.telj)
[1] 7
```

obtivemos a redução de candidatos a empréstimo do *Telejornal* de 53 para 7 unidades léxicas.

A Secção 9.5. retoma este assunto e descreve o método geral que reutiliza a informação existente para reduzir o conjunto de candidatos identificados no novo texto que está a ser processado.

9.2. Identificação de empréstimos no subconjunto de candidatos

Tendo identificado os candidatos a empréstimo, foi preciso analisar os casos manualmente. Depois dessa análise, vimos que os dados revelam que, além dos empréstimos verbais, a lista de verbos candidatos a empréstimo inclui ainda alguns falsos empréstimos, verbos específicos e outros mal escritos. É preciso elidi-los para termos somente os empréstimos verdadeiros.

Em relação aos empréstimos de *A Conjura*, analisamos os 92 candidatos e extraímos 12 empréstimos verbais e a sua respetiva frequência e organizamos os dados na Tabela 9.7. Os empréstimos verbais verdadeiros podem ser vistos na Tabela 9.4, que contém o empréstimo, a frequência e a indicação de dicionarização. Os elementos com traço positivo (+) constam do *Dicionário de Regionalismos Angolanos* (Ribas, 2014). Os que têm traço negativo (-) não constam das entradas deste dicionário nem do *Dicionário da Língua Portuguesa Contemporânea* (Academia das Ciências de Lisboa & Fundação Calouste Gulbenkian, 2001). O étimo, etimologia e a significação dos empréstimos verbais podem ser vistos na lista constante do Anexo 1 desta investigação.

N.º	Verbo empréstimo	Freq.	Dic.	N.º	Verbo empréstimo	Freq.	Dic.
1	baçular	1	+	7	molumbar	1	+
2	coxilar	2	+	8	muturir	1	+
3	facar	1	-	9	ngar	1	-
4	kuatar	1	+	10	sunguilar	1	+
5	malebelember	1	-	11	uandi	1	-
6	massembar	1	+	12	xuaxulhar	2	-

Tabela 9.4 – Empréstimos verbais em *A Conjura*

Constatamos que o lexema *ngar* não está dicionarizado como verbo nos dicionários acima referidos, mas apenas na sua forma da classe nominal *nga*.

Visto que a lista de candidatos a empréstimos continha alguns empréstimos falsos dentre os quais verbos do Português Europeu, verbos específicos e palavras mal escritas, elidimo-los. Os verbos específicos são formas lexicais usadas pelo autor para despertar a atenção.

Verificamos também que havia ainda unidades léxicas que são da classe nominal e que foram classificadas como sendo da classe verbal. Este é precisamente o caso de unidades léxicas como *n'dalatando*, *dande*, *diquixi*, *kunene-bu* e *muhatu*, as quais foram inseridas como empréstimos da classe nominal. Após a sua supressão, organizámo-los. A Tabela 9.5 exemplifica essa organização.

N.º	Port. Europeu	Específicos	Mal Escritos	Mal Classificados
1	cinquentar	imberber	aconselher	dande
2	desconseguir	juditar	afirmer	diquixi
3	inusitar	preconceituar	cabar	kunene-bu
4	magistrar	charar	concordavar	muhatu
5	malcasar	–	consumirar	n'dalatando
6	oitentar	–	descobrar	–
7	safrar	–	...	–
Total	7	4	64	5

Tabela 9.5 – Empréstimos verbais falsos em *A Conjura*

Processamento de outros textos

Este procedimento de elisão de elementos da lista de candidatos foi repetido com os dados do *Jornal de Angola* e do *Telejornal*. A Tabela 9.6 mostra o resultado deste procedimento, cuja lista completa está no Anexo 1. A linha identificada com “Verbo-” indica a quantificação de empréstimos verbais não dicionarizados.

	A Conjura	Jornal de Angola	Telejornal	Total
Verbo	12	52	2	66
Verbo -	5	26	2	33

Tabela 9.6 – Quantificação dos empréstimos verbais

Concernente ao *Jornal de Angola*, notamos que empréstimos verbais como *axiluar*, *calembar*, *candimbar*, *catocar*, *cudila*, etc. (26 no total) não constam dos dicionários acima referidos, conforme podemos observar no Anexo referido.

Relativamente ao *Telejornal*, repare-se que os empréstimos verbais - *ngueza* e *kwanzar* - extraídos não constam dos dicionários acima referidos.

9.3. Extração de empréstimos de outras classes lexicais

Tendo como modelo o processamento computacional acima descrito, nesta secção realizamos estudos relativos a extração de:

- empréstimos nominais;
- empréstimos adjetivais;
- empréstimos de nomes próprios.

Nesta parte, seguimos os passos descritos na Secção 9.1 para a deteção de candidatos a empréstimos e, por isso, não vamos repetir muitos dos pormenores aqui. Apresentamos só algumas tabelas. A Tabela 9.7 mostram os números de lemas sem repetição extraídos dos nossos textos para as quatro classes lexicais.

	Verbo	Nome	Adjetivo	N. Próprio	Total
A Conjura	1.235	2.960	1.008	530	5.733
Jornal de Angola	2.029	12.295	6.277	26.433	47.034
Telejornal	540	1.855	588	389	3.372
Total	3.804	17.110	7.873	27.352	56.139

Tabela 9.7 – Quantificação dos lemas sem repetição nos textos de PA

Observe-se que a Tabela 9.8 apresenta a quantificação de candidatos a empréstimos por classe lexical nos textos do Português de Angola.²¹

	Verbo	Nome	Adjetivo	N. Próprio	Total
A Conjura	92	234	68	227	621
Jornal de Angola	1.407	2.112	1.908	12.142	17.569
Telejornal	53	91	32	261	437
Total	1.552	2.437	2.008	12.630	18.627

Tabela 9.8 – Quantificação dos candidatos a empréstimo por classe lexical

A seguir, apresentamos os resultados relativos às três classes lexicais que ainda não foram apresentados neste capítulo.

²¹ Esta tabela mostra os números antes da redução descrita na Subsecção 9.1.5.

9.3.1. Extração de empréstimos nominais

Nesta parte, descrevemos a extração de nomes que são empréstimos lexicais que não constam do Léxico do Português Europeu. Atentemos nos vinte elementos selecionados do total de 234 com origem no texto de *A Conjura*:

```
. > c.emp.nome.conj[1:20]

[1] "quibuco"      "imbamba"      "maka"         "muxima"      "macala"
[6] "quilumba"    "quindumba"    "quadrzinha"  "machila"     "dicanza"
[11] "cazuela"     "nga"         "kissangua"   "macololo"    "imbondeiro"
[16] "andua"       "quilamba"     "diquixi"     "mujimbu"     "uadila"
```

Fig. 9.7 – Extração de candidatos a empréstimos nominais

Tendo 234 candidatos a empréstimo, é preciso analisar os casos manualmente. Depois dessa análise, vimos que os dados revelam que, além dos empréstimos nominais, a lista de nomes candidatos a empréstimo inclui ainda alguns falsos empréstimos e outros mal escritos. É preciso elidi-los para termos somente os empréstimos verdadeiros. Extraímos 101 nomes e a sua respetiva frequência e organizamos os dados como se ilustra na Tabela 9.9. O étimo, etimologia e significação dos empréstimos nominais podem ser vistos na lista que está no Anexo 1.

N.º	Nome empréstimo	Freq.	Dic.	N.º	Nome empréstimo	Freq.	Dic.
1	andua	1	+	10	macala	1	+
2	calunga-ya-meia	1	-	11	machila	3	+
3	cazuela	1	-	12	macololo	2	-
4	dicamba-dia-ngalafa	1	-	13	maka	4	+
5	dicanza	1	+	14	monangamba	1	+
6	diquixi	2	+	15	mujimbu	1	+
7	gindungo	1	+	16	quilamba	1	+
8	kissangua	1	+	17	quilumba	1	+
9	libata	1	+	18	quindumba	3	+

Tabela 9.9 – Empréstimos nominais em *A Conjura*

Visto que a lista de candidatos a empréstimos continha alguns empréstimos falsos dentre os quais podemos verificar nomes do Português Europeu e palavras mal escritas, elidimo-los. Após a sua supressão, organizámo-los, tal como se exemplifica na Tabela 9.10.

N.º	Port. Europeu	Mal Escritos
1	adivinhadeira	certezar
2	croniqueiro	colonia
3	empacaceiro	idiom
4	ensaiozito	interer
5	ex-tenente	meder
6	maquinazinha	referêncio
7	propagandismo	senhore
8	quadrazinha	superiore
9
Total	51	82

Tabela 9.10 – Empréstimos nominais falsos em *A Conjura*

Processamento de outros textos

Este procedimento de eliminação de elementos da lista de candidatos foi repetido com os dados do *Jornal de Angola* e *Telejornal*. Na Tabela 9.11, podemos observar o resultado deste procedimento, cuja lista completa consta do Anexo 1.

	A Conjura	Jornal de Angola	Telejornal	Total
Nome	101	128	15	244
Nome -	30	112	–	142

Tabela 9.11 – Quantificação dos empréstimos nominais

9.3.2. Extração de empréstimos adjetivais

Nesta parte, descrevemos a extração de adjetivos que são empréstimos lexicais que não constam do Léxico do Português Europeu. Atentemos nos quinze elementos selecionados.

```
> c.emp.adj.conj[1:15]
[1] "muximado"      "cuamato"      "agindungado"  "inverosímil"  "nzua"
[6] "andembo-ya-tata"  "ngo"          "tchibita"     "sapalalo"     "diculo"
[11] "quindumbo"      "uanga"        "cuamato"      "ambaquense"   "muxito"
```

Fig. 9.8 – Extração de adjetivos candidatos a empréstimos

Tendo 68 candidatos a empréstimo, é preciso analisar os casos manualmente. Depois dessa análise, vimos que os dados revelam que, além dos empréstimos de adjetivos, a lista de adjetivos candidatos a empréstimo inclui ainda alguns falsos empréstimos, adjetivos específicos e outros mal escritos. É preciso elidi-los para termos somente os empréstimos verdadeiros. Extraímos 12 adjetivos e a sua respetiva frequência e organizamos os dados na Tabela 9.12. O étimo, a etimologia e a significação dos empréstimos de adjetivos podem ser vistos na lista que está no Anexo 1.

N.º	Adj. empréstimo	Freq.	Dic.	N.º	Adj. empréstimo	Freq.	Dic.
1	agindungado	1	+	7	ngo	1	+
2	ambaquense	1	-	8	nzua	1	+
3	andembo-ya-tata	2	-	9	quindumbo	1	-
4	cuamato	9	+	10	sapalalo	1	-
5	diculu	1	+	11	tchibita	1	-
6	muxito	1	+	12	uanga	1	+

Tabela 9.12 – Empréstimos adjetivais em *A Conjura*

Visto que a lista de candidatos a empréstimos continha alguns empréstimos falsos dentre os quais podemos verificar adjetivos do Português Europeu, específicos e palavras mal escritas, elidimo-los. Após a sua supressão, organizámo-los e veem-se alguns na Tabela 9.13.

N.º	Port. Europeu	Específicos	Mal Escritos
1	desvirtuoso	acredital	independentista
2	galeanista	dal	indígena
3	gorjeante	desditosar	prósperar
4	maniento	donar	rotino
5	precaríssimo	–	telegrafico
6	...	–	...
Total	21	4	31

Tabela 9.13 – Empréstimos adjetivais falsos em *A Conjura*

Processamento de outros textos

Este procedimento de elisão de elementos da lista de candidatos foi repetido com os dados do *Jornal de Angola* e *Telejornal*. A Tabela 9.14 mostra o resultado deste procedimento, cuja lista completa está no Anexo 1.

	A Conjura	Jornal de Angola	Telejornal	Total
Adjetivo	12	39	1	52
Adjetivo -	5	31	–	36

Tabela 9.14 – Quantificação dos empréstimos adjetivais

9.3.3. Extração de empréstimos de nomes próprios

Nesta parte, descrevemos a extração de nomes próprios que são empréstimos lexicais que não constam do Léxico do Português Europeu. Atentemos nos quinze elementos selecionados.

```
> c.emp.nprop.conj[1:15]
[1] "gambos" "quissongo" "quissama" "humbes" "malange"
[6] "ingombotas" "cassoalala" "katunga" "magombala" "mandume"
[11] "camuquembe" "cambambe" "massangano" "ndongo" "mbriz"
```

Fig. 9.9 – Extração de candidatos a empréstimos

Tendo 227 candidatos a empréstimo, é preciso analisar os casos manualmente. Depois dessa análise, vimos que os dados revelam que, além dos empréstimos de nomes próprios, a lista de nomes próprios candidatos a empréstimo inclui alguns falsos empréstimos,

específicos e outros mal escritos. É preciso elidi-los para termos somente os empréstimos verdadeiros. Extraímos 68 nomes próprios e a sua respetiva frequência e organizamos os dados na Tabela 9.15. O étimo, etimologia e a significação dos empréstimos de nomes próprios podem ser vistos na lista do Anexo 2.

N.º	N. Prop empréstimo	Freq.	Dic.	N.º	N. Prop empréstimo	Freq.	Dic.
1	Amboim	1	+	7	Massangano	2	-
2	Bungo	7	+	8	Pacavira	1	-
3	Cambambe	2	-	9	Pungo Andongo	4	+
4	Humbe	8	+	10	Quissama	3	+
5	Kuanhama	1	+	11	Quissongo	4	+
6	Magombala	7	-	12	Zenza	1	-

Tabela 9.15 – Empréstimos de nomes próprios em *A Conjura*

Visto que a lista de candidatos a empréstimos continha alguns empréstimos falsos dentre os quais podemos verificar nomes próprios do Português Europeu e palavras mal escritas, elidimo-los. Após a sua supressão, organizámo-los, estando alguns na Tabela 9.16.

N.º	Port. Europeu	Mal Escritos
1	Apolinário	alfer
2	Barbosa	ámen
3	Geraldo	independent
4	Vasconcelos	sensi
5	...	-
Total	155	4

Tabela 9.16 – Empréstimos de nomes próprios falsos em *A Conjura*

Processamento de outros textos

Este procedimento de elisão de elementos da lista de candidatos foi repetido com os dados do *Jornal de Angola e Telejornal*. A Tabela 9.17 mostra o resultado deste procedimento, cuja lista completa está, igualmente, no Anexo 2.

	A Conjura	Jornal de Angola	Telejornal	Total
Nome Próprio	68	1.286	68	1.422
Nome Próprio -	20	666	26	712

Tabela 9.17 – Quantificação dos empréstimos de nomes próprios

9.4. Sumário de resultados sobre os empréstimos

Após a extração de empréstimos lexicais, descrita na secção anterior, *Análise do subconjunto de candidatos a empréstimo lexical*, constatamos que o *Jornal de Angola* foi mais produtivo em todas as classes lexicais (52 verbos, 128 nomes, 39 adjetivos e 1286 nomes próprios), conforme vemos na Tabela 9.18, que mostra a quantificação de empréstimos no corpus do Português de Angola. Notamos que 923 empréstimos lexicais não constam dos dicionários acima referidos. A sua significação e etimologia estão nos Anexos 1 e 2 desta tese.

	Verbo	Nome	Adjetivo	N. Próprio	Total
A Conjura	12	101	12	68	193
Jornal de Angola	52	128	39	1.286	1.505
Telejornal	2	15	1	68	86
Total	66	244	52	1.422	1.784

Tabela 9.18 – Quantificação dos empréstimos por classe lexical

9.5. Processamento incremental de textos

O estudo acima descrito focou-se na análise de três conjuntos de textos angolanos. Em cada um desses textos, encontramos alguns empréstimos, que introduzimos no nosso léxico de empréstimos. Este processo podia continuar com a análise de outros textos. Nesta secção, apresentamos uma reflexão sobre a forma como este processo deveria ser organizado para minimizar a quantidade de candidatos a empréstimos que têm que ser analisados manualmente.

A ideia básica consiste em adotar o método incremental de processamento (Gama, 2008). No nosso caso, envolve o processamento de apenas uma partição de dados recém-adicionada (um novo texto da variedade do Português de Angola) a um conjunto de dados (outros textos do Português de Angola), quando os dados existentes já foram processados, em

vez de reprocessar o conjunto de dados completo. Este método de processamento melhora a eficiência, eliminando a sobrecarga do sistema de processamento de dados já processados.

Vejam como podemos adaptar este método ao problema de extração de empréstimos. Suponhamos que o primeiro texto a processar é *A Conjura*. Deste texto, extraímos os candidatos a empréstimo com o auxílio do LPE-ESC. Depois da análise manual, criamos os seguintes subgrupos relativos, por exemplo, a verbos do nosso primeiro texto:

- tokens tipo empréstimos
- tokens do Português Europeu
- tokens específicos
- tokens mal escritos
- tokens mal classificados

Em R, cada um destes grupos foi representado em forma de um vetor. Grupos semelhantes serão criados também em relação aos *nomes*, *adjetivos* e *nomes próprios*.

Quando o segundo texto for processado, extraímos os candidatos a empréstimo com o auxílio do LPE-ESC. Mas, antes de passar estes candidatos para a análise manual, usamos os cinco grupos identificados acima (tokens tipo empréstimos, tokens do Português Europeu, etc.) e todos os elementos que aparecem na lista de novos candidatos são suprimidos. Assim, a lista de candidatos a analisar fica, em princípio, mais reduzida do que a lista original. Depois da análise manual e identificação de novos elementos encontrados, usamos para atualizar os cinco grupos (tokens tipo empréstimos, tokens do Português Europeu, etc.). Este procedimento diminui o número de elementos que é preciso analisar manualmente e assim torna o processo mais eficiente.

9.6. Vantagens do processo automático

Recorde-se que o texto de *A Conjura* inclui 5.072 verbos, como se viu na Tabela 9.1, e analisar esse número manualmente representava um esforço bastante grande. Assim, o método descrito neste capítulo permitiu reduzir bastante este número, para 92 verbos, conforme se vê na Tabela 9.3. Tendo em conta estas duas tabelas, constatamos que a taxa de redução de verbos é $92/5072$, i.e. 1.8%. Na Tabela 9.19, está a taxa de redução de elementos lexicais para as 4 classes lexicais e nos 3 textos do Português de Angola.

	Verbo	Nome	Adjetivo	N. Próprio
A Conjura	1.8%	2.7%	2.9%	9.2%
Jornal de Angola	1.9%	1.1%	4.3%	7.5%
Telejornal	0.2%	0.5%	0.6%	5.2%

Tabela 9.19 – Taxa de redução de elementos lexicais por classe lexical

9.7. Extração de cotextos de empréstimos

Nesta parte, por *cotexto* alude-se às relações que as unidades linguísticas estabelecem dentro do texto de forma a contribuir para a fixação da significação de uma unidade léxica. Assim, o nosso objetivo é extrair o cotexto de empréstimos. Este procedimento viabiliza desencadear processos de desambiguação, estabelecendo relações intertextuais e intratextuais, que constituem um dos parâmetros que configuram e definem a noção de texto (Lopes & Carapinha, 2013).

Enquanto se entende por *intratextualidade* a retoma do que acontece na obra do mesmo autor, podemos considerar que, introduzido na literatura por Kristeva (1969), o termo *intertextualidade* designa certamente o conjunto de relações explícitas ou implícitas de ordem retórico-estilística e de ordem semântica que um determinado texto ou grupo de textos mantém com outros textos (Bakhtin, 1999).

Consideramos que o léxico está profundamente ligado ao conhecimento do mundo. Ao contrário do que acontece com a informação proporcionada pelo cotexto, ou contexto linguístico, a informação fornecida pelo contexto situacional é mais difícil de codificar em unidades passíveis de armazenamento e recuperação, uma vez que não existe nenhuma regra que nos permita identificar onde procurar a porção de conhecimento que pode ser útil para a codificação ou descodificação de um determinado enunciado assim como quanto tempo deveria conservar-se essa informação caso fosse necessário utilizá-la. Portanto, o processo de deteção dos contextos situacionais ultrapassa os objetivos desta tese e pode ser abordado em trabalho futuro. Contudo, pensamos que os nossos exemplos de cotextos são bastante úteis para mostrar como os empréstimos são usados (Hutchins & Somers, 1995; Sanromán, 2001).

Efetuamos o estudo de vários tipos de empréstimos, como nomes, verbos e adjetivos. Na subsecção subsequente procedemos à apresentação do método para detetar o cotexto simplificado de nomes.

Cotextos de empréstimos nominais

Atentemos no processamento de extração do cotexto do empréstimo *maka*. Para a concretização deste passo, começamos pela seleção do empréstimo *maka* no conjunto de dados ilustrado na Fig. 9.9. Neste estudo, usamos somente o texto de *A Conjura*. Em virtude de a unidade léxica apresentar flexões no texto, optamos por selecionar a coluna que indica o lema (*lemma*), elidir o NA do *data frame*. Assim, procuramos o lema igual ao empréstimo.

```
> empr <- emp.nome.conj
> sel.empr <- text.annot.conjura$lemma
> sel.empr.sna <- na.omit(sel.empr)
> not.na <- sel.empr %in% sel.empr.sna
> df.sna <- df.conj[not.na,]
> sel.empr <- df.sna$lemma == empr
```

Notamos que o empréstimo *maka* tem 4 ocorrências no corpus. Assim, verificamos as frases em que este ocorre.

```
> df.empr <- df.sna[sel.empr,c("doc_id","paragraph_id","sentence_id")]
> df.empr
  doc_id paragraph_id sentence_id
3508  doc1           1          125
13780  doc1           1           565
29133  doc1           1          1267
38204  doc1           1          1590
```

Extraímos o *sent_id* relativo à primeira frase do pequeno *data frame* com 4 frases.

```
> sent_id <- df.empr[1,"sentence_id"]
> sel.sent <- df.sna$sentence_id == sent_id
> df.empr1 <- df.sna[sel.sent,]
> sel.col <- c("token_id","token","lemma","upos","head_token_id","dep_rel")
> df.empr1.c <- df.empr1[,sel.col]
> df.empr1.c
  token_id token lemma upos head_token_id dep_rel
3504      67 apenas apenas ADV           69 advmod
3505      68 para para SCONJ           69 mark
3506      69 ouvirem ouvir VERB           62 acl
3507      70 as o DET             71 det
3508      71 makas maka NOUN           69 obj
3509      72 dizerem dizer VERB           69 xcomp
3510      73 de de ADP             75 case
3511      74 sua seu DET             75 det
3512      75 opinião opinião NOUN          72 obl
3513      76 . . PUNCT             15 punct
```

Fig. 9.10 – Análise de uma das frases com o empréstimo *maka*

Visto que objetivamos extrair o cotexto de empréstimo, procedemos à identificação do empréstimo no *data frame*, com o auxílio das instruções:

```
> sel.empr1 <- df.empr1.c$lemma == empr
> token.empr <- df.empr1.c[sel.empr1,"token"]
> token.empr
[1] makas
```

Posteriormente, identificamos o *token* e *head_token_id* na linha onde está o empréstimo.

```
> head_token_id.empr <- df.empr1.c[sel.empr1,"head_token_id"]
> head_token_id.empr
[1] 69
```

Podemos notar que se tem de procurar um elemento com este *head_token_id* na coluna *token*, conforme podemos verificar nas instruções.

```
> sel.head.empr1 <- df.empr1.c$token_id == head_token_id.empr
> head_token.empr <- df.empr1.c[sel.head.empr1,"token"]
> head_token.empr
[1] ouvirem
```

A Fig. 9.11 inclui um ciclo que permite processar todas as frases (4 no total) com o empréstimo *maka*. Esta parte inclui algumas instruções que já explicitamos antes e que assinalamos com cor cinzenta.

```
> df.empr <- df.sna[sel.empr,c("doc_id","paragraph_id","sentence_id")]
> ncasos <- dim(df.empr)[1]
> for (i in 1:ncasos) {
  # Extrair sent_id relativa à 1ª frase no pequeno data frame com 4
  frases:
  sent_id <- df.empr[i,"sentence_id"]
  # Extrair todas as linhas relativas a esta frase para um data
  frame auxiliar (df.empr1):
  sel.sent <- df.sna$sentence_id == sent_id
  df.empr1 <- df.sna[sel.sent,]
  sel.col <- c("token_id","token","lemma","upos","head_token_id",
  "dep_rel")
  df.empr1.c <- df.empr1[, sel.col]
  # Encontrar o empréstimo neste pequeno data frame:
  sel.empr1 <- df.empr1.c$lemma == empr
  # Identificar o token e head_token_id na linha onde está o
  empréstimo:
  token.empr <- df.empr1.c[sel.empr1, "token"]
  head_token_id.empr <- df.empr1.c[sel.empr1, "head_token_id"]
  # Procurar um elemento com esse "head_token_id" na coluna "token":
  sel.head.empr1 <- df.empr1.c$token_id == head_token_id.empr
  head_token.empr <- df.empr1.c[sel.head.empr1, "token"]
  # Procurar o token cujo "head_token_id" é igual ao "token_id" do
  empréstimo
  sel.token <- df.empr1.c[sel.empr1, "token_id"]
  sel.token.dep <- df.empr1.c$ head_token_id == sel.token
  token.dep <- df.empr1.c[sel.token.dep, "token"]
  # Imprimir os elementos:
  print(c(as.character(head_token.empr),as.character(token.dep),as.
  character(token.empr)))
}
[1] "ouvirem"      "as"           "makas"
[1] "arranjar"     "uma"          "maka"
[1] "reconstruir"  "alguma"       "passada"      "maka"
[1] "centro"       "de"           "as"           "makas"
```

Fig. 9.11 – Extração de cotexto simplificado do empréstimo *maka*

Uma amostra deste procedimento aplicado à extração de cotexto de empréstimos nominais pode ser observado na Tabela 9.20. A frequência é o critério adotado para a seleção dos cotextos de empréstimos nominais.

Nº	Empréstimo	Frase	Cotexto	Equivalência
1	maka	125	ouvirem as makas	ouvirem os problemas
		565	arranjar uma maka	arranjar um problema
		1267	reconstruir alguma passada maka	reconstruir um litígio
		1590	centro das makas	pessoa problemática
2	buxila	1381	paixão por uma buxila	paixão pela filha de escravos
		1428	tímido e a buxila	tímido e a filha de escravo
		1437	cólera contra a inocente buxila	cólera contra a inocente filha de escravo
		1828	embarque da buxila	embarque da filha de escravo
3	quindumba	373	olhos ao cheiroso corpo quindumba	olhos ao cheiroso corpo de poupa
		463	corpos e as quindumbas	corpos e as elevações de cabelo
		1201	perfumada quindumba	cabelo perfumado
4	machila	167	subido em sua machila	subindo na cadeira régia
		239	carregadores de machila	carregadores da cadeira régia
		244	carregadores de trabalhadores machila	carregadores de trabalhadores da cadeira régia.
5	candengues	293	virava os candengues	virava as crianças
		1853	pedrada com os outros candengues	pedrada com as outras crianças
6	macololo	707	território dos macololos	território de escória
		709	países dos macololos	países de escória
7	muxima	402	briguento, mas de boa muxima	briguento, mas de bom caráter
		776	inquieta muxima	inquieto coração
8	nzua	110	embriagante nzua	embriagante sumo de múcua
		463	preparava a embriagante nzua	preparava o embriagante sumo de múcua
9	matabicho	1834	revirou a matabicho	revirou o pequeno-almoço
10	ngaieta	288	tocador de ngaieta	tocador de harmónica
11	hoxa	275	doente de hoxa	doente de tripanossomíase
12	quilumba	1050	tempos de quilumba	tempos de donzela

Tabela 9.20 – Cotextos de empréstimos nominais

Os cotextos de outros empréstimos nominais estão no Anexo 1 desta dissertação. Esclareça-se que, no cotexto, o empréstimo nominal *muxima* se refere a *coração* e *índole*. Entretanto, podemos verificar que, após a instauração do santuário católico, ocorre uma

extensão semântica desta unidade léxica para topónimo e teónimo, uma vez que equivale respetivamente ao *templo de Fátima* em Portugal e a *Maria*.

Além disso, constatamos que, embora o cotexto indique *embriagante nzua*, esta só o é caso se utilize nos dias subsequentes ou se adicione alguma substância química para alterar as suas propriedades.

Cotextos de empréstimos adjetivais

Efetuamos o mesmo procedimento com os empréstimos adjetivais. Na Tabela 9.21, podemos observar alguns cotextos. Há mais cotextos de empréstimos adjetivais no Anexo 1. Notamos que o empréstimo adjetival *cuamato* também pode ser nominal, uma vez que é um gentílico²². No entanto, inserimos somente o cotexto como adjetivo.

Nº	Empréstimo	Frase	Cotexto	Equivalência
1	muhato	107	nga muhatu	senhora casada à moda tradicional
		1191	nga muhatu	senhora casada à moda tradicional
2	cafuso	115	condenado cafuso	condenado filho de mestiço e preta
		401	inimigo cafuso	inimigo filho de mestiço e preta
3	agindungado	118	rumores bem saborosos agindungados	rumores maliciosos sobre uma celebridade
4	ambaquense	1251	mulata ambaquense	mulata de Ambaca
5	cuamato	1721	guerreiros cuamatos	guerreiros da região de cuamatos
6	diculu	1050	cazuela com alguém diculu	ser logorreico com algum senil
7	sapalalo	580	gasto sapalalo	gasto sabático

Tabela 9.21 – Cotextos de empréstimos adjetivais

Cotextos de empréstimos verbais

Um procedimento semelhante foi aplicado para a extração de verbos empréstimos. Repare-se que a impressão dos resultados depende da transitividade verbal. A extração de cotextos simples envolve o predicador verbal e um dos seus argumentos.

Ao contrário dos cotextos simples, consideramos que os cotextos alargados envolvem o verbo e mais de um argumento do predicador verbal. Entretanto, cingimo-nos aos cotextos simples de empréstimos verbais. Observemos o *data frame* da Fig. 9.12.

²² Para ilustrar, podemos observar a ocorrência de *cuamato* como empréstimo nominal nos seguintes cotextos: *grito dos cuamatos*, *gente do cuamato*, *malditos cuamatos*.

```

> sent_id <- df.empr[1,"sentence_id"]
> sel.sent <- df.sna$sentence_id == sent_id
> df.empr1 <- df.sna[sel.sent,]
> sel.col <- c("token_id","token","lemma","upos","head_token_id","dep_rel")
> df.empr1.verb <- df.empr1[, sel.col]
> df.empr1.verb
  token_id  token  lemma  upos  head_token_id  dep_rel
44166     1   Como   como  SCONJ           5   mark
44167     2     só     só    ADV           3  advmod
44168     3   vavó   vavó   ADJ           4   amod
44169     4   Uandi  Uandi  PROPN          5  nsubj
44170     5 coxilava coxilar  VERB           0   root
44171     6     .     .    PUNCT           5  punct

```

Fig. 9.12 – Extração de frase com empréstimo verbal

Assim, podemos extrair os empréstimos verbais e os seus respetivos argumentos, com o auxílio das instruções

```

> token.empr <- df.empr1.verb[sel.empr1,"token"]
> token.empr
[1] "coxilava"
> head_token_id.empr <- df.empr1.verb[sel.empr1, "head_token_id"]
> token_id.empr <- df.empr1.c[sel.empr1, "token_id"]
> sel.head.empr1 <- df.empr1.c$head_token_id == token_id.empr
> head_token.empr <- df.empr1.c[sel.head.empr1, "token"]
> head_token.empr
[1] "Como" "Uandi"

```

Na Fig. 9.12, podemos notar que o empréstimo verbal *coxilar* tem um predicado unário, ou seja, não seleciona argumento interno, é intransitivo. Portanto, a impressão dos resultados requer que primeiro esteja o *head_token.empr* e só depois o *token.empr*, conforme podemos observar na Fig. 9.13.

```

> ncasos <- dim(df.empr)[1]
> for (i in 1:ncasos) {
  # Extrair sent_id relativa a 1ª frase no pequeno data frame com 4
  frases:
  sent_id <- df.empr[i,"sentence_id"]
  # Extrair todas as linhas relativas a esta frase para um data
  frame auxiliar (df.empr1):
  sel.sent <- df.sna$sentence_id == sent_id
  df.empr1 <- df.sna[sel.sent,]
  sel.col <- c("token_id","token","lemma","upos","head_token_id",
"dep_rel")
  df.empr1.c <- df.empr1[, sel.col]
  # Encontrar o empréstimo neste pequeno data frame:
  sel.empr1 <- df.empr1.c$lemma == empr
  # Identificar o token e head_token_id na linha onde está o
  empréstimo:
  token.empr <- df.empr1.c[sel.empr1, "token"]
  token_id.empr <- df.empr1.c[sel.empr1, "token_id"]
  sel.head.empr1 <- df.empr1.verb$head_token_id == token_id.empr
  head_token.empr <- df.empr1.verb[sel.head.empr1, "token"]
  print(c(as.character(head_token.empr),as.character(token.empr)))
}
[1] "E"      "estás"   "coxilando"
[1] "Como"   "Uandi"   "coxilava"

```

Fig. 9.13 – Extração de cotextos de empréstimos verbais

Note-se que, para um melhor arranjo dos resultados, tratando-se de empréstimos verbais transitivos como *kuatar*, é necessário que primeiro esteja o *token.empr* e só depois o *head_token.empr*. A Tabela 9.22 apresenta uma amostra de cotextos de empréstimos verbais e a sua respetiva equivalência. Outros cotextos de empréstimos verbais podem ser consultados no Anexo 1.

Nº	Empréstimo	Frase	Cotexto	Equivalência
1	coxilar	1805	e estás coxilando	e estás a adormecer
		1806	como Uandi coxilava	como Uandi adormecia
2	xuaxualhar	70	xuaxualhar de asas	farfalhar de asas
		1244	xuaxualhar das folhas	farfalhar das folhas
3	builar	342	buila mu nvunda	soltar urros na luta
4	baçular	305	baçular vertiginosamente raciocínios	aplicar com perícia os raciocínios
5	kuatar	131	kuata os pretos	agarra os pretos
6	massembar	1099	vinham massembando	vinham a danças massemba
7	sunguilar	393	sunguilar no pavilhão	pernoitar no pavilhão

Tabela 9.22 – Cotextos de empréstimos verbais

Em virtude de *A Conjura* ser um texto literário, constatamos que, no cotexto, o empréstimo verbal *xuaxualhar* foi usado no seu sentido de *farfalhar*, embora o mesmo possa remeter-nos para a noção de *galantear*, *seduzir*, *paquerar* e, em alguns casos, *flertar*.

Algo similar ocorre em relação ao empréstimo verbal *baçular*. Recorde-se que se trata de um verbo denominal e refere-se à ideia de aplicar *baçula* a alguém, surrar alguém numa luta. Portanto, depreendemos que o cotexto nos remete para uma situação de logomaquia. Entretanto, entendemos que os nossos exemplos de cotextos de empréstimos nominais, adjetivais e verbais são bastante úteis para elucidar como os empréstimos são usados.

9.8. Síntese do capítulo

Neste capítulo, descrevemos os procedimentos computacionais para a extração de empréstimos lexicais provenientes de Angola encontrados em *A Conjura*, *Jornal de Angola* e *Telejornal de Angola*. Procedemos ao estudo comparativo e à extração de empréstimos por classe lexical com o foco em quatro classes lexicais (verbos, nomes, adjetivos e nomes próprios). Comparamos os resultados obtidos através do método computacional proposto. Apresentamos a organização dos empréstimos por categorização lexical e a sua sumarização. Com efeito, estes dimanam de várias línguas de Angola. Descrevemos os critérios usados para

a determinação do carácter neológico das unidades léxicas. Explicitamos o processamento incremental de texto, que melhora significativamente a eficiência elidindo a sobrecarga do sistema computacional, e descrevemos a extração de cotextos de empréstimos nominais, verbais e adjetivais.

Capítulo 10 – Organização do Subconjunto de Empréstimos Lexicais

Este capítulo tem como base o Capítulo 2, Secção 2.2, *Línguas e surgimento do português em Angola*, o Capítulo 3, Secção 3.3, *Transformação conceptual da dimensão neológica* e o capítulo 4, *Estrutura do léxico e processos de formação de palavras*. Assim, temos decerto quatro objetivos fundamentais. Tendo em vista analisar, organizar e visualizar a informação concernente aos empréstimos lexicais, em primeiro lugar, descrevemos o critério adotado para a determinação do carácter neológico das formas lexicais e procedemos à quantificação do subconjunto de empréstimos por categorização lexical. Posteriormente, explicitamos a determinação da etimologia dos empréstimos lexicais, os quais provêm de línguas de Angola. Apresentamos o seu étimo, etimologia e significação nas listas em Anexo. Em seguida, estruturamos os empréstimos em campos lexicais. Por fim, descrevemos os processos de formação dos empréstimos lexicais. Além disso, apresentamos a proposta de tratamento lexicográfico dos empréstimos.

10.1. Categorização lexical e etimologia dos empréstimos

O critério adotado para a determinação do carácter neológico das formas lexicais foi o lexicográfico. Para a determinação do étimo, etimologia e significação, recorremos ao *Dicionário da Língua Portuguesa Contemporânea* (Academia das Ciências de Lisboa & Fundação Calouste Gulbenkian, 2001) e ao *Dicionário de Regionalismos Angolanos* (Ribas, 2014).

Verificamos a etimologia dos empréstimos e estruturamo-los em tabelas que indicam o étimo, a etimologia e a sua significação, e que constam dos Anexos 1 e 2 deste trabalho. Para ilustrar a organização dos empréstimos, na Tabela 10.1, apresentamos um extrato da lista de empréstimos lexicais que podem ser vistos nos Anexos referidos e que constituem as listas dos empréstimos lexicais extraídos de *A Conjura (2008)*, *Jornal de Angola (2019-2020)* e *Telejornal de Angola (2020)* com o auxílio de métodos computacionais. Nesta Tabela 10.1, apresentamos o étimo, a etimologia, a significação e uma pequena nota, quando necessário, em que se faz um comentário sobre o empréstimo lexical extraído, cuja organização está em ordem alfabética para facilitar o consulente. As unidades léxicas assinaladas com asterisco (*) não constam dos dicionários acima referidos. O processamento incremental efetuado, que diminui a sobrecarga computacional, permite constatar que, quanto aos empréstimos não

dicionarizados, podemos realçar que são um contributo evidente sobretudo para os dicionários referidos e para o Léxico do Português Europeu e outros dicionários.

Empréstimo	Etimologia	Significação	Doc	Frase	Cotexto	Equivalência
buxila	kimbundu	<i>n. m. e f.</i> 1. Filho ou filha de escravos. 2. Filho ou filha de escrava com homem livre, mas nascida na casa em que serve. 3. Criolo.	1	1381	paixão por uma buxila	paixão por filha de escravos
				1428	tímido e a buxila	tímido e a filha de escravo
				1437	cólera contra a inocente buxila	cólera contra a inocente filha de escravo
cambuta	kimbundu	1. adj. unif. Diz-se de pessoa de estatura baixa ou inferior à média. 2. n. m. e f. Pessoa de estatura baixa ou inferior à média. 3. Baixinho.	2	233	ajudante cambuta	ajudante de estatura baixa
				227	esse cambuta	esse baixinho
machila	kimbundu	<i>n. f.</i> 1. Palanquim. 2. Cadeira com tampo e cortinas, a qual suspensa de um bordão de bambu, era transportada aos ombros de dois homens.	1	167	subido em sua machila	subindo na cadeira régia
				239	carregadores de machila	carregadores da cadeira régia
				244	carregadores de trabalhadores machila	carregadores de trabalhadores da cadeira régia .
maka	kimbundu	1. n. f. Conversa; assunto; novidade; discórdia; litígio; conflito; algazarra. 2. Bras. Incolumanca. 3. adj. Problemático.	1	125	ouvirem as makas	ouvirem os problemas
				565	arranjar uma maka	arranjar um problema
				1267	reconstruir alguma maka	reconstruir um litígio passado
				1590	centro das makas	peessoa problemática
mbeji*	kimbundu	<i>n. m.</i> 1. Satélite da terra. 2. Lua. 3. Espaço durante o qual a lua faz a sua revolução em torno da terra. 4. Tempo compreendido entre os dois novilúnios.	2	149	maior mbeji	maior satélite da terra
				187	favorável mbeji	favorável lua
muxima*	kimbundu	<i>n. m. e f.</i> 1. Pessoa de bom carácter. 2. Benfeitor. <i>n. m.</i> 3. Lisonjeio. 4. Fig. Fazer algo com sinceridade. 5.v. Cativar, lisonjear.	1	402	briguento, mas de boa muxima	briguento, mas de bom carácter
				776	inquieta muxima	inquieto coração
quindumba	kimbundu	<i>n. f.</i> 1. Poupa. 2. Elevação de cabelo ou de pena.	1	373	olho ao cheiroso corpo quindumba	atenção ao corpo de poupa
				463	corpos e as quindumbas	corpos e as elevações de cabelo
				1201	perfumada quindumba	cabelo perfumado

Tabela 10.1 – Extrato da lista de empréstimos dimanantes de Angola

Os empréstimos lexicais exemplificados na Tabela 10.1 dimanam de línguas de Angola referidas no Capítulo 2, e as regiões em que estas línguas são faladas podem ser vistas na Fig. 10.1.

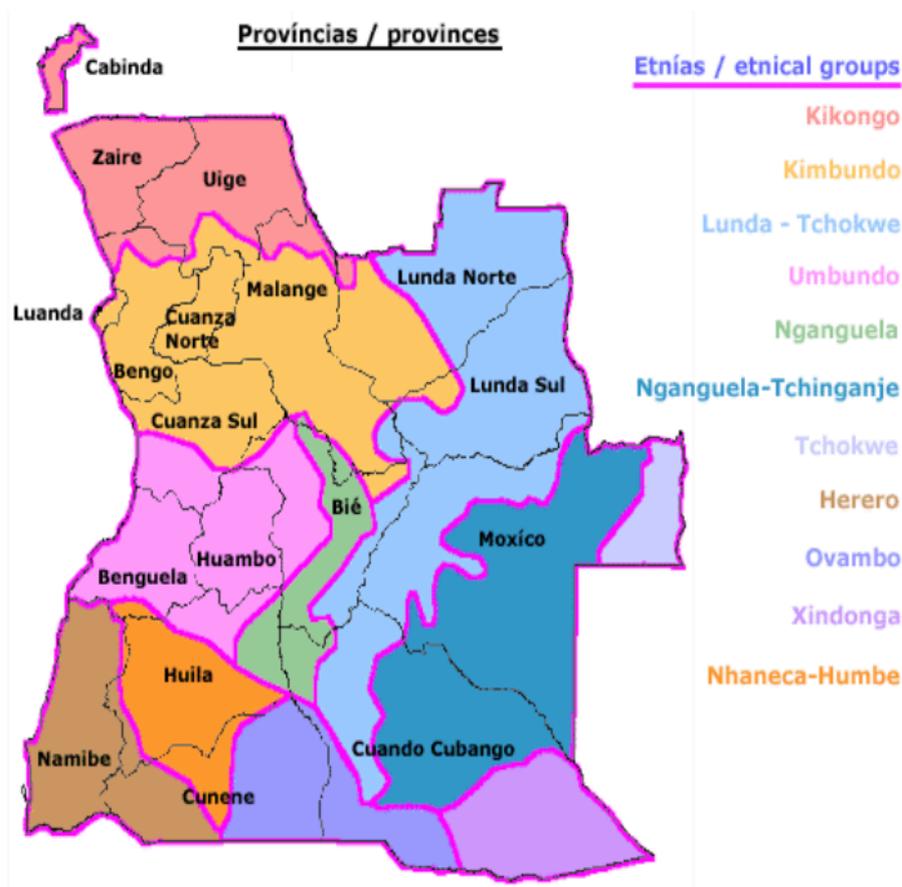


Fig. 10.1 – Algumas línguas de Angola (Edmundo, 2020)

Os dados revelam a deteção de 1.784 empréstimos lexicais. Relativamente à sua categoria lexical, verificamos que há uma maior produtividade a nível da classe dos nomes próprios (1.422), sendo o restante da classe verbal (66), nominal (244) e adjetival (52). A classe dos nomes próprios é a mais produtiva, pois isso acontece provavelmente pela necessidade que sempre se teve de denominar (Trask, 1993).

Quanto à etimologia, a contabilização pelo número de entrada revela que a maioria promana do *kimbundu*, com um total de 872 entradas. Há 1.723 entradas dimanantes de línguas *bantu* e não *bantu*, conforme se verifica na Tabela 10.2, pois algumas palavras pertencem a mais de uma classe lexical.

Uma outra constatação que resulta da análise dos dados é que grande parte dos empréstimos estão aportuguesados (*Njinga* > *jinga*; *Ngunza* > *Gunza*) e há mudança de nomes próprios (*Nguxi* > *Augusto*). Verifica-se também a adaptação de palavras ao sistema gráfico português como *quiocos*, *quissama*, *quissanguela*, *quissongo*, *quissondes* e *quitandeira*. Recorde-se que algumas unidades lexicais não promanam de línguas *bantu* e não *bantu*

(*facar, macala, matabicho, caimaneros, ANAZANGA, CANFEU*), mas constituem inovação lexical típica da variedade do Português de Angola. Estas são denominadas português em Angola. Outras dimanam de línguas de Angola, no entanto os lexicógrafos não conseguiram determinar precisamente a etimologia. Estas são denominadas termo regional (Ribas, 2014). A Tabela 10.2 apresenta a quantificação dos empréstimos lexicais extraídos dos textos que processamos (*A Conjura (2008), Jornal de Angola (2019-2020) e Telejornal de Angola (2020)*), organizados por etimologia.

		A Conjura	J. de Angola	Telejornal	Total	Língua
Etimologia	Kimbundu	146	694	32	872	Bantu
	Umbundu	12	354	14	380	Bantu
	Kikongo	3	117	6	126	Bantu
	Cokwe	3	115	4	122	Bantu
	Ngangela	3	31	–	34	Bantu
	Kwanyama	4	18	2	24	Bantu
	Nyaneka	4	12	2	18	Bantu
	Herero	1	3	–	4	Bantu
	Koisan	–	1	–	1	Não Bantu
	Ngoyo	–	1	–	1	Bantu
	Lingala	–	1	–	1	Bantu
	Português em Angola	7	109	18	134	Latina
	Termo Regional	4	2	–	6	Bantu
Total	187	1.458	78	1.723	3	

Tabela 10.2 – Quantificação das entradas dos empréstimos por etimologia

Se analisarmos os dados da Tabela 10.2, concluímos que, quanto à quantificação dos valores dos empréstimos por etimologia, o *kimbundu* representa a maioria dos empréstimos. Estes dados podem ser vistos no gráfico 10.1. Verificamos que o lexema *cambundo* tem o mesmo étimo, dupla etimologia (*kimbundu* e *kikongo*) e dupla significação, conforme se explicita no Anexo 1 desta dissertação.

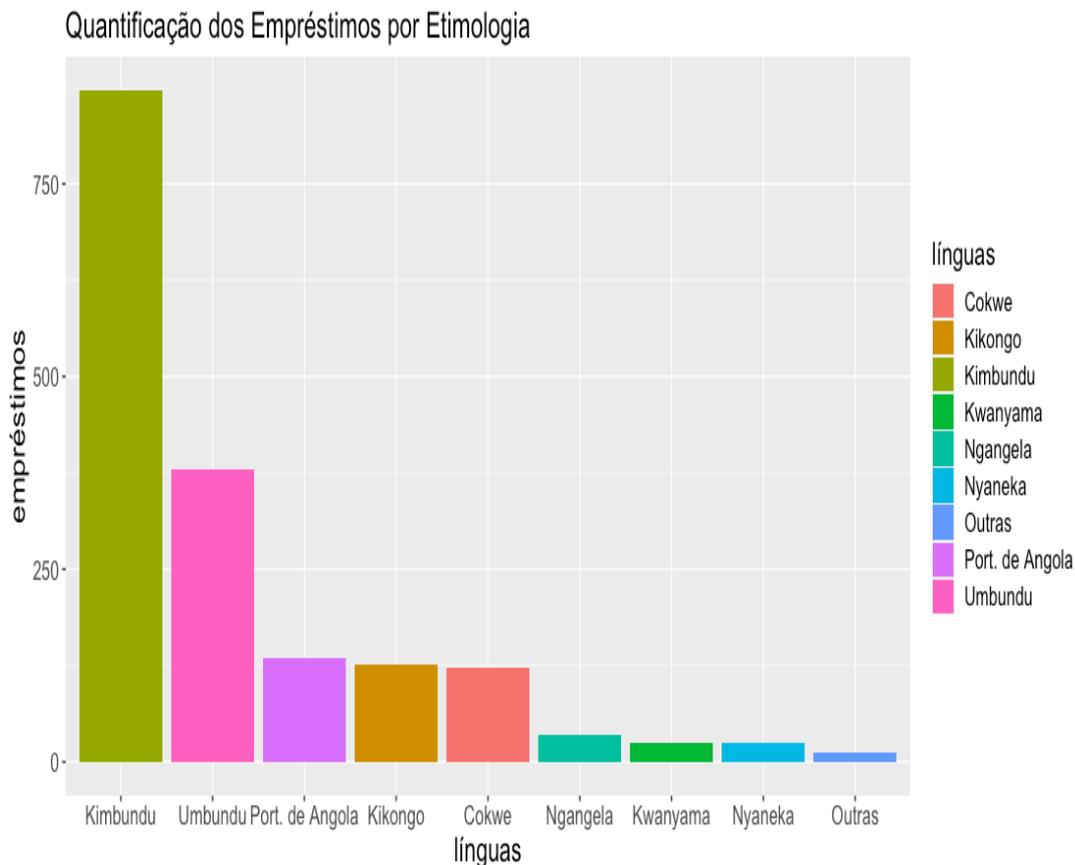


Gráfico 10.1 – Quantificação dos empréstimos por etimologia

10.2. Estruturação dos empréstimos em campos lexicais

Partindo do pressuposto de que o léxico é o conjunto virtual de lexemas que pertencem a um dado sistema linguístico, os lexemas, quando atualizados no discurso, designam-se por vocábulos. O conjunto de vocábulos, as unidades do discurso, constitui o vocabulário. Consideramos que a palavra é detentora de significado lexical, conteúdo referencial ou função gramatical, possui classe lexical em virtude das suas propriedades formais e de um comportamento morfossintático estável e objetivamente observável; corresponde a um nó terminal de uma árvore sintática; desempenha um papel definido na estrutura sintática; tem autonomia mínima dentro da frase; é o resultado ou o domínio da aplicação de processos morfológicos e fonológicos; respeita as restrições fonotáticas da língua como regras de combinação de segmentos em sílabas (Veloso, 2016, pp. 49-51).

Assim, uma unidade lexical pode implicar a existência de um núcleo semântico. Os elementos que conduzem à constituição de campo lexical e campo semântico são, por certo, o arquilexema, lexema, sema e dimensão. O arquilexema é uma unidade que corresponde ao

conteúdo total de um campo lexical. O lexema é a unidade de conteúdo expressa no sistema da língua e que ocupa uma parte do conteúdo do campo lexical. Os semas são unidades menores constituídas por traços distintivos de conteúdo e constitutivas do lexema. A dimensão é o critério implicado por uma oposição, o critério que estabelece o ponto de vista de uma determinada oposição (Vilela, 1979, p. 61).

Entende-se por campo semântico, por um lado, o estudo do vocabulário nos seus vários empregos, tentando uma classificação sistemática. Por outro lado, a delimitação do sentido de um vocábulo num discurso constituído, como por exemplo, num texto literário, pela tentativa da reconstituição do contexto imediato. O campo semântico pode ser organizado, partindo de uma perspectiva semasiológica.

O campo lexical compreende um conceito referente ou não a um domínio extralinguístico. Vilela (1979, p. 60) sublinha que o campo lexical é, na perspectiva estrutural, um paradigma lexical formado pela articulação e distribuição de um contínuo de conteúdo lexical por diversas unidades existentes na língua (palavras) e que se opõem entre si por meio de simples traços de conteúdo. Deduzimos, assim, que a representação de um campo lexical pode ser feita partindo de uma perspectiva onomasiológica ou semasiológica.

Parece evidente que, sob o ponto de vista lexical, os neologismos por empréstimo lexical apresentados remetem-nos para campos como *comércio*, *alimentação*, *habitação*, *status social*, *administração civil e música*. Com efeito, os dados permitem-nos afirmar que, no cômputo geral, os empréstimos encontrados pertencem a campos como:

- Administração civil (*Sekulo, Soba, libata, Andembo, N'gunza, Ngar*);
- Alimentação (*matabicho, funji*);
- Bebida (*nzua, kissangua, Uala*);
- Temperatura (*quindumbo*);
- Botânica (*sobos, gindungo, tacula muxito, imbondeiro*);
- Comércio (*Sambizanga*);
- Belística (*canhangulo*);
- Dança (*massemar*);
- Medicina (*quissongo, molumbar, maculo*);
- Música (*gaieta, dicanza*);
- Necrotério (*Sant'Anna*);
- Indumentária (*bofeta*);
- *Status social* (*monangamba, muene*);
- Obscurantismo (*quianda, quilamba, quimbandice, maquixi, chinguilamento*);
- Topónimos (*Libolo, Ingombotas, Amboim, Cazengo*);

- Antropónimo (*Ambriz, Gambos, Humbes, Kuanhama, Mbunda, Quissama*);
- Zoónimos (*maracachão, gungo, cassacame, andua, Cambambe, Mbiji, Pungo*).

10.3. Processos de formação dos empréstimos

Esta secção, que tem como base o Capítulo 4, tem como objetivo indicar qual é o processo de formação do subconjunto de empréstimos. Com base na observação dos dados, concluímos que os vários processos referidos no Capítulo 4 são ativados: uns são formados por amálgama, alomorfia, derivação prefixal, sufixal e outros são formados por composição morfológica, morfossintática e sintagmática (Spencer, 2001). A fim de analisar e organizar a informação relativa aos processos de formação dos empréstimos lexicais, nesta secção procedemos à análise dos dados no sentido de determinar os empréstimos formados por derivação, composição e formação não concatenativa.

10.3.1. Empréstimos formados por derivação

Esta subsecção tem como base o Capítulo 4, Secção 4.5, *Processos de derivação em português*. Convém sublinhar que, relativamente aos constituintes morfológicos do português, nem todos os morfemas são palavras, contudo todas as palavras são morfemas. Tal dedução advém precisamente do critério de fixidez de posição dos morfemas, os quais, tendo carga semântico-funcional, podem ser presos ou autónomos. É evidente que o morfema preso não pode mudar de posição e não pode, igualmente, ocorrer de forma isolada e alguns têm carga funcional, mas não têm carga semântica. No entanto, o morfema autónomo pode ocorrer por si mesmo como palavra e remeter-nos para o mundo extralinguístico (Rodrigues, 2013).

Se compreendermos a derivação como, além da criação de novas unidades léxicas pelo acréscimo de afixos a unidades pré-existentes, subconjunto completo de estágios que ligam a estrutura profunda de uma sentença à sua estrutura superficial, parece evidente que a maioria dos empréstimos é formado por derivação sufixal (Rio-Torto, 2013). Os dados revelam que alguns empréstimos lexicais encontrados formados por derivação sufixal têm a base dimanante de língua de Angola e o sufixo isocategorial é do português, conforme podemos notar nos seguintes exemplos.

(10.1) [[QUIMBAND]_{RADICAL SIMPLES} [ic]_{SUFIXO}]_{RADICAL COMPLEXO} + e IT]_{TN COMPLEXO},
quimbandice.

(10.2) [[QUIMBAND]_{RADICAL SIMPLES}] + a IT]_{TN SIMPLES}, *quimbanda*.

Notamos que os exemplos 10.1 e 10.2 têm o mesmo radical simples. Visto que há dois temas associados ao mesmo radical, os quais são selecionados em função da concordância com um nome num enunciado, verificamos que estas duas formas lexicais configuram um caso de supletivismo lexical (Mota, 2013a, p. 2812). Atentemos nos seguintes exemplos e notemos o tipo de morfema derivacional.

(10.3) [COXIL]_{RN} e COXIL_{RV a]}_{VT r]}_{TMA}, verbo, *coxilo* e *coxilar*.

(10.4) [ca_{PREF} [LUAND]_{RN SIMPLES}]_{RN COMPLEXO}] + a_{IT}_{TN COMPLEXO}, *caluanda*.

Constatamos que o exemplo (10.3) revela um caso de derivação denominal, cujo sufixo é heterocategorial (*coxilo* > *coxilar*). No exemplo (10.4), notamos que o empréstimo é derivado por prefixação e constatamos que a base e o prefixo isocategorial promanam de língua de Angola. Organizamos os empréstimos formados por derivação prefixal, sufixal e o morfema derivacional nominalizador, adjetivalizador e verbalizador, conforme podemos notar alguns na Tabela 10.3 (Aronoff & Anshen, 2001; Beard, 2001).

N.º	Tipo de derivação	Morfema derivacional	Tipo de afixo ²³	Exemplos
1	Prefixal	<i>a-, ca-</i>	isocategorial	caluanda, Ambriz, Amboim
		<i>-ar, -ado, -</i>		coxilar, massebar, muximar, molumbar, baçular, nzuar,
2	Sufixal	<i>ence, -inha, -ice, eiro</i>	heterocategorial	agindungado, ambaquence, benguelinha, quimbandice, zungueiro

Tabela 10.3 – Empréstimos formados por derivação

Consideramos o lexema como unidade do léxico abstrata, não flexionada e pertencente tipicamente a uma das classes lexicais abertas (Matthews, 1991, p. 26). Sendo um signo, o lexema possui significação própria, assim como forma morfológica e fonológica própria e tem associadas outras propriedades gramaticais na sua matriz lexical. Pelo exposto, na Tabela 10.3, podemos depreender que os dados revelam que os empréstimos formados por derivação são denominais, dentre os quais verbos por conversão de radical. Quanto ao constituinte temático, são da primeira conjugação, conforme se nota em *massebar, coxilar, muximar, molumbar, baçular*.

²³ Cf. (Rio-Torto, 1993).

(10.5a) MASSEMB]_{RV} a]_{VT} r]_{TMA}, *massemar*

(10.5b) MUXIM]_{RV} a]_{VT} r]_{TMA}, *muximar*

(10.5c) MOLUMB]_{RV} a]_{VT} r]_{TMA}, *molumbar*

(10.5d) BAÇUL]_{RV} a]_{VT} r]_{TMA}, *baçular*

Em português, é o radical verbal que sofre conversão e neste processo de adaptação do *kimbundu* é a base que inclui a vogal -a, maior que o radical (Miguel, 2019). Podemos constatar que, quanto ao padrão temático, os empréstimos nominais e adjetivais são de classe temática -a, de classe temática -e, e de classe temática -o, conforme podemos observar em *benguelinha*, *caluanda*, *ambaquence*, *quimbandice*, *agindungado* e *zungueiro*.

(10.6a) BENGUEL]_{RN} inh]_{SUF DERIV AVALIATIVO}]_{RN DER} a]_{IT}]_{TN}, *benguelinha*

(10.6b) AMBAQU]_{RN} enc]_{SUF DERIV}]_{RN DER} e]_{IT}]_{TN}, *ambaquence*

(10.6c) a]_{PREF} GINDUNG]_{RN} ad]_{SUF DERIV}]_{RA DER} o]_{IT}]_{TN}, *agindungado*

Recorde-se que os lexemas *agindungado* e *zungueiro* nos permitem inferir que um mesmo lexema pode apresentar supletivismo temático, sendo que as duas variantes temáticas respondem à propriedade morfosintática da concordância em género requerida em contexto sintático. Além disso, parece que a vogal temática é mais interna do que o índice temático, pois pode preceder o morfema de tempo-modo-aspeto, enquanto o índice temático apenas precede o marcador de número. Um tema nominal ou tema adjetival sofre apagamento do expoente temático no processo de derivação e um tema verbal não o sofre. Encontramos empréstimos lexicais derivados de classe temática -Ø~e e aтемático. Notemos os seguintes exemplos em que se apresentam estes empréstimos lexicais derivados.

(10.7) a]_{PREF} MBRIZ]_{RN} /Ø/]_{IT}]_{TN}, *Ambriz*

(10.8) a]_{PREF} MBOIM]_{RN} -]_{IT}]_{TN}, *Amboim*

Entretanto, a predileção de adotar designações diferentes para estes dois expoentes temáticos – índice temático e vogal temática – contribui para a distinção entre classes lexicais, sem nos olvidarmos de que existem nomes e adjetivos aтемáticos, mas não verbos (Mota, 2013b, pp. 2874-2875).

10.3.2. Empréstimos formados por composição

Esta subsecção tem por base os conceitos definidos no Capítulo 4, Secção 4.8, *Processos de composição em português*. Recorde-se que nos cingimos à tipologia de compostos de Caldas (2016), Rio-Torto e Ribeiro (2016). Tendo em conta a sua formação morfossemântica, detetamos, como podemos notar na Tabela 10.4, dois compostos morfológicos formados por um radical não autónomo grego e dois radicais autónomos dimanantes de língua de Angola, conforme vemos no exemplo (10.9).

(10.9a) HIDR]_{RN1} [o]_{VL} LUACHIM]_{RN2}]_{RN} COMPOSTO o]_{IT}]_{TN} COMPOSTO, *hidroluachimo*

(10.9b) HIDR]_{RN1} [o]_{VL} CHICAP]_{RN2}]_{RN} COMPOSTO a]_{IT}]_{TN} COMPOSTO, *hydrochicapa*

Os empréstimos lexicais formados por composição morfológica ilustrados em (10.9a e 10.9b) são da classe temática *-o* e *-a*, respetivamente (Villalva, 2013). Há dois compostos morfossintáticos formados por duas palavras próprias da variedade do Português de Angola, dentre os quais um é da classe temática *-o* e apresenta a estrutura **verbo-nome** (10.10a); um atemático e apresenta a estrutura **nome-nome** (10.10b); e um verbo que, pela classe temática, pertence à segunda conjugação e apresenta a estrutura **advérbio-verbo** (10.10c).

(10.10a) MATABICH]_{RN} COMPOSTO] o]_{IT}]_{TN}, *matabicho*

(10.10b) DICULUNDUNDU]_{RN} COMPOSTO] –]_{IT}]_{TN}, *diculundundu*

(10.10c) MALEMBELEM]_{RV} COMPOSTO e]_{VT} r]_{TMA}, *malembelember*

Há quatro compostos sintagmáticos, cujo primeiro elemento é formado por duas palavras de línguas de Angola e uma do português. Os empréstimos formados por composição sintagmática são formados por palavras provenientes de línguas de Angola e apresentam a estrutura **nome-preposição-nome** (*dicamba-dia-ngalafa*, *calunga-ya-meia*, *suco-yo-bába*) e **nome-nome-preposição-nome** (*mueze-zambi-ya-mema*), conforme pode ser visto na Tabela 11.4 (Mel'čuk, 2000; Fabb, 2001).

Em síntese, identificamos, nos nossos dados, três tipos de composição na formação de empréstimos lexicais, sintetizados na Tabela 10.4.

N.º	Tipo de composição	Exemplos
1	Morfológica	hidroluachimo, hydrochicapa
2	Morfossintática	malembelember, matabicho, diculundundu
3	Sintagmática	dicamba-dia-ngalafa, calunga-ya-meia, suco-yo-bába, mueze-zambi-ya-mema,

Tabela 10.4 – Empréstimos formados por composição

10.3.3. Empréstimos por construção não concatenativa

Esta subsecção tem como base o Capítulo 4, Secção 4.7, *Processos de construção não concatenativa*. Visto que os traços fundamentais dos lexemas são, como advoga Lewandowski (1995), a sua estruturação morfemática, podemos depreender que a formação de alguns empréstimos nos remete para alguns processos de formação não concatenativa.

Constatamos três casos de alomorfia na base que nos remetem diretamente para o conceito de alofonia. Verificamos dois casos de truncação em que no primeiro se trunca o adjetivo e nome *caluanda* (< KALWANDA), que pode tomar a forma de diminutivo *calu*, com carácter mais valorativo-afetivo. No segundo, trunca-se o nome *sobongo*. No terceiro, trunca-se o nome *kiambamba* e, no quarto, trunca-se o nome *Mutu-ya-Kevela*. Notamos um caso de conversão do nome *muxima* e casos de cruzamento vocabular em que no primeiro notamos *mu* (truncação de *mukua*) + *xi* + *luando*; o segundo é o cruzamento de *mona* + *ngamba*, etc. Podemos ver alguns empréstimos de formação não concatenativa na Tabela 10.5.

N.º	Tipo de formação	Exemplos
1	Alomorfia na base	jindungo/gindungo, imbondeiro/embondeiro, funji/fungi
2	Truncação	calu, sobos, quiamba, mutu
3	Conversão	muximo
4	Cruzamento vocabular	muxiluanda, monangamba, ingombota, chinangol, musangola, mundele, refriango, Textang, bessangana

Tabela 10.5 – Empréstimos por formação não concatenativa

Quanto aos processos de formação não concatenativa, os dados ilustrados na Tabela 10.5 revelam uma grande produtividade de cruzamento vocabular e truncação. Constatamos casos de alomorfia na base por questões fonéticas, o que de facto nos permite verificar a grande tendência para assemelhar as formas lexicais ao sistema linguístico português.

10.4. Protótipo de Dicionário de Regionalismos Angolanos

Esta secção tem como base o Capítulo 3, Secção 3.2, *Língua geral e língua de especialidade*. Visto que pretendemos apresentar o enleio entre a lexicografia e a semântica lexical, nesta secção, procedemos à realização das seguintes propostas:

- Modelo de dicionário;
- Ficha lexicográfica de empréstimos.

10.4.1. Proposta de modelo de dicionário

Do que precede, consideramos, como referimos no primeiro capítulo desta dissertação, a necessidade de criar um dicionário de regionalismos angolanos que, com um pendor didático, possa auxiliar primeiramente os estudantes e professores em Angola e, em segundo lugar, os interessados na compreensão da Literatura e História de Angola. Ao recorrermos ao lexema *regionalismo*, empregamo-lo com o sentido de vocábulo, aceção, expressão própria de uma região, que, no nosso caso, é Angola. Assim, o dicionário pode ser entendido como repositório de amostras de significados (Sanromán, 2012).

Os dados revelam que estes regionalismos emanantes de línguas de Angola também podem passar por mudança semântica, como é o caso de *quindumba*, que denotava somente a ave poupa. Atualmente, a unidade léxica *quindumba* é mais usada para denotar o penteado puxado para o alto ou o cabelo comprido em alusão à ave. Outro exemplo é *muxima*, que outrora denotava coração e, além desta aceção, hodiernamente denota também um topónimo depois de se instaurar o santuário católico. Parece evidente que o lexema *muxima* nos remete para um caso de extensão semântica.

10.4.2. Ficha lexicográfica de empréstimos

Os dados coletados foram organizados e, para facilitar a compreensão, apresentamo-los em ficha lexicográfica, cujo modelo contém variados campos que permitem ter uma visão geral das aceções do empréstimo e sua utilização:

- Entrada: refere-se à unidade lexical;
- Etimologia: a língua de que provém o empréstimo lexical;
- Étimo: a unidade lexical original na língua de que provém o empréstimo;
- Categoria Gramatical: indicação da classe(s) de palavra;
- Variante: as distintas grafias apresentam como se escrevem os empréstimos;
- Significação: explicação breve e clara do empréstimo;
- Fraseologia: permite precisamente ver como a unidade lexical ocorre no eixo sintagmático;

- Imagem: ilustração ou figura que retrata a unidade lexical em causa;
- Nota: breve comentário para apresentar uma explicação.

FICHA LEXICOGRÁFICA			
ID	241	Cat. Gramatical	n. f.
		Abreviatura	-
		Domínio	Justiça
Entrada	Maka	Etimologia	Kimundu
		Fraseologia	O Silva tinha que arranjar uma maka .
Significação	1. n. f. Conversa; assunto; novidade; discórdia; litígio; conflito; algazarra. 2. Bras. Incolumanca. 3. adj. Problemático.		
		Étimo	Maka Variante -
Fonte da significação	<i>Dicionário de Regionalismos Angolanos (2014) e Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa (2001)</i>		
		Cotextos	1. ouvirem as makas . 2. arranjar uma maka . 3. reconstruir alguma maka passada. 4. centro das makas .
Contexto	Na opinião do barbeiro, estaria assim a reconstruir alguma maka passada muito tempo atrás.		
		Equivalências	1. ouvirem os problemas . 2. arranjar um problema . 3. reconstruir algum litígio . 4. pessoa problemática .
Fonte do contexto	<i>A Conjura (2008, p. 9)</i>		
		Imagem	
		Nota	Este lexema foi grafado como maca .

Fig. 10.2 – Ficha lexicográfica da entrada *Maka*

FICHA LEXICOGRÁFICA			
ID	235	Cat. Gramatical	n. f.
		Abreviatura	-
		Domínio	Administração
Entrada	Machila	Etimologia	Kimbundu
Frasesologia	Votaram os carregadores de machila e os trabalhadores das obras.		
Significação	n. f. 1. Palanquim. 2. Cadeira com tampo e cortinas, a qual suspensa de um bordão de bambu, era transportada aos ombros de dois homens.		
Étimo	Machila	Variante	-
Cotextos	<ol style="list-style-type: none"> 1. subido em sua machila 2. carregadores de machila 3. carregadores de trabalhadores machila 		
Fonte da significação	<i>Dicionário de Regionalismos Angolanos (2014) e Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa (2001)</i>		
Equivalências	<ol style="list-style-type: none"> 1. subido na sua cadeira régia. 2. carregadores da cadeira régia. 3. carregadores de trabalhadores da cadeira régia. 		
Contexto	E tendo subido em sua machila, ainda berrou para trás, como quem cospe, os seus sujos quimbundus.		
Imagem			
Fonte do contexto	<i>A Conjura (2008, p. 11)</i>		
Nota	Também é conhecido como tipoia.		

Fig. 10.3 – Ficha lexicográfica da entrada *Machila*

10.5. Síntese do capítulo

Neste capítulo, descrevemos o critério adotado para a determinação do carácter neológico das formas lexicais e procedemos à quantificação do subconjunto de empréstimos por categorização lexical. Explicitamos a determinação da etimologia dos empréstimos lexicais, os quais provêm de línguas de Angola. Apresentamos o seu étimo, etimologia e significação nas listas dos Anexos 1 e 2. Estrutturamos, por amostra, os empréstimos em campos lexicais e descrevemos os processos de formação dos empréstimos lexicais. Por fim, apresentamos o protótipo de dicionário de regionalismos angolanos mediante fichas lexicográficas.

Parte IV

Análise de Estruturas Passivas com Técnicas de Text Mining e Processamento de Linguagem Natural

Capítulo 11 – Estruturas Passivas em Português de Angola e Europeu

Este capítulo tem como suporte teórico o Capítulo 5, *Estruturas passivas e estruturas com o se*. Temos os seguintes objetivos fundamentais:

- Apresentar a metodologia usada para a deteção de estruturas passivas em Português de Angola e Europeu (ver a Secção 11.1);
- Descrever a extração de passivas eventivas e estativas. Além disso, apresentamos a extração de sintagmas com verbo auxiliar e principal (ver a Secção 11.2);
- Apresentar a avaliação do método computacional de extração proposto, com o auxílio de medidas de classificação (ver a Secção 11.3);
- Analisar as frequências de ocorrência de passivas em Português de Angola e Europeu (ver a Secção 11.4).

11.1. Metodologia para o estudo de estruturas predicativas

Objetivamos apresentar uma metodologia de deteção de estruturas passivas eventivas, estativas e resultativas para um estudo comparativo em Português de Angola e Europeu, usando o corpus de *Telejornal de Angola* e *Telejornal de Portugal*.

Para a concretização deste estudo, gravamos o *Telejornal* das 8 da TPA e da TVI disponíveis, respetivamente, no *Youtube* e *TVIPlay*, com o auxílio do *software Audacity* e procedemos à transcrição com o *software Dictate* para processá-lo, conforme descrito no Capítulo 8, Subsecção 8.2.1, *Constituição do corpus do Telejornal de Angola*.

11.1.1. Identificação de frases no corpus Telejornal

Nesta subsecção, descrevemos os procedimentos para a identificação de frases no corpus do *Telejornal de Angola* e *Telejornal de Portugal*. Assim, realizamos os seguintes estudos fundamentais para o processamento dos dados:

- Importação do corpus;
- Anotação dos dados textuais.

Importação do corpus

Procedemos à leitura do corpus *Telejornal de Angola* com o auxílio da instrução `readLines` e verificamos a estrutura de dados, usando a instrução `str`.

```
> corpus.telejornal <- readLines("corpus_telejornal.txt")
> str(corpus.telejornal)
chr [1:5] "Baratas, ratos mortos e muita sujidade. É neste ambiente em
que era produzido o pão que vai à mesa fa"| __truncated__ ...
```

onde podemos constatar que o corpus do *Telejornal de Angola* tem 5 documentos.

Em virtude do nosso objetivo e de modo a fazermos o processamento, formamos um só documento com a junção de todos os documentos de `corpus.telejornal`, usando a instrução

```
> corpus.telejornal <- paste(corpus.telejornal, collapse=" ")
> length(corpus.telejornal)
[1] 1
```

O mesmo procedimento foi aplicado ao corpus do *Telejornal de Portugal*. Assim, obtivemos o `corpus.telejornal.pe`.

Anotação dos dados textuais

Esta subsecção tem como base o procedimento descrito no Capítulo 7, Subsecção 7.3.1, *Part-of-speech tagging*. Para a anotação dos dados textuais, usamos o pacote `udpipe`. Atente-se nos procedimentos computacionais. Na linha 1, é instalado o pacote `udpipe`. Mais adiante, é criado um modelo para *part-of-speech tagging* em português, com o auxílio das instruções `udpipe_load_model` e `udpipe_annotate`. Transformamos os dados do *Telejornal de Angola* em *data frame*, usando a instrução `as.data.frame`, conforme a Fig. 11.1, que é uma parte da estrutura de dados que contém 39.355 linhas e 14 variáveis. A lista das variáveis (colunas) pode ser obtida através da instrução `str(text.annot)`.

```
> install.packages("udpipe")
> library(udpipe)
> udmodel <- udpipes_download_model(language = "portuguese")
> udmodel <- udpipes_load_model(file = udmodel$file_model)
> text.annot <- udpipes_annotate(udmodel, x = corpus.telejornal)
> text.annot <- as.data.frame(text.annot, detailed = TRUE)
> dim(text.annot)
[1] 39355    14
```

Na Fig. 11.1, vemos as colunas do *data frame* (nome da variável e alguns valores).

```
> str(text.annot)
'data.frame': 39355 obs. of 14 variables:
 $ doc_id      : chr  "doc1" "doc1" "doc1" "doc1" ...
 $ paragraph_id : int  1 1 1 1 1 1 1 1 1 1 ...
 $ sentence_id  : int  1 1 1 1 1 1 1 1 2 2 ...
```

```

$ sentence      : chr "Baratas, ratos mortos e muita sujidade."
"Baratas, ratos mortos e muita sujidade." "Baratas, ratos mortos e muita
sujidade." "Baratas, ratos mortos e muita sujidade." ...
$ token_id     : chr "1" "2" "3" "4" ...
$ token        : chr "Baratas" "," "ratos" "mortos" ...
$ lemma        : chr "Baratas" "," "rato" "morto" ...
$ upos         : chr "NOUN" "PUNCT" "NOUN" "ADJ" ...
$ xpos         : chr NA NA NA NA ...
$ feats        : chr "Gender=Fem|Number=Plur" NA
"Gender=Masc|Number=Plur" "Gender=Masc|Number=Plur" ...
$ head_token_id: chr "0" "3" "1" "3" ...
$ dep_rel      : chr "root" "punct" "conj" "amod" ...
$ deps         : chr NA NA NA NA ...
$ misc         : chr "SpaceAfter=No" NA NA NA ...

```

Fig. 11.1 – Tokenização e *Part-of-speech Tagging* de *Telejornal de Angola*

Verificamos o número de frases do *data frame text.annot*, com a instrução

```

> frases.telejornal <- table(text.annot$sentence)
> length(frases.telejornal)
[1] 1707

```

A título de exemplo, apresentamos as frases de 1 a 7 extraídas do documento 1. Mais frases são disponibilizadas no Anexo 3.

- (11.1) «Baratas, ratos mortos e muita sujidade.»
- (11.2) «É neste ambiente em que era produzido o pão que vai à mesa de muitas famílias em Luanda.»
- (11.3) «O legado e a competência profissional de Edgar Cunha marcam as homenagens ao apresentador da TPA.»
- (11.4) «Em Beirute, continua a busca por sobreviventes.»
- (11.5) «A comunidade Internacional continua a mobilizar-se para a ajuda ao Líbano.»
- (11.6) «A Sonangol serve o país na linha da frente há 44 anos.»
- (11.7) «No combate à Covid-19, capacitou um dos seus centros médicos para o rastreio e tratamento da doença e já com casos recuperados.»

O mesmo procedimento foi aplicado ao *Telejornal de Portugal*. O *data frame* que resultou disso tinha 38.106 linhas. Este corpus resultou em 2.271 frases, i.e., mais do que no corpus anterior, que tinha só 1.707 frases. Isso deve-se ao facto de o *Telejornal de Portugal* ter maior duração. Será que, pelo facto de ser mais longo e ter mais frases, quer dizer que há mais frases passivas eventivas, estativas e resultativas? É a isso que pretendemos responder na secção subsequente.

11.2. Extração de passivas

Nesta secção, focamo-nos em todo o corpus do *Telejornal de Angola* de modo a podermos extrair somente as estruturas passivas. Nesta parte, usamos o *data frame* `text.annot` descrito na Subsecção 11.1.1. Seleccionamos todas as frases que têm verbo auxiliar passivo (`dep_rel=="aux:pass"`), com a instrução

```
> text.annot[text.annot$dep_rel=="aux:pass", ]
> sentences.df.passivas <- text.annot[text.annot$dep_rel=="aux:pass",
c("doc_id", "sentence_id", "sentence")]
> dim(sentences.df.passivas)
[1] 3057      3
```

onde pudemos observar que os dados revelam 3.057 linhas com auxiliar passivo. Repare-se que isso se deve a duas causas: (1) existência de linhas com NA no *data.frame*, e (2) ocorrência de mais de uma forma verbal do auxiliar passivo nas frases. Ambos são explicados a seguir com mais detalhe.

Eliminação de NA

Verificamos que, em algumas linhas do *data frame* `text.annot`, há NA, conforme podemos ver na Fig. 11.2.

```
> text.annot[text.annot$sentence_id=="7",c("token_id", "token", "dep_rel")]
      token_id      token dep_rel
103         1          A    det
104         2      Sonangol nsubj
105         3         serve  root
106         4           o    det
107         5         país  obj
108        6-7          na <NA>
109         6           em  case
110         7           a    det
111         8         linha  obl
112        9-10         da <NA>
113         9           de  case
114        10           a    det
115        11         frente nmod
116        12           há  advcl
117        13           44 nummod
118        14         anos   obj
119        15           .  punct
```

Fig. 11.2 – Estrutura com NA do corpus *Telejornal de Angola*

Nota-se que a linha com o `token_id=6-7`, com `dep_rel=<NA>`, envolve `token=na`, i.e. a palavra que aparece no texto. Essa palavra foi decomposta em duas, i.e. a preposição

“em” e o artigo “a”. No caso de token_id=9-10, a palavra também foi decomposta em duas, i.e. a preposição “de” e o artigo “a”.

Para procedermos à eliminação de NA, efetuamos a constituição de um novo *data.frame* sem as linhas que incluem <NA>:

```
> text.annot.sna <- text.annot[!is.na(dep_rel_vector),]
```

Ocorrências de estruturas passivas

Para obtermos todas as estruturas passivas, usamos as instruções apresentadas na Fig.11.3, onde selecionamos de 1 a 7 estruturas passivas.

```
> text.annot.sna <- text.annot[!is.na(dep_rel_vector),]
> cond.linhas <- text.annot.sna$dep_rel=="aux:pass"
> sentences.df.passivas <- text.annot.sna[cond.linhas, ]
> colunas <- c("doc_id","sentence_id", "token_id", "token", "sentence")
> sentences.df.passivas[,colunas] [1:7,]
```

Fig. 11.3 – Estruturas passivas do corpus *Telejornal de Angola*

Na Tabela 11.1, podemos verificar algumas estruturas passivas do *data frame sentences.df.passivas*, que constam do documento 1.

Nº	Sentence_id	Token_id	Token	Sentence
1	2	7	era	É neste ambiente em que era produzido o pão que vai à mesa de muitas famílias em Luanda
2	17	9	era	Em algumas dessas padarias, o pão era fabricado em locais com baratas, ratos mortos e muita sujidade.
3	47	3	ser	Estamos a ser despejados .
4	48	12	seremos	Temos aqui o aviso de notificação que dá conta que nós seremos despejados dentro de 72 horas.
5	61	12	foram	O assalto ocorreu em plena madrugada neste estabelecimento, onde foram retirados 48 telefones.
6	70	19	são	Um dos assaltantes entende contar como são protegidos para efetuarem roubos frequentes sem, no entanto, serem
	70	31	serem	identificados pela polícia.
7	124	5	é	Parte deste gás é transformado em
	124	10	é	LPG e é disponibilizado ao mercado.

Tabela 11.1 – Estruturas passivas do corpus *Telejornal de Angola*

Constatamos que, na Tabela 11.1, algumas frases aparecem mais que uma vez, em virtude de incluírem, na estrutura predicativa, mais do que um verbo auxiliar (i.e. mais que 1 ocorrência de “*dep_rel*==“*aux:pass*””). Dois exemplos disso são as frases 70 e 124:

	doc_id	sentence_id	token_id	token	
	1484	doc1	70	19	são
	1497	doc1	70	31	serem
	2270	doc1	124	5	é
	2275	doc1	124	10	é

Verificamos a dimensão do *data frame* e notamos que os dados revelam que o *Telejornal de Angola* tem 236 estruturas passivas, com o auxílio da instrução

```
> dim(sentences.df.passivas)
[1] 236 14
```

Dado que algumas frases têm mais de um auxiliar passivo e, conseqüentemente, aparecerem mais de uma vez, verificamos a contagem de frases sem repetição.

```
> sentence.pass <- table(sentences.df.passivas[, "sentence_id"])
> length(sentence.pass)
[1] 188
```

O mesmo procedimento foi aplicado ao corpus do *Telejornal de Portugal*, onde pudemos constatar que tem 206 estruturas passivas e vemos algumas delas na Tabela 11.2.

Nº	Sentence_id	Token_id	Token	Sentence
1	23	3	era	O incêndio foi dominado às primeiras horas da manhã.
2	54	14	ser	Trump argumenta que o sistema de voto por correio, que está a ser planeado como medida de segurança devido à pandemia, irá conduzir a uma fraude em massa.
3	130	4	sendo	As medidas vão sendo reforçadas de dia para dia.
4	71	6	foi	Nas últimas horas, foi ultrapassada a marca negra de 150.000 mortos, mais de 4.400.000 infetados.
5	155	4	ser	O teste deve ser feito na origem.
6	142	5	foram	Em alguns hospitais, foram novamente canceladas as consultas e cirurgias para se prepararem para este novo aumento de casos.
7	166	6	será	Quem tiver 38º ou mais será direcionado para estruturas de apoio sanitário no aeroporto.

Tabela 11.2 – Estruturas passivas do corpus *Telejornal de Portugal*

11.2.1. Extração de sintagmas com verbo auxiliar e principal

Tendo em conta a Fig. 11.3, nesta subsecção, pretendemos extrair os sintagmas que consistem em pares *verbo auxiliar + verbo principal* nas frases passivas (*era produzido, era fabricado*). Ao analisarmos o *data frame text.annot*, reparamos que o verbo principal aparece normalmente na linha a seguir depois do verbo auxiliar.

Procedemos efetivamente a utilização do *data frame sentences.df.passivas* e verificamos as seguintes colunas: *doc_id*, *sentence_id*, *token_id*. Usamos isso na procura no *data frame text.annot*. Os primeiros dois elementos têm que ser iguais, mas o terceiro (*token_id*) tem que ser aumentado com +1.

Tem que se criar um ciclo que percorre todas as linhas do *data frame sentences.df.passivas*, permitindo extrair o verbo principal e juntar com o verbo auxiliar, conforme podemos ver na Fig. 11.4.

```
> linha.ult <- dim(sentences.df.passivas)[1]
> for(linha in 1: linha.ult) {
  # Definir a condição para a seleção de elementos (linhas) no data
frame "text.annot":
  xdoc <- sentences.df.passivas[linha, "doc_id"]
  xsent <- sentences.df.passivas[linha, "sentence_id"]
  xtoken <- sentences.df.passivas[linha, "token"]
  xtoken_id <- sentences.df.passivas[linha, "token_id"]
  xtoken_id.next <- as.character(as.numeric(xtoken_id) + 1)
  # Procurar estes elementos (linhas) no data frame "text.annot":
  cond_linhas <- text.annot$doc_id==xdoc &
text.annot$sentence_id==xsent &
text.annot$token_id==xtoken_id.next
  xtoken.next <- text.annot[cond_linhas,"token"]
  print(c(xsent, xtoken, xtoken.next))
}
[1]      "2"          "era"          "produzido"
[1]     "17"          "era"          "fabricado"
[1]     "19"          "são"          "preparados"
[1]     "39"          "ser"          "comercializada"
[1]     "41"          "foram"        "fiscalizados"
[1]     "47"          "ser"          "despejados"
```

Fig. 11.4 – Ciclo de extração de verbo auxiliar e principal

Os dados evidenciam que este método não deteta estruturas passivas eventivas, pois nestas estruturas a unidade lexical que normalmente segue o verbo auxiliar pode ser advérbio, como acontece, por exemplo, na estrutura *foi já implementado*. Neste caso, as nossas instruções iriam retornar *foi já*.

Processo melhorado

O método melhorado é semelhante ao anterior com a seguinte diferença. Após a detecção de *token* na linha a seguir ao verbo auxiliar, é preciso fazer um teste que verifica se esse *token* é um *verbo* e só nesse caso, imprimi-lo.

```
> linha.ult <- dim(sentences.df.passivas)[1]
> for(linha in 1: linha.ult) {
  # Definir a condição para a seleção de elementos (linhas) no data
frame "text.annot":
  < Ver instruções idênticas na Fig. 11.4 >
  # Procurar estes elementos (linhas) no data frame "text.annot":
  xtoken.next <- text.annot[cond_linhas,"token"]
  xupos.next <- text.annot[cond_linhas,"upos"]
  if (xupos.next == "VERB" ) {
    print(c(xsent, xtoken, xtoken.next, xupos.next))
  }
}
[1] "17"      "era"      "fabricado"  "VERB"
[1] "19"      "são"      "preparados" "VERB"
[1] "38"      "ser"      "comercializada" "VERB"
[1] "40"      "foram"    "fiscalizados" "VERB"
[1] "46"      "ser"      "despejados"  "VERB"
[1] "47"      "seremos"  "despejados"  "VERB"
```

Fig. 11.5 – Ciclo de extração de verbo auxiliar e principal melhorado

Extração automática de estruturas passivas estativas e eventivas

Na Fig. 11.6, observamos a extração de estruturas passivas estativas separadamente.

```
> sentence.estar <- text.annot[text.annot$lemma %in% c("estar"), ]
> sentence.estar.sna <- sentence.estar[!is.na(sentence.estar), ]
> cond.linha <- sentence.estar$dep_rel==c("aux:pass")
> sentences.pass.estativas <- sentence.estar[cond.linha, ]
> linha.ult <- dim(sentences.pass.estativas)[1]
> for(linha in 1: linha.ult) {
  # Definir a condição para a seleção de elementos (linhas) no data
frame "text.annot":
  < Ver instruções idênticas a Fig. 11.4 >
  Procurar estes elementos (linhas) no data frame "text.annot":
  xtoken.next <- text.annot[text.annot$doc_id==xdoc &
text.annot$sentence_id==xsent &
text.annot$token_id==xtoken_id.next,"token"]
  print(c(xsent, xtoken, xtoken.next))
}
[1] "262"      "estão"    "deslocadas"
[1] "73"       "estar"    "autorizada"
[1] "119"      "está"     "reunida"
[1] "262"      "estavam"  "internadas"
[1] "138"      "estão"    "comercializadas"
[1] "293"      "está"     "confirmada"
[1] "293"      "estar"    "infetada"
```

Fig. 11.6 – Ciclo de extração de estruturas passivas estativas

Verificamos a dimensão do *data frame* e notamos que os dados revelam que o *Telejornal de Angola* tem 15 estruturas passivas estativas, com o auxílio da instrução

```
> length(table(sentences.pass.estativas$sentence))
[1] 15
```

Efetuamos o mesmo procedimento para as estruturas passivas eventivas.

```
> length(table(sentences.pass.eventivas$sentence))
[1] 173
```

	xsent	xtoken	xtoken.next
[1]	"2"	"era"	"produzido"
[1]	"17"	"era"	"fabricado"
[1]	"19"	"são"	"preparados"
[1]	"39"	"ser"	"comercializada"
[1]	"41"	"foram"	"fiscalizados"
[1]	"48"	"seremos"	"despejados"
[1]	"57"	"fossem"	"apanhados"

Implementamos o mesmo procedimento para o corpus do *Telejornal de Portugal* e notamos que tem 8 estruturas passivas estativas e 196 estruturas passivas eventivas.

11.2.2. Extração de sintagmas com três elementos

Para detetar as passivas eventivas e estativas, usamos o seguinte procedimento, que extrai sintagmas com três elementos. Visto que em algumas estruturas há advérbio entre o verbo auxiliar e o verbo principal, seleciona-se o verbo auxiliar passivo e os dois elementos subsequentes, conforme podemos observar na Fig. 11.7.

```
> linha.ult <- dim(sentences.df.passivas)[1:2]
> for(linha in 1: linha.ult) {
  # Guardar os valores em variaveis:
  xdoc <- sentences.df.passivas[linha, "doc_id"]
  xsent <- sentences.df.passivas[linha, "sentence_id"]
  xtoken <- sentences.df.passivas[linha, "token"]
  xtoken_id <- sentences.df.passivas[linha, "token_id"]
  xtoken_id.next <- as.character(as.numeric(xtoken_id) + 1:2)
  # Procurar estes elementos no data frame "text.annot":
  cond_linhas <- text.annot$doc_id==xdoc &
    text.annot$sentence_id==xsent &
    text.annot$token_id==xtoken_id.next
  xtoken.next <- text.annot[cond_linhas,"token"]
  print(c(xsent, xtoken, xtoken.next))
}
```

[1]	"29"	"foi"	"aqui"	"questionado"
[1]	"33"	"está"	"completamente"	"recolhida"
[1]	"123"	"foi"	"também"	"aprovada"
[1]	"243"	"estava"	"praticamente"	"concluída"

Fig. 11.7 – Ciclo de extração de sintagmas com três elementos

11.2.3. Análise de estruturas passivas eventivas e estativas

As passivas eventivas (*ser + participio pasado*) são caracterizadas por uma componente agentiva e por uma componente eventiva, conforme foi observado no Capítulo 5, Subsecção 5.1.1, *Passiva eventiva*. As estruturas passivas estativas (*estar + participio pasado*) são caracterizadas por não terem componente agentiva nem componente eventiva (Duarte, 2013, p. 444).

Repare-se que a impossibilidade de ocorrência de expressões *em x tempo* revela que as estruturas passivas estativas têm uma natureza não eventiva ou atélica, como vimos no Capítulo 5, Subsecção 5.1.3, *Passiva estativa*.

Estruturas passivas eventivas

Em virtude da Fig. 11.7, podemos constatar que há estruturas com advérbio entre o verbo auxiliar e verbo principal. Tendo em conta o Capítulo 5, Subsecção 5.1.1, *Passivas eventivas*, procedemos à análise de todas as frases que foram fornecidas ao programa e todas as frases detetadas como passivas, para ver se tudo foi feito de maneira correta. Podemos notar, de facto, algumas estruturas com advérbio entre o verbo principal e o verbo auxiliar nos seguintes exemplos de *sentence_id* 29, 150 e 260, respetivamente.

- (11.8) «Por outro lado, também **foi aqui questionado** quantas pessoas vão entrar em quarentena domiciliar.»
- (11.9) «Uma outra referência a **ser já implementada** tem a ver com o desenvolvimento de uma aplicação informática capaz de recolher, tratar e sistematizar a informação para gerar relatórios do Balanço de execução dos projetos de atividades previstas no PDN.»
- (11.10) «Rapidamente, nesta segunda-feira, **foi então levado** à coleta de amostra de zaragatoa enviada a Luanda e confirmou-se, então, o resultado positivo à COVID-19.»

Podemos constatar a relação de dependência de estrutura passiva verbal do exemplo 11.8. Na Fig. 11.8, observamos que se reconhece o verbo *ser* como auxiliar passivo na coluna *dep_rel*, com o auxílio da instrução

```

> text.annot[text.annot$sentence_id=="29",c("token_id","token","dep_rel")]
  token_id      token  dep_rel
1         1         Por       cc
2         2         outro    fixed
3         3         lado     fixed
4         4          ,       punct
5         5         também  advmod
6         6         foi aux:pass
7         7         aqui     advmod
8         8 questionado  advcl
9         9         quantas  det
10        10        pessoas  nsubj
11        11         vão     aux
12        12         entrar  root
13        13          em     case
14        14 quarentena    obl
15        15 domiciliar    amod
16        16          .     punct

```

Fig. 11.8 – Relação de dependência de estrutura passiva eventiva

Estruturas passivas estativas

Como notamos no Capítulo 5, Subsecção 5.1.3, *Passivas estativas*, constatamos que, além de passivas eventivas (*ser + participio passado*), o sistema também deteta passivas estativas (*estar + participio passado*) (Duarte, 2013, pp. 432-444), conforme observamos nos exemplos *sentence_id* 33, 167 e 292 respetivamente.

- (11.11) «É possível de facto ver por estas imagens que Luanda ainda não **está completamente recolhida** pelo menos nesta zona de Cacuaco.»
- (11.12) «**Estavam armados** com uma *star*.»
- (11.13) «E agora **está confirmada** praticamente a transmissão comunitária do vírus em Luanda.»

Tendo em conta o Capítulo 7, Subsecção 7.3.3, *Dependency Parsing*, podemos verificar a relação de dependência da estrutura passiva estativa do exemplo 11.13, conforme podemos observar na Fig. 11.9 na coluna *dep_rel*.

```

> text.annot[text.annot$sentence_id=="292",c("token_id","token","dep_rel")]
  token_id      token  dep_rel
1         1          E        cc
2         2         agora  advmod
3         3         está  aux:pass
4         4   confirmada  root
5         5   praticamente  advmod
6         6          a      det
7         7   transmissão  nsubj:pass
8         8   comunitária  amod
9        9-10         do    <NA>
10        9          de    case
11       10          o      det
12       11         vírus  nmod
13       12          em    case
14       13         Luanda  nmod
15       14          .    punct

```

Fig. 11.9 – Relação de dependência de estrutura passiva estativa

Estruturas mal classificadas

Depois desta extração, constatamos que algumas estruturas (*sentence_id* 186, 363, 43) foram mal classificadas. Analisamos o *data frame* *text.annot* para verificarmos a exatidão da classificação. Atentemos na frase seguinte.

(11.14) «Nós temos agregados familiares também grandes e ali já estamos a ver o que é que pode acontecer.»

Podemos verificar que, na frase 11.14, não há nenhum verbo auxiliar passivo. Na Fig. 11.10, verificamos que a etiqueta (*tag*) na coluna *dep_rel* do *data frame* *text.annot* indica que o verbo *ter* foi classificado como auxiliar passivo.

```

> text.annot[text.annot$sentence_id=="43",c("token_id","token","dep_rel")]
  token_id      token  dep_rel
9467     "1"         Nós  nsuj:pass
9468     "2"         temos  aux:pass
9469     "3"   agregados  root
9470     "4"   familiares  nsuj:pass
9471     "5"         também  advmod
9472     "6"         grandes  advmod

```

Fig. 11.10 – Relação de dependência de estrutura mal classificada

Constatamos que, na Fig. 11.10, não se trata de erro do nosso método, mas do sistema *udpipe*. O sistema entende o verbo *ter* como auxiliar passivo e não como auxiliar do tempo composto, o que também ocorreu em *sentence_id* 363.

(11.15) «Edgar Cunha foi um dos rostos do principal serviço de notícias da estação pública de televisão em Angola, tendo igualmente exercido assinalável atividade enquanto adido de imprensa ao serviço da diplomacia angolana fruto das suas qualidades e empenho.»

Verificamos que, na Fig. 11.10, o sistema classificou o verbo *ter* como auxiliar passivo, quando, na verdade, está como verbo pleno, cuja estrutura argumental é binária, isto é, o verbo *ter* não está como auxiliar. Há duas causas para a má classificação, conforme podemos notar no *Dependency Parsing* da frase. Por um lado, considerou o verbo *ter* como auxiliar passivo. Por outro, considerou *agregados* como verbo, quando está como nome, formando um nome composto sintagmático *agregados familiares*.

Consideramos que o composto sintagmático, como sublinham Mel'čuk (2006, p. 14), Mendevil Giró (2009, p. 84) e Rio-Torto e Ribeiro (2016, pp. 487-489), é um sintagma que funciona semântica e sintaticamente como uma palavra única. Os compostos sintagmáticos são considerados como palavras com estrutura externa, conforme vimos no Capítulo 4, Secção 4.8, *Processos de composição em português*. O sistema entendeu que *agregado* é uma coisa e *familiar* é outra, não considerando como composto sintagmático. Recorde-se que, em Mota (2013, p. 2823), os compostos sintagmáticos são denominados compostos sintáticos. Repare-se em *sentence_id* 186, que podemos observar no exemplo seguinte.

(11.16) «Uma coisa é certa: a morte provoca perdas e despedaça corações.»

O verbo *ser* está como *copulativo*, no entanto foi categorizado como *auxiliar passivo*.

```
> text.annot[text.annot$sentence_id=="186",c("token_id","token","dep_rel")]
  token_id  token  dep_rel
4046    "1"    uma      det
4047    "2"   coisa nsuj:pass
4048    "3"     é    aux:pass
4049    "4"   certa      root
```

11.2.4. Avaliação do processo de extração

Esta secção tem como suporte teórico o Capítulo 7, Subsecção 7.4.1, *Medidas para problemas de classificação*. Com efeito, usamos as equações de precisão, *recall* e medida-F no corpus do *Telejornal de Angola e Telejornal de Portugal*. Quanto às passivas eventivas e estativas, recorde-se que foram detetadas 236 estruturas passivas no *Telejornal de Angola*, dentre as quais 3 são mal classificadas. A Tabela 11.3, apresenta a matriz de confusão destes dados. O símbolo “+” representa as estruturas passivas e o símbolo “-” representa as estruturas ativas.

		Classe prevista	
		+	-
Classe verdadeira	+	233	0
	-	3	1471

Tabela 11.3 – Matriz de confusão do *Telejornal de Angola*

Aplicamos o mesmo procedimento ao *Telejornal de Portugal*. Assim, obtivemos 206 estruturas passivas, dentre as quais 2 delas foram mal classificadas. A Tabela 11.4 indica a matriz de confusão desta estrutura de dados.

		Classe prevista	
		+	-
Classe verdadeira	+	204	0
	-	2	2065

Tabela 11.4 – Matriz de confusão do *Telejornal de Portugal*

Usamos as equações ilustradas na secção acima referida e os dados acima (Tabela 11.3 e Tabela 11.4), para calcular as medidas de desempenho do nosso método automático, conforme se pode ver na Tabela 11.5.

	Precisão	Recall	Medida-F
Telejornal de Angola	98.7%	100%	99.3%
Telejornal de Portugal	99%	100%	99.5%

Tabela 11.5 – Medidas de desempenho do método automático

11.2.5. Estruturas passivas resultativas

Embora o sistema não detete as passivas resultativas, notamos que, para as estruturas passivas resultativas, o sistema indica que o verbo é auxiliar, mas não indica que é auxiliar passivo. A Fig. 11.11 apresenta a extração das passivas resultativas com a instrução

```
> sent.ficar <- text.annot[text.annot$lemma %in% c("ficar"), ]
> sent.ficar.sna <- sentence.ficar[!is.na(sentence.ficar),]
> cond.linha <- sent.ficar$dep_rel==c("aux","root","cop")
> sent.pass.res <- sent.ficar.sna[cond.linha, ]
> colunas <- c("doc_id","sentence_id","token_id","token")
> sent.pass.res[,colunas] [1:3,]
      doc_id sentence_id token_id token
16715 doc2          333         59 ficou
21907 doc3          184         22 ficaram
27566 doc4          160          1 Ficou
```

Fig. 11.11 – Extração do verbo *ficar* como auxiliar

Assim, procedemos à verificação das frases e podemos observar que são estruturas passivas resultativas os exemplos seguintes de *sentence_id* 333, 184 e 160 respetivamente.

(11.17) «Na República Árabs Unidos e da sua comunidade residente, a embaixada descreve que o desaparecimento físico do profissional da comunicação social angolana representa um momento de tristeza e de dor para a prestigiada classe, para a TPA em particular, para a família enlutada em especial, cuja trajetória iniciada nos anos 80 **ficou marcada** com elevado sentido patriótico, espírito de missão e excelência laboral.»

(11.18) «E uma delas é a garantia do regresso a casa dos angolanos que, por motivos diversos, **ficaram retidos** no exterior do país através da realização de voos humanitários.»

(11.19) «**Ficou** ligeiramente **ferido**.»

Na Fig. 11.12, notamos que o *udpipe* não indica o verbo *ficar* como *auxiliar passivo*.

```
> text.annot[text.annot$sentence_id=="160",c("token_id","token","dep_rel")]
      token_id      token      dep_rel
27566         1      Ficou      root
27567         2 ligeiramente  advmod
27568         3      ferido      xcomp
27569         4          .      punct
```

Fig. 11.12 – Relação de dependência de estrutura passiva resultativa

11.3. Análise comparativa de ocorrências de passivas

Nos corpora do *Telejornal de Angola e de Portugal* podemos observar que as ocorrências relativas às frases passivas são bastante diferentes, como se pode ver na Tabela 11.6.

	Sintagmas passivos	Frases passivas	Total de frases	Proporção de sintagmas passivos	Proporção de frases passivas
Telejornal de Angola	233	185	1707	13.64%	10.83%
Telejornal de Portugal	204	204	2271	8.98%	8.98%

Tabela 11.6 – Sumário de ocorrências de passivas

Será que esta diferença é significativa? Para determinar isso, aplicamos o teste estatístico de proporções, descrito no Capítulo 7, Subsecção 7.4.1. Mais pormenores sobre este assunto são apresentados na subsecção subsequente.

11.3.1. Aplicação do teste estatístico de proporções

Nesta parte, vamos analisar as proporções de sintagmas passivos (sintagmas com o verbo auxiliar passivo e verbo principal). Assim, podemos notar duas hipóteses.

```
Corpus do PE  $p_1 = 204/2271 = 0.08982827$   
Corpus do PA  $p_2 = 233/1707 = 0.1364968$   
 $H_0 : p_1 - p_2 = 0$   
 $H_1 : p_1 - p_2 \neq 0$ 
```

Efetuamos o teste de proporções com a instrução:

```
> prop.test(c(204,233),c(2271,1707))  
2-sample test for equality of proportions with continuity correction  
data: c(204, 233) out of c(2271, 1707)  
X-squared = 21.23, df = 1, p-value = 4.073e-06  
alternative hypothesis: two.sided  
95 percent confidence interval:  
-0.06726999 -0.02606703  
sample estimates:  
prop 1 prop 2  
0.08982827 0.13649678
```

Fig. 11.13 – Teste estatístico de proporções de sintagmas passivos

Os dados revelam que se rejeita a hipótese nula (H_0), em virtude do $p\text{-value} = 4.073e-06$, i.e. valor bastante inferior ao limiar habitual de 0.05. Consideramos que a diferença de proporção é significativa. Podemos concluir que, embora o *Telejornal de Portugal* tenha mais frases passivas, o corpus do *Telejornal de Angola* (13,64%) tem mais sintagmas passivos em relação ao corpus do *Telejornal de Portugal* (8,98%).

11.4. Análise comparativa de ocorrências de tipos de passiva

Nos corpora do *Telejornal de Angola* e *de Portugal*, podemos observar que as ocorrências relativas aos sintagmas passivos são diferentes, como se pode ver na Tabela 11.7.

	Sintagmas passivos eventivos	Sintagmas passivos estativos	Total de frases	Proporção de sintagmas passivos eventivos	Proporção de sintagmas passivos estativos
Telejornal de Angola	218	15	1707	12.77%	0.87%
Telejornal de Portugal	196	8	2271	8.63%	0.35%

Tabela 11.7 – Sumário de ocorrências de sintagmas passivos por tipologia

Procuramos verificar se esta diferença é significativa. Assim, aplicamos o teste estatístico de proporções de maneira semelhante como na secção 11.3. Nesta parte, vamos analisar dois casos separadamente: o primeiro envolve as proporções de sintagmas passivos estativos, e o segundo as proporções de sintagmas passivos eventivos (sintagmas com o verbo auxiliar passivo e verbo principal).

Comparação de proporções de sintagmas passivos estativos

Podemos notar duas hipóteses (H_0 , H_1). Na hipótese nula (H_0), considera-se que não há diferença entre a proporção de frases passivas do corpus *Telejornal de Portugal* (p_1) e a proporção de passivas do corpus do *Telejornal de Angola* (p_2). Na outra hipótese (H_1), considera-se que há diferença entre as duas proporções.

```
Corpus do PE  $p_1 = 8/2271 = 0.003522677$   
Corpus do PA  $p_2 = 15/1707 = 0.008787346$   
 $H_0 : p_1 - p_2 = 0$   
 $H_1 : p_1 - p_2 \neq 0$ 
```

O teste de proporções de sintagmas passivos estativos pode ser invocado pela instrução:

```
> prop.test(c(8,15),c(2271,1707))
```

Verificamos que, por um lado, a proporção de estruturas passivas estativas é maior no *Telejornal de Angola*; por outro lado, os resultados comprovam que a proporção ainda é quase significativa, pois está perto do limiar habitual de $p\text{-value} = 0.05042$ normalmente usado.

Comparação de proporções de sintagmas passivos eventivos

Nesta parte, tomamos em consideração as ocorrências de sintagmas passivos eventivos. Assim, podemos notar duas hipóteses.

```
Corpus do PE  $p_1 = 196/2271 = 0.08630559$   
Corpus do PA  $p_2 = 218/1707 = 0.1277094$   
 $H_0 : p_1 - p_2 = 0$   
 $H_1 : p_1 - p_2 \neq 0$ 
```

Procedemos ao teste de proporções com a instrução

```
> prop.test(c(196,218),c(2271,1707))
```

e notamos que os dados revelam que se rejeita a hipótese nula (H_0), em virtude do $p\text{-value} = 2.91e-05$, i.e. valor bastante ao limiar habitual de 0.05. Consideramos que a diferença de proporção é significativa. Podemos concluir que o corpus do *Telejornal de Angola* (12,77%) tem mais sintagmas passivos eventivos em relação ao *Telejornal de Portugal* (8,63%).

11.5. Síntese do capítulo

Neste capítulo, apresentamos a metodologia usada para a deteção de estruturas passivas em Português de Angola e Europeu. Descrevemos a extração de passivas eventivas e estativas e apresentamos a extração de sintagmas com verbo auxiliar e principal. Avaliamos o método computacional proposto, com o auxílio de medidas de classificação. Podemos constatar que o procedimento é fiável, pois o sistema deteta estruturas passivas eventivas e estativas.

Vimos que, embora o *Telejornal de Portugal* tenha mais frases e mais estruturas passivas, o corpus do *Telejornal de Angola* (13,64%) tem mais sintagmas passivos em relação ao corpus do *Telejornal de Portugal* (8,98%) e o teste estatístico confirma que a diferença de proporção é significativa. Podemos notar que o trabalho podia ser melhorado se (1) o sistema detetasse estruturas passivas resultativas e (2) reconhecesse o verbo *ter* como auxiliar do tempo composto e não auxiliar da passiva.

Capítulo 12 – Análise das Estruturas Passivas

Este capítulo é uma continuação do Capítulo 11. Tem como suporte teórico o Capítulo 5, *Estruturas passivas e estruturas com o se*, e a concretização do Capítulo 7, Secção 7.3, *Técnicas avançadas de Text Mining*. Temos, por conseguinte, os seguintes objetivos fundamentais:

- Descrever a estrutura de constituintes em passivas eventivas no Português de Angola e Português Europeu (ver a Secção 12.1);
- Comparar as estruturas passivas em Português de Angola e Português Europeu (ver a Secção 12.2).

12.1. Relação de dependência entre os constituintes

Para a verificação da relação de dependência de constituintes em estruturas passivas eventivas, na Fig. 12.1, procedemos à identificação de classes lexicais e dependências de constituintes de estrutura passiva, com o auxílio das instruções.

```
> sent <- c("Os descobrimentos obviamente foram pensados aqui em Tomar.")
> install.packages("udpipe")
> library(udpipe)
> udmodel <- udpipe_download_model(language = "portuguese")
> udmodel <- udpipe_load_model(file = udmodel$file_model)
> sent.pass.annot <- udpipe_annotate(udmodel, x = sent)
> sent.pass.annot <- as.data.frame(frase.pass.annot, detailed = TRUE)
> sent.pass.annot[,c("token_id", "token", "lemma", "feats", "head_token_id",
"upos", "dep_rel")]
```

	token_id	token	lemma	feats	head_token_id	upos	dep_rel
1	1	Os	o	Definite=Def Gender=Masc Number=Plur PronType=Art	2	DET	det
2	2	descobrimientos	descobrimento	Gender=Masc Number=Plur	5	NOUN	nsubj:pass
3	3	obviamente	obviamente	<NA>	5	ADV	advmod
4	4	foram	ser	Mood=Ind Number=Plur Person=3 Tense=Past VerbForm=Fin	5	AUX	aux:pass
5	5	pensados	pensar	Gender=Masc Number=Plur VerbForm=Part Voice=Pass	0	VERB	root
6	6	aqui	aqui	<NA>	5	ADV	advmod
7	7	em	em	<NA>	8	ADP	case
8	8	Tomar	Tomar	Gender=Masc Number=Sing	5	PROPN	obl
9	9	.	.	<NA>	5	PUNCT	punct

Fig. 12.1 – *Part-of-speech tagging* de estrutura passiva

A construção frásica da Fig. 12.1 indica que se trata de estrutura passiva eventiva e podemos observar a sua relação de dependência de constituintes na Fig. 12.2.

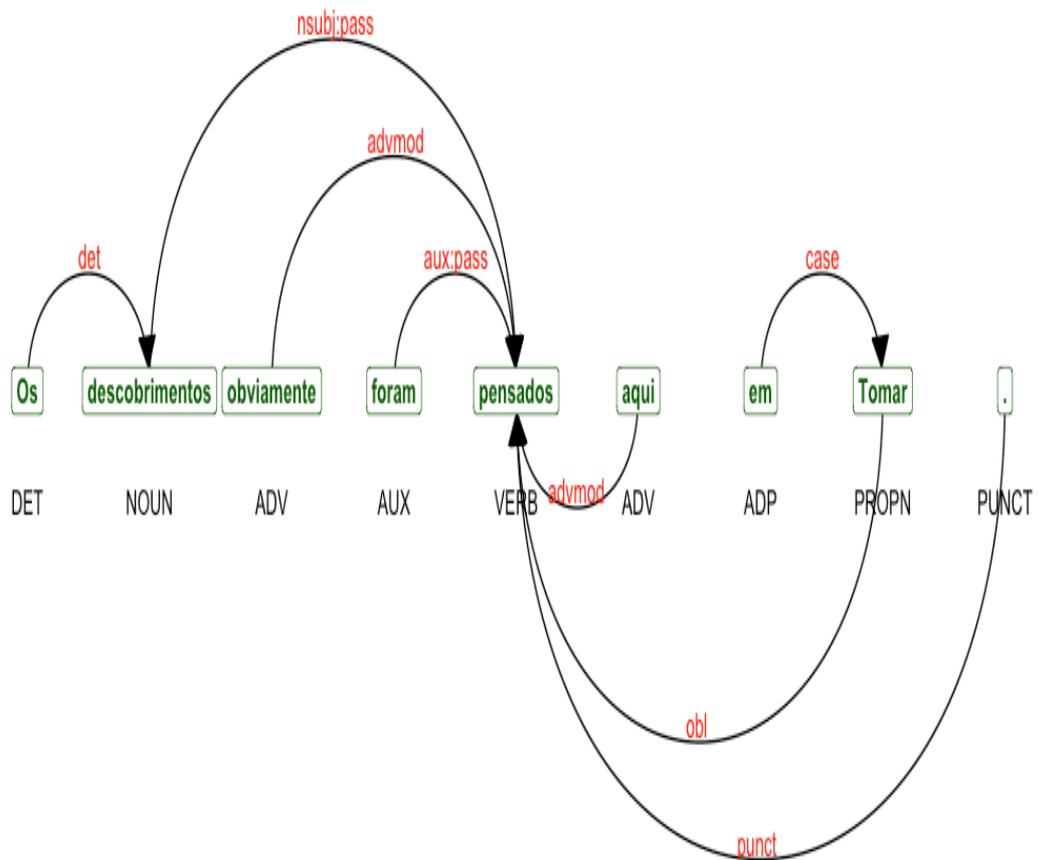


Fig. 12.2 – *Dependency parsing* de passiva eventiva curta no PE

Refletindo com base na estrutura temática, argumental e sintática, constatamos que, com o auxílio de métodos computacionais, é possível distinguir as estruturas passivas das ativas, pois o exemplo da Fig. 12.1, sendo uma passiva eventiva curta, revela que:

- A partir do movimento de constituintes, podemos observar que o argumento interno objeto direto da oração ativa (*os descobrimentos*) passa a ser alinhado com a relação gramatical de sujeito passivo (*nsubj:pass*) na passiva correspondente. Apesar de não estar expresso, o argumento externo (sujeito) da oração ativa passa para uma posição não central, pois passa a estabelecer uma relação gramatical oblíqua. Assim, na passiva, tal constituinte passa a ser introduzido pela preposição de valor agentivo *por* e tem precisamente a relação gramatical de complemento agente da passiva;

- Na frase, ocorre um grupo verbal complexo iniciado pelo verbo auxiliar *ser*, que foi identificado como auxiliar passivo (*aux:pass*), seguido de um particípio com o estatuto verbal (*pensados*) correspondente ao verbo pleno da oração ativa;
- As orações passivas eventivas descrevem de facto eventos, ou seja, situações dinâmicas;
- As passivas eventivas caracterizam-se precisamente por uma componente agentiva e por uma componente eventiva, pois focalizam a fase do processo de mudança de estado, lugar ou posse. Com efeito, vários autores usam a denominação passiva eventiva.
- Na oração passiva eventiva, a estrutura temática mantém-se. Entretanto, há uma estrutura sintática e informacional diferente.
- Nas passivas curtas, o papel temático externo está implícito, como se pode comprovar através da existência de advérbios orientados para agente, como *atentamente*, *propositadamente*, etc.

Atentemos na relação de dependência entre os constituintes da passiva eventiva curta do *Telejornal de Angola*, cujo pronome relativo ‘que’ tem a relação gramatical de sujeito passivo e o seu antecedente é ‘um legado’.

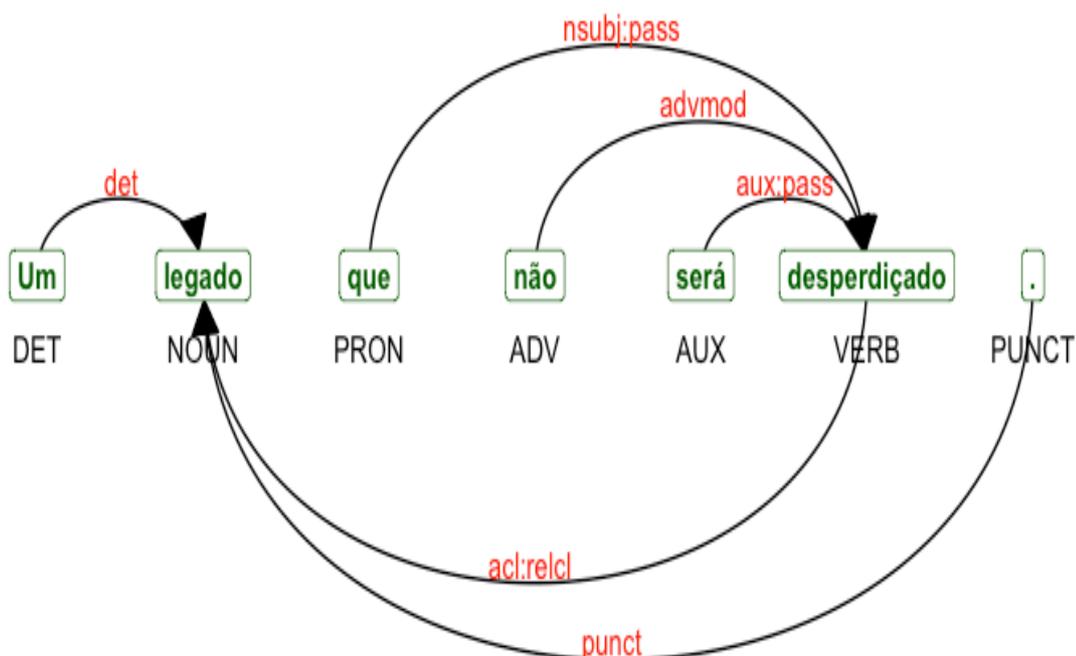


Fig. 12.3 – *Dependency parsing* de passiva eventiva curta no PA

Estruturas passivas eventivas longas

Como vimos na subsecção 5.1.1, *Passivas eventivas*, as orações passivas em que ocorre o complemento agente da passiva são denominadas passivas eventivas longas. A opcionalidade do sintagma *por* poderia levar-nos a pensar que se trataria de um adjunto.

Para construções passivas eventivas, só podem ocorrer em estruturas passivas eventivas os verbos de predicado binário ou ternário em que o argumento que se realiza como sujeito da frase tem o estatuto de argumento direto na entrada lexical do verbo. Semanticamente, as duas frases (ativa e passiva) são essencialmente equivalentes, uma vez que exprimem a predicação básica e os mesmos valores de tempo, aspeto e modo (Gaatone, 1998). Além disso, os predicadores de ambas as frases são formas derivadas do mesmo verbo.

Atentemos nos seguintes exemplos extraídos do *Telejornal de Angola* e *Telejornal de Portugal*, respetivamente.

(12.1a) «O assunto foi debatido pelos deputados.»

(12.1b) «Fomos literalmente proibidos pela Guarda Nacional Republicana.»

A Fig. 12.4 apresenta a estrutura de dependência de constituintes da frase (12.1a).

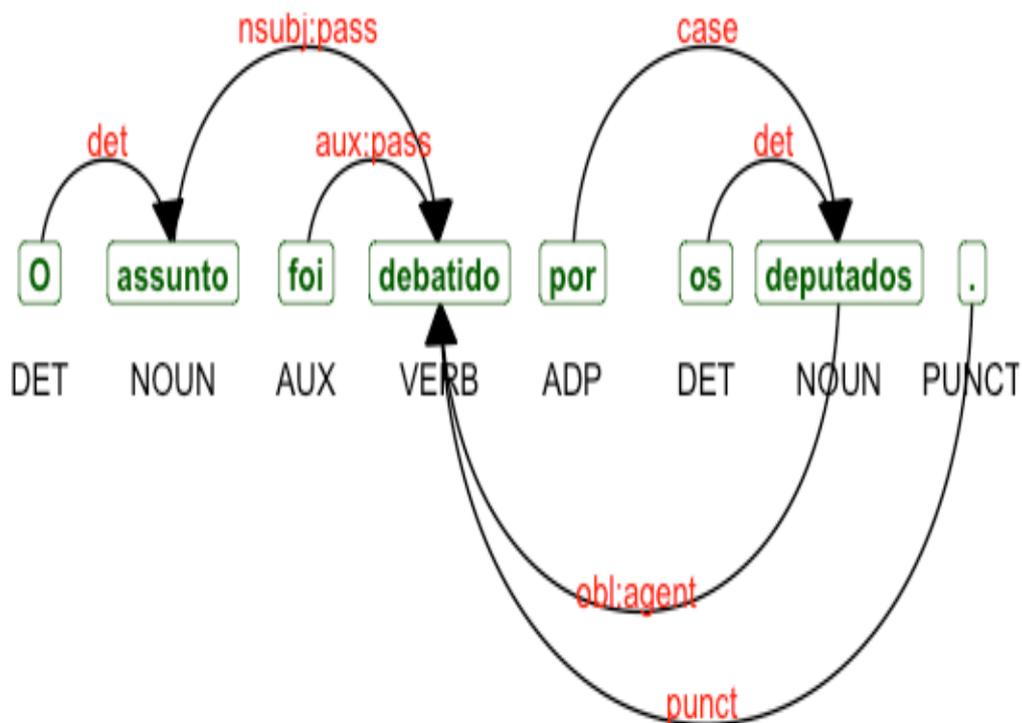


Fig. 12.4 – *Dependency parsing* de passiva eventiva longa no PA

Este procedimento computacional de relação de dependência de estrutura de constituintes foi repetido com a frase (12.1b) extraída do *Telejornal de Portugal*. A Fig. 12.5 mostra o resultado deste procedimento.

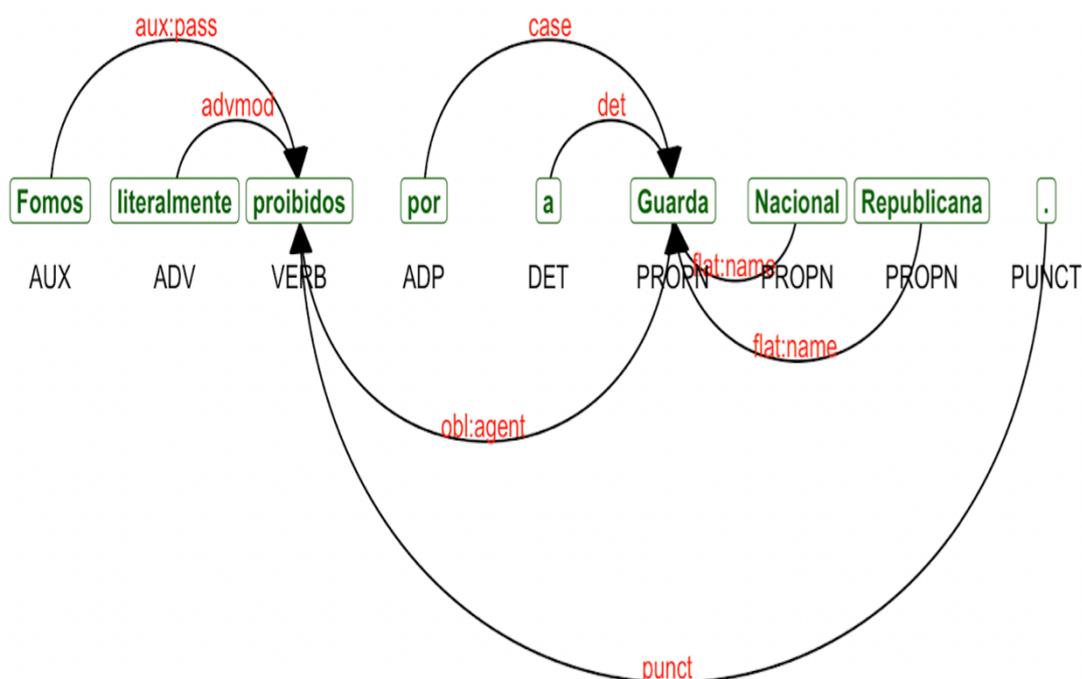


Fig. 12.5 – *Dependency parsing* de passiva eventiva longa no PE

Quando temos culminações ou processos culminados, as passivas eventivas (*ser + participio pasado*) expressam tanto a ação como o estado resultante, construindo-se apenas com verbos transitivos. Nas construções passivas eventivas, a natureza aspectual do participio é irrelevante, uma vez que ocorrem participios eventivos pertencentes a diferentes classes aspectuais básicas, tais como *destruído*, *aberta*, *conduzido*, *apreciado* que indicam processo culminado, culminação, processo e estado, respetivamente (Estrela, 2013).

Parece não ser possível haver elevação do sujeito de uma oração subordinada para sujeito das construções passivas estativas, havendo um comportamento diferente quando se trata da construção passiva eventiva (Mendes, 1994).

(12.1c) «O Reuel foi mandado entrar.»

(12.1d) *«O Reuel está mandado entrar.»

Estruturas passivas estativas

As passivas estativas descrevem situações estativas, mas não incluem a componente da fronteira da passagem ao estado, conforme podemos notar na frase do *Telejornal de Angola*.

(12.2a) «O falso técnico da ENDE já está detido.»

(12.2b) «A Herdade do Mouchão esteve ocupada durante dez anos.»

A Fig. 12.6 apresenta a estrutura de dependência de constituintes da frase (12.2a).

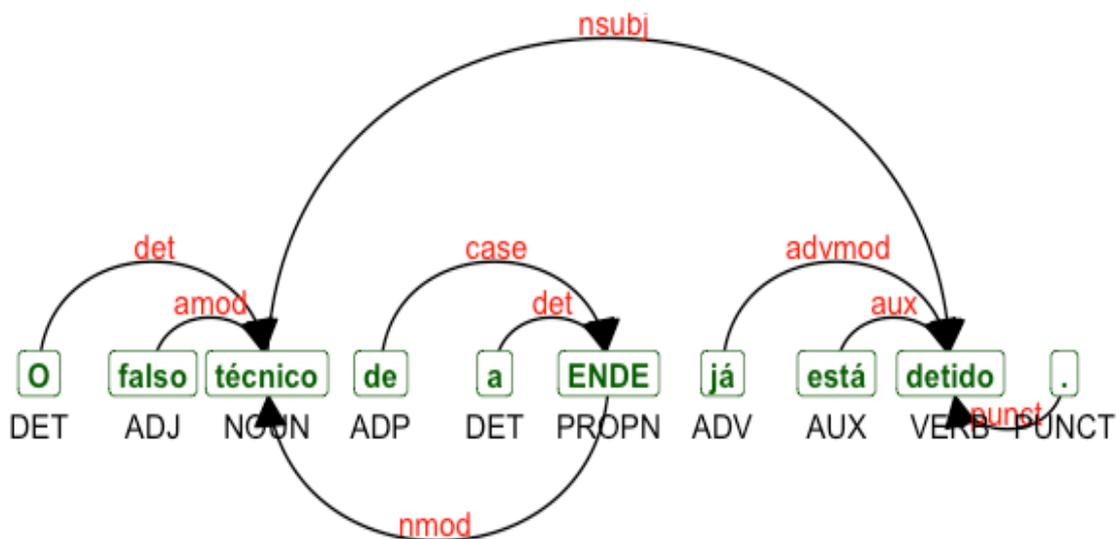


Fig. 12.6 – *Dependency parsing* de passiva estativa no PA

Aplicamos o mesmo procedimento computacional à estrutura passiva do *Telejornal de Portugal* do exemplo (12.2b), cuja estrutura de constituinte podemos observar na Fig. 12.7. Verificamos que a forma participial nas passivas estativas e resultativas é um adjetivo e não um verbo como na passiva eventiva, formado por um processo morfológico de conversão, admitem formas participiais com o prefixo *in-* que não correspondem a verbos em português e formas participiais com sufixos diminutivos (Duarte, 2003b).

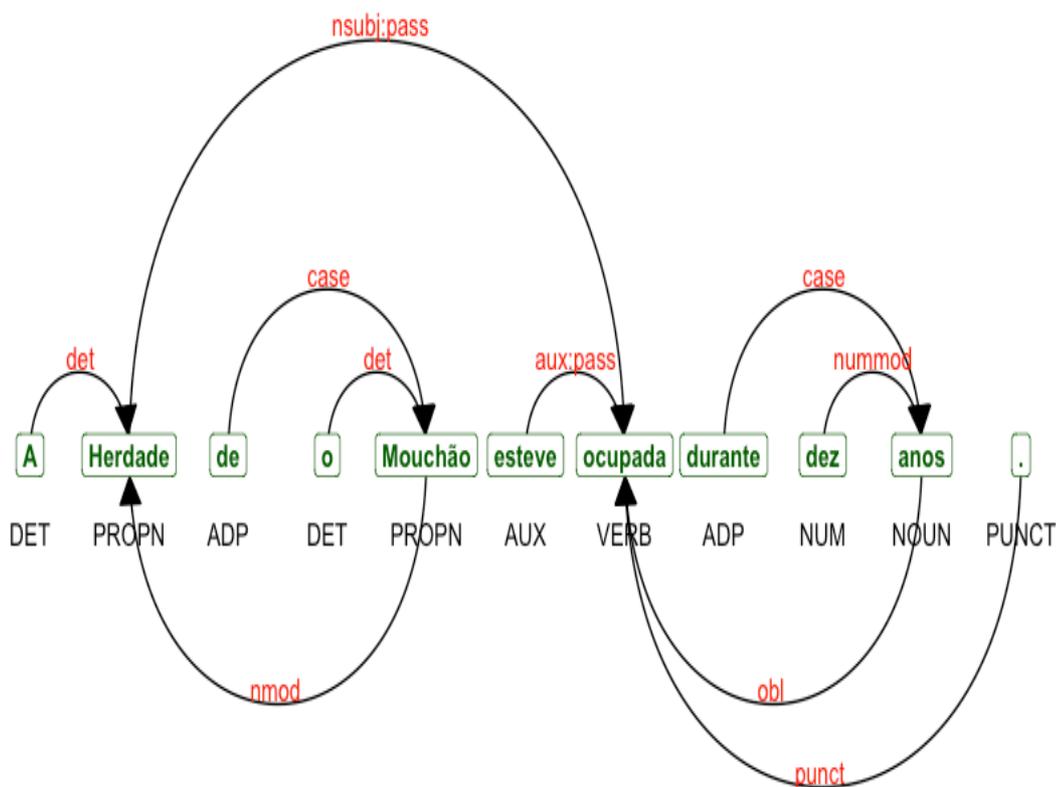


Fig. 12.7 – *Dependency parsing* de passiva estativa no PE

Verificamos que, nas duas variedades, as passivas estativas não têm componente agentiva nem componente eventiva. Como vimos na subsecção 5.1.3, *Passivas estativas*, as passivas estativas não podem coocorrer com a expressão em *x tempo* uma vez que esta expressão é usada essencialmente com eventos télicos, que identificam o intervalos que descrevem a totalidade da situação. Esta característica é reflexo da natureza não eventiva da passiva estativa. Para a formação da passiva estativa, as construções com *estar* não têm valor de processo culminado nem de culminação.

Nas passivas estativas, é obrigatória a ocorrência da forma recategorizada como adjetivo, quando o verbo principal tem duplo participípio. Os complementos preposicionais “pseudoagentivos” em estruturas passivas são compatíveis com a significação estativa dos adjetivos (Bosque, 1999).

(12.2c) «Morpheus estava rodeado **por grupos de agente.**»

Os adjetivos formados a partir de verbos atélicos locativos são legítimos, ocorrendo tipicamente neste caso o constituinte com a interpretação locativa precedido da preposição *de*, conforme o exemplo subsequente.

(12.2d) «Todo perímetro do hotel **estava rodeado de** carros de luxo.»

Além disso, a seleção do verbo auxiliar é sensível ao tipo de predicado: com predicados estáveis (que denotam propriedades estáveis dos indivíduos), utiliza-se o verbo *ser*; enquanto com predicados episódicos (que denotam propriedades transitórias dos indivíduos) se utiliza o verbo *estar* (Duarte, 2013; Duarte & Oliveira, 2010). A passiva eventiva, resultativa e a estativa podem aceitar inserção de um advérbio (*muito*, *pouco*, *bastante*, etc.) para alterar o grau do elemento à sua direita.

(12.2e) «Só que **estão muito escondidos** aqui dentro.»

(12.2f) «Só que **ficaram muito destruídos**.»

(12.2g) «Este carro **foi muito visto** na semana passada.»

Enquanto na passiva eventiva se focaliza a transição sofrida pelo argumento com o papel temático interno direto, na passiva estativa e passiva resultativa, o enfoque recai sobre o estado resultante da transição sofrida.

12.2. Comparação das estruturas passivas

Os dados obtidos permitem constatar que, na realização das estruturas passivas detetadas, se verifica a mesma tendência em PA e em PE no que se refere à sua distribuição, ordenada nos subcorpora, como vimos, da seguinte forma: estrutura passiva eventiva curta – estrutura passiva eventiva longa – estrutura passiva estativa.

É evidente que os dados do Português de Angola indicam que a estrutura de passiva é a mesma, ocorrendo passiva eventiva longa e passiva eventiva curta, conforme verificamos nos exemplos (12.3a – 12.3.c), que são passivas longas, e (12.4a – 12.4.c), que são passivas curtas. Notamos a predominância de passivas eventivas curtas nas duas variedades do português.

(12.3a) «O assunto **foi debatido pelos** deputados.»

(12.3b) «O grupo **foi recebido** esta manhã **pela** Ministra da Saúde.»

(12.3c) «As medidas de prevenção e combate à pandemia **são observadas pelos** dois países.»

(12.4a) «Parte deste gás **é transformado** em LPG e **é disponibilizado** ao mercado.»

(12.4b) «Para o ministro de estado para coordenação económica, é mais um órgão de apoio ao Presidente da República que **acaba de ser criado** para dar resposta aos vários projetos que visam a melhoria da situação social e económica do país.»

(12.4c) «Por outro lado, também **foi aqui questionado** quantas pessoas vão entrar em quarentena domiciliar.»

A ocorrência de passivas, em geral, é frequentemente apontada como uma estratégia do discurso jornalístico, e em particular do noticioso, o que se relaciona com os objetivos comunicativos deste género discursivo e tem consequências ao nível da configuração do fluxo informacional do texto.

Neste contexto, as estruturas passivas eventivas, na sua globalidade, perspetivam a situação a partir de um ponto de partida para a descrição da situação o argumento interno direto, podendo o argumento interno do participio passivo ocorrer como sujeito (passiva pessoal) ou em posição pós-verbal (passiva impessoal). A focalização sobre o processo parece ser a função principal das passivas impessoais, pois a orientação sobre o segundo elemento é impossível. A passiva pessoal não permite essa perspetiva centrada no processo (cf. Estrela, 2013). Verificamos a ocorrência destes dois tipos de passiva nas frases extraídas de ambos os subcorpora, ilustrados em (12.5a – 12.5b) e (12.6a – 12.6b), sendo, por conseguinte, uma estrutura linguística corrente nas duas variedades.

(12.5a) «**O teste deve ser feito** na origem.» (*Telejornal PE – 18*)

(12.5b) «Além de familiares e amigo, **a cerimónia foi presenciada** por diplomatas da embaixada de Angola em Portugal e pela atual administradora da TPA para a área de marketing, Nadir Ferreira.» (*Telejornal PA – 64*)

(12.6a) «**Foram escolhidas** onze personalidades.» (*Telejornal PE – 97*)

(12.6b) «**Foi aprovado** o relatório de atividade de contas, depois de uma longa discussão com dezanove votos a favor e um voto contra.» (*Telejornal PA – 69*)

A opção por uma ou outra pode ser analisada em termos da sua relação com o fluxo informacional do texto (cf. Hawd, 2004). No primeiro caso, o da ocorrência das estruturas passivas pessoais, poderemos, pelo menos em vários dos contextos analisados, associar a sua realização ou não realização lexical ao cumprimento de uma função de tematizar o objeto, contribuindo para a continuidade do tópico. No caso da posposição, a situação é analisada com incidência no próprio processo, sendo o constituinte colocado em posição remática.

No que se refere à ocorrência do complemento agente da passiva, verificamos que, como já foi referido, a não realização desse complemento, opcional em português, domina em ambos os subcorpora, sendo exemplo da sua realização as seguintes frases:

(12.7a) «Algumas lojas começaram a **ser vandalizadas**.» (*Telejornal PA – 40*)

(12.7b) «Parte deste gás **é transformado em LPG e é disponibilizado no mercado**.»
(*Telejornal PA – 48*)

(12.8a) «O ministro britânico dos transportes estava de férias e **acabou por ser apanhado** durante a implementação da quarentena obrigatória no regresso à casa.» (*Telejornal PE – 14*)

(12.8b) «Além do haxixe, **foram apreendidas** cinco viaturas.» (*Telejornal PE – 40*)

Trata-se, como já dissemos, de uma forma básica de ocorrência das passivas eventivas em português, e são várias, as razões que podem motivar a sua não realização, nomeadamente o desconhecimento da entidade denotada pelo argumento agente implícito (cf. (12.7a)), a possibilidade da sua inferência contextual (cf. (12.7b), (12.8b)) ou o interesse em não revelar a sua identidade (12.8a) (cf. Duarte, 2013; Estrela, 2013). Trata-se de razões que, no discurso jornalístico têm sido estudadas, tanto sob o ponto de vista do seu estatuto informacional (cf. Hawd, 2004) quanto no que diz respeito a motivações de natureza mais pragmática ou ideológica (cf. e.o, Dijk 1985).

No caso dos contextos analisados, as duas primeiras razões apontadas parecem ser as que melhor correspondem aos contextos de ocorrência de passivas eventivas curtas. No entanto, é necessária uma análise mais aprofundada que considere outras variáveis para se poder estabelecer com segurança todas as razões subjacentes que podem estar subjacentes a esta preferência.

Igualmente de interesse, na nossa opinião, e a merecer uma análise mais aprofundada em trabalho futuro, é a análise dos contextos em que se produz a realização desse constituinte, o que se verifica igualmente nos dois subcorpora. Nestes casos, e tendo em conta o carácter tipicamente facultativo e mais comum da não realização do constituinte, a sua realização cumpre uma função que, entre outros, pode ser associada à organização informacional do texto. No contexto dos dois subcorpora, verificamos a ocorrência frequente da realização deste complemento através de entidades correspondentes a Pessoa ou Organização, com a função de agente. Ao nível informacional, a sua não omissão pode ser considerada como motivada, no sentido em que fornece informação relevante sobre uma entidade da situação dinâmica descrita, situada em posição focal na oração (cf. Hawd, 2004: 111). Veja-se, a título ilustrativo, os exemplos (12.9a) e (12.9b).

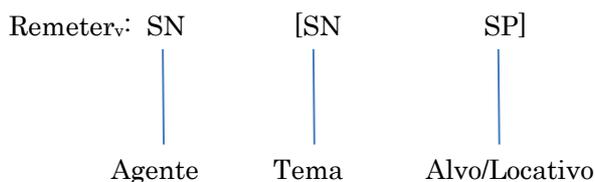
(12.9a) «Nguinamau é **surpreendido pelos efetivos do SIC.**» (*Telejornal PA – 1*)

(12.9b) «Há, neste momento, dois grandes incêndios **a serem combatidos por cerca de 600 bombeiros.**» (*Telejornal PE – 17*)

No corpus de frases extraídas do *Telejornal de Angola*, podemos constatar que há algumas estruturas que apresentam ruturas estruturais, nomeadamente a nível de concordância verbal e nominal, anomalia semântica, ambiguidade, desvios de regência e no uso de estruturas infinitivas, que pretendemos analisar em futuro trabalho. Para ilustrar e porque se trata de casos que se relacionam, ainda que de forma distinta, com o tópico em abordagem, selecionamos, neste capítulo, as construções ocorrentes com o verbo ‘remeter’ e a falta de concordância em algumas frases.

Uso do verbo remeter

Reconhecemos que, quanto à estrutura argumental, o verbo *remeter* é trivalente (12.3a), pois segue o seguinte esquema (Jackendoff, 1991):



1. Agente
2. Tema
3. Alvo/Locativo

Em alguns casos, o predicador verbal *remeter* pode apresentar um predicado binário, conforme vemos em (12.10) retirados do *CETEMPúblico* seguido do código de identificação do exemplo. Visto que o verbo pleno é que determina o número de argumentos e a natureza semântica dos mesmos, constatamos que os dados revelam casos em que este predicador verbal apresenta uma estrutura unária, tornando a estrutura predicativa agramatical, como em (12.11) retirado do *Telejornal de Angola*. Compreendemos que, tendo em conta a regência, poderia ser *remeter a/para/contra*. O mesmo tipo de agramaticalidade é notável em (12.12). Embora se pudesse argumentar que o contexto mais alargado em que é integrada esta frase poderia legitimar esta construção sem que ocorresse agramaticalidade (por exemplo em 12.11 e 12.12) não é isso que se verifica.

(12.10) «Da próxima vez que Guterres for questionado sobre gastos ou sobre porque é que se gasta tão pouco em saúde, poderá **remeter** o impertinente repórter **para** o Bundesbank.» (*par=ext49965-opi-97a-2*)

(12.10b) «Caso sejam detectados indícios de ilícitos susceptíveis de procedimento criminal, o «dossier» **será remetido ao** Ministério Público, para abertura do respectivo inquérito.» (*par=ext591569-eco-95b-1*)

(12.11) «Vão **ser remetidos.**»

(12.12) «Nesse tipo de trabalho, **tem sido retirado.**»

A Fig. 12.3 apresenta os contextos em que o predicador verbal *remeter* é usado.

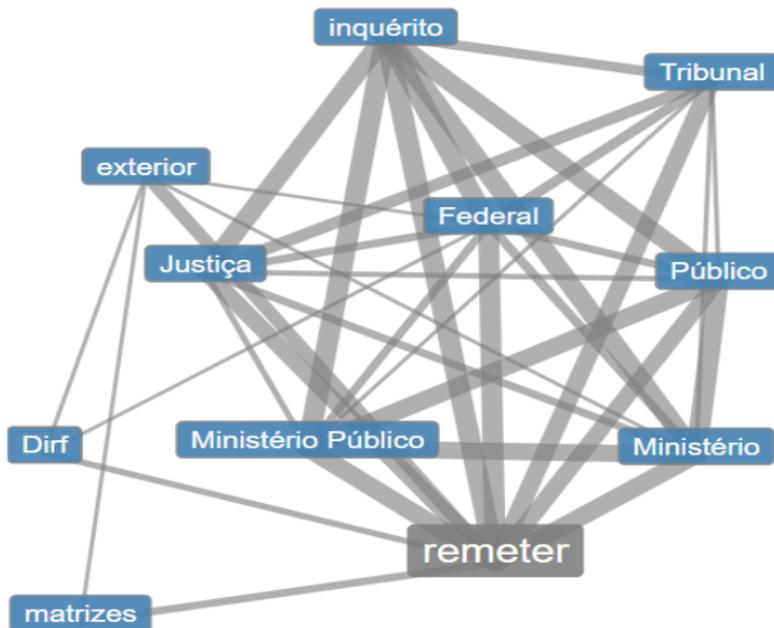


Fig. 12.8 – Contextos do verbo remeter (Leipzig, 2020)

Concordância verbal

Compreende-se que a concordância tem uma base semântica/pragmática e não sintática. Embora a concordância nominal, verbal e predicativa em português sejam assimétricas, uma vez que são, por certo, os valores das propriedades gramaticais em género e número do núcleo nominal, pronominal do sintagma nominal ou verbal que determinam a seleção da forma adequada dos restantes termos concordantes, os dados revelam desvios de concordância, quer no sintagma verbal (Raposo, 2013; Menuzzi, 1994) quer no sintagma nominal em que a concordância é total e uniforme, manifestando-se por constituintes morfológicos (Raposo & Brito, 2013), como se verifica nos exemplos (12.6 a 12.10) do *Telejornal de Angola*.

(12.6) «**Fomos obrigado** também a aliviar algumas medidas por causa da economia que estava a ser sufocada, como exemplo, também para os países, mas estamos a criar medidas de criar os procedimentos para podermos tanto o alívio

como a condição da vida das pessoas pudesse funcionar, mas guardando sempre aquilo que são as recomendações em termos gerais a nível das autoridades sanitárias.»

(12.7) «Que **medidas é que passa a ser tomadas** a partir de agora? (...)»

(12.9) «Esses elementos **estão a ser feita** a perícia de como as pessoas tiveram acesso, mas também começou já ontem no tribunal de Viana o julgamento, elementos que foram encontrados no interior das residências no Zango Zero.»

(12.10) «Em Benguela, **foi apreendido estas pessoas** que invadiram o escritório da Imogestin e subtraíram algumas chaves e comercializavam essas chaves.»

A análise destes e de outros exemplos merecer-nos-á em trabalho futuro de maior análise, também tendo em consideração o subcorpus do jornal. Interessa-nos perceber que razões podem estar subjacentes a este tipo de ocorrências, nomeadamente a modalidade (oral ou escrita), o nível de conhecimento linguístico dos locutores e a processamento da própria construção pelos falantes.

12.3. Síntese do capítulo

Neste capítulo, vimos que os dados revelam que a estrutura de passiva é a mesma em PA e PE, ocorrendo passiva eventiva longa, passiva eventiva curta e passiva estativa. Vimos que algumas razões podem motivar a sua não realização, nomeadamente o desconhecimento da entidade denotada pelo argumento agente implícito, a possibilidade da sua inferência contextual ou o interesse em não revelar a sua identidade. Trata-se de razões que, no discurso jornalístico têm sido estudadas, tanto sob o ponto de vista do seu estatuto informacional no que diz respeito a motivações de natureza mais pragmática ou ideológica. Constatamos que os dados revelam algumas construções com desvios, principalmente estruturais, no Português de Angola quanto ao uso do verbo *remeter* e concordância verbal em estruturas passivas, que deverão ser aprofundados em trabalho futuro.

Capítulo 13 – Conclusões e Trabalho Futuro

Nesta parte, apresentamos as principais conclusões do estudo efetuado. A compreensão da variedade do Português de Angola é essencial para o trabalho numa perspetiva comparativa com o Português Europeu. Visto que o domínio de uma língua é, por conseguinte, o resultado de práticas efetivas, significativas e contextualizadas, esta reflexão revestiu-se de quatro partes principais – enquadramento para o estudo comparativo na área de Léxico e Sintaxe, enquadramento na área de Linguística Computacional, técnicas de *Text Mining* e Processamento de Linguagem Natural em Lexicologia Contrastiva e análise das estruturas passivas com técnicas de *Text Mining* e Processamento de Linguagem Natural.

Partindo deste pressuposto, objetivamos efetuar, primeiro, um estudo comparativo de formas lexicais com vista a extrair empréstimos lexicais dimanantes de línguas de Angola que não constam do léxico do Português Europeu e, segundo, estudo comparativo das estruturas passivas em Português de Angola e Português Europeu com o auxílio de métodos computacionais.

Explicitamos a constituição e caracterização de corpora. Descrevemos o pré-processamento de dados com o auxílio de algumas técnicas de *Text Mining* e Processamento de Linguagem Natural. Em virtude disso, consideramos útil seguir uma abordagem que assenta no uso do pacote *udpipe*, que permite separar as palavras de maneira automática, verificar as categorias sintáticas e gerar as dependências entre elas com o recurso a *dependency parsing*.

Vimos os procedimentos computacionais para a extração de empréstimos lexicais provenientes de Angola encontrados em *A Conjura (2008)*, *Jornal de Angola (2019-2020)* e *Telejornal (2020)*. Procedemos ao estudo comparativo e a extração de empréstimos por classe lexical com o foco para quatro classes lexicais (verbos, nomes, adjetivos e nomes próprios). Com efeito, estes dimanam de várias línguas de Angola. Descrevemos os critérios usados para a determinação do carácter neológico das unidades léxicas. Explicitamos o processamento incremental de texto, que melhora substancialmente a eficiência elidindo a sobrecarga do sistema computacional.

Com o auxílio de métodos computacionais, extraíram-se, por classe lexical, 1.784 empréstimos lexicais dimanantes de línguas de Angola. Estes dividem-se em 66 verbos, 244 nomes, 52 adjetivos e 1.422 nomes próprios, dentre os quais 923 não estão dicionarizados. Além disso, a nossa subsidiação aqui é apresentarmos uma metodologia embasada em métodos computacionais que permite processar outros textos da variedade do Português de Angola para proficientemente identificar novos empréstimos.

Em virtude de a entrada de formas léxicas no português ser um processo natural, explicitamos a determinação da etimologia dos empréstimos lexicais. Apresentamos o seu étimo, etimologia e significação nas listas dos Anexos 1 e 2. Estruturamos os empréstimos em campos lexicais e verificamos os processos de formação dos empréstimos lexicais. Tendo em conta a pertinência da significação dos empréstimos, explicitamos o cotexto como a relação que as unidades linguísticas estabelecem dentro do texto de forma a contribuir para a fixação da significação de uma unidade lexical. Extraímos o cotexto de empréstimos com o auxílio de métodos computacionais. Entretanto, este procedimento viabiliza desencadear processos de desambiguação, estabelecendo relações intertextuais e intratextuais. Verificamos que este estudo pode ser uma subsídio para o vocabulário ortográfico comum do IILP. Na sequência desta análise, apresentamos o protótipo de dicionário de regionalismos angolanos mediante fichas lexicográficas.

Quanto ao nosso segundo objetivo na área de estruturas passivas, consideramos que a passiva eventiva é expressa pelo verbo *ser* como auxiliar. A estrutura temática desta mantém-se; tem, porém, uma estrutura informacional e sintática diferente. As passivas eventivas caracterizam-se por uma componente agentiva e por uma componente eventiva, pois focalizam a fase do processo de mudança de estado, lugar ou posse. As passivas resultativas são expressas pelo verbo *ficar* como auxiliar. Estas são caracterizadas pelo facto de terem uma componente eventiva e ausência de componente agentiva. Com efeito, estas estruturas não admitem geralmente a realização do argumento externo do verbo através de um agente da passiva. As passivas estativas são expressas pelo verbo *estar* como auxiliar e são, na realidade, caracterizadas por não ter componente agentiva nem componente eventiva.

Apresentamos a metodologia usada para a deteção de estruturas passivas em Português de Angola e Português Europeu. Descrevemos os procedimentos computacionais para a extração de passivas eventivas, estativas e resultativas. Destarte, explicitamos a extração de sintagmas com verbo auxiliar e principal. Avaliamos o método computacional proposto, com o auxílio de medidas de classificação. Analisamos as frequências de ocorrência de sintagmas passivos e de estruturas passivas em Português de Angola e Português Europeu. O nosso trabalho indica que o uso de estruturas passivas em Português de Angola é assaz mais frequente do que em Português Europeu. Recorremos a teste estatístico de proporções que comprova a veracidade desta afirmação. Então, por que isso acontece? Será que é efeito de alguma língua autóctone? Tentaremos responder a esta questão no futuro. Uma outra constatação a partir da análise do corpus consistiu em verificar que são usados os mesmos recursos linguísticos para a construção dos tipos de estruturas passivas nas duas variedades. É, no entanto, fundamental o trabalho de análise descritiva e computacional ao

corpus para poder verificar se há outras construções não identificadas nesta fase da investigação.

Recorde-se que estamos, decerto, perante um estudo antessignano, em Angola e em Portugal, que conglera um estudo comparativo de formas lexicais e estruturas passivas em Português de Angola e Português Europeu com vista a identificar empréstimos e diferenças em estruturas passivas com o auxílio de técnicas de *Text Mining* e Processamento de Linguagem Natural. Além disso, o presente estudo comparativo é prógono em apresentar uma metodologia computacional para deteção automática de estruturas passivas eventivas, estativas e resultativas em Português de Angola e Português Europeu.

Trabalho Futuro

No futuro, pretendemos continuar este estudo por ampliar o corpus de extração do Português de Angola e Português Europeu, para a deteção de mais empréstimos lexicais e estruturas passivas, para efetuarmos a constatação de outros aspetos resultantes da análise diferencial nestas duas variedades do português.

Sob esta égide, advogamos que as formas lexicais são o reflexo de realidades histórico-culturais e, como património de valor inestimável, merecem, sob este ponto de vista, alguma reflexão e proteção. Em virtude da necessidade de perquirições sobre as línguas de Angola, tentaremos automatizar de facto a determinação da etimologia dos empréstimos para algumas línguas e engendrar a aplicação do dicionário de regionalismos angolanos, pois, partindo do princípio de que a língua, cultura e história formam um trinómio especial em lexicática, podemos, certamente, relacioná-lo na descrição e compreensão do funcionamento de formas lexicais do Português de Angola e Europeu.

Pelos motivos acima expostos, partimos do pressuposto de que é, de facto, importante perceber que, quanto mais domínio das outras áreas da linguística e não só tiver o lexicólogo, melhor poderá ser o produto que advém do seu labor, porquanto as áreas dela estão seguramente relacionadas com o léxico e onde há uma vontade há um caminho.

No que concerne ao aspeto sintático, além de continuar a investigação sobre as estruturas passivas, contemplando as passivas resultativas e as passivas pronominais, pretendemos expandir o estudo também para a análise diferencial do sintagma nominal e de conectores frásicos nestas duas variedades do português com o auxílio de métodos computacionais. Consideramos que vários fatores concorrem para a mudança do significado das unidades léxicas. Visto que a Sintaxe está intrinsecamente relacionada com a Semântica, se tivermos em consideração o conceito de virtuemá, o qual se compreenda aqui como o

subconjunto do semema que é constituído pelos semas virtuais ou conotativos, pretendemos também efetuar um estudo sobre *word embedding* para verificar diferenças semânticas substanciais de palavras e expressões em Português de Angola e Europeu.

Referências

- Academia das Ciências de Lisboa & Fundação Calouste Gulbenkian. (2001). *Dicionário da Língua Portuguesa Contemporânea*. Lisboa: Verbo.
- Adam, J.-M. (2001). Types de Textes ou Genres de Discours? Comment Classer le Textes qui Disent de et Comment Faire? *Langages*, 10-27.
- Adam, J.-M. (2005). *La Linguistique Textuelle. Introduction à L'analyse Textuelle des Discurs*. Paris: Armand Colin.
- Alexiadou, A., & Anagnostopoulou, E. (2008). Structuring Participles . Em C. Chang, & H. Haynie, *Proceedings of the 26th West Coast Conference on Formal Linguistics* (pp. 33-41). Somerville: Cascadilla Proceedings Project.
- Altuna, R. (2006). *Cultura Tradicional Bantu*. Paulinas: Águeda.
- Alves, I. M. (2002). Neologia Técnico-científica e Análise de Corpus. *Terminologia, Desenvolvimento, e Identidade Nacional Actas do VI Simpósio Ibero-Americano de Terminologia* (pp. 139-149). Lisboa: Colibri.
- Amaro, R., & Mendes, S. (2016). Lexicologia e Linguística Computacional. Em A. M. Martins, & E. Carrilho, *Manual de Linguística Portuguesa* (Vol. 16, pp. 178-199). Berlin/Boston: Walter de Gruyter GmbH.
- António, J., & Osório, P. (2018). História e Variação Morfosintática entre o Quimbundo e o Português na Região de Kwanza Norte, em Angola: Formação do Plural, Concordância e Gênero e Formação do Diminutivo e do Aumentativo dos Substantivos. *Forum lingüístic*, 3289-3302.
- Arim, E., & Freitas, T. (2010). Parassíntese e Conversão: Uma Nova Explicação para um Velho Problema. Em T. Fretas, *Estudos de Corpora - Da Teoria à Prática* (pp. 179-195). Lisboa: Colibri e Instituto de Linguística Teórica e Computacional.
- Aronoff, M., & Anshen, F. (2001). Morphology and the Lexicon: Lexicalization and Productivity. Em A. Spencer, & A. M. Zwicky, *The Handbook of Morphology* (pp. 237-247). Oxford: Blackwell.
- Aronoff, M., & Fudeman, K. (2011). *What is Morphology?* United Kingdom: Wiley-Blackwell Publishing .
- Arquiola, E. F. (2009). Palabras com Estrutura Interna. Em E. De Miguel, *Panorama de la Lexicología* (pp. 51-82). Barcelona: Ariel.
- Baker, M. (2001). Syntax. Em M. Aronoff, & J. Rees-Miller, *The Handbook of Linguistics* (pp. 265-294). Oxford: Blackwell.
- Bakhtin, M. (1999). *A Cultura Popular na Idade Média e no Renascimento: o contexto de François Rabelais*. São Paulo: Hucitec.
- Beard, R. (2001). Derivation. Em A. Spencer, & A. M. Zwicky, *The Handbook of Morphology* (pp. 44-65). Oxford: Blackwell .
- Bechara, E. (1999). *Moderna Gramática Portuguesa*. Rio de Janeiro: Lucerna.
- Bechara, E. (2006). *Gramática Escolar da Língua Portuguesa*. Rio de Janeiro: Lucerna.
- Beedham, C. (2005). *Language and Meaning*. Amsterdam/Filadélfia: John Benjamins.
- Benveniste, É. (1966). *Problèmes de Linguistique Générale I* (Vol. 2). Paris: Gallimard.
- Besnier, J.-M. (1996). *Les Théories de la Connaissance*. S.I.: Flammarion.
- Bijeikienė, V., & Tamošiūnaitė, A. (2013). Quantitative Research: Concepts and definition. Use of corpora . Em V. Bijeikienė, & A. Tamošiūnaitė, *Quantitative and Qualitative Research Methods in Sociolinguistics* (pp. 63-79). Kaunas: Vytautas Magnus University.
- Bonnafous, S., & Tournier, M. (1995). Analyse du Discours, Lexicométrie, Communication et Politique. *Langages* 117, 67-81.

- Bonvini, E. (1994). Angola: Language Situation. Em R. Asher, & J. Simpson, *The Encyclopedia of Language and Linguistics* (pp. 127-128). Oxford, New York, Seoul, Tokyo: Pergamon Press.
- Borer, H. (2005). *Structuring Sense*. Oxford: Oxford University Press.
- Bosque, I. (1999). El Sintagma Adjectival. Modificadores y Complementos del Adjectivo. Adjectivo y Participio. Em I. Bosque, & V. Demonte, *Gramática Descriptiva de la Lengua Española* (pp. 217-310). Madrid: Espasa Calpe.
- Boulanger, J.-C. (1979). Neologie et Terminologie. *Néologie en Marche, serie b: Langue de Spécialités, n° 4, Montreal Office de la Langue Française*.
- Boulanger, J.-C. (2000). Pour dire aujourd'hui. *Infolangue*, 14-15. [ISSN 1206-3975] [RAC].
- Brito, A. M. (2011). Mudança e variação em português: A expressão do objeto indireto. *Cadernos de Literatura Comparada*, 24/25, 27-47.
- Brito, A. M., & Oliveira, F. (1997). Nominalization, aspect and argument structure. Em G. Matos, *Interfaces in linguistic theory*. Lisboa: Colibri.
- Brito, A. M., Duarte, I., & Matos, G. (2003). Tipologia e Distribuição das Expressões Nominais. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 795-867). Lisboa: Caminho.
- Cabré, M. T. (1990). Un Projet Sur La Néologie de Large Difusion – Observatorio de Neologia. *Colóquio de Lexicologia e Lexicografia*. Lisboa: Universidade Nova de Lisboa.
- Cabré, M. T. (1993). *La Terminologia: Teoría, Metodología e Aplicaciones*. Barcelona: Editorial Antartida/Empúries.
- Cabré, M. T. (2002). Una Nueva Teoría de la Terminología: de la Denominación a la Comunicación. *Terminologia, Desenvolvimento, e Identidade Nacional Actas do VI Simpósio Ibero-Americano de Terminologia* (pp. 41-60). Lisboa: Colibri/ Instituto de Linguística Teórica e Computacional.
- Caldas, S. (2016). Lorsque Innovation Linguistique Rime avec Importation Lexicale: Quelques Processus Néologiques d'importation en Portuguais et en Français Contemporain. *L'innovation Lexical dans les Langues Romenes*, 101-118.
- Canosa, X., Varela, X., Lema, P. M., Gamallo, P., Taboada, A. J., & Garcia, M. (2019). Uma Utilidade para o Reconhecimento de Topónimos em Documentos Medievais. *Linguamática*, 3-15.
- Carvalho, D., Marques, M., & Silva, F. (1998). Discurso: Práticas Lexicométricas. *Linguística Computacional: investigação fundamental e aplicações*, 255-262.
- Cielen, D., Meysman, A. D., & Ali, M. (2016). *Introducing Data Science: Big Data, Machine Learning and More, Using Python Tools*. Shelter: Manning Publications.
- Coniam, D. (1993). A Prototype Boundary Marker. Em M. Baker, G. Francis, & E. Tognini-Bonelli, *Text and Technology* (pp. 253-270). Philadelphia/Amsterdam: John Benjamins.
- Correia, M. (2012). *Os Dicionários Portugueses*. Lisboa: Caminho.
- Correia, M., & Lemos, L. (2009). *Inovação Lexical em Português*. Lisboa: Edições Colibri.
- Coseriu, E. (1977). *Principio de Semántica Estructural*. Madrid: Editorial Gredos.
- Costa, T. M. (2015). *Umbundismo no Português de Angola: proposta de um dicionário de umbundismo*. Lisboa: Universidade Nova de Lisboa.
- Dale, R. (2010). Classical Approaches to Natural Language Processing. Em N. Indurkha, & F. Damerau, *Handbook of Natural Language Processing* (pp. 3-7). Boca Raton: Taylor & Francis Group.
- Dang, S., & Ahmad, P. H. (2014). Text Mining : Techniques and its Application. *International Journal of Engineering & Technology Innovations*, 21-25.
- Darmesteter, A. (1967). *Traité de la Formation des Mots Composés de la Langue Française*. Paris: Librairie Honoré Champion.
- Desmet, I. (1990). A Propósito da Neologia Terminológica: O Caso do Empréstimo. *Colóquio de Lexicologia e Lexicografia*. Lisboa: Universidade Nova de Lisboa.

- Desmet, I. (2016). Langues de Spécialité et Foisonnement Néologique en Portugais et en Français: Quelques Réflexions. *L'innovation Lexical dans les Langues Romenes*, 119-136.
- Dias, J. (2003). Caçadores, Artesãos, Comerciantes, Guerreiros: Os Cokwe em Perspetiva Histórica. Em A. Silva, & A. Gonçalves, *Antropologia dos Tshokwe e Povos Aparentados* (pp. 17-47). Porto: Faculdade de Letras da Universidade do Porto.
- Díaz, M. B. (2009). Modelos Estructurales. Em E. Miguel, *Panorama de la Lexicología* (pp. 219-246). Barcelona: Ariel.
- Dijk, T. A. (1985). Structures of News in the Press. Em T. A. Dijk, *Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication* (pp. 69-93). Berlin/New York: Walter de Gruyter.
- Duarte, I. (2003a). Relações Gramaticais, Esquemas Relacionais e Ordem das Palavras. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 275-321). Lisboa: Caminho.
- Duarte, I. (2003b). A Família das Construções Inacusativas. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 507-548). Lisboa: Caminho.
- Duarte, I. (2010). Mudam-se os Tempos, Muda-se a Gramática. Em A. M. Brito, *Gramática: História, Teoria e Aplicações* (pp. 11-28). Porto: Fundação Universidade do Porto.
- Duarte, I., & Oliveira, F. (2010). Particípios Resultativos. Em A. M. Brito, F. Silva, J. Veloso, & A. Fiéis, *Textos Seleccionados, XXV Encontro Nacional da Associação Portuguesa de Linguística* (pp. 397-408). Porto: Associação Portuguesa de Linguística.
- Duarte, I. (2013). Construções Ativas, Passivas, Incoativas e Médias. Em E. B. Raposo, M. F. Nascimento, M. A. Mota, L. Segura, & A. Mendes, *Gramática do Português* (Vol. I, pp. 427-458). Lisboa: Fundação Calouste Gulbenkian.
- Duarte, I., & Brito, A. M. (2003). Predicação e Classes de Predicadores Verbais. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 179-203). Lisboa: Caminho.
- Duarte, I., & Brito, A. M. (2005). Sintaxe. Em I. H. Faria, E. R. Pedro, I. Duarte, & C. Gouveia, *Introdução à Linguística Geral e Portuguesa* (pp. 247-302). Lisboa: Caminho.
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.-B., & Mével, J.-P. (1970). *Dictionnaire de Linguistique*. Paris: Librairie Larousse.
- Edmundo, F. (11 de março de 2020). *Sobre a aprendizagem das línguas nacionais, em Angola*. Obtido de Ciberdúvidas da Língua Portuguesa: <https://ciberduvidas.iscte-iul.pt/artigos/rubricas/lusofonias/sobre-a-aprendizagem-das-linguas-nacionais-em-angola/3032>.
- Edureka (Realizador). (2019). *Text Mining In R, Natural Language Processing, Data Science Certification Training* [Filme].
- Eliseu, A. (2008). *Sintaxe do Português*. Lisboa: Caminho.
- Embick, D. (2004). On the Structure of Resultative Participles in English. *Linguistic Inquiry*, *Volume 35, Number 3*, 355–392.
- Endruschat, A., & Schmidt-Radefeldt, J. (2015). *Introdução Básica à Linguística do Português*. Lisboa: Edições Colibri.
- Estrela, A. P. (2013). *A Aquisição da Estrutura Passiva em Português Europeu*. Lisboa: Tese de Doutoramento. Universidade Nova de Lisboa.
- Ethnologue. (6 de março de 2020). *Angola*. Obtido de Ethnologue: <https://www.ethnologue.com/map/AO>.
- Fabb, N. (2001). Compounding. Em A. Spencer, & A. Zwicky, *The Handbook of Morphology* (pp. 66-83). Oxford: Blackwell.
- Faulstich, E. (2002). Entre a Sincronia e a Diacronia: Variação Terminológica no Código e na Língua. *Terminologia, Desenvolvimento, e Identidade Nacional Actas do VI Simpósio*

- Ibero-Americano de Terminologia* (pp. 61-74). Lisboa: Colibri/Instituto de Linguística Teórica e Computacional.
- Feinerer, I. (30 de dezembro de 2019). *Introduction to the tm Package Text Mining in R*. Obtido de cran.r-project.org: <https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>
- Ferreira, F. (2 de março de 1854). Memórias sobre Cassange. *Boletim do Conselho Ultramarino*.
- Fonseca, D. (2012). As Línguas Nacionais e o Prestigioso Português em Angola. *Simpósio Internacional de Ensino de Língua Portuguesa. Volume 2, Número 1*.
- Foucault, M. (1966). *As Palavras e as Coisas: Uma Arqueologia das Ciências Humanas*. Lisboa: Edições 70.
- Foucault, M. (1969). *A Arqueologia do Saber*. Lisboa: Edições 70.
- Foucault, M. (1997). *A Ordem do Discurso. Aula Inaugural no Collège de France, pronunciada em 2 de Dezembro de 1970*. Lisboa: Relógio D'Água Editores.
- Freitas, T., Ramilho, M. C., & Arim, E. (2010). Os Neologismos nos Meios de Comunicação Social Portuguesa. Em T. Freitas, *Estudos de Corpora - Da Teoria à Prática* (pp. 129-144). Lisboa: Colibri e Instituto de Linguística Teórica e Computacional.
- Freitas, T., Ramilho, M. C., & Arim, E. (2010a). O Processo de Integração dos Estrangeirismos no Português Europeu. Em T. Freitas, *Estudos de Corpora - Da Teoria à Prática* (pp. 145-162). Lisboa: Colibri e Instituto de Linguística Teórica e Computacional.
- Gaatone, D. (1998). *Le Passif en Français*. Paris/Bruxelles: Editions Duculot.
- Gama, J. (2008). *Knowledge Discovery from Data Streams*. Porto: Universidade do Porto.
- Gama, J., Faceli, A., Lorena, A. C., & Oliveira, M. (2017). *Extração de Conhecimento de Dados: Data Mining*. Lisboa: Edições Sílabo.
- Gamallo, P. (s.d). *Léxico para Português Europeu*. Santiago de Compostela: Univesidade de Santiago de Compostela.
- Gamallo, P., & Garcia, M. (2011). A Resource-Based Method for Named Entity Extraction and Classification. Em L. Antunes, & S. Pinto, *Progress in Artificial Intelligence* (pp. 610–623). 15th Portuguese Conference on Artificial Intelligence (EPIA), vol. 7026: Springer.
- Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Giraldo Ortiz, J. (2010). Hacia una Revisión del Concepto de Siglación. *Panace@: 11 (31)*, 70-76.
- Gross, G. (1998). Degré de Figement des Noms Composés. *Langages n° 90*, 57-72.
- Gross, M. (1997). The Construction of Local Grammars. Em E. Roche, & Y. Schabes, *Finite-state language processing, Language, Speech, and Communication* (pp. 329–354). Cambridge: Mass.
- Guilbert, L. (1975). *La Créativité Lexicale*. Paris: Larousse.
- Guthrie, M. (1971). *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. London: Gregg International Publishers.
- Hagemeijer, T. (2016). O Português em Contacto em África. Em A. M. Martins, & E. Carrilho, *Manual de Linguística Portuguesa* (pp. 43-67). Berlin/Boston: Walter de Gruyter.
- Hall, C. (2000). Prefixation, suffixation and circumfixation. Em G. Booij, C. Lehmann, J. Mugdan, W. Kesselheim, & S. Skopeteas, *Morphology. An international handbook on inflection and word formation. Handbooks of Linguistics and Communication Science* (pp. 535-545). Berlin: De Gruyter.
- Halliday, M. (2004). Lexicology. Em M. Halliday, W. Teubert, C. Yallop, & A. Cermáková, *Lexicology and Corpus Linguistics*. London: Continuum.
- Hartmann, R., & James, G. (2001). *Dictionary of Lexicography*. London: Routledge.
- Haspelmath, M. (2002). *Understanding Morphology*. London: Arnold.

- Hawd, H. F. (2004). A Voz Verbal e o Fluxo Informacional do Texto. *D.E.L.T.A*, 20(1), 97-121.
- Henriques, C. (2007). *Morfologia: Estudos Lexicais em Perspetiva Sincrônica*. Rio de Janeiro: Elsevier.
- Hippisley, A. (2010). Lexical Analysis. Em N. Indurkha, & F. J. Damerou, *Handbook of Natural Language Processing* (pp. 31-58). Boca Raton: Chapman & Hall.
- Hock, H., & Joseph, B. (1996). *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.
- Hutchins, W. J., & Somers, H. L. (1995). *Introducción a la Traducción Automática*. Madrid: Visor.
- Inverno, L. (2008). Transição de Angola para o Português: uma história sociolinguística. Em L. Torgal, F. Pimenta, & J. Sousa, *Comunidades Imaginadas* (pp. 169-181). Coimbra: Imprensa da Universidade de Coimbra.
- Inverno, L. (2009). *Contact-induced Restructuring of Portuguese Morphosyntax in Interior Angola: Evidence from Dundo*. Coimbra: Universidade de Coimbra.
- Inverno, L. (2018). Angolan Portuguese: Its historical development and current sociolinguistic setting. Em L. Álvarez López, P. Gonçalves, & J. Ornelas de Avelar, *The Portuguese Language Continuum in Africa and Brazil* (pp. 111-134). Amsterdam/Philadelphia: John Benjamins.
- Iriguti, A., & Feltrim, V. (2019). Evaluating Features for Rhetorical Structure Classification in Scientific Abstracts. *Linguamática*, 41-53.
- Jackendoff, R. (1991). Grammatical Relations and Functional Structure. Em M. H. Campos, & M. F. Xavier, *Sintaxe e Semântica do Português* (pp. 131-154). Lisboa: Universidade Aberta.
- Jackendoff, R. (2002). *Foundations of Language (Brain, Meaning, Grammar, Evolution)*. Oxford: Oxford University Press.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Londres: Longman.
- Konkol, I. M. (2015). *Named Entity Recognition*. Pilsen: PhD thesis, University of West Bohemia.
- Kristeva, J. (1969). *Introdução à Seminálise*. São Paulo: Debates.
- Labov, W. (1999). *Principles of Linguistic Change*. Oxford: Blackwell.
- Lehman, A., & Martin-Berthet, F. (2000). *Introduction à La Lexicologie Sémantique et Morphologie*. Paris: Éditions Nathan.
- Leipzig, U. (4 de 09 de 2020). <https://wortschatz.uni-leipzig.de/en/>. Obtido de wortschatz.uni-leipzig: https://corpora.uni-leipzig.de/en/res?corpusId=por-br_newscrawl_2011&word=remeter.
- Lerat, P. (1989). Les Fondements Theoriques de La Terminologie. *La Banque des Mots*, 52-62.
- Lewandowski, T. (1995). *Diccionario de Lingüística*. Madrid: Cátedra.
- Lieber, R. (1992). *Deconstructing Morphology*. Chicago: The University of Chicago.
- Lightfoot, D. (2006). *How New Languages Emerge*. Cambridge: Cambridge University Press.
- Lino, M. T. (1979). Importância de uma Lexicologia Constrativa. *Letras Soltas 1*, 11-16.
- Lino, M. T. (1984). Análise Léxico-Semântica. *Letras Soltas 2*.
- Lino, M. T., & Dechamps, C. (2016). Langue Juridique et Créativité Terminologique: une perspective français-portugais. *L'innovation Lexicale dans les Langues Romanes*, 83-99.
- Lino, M. T., Mocho, M., Costa, M. R., & Desmet, I. (1991). *Terminologia da Lexicologia, da Lexicografia, da Terminologia e da Terminografia*. Lisboa: Cosmo.
- Llamazares, M. (2008). Lingüística con Corpus. *Filología*, 329-349.
- Lopes, A. C., & Carapinha, C. (2013). *Texto, Coesão e Coerência*. Coimbra: Almedina.

- López Rúa, P. (2010). *English Acronyms and Alphabetisms. A Prototypebased Approach with Special Reference to Their Method of Formation, Realization, and Connections with Other Morphological Devices*. . Santiago de Compostela: Facultad de Filosofía.
- Luschützky, H. C. (2000). Morphem, Morph und Allomorph. Em G. Booij, C. Lehmann, J. Mugdan, & S. Skopeteas, *Morphologie/Morphology. An International Handbook on Inflection and Word-Formation* (pp. 451-462). Berlin/New York: Walter de Gruyter.
- Lynch, D. A. (6 de março de 2020). *Colonization*. Obtido de Lynch's PSGS Hub: <https://sites.google.com/a/smumn.edu/lynch/maps-globalization-non-trade/colonization>.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- Maingueneau, D. (1997). *Novas Tendências em Análise do Discurso*. São Paulo: Universidade Estadual de Campinas.
- Marques, M. M. (1998). *Passivas com Verbos Psicológicos da Família de Preocupar*. Porto: Universidade do Porto.
- Martinet, A. (1967). Sintagme et Synthème. *Linguistique 2*.
- Mascarenhas, F. (2008). *Memórias de Icolo e Bengo: Figuras e Família*. Luanda: EAL/Edições de Angola.
- Mateus, M. H., & Cardeira, E. (2007). *Norma e Variação*. Lisboa: Caminho.
- Matos, J. M. (9 de Dezembro de 1921). Decreto n° 77. *Boletim Oficial de Angola*, pp. n°50, 1ª Série.
- Matthews, P. H. (1991). *Morphology*. Cambridge: Cambridge University Press.
- Mejri, S. (2009). Du morphème à la signification : le flou définitionnel des catégories linguistiques. *L'information grammaticale 122*, 4-9.
- Mel'čuk, I. (2000). Morphological Processes. Em G. L. Booij, *Morphology. An international handbook on inflection and word formation* (pp. 523-535). Berlin: De Gruyter.
- Mel'čuk, I. (2006). Colocaciones en el Diccionario. Em M. Alonso Ramos, *Diccionarios y Fraseología* (pp. 11-43). A Coruña: Universidade de Coruña.
- Mendes, A. (1994). *Análise Sintáctica dos Verbos Psicológicos do Português*. Lisboa: Dissertação de Mestrado. Faculdade de Letras da Universidade de Lisboa.
- Mendes, A., & Estrela, A. (2008). Constructions with Se in African Varieties of Portuguese. *Phrasis*, 83-107.
- Mendevil Giró, J. L. (2009). Palabras con Estructura Externa. Em E. De Miguel, *Panorama de la Lexicología* (pp. 83-113). Barcelona: Ariel.
- Mendikoetxea, A. (1999). Construcciones con Se: Medias, Pasivas e Impersonales. Em I. Bosque, & V. Demonte, *Gramática Descriptiva de la Lengua Espanhola* (pp. 1631-1722). Madrid: Espasa Calpe.
- Menuzzi, S. (1994). Adjectival positions inside DP. Em R. Bok-Bennema, & C. Cremers, *Linguistics in the Netherlands 11* (pp. 127-138). Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Miguel, A. (2019). *Integração Morfológica e Fonológica de Empréstimos Lexicais Bantos no Português Oral de Luanda*. Lisboa: Tese de Doutoramento. Universidade de Lisboa.
- Miguel, E. (2009). La Teoría del Lexicón Generativo. Em E. Miguel, *Panorama de la Lexicología* (pp. 337-368). Barcelona: Ariel.
- Mingas, A. (2000). *Interferência do Kimbundu no Português Falado em Lwanda*. Luanda: Chá de Caxinde.
- Miranda, P. (2009). Neología y Pérdida Léxica. Em E. Miguel, *Panorama de la Lexicología* (pp. 133-158). Barcelona: Ariel.
- Móia, T. (2016). Semântica e Pragmática. Em A. Martins, & E. Carrilho, *Manual de Linguística Portuguesa* (pp. 308-344). Berlin/Boston: Walter de Gruyter GmbH.
- Mota, M. A. (2013). Introdução à Morfologia. Em E. B. Raposo, M. F. Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 2787-2831). Lisboa: Fundação Calouste Gulbenkian.

- Mota, M. A. (2013a). Introdução à Morfologia. Em E. B. Raposo, M. F. Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 2787-2831). Lisboa: Fundação Calouste Gulbenkian.
- Mota, M. A. (2013b). Morfologia do Nome e do Adjetivo. Em E. B. Raposo, M. F. Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 2833-2930). Lisboa: Fundação Calouste Gulbenkian.
- Muhongo, T. (2017). *Empréstimos de Origem Angolana em Voz de Angola Clamando no Deserto*. Lisboa: Dissertação de Mestrado/ FCSH - UNL.
- Muhongo, T. (2019a). Sobre Valores de “se” no Português Europeu e no Português em Angola. *Diseminaciones. Revista de Investigación y Crítica en Humanidades y Ciencias Sociales*, 2(3), 123-141.
- Muhongo, T. (2019b). Sobre Dialogismo e Construção da Força Discursiva em Solémnia Verba. *Portuguese Language Journal* (13), 164-183.
- Neto, M. (2006). *Os Sacrificados de Icolo e Bengo*. Luanda: Instituto Nacional do Livro e do Disco (INALD).
- Neto, M. (2012). *Aproximação Linguística e Experiência Comunicacional: O Caso da Escola de Formação Garcia Neto*. Luanda: Mayamba.
- Oliveira, F. (2003). Tempo e Aspecto. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 127-178). Lisboa: Caminho.
- Oliveira, P. F., Guerra, S., & McDonnell, R. (2018). *Ciência de Dados com R – Introdução*. Brasília : IBPAD.
- Palmer, D. (2010). Text Preprocessing. Em N. Indurkha, & F. J. Damerau, *Handbook of Natural Language Processing* (pp. 9-30). Boca Raton: Chapman & Hall/CRC.
- Pavel, S., & Nolet, D. (2002). *Manual de Terminologia*. Canadá: Direção de Terminologia e Normalização Departamento de Tradução do Governo Canadense.
- Pereira, I. (2013). Formação de verbos. Em G. Rio-Torto, A. S. Rodrigues, I. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 265-319). Coimbra: Imprensa da Universidade de Coimbra.
- Pereira, I. (2016). Processos de Construção não Concatenativa. Em G. Rio-Torto, A. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 521-553). Coimbra: Imprensa da Universidade de Coimbra.
- Pereira, R. (2016). Formação de Verbos. Em G. Rio-Torto, A. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 297-355). Coimbra: Imprensa da Universidade de Coimbra.
- Peres, J. A., & Mória, T. (1995). *Áreas Críticas da Língua Portuguesa*. Lisboa: Caminho.
- Perini, M. (2005). *Gramática Descritiva do Português*. São Paulo: Ática.
- Piana, C. F., Machado, A. A., & Selau, L. P. (2009). *Estatística Básica*. Pelotas: Universidade Federal de Pelotas .
- Picoche, J. (1977). *Précis de Lexicologie Francese*. Paris: Nathan Université.
- Piera, C. (2009). Una Idea de la Palabra. Em E. d. Miguel, *Panorama de la Lexicología* (pp. 25-49). Barcelona: Ariel.
- Pinto, A., Alves, A., & Oliveira, H. (2016). Comparing the Performance of Different NLP Toolkits in Formal and Social Media Text. *5th Symposium on Languages, Applications and Technologies (SLATE)*, vol. 51 (pp. 3:1–3:16). Mariboru: Elsevier.
- Pires, C. (6 de março de 2020). *Povos e Línguas de Angola*. Obtido de cpires.planetaclix: http://cpires.planetaclix.pt/cpires_angola_povos.html
- Plag, I. (2003). *Word-formation in English*. Cambridge : Cambridge University Press.
- Portal Action. (27 de Maio de 2020). Obtido de Portal Action: <http://www.portalaction.com.br/tabela-de-contingencia/teste-de-mcnemar-para-frequencias-correlacionadas>
- Pruvost, J., & Sablayrolle, J. (2003). *Le Néologismes*. Paris: Presse Universitaire de France.

- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge: Massachusetts Institute of Technology.
- Pustejovsky, J., & Boguraev, B. (1994). Lexical Knowledge Representation and Natural Language Processing. Em F. C. Pereira, & B. J. Grosz, *Natural Language Processing* (pp. 193-223). Cambridge: The MIT Press.
- Raposo, E. B. (1992). *Teoria da Gramática. A Faculdade da Linguagem*. Lisboa: Caminho.
- Raposo, E. B. (2013). Concordância Verbal. Em E. B. Raposo, M. F. Bacelar do Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 2425-2495). Lisboa: Fundação Calouste Gulbenkian.
- Raposo, E. B., & Brito, A. M. (2013). Concordância Nominal. Em E. B. Raposo, M. F. Bacelar do Nascimento, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 2499-2509). Lisboa: Fundação Calouste Gulbenkian.
- Reinhart, T., & Siloni, T. (2005). The Lexicon-Syntax Parameter: Reflexivization and Other Arity Operations. *Linguistic Inquiry*, 389-436.
- Rey, A. (1970). *La Lexicologie*. Paris: Klincksieck.
- Rey, A. (1977). *Le Lexique: Image et Modèles*. Paris: Armand Colin.
- Ribas, Ó. (2014). *Dicionário de Regionalismos Angolanos*. Lisboa: Mercado de Letras.
- Ribeiro, S. I. (2011). *Estruturas com Se Anafórico, Impessoal e Decausativo em Português*. Coimbra: Tese de Doutoramento. Faculdade de Letras da Universidade de Coimbra.
- Rio-Torto, G. (1993). *Formação de Palavras em Português. Aspectos da construção de avaliativos*. Coimbra: Tese de Doutoramento: Universidade de Coimbra.
- Rio-Torto, G. (2013). Derivação. Em E. B. Raposo, M. F. Bacelar do Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 3029-3152). Lisboa: Fundação Calouste Gulbenkian.
- Rio-Torto, G. (2013). Formação de Nomes. Em G. Rio-Torto, A. Rodrigues, I. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 117-211). Coimbra: Imprensa da Universidade de Coimbra.
- Rio-Torto, G. (2016a). Formação de Avaliativos. Em G. Rio-Torto, A. S. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 357-389). Coimbra: Imprensa Universidade de Coimbra.
- Rio-Torto, G. (2016b). Prefixação. Em G. Rio-Torto, A. S. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 411-459). Coimbra: Imprensa da Universidade de Coimbra.
- Rio-Torto, G., & Ribeiro, S. (2016). Composição. Em G. Rio-Torto, A. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (p. 461). Coimbra: Imprensa da Universidade de Coimbra.
- Rocha, C., Jorge, A., Sionara, R., Brito, P., Pimenta, C., & Rezende, S. (2016). *PAMPO: using pattern matching and pos-tagging for effective Named Entities recognition in Portuguese*. arXiv:1612.09535.
- Rocha, M., & Ferreira, P. (2017). *Análise e Exploração de Dados com R*. Lisboa: FCA.
- Rodrigues, A. (2013). Introdução. Em G. Rio-Torto, A. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 29-116). Coimbra: Imprensa da Universidade de Coimbra.
- Rodrigues, A. (2016). Noções Basilares sobre a Morfologia e o Léxico. Em G. Rio-Torto, A. S. Rodrigues, I. Pereira, R. Pereira, & S. Ribeiro, *Gramática Derivacional do Português* (pp. 35-134). Coimbra: Imprensa da Universidade de Coimbra.
- Rosa, M. (2006). *Introdução à Morfologia*. São Paulo : Contexto.
- Sablajrolles, J.-F. (2000). *La Néologie en Français Contemporain: Examen du Concept et analyse de Productions Néologiques Récentes*. Paris: Honoré Champion Éditeur.
- Sablajrolles, J.-F. (2016). Prolégomènes aux Analyses Néologiques Contrastive. *L'innovation Lexical dans les Langues Romenes*, 71-82.

- Sacanene, B. (2019). Análise dos Angolanismos no Dicionário da Língua Portuguesa Contemporânea. *Diacrítica Journal*, 485-503.
- Sager, J. (1990). *Curso Práctico sobre el Procesamiento de la Terminología*. Madrid: Fundación Germán Sánchez Ruipérez.
- Sanromán, A. (2001). *A Unidade Lexicográfica: Palavras, Colocações, Frasemas, Pragmatemas*. Braga: Centro de Estudos Humanísticos da Universidade do Minho.
- Sanromán, A. (2012). O Dicionário de Tradução: um repositório de amostras de significados. Em O. D. Fouces, *Olhares & Miradas. Reflexiones sobre la traducción portugués-español y su didáctica* (pp. 137-151). Granada: Editorial Atrio.
- Schmid, H. (2010). Decision Trees. Em A. Clark, C. Fox, & S. Lappin, *The Handbook of Computational Linguistics and Natural Language Processing* (pp. 180-196). Chichester: Wiley-Blackwell.
- Silva, A. P. (2015). *Lexicografia Bilingue de Especialidade: e-dicionário de português-kimbundu no domínio da saúde*. Lisboa: Universidade Nova de Lisboa.
- Silvestre, J. P. (2016). Lexicografia. Em A. M. Martins, & E. Carrilho, *Manual de Linguística Portuguesa* (pp. 200-223). Berlin/Boston: Walter de Gruyter GmbH.
- Sinclair, J. (2005). Corpus and Text — Basic Principles. Em M. Wynne, *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books.
- Spencer, A. (2001). Morphology. Em M. Aronoff, & J. Rees-Miller, *The Handbook of Linguistics* (pp. 213-237). Oxford: Blackwell.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Teixeira, C., Fuccio, R., & Oliveira, T. (27 de Maio de 2020). *Universidade Federal do Paraná*. Obtido de www.leg.ufpr.br: http://www.leg.ufpr.br/lib/exe/fetch.php/disciplinas:ce001:teste_de_mcnemar_pronto.pdf.
- Torgo, L. (2009). *A Linguagem R: Programação para a Análise de Dados*. Lisboa: Escolar Editora.
- Torrueja, J., & Llisterri, J. (1999). Diseño de corpus textuales y orales. Em E. Milenio (Ed.), *Seminari de Filologia i Informàtica, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona* (pp. 45-77). Barcelona: Editorial Milenio.
- Trask, R. (1993). *A Dictionary of Grammatical Terms in Linguistic*. London: Routledge.
- Ullmann, S. (1964). *Semântica: Uma Introdução à Ciência do Significado*. Lisboa: Fundação Calouste Gulbenkian.
- United Nations. (6 de março de 2020). *Angola - The United Nations*. Obtido de United Nation: <https://www.un.org/Depts/Cartographic/map/profile/angola.pdf>
- Veloso, J. (2016). *Verba Manent: A Palavra como Unidade Pertinente para a Descrição Linguística do Português e de outras Línguas Flexionais*. Porto Alegre: EDIPUCRS.
- Vilela, M. (1979). *Estruturas Léxicas do Português*. Coimbra: Almedina.
- Vilela, M. (1994). *Estudos de Lexicologia do Português*. Coimbra: Almedina.
- Villalva, A. (2000). *Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português*. Lisboa: Fundação Calouste Gulbenkian.
- Villalva, A. (2003a). Estrutura Morfológica Básica. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, & F. Oliveira, *Gramática da Língua Portuguesa* (pp. 917-938). Lisboa: Caminho.
- Villalva, A. (2003b). Formação de Palavras: Afixação. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 939-983). Lisboa: Caminho.
- Villalva, A. (2003c). Formação de Palavras: Composição. Em M. H. Mateus, A. M. Brito, I. Duarte, I. H. Faria, S. Frota, G. Matos, . . . A. Villalva, *Gramática da Língua Portuguesa* (pp. 969-983). Lisboa: Caminho.

- Villalva, A. (2013). Composição. Em E. B. Raposo, M. F. Bacelar do Nascimento, M. A. Mota, L. Segura, A. Mendes, & A. Andrade, *Gramática do Português* (Vol. III, pp. 3153-3210). Lisboa: Fundação Calouste Gulbenkian.
- Villalva, A., & Silvestre, J. P. (2014). *Introdução ao Estudo do Léxico: Descrição e Análise do Português*. Rio de Janeiro: Vozes.
- Wüster, E. (1998). *Introducción a la Teoría General de la Terminología y a la Lexicografía Terminológica*. Barcelona: Institut de Lingüística Aplicada (IULA).
- Wasow, T. (2001). Generative Grammar. Em M. Aronoff, & J. Rees-Miller, *The Handbook of Linguistics* (pp. 295-318). Oxford: Blackwell.
- Weinrich, H. (1979). A Verdade dos Dicionários. Em M. Vilela, *Problemas da Lexicologia e Lexicografia* (p. 314). Porto: Livraria Civilização.
- Weiss, S., Indurkha, N., & Zhang, T. (2015). *Fundamentals of Predictive Text Mining*. London: Springer.
- Welbers, K., Atteveldt, W. V., & Benoit, K. (2017). Text Analysis in R. *Routledge*, 11, 245–265.
- Wickham, H. (2015). *Advanced R*. Boca Raton: Taylor & Francis.
- Wickham, H., & Grolemund, G. (2017). *R for Data Science : Import, Tidy, Transform, Visualize, and Model Data*. Sebastopol: O'Reilly.
- Wijffels, J. (2019). *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*. R package version 0.8.3. <https://CRAN.R-project.org/package=udpipe>.

Anexos da Tese

Anexo I – Cotexto de empréstimos nominais, verbais e adjetivais

Nº	Empréstimo	Étimo	Etimologia	Significação	Doc	Frase	Cotexto	Equivalência
1	agindungado	jindungo	kimbundu	<i>adj.</i> 1. Excitado com jindungo; malicioso. 2. <i>Fig.</i> Alterado.	1	118	rumores bem saborosos agindungados	rumores alterados sobre uma celebridade
2	aka	aka	umbundu	1. <i>Interj.</i> Exclamação que exprime espanto, surpresa *(oh, ah, caramba, etc). 2. <i>adj. determ. e pron. demomonstr.</i> 3. Esta; este. 4. Abreviação de <i>kaka</i> . 5. <i>n. f.</i> Arma de fogo.	2	522	aka , mano	caramba , mano!
3	akixi*	akixi	cokwe	1. <i>adj.</i> Mascado. 2. <i>n. m.</i> Ídolo de pau. 3. O mesmo que <i>mukixi</i> . 4. Fetiche.	2	38 43	iniciar akixi mulaza de uma akixi	iniciar o fetiche mulaza de um ídolo de pau
4	ambaquense	mbaka	kimbundu	<i>adj.</i> 1. Indivíduo natural da região de Ambaca; ambaquista. 2. Que reclama muito.	1	1251	mulata ambaquense	mulata de Ambaca
5	andua	ndua	kimbundu	<i>n. f.</i> Ave de plumagem tirante à do papagaio, olhos vermelhos e corpulência.	1	9	transforme em andua	transforme em ave corpulenta
6	atumba	atundua	nganguela	<i>n. f.</i> 1. Erva vivaz, rizomatosa. 2. Fruto de uma baga ovóide de 7cm, de pericarpo rijo, carnudo, grosso, aromático e dum vermelho vivo na maturação, mesocarpo polposo, branco, agridoce, comestível e muito procurado pelo indígena. 3. Semente preta.	2	50	ração de atumba	ração de erva rizomatosa
7	axiluandar*	akuaxiluanda	kimbundu	1. <i>v.</i> Torna-se natural da ilha adjacente à cidade de Luanda. 2. <i>p.ext.</i> Ser pescador dessa ilha. 3. <i>adj. unif.</i> Relativo a essa população. 4. Plural de muxiluanda.	2	6	superior em axiluandar	superior por ser natural de Luanda
8	baçular	kubasula	kimbundu	<i>v. intr.</i> 1. Aplicar a bassula. 2. Executar com perícia essa manobra de briga.	1	305	baçular vertiginosamente raciocínios	aplicar com perícia os raciocínios

9	bambila*	mbambi	kikongo	<i>n. m.</i> 1. Cabra selvagem. 2. Apito de chifre desse antílope.	2	23	passos de bambila	passos de cabra selvagem
10	bangão	kudibanga	kimbundu	<i>adj. e n. m.</i> Pessoa que tem presunção, vaidade, ostentação, garbo, elegância.	2	44	bangão marca sempre	elegante marca sempre
11	bassula	kubasula	kimbundu	<i>n. f.</i> 1. Golpe de briga que consiste em derrubar o adversário, erguendo-o e arremessando-o ao chão. 2. Surra.	2	42	foi uma bassula	foi uma surra
12	bembe	mbêmba	kimbundu	<i>n. f.</i> 1. Ave de rapina semelhante à águia, mas de plumagem branca e preta. 2. Açor.	2	93	era bembe	era ave de rapina
13	bengalim*	engalim	kimbundu	<i>n. m.</i> Pássaro de cor vermelha.	1	1158	vestido de bengalim	vestido como pássaro de cor vermelha
14	benguelinha*	kubengeleka	kimbundu	<i>n. f.</i> Acompanhante de cama de alguém. Meretriz.	1	69 106	pímulas, benguelinhas amizade do marido por uma benguelinha	pímulas, meretrizes amizade do marido por uma meretriz
15	benje*	mbeji	kimbundu	<i>n. m.</i> 1. Satélite da terra. 2. Lua. 3. Espaço durante o qual a lua faz a sua revolução em torno da terra. 4. Tempo compreendido entre os dois novilúnios.	2	402	10.875 benje	10.875 satélites
16	bessangana	besángana	kimbundu	<i>n. f.</i> 1. Mulher distinta, da esfera do traje de panos. 2. Senhora africana da classe dessa vestimenta. 3. Traje tradicional.	2	55	grato à bessangana	grato pelo traje tradicional
17	bilar	kubila	kimbundu	<i>v. intr.</i> Lutar, brigar, pelejar.	2	9 42	favorável bilar acções de bilar	favorável lutar acções de lutar
18	bofeta	bukufeta	kimbundu	<i>n. f.</i> 1. Pano preto com função de capa, que cobre da cabeça ao meio da perna, descaindo as pontas de cima para a frente. 2. Anágua.	1	56 1205	rasgou a sua bofeta enfim as graves bofetetas	rasgou a sua anágua enfim as graves anáguas
19	buange*	mbuanga	kimbundu	<i>n. m.</i> 1. Peixe miúdo teleósteo silúrida. 2. Bagre pequeno abundante no rio Kwanza. 3. Engano; equívoco.	2	6	pescavam buange	pescavam peixe miúdo

20	builar	kubuila	kimbundu	<i>v. intr.</i> 1. Urrar. 2. Rugir. 3. Soltar urros ou uivos.	1	342	buila mu nvunda	soltar urros na luta
21	bulunganga	ombulunga	nhaneke	<i>n. f.</i> 1. Espécie de cerveja. 2. O mesmo que <i>txicundo</i> entre os Humbes e Handas.	1	49	era bulunganga	era espécie de cerveja
22	bundo	mbundu	kimbundu	<i>n. m.</i> 1. Peixe do mar semelhante à tainha, mas de menor porte. 2. <i>Pl. híbr.</i> Jimbundos.	2	22	na compra de bundos	na compra de peixe similar à tainha
23	buxila	mbuxila	kimbundu	<i>n. m. e f.</i> 1. Filho ou filha de escravos. 2. Filho ou filha de escrava com homem livre, mas nascida na casa em que serve. 3. Criolo.	1	1381 1428 1437 1828	paixão por uma buxila tímido e a buxila cólera contra a inocente buxila embarque da buxila	paixão por filha de escravos tímido e a filha de escravo cólera contra a inocente filha de escravo embarque da filha de escravo
24	cabingado*	kambinda	kikongo	<i>adj.unif.</i> Que tem feições de Cabinda.	2	14	negligenciaram cabingado	negligenciaram o de feições de Cabinda
25	cabucado*	kabucadu	kimbundu	<i>n. m.</i> 1. Porção de alimento que se pode colocar de uma só vez na boca; aquilo que se trinca de uma só vez. 2. Fragmento ou pequena quantidade de um corpo, de uma. Substância separada do todo. 3. Pouco.	2	65	cabucado de história	um pouco de história

26	cachimbo	kishimba	kimbundu	<i>n. m.</i> 1. Objeto usado pelo fumador, constituído por um recipiente onde arde o tabaco ou outra substância fumável, e. por um tubo por onde se aspira o fumo. 2. Buraco do castiçal onde se encaixa a vela. 3. Peça de ferro fêmea da dobradiça onde entra e gira o gonzo ou o espigão dos lemes das portas ou das janelas. 4. Náut. Ferragem que recebe o macho de outra ferragem. 5. Peça rotativa para contactos elétricos sucessivos na distribuição de corrente a diversos circuitos. 6. Porção de terra, de forma cónica, deixada nos desteros para calcular a quantidade de metros cúbicos de terra retirada. 7. Dispositivo colocado na boca dos bezerros e cavalos, para impedir que mamem. 8. Náut. Cada um dos paus curvos que formam a proa e a popa das canoas. 9. Bras. Bebida preparada com aguardente e mel. 10. pl. pop. pés	2	33	agarrou o cachimbo	agarrou o objeto de fumar
27	cacumbo	kahumbu	kimbundu	<i>adj. unif. e n. m. e f.</i> 1. Forma diversificada de cahumbo. 2. Mutilado de um membro ou de outra parte do corpo.	2	136	experientes cacumbos	mutilados experientes
28	cacungula*	kasumbula	kimbundu	<i>n. m.</i> Jogo infantil que consiste em cada jogador, de súbito, apoderar-se de uma coisa que outrem detém.	2	56	valor associado a cacungula	valor associado ao jogo infantil
29	cacusso	kikusu	kimbundu	<i>n. m.</i> 1. Género de peixe da família dos Pércidas, de que existem duas variedades: uma do rio outra do mar. 2. Perca.	3	171	saudade do cacusso	saudade do peixe dos pércidas
30	cafrique	cafíriqui	kimbundu	<i>n. m.</i> 1. Golpe ginástico ofensivo. 2. O mesmo que <i>capanga</i> .	2	71	recusam cafrique	recusam golfe ginástico ofensivo

31	cafuso	kufunzaka	kimbundu	<i>n. m.</i> Filho de mestiço e preta, ou vice-versa.	1	115 401	condenado cafuso inimigo cafuso	condenado o filho de mestiço e preta inimigo filho de mestiço e preta
32	caimaneros	caimaneros	português em angola	<i>n. m.</i> Denominação da associação dos ex-estudantes angolanos em Cuba.	2	327	chegaram os caimaneros	chegaram os ex-estudantes angolanos em Cuba
33	canjunga*	kunjondoka	kimbundu	<i>adj. unif. e n. m. e f.</i> 1. Que tem falta de um membro. 2. Mutilado. 3. <i>Cahumbo</i> .	2	38	também o canjunga	também o mutilado
34	cahama	cahama	umbundu	<i>n. f.</i> 1. Antílope com a configuração de uma vaca. 2. Vaca do mato.	2	49	este cahama	este antílope semelhante a uma vaca
35	cahombo	cahombo	kimbundu	<i>n. m.</i> 1. variedade de jindungo. 2. Forma abreviada de <i>jindungo-de-cahombo</i> . 3. <i>Bras.</i> Pimenta de cheiro.	2	26	estava no cahombo	estava no piri-piri
36	caiombo*	kuhumbuka	kimbundu	<i>adj. unif. e n. m. e f.</i> 1. Mutilado de um membro ou de outra parte do corpo. 2. O mesmo que cacumbo.	2	57 234	conseguiu caiombo Augusto caiombo	conseguiu o mutilado Augusto mutilado
37	calembela*	calembela	kimbundu	<i>n. f.</i> 1. Antiga dança carnavalesca dos musseques de Luanda aparecida por volta de 1910. 2. O mesmo que <i>calembe</i> .	2	40	experiência da calembela	experiência da dança carnavalesca
38	calongo	calongo	kimbundu	<i>adj. unif. e n. m. e f.</i> Ave de rapina agoirenta.	2	11	estudantes de calongo	estudantes de ave de rapina agoirenta
39	calulense*	kulûla	umbundu	<i>n. m. e f.</i> Indivíduo natural de Calulo, na província do Kwanza-Sul.	2	24	pagamentos calulenses	pagamentos de Calulo
40	calundo	calundo	umbundu	<i>n. m.</i> 1. Cemitério. 2. Ser sepultado.	2	21 22	seguida calondo esse calondo	ser sepultado de seguida esse cemitério
41	caluanda	kaluanda	kimbundu	1. n. m. e f. Indivíduo natural ou habitante da cidade de Luanda. 2. adj. unif. Relativo a essa população.	1	1874	jovens caluandas	jovens naturais de Luanda
42	calundu	kalundu	kimbundu	<i>n. m.</i> Espírito.	1	94 1048 1329	reduzia os calundus interrogar os calundus poder dos calundus	reduzia os espíritos interrogar os espíritos poder dos espíritos
43	calunga-ya-meia*	kalunga	kimbundu	<i>n. m.</i> 1. Mar, oceano, além-túmulo, eterno, Deus. 2. Interj. Que horror! Santo Deus!	1	972	chamam calunga-ya-meia	clamem a Deus dos grandes rios

44	caluyombo*	uyombe	umbundu	<i>n. m.</i> 1. Pequena planta medicinal. 2. <i>v.</i> Recorrer a planta para gerar filho.	2	25	via-se caluyombo	via-se a planta medicinal
45	camacove*	camacove	kimbundu	<i>n. f.</i> Couve pequena, geralmente murcha.	2	49 52	camacoves internacionais promover o camacove	couves internacionais promover a couve
46	camassete*	kamása	kimbundu	<i>adj.</i> 1. Do milho ou a ele relativo. 2. <i>n. Bot.</i> Planta esterculiácea. 3. Milharal.	2	34	chegados do camassete	chegados do milharal
47	cambalanganja*	mbalanganza	kimbundu	<i>n. Zoo.</i> 1. Inseto cujo abdómen termina em forma de tenaz. 2. Tesoura.	2	43	trabalhar cambalanganja	trabalhar com tesoura
48	cambamba*	kámbamba	kimbundu	1. <i>n.</i> Varinha; chibata; vareta. 2. <i>adj.</i> Contínuo; completo. 3. Do princípio ao fim.	2	13	era a sua cambamba	era a sua vareta
49	cambala	kimbala	kimbundu	<i>n. f.</i> 1. Embarcação de inferior dimensão da quimbala. 2. Pequena canoa; piroga.	2	51	margem da cambala	margem da piroga
50	cambinga*	kubonga	kimbundu	<i>n. m. e f.</i> Garoto; petiz; rapazote.	2	129	era cabinga	era um rapazote
51	cambondo	kambóndo	kimbundu	<i>n.</i> 1. Pequeno embondeiro. 2. <i>n.</i> Relativo a adansónia.	2	526	deve ser cambondo	deve ser pequeno imbondeiro
52	cambundo	cambundo	kimbundu	<i>n. m.</i> 1. Grão acinzentado, semelhante, pelo tamanho, ao grão-de-bico, utilizado em colares. 2. O que, no ritual da circuncisão, acompanha os circuncisos à caça.	2	218	vinte cambundos	vinte acompanhantes dos circuncisos
53	cambungo	cambungo	nganguela	<i>n. m.</i> 1. Arbusto de folhas persistentes. 2. Caules lenhosos, geralmente com 1cm de diâmetro, ou pouco mais, e muito ramificado nas extremidades, atingindo 2m de altura total. 3. Seiva lactescente coagulando no ritidoma dos caules ao contacto com o ar.	2	50	dias de cambungo	dias de arbusto
54	cambuta	kubuta	kimbundu	1. <i>adj. unif.</i> Diz-se de pessoa de estatura baixa ou inferior à média. 2. <i>n. m. e f.</i> Pessoa de estatura baixa ou inferior à média. 3. Baixinho.	2	233 227	ajudante cambuta esse cambuta	ajudante de estatura baixa esse baixinho
55	camoço*	kamóso	kimbundu	<i>n.</i> 1. Rapaz. 2. Criado; servente.	2	154 32	caixa camoço considerou o camoço	caixa do rapaz considerou o servente

56	candengue	kandengue	kimbundu	<i>n. m. e f.</i> 1. Criança. 2. O mesmo que <i>cambonga</i> . 3. O mais novo dos irmãos.	1	293 1853	virava os candengues pedrada com outros candengues	virava as crianças pedrada com as outras crianças
57	candando	kandandu	kimbundu	<i>n. m.</i> Abraço; amplexo.	2	27	um candando	um abraço
58	candimba*	kandimba	kimbundu	<i>n. m.</i> 1. Variedade de lebre. 2. Recipiente. 3. Personagem principal de contos populares. 4. <i>v.</i> Agir como personagem principal.	2	62	água na candimba	água no recipiente
59	candingo	kandingo	kimbundu	<i>n. f.</i> Mandioca.	2	164	leite e candingo	leite e mandioca
60	candumba*	kandumba	kimbundu	<i>n. f.</i> Farinha de batata doce.	2	3	compra-se candumba	compra-se farinha de batata doce
61	candumbo	candumbo	kimbundu	<i>n. m.</i> 1. Cobertor ordinário. 2. Manta grosseira. 3. Agasalho reles.	2	38	envolvidos em candumbo	envolvidos em manta
62	canhangulo	kunyanga + ngulo	kimbundu	<i>n. m.</i> 1. Espingarda gentílica de carregar pela boca. 2. <i>Gír.</i> Caneca ou copo grande, cheio de cerveja, contendo a quantidade de uma garrafa grande.	1	897	grosseiros canhangulos	grosseiras espingardas
63	capingalã*	kapingalã	umbundu	<i>n.</i> Substituto; sucessor.	2	153	terminais capingalã	terminais substitutos
64	capoco*	kapoko	umbundu	<i>n. m.</i> 1. Árvore de algodão. 2. O mesmo que <i>utele</i> . 3. Faca pequena. 4. Navalhinha. 5. Canivete.	2	57	melhor que capoco	melhor que faca pequena
65	capuita*	kubuita	kimbundu	<i>n. f.</i> 1. Tambor. 2. Quipuita. 3. Dança antiga carnavalesca de Luanda anterior a 1874.	2	10	a nossa capuita	o nosso tambor
66	caricoco*	karikoko	kimbundu	<i>n.</i> Coquilho.	2	51	continuam caricoco	continuam coquilho
67	cassacambe	kukasa + ukamba	kimbundu	<i>n. m.</i> Grande ave fabulosa, resultante da metamorfoseação humana, figurando nas histórias populares.	1	1232	maquixi, cassacambes	monstros, ave resultante da metamorfose humana
68	cassule	kusuluka	kimbundu	1. <i>adj. unif.</i> O mais novo. 2. <i>n. m. e f.</i> Último filho, nome dado a tal indivíduo. 3. Ultimogénito.	2	117	uma cassule	uma ultimogénita

69	cassungu	okasungu	umbundu	<i>n. m.</i> 1. Missanga miúda. 2. Colar de tais missangas.	2	236	levou o cassungu	levou o colar de missanga
70	catambi*	katambi	umbundu	<i>n. f.</i> Dança criada na década de setenta na comuna da Gangula, município do Sumbe, província do Kwanza-Sul.	2	60	preguiça, catambi	preguiça, dança
71	catembo*	katembo	kimbundu	<i>n.</i> Passarinho dentirrosto de rabo comprido conhecido por “viúva”.	2	89	dia catembo	dia de passarinho
72	catocar*	katoca	cokwe	<i>v.</i> Trabalhar na mina de diamantes em operação desde 1996, localizada no município de Saurimo, província da Lunda-Norte.	2	47	sempre a catocar	sempre a trabalhar em Catoca.
73	catolo*	katurila	umbundu	<i>n. m.</i> 1. Médiun (em relação espírito). 2. Opõe-se a xinguilador, que o é em relação ao vidente. 3. <i>adj.</i> mediúnico.	2	89	final catolo	final mediúnico
74	catumua*	ntúmua	kikongo	<i>n. m.</i> Mensageiro.	2	652	estava o catumua	estava o mensageiro
75	caxexe	kuxexa	kimbundu	1. <i>n. m. e f.</i> Pássaro de coloração azul-celeste. 2. <i>adj.</i> Improvisada.	2	12	defensiva de caxexe	defensiva improvisada
76	caxinde	caxinde	kimbundu	<i>n. m.</i> Gramínea originária da Índia, provavelmente vivaz; folhas do colmo com bainha mais curta que o intervalo do nó, colmos em conjunto com a panícula atingindo 1,5m.	2	52	muito de caxinde	muito de gramínea
77	cazucuta	kuzukuta	kimbundu	<i>n. f.</i> 1. Antiga dança carnavalesca de Luanda, aparecida por volta de 1920. 2. Inflamação da parótida. 3. Papeira. 4. Anarquia; não cumprimento do dever. 5. Desrespeito pelas obrigações do trabalho.	2	483	ensaiam a cazucuta	ensaiam a dança carnavalesca
78	cazuela*	zuela	kimbundu	1. <i>n. m. e f.</i> Falador, verborrágico. 2. <i>v.</i> conversar.	1	1050	sabia que a velha estava de cazuela	sabia que a velha estava disposta a conversar
79	cihongo*	cihongo	cokwe	<i>n. m.</i> Máscara da cultura Cokwe.	2	36 39	uma cihongo entidade, cihongo	uma máscara entidade, máscara

					35 43	vendo cihongo é cihongo	vendo máscara é máscara	
80	chandula*	chandula	português em Angola	<i>n. f.</i> Sandes.	2	61	uma chandula	uma sandes
81	chibinda	chibinda ilunga	cokwe	<i>n. f.</i> Figura ancestral de madeira.	2	238	identificação de chibinda	identificação de figura ancestral
82	chiminha	ximinha	kimbundu	<i>n. f. Deprec.</i> Mulher de vestuário de pano.	2	53	popularizar chiminha	popularizar a mulher de pano
83	chimuco*	tchimuco	umbundu	1. <i>n. m.</i> Ratazana. 2. <i>adj. unif.</i> Relativo à ratazana.	2	207 203 130	chimuco na caprintaria sagrada chimuco próximo chimuco	ratazana na caprintaria sagrada ratazana próximo ratazana
84	chindondo*	otchindundu	umbundu	<i>n. m.</i> Escorpião, lacrau.	2	84	andar de chindondo	andar de escorpião
85	chinduva*	nduva	umbundu	<i>n. m.</i> Pássaro grande.	2	9	comprou chinduva	comprou pássaro grande
86	chinguilamento	kuxingila	kimbundu	<i>n. m.</i> 1. Atuação de um ser espiritual. 2. Incorporação, transe mediúnico.	1	94	calundus, chinguilamentos	espíritos, transe mediúnico
87	chipalanga*	tchipala	umbundu	<i>n. m.</i> Rosto, face.	2	35	superior de chipalanga	superior do rosto
88	chipepe	xipepa	umbundu	<i>adj. unif.</i> Bom. Excelente. Gostoso.	2	44	marca chipepe	marca excelente
89	chipilica*	tchipilika	umbundu	1. <i>v.</i> Insistir; perseverar; teimar. 2. <i>n. m. e f.</i> Insistente, perseverante.	2	101	pensamento chipilica	pensamento insistente
90	chipululu*	oxipululu	umbundu	<i>n. m.</i> 1. Inveja. 2. Ciúme, despeito.	2	208	reparou chipululu	reparou a inveja
91	chiquengue*	tchikenge	umbundu	<i>n. m.</i> Alicate; papagaio.	2	652	comprou chiquengue	comprou alicate
92	cobele*	cobele	português em Angola	<i>n. m. e f.</i> Cobrador de táxi.	2	48	passada de cobele	passada de cobrador de táxi
93	condeco*	condeco	português em Angola	<i>n. m.</i> 1. Pessoa usada para a transmissão de recados. 2. Porta-voz. 3. Mensageiro.	2	20	atendimento do condeco	atendimento do mensageiro
94	coxilar	kukoxila	kimbundu	<i>v. intr.</i> 1. Dormitar, cabecear. 2. Toscanear.	1	1805	e estás coxilando	e estás a adormecer

					1806	como Uandi coxilava	como Uandi adormecia	
95	cuamato	cuamatui	português em angola	1. <i>adj. unif.</i> Que diz respeito ou pertence aos Cuamatos, povo que habita o sul de Angola, no baixo Cunene. <i>n. m. e f.</i> 2. Indivíduo pertencente ao povo dos Cuamatos. <i>n. m.</i> 3. Forma deturpada de cuamatui, embora mais usual. 4. Dialeto falado nessa região pertencente ao grupo étnico dos Ambós.	1	1721	guerreiros cuamatos	guerreiros da região de cuamatos
96	cudila*	kudila	kimbundu	1. <i>v.</i> Chorar. 2. <i>n.</i> Choro, lamentação	2	200 218	a cudila e cudila	a lamentação e chorava
97	cuilo	kukuila	kimbundu	<i>n. m.</i> 1. Contrapeso. 2. Quebrar. 3. Brinde dado ao comprador da natureza do produto adquirido.	2	62	reconhecimento cuilo	reconhecimento de contrapeso
98	culumbo*	kûlu	kimbundu	<i>n. m.</i> 1. Espírito de antigo marinheiro da Serra Leoa. 2. Kulumanu.	2	52	afastar de culumbo	afastar de espírito de marinheiro
99	cumbu	ukumbu	kimbundu	<i>n. m.</i> 1. <i>Gír.</i> Dinheiro. 2. Vaidade.	2	101	reparado o cumbu	reparado o dinheiro
100	cunga	kunga	kimbundu	<i>n. m.</i> Trabalho árduo.	2	136	foi uma cunga	foi um trabalho árduo
101	curihingana*	kuriunga	kimbundu	<i>v.</i> 1. Estar isolado; só. 2. Afastar-se da convivência.	2	20	ficar curinhingana	ficar isolado
102	cutar*	cutar	português em Angola	1. <i>v.</i> Travar. 2. <i>n.</i> Entrave	2	29	provocaram cutadas	provocaram entraves
103	cuteta*	cuteta	português em Angola	<i>n. m.</i> Navio industrial angolano com capacidade para captar 800 toneladas de peixe.	2	54	dois cutetas	dois navios industriais
104	cuvula*	kuvula	kimbundu	<i>v.</i> 1. Abundar; ter muito; haver em excesso, em quantidade. 2. Ser numeroso. 3. <i>n.</i> Abundância. 4. <i>adv.</i> Em grande número.	2	4	cultural cuvula	cultura em abundância

105	damba	ndamba	umbundu	<i>n. m.</i> 1. Depressão de terreno onde geralmente corre um riacho ou existe um pântano. 2. Vale. 3. Córrego.	2	62	grande damba	grande depressão de terreno
106	diala*	diala	kimbundu	1. n. m. Homem; pessoa do género masculino. 2. adj. unif. Relativo a homem ou pessoa do género masculino.	2	44	água e diala	água e homem
107	dicamba-diá-ngalafa*	dicamba-diá-ngalafa	kimbundu	<i>adj.</i> 1. Amigo da garrafa. 2. Beberrão.	1	1590	posto dicamba-diá-ngalafa	posto o beberrão
108	dicanza	kukanzana	kimbundu	1. n. f. Espécie de chocalho de bordão, de cerca de um metro de comprimento. 2. v. Tocar dicanza.	1	1099	raspando dicanzas	raspando chocalhos de bordão
109	diculu	ukûlu	kimbundu	1. n. m. Espírito de ancião; 2. adj. Entidade espiritual que, na vida terrena, atingiu avançada idade.	1	1050	cazuela com alguém diculu	ser logorreico com algum espírito de ancião
110	diculundundu	ukulundundu	kimbundu	<i>n. m.</i> 1. Espírito de ancião de avançada idade. 2. Entidade espiritual que, na vida terrena, ultrapassou o período existencial do diculo.	1	1637	herdara o diculundundu	herdara o espírito de ancião de avançadíssima idade
111	dikota	kota	kimbundu	<i>n. m. e f.</i> 1. Pop. Geralmente entre adolescentes, pessoa mais velha que aquela que profere a designação. 2. Pessoa que encontra na meia idade.	2	3	corpo da dikota	corpo de pessoa de meia idade
112	dipanda*	dipanda	kimbundu	<i>n. f.</i> 1. Independência, <i>ufolo</i> . 2. <i>Wukule</i> em Cokwe. 3. <i>Eyovo</i> em umbundu. 4. <i>Emanguluku</i> em kwanyama. 5. <i>Uwuvoke</i> em nganguela. 6. <i>Kimpwanza</i> em kikongo.	2	67	dia da dipanda	dia da independência
113	diquixi	kuxiba	kimbundu	1. n. m. Monstro de grande cabeça, a qual uma vez despedaçada, se reproduz. 2. v. Sorver; alusão ao devoramento de pessoas e coisas.	1	1246	surgissem as cabeças de diquixi	surgissem as cabeças de monstro
114	dixinde	dixindi	kimbundu	<i>n. m.</i> 1. Abrolho. 2. Planta prostrada zigofilácea. 3. Pl. Maxindes.	2	291	encontrar dixinde	encontrar abrolho

115	dondi*	dondi	kimbundu	<i>n. m.</i> 1. Espécie de banana, geralmente usada para fazer o bananal. 2. Região da província do Kwando Kubango. 3. Banana da madeira.	2	69	apreciação de dondi	apreciação de banana da madeira
116	djavela*	ondjavela	umbundu	<i>n. m.</i> Milho cozido não descascado.	2	53	vendia djavela	vendia milho cozido
117	dongala*	ndungula	umbundu	<i>n. m.</i> 1. Profeta. 2. O antropónimo é atribuído a quem assume cargo de conselheiro.	2	26	último dongala	último profeta
118	dumba	odumba	nganguela	<i>n. m. e f.</i> 1. Leão, leoa. 2. O mesmo que curica.	2	31 52	intenção de dumba uma dumba	intenção de leão uma leoa
119	dumbi	ndumbe	kimbundu	<i>n. m. e f.</i> 1. Médiu. 2. Xinguilador. 3. Mulher de calundu. 4. Mulher de escravidão.	2	187	santo dumbi	santo médiu
120	ecundi	ecundi	nyaneka	<i>n. m.</i> 1. Bebida constituída de quimbombo com mel. 2. Hidromel de quimbombo.	2	59	relação com ecundi	relação com hidromel
121	escanzelar*	escanzelar	português em Angola	<i>v.</i> Emagrecer.	2	16	terem escanzelados	terem emagrecido
122	facar*	faca	português em angola	<i>v. tr.</i> Repelir, revogar, anular.	2	342	facar ele não	não o revogues .
123	funji	funji	kimbundu	1. n. m. massa cozida de farinha denominada fuba, geralmente de milho, massambala, massango, mandioca ou bata doce. 2. impr. Pirão. 3. Bras. Angu.	1	1810	comeste do nosso funji	comeste do nosso pirão
124	ecunha	ecunha	umbundu	<i>n. f.</i> 1. Novilho ou novilha desmamados, ou prestes a sê-lo. 2. Cria de vaca que já deixou de mamar ou está para deixar.	2	49	só ecunha	só novilho
125	ekumbi*	ekumbi	umbundu	<i>n. m.</i> 1. Sol. 2. Dia.	2	132 25	pequenos ekumbi investigação dos ekumbi	pequeno dias investigação dos dias
126	gindungo	ndungu	kimbundu	<i>n. m.</i> 1. Fruto do jindungueiro. 2. Malagueta pequena.	1	228	bago de gindungo	bago de piripiri .

127	gungo	ngûngu	kimbundu	1. <i>n. m.</i> Pássaro de coloração castanha clara pintalgada de branco, tanto dorsal como ventralmente, bico e pés vermelhos. 2. Pessoa de grande relevo social. 3. Aquele que é importante, quer pelo saber, quer pelo cargo, quer pelo dinheiro. 4. Personalidade notável.	1	69	até mesmo um gungo	até mesmo um pássaro de coloração castanha clara pintalgada de branco
128	ginguba	nguba	kimbundu	<i>n. f.</i> 1. Forma errónea de jinguba. 2. Amendoim.	2	26 23	grau de ginguba apelo em ginguba	grau de amendoim apelo em amendoim
129	gombala*	mbála	kimbundu	<i>n. m.</i> 1. Tira de couro para cingir ou atar. 2. Correia.	2	86	próxima gombala	próxima tira de couro
130	gombe	ngombo	nganguela	<i>n. m.</i> 1. Adivinho, necromante. 2. Foragido, desertor, bandido, escravo fugitivo.	2	53	estado de gombe	estado de adivinho
131	hoxa	kukoxa	kimbundu	<i>n. f.</i> 1. Sonolência. 2. Doença de sono. 3. Tripanossomiase.	1	275	doente de hoxa	doente de tripanossomiase
132	huilano	híla	umbundu	<i>adj.</i> Relativo a essa população.	2	29	responsável huilana	responsável proveniente da Huíla
133	humba*	humba	umbundu	<i>n. f.</i> 1. Cesto, quinda. 2. Vasilha feita de coiro de certos animais para condução de líquido. 3. Odre.	2	231	nova humba	novo cesto
134	hungulo	kungunguma	kimbundu	1. <i>n. m.</i> Instrumento musical, formado de uma vara arqueada, retesada por uma corda de arame de cobre, possuindo, quase no extremo do arco, para funcionar como caixa de ressonância, uma aplicação de cabacinha. 2. Urucungu. 3. Bailado executado ao som deste instrumento.	2	196	tocava hungulo	tocava instrumento musical
135	idimakaji	idimakaji	kimbundu	<i>n. m.</i> 1. Trabalhador. 2. Grupo de música folclórica angolana.	2	28 30	educação do idimakaji tem todos idimakaji	educação do trabalhador tem todos os trabalhadores

136	imbamba	kimbamba	kimbundu	<i>n. f.</i> 1. Coisas; móveis; bagagem. 2. Tudo o que pertence a alguém. 3. O mesmo que <i>bicuatas</i> , em linguagem de formação umbundu.	1	1813	tuas imbambas	tudo que te pertence
137	imbondeiro	mbondo	kimbundu	<i>n. m.</i> Forma diversificada de embondeiro, constituindo a grafia mais usual.	1	171 1771	gigantesco imbondeiro imponentes imbondeiros	gigantesco imbondeiro imponentes imbondeiros
138	issukussuku*	issukussuku	kimbundu	<i>n.</i> 1. Chuvisco; chuva miúda. 2. Gotas espaçadas de chuva. 3. Orvalho.	2	9	com issukussuku	com chuvisco
139	jianju*	jihanji	kimbundu	1. <i>n. m.</i> Desejo; apetite; ambições; vontades. 2. <i>adj.</i> Apetitoso.	1	1567 1578	Ximinha Jianju idem	Ximinha apetitosa idem
140	kadima*	kahima	kimbundu	<i>n. m.</i> 1. Macaquinho. 2. Saguim. 3. <i>Adj.</i> Amacacado; simiesco.	2	67	é kadima	é macaquinho
141	kalombo	kalombo	kimbundu	<i>n. f.</i> 1. Espírito feminino que promove a esterilidade. 2. Entidade espiritual da esterilidade.	2	62	em kalombo	em espírito de fertilidade
142	kalule*	kalûle	kimbundu	1. <i>adj.</i> Que tem sabor amargo. 2. <i>n.</i> Peixe do mar de sabor amargo. 3. Cachucho.	2	12	música kalule	música amarga
143	kalumba	kalumba	kimbundu	<i>n. f.</i> 1. <i>Deprec.</i> Rapariga de baixa condição social. 2. Mulherzinha. 3. <i>Restr.</i> Nome que, por desdém dessa inferioridade, as antigas amas impunham a uma filha de escravos ou serviçais.	2	23	kalumba de confiança	rapariga de confiança
144	kalunga	kalunga	kimbundu	<i>n. m.</i> 1. Mar, oceano, além-túmulo, eternidade, Deus. 2. <i>Interj.</i> Que horror! Santo Deus!	2	4 25	teve kalunga melhor é kalunga	teve eternidade melhor é além-túmulo
145	kamaka*	kamaka	kimbundu	<i>n. m. e f.</i> Pessoa problemática.	2	12	for kamaka	for peessoa problemática
146	kamanga	kamanga	termo regional	<i>n. f.</i> 1. Negócio ilícito de diamantes. 2. <i>P. ext.</i> Venda clandestina de pedras preciosas. 3. <i>n. m.</i> diamante.	2	166	era mesmo kamanga	era mesmo diamante
147	kamba	rikamba	kimbundu	<i>n. m.</i> Amigo, camarada, companheiro.	2	14	espaço de kamba	espaço de amigo

						38	suficiente kamba	suficiente companheiro
148	kambala	kimbala	kimbundu	<i>n. m.</i> Pequena canoa.	2	42	momento de kambala	momento de pequena canoa
149	kambadiami	kambadiami	kimbundu	1. n. m. e f. Expressão coloquial que quer dizer “meu amigo”. 2. v. Comportar-se como um amigo.	2	6 7	suportar kambadiami investi kambadiami	suportar o meu amigo investi, meu amigo
150	kambambi	mbambi	kimbundu	<i>n. m.</i> Diminutivo de gazela, cervo.	2	16	comprou kambambi	comprou a gazela
151	kambinga*	kambinga	kimbundu	<i>n.</i> Cabacinha.	2	43	recursos kambinga	recursos de cabacinha
152	kambundo	kambundo	kimbundu	<i>n. m.</i> 1. Grão acinzentado, semelhante, pelo tamanho, ao grão-de-bico, utilizado em colares. 2. O que, no ritual da circuncisão, acompanha os circuncisos à caça.	2	191	sob kambundo	sob acompanhantes dos circuncisos
153	kambuta	kubuta	kimbundu	1. adj. unif. Baixo (ref. a pessoa). 2. n. m. e f. Indivíduo de pouca altura. 3. Nome familiar de um homem ou mulher de com essa estatura.	2	214	Lázaro é kambuta	Lázaro é baixinho
154	kamuanga*	kamuánze	kimbundu	<i>n. Bot.</i> Árvore da família das leguminosa, utilizada para sombra.	2	42	é kamuanga	é árvore leguminosa
155	kandambu*	kandambia	kimbundu	<i>n.</i> 1. Erva miúda. 2. Relva.	2	33	sobretudo kandambu	sobretudo relva
156	kandengue	kandengue	kimbundu	<i>n. m. e f.</i> 1. Criança. 2. O mesmo que <i>cambonga</i> . 3. O mais novo dos irmãos.	2	34	coisas de kandengue	coisas de criança
157	kamdimba*	kamdimba	kimbundu	<i>n. f.</i> Unidade de medida equivalente a meio quilo.	2	17	em kamdimba	em meio quilo
158	kangandjo	kanganji	nganguela	<i>n. m.</i> 1. Homem mascarado ritualmente. 2. O mesmo que <i>muquíxi</i> .	2	437	vimos o kangandjo	vimos o homem mascarado
159	kanhamei*	kanhamu	kimbundu	<i>n.</i> Ressentimento.	2	54	inevitáveis para a kanhamei	inevitáveis para o ressentimento
160	kanhangulo	kunyanga (matar) + ngulo (porco)	kimbundu	<i>n. m.</i> 1. Espingarda gentílica de carregar pela boca. 2. Gir. Caneca ou copo grande.	2	35	peças de kanhangulo	peças de espingarda

161	kanhinguquine*	kanhenge	kimbundu	1. Delgado; fino; esguio; de pouca grossura. 2. Delicado.	2	58	forma kanhinguquine	forma delgada
162	kanjongo*	kanjungu	kimbundu	<i>n. m.</i> 1. Diminutivo de <i>njungu</i> . 2. Lourinho.	2	62	era kanjongo	era lourinho
163	kapoco*	kaphoko	kimbundu	<i>n. m.</i> Faca pequena.	2	161	o kapoco	o faca pequena
164	kapuka*	kapuka	kimbundu	<i>n. f.</i> Bebida feita de cana de açúcar.	2	55 10	bases de kapuka antiga kapuka	bases de bebida de cana antiga bebida de cana
165	kasola*	kasola	kimbundu	<i>n. f.</i> Oportunidade.	2	98	minha kasola	minha oportunidade
166	kassinda*	kasinda	umbundu	<i>v.</i> 1. Impelir. 2. Empurrar.	2	12	não kassinda	não empurres
167	kassoma*	kasoma	umbundu	<i>n. m.</i> 1. Substituto. 2. Orador sagrado.	2	30 77	destacar o kassoma foi kassoma	destacar o sagrado foi substituto
168	kassonde*	kusondoloka	kimbundu	<i>n. m. e f.</i> Formigão avermelhado, de mordedura dolorosa.	2	24 27	provocou kassonde qualquer kassonde	provocou o formigão qualquer formigão
169	katale*	kata	umbundu	<i>n. m.</i> Espécie de alicate.	2	54	antigo katale	antigo alicate
170	katunga	katunga	kimbundu	<i>n. m.</i> 1. Bicho dos pés. 2. Pulguedo.	1	707	katunga e fortificados lugares	pulguedo e fortificados lugares
171	katungui*	katungu	kimbundu	<i>n.</i> Exceção; exclusão.	2	26	reunião e katungui	reunião e exclusão
172	kayaya*	kaiáia	kimbundu	<i>n.</i> Peixe miudinho do rio Kwanza.	2	41	demais kayaya	demais peixes miudinhos
173	kenzo*	kienzu	kimbundu	<i>n.</i> 1. Guarda de passarinho. 2. Ninho. 3. Abrigo; refúgio.	2	41	um kenzo	um ninho
174	kiala*	kiala	kimbundu	<i>n. f.</i> 1. Unha. 2. Cada uma das unhas das aves de rapina ou das feras carniceiras. 3. Garra. 4. Extremidade de algumas ferramentas e utensílio.	2	20	geral a kiala	geral a unha
175	kiambi*	kiámbi	kimbundu	<i>n. m.</i> 1. órgão glandular do lado esquerdo. 2. Baço. 3. O lado exterior esquerdo do abdómen.	2	36	entregue kiambi	entregue o baço

176	kiamuxinda*	muxinda	kimbundu	<i>n. m.</i> 1. Traço, linha (vertical ou longitudinal). 2. Via láctea.	2	439	chamam kiamuxinda	chamam via láctea
177	kianda	kuanda	kimbundu	1. n. m. Entidade sobrenatural das águas. Sereia. 2. P. ext. Criatura fisicamente anormal. 3. Bras. Iemanjá.	2	257	efeito de kianda	efeito de sereia
178	kibabo	kibabo	kimbundu	<i>n. m.</i> 1. Tabefe. 2. Bofetada fraca.	2	10	o kibabo	a bofetada
179	kicangue	kukanguka	kimbundu	<i>n. m.</i> 1. Manchas brancas que aparecem na pele, especialmente no rosto e pescoço, ocasionadas pela puberdade, gravidez ou mau funcionamento do fígado. 2. Pano.	2	125	levou o kicangue	levou o pano
180	kilape*	kilapi	kimbundu	<i>n. m.</i> Crédito, dívida.	2	43	maka e outros kilapes	problema e outras dívidas
181	kilombelombe	kulomba	kimbundu	<i>n. m.</i> 1. Corvo. 2. Deprec. Agente de casa funerária. 3. Editora angolana.	2	42	difícil de kilombelombe	difícil de corvo
182	kilombo	xilombo	kimbundu	<i>n. m.</i> Sanzala de trabalhadores agrícola, junto à área da lavoura; acampamento.	2	19 64	aumento do kilombo comercializam em kilombo	aumento do acampamento comercializam na sanzala
183	kima*	kima	kimbundu	<i>n. m.</i> Começo.	2	24	última kima	último começo
184	kimbo	kimbo	umbundu	<i>n. m.</i> 1. Povoado. 2. Aldeamento. 3. Sanzala. 4. Bairro indígena. 5. Casa isolada, ou bloco de casas constituindo um só lar.	2	29 60	diagnóstico de kimbo apresenta o kimbo	diagnóstico do povoado apresenta a aldeia
185	kinda	kuinda	kimbundu	<i>n. f.</i> 1. Espécie de cesta. 2. Cesta regional angolana.	2	412	pegou a kinda	pegou a cesta
186	kindala*	kindala	kimbundu	<i>adv.</i> 1. Já. 2. Sem demora. 3. Neste instante. 4. Imediatamente. 5. adj. Célere; imediato.	2	37	para kindala	para este instante
187	kinguila*	kingila	kimbundu	1. n. m. e f. Pessoa que cambiam moeda nas ruas. 2. Cambista. 3.v. Esperar.	2	31	através da kinguila	através do cambista
188	kingungo	kungonga	kimbundu	<i>n. m. e f.</i> Nome dado a gémeo.	2	39	jornalista kingungo	jornalista gémeo

189	kissangua	kisangua	umbundu	<i>n. f.</i> Cerveja de milho.	1	1810	bebeste da nossa kissangua	bebeste da nossa cerveja de milho
190	kissanguela	kusangela	kimbundu	<i>n. f.</i> Sociedade, parceria, cotização, contribuição, coleta.	2	24	reconhecerem kissanguela	reconhecerem a sociedade
191	kissanji	kisanji	kimbundu	<i>n. m.</i> Instrumento músico de hastes metálicas.	2	28	clube de kissanji	clube de instrumento musical
192	kissueias*	kissueias	kimbundu	<i>n. m.</i> Inimigo, bandido, vampiro.	2	68	eram kissueias	eram inimigos
193	kitembo	kitembo	kimbundu	<i>n. m.</i> Divindade.	2	22	reconhece kitembo	reconhece a divindade
194	kitoco	kutokoka	kimbundu	<i>adj. unif.</i> 1. Esbelto, elegante, bonito. 2. <i>Bras.</i> Xiruba.	2	218	Jurema kitoco	Jurema elegante
195	kitongo	kutunga	kimbundu	<i>n. m.</i> 1. Casebre. 2. Habitação arruinada. 3. Pousada tosca em lavra.	2	60	desconcentração de kitongo	desconcentração de casebre
196	kitumba*	kitumba	umbundu	<i>adj. unif.</i> Relativo ao feitiço.	2	34	final de kitumba	final de feitiçaria
197	kizembua	kuzembeza	kimbundu	<i>n. m.</i> 1. Alpendre mortuário. 2. Recinto com toldo, para exposição ritual de cadáveres. 3. <i>Fig.</i> Óbito.	2	142	retirar o kizembua	retirar o alpendre mortuário
198	kizomba	kuzomba	kimbundu	<i>n. f.</i> 1. Batuque. 2. Bailarico. 3. Diversão dançante. 4. Folgado.	2	173	basculante de kizomba	basculante de bailarico
199	kota	kota	kimbundu	<i>n. m. e f.</i> 1. <i>Pop.</i> Geralmente entre adolescentes, pessoa mais velha que aquela que profere a designação. 2. Pessoa que se encontra na meia idade. 3. Adulto. 4. <i>adj. unif.</i> Que se encontra na meia idade.	2	8 26	anúncio de kota personalidades kotas	anúncio de adulto personalidades adultas
200	kotukua	kotukua	kikongo	<i>v.</i> Acordar.	2	541	já kotukua	já a acordar
201	kuabar*	kuâbu	kimbundu	<i>interj.</i> Não mais! Basta! 2. <i>adj.</i> Suficiente. 3. <i>loc. adj.</i> Quanto basta; tanto; quanto possível.	2	23	aconselhou kuaba	aconselhou o suficiente
202	kuandunda*	kuandunda	kimbundu	<i>v.</i> 1. Ruminar; remoer. 2. tornar vivo.	2	31	garantir para kuandunda	garantir para tornar vivo

203	kuatar	kukuata	kimbundu	<i>v.tr.</i> Agarrar, segurar.	1	131	kuata os pretos	agarra os pretos
204	kuburisékima*	kubunjikila	kimbundu	1. v. Dobrar; ajustar por outrem. 2. Ajuste.	2	15	precedente kuburisékima	precedente ajuste
205	kundi*	kundi	kimbundu	<i>n. m.</i> 1. Feijão frade; verde. 2. Abreviação de rikunde	2	16	chegou kundi	chegou feijão frade
206	kunene-bu*	kunene-bu	kikongo	<i>Exp.</i> Não defeques onde comes.	1	168	bu kunene-bu	não cagues onde comes
207	kupapata*	kupapata	umbundu	<i>n. f.</i> 1. Motociclo de três rodas usada para o transporte pago de passageiros ou mercadorias. 2. Mototáxi. 3. n. m. e f. Pessoa que conduz uma motocicleta, assegurando um serviço pago de transporte tanto de passageiros como de mercadorias. 4. Mototaxista.	2	4	passam a kupapata	passam o motociclo de três rodas
208	kuribeka*	kuribeka	kimbundu	1. v. Vir sem ser chamado. 2. Apresentar-se; oferecer-se; entregar-se; exibir-se.	1	196	pedreiros-livres, kuribekas	pedreiros-livres, peessoas não convidadas
209	kususa*	kususa	kimbundu	1. v. intr. Urinar; mijar. 2. n. Urina, mijo.	2	31	que kususa	que urina
210	kusseletala*	kuséla	kimbundu	<i>v.</i> 1. Recalcar; repisar. 2. Enchamelar. 3. Fazer-se lavar. 4. Purificar-se.	2	45	e kusseletala	e recalca
211	kussuku*	kusukuisa	kimbundu	<i>v.</i> 1. Mandar lavar. 2. Passar pela água. 3. Lavar. 4. Abluir.	2	9	oposição desses kussuku	oposição dessas passagens pela água
212	kuxixima*	kuxixima	kimbundu	1. v. intr. Arder. 2. n. f. Ardência, desventura, infelicidade.	2	29	está em kuxixima	está em desventura
213	kwacha	kwaxa	kimbundu	<i>n. m. e f.</i> Militante ou simpatizante do movimento guerrilheiro União Nacional para a Independência Total de Angola.	2	64	foram kwachas	foram militantes da UNITA
214	kwazambi*	nzâmbi	kimbundu	<i>n.</i> 1. Deus. 2. Santo; divindade de qualquer religião. 3. Cada um dos membros da trindade cristã. 4. O Ser Supremo; O Todo Poderoso.	2	69	apreciação em kwazambi	apreciação no Ser Supremo

215	kwanzar*	kwanza	umbundu	v. Usar a unidade monetária Kwanza.	3	253	sempre a kwanzar	sempre a usar a moeda kwanza
216	lépi*	elépi	kimbundu	n. f. Coreografia.	2	18 19	correspondem a lépi assegurar a lépi	correspondem à coreografia assegurar a coreografia
217	liambar	oliamba	umbundu	n. f. 1. Fumar uma espécie de cânhamo. 2. Diamba	2	26	levar liamba	levar diamba
218	libata	libata	umbundu	n. f. Aglomerado de habitações envolvidas por uma sebe narcótica e pertencentes a um mesmo senhor, nas quais vive ele, as suas mulheres e demais familiares.	1	736	ele e a sua libata	ele e o seu aglomerado de habitações
219	livongue*	livonge	umbundu	n. m. 1. Alcinha de sobas. 2. Que surge depois de muitos nados mortos.	2	17 122	agrupamento do livongue empregada livongue	agrupamento de sobas empregada que surge depois de nados mortos
220	lobitanga*	lobitanga	umbundu	n. m. e f. Indivíduo natural do Lobito.	2	81 82	primeira cria lobitanga queixaram essa lobitanga	primeira cria natural do Lobito queixaram essa natural do Lobito
221	luandar*	luanda	kimbundu	v. Falar à maneira de Luanda.	2	140	está a luandar	está a falar como um luandense
222	lubongo*	lúbongo	kimbundu	n. m. 1. Credencial. 2. Licença.	3	234	recebeu o lubongo	recebeu a credencial
223	lucamba*	lukamba	umbundu	n. m. 1. Amigável. 2. Destemido.	2	76	melhor é lucamba	melhor é destemido
224	lueno	valuena	cokwe	1. n. m. Indivíduo natural de uma área na província do Moxico, a leste de Angola. Dialeto falado nessa região. 2. pl. população dessa área pertencente ao grupo étnico dos nganguelas. 3. adj. unif. Relativo a essa população. 4. pl. híbr. Valuenas.	1	972	chamam os luenos	chamam os naturais de Luena
225	lukombo*	lukombo	kimbundu	n. m. 1. Agência. 2. v. Compra de negócio por atacado.	2	89	estava no lukombo	estava na agência
226	lunguela	lunguila	kikongo	n. f. Vinho de cana de açúcar.	2	67	era muita lunguela	era muito vinho de cana de açúcar

227	mabangas	dibânga	kimbundu	<i>n. f.</i> Marisco assado na própria casca ou ensopado com quiabo, abóbora, óleo de palma e temperos, servido com funje de milho.	2	28	era de mabanga	era de marisco
228	maboko*	maboko	kikongo	<i>n. m.</i> Palmas, aplausos.	2	45	viram maboko	viram os aplausos
229	mabuba	mabuba	kimbundu	<i>n. f.</i> Catarata, queda-d'água.	2	87	beleza da mabuba	beleza da catarata
230	mabululu	dibululu	kimbundu	1. n. m. Confins, terras interiores, matos. 2. adj. unif. Relativo a uma terra distante.	2	83	dificuldades a mabululu	dificuldades às terras interiores
231	macala	macala	português em angola	<i>n. f.</i> 1. Mulher de cor negra. 2. Carvão	1	106	preto macala	preto carvão
232	macanha	kukenya	kimbundu	<i>n. f.</i> 1. Tabaco, rapé. 2. Peixe marinho de tamanho médio, revestido de escamas finas. 3. O mesmo que peixe-macanha.	2	22 25	chegou macanha contas de macanha	chegou tabaco contas de tabaco
233	mucanda	mucanda	kimbundu	<i>n. f.</i> 1. Montanha; monte. 2. Carta com valor de dinheiro, utilizada geralmente pelos comerciantes. 3. Carta, bilhete, requerimento. 4. Livro, jornal, mapa. 5. Qualquer escrito (manuscrito ou impresso).	2	97	quintal na mucanda	quintal no montanha
234	muceque*	museke	kimbundu	<i>n. m.</i> 1. Terreno arenoso, mas agricultável, situado fora da orla marítima, em planície de altitude. 2. P. ext. Quinta, xitaca. 3. Pl. fig. Aldeolas; confins; berças. 4. Bras. Chácara.	2	39 41	três muceques metidas de muceque	três terrenos arenosos metidas de terreno arenoso
235	machila	maxila*	kimbundu	<i>n. f.</i> 1. Palanquim. 2. Cadeira com tampo e cortinas, a qual suspensa de um bordão de bambu, era transportada aos ombros de dois homens.	1	167 239 244	subido em sua machila carregadores de machila carregadores de trabalhadores machila	subindo na cadeira régia carregadores da cadeira régia carregadores de trabalhadores da cadeira régia .
236	macololo*	makolôlo	kimbundu	<i>n. m.</i> 1. Escória, ralé. 2. Os resíduos do fundo da panela.	1	707 709	território dos macololos países dos macololos	território de escória países de escória

237	macongo	dikongo	kimbundu	1. <i>n. m.</i> Crédito, dívida. 2. <i>adj. unif.</i> Que diz respeito à dívida.	2	41 76	entregues a macongo acabou o macongo	entregues a crédito acabou a dívida
238	maculo	kukulula	kimbundu	1. <i>n. m.</i> Afeção do ânus, motivada pela fixação local de pequeníssimos parasitas. Proctite. 2. <i>Bras.</i> Corrupção, mal-de-bicho.	1	1567	do maculo	de afeção de ânus
239	macunge*	makunji	kimbundu	<i>n. m. pl.</i> 1. Conjunto de focas. 2. Pancadas dadas com chicote de cavalo-marinho. 3. Chicotada.	2	139	era no mucunge	era no conjunto de focas
240	madezo	madezo	kikongo	<i>n. m.</i> Mistura de kizaka com feijão.	2	258	comem madezo	comem kizaka com feijão
241	maka	maka	kimbundu	1. <i>n. f.</i> Conversa; assunto; novidade; discórdia; litígio; conflito; algazarra. 2. <i>Bras.</i> Incolumanca. 3. <i>adj.</i> Problemático.	1	125 565 1267 1590	ouvirem as makas arranjar uma maka reconstruir alguma passada maka centro das makas	ouvirem os problemas arranjar um problema reconstruir um litígio passado pessoa problemática
242	makongo	dikongo	kimbundu	<i>n. m.</i> Crédito, dívida.	2	38 54	aquele e este makongo dois makongos	aquele e este crédito duas dívidas
243	malamba	dilamba	kimbundu	<i>n. m.</i> 1. Amargura. 2. Tormento, penar. 3. <i>Pl.</i> Apertos do coração.	2	49	zona de malamba	zona de amargura
244	malanjinho	malanji	kimbundu	<i>n. m.</i> Indivíduo natural da província de Malanje, localizada a nordeste de Angola.	2	135	bairro de malanjinho	bairro de naturais de Malanje
245	malembe*	malembe	kimbundu	1. <i>n. m.</i> Calmaria. 2. <i>Interj.</i> Cuidado! Tem calma!	2	40 52	através de malembe estão de malembe	através de calmaria estão de calmaria
246	malembe- malembe*	malembe	kimbundu	<i>n. m.</i> 1. Devagar se vai ao longe. 2. Pouco a pouco se vai ao longe.	2	47	bola malembe-malembe	bola vagarosa
247	malembelember	malembe	kimbundu	1. <i>v.</i> Ir de vagar. 2. <i>adv.</i> Lentamente.	1	966	corre muito tempo malembelember	corre muito de vagar
248	malundo	malunda	kikongo	<i>n. f.</i> Colar ou pulseira feitos com o enfiamento de caroços do fruto do mubafo, depois de polidos e caracterizados com desenhos.	2	100	empreitada de malundo	empreitada de colar
249	malungo	dilunga	kimbundu	<i>n. f.</i> 1. Espírito da simpatia. 2. Entidade espiritual que, por afeição, propicia amparo. 3. <i>n. f.</i> Vinho de palmeira-leque.	2	77 158	inaugurada malungo expressão de malungo	inaugurado vinho de palmeira expressão de espírito de simpatia

250	mambu	mámbu	kimbundu	<i>n. m.</i> 1. Doutrina, preceito, regra, praxa. 2. Orações, rezas, hinos sacros.	2	57	mesmo mambu	mesma regra
251	mamiondona	kukondona	kimbundu	<i>n. f.</i> Espíritos tutelares.	1	775	tenho mamiondona	tenho espíritos tutelares
252	mamuíla	hila	nyaneka	1. <i>n. m. e f.</i> Indivíduo natural da Huíla, a sudoeste de Angola, pertencente ao grupo Nyaneka 2. <i>n. m.</i> Dialeto falado nessa região, pertencente ao grupo Nyaneka.	1	1895	desdém de uma velha mamuíla	desdém de uma velha pertencente ao grupo Nyaneka
253	mangonheiro*	mangonha	kimbundu	<i>n. f.</i> 1. Preguiça. 2. Esperteza. 3. Patifaria. 4. <i>adj.</i> Preguiçoso.	2	57	anos mangonheiros	anos preguiçosos
254	maquixi*	kuxiba	kimbundu	<i>n. m.</i> Monstro de grande cabeça, a qual, uma vez decepada, se reproduz limitadamente, segundo uns; ou com muitas cabeças simultaneamente, em número variável, segundo outros.	1	1232	ocorrência de maquixi	ocorrência de monstro
255	maracachão	mbalakaxongo	kimbundu	<i>n. m.</i> Pássaro canoro, de plumagem cinzenta escura, no dorso, mas salpicada de branco, e cinzenta clara no peito e ventre, com bico e pés avermelhados.	1	69	pírulas, maracachões	pírulas, pássaros canoros
256	massambalar	masa-ma- mbâla	kimbundu	<i>v.</i> 1. Comer erva não rizomatosa, persistente. 2. Comer ou confeccionar sorgo, milho moído. 3. <i>n. f.</i> Milho moído.	2	10	volta de massambala	volta de milho moído
257	massemba	disemba (de kusemba)	kimbundu	<i>n. f.</i> 1. Bailado angolano caracterizado por semba ou umbigadas. 2. Rebita. 3. <i>v.</i> Requebrar-se.	2	42	questões de massemba	questões de rebita
258	massembar	disemba, kusemba	kimbundu	<i>v. intr.</i> Dançar a massemba; requebrar-se.	1	1099	vinham massembando	vinham a dançar massemba
259	massumbo*	músumbo	kimbundu	<i>adj.</i> Suscetível de compra.	2	298	era de massembo	era suscetível de compra
260	massoxi*	masôxi	kimbundu	<i>n. m.</i> 1. Jorramento. 2. O que verte em abundância.	2	43 46	ocorre para massoxi mais por massoxi	ocorre para jorramento mais por abundância

261	massumba*	músumba	kimbundu	<i>n. f.</i> 1. Bot. Árvore da família das leguminosas, de madeira muito resiliente. 2. Planta têxtil e ornamental.	2	238	própria de massumba	própria de árvore leguminosa
262	matabicho	matabicho	português em angola	1. <i>n. m.</i> Pequeno-almoço. Primeira refeição da manhã. 2. <i>P. ext.</i> Gratificação. Gorjeta. 3. <i>Deprec.</i> Remuneração ridícula. Insignificante retribuição de favor. 4. <i>Bras.</i> Jabaculê, xixica. (Ref. A dinheiro).	1	1834	revirou a matabicho	revirou o pequeno-almoço
263	mateta*	matete (de kutela)	kimbundu	<i>n. m.</i> 1. Papas. 2. Massa de farinha cozida, mas inconsistente. 3. <i>Fig.</i> Indivíduo sem ação, preguiçoso.	2	37 45	centro de mateta vêm mateta	centro de papa vêm papas
264	mateba	mateba	kimbundu	<i>n. f.</i> Fibra da madeira.	2	31 238	envolveu mateba identificação de mateba	envolveu fibra da madeira identificação de fibra da madeira
265	matu*	matu (de matumbu)	kimbundu	<i>n. m.</i> 1. Sertão. 2. Região afastada da costa.	2	27 37 53	antigo matu avançou no matu chamada para o matu	antigo sertão avançou no sertão chamada para o sertão
266	matuba	kutubula	kimbundu	<i>n. f.</i> 1. Testículo. 2. Cada uma das glândulas contidas no escroto.	1	303 557	tinha as matubas tem as matubas	tinha os testículos tem os testículos
267	matubacana*	dituba (resultado de kutubula)	kimbundu	<i>n. f.</i> 1. Testículo. 2. Cada uma das glândulas contidas no escroto.	2	20	tinha matubacana	tinha testículo
268	matuti*	ditute	kimbundu	<i>n. m.</i> Pernas engrossadas pela elefantíase.	2	264	tem matute	tem elefantíase
269	mavungo*	mavunzu	kimbundu	<i>n. m.</i> 1. Turvação. 2. Lia de líquido. 3. Pouca clareza.	2	8 11 76	foi de mavungo diante de mavungo pessoas mavungo	foi turva diante de pouca clareza pessoas com pouca clareza
270	maximbombo	maximbombo (de kuximba)	kimbundu	<i>n. m.</i> 1. Ônibus, autocarro. 2. <i>irón.</i> Pessoa bastante nutrida.	2	35	ainda maximbombos	ainda autocarros
271	mayala	mayala	kikongo	<i>n. m. pl.</i> Homens.	2	3 5	música mayala havendo mayala	música de homens havendo homens

272	mayuyar*	mayuyar	português em Angola	<i>v.</i> Falsificar, alterar para enganar.	2	43	está a mayuyar	está a falsificar
273	mazembe	mazombo (de muzombo)	kimbundu	<i>adj.</i> 1. Acanhado, tímido. 2. Sem espírito de desenvolvimento. 3. <i>n. m.</i> O que é envergonhado, pouco expedito.	2	26 39 48	conselho mazembe em evitar mazembe começou mazembe	conselho tímido em evitar vergonha começou a timidez
274	mbambi*	mbambi	umbundu	<i>n. m.</i> Cabra do mato.	2	62 63	milhões de mbambi dissipassem mbambi	milhões de cabra do mato dissipassem cabra do mato
275	mbandua	ombandua	umbundu	<i>n. f.</i> 1. Instrumento de adivinhação material. 2. Peça de madeira com um sulco longitudinal ao centro onde, por deslizamento de um pauzinho comprido, se formam perguntas divinatórias.	2	34 43 45 73	uma mbandua acusação mbandua organizar mbandua casos de mbandua	um instrumento de adivinhação acusação de adivinhação organizar pedaço de madeira casos de adivinhação
276	mbanza	mbánza (resultado de kubanza)	kimbundu	<i>n. m.</i> 1. Espécie de viola. 2. Instrumento musical com três cordas que se toca com palheta de pau. 3. População residencial do Soba. 4. Moradia do Soba. 5. Palácio régio gentílico. 6. Embala, pela. 7. <i>n. f.</i> Casca de bordão, sua lasca. 8. <i>P. ext.</i> Lasca de ramo de palmeira, coqueiro, matebeira. 9. O mesmo que catandu entre os povos de língua kimbundu.	2	44	apuramento mbanza	apuramento de viola
277	mbeji*	mbeji	kimbundu	<i>n. m.</i> 1. Satélite da terra. 2. Lua. 3. Espaço durante o qual a lua faz a sua revolução em torno da terra. 4. Tempo compreendido entre os dois novilúnios.	2	149 187	maior mbeji favorável mbeji	maior satélite da terra favorável lua
278	mbemba	mbêmba	kimbundu	<i>n. f.</i> Ave de rapina semelhante a águia, mas de plumagem branca e preta, açor.	2	21 26	equipa de mbemba em mbemba	equipa de ave de rapina em ave de rapina
279	mbiji*	mbiji	kimbundu	<i>n. m.</i> Peixe.	1	1161	mbiji ya ukange	peixe frito
280	mbimbi*	mbimbi	kimbundu	<i>n. m.</i> 1. Timbre metal de voz. 2. Fala. 3. Modo de se exprimir. 4. Linguagem.	2	51	tem mbimbi	tem modo de se exprimir

281	mboba*	mboba	kimbundu	<i>adv.</i> 1. Aqui. 2. Neste lugar. 3. Cá.	2	40	joelho a mboba	joelho a aqui
282	mbombo	kubomba	kimbundu	<i>n. m. e f.</i> 1. Segundo filho imediato aos gémeos. 2. Nome dado a esse indivíduo.	2	65	a mbombo	a segunda filha imediata aos gémeos
283	mbunda	mbunda	kimbundu	1. <i>n. f.</i> Nádegas, rabo, traseiro (<i>Pop.</i>) 2. <i>n. m. e f.</i> Pessoa reles, ordinária, pessoa insignificante, sem importância. 3. <i>adj. m. e f.</i> Que pertence ou diz respeito aos Bundas, tribo banta de Angola. 4. <i>n. m. e f.</i> Indivíduo que pertence aos Bundas, uma das tribos bantas de Angola. 5. <i>n. m.</i> Bundo.	1	56	nga mbunda	senhora pertencente à tribo dos Bundos
284	miondona	kukondona	kimbundu	<i>n. f.</i> Espíritos tutelares.	1	94 1634	espíritos, as miondonas conversando compridamente de miondonas	espíritos, os espíritos tutelares conversando compridamente de espíritos tutelares
285	missongo	misóngo	kimbundu	<i>n. m.</i> 1. Raios, frechas, espeto. 2. Parte de um exército armado de lanças. 3. As pontas aguçadas. 4. A pontas aguçadas de que se reveste o cacho de dendém.	2	55	em missongo	em raios
286	molumba	kulumbula	kimbundu	1. <i>n. f.</i> Giba. Corcunda. Marreca. 2. <i>v.t.</i> Dar forma de mulumba a. 3. <i>v. intr.</i> Criar giba, corcovar-se.	1	319	parecia de corcunda molumba	parecia de corcunda acentuada
287	monangamba	monangamba	kimbundu	<i>n. m. e f.</i> 1. Carregador, moço de fretes. 2. <i>Deprec.</i> Servente. Homem ordinário ou sem modo de vida.	1	132	houve monangamba	houve servente
288	môngua*	múngua	kimbundu	<i>n. m.</i> Sal.	2	10 48	mistura môngua muita môngua	mistura sal muito sal
289	mopessela	pesela	umbundu	<i>v.</i> 1. Deitar algo. 2. Livrar-se de alguma coisa.	2	98	mopessela , mano.	livra-te disso , mano.
290	mpemba	kubembula	kimbundu	<i>n. m.</i> 1. Calcário amargo, caulim. 2. Espécie de gesso. 3. Variedade de caracterizador ritual. 4. O mesmo que <i>xiquela</i> , em língua umbundu.	2	9 10 19	história de mpemba seu mpemba falta mpemba	história de calcário seu calcário falta calcário

291	muata	muata	cokwe	<i>n. m.</i> 1. Régulo. 2. O mesmo que Soba entre os povos de língua Kimbundu.	2	14	anunciou a muata	anunciou ao régulo
292	mucanda	mucanda	kikongo e cokwe	<i>n. f.</i> 1. Recinto da circuncisão. 2. O mesmo que lônua, entre os povos de língua umbundu.	2	83 97	colchões e mucanda precisa de mucanda	colchões e cinto de circuncisão precisa de cinto de circuncisão
293	mucano*	mukano	kimbundu	<i>n. m.</i> Litígio.	2	73 291	lugares a mucano não a mucano	lugares a litígio não a litígio
294	muene*	muene	nganguela e kimbundu	<i>n. m.</i> Autoridade, rei, proprietário.	1	1044	combate contra o muene puto	combate contra a autoridade portuguesa
295	mueneputo*	mueneputu	kimbundu	<i>n. m.</i> Autoridade portuguesa (qualquer que seja a sua hierarquia): governador, juiz, militar, chefe de posto, etc.	2	18	pediam a mueneputo	pediam à autoridade portuguesa
296	muesseque*	museke	kimbundu	<i>n. m.</i> 1. Terreno arenoso, mas agricultável, situado fora da orla marítima, em planície de altitude. 2. <i>P. ext.</i> Quinta, xitaca. 3. <i>Pl. fig.</i> Aldeolas; confins; berças. 4. <i>Bras.</i> Chácara.	2	59	formação de muesseque	formação de terreno arenoso
297	mueze-zambi-ya-mema	mueze-zambi-ya-mema	kimbundu	<i>n. m.</i> 1. Báculo. 2. Autoridade.	1	972	bundos mueze-zambi-ya-mema	báculo dos bundos
298	mufilo*	mufito	kwanyama	<i>n. m.</i> Floresta dos areais.	1	1719	nome de mufilo	nome de floresta dos areais
299	muhatu*	muhâtu	kimbundu	1. <i>n. f.</i> Mulher; fêmea; filha. 2. <i>v.t.</i> Chefiar, comandar, dirigir.	1	107 1191	nga muhatu idem	senhora casada à moda tradicional idem
300	mujimbu	mujimbu	kimbundu	<i>n. m.</i> 1. Rumor; murmuro. 2. Especulação de um assunto.	1	647	eco de mujimbus	eco de especulação de um assunto
301	mukenge*	múkenge	kimbundu	<i>n. m.</i> 1. Raposa. 2. Gato bravo.	2	265	viu mukenge	viu raposa

302	mukije	mukixi	kikongo	<i>n. m.</i> 1. Forma errônea de muquíxi. 2. Máscara de ritual. 3. <i>P. ext.</i> Homem mascarado com semelhante velatura. 4. Dançarino de determinadas celebrações rituais. 5. Droga medicinal ou nociva. 6. Sortilégio para proteger contra o mal, ou inversamente, para ocasionar um dano. 7. Bruxedo. 8. Objeto dotado de poder sobrenatural.	3	344	coluna de mukije	coluna de máscara de ritual
303	mukonda*	mukonda	kimbundu	1. <i>pron. Interrog.</i> Porquê?; falso; causa ou motivo. 2. <i>n.</i> Motivo; razão; causa.	2	7	apontando muconda climática	apontando a causa climática
304	mukongo*	mukóngo	kimbundu	<i>n. m.</i> Caçador.	2	982	era mukongo	era caçador
305	mukwambi	mukwambi	umbundu	<i>n. m. e f.</i> 1. Que só fica em casa; pessoa caseira. 2. Pessoa parada.	2	75	grande mukwambi	muito parado
306	mulemba	mulemba	kimbundu	<i>n. f.</i> árvore de seiva leitosa, atingindo 25m, copa volumosa, hemisférica, muito ramificada.	2	25	impróprios para mulemba	impróprios para árvore de seiva leitosa
307	mumbanda	mumbanda (de mulumba)	kimbundu	<i>adj.</i> Giboso, corcovado, que tem molumba.	2	24	aceita mumbanda	aceita o que tem mulumba
308	mundele	mu (abreviação de múkua) + ndele (alma)	kimbundu	<i>n. m. e f.</i> 1. Indivíduo de raça branca. 2. <i>P.ext.</i> Que se traja à europeia. 3. <i>Pl. híbr.</i> Mindeles.	2	32	filho de mundele	filho de branco
309	mundombe	ndombe	umbundu	<i>n. f.</i> 1. Título feminino de pedras. 2. Bagre.	2	18 43	imprensa mundombe dar todos mundombes	imprensa pedregosa dar todos bagres
310	muntungila	mutúngila	kikongo	<i>n. m.</i> 1. Esqueleto de casa. 2. Armação exterior de paus constituindo a ossatura das paredes de uma cubata.	2	54	fez a muntungila	fez o esqueleto da casa
311	muondo	muondo	umbundu	<i>n. f.</i> 1. Espírito tutelar. 2. <i>Pl. híbr.</i> Muondonas	2	33	possuir muondo	possuir espírito tutelar
312	mussaco	musako	kimbundu	<i>n. m.</i> tambor usado em julgamentos judiciais populares.	2	3	sítios de mussaco	sítios de tambor de julgamento

313	mussalo	kusala	kimbundu	<i>n. m.</i> 1. Peneira de palha fina em formato de jarro, com o fundo crivado. 2. O mesmo que mussualo. 3. <i>Pl. híbr.</i> Missalos	2	4 8 49 56	mão no mussalo trabalharam com mussalo primeiro mussalo iniciou o mussalo	mão na peneira trabalharam com peneira primeira peneira iniciou a peneira
314	mussenga	musenga	kimbundu	<i>n. m.</i> 1. Vara aguçada para a sementeira. 2. Esburacador agrícola. 3. <i>Deprec.</i> Perna delgada, sem beleza muscular. 4. <i>Pl. híbr.</i> Missengos.	2	24	de sua mussenga	de sua vara aguçada
315	mussuco	musuco	herero	<i>n. m.</i> 1. Moça que se submete, ou se submeteu recentemente, ao rito da puberdade ou de nubilidadade. 2. Observante do cerimonial do essuco. 3. O mesmo que <i>mufico</i> entre os Humbes e <i>muhico</i> entre os Nyaneikas.	2	206	por mussuco	por moça depois do rito da puberdade
316	mussula	mussula	kimbundu	<i>n. f.</i> 1. Fenda; racha. 2. <i>Fig.</i> Vagina.	2	22	guarda a mussula	guarda a fenda
317	mussunda	mussunda nganga	kikongo	<i>n. f.</i> 1. Espírito feminino de procedência conguesa. 2. Entidade espiritual da pacificação.	2	38	válido a mussunda	válido a espírito feminino
318	mussungo	kusunga	kimbundu	<i>n. m.</i> Crédito; abono.	2	22 27 58	o mussungo acolher o mussungo em mussungo	o crédito acolher o crédito em abono
319	mutunda	omutunda	nyaneika	<i>n. f.</i> Arbusto ou pequena árvore de copa compacta e folhas persistentes, de um verde acinzentado.	2	56 127	lidar e mutunda condições de mutunda	lidar e arbusto condições do arbusto
320	muturi	kutula	kimbundu	1. <i>n. m. e f.</i> Viúvo ou viúva. 2. <i>n. m.</i> Nome dado a indivíduo do sexo feminino, em homenagem a criatura nesse estado e, o qual, por afeto, se reduplica em <i>Tuturi</i> . 3.* <i>v.</i> Enviivar.	1	1636	reconhecer o rosto da nga muturi	reconhecer o rosto da senhora viúva

321	muxiluanda	muxiluanda	kimbundu	1. <i>n. m. e f.</i> Indivíduo natural da ilha adjacente à cidade de Luanda. 2. <i>p.ext.</i> Pescador dessa ilha. 3. <i>adj. unif.</i> Relativo a essa população. 4. <i>p. híbri.</i> Axiluandas.	1	238	acima de duzentos muxiluandas	mais de 200 habitantes da ilha de Luanda
322	muxima*	muxima	kimbundu	<i>n. m. e f.</i> 1. Pessoa de bom carácter. 2. Benfeitor. <i>n. m.</i> 3. Lisonjeio. 4. <i>Fig.</i> Fazer algo com sinceridade. 5.v. Cativar, lisonjear.	1	402 776	briguento, mas de boa muxima inquieta muxima	briguento, mas de bom carácter inquieta coração
323	muxito	kuxíta	kimbundu	1. <i>adj.</i> Selvagem.* 2. <i>n. m.</i> bosque; mata; floresta; selva; brenha. <i>Bras.</i> Caetê.	1	1071	escuros muxitos	escuros bosques
324	muzumbi*	muzumbi	kimbundu	<i>n. m.</i> 1. Lábios grossos. 2. Beiçola. 3. <i>Pl. híbr.</i> Mizumbos.	2	26	preparação de qualquer muzumbi	preparação de qualquer beiçola
325	mwana*	mwana	kimbundu	<i>n. m. e f.</i> 1. Filho. 2. <i>adj.</i> Relativo aos filhos.	2	35 43	participação de mwana diante do mwana	participação de filho diante do filho
326	ndengue	kandengue	kimbundu	<i>n. m. e f.</i> 1. Criança. 2. Garoto. 3. O mesmo que <i>cambonga</i> .	2	17 28	permitiu o ndengue o ndengue	permitiu a criança a criança
327	ndolo	ndôlo	kimbundu	<i>n. m.</i> 1. Espécie de búzio. 2. Pequena concha. 3. Planta medicinal da família das compostas. 4. Planta ornamental. 5. <i>n. f.</i> Dor.	2	8 9	terceiro ndolo ter ndolo	terceira dor ter pequena concha
328	ndonga	ndonga	kimbundu	<i>n. f.</i> 1. Árvore da família das leguminosas das leguminosas, cuja raiz ralada é utilizada em casos de inchação. 2. Rio nas gangueles. 3. Grupo; muita gente no mutemo.	2	16	tivesse no ndonga	tivesse no rio
329	ndua*	ndua (resultado de kulua, conquistar)	kimbundu	<i>n. f.</i> 1. Sangue nasal de cadáver de médium. 2. Testemunho mortuário de restituição do sangue dos sacrifícios, sorvido no cerimonial de aptificação mediúnica.	2	925	controlo de ndua	controlo de sangue nasal
330	ndumbo*	ondumba	nganguela	<i>n. m. e f.</i> 1. Leão, leoa. 2. O mesmo que curica.	2	66	o mesmo ndumbo	o mesmo leão

331	ndundum*	ndungu	kimbundu	<i>n. m. infant.</i> 1. Jindungo; condiment picante. 2. Tudo o que, como a malagueta ou certo tipo de conservas, é de sabor ardente.	2	35	que ndundum	que condimento picante
332	ndunduma*	ndunduma	kimbundu	<i>n. m.</i> Estrondo; estampido.	1	737	sorte do orgulhoso ndunduma	sorte do orgulhoso estrondo
333	ndungidi*	ndungidi	kikongo	<i>n. f.</i> Razão, justiça, vitória.	2	20 71	valer ndungidi campo, ndungidi	valer a razão campo, vitória
334	nga*	nga	kimbundu	1. <i>n. m. e f.</i> Abreviação de <i>Ngana</i> . 2. Senhor ou senhora. 3. Pessoa que exerce domínio como senhor sobre determinado território ou povo. 4. <i>adj. unif.</i> Relativo a senhor ou senhora. 5. <i>v.</i> Exercer domínio como senhor sobre determinado território ou povo.	1	56 107 257 291 698 1260 1636	súbito nga Mbunda vida de nga muhatu leite da nga Uála nga Féfa influência de nga Fefa palácio de nga Andembo rosto da nga	súbito senhora Mbunda vida de senhora casada leite da senhora Uála senhora Josefa influência da senhora Josefa palácio de senhor Andembo rosto da senhora
335	ngaieta	gaita	kimbundu	<i>n. f.</i> 1. Forma popular de gaita. Instrumento de fole. 2. Harmónica. Concertina. 3. <i>P. ext.</i> A dança <i>massemba</i> , por meio do acompanhamento da harmónica.	1	288	tocador de ngaieta	tocador de harmónica
336	ngala*	ngála	kimbundu	<i>adj.</i> 1. Aliviado. 2. Que tem pouco peso. 3. Livre. 4. Vestimenta que não é de luto. 5. Garrafa.	2	94	banho ngala	banho aliviado
337	ngalula*	ngalulu	kimbundu	<i>n. m.</i> 1. Garfo. 2. O mesmo que musandu.	2	239	de galula	de garfo
338	ngana	ngana	kimbundu	<i>n. m. e f.</i> 1. O que possui honras ou dinheiro. 2. Fidalgo. 3. Senhor; dono; possuidor; dominador. 4. Título que se dá a pessoas que não se tutelam. 5. O dono ou dona de casa em relação aos criados. 6. Aquele que tem domínio sobre o outro.	2	22	colocados de ngana	colocados do senhor

339	ngandu*	ngându	kimbundu	<i>n. m.</i> 1. Anfíbio carnívoro e perigoso, que habita os mais importantes rios do continente africano. 2. Crocodilo; caimão; jacaré. 3. Aligator.	2	46	consultas a ngandu	consultas a aligator
340	nganga	nganga	kimbundu	<i>n.</i> 1. Salina; lagamar. 2. Lugar à beira mar onde se fabrica o sal. 3. Bot. Nome por que no Seles e Amboim é conhecida a planta <i>mutuxi</i> . 4. <i>adj.</i> Sacerdote. Profeta. 5. Que tem ou revela grande saber. 6. Douto; mestre.	2	242	caveto nganga	caveto de mestre
341	ngango	ngángu	kimbundu	<i>adj.</i> 1. Encarniçado. 2. Ruívo.	2	21	pontos gango	pontos encarniçados
342	ngangula	ngangúla	kimbundu	<i>n. m.</i> 1. Operário que trabalha em ferro. 2. Ferreiro	2	17	apelidado ngangula	apelidado ferreiro
343	ngila	ngila	kimbundu	<i>conj. advers.</i> Mas, porém, todavia.	2	49	há compra, ngila	há compra, mas
344	ngo	ngo-ê/goê	kimbundu	1. <i>adj.</i> Depressivo; deprimido.* 2. <i>interj.</i> Só! Ai, que solidão.	1	1384	doce ngo	doce depressivo
345	ngoma	ngoma e ongoma	kimbundu	<i>n. f.</i> 1. Tambor comprido. 2. <i>P. ext.</i> Bombo.	2	24 216	a ngoma não há ngoma	não há tambor comprido o tambor comprido
346	ngongo*	ngongo	kimbundu	<i>v.</i> 1. Compadecer-se de; 2. <i>n.</i> Dor; padecimento; enfermidade. 3. <i>n. m.</i> Sofrimento; tribulação; padecimento; martírio; mundo. 4. <i>adj. e n.</i> Irmão gêmeo. 5. Um dos primeiros filhos nascidos do mesmo parto. 6. O par; o semelhante; gêmeo; <i>jingongo</i> .	2	23 90	na ngongo começar a ngongo	na dor começar a compadecer-se
347	ngonguembo*	ngongembu	kimbundu	<i>n. m.</i> 1. Pássaro noctívago. 2. Espécie de murcego.	2	18 31	a ngonguembo uma cimeira de ngonguembo	a pássaro noctívago uma cimeira de pássaro noctívago
348	nguami*	nguami	kimbundu	1. <i>Adv.</i> Não quero; não consinto; não concordo; não. 2. <i>n. f.</i> Recusa; negação.	2	53 79	apareceu nguami furos nguami	apareceu a recusa furos de negação

349	ngueji*	ngânji	kimbundu	1. <i>n. m. e f.</i> Julgador; pessoa que julga; juiz; desembargador; arbitro. 2. <i>Bot.</i> Nome por que na região do Zaire se conhece a planta <i>muhinji</i> . 3. Designação vaga de uma pessoa incerta ou daquela cujo nome se ignora. 4. Fulano; cicrano; beltrano.	2	286	era ngueji	era juiz
350	nguelenga*	ngelenge	kimbundu	<i>n. m.</i> 1. Antílope da corpulência do boi. 2. O mesmo que <i>ngalange</i> .	2	46	mas vimos nguelenga	mas vimos antílope da corpulência do boi
351	nguenda*	ngenda	kimbundu	<i>n. f.</i> 1. Cova de feras. 2. Covil.	2	98	parecia nguenda	parecia cova de feras
352	ngueza*	ngueza	kimbundu	<i>v.</i> Venham.	3	390	agora ngueza	venham agora
353	ngulo	ngulo	kimbundu	<i>n. m.</i> Leitão; bócoro.	2	53 12	vê ngulo venceu ngulo	vê o leitão venceu o leitão
354	ngumba*	ngúmba	kimbundu	<i>n.</i> 1. Arma branca, curta e perfurante. 2. Punhal. 3. Faca de dois gumes. 4. Estilete. 5. Choupa.. 6. <i>Adj. e n</i> Que trabalha. 7. Que faz obras; 8. Mestre. 9. Trabalhador. 10. Obreiro. 11. Amigo do trabalho. 12. Operário; oficial.	2	5 6	a ngumba projecto ngumba	o estilete projecto estilete
355	ngunga*	ngunga	kimbundu	<i>n. m.</i> 1. Qualquer instrumento sonoro. 2. Sino. 3. Som produzido pela pancada dada ao sino. 4. Badaladas.	2	56	aquele ngunga	aquele instrumento de sopro
356	nguvengue*	nguvenge	umbundu	<i>n. m. e f.</i> 1. Bébado. 2. Embriagado.	2	67	era nguvengue	era bébado
357	nguvulo*	nguvulo	kimbundu	<i>n. m. e f.</i> 1. Aquele a quem se confia o governo de um distrito ou província. 2. Governador. 3. Chefe do governo.	2	237	um nguvulo	um chefe do governo

358	njango	ondjangu	umbundu	<i>n. m.</i> 1. Refeitório tribunal da comunidade de uma sanzala. 2. Lugar onde as pessoas de um povoado se reúnem para comer e julgar, em primeiras instâncias, os seus próprios litígios. 3. Serão.	2	167	garagem do njango	garagem do serão
359	nhanga*	nhânga	kimbundu	1. <i>n. m.</i> Abreviatura de rinhangá; caçador. 2. <i>v.</i> Adestrar; habilitar.	2	21 82	concessão de nhanga em frente do nhanga	concessão de caçador em frente do caçador
360	nvula*	nvula	kimbundu	<i>n. f.</i> Aguaceiro, chuva; maré.	2	2 25 67	demais nvula associada em nvula segunda nvula	demais chuva associada a aguaceiro segunda chuva
361	nvumbi*	nvumbi	kikongo	<i>n. m. e f.</i> Morto; cadáver.	2	236 247	categoria de nvumbi o nvumbi	categoria de nvumbi o cadáver
362	nvunda*	nvunda	kimbundu	<i>n.</i> 1. Briga, luta, zaragata. 2. Desordem.	1	342	doutor, poco ya nvunda	doutor, faca de luta
363	nzamba*	nzamba	kimbundu	1. <i>n.</i> Elefante; sociedade; associação. 2. <i>adv.</i> A meias; em partes iguais.	2	52	estrutura e nzamba	estrutura e elefante
364	nzambi	nzambi	kimbundu	<i>n.</i> Deus; Santo; Divindade de qualquer religião; cada um dos membros da trindade cristã.	2	256	ngana Nzambi	meu Deus
365	nzukulu*	nzukulu	kikongo	<i>n. m. e f.</i> 1. Amigo; companheiro. 2. O mesmo que <i>kamba</i> em <i>Kimbundu</i> .	2	443	este e nzukulu	este e amigo
366	nzua	nzuua	kimbundu	<i>n. f.</i> 1. Hidromel. 2. Bebida composta de água e mel. 3. Alusão ao mel. 4. Sumo de múcua. 5. <i>v.</i> Fazer nzua, hidromel.	1	110 463	embriagante nzua preparava a embriagante nzua	embriagante sumo de múcua preparava o embriagante sumo de múcua
367	ogamba*	ogamba	umbundu	<i>n. f.</i> Andorinha.	2	40	possuir ogamba	possuir andorinha
368	pango	kubangulula	kimbundu	<i>n. m.</i> 1. Maneira de efetuar processo. 2. Moda. 3. Novidade.	2	20	morada e pango	morada e moda
369	pembe	pémbe	kimbundu	<i>adj.</i> 1. Ebúrneo que tem a cor de marfim. 2. Homo ia – : sem brilho.	1	1478 1643 1694	vau do pembe idem idem	vau ebúrneo cor de marfim idem idem
370	perar	perar		<i>v.</i> Ter relação sexual.	2	70	monta de perar	monta de ter relação sexual

			português em Angola			37	na condução, peram	na condução, têm relação sexual
371	pindali*	pindi	umbundu	<i>n. m.</i> Parte traseira da perna.	2	45	notou pindali	notou a parte traseira da perna
372	piô*	piô	português em Angola	1. n. m. e f. Criança. 2. Gír. Amante.	2	698	aquela piô	aquela amante
373	quiala	kiiala	kimbundu	<i>n. m.</i> 1. Indivíduo de ínfima condição de raça negra. 2. Selvagem. 3. Homem boçal. 4. Deprec. Aquele que, por falta de educação, procede grosseiramente.	2	19	revelações quiala	revelações de pessoa boçal
374	quianda	kuanda	kimbundu	1. n. m. Entidade sobrenatural das águas. Sereia. 2. P. ext. Criatura fisicamente anormal. 3. Bras. Iemanjá.	1	94 971 1901	calundus, quiandas falava das quiandas nem maquixis e nem quiandas	espíritos, sereias falava das sereias nem monstros e nem sereias
375	quiamba*	kiambamba	kimbundu	<i>n. m.</i> 1. Peso; carga. 2. Fig. Compromisso, responsabilidade, encargo. 3. kimbamba	2	176	levar a quiamba	levar o peso
376	quianga	kiamba*	kimbundu	<i>n. m.</i> Magia que concede a faculdade da imunidade contra ferimentos acidentais ou por agressão.	2	591	aquela quianga	aquela magia
377	quiba*	kíba	kimbundu	<i>n. m.</i> 1. Pele; cabedal; couro. 2. A pele de um animal separada do corpo.	2	66	estava no quiba	estava no couro
378	quibala	kibala	kimbundu	<i>n. f.</i> 1. Mandioca seca fermentada; Quifaça. 2. Variedade de feijão. 3. adj. m. e f. Ling. Que é relativo ao quibala, língua falada em Angola.	2	16 60	nomeadamente quibala ser quibala	nomeadamente mandioca seca ser feijão
379	quibila*	kibila	kimbundu	<i>n. f.</i> Sombra ou vulto que por nós, ou ao alcance da nossa vista, passa despercebida ou rapidamente.	2	126	bomba quibila	bomba de vulto
380	quibuco*	kibúko	kimbundu	<i>n. m.</i> 1. Ventura, sorte, felicidade. 2. Caravana.	1	239 2052	andrajosos das quibucas cruzavam com as quibucas	andrajosos das caravanas cruzavam com as caravanas
381	quilamba	kulambula	kimbundu	<i>n. m. e f.</i> 1. Intérprete das sereias. 2. Sacerdote do culto de tais entidades.	1	971	falava da quilamba	falava da intérprete das sereias

382	quilango*	kilánga	kimbundu	<i>n.</i> 1. Papada. 2. A maçã do peito das rezes. 3. Barbela; papeira. 4. Broncócele; bócio. 5. Papo; inflamação ou tumor na papada.	2	326	domínio de quilango	domínio de barbela
383	quilenda*	kilénda	kimbundu	<i>n. f.</i> 1. Prosperidade. 2. Felicidade; ventura. 3. Posto administrativo do concelho de Amboim, província do Kwanza-Sul.	2	55	outra quilenda	outra prosperidade
384	quilombo	kuanda	kimbundu	<i>n. m.</i> 1. Aldeia de escravos fugidos. 2. <i>adj.</i> Pessoa albina.	1	1021	espécie de quilombo	espécie de aldeia de escravos fugidos
385	quilumba	kulumbila	kimbundu	<i>n. f.</i> Moça. Rapariga nova. Donzela.	1	1050	tempos de quilumba	tempos de donzela
386	quimbamba	kimbamba	kimbundu	<i>n. f.</i> 1. Carga. 2. Cruz de Cristo. 3. Carga sagrada. 4. Nossa Senhora dos Passos. 5. <i>n. m.</i> Inseto; bicho; piolho; animáculo.	2	39	capacidade de quimbamba	capacidade de carga
387	quimbandice	kûbanda	kimbundu	<i>n. f.</i> 1. Sabedoria do quimbanda. 2. Ritual de quimbanda. 3. Bruxaria. 4. Tratamento de enfermidade prescrito por quimbanda.	1	1902	práticas mágicas e quimbandice	práticas mágicas e bruxaria
388	quimbazi*	kimbanzu	kimbundu	<i>n. m.</i> 1. Estiagem; falta de chuva. 2. Tempo seco.	2	57	a quimbazi	a estiagem
389	quimbo	ko + imbo	umbundu	<i>n. m.</i> Povoado. Sanzala. Casa isolada ou blocos de cas constituindo um só lar.	1	1071	passara novos quimbos	passara novos povoados
390	quiminha	kimínha	kimbundu	<i>adj. e n.</i> 1. Voraz. 2. Devorador. 3. Que traga ou engole. 4. O que devora ou arruina bens ou fortuna. 5. Sorvedouro. 6. <i>Bot.</i> Planta utilizada como contra-peçonha.	2	8	mão da quiminha	mão do devorador
391	quimoma	kimóma	kimbundu	<i>n.</i> 1. Aumentativo de <i>móma</i> . 2. Piton.	2	4	passam quimomas	passam pitons
392	quinanga*	kinánga	kimbundu	<i>n. m.</i> 1. Lugar de estar. 2. Largo; praça. 3. <i>Fig.</i> Mercado; feira.	2	57	contornou o quinanga	contornou o largo

393	quinda*	kínda	kimbundu	<i>n. f.</i> 1. Artefacto de mateba em forma de bacia. 2. Teiga. 3. Cesto para a condução de materiais de construção. 4. Cestinha de palha enfeitada a cores. 5. <i>adj. num. ord.</i> Quinta. 6. O dia de quinta-feira. 7. Espécie de cesta. 8. Cesta regional angolana.	2	369	aqui vemos a quinda	aqui vemos o cesto tradicional
394	quindumba	kulumba	kimbundu	<i>n. f.</i> 1. Poupa. 2. Elevação de cabelo ou de penas.	1	373 463 1201	olho ao cheiroso corpo quindumba corpos e as quindumbas perfumada quindumba	olhos ao cheiroso corpo de poupa corpos e as elevações de cabelo perfumado cabelo
395	quindumbo	kulumbula	kimbundu	1. <i>adj.</i> Calorífero, soalheiro. * 2. <i>n. m.</i> Cobertor; manta; agasalho.	1	1352	tempo de quindumbo	tempo de soalheiro
396	quingando*	kingáнду	kimbundu	<i>n. m.</i> 1. <i>Bot.</i> Planta da família das leguminosas de propriedades narcóticas. 2. Arurão. 3. Arrebatador; destruidor; sanguinário. 4. Cobarde.	2	22	acertou o quingando	acertou o arurão
397	quinguela*	kingelengende	kimbundu	<i>n. m.</i> Caco de castanha de dendém.	2	163	venha quinguela	venha caco de castanha de dendém
398	quinhamel*	kinhami	kimbundu	<i>n. m.</i> 1. Sentimento de atração moral. 2. Afeição; apego; simpatia de uma filho por seus pais. 3. Amizade; amor.	2	56	sobre o quinhamel	sobre o amor
399	quinzari	kuzana + kúria	kimbundu	<i>n. m.</i> 1. Pantera. 2. <i>P. ext.</i> Monstro com o corpo de fera, mas possuindo os pés humanos. 3. <i>Fig.</i> Pessoa furiosa e má.	1	986 1232	conhecido por quinzari marcado, e quinzaris	conhecido por pantera marcado, e pantera
400	quioco	txokwe	cokwe	<i>adj. e n. m.</i> 1. Forma deturpada de txokwe, usual na época colonial. 2. Autóctone da Lunda, a nordeste de Angola, constituindo membro da tribo dessa região.	1	972	Deus muene-zambi-yameia quiocos	Deus muene-zambi-yameia quiocos
401	quipembe*	kupemba	umbundu	<i>v.</i> 1. Assoar. 2. Falar quando necessário.	2	191	devida quipembe	devida fala oportuna
402	quipindi*	pindi	umbundu	<i>n.</i> Parte traseira da perna.	2	24	arranhou no quipindi	arranhou na parte traseira da perna
403	quipuna*	kipuna	kimbundu	<i>n. m.</i> 1. Cotovelo. 2. Parte da manga que cobre o cotovelo.	2	53	tenção por quipuna	tenção pelo cotovelo

404	quirima*	kirima	kimbundu	<i>n.</i> 1. Ato de andar para trás. 2. Recuo. 3. Lavra e seus produtos. 4. Herdade; granja; horta. 5. <i>adj.</i> Improdutivo.	2	28	ensaio quirima	ensaio quirima
405	quirindo*	kirinda	kimbundu	<i>adj.</i> 1. Aleijado. 2. Deformado. 3. Diz-se da pessoa que não pode andar, ou que o faz de rastos. 4. Imperfeito; inerte; 5. Pessoa de membros estropiados.	2	41	do quirindo	do aleijado
406	quissanguela	kusangela	kimbundu	<i>n. f.</i> 1. Sociedade. Parceria. Cotização. Contributo. Coleta. 2. <i>loc.</i> Homem de _: o que pertence a mais de uma mulher. Mulher de _: a que pertence a mais de um homem.	1	377	mulher de quissanguela	mulher que tem a mais de um homem
407	quissonde	kusondoloka	kimbundu	<i>n. m. e f.</i> 1. Formigão avermelhado, de mordedura dolorosa. 2. <i>Bras.</i> Crauçanga. Murupeteca. Taoca.	1	1025	ninho de quissondes	ninho de formigão
408	quissongo	kusonga	kimbundu	1. <i>n. m.</i> Cancro. Enfermidade cancerosa. Úlcera de procedência sífilítica ou venérea. 2. <i>n. m. e f.</i> Indivíduo natural de Quissongo, na região do Libolo, a sul do rio Kwanza. 3. <i>Pl.</i> População dessa área, mas de reduzida importância, pertencente ao grupo étnico do Kimbundu. 4. <i>adj. unif.</i> relativo a essa população.	1	333	abria quissongo	abria úlcera
409	quissua*	kisua-njila	kimbundu	<i>n. m.</i> 1. Maltrapilho; farroupilha. 2. Pelintra.	2	620	quissuas do panguila	maltrapilhos do panguila
410	quissueia*	kisueia	kimbundu	<i>n. m. e f.</i> 1. Nome genérico dos mamíferos carnívoros. 2. Fera. 3. <i>Fig.</i> Pessoa cruel. 4. <i>adj.</i> Sanguinário; déspota.	2	430	governo de quissueia	governo de sanguinários
411	quitari	kutarika	kimbundu	<i>n. m.</i> 1. Dinheiro. 2. O mesmo que lombongo em deturpação de olombongo. 3. <i>Bras.</i> Itajuba; tutu (<i>gír.</i>).	2	353	todo quitari	todo dinheiro

412	quiteumbo	kutembuka	kimbundu	<i>n. m.</i> Peido; bufa.	2	54	sentiu o quiteumbo	sentiu o peido
413	quiteque	kutejeja	kimbundu	<i>n. m.</i> 1. Ídolo. 2. Qualquer imagem de madeira ou barro. 3. Manipanso. 4. <i>Deprec.</i> Indivíduo feio e mal feito.	2	15	reforço de quiteque	reforço de ídolo
414	quitoto	kutotoka	kimbundu	<i>n. m.</i> Variedade de cerveja de milho.	1	769	semana, quitoto	semana, variedade de cerveja de milho
415	raboteiro*	raboteiro	português em Angola	<i>n. m.</i> 1. Pessoa que transporta mercadorias ou materiais empurrando um veículo, geralmente de madeira, com apenas uma só roda. 2. Cangulo.	2	62	expressão de raboteiro	expressão de transportador de mercadorias
416	rivungo*	rivungo	kimbundu	<i>n. m.</i> Casacão, capote.	3	145	avaliado rivungo	avaliado o casacão
417	sabalo*	sabalu	kimbundu	<i>n. m.</i> 1. Sapato. 2. <i>Kikoto</i> .	2	262	categoria sabalo	categoria de sapato
418	sacupalica*	kuparika	kimbundu	<i>v.</i> 1. Entalar. 2. Apresilhar; engate. 3. Entalação.	2	57	arrecadam para sacupalica	arrecadam para entalação
419	sapalalo*	sapalalo	kimbundu	<i>n. m.</i> Indivíduo nascido no sábado.	1	580	gasto sapalalo	gasto sabático
420	sekulo	osekulu	umbundu	<i>n. m.</i> 1. Conselheiro do soba. Maioral do quimbo. Varão de respeitabilidade, quer pela posição social, quer pela idade. 2. O mesmo que <i>macota</i> entre os povos da língua quimbundo.	1	1810	amanhã um sekulo	amanhã um conselheiro do soba
421	sembar	kusemba	kimbundu	1. <i>n.</i> Dança tradicional angolana. 2. <i>v.</i> Dar sembas; dançar a massemba.	2	27	têm sido semba	têm sido semba
422	soalala*	sualala	kimbundu	<i>n.</i> 1. Salalé; caprim; formiga branca. 2. Termite. 3. <i>Fig.</i> Inimigo; traidor.	2	32	o que era soalala	o que era formiga branca
423	soba	soba	kimbundu	<i>n. m.</i> 1. Chefe de tribo africana; régulo. 2. Pessoa que, pela sua riqueza ou por outro motivo, exerce poder económico ou político sobre a população menos informada. 3. <i>Iron.</i>	1	429 1022	algum soba da Quissama fugidos a sobas despóticos	algum régulo da Quissama fugidos à autoridade tradicional

				Indivíduo de maior preponderância num meio. 4. Chefe de família. 5. O que possui muitas mulheres. 6. <i>adj. unif.</i> O maior; o principal.		1034	trezentos homens de soba	300 homens da autoridade tradicional
						1041	fúria de sobas	fúria de autoridades tradicionais
424	sobos	sobongo	kimbundu	<i>n. m.</i> 1. Arbusto de fruto medicinal. 2. Sobongo.	2	1233	com sobos	com arbusto de fruto medicinal
425	songo	kusongoloka	kimbundu	<i>n. m.</i> Formigão preto, mais pequeno que o zeu.	1	346	causa dos songos , praguejava	causa dos formigão preto , praguejava
426	sonhi	sonhi	kimbundu	<i>n.</i> Pejo; pudor; vergonha.	2	54	wa kambo sonhi	não tem vergonha
427	sukuama	kukuama	kimbundu	<i>interj.</i> 1. (designativa de repulsa) Poças! Pôpilas! 2. Exprime repulsa, com os demónios!, apre!, basta!	1	534 538 1808	sukuama! idem idem	poças! idem idem
428	sunguilar	kusungila	kimbundu	<i>v. intr.</i> Seroar, passar a noite, pernoitar.	1	393	sunguilar no pavilhão	pernoitar no pavilhão
429	tacula	kuta + kula	kimbundu	<i>n. f.</i> Árvore bastante elevada que se encontra nas ravinas de Icolo e Bengo, florestas do Kwanza-Norte e Kwanza-Sul.	1	1811	feito de tacula	feito de árvore bastante elevada
430	tantar*	tantar	português em Angola	<i>v.</i> Enlouquecer; estar demente.	2	215	estava a tantar	estava a enlouquecer
431	tchambuluka*	tchambula	umbundu	<i>n. m.</i> Adivinho; feiticeiro; kimbanda.	2	158	viu o tchambuluka	viu o adivinho
432	tchibita*	tchibita	kimbundu	<i>n. f.</i> Flauta.	1	472	verdiano tchibita	flauta esverdeada.
433	tchikanda*	tchitanda	umbundu	<i>n.</i> Feira; negócio.	2	129	vêm da tchicanda	vêm da feira
434	tchilima*	tchilima	umbundu	1. <i>n. m.</i> Luz da lua. 2. <i>adj.</i> Incircunciso.	2	228	era tchilima	era incircunciso
435	tchinguvo	xingufo	cokwe	<i>n. m.</i> 1. Forma diversificada de xingufo. 2. Variedade de tambor.	3	27	mina a custo tchinguvo	mina a custo de tambor
436	tchimpande*	tchimpandi	umbundu	<i>n. m.</i> Sítio ou lugar sem vegetação.	2	21	trabalhar nos tchimpande	trabalhar nos lugares sem vegetação

437	tchipa	xipala (de oxipala)	umbundu	<i>n. f.</i> 1. Cara; rosto; semblante. 2. <i>P.ext.</i> Fotografia.	2	35	registo atualizado de tchipa	registo atualizado da fotografia
438	tchiputo*	tchipuku	umbundu	<i>n. m.</i> Desprezo.	2	438	era mesmo tchiputo	era mesmo desprezo
439	tchitue*	tchitwe	umbundu	<i>n. m.</i> 1. Cabeça grande; 2. <i>n. m. e f.</i> Líder.	2	24	era nosso tchitue	era nosso líder
440	tchitungo	tchitungo	umbundu	<i>n. m.</i> Feixe	2	187	água, luz e tchitungo	água, luz e feixe
441	telo*	telo	português em angola	<i>n. m.</i> Telefone público, cuja utilização é paga.	2	140	está no telo	está ao telefone público
442	tujir*	túji	kimbundu	1. <i>n. m.</i> Excremento. Imundície. 2. <i>v. intr.</i> Expelir naturalmente os excrementos.	1	63	tuji , oh porra!	defeca , oh porra!
443	tuga*	português	português em angola	<i>n. m.</i> 1. Portugal; Autoridade portuguesa. 2. Maneiras portuguesas. 3. <i>adj. unif.</i> Que diz respeito à Portugal.	2	28	observador baixo tuga	observador baixo português
444	twabixila*	twabixila	kimbundu	<i>v.</i> Já chegamos.	2	523	dizia twabixila	dizia já chegamos
445	uadila*	uadila	kikongo	<i>v.</i> Chorar, lamentar.	1	168	bu uadila	onde chorava
446	uala*	kualalesa	kimbundu	<i>n. f.</i> Cerveja de milho.	1	1072	vavó Uala	avó Uala
447	ualar*	kualalesa	kimbundu	<i>v.</i> Aquecer ou beber cerveja de milho.	2	25	aprender a uala	aprender a beber a cerveja de milho
448	uála-das-ingombotas*	uála-das-ingombotas	kimbundu	<i>n. f.</i> Cerveja de milho produzida no bairro da cidade de Luanda onde se aceitavam escravos foragidos, localizado na encosta fronteira à Igreja do Carmo.	1	1072	uála-das-Ingombotas diria	diria que é cerveja de milho produzida no bairro das Ingombotas
449	ulombe	kulomba	kimbundu	<i>n. m.</i> 1. Pó preto de folhas vegetais. 2. Variedade de caraterizador ritual.	2	542	levou o ulombe	levou o pó preto de folhas vegetais
450	undengueuami*	undengueuami	kimbundu	<i>n. m. e f.</i> Meu menino; minha menina.	2	15	virem undengueuami	virem o meu filho
451	uandi*	uandi	kikongo	<i>v.</i> Chorar, lamentar.	1	1806	como uandi coxilava	como chorava e adormecia

452	uanga	ouanga	termo regional	1. <i>adj.</i> Propinado por ocultistas. 2. <i>n. m. e f.</i> Ciência do feiticeiro. 3. Feitiço; bruxaria; malefício; veneno ou droga nociva.	1	1813	tua uanga	tua bruxaria
453	ukamba*	ukámba	kimbundu	<i>n.</i> Amizade; camaradagem; companhia.	2	15	via a ukamba	era notável a camaradagem
454	wabiluka	wabiluka	kimbundu	<i>v.</i> Transformar.	2	20	wabiluka monami	transformar o meu filho
455	wami*	wâmi	kimbundu	<i>adj. e pron. pess.</i> 1. Meu; minha. 2. Adv. Também; igualmente.	2	140	wami maka	meu problema
456	wanga*	ouanga	kimbundu	<i>n. m. e f.</i> 1. Ciência de feiticeiro. 2. Feitiço; bruxaria. 3. Malefício. 4. Veneno ou droga nociva, propinados por ocultistas.	2	162	centro wanga	centro de feitiçaria
457	wakala	kala	kimbundu	<i>adj.</i> 1. Qualquer (entre dois ou mais). 2. Cada; cada um; cada qual. 3. Do mesmo modo que. 4. <i>Interj.</i> Esteja; fique.	2	59	assumir o wakala	assumir qualquer
458	ximinha	de muxima	kimbundu	<i>n. f.</i> 1. <i>Deprec.</i> Mulher de vestuário de panos. 2. Senhora de mentalidade atrasada, obrando segundo os princípios tradicionais.	1	1567 1578	Ximinha Jianju idem	mulher vestida de panos apetitosa idem
459	xinguilar	kuxinguila	kimbundu	<i>v.</i> Estar sob influência de entes sobrenaturais com agitação nervosa.	2	56	as democratas a xinguilar	as democratas sob agitação nervosa
460	xuaxualhar	xuaxualhar	kimbundu	1. <i>v.</i> Galantear, seduzir, paquerar, farfalhar. 2. <i>n.</i> Galanteio, sedução.	1	70 1244	xuaxualhar de asas xuaxualhar das folhas	farfalhar de asas farfalhar das folhas
461	zage*	nzaji	kimbundu	<i>n. f.</i> Fásca elétrica; corisco.	2	18	parecia zage	parecia fásca elétrica
462	zambi	nzambi	kimbundu	<i>n.</i> Deus; Santo; Divindade de qualquer religião; cada membros da trindade cristã.	2	42	resultante interpretação de Zambi	resultante interpretação de Deus
463	zemba	nzêmba	kimbundu	<i>n. f.</i> 1. Pano suspenso do pescoço para o peito, no qual se acomoda uma criancinha. 2. Cinta ritual fechada, contendo, segundo a entidade a que se consagra, ingredientes vários ou aparas de cadáver.	2	9 34	entrar de zemba sem esquecer a zemba	entrar com cinta ritual sem esquecer a cinta ritual
464	zuela*	kuzuela	kimbundu	<i>v.</i> 1. Falar. 2. Comunicar-se.	2	46	módulos de zuela	maneiras de se comunicar

465	zungueiro*	nzunga	kimbundu	<i>adj.</i> 1. Aquele que anda sem objetivo. 2. Aquele que pratica a venda ambulante.	3	25	mamã zungueira	mamã vendedora ambulante
-----	------------	--------	----------	---	---	----	-----------------------	---------------------------------

Anexo II – Lista de empréstimos nominais próprios

Nº	Empréstimo	Étimo	Etimologia	Doc	Significação	Nota
1	Amboim	Mbui	Umbundu	1	1. <i>n. m. e f.</i> Indivíduo natural ou habitante do Amboim na região do Kwanza-Sul. 2. <i>n. m.</i> Língua bantu do subgrupo umbundu falada na região de Amboim. 3. <i>Pl.</i> População dessa área, pertencente ao grupo étnico dos Umbundos. 4. <i>adj. unif.</i> Relativo a essa população.	
2	Ambriz*	Mbriz	Kimbundu	1	<i>n. m.</i> Região ao Norte de Angola entre o <i>Kisembu</i> e o <i>Musulu</i> .	
3	Ambuela	Ambuela	Kimbundu	2	1. <i>adj. m. e f.</i> Que diz respeito ou pertence aos Ambuelas, grupo angolense do alto Cubango. 2. <i>n. m. e f.</i> Indivíduo pertencente ao povo dos Ambuelas. 3. <i>n. m.</i> Língua de Angola.	
4	ANANGOLA*	ANANGOLA	Kimbundu e português	2	<i>n. f.</i> Sigla da Associação dos Naturais de Angola.	
5	ANAZANGA*	ANAZANGA	Português em Angola	2	<i>n. f.</i> Sigla da Associação dos Naturais e Amigos da Ilha do Cabo.	
6	Andembo	Ndembu	Kimbundu	1	1. <i>n. m. e f.</i> Indivíduo natural de Dembos ao Norte do rio Kwanza. 2. <i>n. m.</i> Medicamento externo gorduroso. Droga manipulada com folhas de óleo vegetal, sobretudo de palma, e em certos casos com essências aromáticas. Unguento. Fricção. Paliativo.	Lematizado como dembo.
7	Andembo-ya-Tata*	Ndembu-ya-Tata	Kimbundu	1	<i>n. m.</i> Autoridade suprema tradicional da região dos Dembos. Régulo. Soba que tem sob a sua jurisdição outros sobas.	
8	Andongo*	Ndongo	Kimbundu	1	<i>n. m.</i> 1. Variedade de embarcação utilizada na pesca, constituída por uma só peça de madeira; espécie de canoa. 2. Pimenta ou pimentão usado como condimento culinário. 3. Adereço de pescoço, feito de búzios e pedaços de ovos de avestruz.	
9	Andulo	Ndulu	Umbundu	2	1. <i>n. m. e f.</i> Natural de Andulo, entre os rios Cunhinga a este e Cutato a oeste, no planalto do Bié. 2. <i>Pl.</i> População dessa área pertencente ao grupo étnico dos Umbundos. 3. <i>adj. unif.</i> Relativo a essa população.	
10	ANEP	ANEP	Português em Angola	2	<i>n. f.</i> Sigla da Associação Nacional do Ensino Particular.	
11	ANGOP*	ANGOP	Português em Angola	2	<i>n. f.</i> Sigla da Agência Angola Press, agência de notícia do Estado Angolano, fundada em 1975 ex-aliada da extinta agência de notícia oficial da União Soviética, Agência Telegráfica da União Soviética.	
12	Assango*	Asango	Umbundu	2	<i>n. m.</i> Região da província do Kwanza-Sul.	
13	Bakama	Bakama	Kikongo	2	<i>s. m.</i> Grupos culturais localizados na província de Cabinda, conglomerando os do Tchizo, Tchinzazi, Susu e do Ngoyo.	
14	Balombo*	Balombo	Umbundu	2	<i>n. m.</i> Município da província de Benguela	
15	Bendinha	Mbende	Kimbundu	2	<i>n. m.</i> 1. Antiga moeda de cobre do valor de 15 reis. 2. Meia macuta.	

16	Bengo*	Mbengu	Kimbundu	2	<i>n. m.</i> Nome dado à região ao Norte de Angola em homenagem ao poder dos sobas <i>Mani Mbengu</i> (que reinou a região até a conquista dos portugueses) e <i>Mani Ikolo</i> (sendo vizinho); esta região é banhada pelos rios <i>Kwanza</i> , a Sul, e <i>Nzenza</i> (nome vernáculo do rio) ou <i>Mbengu</i> (nome dado ao mesmo rio mais tarde em homenagem ao soba <i>Mani Mbengu</i>), ao Norte.
17	Benguela	Venga, Mbenga*	Umbundu*	2	1. * <i>n. f.</i> Província da República de Angola. 2. <i>adj. m. e f.</i> Que diz respeito ou pertence aos Benguelas, povo bantu que habita a região de Benguela. 3. <i>m. e f.</i> Indivíduo pertencente ao povo dos Benguelas. 4. * <i>v.</i> Ser turva (com respeito à água do rio ou lagoa)
18	Bengui*	Bênge	Kimbundu	2	<i>adj.</i> 1. Diferente. 2. Em outro lugar; em outra parte. 3. Em sítio diferente.
19	Bibala*	Bibala	Umbundu	2	<i>n. f.</i> Município da província do Namibe.
20	Bicuar*	Bicuar	Umbundu	2	<i>n. m.</i> Parque nacional de Angola, localizado na província da Huíla.
21	Bié*	Wye ou vye	Umbundu	1	1. <i>n. m.</i> Província localizada no planalto central de Angola. 2. Natural do Bié, no planalto central de Angola. 3. Dialeto falado nessa região. 4. <i>Pl.</i> População da região do Bié, pertencente ao grupo étnico dos Umbundos.
22	Bimbi*	Ombindi	Nhaneka	2	<i>n. m.</i> 1. Designação genérica de folhas tenras ou ervas, com aplicação em esparregados regionais. 2. O mesmo que nômbo entre os Humbes.
23	BNA*	BNA	Português em Angola	2	<i>n. m.</i> Sigla do Banco Nacional de Angola.
24	Bocoio*	Mbocoio	Umbundu	2	<i>n. m.</i> Município da província de Benguela.
25	Bolombo*	Mbalombo	Umbundu	2	<i>n. m.</i> Município da província de Benguela.
26	Bolongongo*	Mbolongongo	Kimbundu	2	<i>n. m.</i> Município da província do Kwanza-Norte.
27	Buangongo*	Nambuanguongo	Kimbundu	2	<i>n. m.</i> Cidade do município do Bengo, província de Luanda.
28	Buco-Zau*	Buco-Zau	Kikongo	2	<i>n. m.</i> Cidade e município da província de Cabinda.
29	Buíla*	Bwela	Kikongo	2	<i>n. f.</i> Vila do município de Cuimba, na província do Zaire.
30	Bunga*	Mbunga	Kikongo	2	<i>n.</i> Região da província do Uíje.
31	Bungo	Mbungu	Kimbundu	1	1. <i>n. m.</i> Magia para obtenção de grandes proventos e prontas facilidades. 2. <i>n. m.</i> Antigo bairro da cidade de Luanda, localizado na zona da Baixa, nas mediações da estação do caminho-de-ferro. 3. <i>n. m. e f.</i> Indivíduo natural da região do Bungo, na região da Damba, ao norte de Angola. 4. <i>Pl.</i> População dessa área pertencente ao grupo étnico dos Quicongos, possuindo afinidades com Sossos e Muxicongos. 5. <i>adj. unif.</i> Relativo a essa população.

32	Bwila*	Bwela	Kikongo	2	<i>n. f.</i> Vila do município de Cuimba, na província do Zaire.	
33	Caála*	Kahala	Umbundu	2	1. n. f. Município da província do Huambo. 2. n. m. e f. População dessa área, pertencente ao grupo étnico Umbundu. 3. adj. unif. Relativo à essa população.	
34	Cabenda	Kambinda	Kikongo	2	1. adj.m. e f. Que diz respeito ou pertence aos Cabindas, grupo bantu que habita Cabinda, província costeira de Angola. 2. n. m. e f. Indivíduo pertencente ao povo dos Cabindas. 3. n. m. e f. Negro apresentado na costa norte de Angola e trazido como escravo para o Brasil. 4. n. m. Língua do grupo quicongo falada por este povo.	Lematizado como cabinda.
35	Cabindês	Kambinda	Kikongo	2	1. n. m. e f. Natural de Cabinda, ao norte de Angola, na margem direita do Zaire. 2. adj.unif. Relativo a população de Cabinda. O mesmo que Cabinda.	Lematizado cabindense
36	Cabingano*	Kabinganu	Kimbundu	2	<i>adj.</i> 1. Sucessor; substituto. 2. Que pode suceder a outrem. 3. Aquele que passa a ter os mesmos direitos.	
37	Cabolombo*	Kabolombo	Kimbundu	2	<i>n. m.</i> Distrito urbano do município de Balas, na província de Luanda.	
38	Cachiungo*	Katchiungu	Umbundu	2	1. n. m. Município da província do Huambo. 2. n. m. e f. Indivíduo pertencente ao povo do Cachiungo.	
39	Cacoma*	Kakoma	Umbundu	2	<i>n. m.</i> Vila da província do Huambo.	
40	Caconda	kukondojoaka	Kimbundu	2	1. adj. m. e f. Que diz respeito ou pertence aos cacondas, grupo étnico que habita Caconda no planalto central de Angola. 2. n. m. e f. Indivíduo pertencente aos Cacondas. 3. Hist. Negro escravizado, trazido ao Brasil a partir da região de Caconda. 4. Ling. Língua bantu do grupo Umbundu falada pelo povo de Caconda.	
41	Caconga*	Cacongo	Cokwe	2	<i>adj. m. e f.</i> Que diz respeito ou pertence aos Cacongos, grupo étnico que habita a margem direita do rio Zaire, no Sudoeste de Angola.	
42	Cacongo	Cacongo	Cokwe	2	<i>n. m. e f.</i> 1. Indivíduo pertencente ao povo dos Cacongos. 2. Natural da área da Lunda, no sobado de Txicolondo. 3. População dessa área pertencente ao grupo Lunda-Cokwe. 4. n. m. Ling. Língua falada pelos Cacongos. 5. Ang. Zool. Nome vulgar de uma ave (<i>Totanus glareola</i>). 6. Zool. Designação de certa variedade africana de salmão.	
43	Cacuaco*	Kakuaku	Kimbundu	2	1. n. m. Município da província de Luanda. 2. n. m. e f. Indivíduo pertencente ao povo de Cacuaco. 3. n. m. e f. Indivíduo natural de Cacuaco. 4. População pertencente à área de Cacuaco. 5. n. m. Não é aqui. 6. Mão.	
44	Caculama*	Kakulama	Kimbundu	2	<i>n. f.</i> Também conhecido por Mucari, é um município da província de Malanje.	
45	Caculo	Kukûla	Kimbundu	2	<i>n. m.</i> 1. Variedade de feijão. 2. Eva. 3. adj. Gémeo nascido em primeiro lugar.	
46	Cacuso	Kakuzu	Kimbundu	3	<i>n. m.</i> Município da província de Malanje.	

47	Cafala*	Cafala	Kimbundu	2	<i>n. m.</i> Grupo musical angolano.
48	Cafunfu*	Kafunfu	Cokwe	2	<i>n. m.</i> 1. Vila da província da Lunda-Norte. 2. População pertencente à área de Cafunfu.
49	Cafuta*	Cafuta	Kimbundu	2	1. n. m. Empresa angolana. 2. adj. unif. Relativo a essa empresa.
50	Cahala*	Cahala	Umbundu	2	<i>n. f.</i> Município da província do Huambo.
51	Cahango*	Cahango	Kimbundu	2	<i>n. m.</i> Comuna do município de Icolo e Bengo, na província de Luanda.
52	Cahenda*	Cahenda	Kimbundu	2	<i>n. m.</i> Vila da comuna de Bula-Atumba, do município do Bengo, na província de Luanda.
53	Caimbambo*	Caimbambo	Umbundu	2	<i>n. m.</i> Município da província de Benguela.
54	Cainde*	Cainde	Nyaneka	2	<i>n. m.</i> Vila do município de Virei, na província do Namibe.
55	Caingo	Caingo	Kikongo	2	<i>n. m.</i> Vila da província do Uíge.
56	Caiundo*	Caiundo	Nganguela	2	<i>n. m.</i> Vila da província do Kwando-Kubango.
57	Calandula	Calandula	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de Calandula, ex-Duque de Bragança, na região de Malanje a sul de Camabatela. 2. Pl. População dessa área pertencente ao grupo étnico Kimbundu.
58	Calawenda*	Calawenda	Kimbundu	2	<i>n. m.</i> Distrito urbano do município do Cazenga, na província de Luanda.
59	Calemba	Kalemba	Kimbundu	3	<i>n. f.</i> 1. Bairro da província de Luanda. 2. Ondulação fortemente agitada, quer do mar, quer do rio. 3. Marulhada. 4. O mesmo que calema em deturpação portuguesa. 5. v. Agir incoerentemente.
60	Calenga*	Calenga	Umbundu	2	<i>n. m.</i> Vila do município da Caála, na província do Huambo.
61	Calombolo*	Calombolo	Umbundu	2	<i>n. m.</i> 1. Rio de Angola. 2. Salina localizada na província de Benguela.
62	Calú	Caluanda	Kimbundu	1	1. n. m. e f. Indivíduo natural ou habitante da cidade de Luanda. 2. adj. unif. Relativo a essa população.
63	Calualua	Calualua	Kimbundu	2	<i>n. m.</i> Rio de Angola, na província de Benguela.
64	Caluanda	Caluanda	Kimbundu	2	1. n. m. e f. Natural ou habitante de Luanda. 2. adj. unif. Relativo a essa população.
65	Caluanga	Caluanga	Umbundu	2	<i>n. m.</i> Vila da província do Kwanza-Sul.
66	Calueque	Calueque	Kwanyama	2	<i>n. f.</i> 1. Povoação da comuna de Naulila, município de Ombadja na província do Cunene. 2. Barragem construída no leito do rio Cunene.
67	Calulo*	Kulûla	Umbundu	2	<i>n. m.</i> Município localizado na província do Kwanza-Sul.
68	Calumbo	Calumbo	Kimbundu	3	<i>n. f.</i> 1. Espírito feminino que promove a esterilidade. 2. Entidade espiritual da esterilidade. 3. Comuna do município de Viana na província de Luanda, onde se localiza o santuário denominado S. José de Calumbo.
69	Calumbula*	Kulumbula	Umbundu	2	<i>n. f.</i> Região localizada na província do Kwanza-Sul.

70	Calundilili*	Calundilili	Kimbundu	2	<i>n. m.</i> Escola do município de Chinguar, província do Bié.
71	Calundungo*	Calundungo	Cokwe	2	<i>n. m.</i> Região da província do Moxico.
72	Caluquembe*	Kalukembe	Umbundu	2	<i>n. m.</i> Município da província da Huíla.
73	Camabatela*	Kamabatela	Kimbundu	2	<i>n. f.</i> 1. Comuna do município de Ambaca, na província do Kwanza-Norte. 2. O mesmo que Mufongo.
74	Camama*	Camama	Português em Angola	3	1. n. m. Comuna do município de Bela. 2. n. f. Cemitério localizado na província de Luanda. 3. f. Falsidade, mentira, decepção.
75	Camama*	Kamama	Kimbundu	2	<i>n. m.</i> 1. Distrito urbano do município de Belas, na província de Luanda. 2. n. m. e f. Indivíduo natural de Camama, na província de Luanda. 3. n. m. Cemitério localizado na província de Luanda.
76	Camame*	Kamame	Kimbundu	2	<i>n. m.</i> Vila do município de Gonguambo, na província do Kwanza-Norte.
77	Camanongue*	Kamanongue	Kimbundu	2	1. n. m. Município localizado na província do Moxico. 2. n. m. e f. Indivíduo natural ou residente de Camanongue, província do Moxico.
78	Camaxilo	Kamaxilo	Cokwe	2	<i>n. m.</i> Vila do município de Caungula, na província da Lunda-Norte.
79	Cambala	Kimballa	Kimbundu	2	1. n. m. e f. Indivíduo natural ou residente da localidade da margem do Cassai, na região da Lunda, a nordeste de Angola. 2. Pl. População desta área pertencente ao grupo étnico dos Lunda-Txokwe. 3. adj. unif. Relativo a essa população.
80	Cambamba*	Kambamba	Kikongo	2	<i>n. m.</i> Vila do município do Kitexe, na província do Uíge.
81	Cambambe*	Kambambi	Kimbundu	1	<i>n. m.</i> 1. Município da província do Kwanza-Norte. 2. Derivado do <i>kimbundu mbambi</i> (<i>gazela</i>), o lexema <i>kambambi</i> é o diminutivo de <i>gazela</i> , cervo.
82	Cambambe*	Kambambi	Kimbundu	2	<i>n. m. Zool.</i> Corça; veado pequeno; gazela. 2. Fig. Pessoa esperta; ladina.
83	Cambanda*	Kambanda	Umbundu	2	<i>n. f.</i> Município da província de Benguela.
84	Cambinda*	Kambinda	Kikongo	2	<i>n. m. e f.</i> Indivíduo natural ou residente de Cabinda.
85	Cambondo*	Kambóndo	Kimbundu	2	<i>n. m.</i> Vila do município de Gulungo Alto, na província do Kwanza-Norte.
86	Cambuanda*	Kambuanda	Umbundu	2	<i>n. f.</i> Região da província do Bié.
87	Camoco*	Kamóco	Umbundu	2	<i>n. m.</i> Vila localizada na província do Bié.
88	Camongua*	Kamôngua	Umbundu	2	<i>n. m.</i> 1. Vila localizada no município de Bula Atumba, na província do Kwanza-Norte. 2. Rio de Angola.
89	Camuanga*	Kamuanga	Kikongo	2	<i>n. f.</i> Vila do município do Alto Zaza, na província do Uíje.
90	Camuaxi*	Kamuaxi	Kimbundu	2	<i>n. m.</i> Fazenda localizada na província do Kwanza-Norte.
91	Camuine*	Kamuine	Umbundu	2	<i>n. f.</i> Vila do município do Chongoroi, na província de Benguela.
92	Camuiza*	Kamuiza	Kimbundu	2	<i>n. f.</i> Vila do município do Kazengo, na província do Kwanza-Norte.

93	Camulemba*	Camulemba	Cokwe	2	<i>n. f.</i> Município da província da Lunda-Norte.
94	Camuquembi*	Camuquembi*	Umbundu	1	<i>n. m.</i> Município da província da Huíla.
95	Camutue*	Kamutue	Kimbundu	2	<i>n. m.</i> 1. Mina localizada no município do Lucapa, na província da Lunda-Norte. 2. Cabeça pequena; cabeça de criança.
96	Candundo*	Candundo	Cokwe	2	<i>n. m.</i> Vila localizada no município de Alto Zambeze, na província do Moxico.
97	CANFEU*	CANFEU	Português em Angola	2	<i>n. m.</i> Sigla de Campo. Nacional de Férias de Estudantes Universitários.
98	Cangandala*	Kangandala	Kimbundu	2	<i>n. f.</i> município da província de Malanje.
99	Canganjo*	Kanganjo*	Umbundu	2	<i>n. m.</i> Vila do município de Camacupa, na província do Bié.
100	Cangombe*	Kangombe	Cokwe	2	<i>n. m.</i> Vila do município de Luchaze, na província do Moxico.
101	Canguenda*	Kanguenda	Umbundu	2	<i>n. f.</i> Vila do município de Waku Kungu, na província do Kwanza-Sul.
102	Canguia*	Kanguia	Kimbundu	2	<i>n. f.</i> Comuna da província de Malanje.
103	Cangumbe*	Cangumbe	Cokwe	2	<i>n. f.</i> Comuna da província de Moxico.
104	Cangungo*	Kangungu	Kimbundu	2	1. <i>n. f.</i> Vila do município de Kazengo, na província do Kwanza-Norte. 2. <i>n. m. e f.</i> Natural ou residente de Cangungo.
105	Canhala*	Canhala	Umbundu	2	<i>n. m.</i> Estádio localizado no município da Caála, na província do Huambo.
106	Canhanga*	Canhanga	Umbundu	2	<i>n. f.</i> Vila do município de Cassongue, na província do Kwanza-Sul.
107	Canhime*	Canhime	Umbundu	2	<i>n. f.</i> Comuna da província de Benguela.
108	Canhoca*	Canhoca	Kimbundu	2	<i>n. f.</i> Vila do município de Kazengo, na província do Kwanza-Norte.
109	Caninguili*	Caninguili	Umbundu	1	<i>n. m.</i> Vila do município de Mungo, província do Huambo.
110	Canjanja*	Canjanja	Cokwe	2	<i>n. f.</i> Vila localizada na província do Moxico.
111	CAP*	CAP	Português em Angola	3	<i>n. m.</i> Sigla de Comité de Ação do Partido.
112	Capamba*	Capamba	Cokwe	2	<i>n. f.</i> Vila do município de Dala, na província da Lunda-Sul.
113	Capanda*	Kapanda	Kimbundu	2	<i>n. f.</i> 1. Vila do município de Cacuso, na província da Malanje. 2. Central hidroelétrica localizada entre a província de Malanje e Kwanza-Sul.
114	Capande*	Kapanda	Kimbundu	2	<i>n. f.</i> O mesmo que Capanda.
115	Capari*	Kapari	Kimbundu	2	<i>n. f.</i> 1. Vila do município do Bengo, na província da Luanda. 2. Centralidade localizada no Bengo.
116	Capenda*	Capenda	Cokwe	2	<i>n. f.</i> Município da Lunda-Norte.
117	Capitango*	Kapitango	Umbundu	2	<i>n. m.</i> 1. Vila do município de Tchindjenje, província do Huambo. 2. Capitão. 3. Oficial do exército ou da justiça.
118	Capolo*	Kapolo	Kimbundu	2	<i>n. m.</i> 1. Distrito urbano do município do Kilamba Kixi, na província de Luanda. 2. Carinha; faceta.
119	Capunda*	Kapunda	Kimbundu	2	<i>n. f.</i> Vila do município do Luquembo, na província da Malanje.

120	Caputo	Okaputu	Kimbundu	2	<i>n. m.</i> 1. Governo. 2. O mesmo que <i>Muene-Puto</i> . 3. Autoridade portuguesa.
121	Caquarta*	Kakwarta	Kimbundu	2	<i>n. m. e f.</i> Indivíduo nascido numa quarta-feira.
122	Caquepa*	Kakepa	Umbundu	2	<i>n. f.</i> Município da província de Benguela.
123	Caquinta*	Caquinta	Kimbundu	2	<i>n. m. e f.</i> Indivíduo nascido numa quinta-feira.
124	CASA-CE*	CASA-CE	Português em Angola	2	1. n. f. Sigla de Convergência Ampla para a Salvação de Angola – Coligação Eleitoral. 2. n. m. e f. Indivíduo que pertence a esse partido político. 3. n. m. e f. Relativo à população que pertence a esse partido político.
125	Cassabalo	Cassabalo	Kimbundu	2	<i>n. m. e f.</i> Indivíduo nascido num sábado.
126	Cassaca*	Casseça	Kimbundu	2	<i>n. f.</i> 1. Vila do município do Cazengo, província do Kwanza-Norte. 2. Rio de Angola.
127	Cassamano	Kasamanu	Kimbundu	2	<i>adj. num. ord.</i> Sexto; Que está em sexto lugar; a sexta parte.
128	Cassambi	Cassambi	Cokwe	2	<i>n. f.</i> Vila do município do Luchaze, província do Moxico.
129	Cassâmbua*	Kasâmbua	Kimbundu	2	<i>n.</i> 1. Samba pequena. 2. adj. Da savana.
130	Cassanga*	Kasanga	Kimbundu	2	1. n. f. Região do município do Bengo, na província de Luanda. 2. n. m. Planta amarantácea conhecida por agimbôa brava.
131	Cassanje	Kisanji	Kimbundu	2	<i>n. m.</i> 1. Região em Malanje entre o Bondo, Songo e o rio Kuangu ou Zaire. 2. Reino de Angola. 3. Frango. Galinha pequena.
132	Cassenda	Cassenda	Português em Angola	2	1. n. m. Distrito urbano da província de Luanda. 2. n. m. e f. Relativo à população que pertence a esse distrito urbano.
133	Casseque*	Kaseke	Umbundu	2	1. n. m. Comuna do município de Ganda, na província de Benguela. 2. n. Planta da família das leguminosas de utilidade medicinal e madeira própria para construções.
134	Cassequel	Cassequel	Português em Angola	2	<i>n. m.</i> Bairro do município da Maianga, na província de Luanda.
135	Cassessa*	Kasesa	Kimbundu	2	<i>n. f.</i> 1. Vila do município do Cazengo, província do Kwanza-Norte. 2. Rio de Angola.
136	Cassoalala*	Kasualala	Kimbundu	2	<i>n. f.</i> Vila do município do Cambambe, província do Kwanza-Norte.
137	Cassoma*	Kasoma	Umbundu	2	1. n. f. Vila localizada no município do Caála, província do Huambo. 2. n. Substituto. 3. Orador sagrado. 4. n. Confluente da margem esquerda do rio <i>Kuiji</i> , no antigo concelho e hoje posto de Pungo Andongo.
138	Cassombe*	Kasombe	Cokwe	2	<i>n. f.</i> Vila localizada no município do Chitato, província da Lunda-Norte.
139	Cassongo*	Okasongu	Umbundu	2	<i>n. f.</i> Vila localizada na província do Huambo.
140	Cassua	kusuama	Kimbundu	2	1. n. m. e f. Nome dado a tal indivíduo. 2. adj. unif. Raquítico. 3. n. m. e f. O que é enfezado.
141	Cassueca	Kasueca	Umbundu	2	<i>n. f.</i> Região da província do Kwanza-Sul.

142	Cassumba*	Kasumba	Umbundu	2	<i>n. f.</i> Município da província do Bié.
143	Catala*	Etala	Kwanyama	2	1. n. m. Vila da comuna de <i>Kisama</i> , município do Bengo, província de Luanda. 2. n. f. Braço de rio que se estende pela terra dentro. 3. Vala ou regueiro que comunica com um rio, enchendo-se na época das chuvas. 4. n. Antiga população do concelho de Muxima, província de Luanda, margem direita do rio Longa.
144	Catchiungo*	Katchiungo	Umbundu	2	<i>n. m.</i> Município localizado na província do Huambo.
145	Catembo*	Katembo	Kimbundu	2	<i>n. m.</i> Município localizado na província de Malanje.
146	Catenda*	Katenda	Kimbundu	2	<i>n. m.</i> Região da província de Malanje.
147	Catepa*	Katepa	Kimbundu	2	<i>n. f.</i> Município localizado na província de Malanje
148	Catetão*	Catetão	Português em Angola	2	<i>n. m.</i> Campo do Futebol Clube do Petro de Luanda.
149	Catiavala*	Katiavala	Umbundu	2	<i>n. m.</i> Nome dos Somas do Reino do Bailundo
150	Catinton*	Catinton	Português de Angola	2	<i>n. m.</i> Mercado situado no bairro da Gamek, província de Luanda.
151	Catiolo*	Katiolo	Umbundu	2	<i>n. m.</i> Rio de Angola localizado na província do Kwanza-Sul. 2. Tipo de feijão branco miúdo, conhecido também por <i>Kaluwawala</i> .
152	Catoca*	Katoca	Cokwe	2	1. n. f. Sociedade mineira que opera na província da Lunda-Sul. 2. Mina de diamante em operação desde 1996, localizada no município de Saurimo, província da Lunda-Norte.
153	Catombela*	Katombelwa	Umbundu	2	1. n. f. Município da província de Benguela, localizado na foz do rio com o mesmo nome. 2. n. m. e f. Indivíduo natural ou habitante da Vila de Catumbela. 3. n. m. Adágio em <i>umbundu</i> “ <i>Ulume Katombelwa kotchahé</i> ” (o homem não pode ou não deve ser subjugado naquilo que é seu). / Nome do primeiro soberano da região da Catumbela.
154	Catumbo*	Katumbo	Kimbundu	2	1. n. m. Vila do município do Dande, Bengo, província de Luanda. 2. Tranças que já estão a muito tempo e com mal cheiro.
155	Cauaia*	Kawaia	Umbundu	2	<i>n. m.</i> Município da província do Bié.
156	Caúmbi*	Kaúmbi	Umbundu	2	<i>n. m.</i> Vila da província do Kwanza-Sul.
157	Caungo*	Kaungu	Kimbundu	2	<i>n. m.</i> Vila da província de Malanje.
158	Caungula*	Kaungula	Cokwe	2	<i>n. m.</i> Município da província da Lunda-Norte.
159	Caviombo	Kaviombo	Umbundu	2	1. n. m. Vila do município do Lubango, província da Huíla. 2. n. m. e f. Indivíduo nascido nessa vila.
160	Cazanga	Kanzenga	Kimbundu	2	1. n. m. Município da província de Luanda. 2. n. m. e f. Indivíduo que pertence ao município do Cazanga, na província de Luanda. 3. adj. unif. Relativo à população do município do Cazanga.

161	Cazengo	Kanzenga	Kimbundu	2	1. <i>n. m.</i> Região agrícola da província do Kwanza-Norte, de cuja produção sobressai o café. 2. Famosa região agrícola da província do Kwanza-Norte, de cuja produção sobressai o café.
162	CCDA*	CCDA	Português em Angola	2	<i>n. f.</i> Sigla da Companhia de Dança Contemporânea de Angola.
163	Chambungo*	Chambungo	Umbundu	2	<i>n. m.</i> Município da província de Benguela.
164	Chamuanga*	Chamuanga	Umbundu	2	<i>n. m.</i> Município da província do Huambo.
165	Chamume*	Chamume	Umbundu	2	<i>n. f.</i> Salina localizada no município da Baía Farta, província de Benguela.
166	Chianga*	Tchianga	Umbundu	2	<i>n. f.</i> Região da província do Huambo.
167	Chibia*	Xibia	Umbundu	2	1. <i>n. f.</i> Município da província da Huíla. 2. <i>n. m. e f.</i> Indivíduo natural ou habitante dessa região.
168	Chicala	Chicala	Português em Angola	2	<i>n. f.</i> Bairro costeiro localizado no centro da cidade de Luanda.
169	Chicapa	Chicapa	Cokwe	2	<i>n. m.</i> Vila do município de Saurimo, província da Lunda-Sul.
170	Chicomba*	Chicomba	Umbundu	2	<i>n. f.</i> Município da província da Huíla.
171	Chiculo*	Chiculo	Umbundu	2	<i>n. m.</i> Região da província de Benguela.
172	Chicunga*	Chicunga	Umbundu	2	<i>n. f.</i> Vila do município do Tchindjenje, província do Huambo.
173	Chiengo*	Chiengo	Umbundu	2	<i>n. m.</i> Região da província do Bié.
174	Chihumbue	Chihumbue	Cokwe	2	<i>n. m.</i> Hidroelétrica da comuna de Dala, província da Lunda-Sul.
175	Chikoti	Kuxokotiala	Kimbundu	2	<i>n. m.</i> 1. Importunar com perguntas. 2. Fazer arrelhar com impertinências.
176	Chilala	Chilala	Umbundu	2	<i>n. f.</i> Região da província do Huambo.
177	Chilembo	Chitembo	Umbundu	2	<i>n. m.</i> Município da província do Bié.
178	Chiloango	Chiloango	Kikongo	2	<i>n. m.</i> Rio principal da província de Cabinda.
179	Chiluage	Chiluage	Cokwe	2	<i>n. f.</i> Vila do município de Muconda, província da Lunda-Sul.
180	Chilungo	Chilungo	Umbundu	2	<i>n. m.</i> Vila do município do Lobito, província de Benguela.
181	Chimboco*	Tchimboco	Umbundu	2	<i>n. m.</i> Vila do município do Sumbe, província do Kwanza-Sul.
182	Chimbundo*	Chimbundo	Cokwe	2	<i>n. f.</i> Região da província da Lunda-Sul.
183	Chimbungo	Chimbungo	Umbundu	2	<i>n. m.</i> Vila do município de Cassongue, província do Kwanza-Sul.
184	Chimuanga	Tchimuanga	Umbundu	2	<i>n. m.</i> Região da província do Bié.
185	Chinangol*	Chinangol	Português em Angola	2	<i>n. f.</i> Empresa de construção civil.
186	Chindembe*	Chindembe	Umbundu	2	<i>n. m.</i> Vila localizada no município de Seles, província do Kwanza-Sul.
187	Chinduva*	Chinduva	Umbundu	2	<i>n. m.</i> Região da província do Benguela.
188	Chingango*	Chingango	Umbundu	2	<i>n. m.</i> Empresa angolana.
189	Chinganguela*	Ngangela	Nganguela	2	<i>n. m. e f.</i> 1. Membro do grupo étnico dos ganguelas, disseminados nas bacias do Cuando e Cubango. 2. Indivíduo desse povo.

					3. <i>adj. m. e f.</i> Que é relativo ao ganguela, língua falada em Angola. 4. <i>n. m.</i> Língua falada na zona leste de Angola e oeste da Zâmbia.
190	Chingila*	Chingila	Cokwe	2	<i>n. f.</i> Município da província do Moxico.
191	Chingongo*	Chingongo	Umbundu	2	<i>n. m.</i> Vila localizada no município de Balombo, província de Benguela.
192	Chinguar*	Chinguar	Umbundu	2	1. <i>n. m.</i> Município da província do Bié. 2. <i>n. m. e f.</i> Indivíduo natural ou residente dessa região.
193	Chinguari*	Chinguar	Umbundu	2	<i>n. m.</i> O mesmo que Chinguar.
194	Chingufo	Oxingufo	Umbundu	2	<i>n. m.</i> 1. Tambor. 2. Município da província da Lunda-Norte.
195	Chinhama*	Xinhama	Umbundu	2	<i>n. m.</i> Vila localizada no município de Cachiungo, província de Huambo.
196	Chinhundo	Tchinhundo	Cokwe	2	<i>n. m.</i> Região da província do Moxico.
197	Chinongue	Chinongue	Cokwe	2	<i>n. m.</i> Região da província da Lunda-Sul.
198	Chipenda	Xipenda	Umbundu	2	<i>n. m.</i> Um dos principais intervenientes da Guerra de Independência de Angola.
199	Chipende*	Xipende	Umbundu	2	<i>n. m.</i> Vila localizada no município de Cassongue, província de Kwanza-Sul.
200	Chipeta*	Tchipeta	Umbundu	2	<i>n. f.</i> Comuna do município de Catabola, província do Bié.
201	Chipoia*	Xipoia	Cokwe	2	<i>n. f.</i> Região da província do Moxico.
202	Chissamba	Xissamba	Umbundu	2	<i>n. f.</i> Região da província do Bié.
203	Chissambo*	Xissambo	Kikongo	2	<i>n. m.</i> Lago da província de Cabinda.
204	Chissanda*	Xissanda	Cokwe	2	<i>n. m.</i> Região da província da Lunda-Norte.
205	Chissengue*	Xisenge	Umbundu	2	<i>n. m.</i> Vila localizada no município de Tchicala-Tcholo, província do Huambo.
206	Chissola*	Tchisola	Umbundu	2	<i>n. f.</i> Amor; caridade; bondade.
207	Chissuata*	Xisuata	Umbundu	2	<i>n. f.</i> Vila localizada no município de Chitembo, província do Bié.
208	Chitato*	Xitato	Cokwe	2	<i>n. m.</i> Município da província da Lunda-Norte.
209	Chiteculo*	Tchiteculo	Umbundu	2	<i>n. m.</i> 1. Vila localizada no município de Cassongue, província de Kwanza-Sul. 2. Benfeitor; pessoa bondosa.
210	Chitembo	Xitembo	Umbundu	2	<i>n. m.</i> Município da província do Bié.
211	Chitotolo*	Xitotolo	Umbundu	2	<i>n. m.</i> Sociedade Mineira de Angola.
212	Chitula*	Xitula*	Kimbundu	2	<i>n. f.</i> Região da província de Malanje.
213	Chitundo*	Xitundo	Umbundu	2	<i>n. m.</i> Morro granítico situado na província do Namibe.
214	Chiuca	Chiuca	Umbundu	2	<i>n. f.</i> Vila localizada no município de Catabola, província do Bié.
215	Chiula	Chiula	Umbundu	2	<i>n. m.</i> Região da província da Kwanza-Sul.
216	Chiulo	Chiulo	Umbundu	2	<i>n. m.</i> Vila localizada no município de Ombadja, na província do Cunene.
217	Chivaulo*	Chivaulo	Umbundu	2	<i>n. m.</i> Vila localizada no município de Andulo, na província do Bié.

218	Chivela*	Tchivela	Umbundu	2	<i>n. m.</i> 1. Doente. 2. Ferro.
219	Chivembe*	Oxivembe	Umbundu	2	<i>n. m.</i> Vila localizada no município de Chitembo, província do Bié.
220	Chivinda*	Tchivinda	Umbundu	2	1. <i>n. m.</i> Região da província do Bié. 2. <i>n. m. e f.</i> Indivíduo dessa região. 3. <i>n. m. e f.</i> Ferreiro, artífice.
221	Chivukuvuku*	Chivukuvuku	Umbundu	2	<i>n. m.</i> Líder do grupo parlamentar CASA-CE.
222	Chivunda*	Chivunda	Cokwe	2	<i>n. f.</i> Vila localizada no município do Leua, na província do Moxico.
223	Cholohanga*	Cholohanga	Umbundu	2	<i>n. m.</i> Município da província do Huambo.
224	Chongoroi*	Xongoroi	Umbundu	2	1. <i>n. m.</i> Município da província de Benguela. 2. <i>n. m. e f.</i> Indivíduo natural ou residente desse município. 3. <i>adj. unif.</i> Relativo à população dessa região.
225	CIF*	CIF	Português em Angola	2	<i>n. m.</i> Sigla de Fundo Internacional da China.
226	Codiango	Codiango	Português em Angola	2	<i>n. m.</i> Instituição formada para criar código de barra de Angola.
227	Cokwe	Cokwe*	Cokwe*	2	1. <i>n. m.</i> Grupo etnolinguístico de Angola.* 2. <i>n. m. e f.</i> Natural da área da Lunda a nordeste de Angola. 3. <i>n. f.</i> Língua falada pelo povo Cokwe.* 4. <i>adj. unif.</i> Relativo a esse povo.* 5. <i>n. m. e f.</i> Natural da área da Lunda. A nordeste de Angola. 6. O mesmo que Quioco.
228	Combo*	Combo	Português em Angola	2	<i>n. m.</i> Empresa angolana especializada em venda de cabos elétricos.
229	Corimba*	Corimba	Português em Angola	2	1. <i>n. f.</i> Bairro dos distrito da Samba, província de Luanda. 2. <i>n. m. e f.</i> Natural ou residente desse bairro. 3. <i>adj. unif.</i> Relativo à população desse bairro.
230	Cuacra	Cuacra	Umbundu	2	1. <i>n. f.</i> Região da província do Kwanza-Sul. 2. <i>adj. unif.</i> Relativo à população dessa região.
231	Cuamato	Kuamátui	Termo regional	1	<i>n. m. e f. e adj. unif.</i> 1. Forma deturpada de cuamatui, embora mais usual. 2. <i>adj. unif.</i> Que diz respeito ou pertence aos Cuamatos, povo que habita o sul de Angola, no baixo Cunene. 3. <i>n. m e f. e adj. unif.</i> 1. Indivíduo natural de Cuamátui a sudeste de Angola. 4. Dialeto falado nesta região, pertencente ao grupo étnico dos Ambós. 5. <i>n. m. e f.</i> Forma deturpada de cuamatui, embora mais usual.
232	Cuanavale	Cuanavale	Nganguela	2	<i>n. m.</i> Município da província do Kwando Kubango.
233	Cuando-Cubango*	Kwando Kubango	Nganguela	2	<i>n. m.</i> Província de Angola na fronteira com a Zâmbia e a Namíbia.
234	Cuando*	Kwando	Umbundu	2	<i>n. m.</i> 1. Rio de Angola que nasce no planalto central. 2. Região ao lado deste rio. 3. <i>adj. unif.</i> Relativo à população dessa região.
235	Cuangar*	Kwanger	Nganguela	2	<i>n. m.</i> Município da província do Kwando Kubango.
236	Cuango*	Cuango	Cokwe	2	<i>n. m.</i> Município da província da Lunda Norte.
237	Cuanza	Kwanza	Umbundu	2	<i>n. m.</i> Rio de Angola, cuja foz é 12 km ao noroeste da povoação de Katota, do posto

				mesmo nome circunscção do Alto Kwanza – Chitembo, na província do Bié.	
238	Cuatir*	Kwatiri	Nganguela	2	<i>n. m.</i> Região da província da Kwando Kubango.
239	Cubal	Kukuba	Umbundu	3	<i>n. m. e f. e adj. unif.</i> 1. Forma diversificada de Cuvale. 2. Indivíduo natural deste município da província de Benguela. 3. <i>m. m.</i> Dialeto falado nessa região.
240	Cubango*	Kubango	Nganguela	2	<i>n. m.</i> 1. Rio de Angola que nasce no planalto central. 2. Região ao lado deste rio. 3. <i>n. m. e f.</i> Indivíduo natural ou residente do Kwando Kubango. 4. <i>adj. unif.</i> Relativo à população da região do Kwando Kubango.
241	Cuchi*	kuxi	Nganguela	2	<i>n. m.</i> Município da província do Kwando Kubango.
242	Cuemba*	Kwemba	Umbundu	2	<i>n. m.</i> Município da província do Bié.
243	Cuimba*	Kuimba	Kikongo	2	<i>n. m.</i> Município da província do Zaire.
245	Cuito*	Kuito	Umbundu	2	1. <i>n. m.</i> Município da província do Bié. 2. <i>n. m. e f.</i> Indivíduo que é natural ou residente desse município. 3. <i>adj. unif.</i> Relativo à população da região pertencente ao grupo Umbundu. 4. <i>m. n.</i> Amarrados fortemente.
246	Cuíto*	Kuito	Umbundu	2	<i>n. m.</i> O mesmo que Cuito.
247	Culembe*	Kulembe	Umbundu	2	<i>n. m.</i> Região da província do Kwanza-Sul.
248	Culimahãla*	Kulimahãla	Umbundu	2	<i>n. m.</i> Rio de Angola que está na província do Huambo.
249	Culungo*	Kulungo	Kikongo	2	<i>n. m.</i> 1. Região da província do Zaire. 2. Rio de Angola.
250	Cumbi*	Kumbi	Kikongo	2	<i>n. m.</i> Vila do município de Mbanza Congo.
251	Cundueji*	Cundueji	Cokwe	2	<i>n. m.</i> Rio de Angola situado no município do Chitato, província da Lunda-Norte.
252	Cunene*	Kunene	Kwanyama	2	1. <i>n. m.</i> Província da República de Angola. 2. <i>n. m. e f.</i> Indivíduo que é natural ou residentes dessa província.
253	Cunhinga*	Cunhinga	Umbundu	2	<i>n. m.</i> Município da província do Bié.
254	Cunhoŋgamua*	Cunhoŋgamua	Umbundu	2	<i>n. m.</i> Rio de Angola, localizado na província do Huambo.
255	Cunje*	Kunji	Umbundu	2	<i>n. m.</i> Vila do município do Cuito, província do Bié.
256	Cuononoca*	Kuononoka	Umbundu	2	<i>n. m.</i> Antropólogo e político angolano.
257	Cuquema	Kukema	Umbundu	2	<i>n. f.</i> Cadeia localizada no município do Cuito, província do Bié.
258	Curoca	Kuroka	Kwanyama	2	<i>n. m.</i> 1. Município da província do Cunene. 2. Rio de Angola.
259	Cutembo*	Kutembo	Umbundu	2	<i>n. m.</i> 1. Vila do município do Quilengue, província da Huíla. 2. Rio de Angola.
260	Cuvango*	Kuvango	Umbundu	2	<i>n. m.</i> 1. Município da província do Huíla. 2. <i>n. m. e f.</i> Indivíduo natural ou residente desse município. 3. <i>adj. unif.</i> Relativo à essa população.
261	Cuvelai*	Kuvelai	Kwanyama	2	<i>n. m.</i> Município da província do Cunene.

262	Damba	Ndamba	Umbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural da Damba, na região do Uíje, ao norte de Angola. 2. <i>Pl.</i> População dessa área.
263	Dande	N'danji	Kikongo	3	1. <i>n. m. e f.</i> natural de Ndanje, na região do Púri, ao norte de Angola. 2. <i>adj.</i> Relativo à população dessa área.
264	Dangala*	Dangala	Umbundu	2	<i>n. m.</i> Vila da província do Kwanza-Sul.
265	Dange	Ndanji	Kikongo	2	<i>n. m. e f.</i> 1. Natural de N'Danje, na Região do Púri, ao norte de Angola. 2. <i>Pl.</i> População dessa área pertencente ao grupo étnico dos Kikongos.
266	Diamang*	Diamang	Português em Angola	2	<i>n. f.</i> Companhia de diamante de Angola.
267	Diassonama	Diassonama	Português em Angola	2	1. <i>n. m.</i> Grupo teatral angolano. 2. <i>n. m. e f.</i> Indivíduo pertencente a esse grupo teatral. 3. <i>adj. unif.</i> Relativo a esse grupo teatral.
268	Dimuca*	Dimuka	Kikongo	2	<i>n. m.</i> Vila da província do Uíge.
269	Do-itombe*	Zenza do Itombe	Kimbundu	2	<i>n. m.</i> Vila do município de Cambambe, província do Kwanza Norte.
270	Dombe	Ndombe	Umbundu	2	1. <i>n. m. e f.</i> Indivíduo natural do Dombe, na região de Benguela, a sul de Angola. 2. <i>n. m.</i> dialeto falado nessa região. 3. <i>Pl.</i> População dessa área pertencente ao grupo étnico Umbundu. 4. <i>adj. unif.</i> Relativo à essa população. 5. O mesmo que Mondombe e Mundombe. 6. <i>Pl. híbr.</i> Vandombes.
271	Dombele*	Ndembele	Umbundu	2	<i>n. f.</i> União.
272	Dombolo	Dombolo	Kikongo*	2	<i>n. m.</i> Roedor do tamanho do tamanho aproximado de um coelho, com patas posteriores muito muito desenvolvida, adaptadas para saltar frequente no centro e sul de África.
273	DRM*	DRM	Português em Angola	2	<i>n. m.</i> Sigla de Direção Regional Militar.
274	Dundo	Ndundu	Kimbundu	2	1. <i>n. m.</i> Município da província da Lunda Norte. 2. <i>adj. unif.</i> Relativo à população dessa região.
275	Dundu	Ndundu	Kimbundu	2	<i>n. m.</i> 1. Soco. 2. Murro. 3. <i>Pl. híbr.</i> Jundundos.
276	Ebanga	Ebanga	Umbundu	2	<i>n. m.</i> Vila do município de Ganda, província de Benguela.
277	Eka	Eka	Kimbundu	2	<i>n. f.</i> Cerveja de Angola
278	Ekuike	Ekwikwi	Umbundu	2	<i>n. m.</i> Título dado ao Soma Inene do Reino do Bailundo.
279	Elinga*	Elinga	Kimbundu	2	<i>n. m.</i> Grupo teatral angolano.
280	ENDE*	ENDE	Português em Angola	3	<i>n. f.</i> Sigla de Empresa Nacional de Distribuição de Energia Elétrica.
281	Endiama*	Endiama	Português em Angola	2	<i>n. f.</i> Sigla da Empresa Nacional de Diamantes de Angola.
282	Epalanca*	Epalanca	Umbundu	2	1. <i>n. m.</i> Rio de Angola. 2. <i>n. m. e f.</i> Assistente, auxiliar. 3. Irmão mais novo; ultimogénito.
283	Ex-satec *	Satec	Português em Angola	2	<i>n. f.</i> Fábrica têxtil angolana, localizada no município do Dondo, na província do Kwanza-Norte.

284	FAB*	FAB	Português em Angola	3	<i>n. f.</i> Sigla de Federação Angolana de Basquetebol.
285	Fefa*	Josefa	Português em Angola	1	<i>n. f.</i> Hipocorístico de Josefa.
286	FENACULT*	Fenacult	Português em Angola	2	<i>n. f.</i> Sigla de Festival Nacional de Cultura Angolana.
287	FNLA*	FNLA	Português em Angola	2	1. n. f. Sigla de Frente Nacional de Libertação de Angola. 2. n. m. Partido político angolano. 3. n. m. e f. Indivíduo pertencente a esse partido político.
288	Futla*	Futla	Kikongo	2	<i>n. 1.</i> Região da província de Cabinda. 2. Polo industrial situado na província de Cabinda.
289	Futungo*	Futungo	Português em Angola	2	<i>n. m.</i> Distrito urbano do município de Belas, província de Luanda.
290	Gabela*	Gabela	Umbundu	2	<i>n. f.</i> Vila do município de Amboim, província do Kwanza-Sul.
291	Gambos	Ngambue	Nyaneka	1	1. n. m. e f. Indivíduo natural dos Gambos, a sudeste de Angola. 2. n. m. dialeto falado nessa região. 3. Pl. População dessa área pertencente ao subgrupo dos nhanecas.
292	Girabola*	Girabola	Português em Angola	2	<i>n. m.</i> Campeonato angolano de futebol.
293	Golungo	Ngulungu e Ongulungu	Kimbundu	1	<i>n. m. 1.</i> Antílope de corpulência maior que a da corsa. 2. Veado. 3. n. m. Forma errónea de gulungo. 4. n. m. Antílope de corpulência maior que a da corsa; veado.
294	Grafanil*	Grafanil	Português em Angola	2	<i>n. m.</i> Distrito urbano do município de Viana, província de Luanda.
295	Guimbe	Gimbe	Kimbundu	2	<i>n. f.</i> Vila do município de Icolo e Bengo, província do Kwanza-Norte.
296	Gunza	Ngunza	Kimbundu	2	<i>n. m.</i> Diferente denominação da entidade espiritual Mutakalombo.
297	Guvulo	Guvulu	Cokwe	2	<i>n. m.</i> Município da província do Moxico.
298	Hebo	Hebo	Kimbundu	2	<i>n. m. e f. 1.</i> Indivíduo de longa gestação. 2. Nome dado a tal criança.
299	Henda*	Henda	Kimbundu	2	<i>n. 1.</i> pena solícita pela desgraça ou males alheios, que sentimos como nosso. 2. Misericórdia; devoção; piedade. 3. Piedade; compaixão. 4. n. f. Conversa, assunto, novidade, discórdia, questão, litígio, conflito.
300	Hequele*	Hequele	Umbundu	2	<i>n. m.</i> Vila da província do Huambo.
301	Herero	Helelu	Herero	1	1. n. m. e f. Membro do grupo étnico dos Hereros, disseminados a sudoeste de Angola. 2. n. m. Língua falada nessa região.
302	Hidrochicapa	Hidrochicapa	Português e Umbundu	2	<i>n. f.</i> Barragem sobre o rio Chicapa, no município de Saurimo, província da Lunda Sul.
303	Hidroluachimo	Hidroluachimo	Português e Umbundu	2	<i>n. f.</i> Barragem sobre o rio Luachimo, no município do Dundo, província da Lunda Norte.
304	Hoji-ya-henda	Hoji-ya-henda	Kimbundu	2	<i>n. m.</i> Distrito urbano do município do Cazenga, província de Luanda.
305	Huambo	Hambo	Umbundu	2	<i>n. m. 1.</i> Província de Angola. 2. Deturpação de Hambo nome do Soba que originou o topónimo. 3. n. m. e f. Indivíduo natural do Huambo, no planalto central de Angola. 4.

				<i>Pl.</i> População dessa área pertencente ao grupo étnico dos Umbundu. 5. <i>n. m.</i> Dialeto falado nessa região. 6. <i>adj. unif.</i> Relativo à essa população. 7. <i>Pl. híbr.</i> Vahuambos.	
306	Huíla	Híla	Nyaneka	1	1. <i>n. f.</i> Província de Angola. 2. <i>m. e f.</i> Indivíduo natural da Huíla, a sudoeste de Angola. 2. <i>n. m. e f.</i> O mesmo que <i>Muíla</i> e <i>Mamuíla</i> . 3. Língua falada em Angola, pertencente ao grupo nyaneka. 4. <i>adj.</i> Relativo a essa população.
307	Huilano	Híla	Umbundu	2	<i>n. m.</i> Natural ou habitante da Huíla, no planalto do sudeste de Angola.
308	Humabo*	Hambo	Umbundu	2	<i>n. m.</i> O mesmo que Huambo.
309	Humbe	Nkumbi	Nyaneka	2	1. <i>n. m. e f.</i> Membro do subgrupo étnico dos Humbes, disseminados a sudoeste de Angola. 2. <i>n. m.</i> Dialeto falado nessa região. 3. <i>Pl.</i> Povos bantos, compreendendo os Dongoenas ou Dongonas, Hingas, Cuâncuas, Handas Quipungos, Quilengues-Humbes e Quilengues-Muzós.
310	Humbi	Nkumbi	Nyaneka	2	<i>n. m. e f. e adj. unif.</i> O mesmo que Humbe, mas de forma mais aproximada ao vernáculo.
311	Humpata*	Humpata	Umbundu	1	<i>n. f.</i> Município da província da Huíla.
312	Iaca	Iaca	Termo regional	2	1. <i>n. m. e f.</i> Natural de Iaca, ao norte de Angola. 2. <i>n. m.</i> Dialeto falado nessa região. 3. <i>Pl.</i> População dessa área, pertencente ao grupo étnico dos Kikongos. 4. <i>adj. unif.</i> relativo a essa população. 5. O mesmo que <i>Muiaca</i> e <i>Quiaca</i> . 6. <i>Pl. híbr.</i> Baiacas, ou impropriamente, Maiacas.
313	Icolo*	Ikolo	Kimbundu	2	<i>n. m.</i> Nome dado à região ao Norte de Angola em homenagem ao poder dos sobas <i>Mani Mbengu</i> (que reinou a região até a conquista dos portugueses) e <i>Mani Ikolo</i> (sendo vizinho); esta região é banhada pelos rios <i>Kwanza</i> , a Sul, e <i>Nzenza</i> (nome vernáculo do rio) ou <i>Mbengu</i> (nome dado ao mesmo rio mais tarde em homenagem ao soba <i>Mani Mbengu</i>), ao Norte.
314	IGAE	IGAE	Português em Angola	2	<i>n. m.</i> Sigla de Inspeção Geral da Administração do Estado.
315	IGAPE*	IGAPE	Português em Angola	2	<i>n. m.</i> Sigla de Instituto de Gestão de Ativos e Participações do Estado.
316	IGCA	IGCA	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Geográfico e Cadastral de Angola.
317	INAAREES*	INAAREES	Português em Angola	2	1. <i>n. m.</i> Sigla do Instituto Nacional de Avaliação Acreditação e Reconhecimento de Estudos do Ensino Superior. 2. <i>adj.</i> Relativo a esta instituição.
318	INACOM*	INACOM	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Angolano das Comunicações.
319	INADEC*	INADEC	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Nacional de Defesa do Consumidor.
320	INAMET*	INAMET	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Nacional de Meteorologia e Geofísica.

321	INAVIC*	INAVIC	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Nacional de Aviação Civil.
322	INEA*	INEA	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Nacional de Estradas de Angola.
323	INEJ*	INEJ	Português em Angola	2	<i>n. m.</i> Sigla de Instituto Nacional de Estudos Judiciários.
324	Ingombota	Ngombo + kuta	Kimbundu	2	<i>n. f.</i> 2. Antigo bairro africano da cidade de Luanda, localizado na encosta fronteira à Igreja do Carmo. 2. Lugar da cidade de Luanda onde se aceitavam escravos foragidos, localizado na encosta fronteira à Igreja do Carmo. 3. adj. unif. Relativo a este bairro.
325	Inhuca*	Inhuca	Kikongo	2	<i>n. f.</i> Vila da província de Cabinda.
326	Iona*	Iona	Umbundu	2	<i>n. m.</i> Parque Nacional localizado na província do Namibe.
327	Issenguele*	Isenga	Kimbundu	2	<i>n.</i> Farelo; parte grossa que fica na peneira depois de peneirada a farinha; milho.
328	Itumbo*	Itumbo	Umbundu	2	<i>n. m.</i> Vila da província do município do Sumbe, província do Namibe.
329	JMPLA*	JMPLA	Português em Angola	2	1. n. m. e f. Indivíduo pertencente a esta facção do partido político. 2. n. m. Sigla de Juventude do Movimento Popular de Libertação de Angola.
330	Kabocomeu	Kabocomeu	Kimbundu	2	<i>n. m.</i> Grupo carnavalesco angolano.
331	Kachiungo	Kachiungo	Umbundu	2	<i>n. m.</i> Município da província do Huambo.
332	Kafala	Kafala	Kimbundu	2	<i>n. m.</i> Grupo musical angolano.
333	Kahanganhi	Kahanganhi	Cokwe	2	<i>n. m.</i> Bairro da província do Moxico.
334	Kahombo*	Kahombo	Kimbundu	2	<i>n. m.</i> 1. Município da província de Malanje. 2. Tipo de jindungo.
335	Kalawenda*	Kalawenda	Kimbundu	2	<i>n. m.</i> Distrito urbano do município do Cazenga, província de Malanje.
336	Kalembe	Kalembe	Kimbundu	2	<i>n. f.</i> 1. Ondulação fortemente agitada, quer do mar quer do rio. 2. Marulhada. 3. O mesmo que calema, em deturpação portuguesa.
337	Kalengue*	Kalenge	Kimbundu	2	1. n. m. Vila do município de Icolo e Bengo, província de Luanda. 2. n. m. e f. Indivíduo natural ou habitante desta região.
338	Kaluanda	Kaluanda	Kimbundu	2	<i>n. m. e f.</i> Indivíduo natural ou habitante de Luanda.
339	Kalucango*	Kalucango	Kimbundu	2	<i>n. m.</i> Parque localizado em Luanda.
340	Kalueio*	Kalueio	Umbundu	2	<i>n. m.</i> Bairro do município do Bailundo, província do Huambo.
341	Kalumbo*	Kalumbo	Kimbundu	2	<i>n. m.</i> 1. Bairro do município de Viana, província de Luanda. 2. Centro de saúde localizado neste bairro.
342	Kalunga-Mata*	Kalunga-Mata	Kimbundu e Português	2	<i>n. m.</i> Grupo de cantores angolanos.
343	Kaluvundo*	Kaluvundo	Umbundu	2	<i>n. m.</i> Vila do município de Kamucuiu, província do Namibe.
344	Kama*	Kalyatu	Cokwe	2	<i>n. f.</i> Nome da rainha Lunda Cokwe, Ngulia Kama, cujas exéquias foi a 22 de julho de

				2017, em Cafunfo, município do Cuango, província da Lunda-Norte.	
345	Kambinda*	Kambinda	Kikongo	2	<i>n. m. e f.</i> Indivíduo natural de Cabinda, província ao norte de Angola, na margem direita do rio Zaire.
346	Kambolo	Kambolo	Kimbundu	2	<i>n. f.</i> Nome da antiga rainha da Matamba, Mulundo Kambolo.
347	Kanda	Kanda	Kikongo	2	<i>n. f.</i> 1. Clã. 2. Instituição de carácter familiar e social. 3. Tribo formada por várias famílias subordinadas a um chefe. 4. O mesmo que eanda entre os Nyanekas, e epata entre os kwanyamas.
348	Kanguimbo*	Kanguimbo	Umbundu	2	<i>n. f.</i> Escritora angolana e membro da União dos Escritores Angolanos.
349	Kaombo*	Kahombo	Kimbundu	2	<i>n. m.</i> 1. Município da província de Malanje. 2. Tipo de jindungo.
350	Kapalanga*	Kapalanga	Kimbundu	2	<i>n. m.</i> Bairro do município de Viana, província de Luanda.
351	Kapanda*	Kapanda	Kimbundu	2	<i>n. f.</i> Barragem hidroelétrica construída sobre o rio Kwanza, no município de Cacuso, província de Malanje.
352	Kapango*	Kapango	Umbundu	2	<i>n. m.</i> Aeroporto do Kuito, província do Bié.
353	Kapolo*	Kapolo	Umbundu	2	<i>n. m.</i> 1. Distrito urbano do município do Kilamba Kiaxi, província de Luanda. 2. Vila do município do Porto Amboim, província do Kwanza-Sul.
354	Kapossoka*	Kaposoka	Kimbundu	2	<i>n. m.</i> Terminal marítimo de passageiros de Luanda, que dá acesso ao Mussulo.
355	Kasai*	Kasai	Kimbundu	2	<i>n. m.</i> Rio de Angola.
356	Kassange*	Kasanji	Kimbundu	2	<i>n. m.</i> Região em Malanje entre o Bondo, Songo e o rio <i>Kuangu</i> ou <i>Zaire</i> .
357	Kassanje*	Kasanji	Kimbundu	2	<i>n. m.</i> O mesmo que Kassange.
358	Katepa*	Katepa	Kimbundu	2	<i>n. f.</i> Município da província de Malanje.
359	Katiavala*	Katiavala	Umbundu	2	<i>n. m.</i> Nome dos Somas do Reino do Bailundo
360	Katinton*	Katinton	Português em Angola	2	<i>n. m.</i> Mercado situado no bairro da Gamek, província de Luanda.
361	Katyavala	Katiavala	Umbundu	2	<i>n. m.</i> Nome dos Somas do Reino do Bailundo
362	Kawazanga*	Kawazanga	Cokwe	2	<i>n. m.</i> Bairro do município do Saurimo, província da Lunda-Sul.
363	Keno*	Kueno	Kikongo	2	<i>n. m.</i> Músico e compositor angolano.
364	Kessongo	Okasongu	Umbundu	2	<i>n. f.</i> Vila localizada na província do Huambo.
365	Kessongo	Okasongu	Umbundu	2	<i>n. m. e f.</i> Porta-voz.
366	Keve*	Keve	Umbundu	2	<i>n. m.</i> 1. Rio de Angola localizado na província do Kwanza-Sul. 2. Banco angolano.
367	Khoisans	Khoisan	Khoisan	2	<i>n. m. e f.</i> 1. Membro do grupo étnico dos Khoisan, disseminados a sul de Angola. 2. <i>n. m.</i> Língua falada nessa região. 3. <i>Pl.</i> Povos não-bantu e não de raça negra, demograficamente constituindo o 9º agrupamento angolano compreendendo duas tribos principais: Boximanes e Cazamas.

368	Kiaku*	Kiáku	Kikongo	2	1. <i>Pron. possess.</i> Teu. 2. <i>n. m.</i> Lenço. 3. Tecido próprio para a pessoa assoar.
369	Kiaxi	Kiaxi	Kimbundu	2	<i>n. m.</i> Município de Luanda, conhecido como Kilamba Kiaxi.
370	Kibala	Kubala muiji	Kimbundu	3	<i>n. m. e f.</i> 1. Espírito gerador de cada família. 2. Entidade espiritual da génese familiar.
371	Kibanda	Otyibanda	Umbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de Quibanda a sul de Angola, a sul do curso superior do Cuvo. 2. População dessa área pertencente ao grupo étnico dos umbundos. 3. <i>adj. unif.</i> Relativo a essa população.
372	Kicolo	Kúkola	Kimbundu	2	<i>n. m.</i> 1. Bairro do município de Cacuaco, província de Luanda. 2. Acampamento de circuncisão. 3. Aposento onde permanece um paciente dessa operação.
373	Kicombo*	Kikombo	Umbundu	2	<i>n. m.</i> Vila do município do Sumbe, província do Kwanza-Sul.
374	Kicuxi*	Kikuxi	Kimbundu	2	1. <i>n. m. e f.</i> Indivíduo natural ou habitante dessa vila. 2. <i>n. m.</i> Vila do município de Viana, província de Luanda.
375	Kienda*	Kienda	Kimbundu	2	<i>n. m.</i> Denominação de uma avenida de Luanda em alusão ao Comandante Kima Kienda.
376	Kiesse*	Kiesse	Kimbundu	2	<i>n. m.</i> Denominação de um centro infantil de Luanda.
377	Kiezo	Kiezo	Kimbundu	2	1. <i>n. m.</i> Grupo de música folclórica angolana. 2. <i>adj. unif.</i> Relativo a esse grupo musical.
378	Kifangondo*	Kifangondo	Kimbundu	2	<i>n. m.</i> Vila do município de Cacuaco, província de Luanda.
379	Kifica*	Kifica	Kimbundu	2	<i>n. m.</i> Vila do bairro Benfica, município de Belas, província de Luanda.
380	Kikagil	Kikagil	Kimbundu e Português	2	<i>n. m.</i> Bairro do município de Belas, província de Luanda.
381	Kikolo	Kúkola	Kimbundu	2	<i>n. m.</i> 1. Desfiladeiro, vale. 2. Lugar ou casa em que se pratica a circuncisão. 3. Sabugo de milho nos Dembos. 4. Bairro do município de Cacuaco, província de Luanda. 5. Acampamento de circuncisão. 6. Aposento onde permanece um paciente dessa operação.
382	Kikuxi*	Kikuxi	Kimbundu	2	<i>n. m.</i> Vila do município de Viana, província de Luanda.
383	Kilamba	Kulambula	Kimbundu	2	1. <i>n. m.</i> Centralidade do município de Belas, província de Luanda. 2. <i>adj. unif.</i> Relativo à essa centralidade. 3. Relativo ao culto de tais entidades. 4. <i>n. m. e f.</i> Intérprete das sereias. Sacerdote do culto de tais entidades.
384	Kilamba Kiaxi	Kilamba Kiaxi	Kimbundu	3	<i>n. m.</i> Município da província de Luanda.
385	Kiluange*	Kiluanji	Kimbundu	2	<i>n. m.</i> distrito urbano da província do Luanda.
386	Kiluanje*	Kiluanji	Kimbundu	2	<i>n. m.</i> O mesmo que Kiluange.
387	Kiluanji*	Kiluanji	Kimbundu	2	<i>n. m.</i> Soberano do Reino Ngola, conhecido como Ngola Kiluanji.
388	Kimbanguista*	Kimbanguista	Kikongo e Português	2	1. <i>n. f.</i> Denominação religiosa. 2. <i>n. m. e f.</i> Pessoa dessa denominação religiosa.

389	Kimbundo	Kimunu, Ambundu	Kimbundu	2	1. <i>n. m.</i> Membro do grupo étnico dos Kimbundu, disseminados, na quase totalidade ao norte do rio Kwanza. 2. <i>adj. unif.</i> Relativo a esse povo ou a essa língua.
390	Kimbundu	Kimunu, Ambundu	Kimbundu	2	<i>n. m.</i> Língua do povo Kimbundu.
391	Kimpa*	Kimpa	Kikongo	2	<i>n. f.</i> Profetiza do reino do Kongo.
392	Kimuezo*	Kimuézu	Kimbundu	2	<i>n. m.</i> 1. Bárbaro, barbudo. 2. De barbas compridas. 3. Barba crescida e mal cuidada. 4. Barbaçana.
393	Kimuezu*	Kimuézu	Kimbundu	2	<i>n. m.</i> O mesmo que Kimuezo.
394	Kinanga*	Kinanga	Kimbundu	2	<i>n. f.</i> Bairro do município da Ingombota, província de Luanda.
395	Kinaxixi*	Kinaxixi	Kimbundu	2	<i>n. f.</i> Bairro do município da Maianga, província de Luanda.
396	Kindongo	Kindongo	Kimbundu	2	<i>adj.</i> 1. De rápido desenvolvimento. 2. Que cresce muito e facilmente. 3. Agitado; grande. 4. <i>n.</i> Pessoa alta e robusta.
397	Kingury*	Kingury	Kimbundu	2	<i>n. m.</i> Pseudónimo de Celso José, académico angolano.
398	Kipemba*	Kipembe	Kimbundu	2	<i>n. m.</i> Região do município de Golungo Alto, província do Kwanza-Norte.
399	Kipungo	Txipungo	Nyaneka	3	1. <i>n. m.</i> Natural do Quipungo, a sudoeste de Angola. 2. <i>n. m.</i> Dialeto falado nessa região. 3. <i>Pl.</i> População dessa área, pertencente ao subgrupo étnico dos Humbes.
400	Kiquiemba*	Kiquiemba	Umbundu	2	<i>n. f.</i> Comuna do município de Bolongongo, província do Kwanza-Sul.
401	Kissama	Kusasa	Kimbundu	2	<i>n. m.</i> 1. Língua do grupo quimbundo falada em Angola. 2. <i>Bras.</i> Mingau de mandioca. 3. Variedade de cana-de-açúcar. 4. <i>adj. m. e f.</i> Que diz respeito ou pertence aos Quiçamas, povo angolano que vive na margem esquerda do rio Kwanza. 5. <i>n. m. e f.</i> Indivíduo desse povo. 6. <i>Líng.</i> Que é relativo ao quiçama, língua falada em Angola.
402	Kiteculo	Tchitekulu	Umbundu	2	<i>adj.</i> 1. Benfeitor. 2. Pessoa bondosa e prestável
403	Kituxe	Kituxi	Kimbundu	2	<i>n. m.</i> Músico e compositor e angolano.
404	Kizua*	Kizua	Kimbundu	2	<i>n. m.</i> Músico angolano.
405	Kuango*	Kuangu	Kimbundu	2	<i>n. m.</i> Município da província da Lunda-Norte.
406	Kuangu*	Kuangu	Kimbundu	2	<i>n. m.</i> O mesmo que Kuango.
407	Kuanhama	Kwanyama	Kwanyama	1	1. <i>adj. m. e f.</i> que diz respeito ou pertencente aos Cuanhamas, povo do sudoeste de Angola. 2. <i>n. m. e f.</i> Indivíduo pertencente ao povo dos Cuanhamas. 3. <i>Líng.</i> Língua falada na região e pertencente ao grande grupo Ambó.
408	Kubango*	Kubango	Nganguela	2	<i>n. m.</i> Rio de Angola.
409	Kudima	Kudima	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de Quirima, na região de Malanje, ao norte do rio Kwanza. 2. <i>Pl.</i> População dessa área, pertencente ao grupo étnico Kimbundu.

410	Kueno*	Kueno	Kikongo	2	1. <i>n. m.</i> Cantor e compositor angolano. 2. <i>adj. unif.</i> Relativo a este cantor e compositor.	
411	Kufuna*	Kufuna	Cokwe	2	<i>n. m.</i> Estádio de futebol localizado no Luena, província do Moxico.	
412	Kuito	Kuito	Umbundu	2	<i>n. m.</i> Município da província do Bié.	
413	Kukuxi	Kukuxi	Kimbundu	2	<i>n. m.</i> Região da província de Luanda.	
414	Kulaxingo	Kulaxingo	Kimbundu	2	<i>n. m.</i> Rei de Cassanje, reino de Angola.	
415	Kunene	Kunene	Kwanyama	1	<i>n. m.</i> Província da República de Angola.	
416	Kungo	Waku Kungo	Umbundu	2	<i>n. m.</i> Município localizado na província do Kwanza-Sul.	
417	Kwanza	Kwanza	Umbundu	3	<i>n. m.</i> 1. Rio de Angola, cuja foz é 12 km ao noroeste da povoação de Katota, do posto mesmo nome circunscrição do Alto Kwanza – Chitembo, na província do Bié. 2. Unidade monetária de Angola, constituída em 1976 e posta em circulação a 8 de janeiro de 1977. Tem como fração o lwei.	
418	Kwanza-Norte*	Kwanza	Umbundu	3	1. <i>n. m.</i> Província da República de Angola. 2. <i>n. m. e f.</i> Indivíduo natural ou residente dessa província. 3. <i>adj. unif.</i> Relativo à população dessa província.	
419	Kwanza-Sul*	Kwanza	Umbundu	3	1. <i>n. m.</i> Província da República de Angola. 2. <i>n. m. e f.</i> Indivíduo natural ou residente dessa província. 3. <i>adj. unif.</i> Relativo à população dessa província.	
420	Kwenha*	Kwenha	Nganguela	2	<i>n. m.</i> Aeroporto localizado na província de Menongue.	
421	Laúca*	Laúca	Kimbundu	3	<i>n. f.</i> 1. Hidroelétrica construída sobre o rio Kwanza entre as províncias de Malanje e Kwanza-Sul. 2. Central hidroelétrica entre a província de Malanje e Kwanza-Sul.	Conhecido por Comandante Kwenha.
422	Leba	Leba	Umbundu	2	1. <i>n. f.</i> Formação montanhosa na província da Huíla. 2. <i>adj. unif.</i> Relativo a essa região.	
423	Lemba	Lemba	Kimbundu	2	<i>n. f.</i> 1. Espírito feminino que promove a procriação. 2. Entidade espiritual da procriação. 3. Nome dado a quem, por influência dessa entidade, nasce com uma anomalia específica. 4. <i>n. m. e f.</i> Indivíduo natural de Lemba, a sul de Angola a norte do rio Cúndji. 5. População dessa área, pertencente ao grupo étnico Umbundu. 6. <i>n. f.</i> Deusa protetora das mulheres grávidas.	
424	Leúa*	Leúa	Cokwe	2	<i>n. f.</i> Município da província do Moxico.	Dicionário de Kimbundu-Português.
425	Libolo	Lubolo	Kimbundu	1	1. <i>n. m. e f.</i> Natural de Libolo, na região do Kwanza-Sul. 2. <i>n. Ling.</i> Língua de Angola de um grupo Kimbundu da região do Kwanza-Sul. 3. <i>Pl.</i> População dessa área, pertencente ao grupo étnico dos Umbundos. 4. <i>adj. unif.</i> Relativo a essa população.	
426	Lilunga	Lilunga	Umbundu	2	<i>n. f.</i> Região localizada na província do Bié.	
427	Livulo*	Livulu	Kimbundu	2	<i>n. m.</i> 1. Vila da província de Malanje. 2. Livro.	

428	Livulu*	Livulu	Kimbundu	2	<i>n. m.</i> O mesmo que Livulo.	
429	Loanda*	Luuanda	Kimbundu*	1	1. <i>n. m. e f.</i> Indivíduo natural da cidade de Luanda, designadamente o procedente do cruzamento de etnias diferentes. 2. <i>Ling.</i> Variedade do Kimbundu falada na região de Luanda, em Angola.	
430	Lobinave*	Lobinave	Português em Angola	2	<i>n. m.</i> Sigla de Estaleiro Naval do Lobito.	
431	Lobito*	Lobito	Umbundu	2	1. <i>adj. unif.</i> Referente à essa região. 2. Município da província de Benguela.	
432	Londuimbali*	Londuimbali	Umbundu	2	<i>n. m.</i> Município da província do Huambo.	
433	Longonjo*	Longonjo	Umbundu	2	<i>n. m.</i> Município da província do Huambo.	
434	Lossambo*	Losambo	Umbundu	2	<i>n. m.</i> Centralidade da província do Huambo.	
435	Luacano*	Luakano	Umbundu	2	<i>n. m.</i> Vila da província do Moxico.	
436	Luachimo*	Luaximo	Umbundu	2	<i>n. m.</i> Vila da província do Moxico.	
437	Luaco*	Luaco	Cokwe	2	<i>n. m.</i> Vila do município de Cambulo, na província da Lunda-Norte.	
438	Luamba*	Luamba	Cokwe	2	<i>n. f.</i> Vila do município de Ambaca, na província da Lunda-Norte.	
439	Luanda	Luuanda	Kimbundu	1	<i>n. m.</i> 1. Indivíduo natural da cidade de Luanda, designadamente o procedente do cruzamento de etnias. 2. <i>n. f.</i> Audana. 3. O que uma pessoa paga ao soberano pelo exercício do seu comércio. 4. <i>Ling.</i> Variedade do Kimbundu falada na região de Luanda, em Angola.	
440	Luanda-Sul*	Luuanda-Sul	Kimbundu	2	<i>n. m.</i> Bairro do município de Viana, província de Luanda.	
441	Luanganji*	Luanganji	Cokwe	2	<i>n. m.</i> Rio de Angola, localizado na província do Moxico	
442	Luau*	Luau	Cokwe	2	<i>n. m.</i> Município da província do Moxico.	
443	Luavur*	Luavur	Cokwe	2	<i>n. m.</i> Bairro do município de Saurimo, província da Lunda-Sul.	
444	Luaxe*	Luaxe	Cokwe	2	<i>n. m.</i> Mina de diamante de Angola localizada na província da Lunda-Sul.	
445	Lubalo*	Lubalo	Cokwe	2	<i>n. m.</i> Município da província da Lunda-Norte.	
446	Lubango*	Lubango	Umbundu	1	1. <i>n. m.</i> Município da província da Huíla. 2. <i>n. m.</i> Município e capital da província da Huíla. 3. <i>adj. unif.</i> Que lhe diz respeito.	
447	Lucala*	Lukala	Kimbundu	3	<i>n. m.</i> Município da província do Kwanza-Norte.	
448	Lucapa*	Lucapa	Cokwe	2	1. <i>n. m.</i> Município e capital da província da Lunda Norte. 2. <i>adj. unif.</i> Que lhe diz respeito.	
449	Luchaz*	Lutyaji	Nganguela	1	1. <i>n. m. e f.</i> Indivíduo natural de uma área da província do Moxico, a leste de Angola. População dessa área pertencente ao grupo étnico dos Ganguelas. 2. <i>adj. unif.</i> Relativo a essa população 3. <i>pl. híbr.</i> Valuxazes ou Valutxazes.	
450	Luchazes*	Luchaze	Cokwe	2	<i>n. m.</i> Município da província do Moxico.	Variação de Luxaz, Lutxaz ou Balutxazes.

451	Lucinga*	Lucinga	Umbundu	2	<i>n. f.</i> Vila da província de Benguela, conhecida por Chá Lucinga.
452	Lucunga	Lucunga	Kikongo	2	<i>n. f.</i> Vila da província do Uíje.
453	Lueji*	Lueji	Cokwe	3	<i>n. f.</i> Rainha do reino Lunda Cokwe.
454	Luena	Luena	Cokwe	2	<i>n. m.</i> Município e capital da província do Moxico.
455	Luena	Luena	Cokwe	2	<i>n. m. e f.</i> 1. Indivíduo natural de uma área na província do Moxico, a leste de Angola. 2. <i>n. m.</i> Língua integrada no grupo lunda-quioico, falada em Angola. 3. <i>Pl.</i> População dessa área, pertencente ao grupo étnico dos ganguelas. 4. <i>adj. unif.</i> Que é relativo ao Luena, língua falada em Angola.
456	Luengue*	Luenge	Nganguela	2	<i>n. m.</i> Vila da província do Kwando Kubango.
457	Lueno	Valuena	Cokwe	1	1. <i>n. m.</i> Indivíduo natural de uma área na província do Moxico, a leste de Angola. Dialeto falado nessa região. 2. <i>pl.</i> população dessa área pertencente ao grupo étnico dos nganguelas. 3. <i>adj. unif.</i> Relativo a essa população. 4. <i>pl. híbr.</i> Valuenas.
458	Lufuiji*	Lufuiji	Cokwe	2	<i>n. m.</i> Rio de Angola localizado na província do Moxico.
459	Lumbiji*	Lumbiji	Kimbundu	2	<i>n. m.</i> Rio de Angola localizado na província do Kwanza-Norte.
460	Lumege*	Lumeji	Cokwe	2	<i>n. m.</i> Município da província do Moxico, conhecido como Lumeji Kameia.
461	Lumeji*	Lumeji	Cokwe	2	<i>n. m.</i> O mesmo que Lumege.
462	Lumueno*	Lumueno	Kikongo	2	<i>n. m.</i> Vila do município de Cuimba, na província do Zaire.
463	Lunda	Lunda	Cokwe	2	<i>adj. m. e f.</i> 1. Que diz respeito ou pertence à região da Lunda, em Angola. 2. Que diz respeito ou pertence aos Lundas, povo do Este de Angola, Zaire e Zâmbia, cujo império expansionista durou do século XVII ao século XIX. <i>n. f.</i> 3. Lugar que, por calamidade de mortes, foi relegado ao abandono. 4. Lugar que, por interdição de males, deixou de ser habitado. 5. <i>n. f. e m.</i> Indivíduo de uma área da Lunda, a nordeste de Angola. 6. <i>n. LÍng.</i> Que é relativo ao lunda, língua falada em Angola.
464	Lunda-Norte*	Lunda	Cokwe	2	<i>n. m.</i> Província da República de Angola
465	Lunda-Sul*	Lunda	Cokwe	2	<i>n. m.</i> Província da República de Angola
466	Lundas	Lunda	Cokwe	2	<i>adj. m. e f.</i> 1. Que diz respeito ou pertence à região da Lunda, em Angola. 2. Que diz respeito ou pertence aos Lundas, povo do Este de Angola, Zaire e Zâmbia, cujo império expansionista durou do século XVII ao século XIX.
467	Lundoloki*	Lundoloki	Kikongo	2	<i>n. m.</i> Atleta angolano.
468	Lundungo*	Lundungo	Kimbundu	2	<i>n. m.</i> Riacho da província de Malanje.
469	Lunga	Lunga	Kimbundu	2	1. <i>n. m. e f.</i> Indivíduo natural de Lunga, na região de Ambriz, ao norte de Angola. 2.

				População dessa área, pertencente ao grupo étnico kimbundu. 3. adj. unif. Relativo a essa população.	
470	Lunhinga	Lunhinga	Cokwe	2	<i>n. m.</i> 1. Rio de Angola localizado na província da Lunda-Sul.
471	Luquemba	Lukembo	Kimbundu	2	<i>n. f.</i> O mesmo que Luquembo
472	Luquembo	Lukembo	Kimbundu	2	1. n. m. Município da província de Malanje. 2. n. m. e f. Indivíduo natural ou habitante de Lukembo. 3. adj. unif. Que diz respeito a região de Lukembo.
473	Lussivi	Lusivi	Cokwe	2	<i>n. m.</i> Vila localizada na província do Moxico.
474	Lutucuta*	Lutukuta	Umbundu	2	<i>n. m.</i> Ministra da Saúde em Angola.
475	Luvale	Oluvale	Nyaneka	2	1. n. m. e f. Indivíduo natural de Luvale, a sueste de Angola. 2. n. m. Dialeto falado nessa região. 3. Pl. População dessa área pertencente ao grupo étnico dos Ganguelas. 4. adj. unif. Relativo a essa população. 5. n.m. Amassador de funji. 6. Espátula. 7. O mesmo que guíço ou muxarico entre os povos de língua kimbundu.
476	Luvambo*	Luvambo	Umbundu	2	<i>n. m.</i> 1. Vila da província de Benguela. 2. Rio de Angola.
477	Luvando*	Luvando	Umbundu	2	<i>n. m.</i> Vila do município de Chicomba, na província da Huíla.
478	Luvumbo	Luvumbo	Kikongo	2	<i>n. m.</i> O que, no recinto da circuncisão, prepara as máscaras para a celebração da saída dos circuncisos.
479	Luyindula*	Luyindula	Kimbundu	2	<i>n. m.</i> Centro médico localizado em Luanda.
480	Lwazi*	Lwazi	Kikongo	2	<i>n. f.</i> Atleta de natação.
481	Lwini*	Lwini	Português em Angola	2	<i>n. m.</i> Fundo de Solidariedade Social criado a 30 de junho de 1998, logo após a visita da princesa Diana em janeiro de 1997, e tem como objeto social a angariação de fundos e a execução de acção de apoio às vítimas civis de mina terrestre.
482	Mabor*	Mabor	Português em Angola	2	<i>n. m.</i> Bairro do município do Cazenga, província de Luanda.
483	Mabuba	Mabuba	Kimbundu	3	<i>n. f.</i> 1. Vila do município do Bengo. 2. Catarata, queda de água.
484	Mabuquila*	Mabu	Kimbundu	2	1. n. m. Grupo de papi. 2. Haste de papi. 3. n. f. Planta com peiácea.
485	Macondo	Macondo	Cokwe	2	<i>n. m.</i> Vila da província do Moxico.
486	Maculungungo	Makulungungo	Nganguela	2	<i>n. m.</i> Catarata localizada na província do Kwando Kubango.
487	Maculusso	Dikulusu	Kimbundu	2	<i>n. m.</i> 1. Antigo bairro africano da cidade de Luanda logo acima da Ingombota, no qual, outrora, funcionou um cemitério. 2. Antiga designação de tecido que, às riscas ou aos quadrados, se denomina por riscado. 3. n. m. pl. Cruzes, cemitério.
488	Macuma	Makuba	Kimbundu	2	<i>n. f.</i> 1. Rel. Religião comum no Brasil, que associa elementos do cristianismo, do animismo africano e de crenças ameríndias. 2. Rel. Ritual dessa religião que inclui

					danças e cantos ao som de tambor. 3. Prática de magia negra, bruxaria ou feitiçaria. 4. <i>Mús.</i> Instrumento musical de percussão, de origem africana, que produz um som de rapa
489	Macussa*	Makuse	Nganguela	2	<i>n. m.</i> Vila da província do Kwando Kubango.
490	Macuto	Makuta	Kimbundu	2	<i>n. m.</i> 1. Antiga moeda de cobre do valor de 30 reis. 2. Fracção do angolar, antiga moeda monetária de Angola, instituída em 1928 no governo do Alto Comissário António Vicente Ferreira, valendo 5 centavos.
491	Magombala*	Mbála	Kimbundu	1	<i>n. m.</i> 1. Tira de couro para cingir ou atar. 2. Correia. 3. Bernardo; Bernarda.
492	Maianga	Mayanga	Kimbundu	3	1. <i>n. m.</i> Antigo bairro africano da cidade de Luanda localizado na parte baixa. 2. <i>n. m. e f.</i> Indivíduo natural da Maianga. 3. <i>adj. unif.</i> Que diz respeito a Maianga. 4. <i>n. f.</i> Lagoas, onde o povo que vivia na ilha de Luanda buscava água para beber.
493	Maianga	Manyânga	Kimbundu	2	1. <i>n. m.</i> Antigo bairro africano da cidade de Luanda localizado na parte baixa. 2. <i>n. m. e f.</i> Indivíduo natural da Maianga. 3. <i>adj. unif.</i> Que diz respeito a Maianga.
494	Maiombe	Baiaka	Kikongo	2	<i>n. m.</i> Indivíduo natural do Iombe, a nordeste de Cabinda, ao norte de Angola, pertencente ao grupo étnico Kikongo
495	Makiesse*	Makiese	Kimbundu	2	<i>n. m.</i> Empresa angolana.
496	Makila*	Makila	Kimbundu	2	<i>n. m.</i> Humorista angolano.
497	Malange*	Malanji	Kimbundu	1	1. <i>n. f.</i> Província localizada a nordeste da República de Angola. 2. <i>n. m. e f.</i> Indivíduo natural de Malanje, localizada a nordeste de Angola.
498	Malengue*	lengue	Kimbundu	2	<i>n. m.</i> Indivíduo natural de Lengue, na margem do Cambo, a nordeste de Angola, pertencentes ao grupo étnico do Kimbundu.
499	Malesso*	Maleso	Kimbundu	2	<i>n. m.</i> Vila do município do Golungo Alto, na província do Kwanza-Norte.
500	Malongo*	Malongo	Kikongo	2	<i>n. m.</i> Terminal portuário de Cabinda.
501	Malueka*	Malueka	Kimbundu	2	<i>n. m.</i> Bairro do município do Cazenga, província de Luanda.
502	Mambuco*	Mpu-nkanda	Ngoyo	2	<i>n. m.</i> Governador do litoral do reino Ngoyo que usa um barrete confeccionado de peles de felinos que lhe confere autoridade.
503	Mampuya	Mampuya	Lingala	2	<i>n. m.</i> Atelier localizado no Zango, província de Luanda.
504	Mamuila	Hila	Nyaneca	1	1. <i>n. m. e f.</i> Indivíduo natural da Huíla, a sudoeste de Angola. 2. <i>n. m.</i> Dialeto falado nessa região, pertencente ao grupo nhaneca.
505	Mandjeno	Mandjeno	Kikongo	2	<i>n. m.</i> Vila do município do Cacongo, província de Cabinda.
506	Mandume*	Mandume	Kwanyama	1	<i>n. m.</i> Rei do grupo etnolinguístico Kwanyama.
507	Mangumbala*	Mangumbala	Kimbundu	2	<i>n. f.</i> Vila da província de Malanje.
508	Manguxi*	Manguxi	Kimbundu	2	<i>n. m.</i> Pseudónimo de António Agostinho Neto, primeiro presidente da República de Angola.

509	Manhinga*	Manhinga	Cokwe	2	<i>n. f.</i> Região da província do Moxico.
510	Mapunda	Mapunda	Umbundu	2	Bairro dos arredores do município de Lubango, província da Huíla.
511	Maquelo	Makelo	Umbundu	2	<i>n. m.</i> Região da província da Huíla.
512	Marçal	Marçal	Português em Angola	2	<i>n. m.</i> Bairro do município do Rangel, na província de Luanda.
513	Marimbate	Marimbate	Kimbundu	2	<i>n. m.</i> Grupo carnavalesco do município do Luquembo, província de Malanje.
514	Massabi	Massabi	Kikongo	2	<i>n. m.</i> Vila do município do Cacongo, província de Cabinda.
515	Massabi	Massabi	Kikongo	2	<i>n. m. e f.</i> Indivíduo natural de Massabi, município de Cacongo, na província de Cabinda.
516	Massangano*	Masa ngana	Kimbundu	1	1. n. m. Município da província do Kwanza-Norte. 2. n. m. e f. Indivíduo natural de Massangano, na província do Kwanza-Norte.
517	Massongue*	Massongue	Nganguela	2	<i>n. m.</i> Vila da província do Kwando Kubango.
518	Massunguna*	Massunguna	Kimbundu	2	<i>n. m.</i> Jogador angolano de futebol.
519	Matadi*	Matadi	Kikongo	2	<i>n. m.</i> Antigo aristocrata do reino do Congo, conhecido por Bula Matadi.
520	Matala	Matala	Umbundu	2	<i>n. f.</i> Município da província da Huíla.
521	Matamba*	Matamba	Kimbundu	2	<i>n. m.</i> Reino de Angola localizado na Baixa de Cassanje, província de Malanje.
522	Matenda*	Matenda	Kikongo	2	<i>n. m.</i> Vila do município de Songo, província do Uíge.
523	Matoso*	Mattoso	Kimbundu	1	<i>n. m.</i> Homem que não gosta de trabalhar.
524	Mavacala*	Mavacala	Kimbundu	2	<i>n. f.</i> Centralidade do município do Soyo, província do Zaire.
525	Mavambo*	Mavambo	Kikongo	2	<i>n. m.</i> Nkisi cultuado no Candomblé em Angola e no Congo como guardião dos caminhos.
526	Mavinga*	Mavinga	Nganguela	2	<i>n. f.</i> Município do Kwando Kubango.
527	Maxinde*	Maxinde	Kimbundu	2	<i>n. f.</i> Bairro da província de Malanje.
528	Mayembe*	Mayembe	Kikongo	2	<i>n. m. e f.</i> Indivíduo que é bom em muitas coisas.
529	Mayombe	Mayombe (do vernáculo Baiaka)	Kikongo	2	1. n. m. Indivíduo natural do Iombe, a nordeste de Cabinda, pertencente ao grupo do Kikongo. 2. n. f. Floresta de Angola.
530	Mbadi	Mbandi	Kimbundu	2	<i>n. f.</i> Heroína e rainha do Ndongo e da Matamba, conhecida por Nzinga Mbandi (1581 – 1663).
531	Mbala*	Mbála	Kimbundu	2	1. n. m. Região da província do Kwanza-Sul. n. m. 2. Tira de couro para cingir ou atar. 3. Correia. 4. Bernardo; Bernarda.
532	Mbandi	Mbandi	Kimbundu	2	<i>n. f.</i> Heroína e rainha do Ndongo e da Matamba, conhecida por Nzinga Mbandi (1581 – 1663).
533	Mbanza-Kongo	Mbanza-Kongo	Kikongo	2	<i>n. m.</i> Município e capital da província do Zaire.
534	Mbave*	Mbave (Mbawe)	Umbundu	2	<i>m. m.</i> Vila do município de Chicala-Choloanga, na província do Huambo.
535	Mbiavanga	Mbi ca venge	Kikongo	2	<i>n. m. Exp.</i> Que mal é que eu fiz?

536	Mbidi*	Mbidi	Kikongo	2	<i>n. m.</i> Região da província do Uíje.	
537	Mbridge*	Mbrige	Kikongo	2	<i>n. m.</i> Rio de Angola localizado na província do Zaire.	
538	Mbrige*	Mbrige	Kikongo	2	<i>n. m.</i> Região da província do Zaire.	
539	Mbundo	Mbundu	Kimbundu	2	<i>n. m.</i> 1. Grupo étnico bantu que vive em Angola, na região que se estende da capital de Luanda para leste. 2. Peixe do mar, mas de menor porte. 3. <i>Pl. híbr.</i> Jimbundos.	
540	Mbunga*	Mbunga	Kimbundu	2	<i>n. m.</i> Região da província do Uíje.	
541	Mbuyo*	Mbuyo	Umbundu	2	<i>n. f.</i> Vencedora da Taça Sayovo em feminino.	
542	Menongue	Menongue	Nganguela	2	1. n. m. Capital e município da província do Kwando Kubango. 2. adj. unif. Que diz respeito a esse município.	
543	Messala	Mesala	Kikongo	2	<i>n. m.</i> Região da província do Uíje.	
544	Messele	Mesele	Umbundu	2	<i>n. m.</i> Aldeia da comuna de Chipipa, província do Huambo.	
545	Messene	Messene	Português em Angola	2	<i>n. f.</i> Empresa especializada na criação e transformação de conteúdos em plataforma e-learning.	
546	Micolo*	Micolo	Kimbundu	2	<i>n. m.</i> 1. Vila localizada na província de Malanje. 2. n. f. Filadeira.	
547	Milunga*	Milunga	Kikongo	2	<i>n. f.</i> Vila localizada na província do Uíje.	
548	Mirabilis*	Ontumba	Umbundu	2	<i>n. f.</i> Planta cuja existência só há em Angola, no deserto do Namibe.	
549	Miramar*	Miramar	Português em Angola	2	<i>n. m.</i> Bairro localizado na província de Luanda.	
550	Missende*	Missende	Umbundu	2	<i>n. m.</i> Vila localizada no município de Catabola, província do Bié.	
551	Missombo*	Misombo	Nganguela	2	<i>n. m.</i> Comuna da província do Kwando Kubango.	
552	Missongo*	Misóngo	Kimbundu	2	<i>n. m.</i> Vila localizada no município de Luquembo, província de Malanje.	
553	Mixinge	Mixinje	Kimbundu	2	<i>n. m.</i> Escritor angolano.	
554	Môngua*	Môngua	Kwanyama	2	<i>n. f.</i> Vila do município de Cuanhama, província do Cunene.	
555	Moxico*	Moxico	Cokwe	2	<i>n. m.</i> Província da República de Angola.	
556	MPLA	MPLA	Português em Angola	2	1. n. m. Sigla de Movimento Popular de Libertação de Angola. 2. n. m. e f. Indivíduo pertencente a esse partido.	
557	Muacavula	Muakavula	Cokwe	2	<i>n. f.</i> Região localizada na província da Lunda Norte.	
558	Muambumba	Muambumba	Cokwe	2	1. n. m. Vila do município de Cacolo, província da Lunda Sul. 2. n. m. e f. Indivíduo natural ou habitante de Muambumba.	
559	Muanga	kubangama	Kimbundu	2	<i>n. f.</i> árvore espinhosa, atingindo 12 m, raras vezes mais; troncos eretos e copa volumosa mas não dilatada, folhas geralmente caducas.	
560	Muangala	Muangala	Cokwe	2	<i>n. f.</i> Vila localizada na província da Lunda Norte.	Lematizada como mubanga.
561	Muangolé	Muangolé	Português de Angola	2	<i>n. m. e f.</i> Indivíduo natural de Angola.	

562	Muangunza	Muangunza	Cokwe	2	<i>n. m.</i> Região localizada na província da Lunda Sul.
563	Muanha	Muanha	Cokwe	2	<i>n. f.</i> Vila localizada no município de Xa-Muteba, província da Lunda Norte.
564	Muanza	Muanza	Kimbundu	2	<i>n. f.</i> Árvore atingindo 30m, tronco cilíndrico erecto de cor castanha estiada de branco.
565	Muassangue	Muassangue	Kimbundu	2	<i>n. m. e f.</i> Indivíduo natural ou habitante de Massangano.
566	Muazaza*	Muazaza	Cokwe	2	<i>n. m.</i> Vila localizada no município de Muconda, província da Lunda-Sul.
567	Mucare*	Mucari	Kimbundu	3	<i>n. m.</i> Município da província de Malanje.
568	Mucari*	Mucari	Kimbundu	2	<i>n. m.</i> Município da província de Malanje.
569	Mucazo*	Mucazo	Kikongo	2	<i>n. m.</i> Região da província do Uíje.
570	Muchinga*	Muxinga	Kikongo	2	<i>n. m.</i> Região da província do Uíje.
571	Muconda*	Mukonda	Kimbundu	2	1. n. f. Município da província da Lunda Sul. 2. pron. Interrog. Porquê?; Por qual razão; causa ou motivo. 3. n. Motivo; razão; causa
572	Mucuenda*	Mukuenda	Kimbundu	2	<i>n. m.</i> Região da província de Malanje.
573	Mucundueji*	Mukundueji	Kimbundu	2	<i>n. m.</i> Rio de Angola.
574	Mucuta*	Múkuta	Kimbundu	2	<i>n. m.</i> 1. Antiga moeda de prata equivalente a dez paninhos de mabela. 2. Moeda de cobre no valor de 30 reis (0,003). 3. Unidade de conta da moeda angolana. 4. Macuta. 5. Bot. Planta ampelídea sarmentosa de frutos comestíveis.
575	Muekália	Muekália	Umbundu	2	<i>n. m.</i> Escritor e nacionalista angolano.
576	Muembeje	Muembeje	Kimbundu	2	<i>n. m.</i> Rio de Angola.
577	Muginga*	Muginga	Nganguela	2	<i>n. f.</i> Vila do município do Cuango, província do Kwando Kubango.
578	Muhongo	Muhongo	Kimbundu	2	1. n. m. Vila da Província de Malanje, também conhecida como Tala Mungongo. 2. n. m. e f. Denominação do indivíduo anormal que se revela por um desmaio, quando, ainda de colo, se acha com a mãe ou outra pessoa, num lugar de morte, quer num caminho com sepulturas, quer numa casa com defuntos. 3. adj. Egrégio; magnânimo; excelente.
579	Mukinda*	Mukinda	Umbundu	2	<i>n. m.</i> Deputada da Assembleia da República de Angola.
580	Mulaza	Mulaza	Kimbundu	2	<i>n. m.</i> Região localizada na província de Malanje.
581	Mulende*	Mulende	Kimbundu	2	<i>n. m.</i> Vila do município de Cambambe, província do Kwanza-Norte.
582	Mulenvo*	Mulenvo	Kimbundu	2	<i>n. m.</i> Bairro do município de Cacuaco, província de Luanda.
583	Mulombe	Mulombe	Cokwe	2	<i>n. m.</i> Região localizada na província da Lunda Sul.
584	Mumuíla	De Híla	Umbundu	2	<i>n. m. e f. e adj. unif.</i> Forma diversificada de Muíla.
585	Mundomba*	Ndombe	Umbundu	2	<i>n. f.</i> 1. Indivíduo natural do N'Dombe, na região de Benguela. 2. O mesmo que Dombe.

586	Mundombe	Ndombe	Umbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural do N'Dombe, na região de Benguela. 2. O mesmo que Dombe.	
587	Mundunduleno*	Mundunduleno	Cokwe	2	<i>n. m.</i> Estádio de futebol localizado no município do Luena, província do Moxico.	
588	Munjanga	Munjanga	Umbundu	2	<i>n. f.</i> Região da província da Huíla.	
589	Mupila	Mupila	Cokwe	2	<i>n. f.</i> Região da Lunda Norte.	
590	Musangola	Musangola	Português em Angola	2	<i>n. m.</i> Grupo musical angolano.	
591	Mussende	Musende	Kimbundu	2	1. <i>n. m. e f.</i> Indivíduo natural do Mussende, na Quibala, na província do Kwanza-Sul. 2. <i>n. m.</i> Dialeto falado nessa região. 3. <i>Pl.</i> População dessa área pertencente ao grupo étnico Kimbundu. 4. <i>adj. unif.</i> Relativo a essa população. 5. O mesmo que Sende.	
592	Mussuco	Musuco	Kikongo	2	1. <i>n. m. e f.</i> Indivíduo natural de Nsuku, a nordeste da Lunda Norte. 2. <i>n. m.</i> Dialeto falado nessa área. 3. População dessa localidade, pertencente ao grupo étnico Kikongo. 4. <i>adj. unif.</i> Relativo a essa população; o mesmo que Suco. 5. <i>Pl. híbr.</i> Bassucos.	
593	Mussulu*	Músulu	Kimbundu	2	<i>n. m.</i> 1. Península e comuna do município de Belas, província de Luanda. 2. Povo de pescadores ao sul da barra da Corimba, fronteira à ilha de Cazanga, província de Luanda.	
594	Mutamba	Mutamba	Kimbundu	2	<i>n. f.</i> 1. Bairro antigo de Luanda situado na baixa, ao fundo do edifício da Câmara Municipal. 2. Pequeno esqualo, mas de cabeça alongada. 3. Espécie de cação. 4. Arbusto de 3m de altura, com folhas persistentes.	
595	Muteba	Muteba	Cokwe	2	<i>n. f.</i> Município da província da Lunda Norte, também chamado de Xá-Muteba.	
596	Muto*	Mutu	Umbundu	2	<i>n. m.</i> Largo localizado província de Luanda.	
597	Mutu-Ya-Kevela*	Mutu-Ya-Kevela	Umbundu	2	<i>n. m.</i> Comandante militar e vigésimo Soma Inene do reino do Bailundo.	
598	Mutu*	Mutu	Umbundu	2	<i>n. m.</i> Relativo a Mutu-Ya-Kevela.	
599	Mutunda	Omutunda	Nyaneka	2	<i>n. m. e f.</i> 1. Indivíduo natural de Tunda, a sul de Angola, a sul do rio Cuvo. 2. O mesmo que Tunda.	
600	Muzangola*	Musangola	Português de Angola	2	<i>n. m.</i> Grupo musical angolano.	
601	Muzombo	Munzombo	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural da região compreendida entre Maquela e Damba, ao norte de Angola. 2. O mesmo que Zombo.	
602	N'dalatando	Ndanla tandu	Kimbundu	1	<i>n. m.</i> 1. Cidade localizada na província do Kwanza-Norte. 2. Indivíduo natural ou habitante da cidade de N'Dalatando, localizada na província do Kwanza-Norte.	
603	N'gola	N'gola	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de uma área de Malanje, ao norte do rio Kwanza. 2. O mesmo que Angola e Jinga. 3. <i>n.</i> Denominação dada aos reis do reino do	Era designada de Salazar, epitetada de

				Ndongo. 5. n. Denominação dada aos reis do reino do Ndongo.	“Princesa do Cazengo”.
604	Namacunde*	Namakunde	Kwanyama	2	<i>n. m.</i> Município da província do Cunene.
605	Nambuanguongo*	Nambuanguongo	Kimbundu	2	<i>n. m.</i> Vila localizada no município do Bengo, província de Luanda.
606	Namibe*	Namibe	Nyaneka	2	<i>n. m.</i> 1. Província da República de Angola. 2. Corresponde a Moçâmedes, antiga denominação substituída em 1982 por Namibe. 3. n. m. e f. Indivíduo natural ou habitante de província do Namibe, a sudoeste de Angola. 4. Namibense. 5. adj. unif. Que diz respeito à essa população.
607	Nancova*	Nancova	Ngangela	2	<i>n. f.</i> Município da província do Kwando Kubango.
608	Nangombe*	Nangombe	Umbundu	2	<i>n. m.</i> 1. Rio de Angola, localizado na província da Huíla. 2. n. m. Mercado localizado no município do Lubango, província da Huíla.
609	Nanguanza	Nanguanza	Cokwe	2	<i>n. m.</i> Vila da província da Lunda Sul.
610	Nassoma*	Nasoma	Umbundu	2	1. n. m. Rio de Angola, localizado na província do Huambo. <i>n. f.</i> 2. Rainha. 3. Mãe cujo primeiro filho se chama <i>Kasoma</i> . 4. Mulher de rei.
611	Naumba*	Naumba	Cokwe	2	<i>n. f.</i> Vila do município do Cuangar, na província da Lunda Norte.
612	Ndala	Ndala	Kimbundu	3	<i>n. m.</i> 1. Referente à região de Ndala Tandu. 2. Cobra de mordedura fatal, constituindo, pelo seu elevado poder mortífero, uma das piores serpentes da selva africana, e geralmente vivendo nas ramarias. <i>n. f.</i> 1. Cobra voadora. 2. Coisa ou comida do dia anterior. 3. Região do município da Baía Farta, na província de Benguela. 4. n. m. Vila da província do Kwanza-Sul, conhecida como Dala Kachibo.
613	Ndalatando	Ndala tandu	Kimbundu	2	<i>n. m.</i> Indivíduo natural ou habitante da cidade de N'Dalatando, localizada na província do Kwanza-Norte.
614	Ndalo	Ndálu	Kimbundu	2	<i>n. m.</i> 1. Forma aportuguesada de ndalu. 2. Fogo. 3. Carácter ardente.
615	Ndalu	Ndálu	Kimbundu	2	<i>n. m.</i> 1. Vestimenta interior tecida de fibra de embondeiro. 2. Cota. 3. Saia curta feita desta fibra. 4. n. m. Pseudónimo do General António Santos França, que foi ministro no antigo regime angolano. 5. n. f. Empresa angolana do ramo de navegação e logística empresarial.
616	Ndimbo	Ndimbo	Herero	2	1. adj. m. e f. Que é relativo ao dimba, língua falada em Angola. 2. n. m. Língua do grupo Herero.
617	Ndombe	Ndombe	Umbundu	2	1. n. m. e f. Natural do Dombe, na região de Benguela, a sul de Angola. 2. n. m. dialeto falado nessa região. 3. Pl. População dessa área pertencente ao grupo étnico Umbundu. 4. adj. unif. Relativo à essa população. 5. O mesmo que Mondombe e Mundombe. 6. Pl. híbr. Vandombes.

618	Ndongo*	Ndongo	Kimbundu	2	<i>n. f.</i> Antiga corte dos reis do Ndongo (Ngola).
619	Negage*	Negage	Kikongo	3	<i>n. m.</i> Município da província do Uíje.
620	Negaje*	Negaje	Kikongo	2	<i>n. m.</i> O mesmo que Negage.
622	Ngambaxi*	Nga mbáxi	Kimbundu	2	<i>n. m.</i> Senhor Sebastião.
623	Ngandavila*	Ngandavila	Umbundu	2	<i>n.</i> Região da província do Huambo.
624	Nganete*	Nga nete	Kimbundu	2	<i>n. f.</i> Senhora Anete.
625	Ngazuzé	Nga Zuzé	Kimbundu	2	<i>n. m.</i> Senhor José.
626	Nginga	Njinga Mbandi	Kimbundu	2	<i>n. f.</i> Heroína e rainha do Ndongo e da Matamba, conhecida por Nzinga Mbandi (1581 – 1663).
627	Ngola	N'gola	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de uma área de Malanje, ao norte do rio Kwanza. 2. O mesmo que Angola e Jinga.
628	Ngonguito*	Ngongo	Kimbundu	2	<i>n. m.</i> 1. Diminutivo de Ngongo. 2. Mundo; mundano.
629	Ngueve*	Ngeve	Umbundu	2	<i>n. f.</i> 1. Região da província do Huambo. 2. Segunda gémea. 3. <i>n. m.</i> Hipopótamo.
630	Ngunza	Ngunza	Kimbundu	2	<i>n. m.</i> Diferente denominação da entidade espiritual Mutakalombo.
631	Nguxi*	Nguxi	Kimbundu	2	<i>n. m.</i> Augusto.
632	Nhamuana*	Nhamuana	Cokwe	2	<i>n. f.</i> Vila do município de Alto Zambeze, província do Moxico.
633	Njinga	Njinga	Kimbundu	2	1. <i>n. m. e f.</i> Indivíduo natural de uma área de Malanje, ao norte do rio Kwanza. 2. <i>n. m.</i> dialeto falado nessa localidade. 3. <i>Pl.</i> População pertencente ao grupo étnico Kimbundu. 4. <i>adj. unif.</i> Relativo a essa tribo. 5. O mesmo que <i>Ngola</i> ou <i>Angola</i> .
634	Nkonde*	Nkonde	Cokwe	2	<i>n. m. e f.</i> Relativo a Lueji A'Nkonde.
635	Nkondi	Lueji	Cokwe	3	<i>n. f.</i> Rainha do reino Lunda Cokwe.
636	Nkunga*	Nkunga	Kikongo	2	<i>n. f.</i> Vila do município de Mbanza Kongo, província do Zaire, conhecida como Nkunga Paza.
637	Nocal	Nocal	Português em Angola	2	<i>n. f.</i> 1. Cerveja de Angola. 2. Bairro do município do Cazenga, província de Luanda.
638	Nóqui*	Nóqui	Kikongo	2	<i>n. m.</i> Município da província do Zaire.
639	Nsoki*	Nsoki	Kikongo	2	<i>n. f.</i> Cantora angolana.
640	Nsumbi*	Nsumbi	Umbundu	2	<i>n. m.</i> Município da província do Kwanza-Sul.
641	Nungulo*	Nungulo	Umbundu	2	<i>n. m.</i> Região da província do Huambo.
642	Nvundo*	Nvunda	Kimbundu	1	<i>n.</i> 1. Briga, luta, zaragata. 2. Desordem.
643	Nzagi	Nzangi	Kimbundu	2	<i>adj. unif. e n. m. e f.</i> 1. Boa pessoa. 2. Que é essencialmente bom. 3. Mártir. 4. Faisca elétrica; corisco.
644	Nzeto	Nzetu	Kikongo	2	1. <i>n. m.</i> Município da província do Zaire. 2. <i>n. m. e f.</i> Indivíduo natural ou habitante do município de Nzeto, província do Zaire. 3. <i>adj. unif.</i> Relativo a essa população.
645	Nzinga	Njinga	Kimbundu	2	1. <i>n. m. e f.</i> Indivíduo natural de uma área de Malanje, ao norte do rio Kwanza. 2. <i>n. m.</i>

				dialeto falado nessa localidade. 3. Pl. População pertencente ao grupo étnico Kimbundu. 4. adj. unif. Relativo a essa tribo. 5. O mesmo que Ngola ou Angola.	
646	Nzoji	Nzoji	Kimbundu	2	<i>n. m.</i> Município da província de Malanje.
647	Nzola*	Nzola	Kikongo	2	<i>n. f.</i> Primeira mulher a vencer o Luanda Slam.
648	Nzolani*	Nzolani	Kikongo	2	<i>n.</i> Campo polivalente do município de Mbanza Kongo, província do Zaire.
649	Nzuzi*	Nzuzi	Kikongo	2	<i>n. m. e f.</i> Nome dado ao segundo gêmeo à nascença, pois é considerado o primeiro, qua se chama Msimba.
650	O'mbaka*	O'mbaka	Umbundu	2	<i>n. m.</i> Estádio da província de Benguela.
651	Okavango*	Okavango	Umbundu	2	<i>n. m.</i> 1. Rio de Angola, localizado na província do Huambo. 2. O mesmo que Cubango. 3. adj. unif. Relativo a este rio.
652	Olombangui*	Olombangui	Umbundu	2	<i>n. m.</i> Grupo teatral angolano da província do Bié.
653	OMA	OMA	Português em Angola	3	<i>n. f.</i> Sigla de Organização Mulher Angolana.
654	Ombadja*	Ombadja	Kwanyama	2	<i>n. f.</i> Município da província do Cunene.
655	Ombaka*	Ombaka	Umbundu	2	<i>n.</i> Campo polivalente da província de Benguela.
656	Ombandja*	Ombandja	Kwanyama	2	<i>n. f.</i> Município da província do Cunene.
657	Onahumba*	Onahumba	Kwanyama	2	<i>n. f.</i> Bairro da cidade de Ondjiva, município do Kwanyama, província do Cunene.
658	Ondjiva*	Ondjiva	Kwanyama	2	1. n. f. Cidade do município do Kwanyama, província do Cunene. 2. n. m. e f. Indivíduo natural ou habitante de Ondjiva, província do Cunene. 3. adj. unif. Relativo a esta cidade.
659	Oshimolo*	Oshimolo	Kwanyama	2	<i>n. m.</i> Vila do município do Kwanyama, província do Cunene.
660	Ovinjenje	Ovinjenje	Kwanyama	2	<i>n. m.</i> Grupo carnavalesco da província do Huambo.
661	Pacavira	Mpaka	Kimbundu	1	1. n. m. e f. Indivíduo natural de Paca, ao norte de Angola. 2. n. m. Dialeto falado nessa região. 3. Pl. População dessa área, pertencente ao grupo étnico dos Kikongos. 4. adj. unif. Relativo a essa população, o mesmo que Quipaca. 5. n. f. curral, pequena choupana para resguardo da criação. O mesmo que <i>quibanga</i> entre os povos de língua kimbundu. 6. n. f. mamífero roedor da família dos caviídeos, medindo 70 cm de comprimento e 35 cm de altura.
662	Palanquinhas*	Palanga	Kimbundu	2	1. n. f. Antílope de corpulência de um cavalo, com chifres compridos, em curvatura para trás, medindo cerca de 1m, sendo a cauda curta. 2. n. m. e f. Indivíduo pertencente a seleção nacional de Angola. 3. Símbolo da República de Angola. 4. n. f. Esteira sobre a qual se realiza a cerimónia do xinguilamento.

663	Pangala*	Pangala	Kikongo	2	<i>n. f.</i> Vila do município do Soyo, província do Zaire.
664	Panguila*	Pangila	Kimbundu	2	1. n. m. Região do município do Bengo, província de Luanda. 2. n. m. e f. Indivíduo natural ou habitante do Panguila, município do Bengo, província de Luanda. 3. n. corog. Grande lago no caminho para o Dande, a 3 km de Kifangondo.
665	Panzo	Panza	Kimbundu	2	<i>n. f. Gír.</i> Jogo de azar. Batota.
666	Peliganga*	Peliganga	Kikongo	2	<i>n. m.</i> Advogada e política angolana.
667	Petrangol*	Petrangol	Português em Angola	2	<i>n. f.</i> Sigla de Companhia Angolana de Petróleo.
668	Petro*	Petróleo	Português em Angola	2	1. n. m. Clube desportivo de Luanda. 2. adj. unif. Relativo ao clube desportivo.
669	PGR	PGR	Português em Angola	3	<i>n. f.</i> Sigla de Procuradoria Geral da República.
670	Púcuta*	Púkuta	Kikongo	2	<i>n. f.</i> Rio de Angola, localizado na província de Cabinda.
671	Pululu*	Pululu	Umbundu	2	<i>adj. unif.</i> Limpo, transparente, saudável.
672	Pungo	Pungo	Kimbundu	3	<i>n. m.</i> 1. Referente à região do <i>Kipungo</i> . 2. Peixe marinho de grande corpulência ordinariamente aparecendo na época do Cacimbo.
673	Pungo Aluquém*	Pungo Aluquém	Kimbundu	2	<i>n. m.</i> 1. Comuna do município do Bengo, província de Luanda. 2. n. m. e f. Indivíduo natural ou habitante dessa região.
674	Pungo Andongo	Pungo Andongo	Kimbundu	1	<i>n. m.</i> 1. Extensão do planalto de Cacusó. 2. Peixe marinho de grande corpulência, ordinariamente aparecendo na época do Cacimbo.
675	Pungo-a-Ndongo*	Pungo ia Ndongo	Kimbundu	2	<i>n. m.</i> 1. Comuna da província de Malanje. <i>n. m.</i> 2. Antigo presídio então conhecido pelo nome das "Pedras Negras" fundado em 1397 e situado a 9° 4' 42" de lat. S, I 020 mtr. De altitude. 3. Antiga corte dos reis do Ndongo (Ngola).
676	Punguila*	Pangila	Kimbundu	2	<i>n. m.</i> O mesmo que Panguila.
677	Quemba	Kemba	Cokwe	2	1. n. f. Vila do município de Alto Zambeze, na província do Moxico. 2. n. m. e f. Indivíduo natural ou habitante dessa vila.
678	Quiamafulo*	Kiamafulo	Kimbundu	2	<i>n. m.</i> Região da província de Malanje.
679	Quibala	Kibala	Kimbundu	2	<i>n. m.</i> Língua do grupo kimbundu falada em Angola.
680	Quiçama	Kisama*	Kimbundu	2	1. adj. m. e f. Que diz respeito ou pertence aos kisamas, povo angolano que vive na margem esquerda do rio Kwanza. 2. n. m. e f. Indivíduo desse povo. 3. n. m. Ling. Língua do grupo Kimbundu falada em Angola. 4. Bras. Mingau de mandioca. 5. Bras. Variedade de cana de açúcar.
681	Quicanha*	Kikanha	Umbundu	2	<i>n. f.</i> Região da província do Kwanza-Sul.
682	Quicoca*	Kikoka*	Kimbundu	2	<i>n. f.</i> Região da comuna do Dande, município do Bengo, província de Luanda.
683	Quicula*	Kikula	Kikongo	2	<i>n. m.</i> Região da província do Zaire.

684	Quiculungo	Kikulungu	Kimbundu	2	<i>n. m.</i> 1. Pelicano. 2. Povo e sede do posto civil do concelho de Ambaca, província do Kwanza-Norte. 3. <i>n. m.</i> Município do Kwanza-Norte.
685	Quiufucussa*	Kifukusa	Kimbundu	2	1. <i>n. f.</i> Vila do município de Kiwaba Nzogi, província de Malanje. 2. <i>n. m.</i> Grupo carnavalesco angolano da província de Malanje.
686	Quihita*	Kihita	Umbundu	2	<i>n. f.</i> Vila do município da Chibia, província da Huíla.
687	Quihungo*	Kihungo	Kikongo	2	1. <i>n. m.</i> Variante da língua kikongo, falada na província do Uíje. 2. <i>n. m.</i> Região do município do Bengo, província de Luanda.
688	Quilevo*	Kileva	Kikongo	2	<i>n.</i> Região localizada no município do Uíje, província do Uíje.
689	Quiluange*	Kiluanji	Kimbundu	2	<i>n. m.</i> Soberano do Reino Ngola, conhecido como Ngola Kiluanji.
690	Quinanga*	Kinánga	Kimbundu	2	<i>n. f. top.</i> Lugarejo à beira da praia, ao sul da cidade de Luanda.
691	Quinfuquena*	Kinfuquena	Kikongo	2	<i>n. f.</i> Praia do Município do Soyo, província do Zaire.
692	Quioco	Txokwe	Cokwe	1	<i>adj. e n. m.</i> Forma deturpada de txokwe, usual na época colonial. Autótone da Lunda, a nordeste de Angola, constituindo membro da tribo dessa região.
693	Quipaca	Kimpaca	Kimbundu	2	1. <i>n. f.</i> Antiga moeda de cobre do valor de 10 reis, ou, modernamente, 1 centavo. 2. <i>n. m. e f.</i> Forma diversificada de paca.
694	Quipembe*	Kipembe	Kikongo	2	<i>n. m.</i> Vila do município de Mucaba, província do Uíje.
695	Quipungo	Txipungo	Umbundu	2	1. <i>n. m. e f.</i> Indivíduo natural de Kipungu, a sudoeste de Angola. 2. <i>n. m.</i> Dialeto falado nessa região.
696	Quirima	Kirima*	Kimbundu*	2	1. <i>n. m. e f.</i> Indivíduo natural de Kirima, na região de Malanje, ao norte do rio Kwanza. 2. <i>Pl.</i> População dessa área pertencente ao grupo étnico Kimbundu. 3. <i>adj. unif.</i> Relativo à essa população.
697	Quissama	Kisama*	Kimbundu*	1	1. <i>n. m. e f.</i> Indivíduo natural de Quissama, na província do M'bengo, ao sul do rio Kwanza. 2. <i>Pl.</i> População dessa área, pertencente ao grupo étnico Kimbundu. 3. <i>n. m. e f.</i> Indivíduo desse povo. 4. <i>adj. m. e f.</i> Que diz respeito ou pertence aos kisamas, povo angolano que vive na margem esquerda do rio Kwanza. 5. <i>n. m. Ling.</i> Língua do grupo Kimbundu falada em Angola. 6. <i>Bras.</i> Mingau de mandioca. 7. <i>Bras.</i> Variedade de cana de açúcar.
698	Quissanga*	Kisanga	Kimbundu	2	<i>n. m.</i> 1. Achadouro. 2. Enseada; pequena baía. 3. Pénínsula. 4. Ilha junto de terra. 5. <i>adj.</i> Achadiço. 6. Que está à vista; fácil de encontrar. 7. Achado; encontrado. 8. <i>adj. unif.</i> Povo da ilha deste nome na região do rio Nzenza, circunscrição de Icolo e Bengo, província de Luanda.
699	Quissoa*	Kisoa	Kikongo	2	<i>n. f.</i> Região da província do Uíje.

700	Quissolene*	Kisole	Kimbundu	2	<i>adj.</i> 1. Escolhido; distinguido. 2. <i>n.</i> O que pode ou deve ser escolhido. 3. <i>Bot.</i> Árvore autocárpia do género <i>figus</i> .
701	Quissolone*	Kisole	Kimbundu	2	<i>n. m.</i> Povo e região localizada na província de Malanje.
702	Quissoma*	Kisoma	Kimbundu	2	<i>n. f. Bot.</i> Planta euforbiácea de fruto corrosivo, utilizada para sebe e ornamento.
703	Quissonde	Kusondoloka	Kimbundu	1	<i>n. m. e f.</i> 1. Formigão avermelhado, de mordedura dolorosa. 2. <i>Bras.</i> Crauçanga. Murupeteca. Taoca.
704	Quitamba*	Kitamba	Kimbundu	2	<i>n. f.</i> 1. Vila do município de Massango, na província de Malanje. 2. Fogueira de lenha para fazer carvão. 3. Pira. 4. <i>Fig.</i> Brasileiro.
705	Quiteculo	Kitekulu	Kimbundu	2	<i>n. m.</i> Região da província do Kwanza-Norte.
706	Quitexe*	Kitexi	Kimbundu	2	1. <i>n. m.</i> Município da província do Uíje. 2. <i>n. m. e f.</i> Indivíduo natural ou habitante de Kitexi.
707	Quitumba*	Kitumba	Cokwe	2	1. <i>n. m.</i> Vila do município de Xá-Muteba, província da Lunda Norte, conhecida por Cassange Quitumba. <i>adj.</i> 2. Mata expeça de plantas de pouca altura. 3. Entravado. 4. Autómato; cego. 5. Palhote. 6. <i>Fig.</i> Armação de madeira.
708	Quivuenga*	Kivuenga	Kikongo	2	<i>n. m.</i> Vila localizada na província do Uíje.
709	Ramiros*	Ramiros	Português em Angola	2	<i>n. m.</i> Comuna do município de Belas, província de Luanda.
710	Rangel*	Rangel	Português em Angola	2	<i>n. m.</i> Município da província de Luanda.
711	Refriango*	Refriango	Português em Angola	2	<i>n. f.</i> Empresa angolana especializada na produção e distribuição de refrigerantes sumos, águas, bebidas energéticas e bebidas alcoólicas.
712	Rivungo*	Rivungo	Ngangela	2	<i>n. m.</i> Município da província do Kwando Kubango.
713	RNA*	RNA	Português em Angola	2	<i>n. f.</i> Sigla de Rádio Nacional de Angola.
714	Sabalo*	Sabalu	Kimbundu	2	<i>n. m.</i> Sábado.
715	Sacaluile*	Sacaluile	Cokwe	2	<i>n. m.</i> Bairro do município da Muconda, província da Lunda Sul.
716	Sachambula	Sachambula	Umbundu	2	<i>n. m.</i> Advogado da Sonangol.
717	Sachilombo*	Sachilombo	Cokwe	2	<i>n. m.</i> Região do município do Alto Zambeze, província do Moxico.
718	Sachipando*	Sachipando	Umbundu	2	<i>n. m.</i> Diretor da Direção Nacional de Investigação e Ação Penal.
719	Sachuma*	Sachuma	Cokwe	2	<i>n. f.</i> Região da província da Lunda Norte.
720	Saco-mar*	Saco-mar	Umbundu	2	<i>n. m.</i> Distrito urbano da província do Namibe.
721	Sakuanda*	Sakuanda	Cokwe	2	<i>n. f.</i> Região da província do Moxico.
722	Salaquiaco*	Salakiako	Kimbundu	2	<i>col.</i> Trabalha o que é teu.
723	Salumbo	Salumbo	Umbundu	2	<i>n. m.</i> Região da província do Huambo.
724	Samakuva	Samakuva	Umbundu	2	<i>n. m.</i> Líder do partido União Nacional para a Independência Total de Angola.

725	Samalaca	Samalaca	Cokwe	2	<i>n. m.</i> Região da Lunda Norte.
726	Sambaca	Samalaca	Cokwe	2	<i>n. m. e f.</i> Indivíduo natural ou habitante dessa região.
727	Sambala	Samalaca	Cokwe	2	<i>adj. unif.</i> Que lhe diz respeito.
728	Sambambe	Sambambe	Umbundu	2	<i>n. m.</i> Região do Tchikala-Tcholo, província do Huambo.
729	Sambila	Sambila	Português em Angola	2	<i>n. m.</i> O mesmo que Sambizanga.
730	Sambingo*	Sambingo	Umbundu	2	<i>n. m.</i> Região da província do Huambo.
731	Sambinzanga*	Sambizanga	Kimbundu	2	<i>n. m.</i> O mesmo que Sambizanga.
732	Sambizanga*	Sambizanga	Português em Angola	3	<i>n. m.</i> Município da província de Luanda.
733	Sambuquila*	Sambukila	Umbundu	2	<i>v.</i> Contagiar; contaminar.
734	Samussanda	Samusanda	Cokwe	2	<i>n. m.</i> Queda e região turística da província da Lunda-Sul
735	Sandemba*	Sandemba	Cokwe	2	<i>n. f.</i> Vila do município de Saurimo, província da Lunda Sul.
736	Sangabi*	Sangabi	Kimbundu	2	<i>n. f.</i> Vila do município de Golungo Alto, província do Kwanza-Norte.
737	Sangano*	Sangano	Kimbundu	2	<i>n. m.</i> Praia localizada na província de Luanda.
738	Sango	Sango	Kimbundu	2	<i>n. m.</i> Festim consagrado a alma de um extinto, a fim de a tranquilizar das faltas terrenas.
739	Sanguveve*	Ngeve	Umbundu	2	<i>n. m.</i> Hipopótamo.
740	Sant'ana*	Santa Ana	Português em Angola	2	<i>n. f.</i> Cemitério localizado na província de Luanda.
741	Sapú*	Sapú	Português em Angola	2	<i>n. m.</i> Bairro do município do Kilamba Kiayi.
742	Sassoma*	Sasoma	Umbundu	2	<i>n. m.</i> Pais de Soma.
743	Satambi	Satambi	Cokwe	2	<i>n. m.</i> Região da província da Lunda Norte.
744	Saurimo*	Saurimbu	Umbundu	3	1. n. m. Município da província da Lunda-Sul. 2. adj. unif. Relativo a essa população.
745	Saviemba*	Saviemba	Cokwe	2	<i>n. m.</i> Vila do município de Lumbala Nguimbo, província do Moxico.
746	Sequele*	Sekele	Kimbundu	2	1. n. m. Centralidade da província de Luanda. 2. n. m. e f. Indivíduo que habita nesta centralidade.
747	SIAC*	SIAC	Português em Angola	2	<i>n. m.</i> Serviço Integrado de Atendimento ao Cidadão.
748	Simulambuco	Simulambuko	Kikongo	2	<i>n. m.</i> Tratado entre Portugal e o Reino do Nêgoio.
749	Sodiam*	Sodiam	Português em Angola	2	<i>n. f.</i> Sigla de Sociedade de Comercialização de Diamantes.
750	Sombo*	Sombo	Kimbundu	2	<i>n. m.</i> Planta de casca taminosa da família das combretáceas, utilizada para curtimentos.
751	Sonagás*	Sonagás	Português em Angola	3	<i>n. f.</i> Sigla de Sonangol Gás Natural.
752	Sonagaz*	Sonagaz	Português em Angola	2	<i>n. f.</i> Empresa de gás natural subsidiária da Sonangol.

753	Sonair*	Sonair	Português em Angola	2	<i>n. f.</i> Empresa de aviação angolana que pertence totalmente à empresa petrolífera estadual Sonangol.
754	Sonamet	Sonamet	Português em Angola	2	<i>n. f.</i> Empresa angolana fabricante de estruturas metálicas para a produção e prospeção de petróleo, constituída pela Sonangol e a Acergy.
755	Sonangol*	Sonangol	Português em Angola	2	1. n. f. Sigla de Sociedade Nacional de Combustível de Angola. 2. adj. unif. Que lhe diz respeito.
756	Songo	Kusongoloka	Kimbundu	2	<i>n. m. e f.</i> 1. Indivíduo natural de Songo, na região de Malanje, ao norte do rio Kwanza. 2. n. m. Dialeto falado nessa região. 3. Pl. População dessa área pertencente ao grupo étnico Kimbundu. 4. adj. unif. Relativo a essa população. 5. n. m. 1. Formigão preto mais pequeno que o zeu. 6. n. m. Sortilégio que ocasiona uma longa e cruciante pontada, ordinariamente mortal num curto espaço de tempo.
757	Soyo	Soyo	Kikongo	2	1. n. m. Município da província do Zaire. 2. n. m. e f. Indivíduo natural ou habitante deste município.
758	Sumbe	Sumbe	Umbundu	3	1. n. m. Município da província do Kwanza-Sul. 2. n. m. e f. Indivíduo natural do Sumbe em relação à tribo, localizado no Kwanza-Sul. 3. Dialeto falado nessa região. 4. Pl. População dessa área pertencente ao grupo étnico dos Umbundos. 5. adj. unif. Relativo a essa população. 6. O mesmo que <i>Mussumbe</i> e <i>Pinda</i> ou <i>Mupinda</i> .
759	TAAG	TAAG	Português em Angola	3	<i>n. f.</i> Sigla de Transportes Aéreos de Angola.
760	Tala-Hady*	Tala-hady	Kimbundu	2	<i>n. m.</i> 1. Bairro do município do Cazenga, província de Luanda. 2. col. Olha o sofrimento.
761	Tala*	Tala	Kimbundu	2	<i>n. f.</i> 1. Berço de rio que se estende pela terra dentro. 2. Vala ou regueiro que comunica com um rio, enchendo-se na época das chuvas. 3. Lagoa. 4. Abreviação de ritala. 5. adv. Ao alto; em cima. 6. interj. Olha; toma cuidado.
762	Talatona*	Talatona	Kimbundu	3	1. n. m. Bairro do município de Belas, província de Luanda. 2. adj. unif. Que lhe diz respeito.
763	Tandu*	Tandu	Kimbundu	3	<i>n. m.</i> 1. Referente à região de Ndala Tandu. 2. Parte superior de um objeto elevado.
764	Tchicolondo*	Tchicolondo	Kikongo	2	<i>n. m.</i> Ponto fronteiro entre Angola e República Democrática do Congo.
765	Tchiela	Tchiela	Kikongo	2	<i>n. f.</i> Região da província de Cabinda.
766	Tchifuchi*	Tchifuchi	Cokwe	2	<i>n. m.</i> Bairro do município do Luena, província do Moxico.
767	Tchilombo*	Tchilombo	Umbundu	2	<i>n. m.</i> 1. Acampamento. 2. Tenda.
768	Tchilunda*	Tchilunda	Umbundu	2	<i>n. f.</i> Região da província da Lunda Norte.
769	Tchimbala*	Tchimbali	Umbundu	2	<i>n. m.</i> Negro assimilado.

770	Tchimbonde*	Tchimbonde	Umbundu	2	<i>n. m.</i> região da província do Huambo.
771	Tchingando*	Tchingando	Umbundu	2	<i>n. m.</i> Região da província do Bié.
772	Tchingi*	Tchingi	Cokwe	2	<i>n. m.</i> Região da província da Lunda Norte.
773	Tchinhama*	Tchinhama	Umbundu	2	<i>n. f.</i> Comuna da província do Huambo.
774	Tchinjenje*	Tchingenge	Umbundu	2	<i>n. m.</i> 1. Milho. 2. Vinco.
775	Tchipungo	Tchipungo	Umbundu	2	<i>n. m.</i> Município da província da Huíla.
776	Tchiteculo*	Tchitekulu	Umbundu	2	<i>n. m.</i> 1. Bem feitor. 2. Pessoa bondosa e prestativa.
777	Tchitondotondo	Tchitongotongo	Umbundu	2	<i>n. m.</i> O mesmo que Tchitongotongo.
778	Tchitongotongo*	Tchitongotongo	Umbundu	2	<i>n. m.</i> Bairro da comuna do Kihita, província da Huíla.
779	Tchitunda*	Tchitunda	Umbundu	2	<i>n. m.</i> Líder religioso.
780	Tchitundu*	Tchitundu-Hulu	Umbundu	2	<i>n. m.</i> Morro granítico situado no município de Virei, província do Namibe.
781	Tchiutuka*	Tchiutuka	Umbundu	2	<i>n. m.</i> Grupo carnavalesco da província do Huambo.
782	Tchivinda*	Tchivinda	Umbundu	2	<i>n. m.</i> 1. Região da província do Huambo. 2. Ferreiro; artífice.
783	Tchiwana*	Tchiwana	Umbundu	2	<i>n. f.</i> Fazenda localizada no município do Chongoroi, província de Benguela.
784	Tchiyaka*	Tchiyaka	Umbundu	2	<i>n. m.</i> Antigo reino de Angola, localizado na província do Huambo.
785	Tecassala*	Tecassala	Kikongo	2	<i>n.</i> Estilista angolano.
786	Tembo*	Tembo	Kimbundu	2	<i>n. m.</i> Comuna do município de de Malanje, província de Malanje.
787	Textang*	Textang	Português em Angola	3	<i>n. f.</i> Empresa industrial têxtil de Angola.
788	Tomboco*	Tomboco	Kikongo	2	<i>n. m.</i> 1. Município da província do Zaire. 2. <i>n. m. e f.</i> Indivíduo natural deste município.
789	Tombua*	Tômbwa	Umbundu	2	<i>n. m.</i> O mesmo que Tômbwa.
790	Tômbwa*	Tômbwa	Umbundu	2	<i>n. m.</i> 1. Município da província do Namibe. 2. <i>n. m. e f.</i> Indivíduo natural ou habitante do Tômbwa. 3. <i>adj. unif.</i> Que lhe diz respeito.
791	TPA*	TPA	Português em Angola	2	<i>n. f.</i> Sigla de Televisão Pública de Angola.
792	Tubiyáxikelela*	Tubiyáxikelela	Umbundu	2	<i>n. m.</i> Grupo musical angolano.
793	Tunguno*	Tungunu	Kimbundu	2	<i>n. m.</i> 1. O primeiro na ordem das coisas. 2. O mais votado; o mais velho; o mais antigo. 3. Principal. 4. Que precede a todos.
794	Twafundumuca*	Twafundumuka	Kimbundu	2	<i>n. m.</i> Grupo carnavalesco do Rangel, província de Luanda.
795	Txissola	Tchisola	Umbundu	2	<i>n. f.</i> Amor; caridade; bondade.
796	Uanhenga	Uanhenga	Kimbundu	2	<i>n. m.</i> O poder é odiado.
797	Uanhenguiano	Uanhenga	Kimbundu	2	<i>adj.</i> Que diz respeito a Uanhenga Xitu.
798	Úcua	Úcua	Kimbundu	2	<i>n. m.</i> 1. Vila da comuna do Dande, município do Bengo, província de Luanda. 2. <i>adj. unif.</i> Relativo à vila do Úcua.

799	Uíje	Uíje	Kikongo	2	1. <i>n. m.</i> Província da República de Angola. 2. <i>n. m. e f.</i> Indivíduo natural ou habitante do Uíje. 3. <i>adj. unif.</i> Relativo a província do Uíje.
800	Uigense	Uíje	Kikongo	2	<i>n. m. e f.</i> 1. Indivíduo natural ou habitante do Uíje. 2. <i>adj. unif.</i> Relativo à essa população.
801	Uíje*	Wizi	Kikongo	3	<i>n. m.</i> 1. Chegada. 2. Província da República de Angola. 3. Indivíduo natural dessa província.
802	Ukuandumbo*	Ndumbo	Kimbundu	2	<i>n.</i> Nome genérico de vários frutos das plantas paperáceas de diversas espécies.
803	Ulengo*	Ulengelu	Kimbundu	2	<i>n.</i> 1. Maneira de debandar; fugir. 2. Qualidade de corrida.
804	Umbundu	Mbundu	Umbundu	2	1. <i>n. m. e f.</i> Membro do grupo étnico Umbundu, disseminados ao sul do rio Kwanza, ocupando a zona meridional e central de Angola. 2. <i>adj. m. e f. Ling.</i> Que é relativo ao umbundu, língua falada em Angola. 3. <i>n. m.</i> Língua bantu falada pelos Ovimbundu, habitantes da zona meridional e central de Angola.
805	UNAC-SA	UNAC-SA	Português em Angola	2	<i>n. m.</i> Sigla de União Nacional dos Artistas e Compositores – Sociedade de Autores.
806	UNACA*	UNACA	Kimbundu	2	<i>n. f.</i> Sigla da Confederação das Associações de Camponeses e Cooperativas Agropecuárias de Angola.
807	UNITA	UNITA	Português em Angola	2	1. <i>n. f.</i> Sigla de União Nacional para a Independência Total de Angola. 2. <i>n. m. e f.</i> Indivíduo pertencente a este partido político. 3. <i>adj. unif.</i> Relativo a esse partido.
808	Uqueta*	Uketa	Umbundu	2	<i>n. f.</i> Região da província do Kwanza-Sul.
809	Utxocwe*	Cokwe	Cokwe	2	<i>n. m. e f.</i> Indivíduo natural da área da Lunda a nordeste de Angola.
810	Vemba*	Nvemba	Kikongo	2	<i>n. m.</i> 1. Açor. 2. O mesmo que Bemba, entre os povos de língua kimbundu.
811	Viana*	Viana	Kimbundu	2	<i>n. f.</i> Município da província de Luanda.
812	Vidrul*	Vidrul	Português em Angola	2	<i>n. f.</i> Unidade de produção de garrafa.
813	Vindongo	Ovindonga	Umbundu	2	<i>n. m.</i> 1. Indivíduo natural de uma área compreendida entre os cursos do Cubango e do Cuando, a sul de Angola. 2. O mesmo que xindongas.
814	Vunge	Nvunji	Kimbundu	2	<i>n. f.</i> 1. Espírito feminino que administra a justiça. 2. Entidade espiritual da justiça. 3. <i>adj. unif.</i> Aquele cuja conceção, por influência desse espírito é caracterizada pela ausência do cataménio, desde o nascimento do último filho ao referente a este caso. 4. <i>n. m. e f.</i> Nome dado a tal indivíduo.
815	Wambo	Wambo	Umbundu	2	<i>n. m.</i> O mesmo que Huambo, província de Angola.
816	Welwitchia	Ontumba	Umbundu	2	<i>n. f.</i> 1. Designação botânica. 2. Planta cuja existência só há em Angola, no deserto do Namibe.
817	Wenda*	Wendo	Umbundu	2	<i>n. m.</i> Pessoa que nasce depois de muitos abortos.

818	Weza*	Weza	Kimbundu	2	<i>n. m. e f.</i> 1. Veio. 2. Chegou.
819	Xá-muteba*	Xá-muteba	Cokwe	2	1. n. m. Município da província da Lunda Norte. 2. Relativo a população dessa região.
820	Xirimimbi	Xirimimbi	Umbundu	2	<i>n. m.</i> Ex-Ministro das Pescas de Angola.
821	Xitu	Xitu	Kimbundu	2	<i>n. m.</i> Pseudónimo do escritor angolano Mendes de Carvalho, Uanhenga Xitu.
822	Yaka	Yaka	Kikongo	2	1. n. m. e f. Individuo natural de Yaka, ao norte de Angola. 2. n. m. Dialeto falado nessa região. 3. População dessa área pertencente ao grupo étnico Kikongo. 4. Adj. unif. relativo a essa população; o mesmo que Muiaca e Quiaca. 5. Pl. híbr. Baiacas.
823	Yembe*	Yembe	Kimbundu	2	<i>n. m.</i> Vila da comuna do Tabi, município do Ambriz, distrito do Bengo, província de Luanda.
824	Yetu*	Yetu	Umbundu	2	<i>n. m.</i> 1. Instituição bancária angolana. 2. Pron. poss. Nosso.
825	Zaire	Zaire	Kikongo	2	<i>n. m.</i> Província da República de Angola.
826	Zango	Zango	Português em Angola	2	<i>n. m.</i> 1. Projeto de habitação social, situado a sudeste de Luanda. 2. Bairro do município de Viana, província de Luanda.
827	Zavula	Zavula	Kimbundu	2	<i>n.</i> Vila do município do Cazengo, província do Kwanza-Norte.
828	ZEE*	ZEE	Português em Angola	2	<i>n. f.</i> Sigla de Zona Económica Especial.
829	Zenza*	Zenza	Kimbundu	1	<i>n. m.</i> 1. Rio de Angola. 2. Empresa de venda de carros e acessórios.
830	Zongo	Zongo	Kikongo	2	<i>n. m.</i> Região do município da Damba, província do Uíje.

Anexo III – Estruturas passivas do *Telejornal de Angola*

Estruturas passivas eventivas longas

1. Nguinamau Luvuvamo é surpreendido pelos efetivos do SIC.
2. Foram incluídos no plano apenas os projetos já com o financiamento por linha de crédito internas e externas garantidos.
3. Hélder Silva foi acusado na quinta-feira pelo serviço de investigação criminal dos crimes de associação criminosa e burla por defraudação.
4. A comissão eleitoral é composta por Elizabeth Pedro, Joana Raimundo e Paulo Daniel.
5. Os beneficiários dos imóveis comercializados pela Imogestin já estão a ser descontados por débito direto bancário mensalmente em regime de propriedade de renda resolúvel, mas não habitam nas residências porque foram arrombadas e comercializadas para terceiros que as ocupam ilegalmente.
6. Não se pode de forma alguma que os imóveis do estado estejam sendo comercializados por entidades na qual não têm responsabilidade nenhuma.
7. É muito bem as forças de investigação criminal em Benguela que logo de pronto travaram essa onda e essas pessoas serão julgadas também pelo delito que assim cometeram.
8. E no quadro da prevenção e combate a COVID-19, várias são as medidas que têm sido anunciadas pela comissão intersectorial criada para o efeito.
9. O número que foi divulgado pelos diferentes órgãos de informação privada de duzentos e tal mil postos de trabalho não corresponde à verdade.
10. Onze anos depois os empreendimentos hidroelétricos do Lubão na província de Benguela e das Mabubas no Bengo voltam a ser geridos pelo estado.
11. De forma voluntária, eles entregaram à polícia dez armas de fogo, mas no momento em que tudo isso acontecia, dois cidadãos também na Calemba 2 foram acusados de roubar motorizadas e foram detidos na hora pela polícia.
12. O conflito na Igreja Universal em Angola é de longe, mas acentuou-se quando em novembro de 2019 bispos e pastores angolanos assinaram um manifesto pastoral defendendo que a instituição passasse a ser administrada apenas por angolanos.
13. O grupo foi recebido esta manhã pela Ministra da Saúde.
14. E que foi absorvido por essa empresa internacional renomada e que já tem estado a dar apoio internacional em vários países com construção e implementação de projeto de laboratório desta natureza.
15. As medidas de prevenção e combate à pandemia são observadas pelos dois países.
16. O assunto foi debatido pelos deputados.
17. Uma grande iniciativa e nós temos que reconhecer o esforço que, como sabem, a iluminação pública é um trabalho que até ao momento é feito pela administração municipal de distrito assim como o governo da província, mas naturalmente nós, ENDE, estamos sempre aqui para dar aquilo que é o apoio naquilo que for possível.
18. Hectares de cana-de-açúcar da Biocom foram destruídos desde junho por queimadas anárquicas perpetradas por caçadores furtivos, causando prejuízos que vão para lá dos 200000000 kwanzas.
19. A novidade é aguardada pelas vendedoras da pracinha do Papá Simão e do Kikolo, em Cacucaco.
20. A maior parte dos casos são em termos geográficos para Luanda e províncias fronteiras. Também foi dito aqui já pelo senhor comandante geral da polícia.
21. O último adeus ao general Kundi Paihama foi marcado por dor, consternação e mensagens que retratam o legado deixado pelo malogrado para a família, amigos, camaradas e ao país.
22. A LIMA está reunida aqui em Luanda.
23. Especialistas disseram ainda ser nesta época que parte dos cardumes deslocam-se até as zonas dos trópicos, sendo seguidos pelas baleias.
24. O que parecia apenas um caso de polícia, eis que um dos assaltantes entende contar como são protegidos para efetuarem roubos frequentes sem, no entanto, serem identificados pela polícia.

25. Com Luanda já com transmissão comunitária, o que se pode esperar das imagens que vão ver a seguir é que todos os dias há pessoas a serem infetadas pelo novo coronavírus.
26. A ministra de estado, Carolina Cerqueira, disse hoje que o governo está a acompanhar o diferendo que existe na Igreja Universal do Reino de Deus em Angola e considera que existem situações que configuram crimes e estão a ser tratadas pelos tribunais.
27. Esta foi a posição que foi defendida pelo ministro da justiça e dos direitos humanos, depois do encontro, que analisou a crise na igreja universal, realizado entre membros do executivo e deputados da sétima comissão à assembleia nacional.
28. Sabemos que este corrimão é usado por milhares de pessoas que passam aqui diariamente.
29. Mais de 100 apartamentos de 480 existentes no bloco 8 foram ocupados por várias famílias.
30. Há casos com indícios de crimes que têm que ser tratados no fórum próprio e pelas entidades competentes.

Estruturas passivas eventivas curtas

1. É neste ambiente em que era produzido o pão que vai à mesa de muitas famílias em Luanda.
2. Em algumas dessas padarias, o pão era fabricado em locais com baratas, ratos mortos e muita sujidade.
3. Nesta reportagem, a TPA flagrou água imprópria para o consumo, baratas e ratos mortos em locais onde são preparados os pães.
4. Estamos a falar de 830 sacos de fuba da marca mamã fuba que estava a ser comercializada com gorgulhos – vulgo bichos – e também larvas.
5. A fuba em que os armazéns compraram aqui já compraram há muito tempo e como eles fizeram queixa no sentido de que a fuba compraram dentro do armazém então o produto mesmo lá já está meio fora do prazo lá nos armazéns então eles como foram fiscalizados lá, mas trouxeram tudo aqui.
6. Estamos a ser despejados.
7. Temos aqui um aviso de notificação que dá conta que nós seremos despejados dentro de 72 horas a contar com dia de ontem.
8. O grupo confessou ainda que recorreu a forças ocultas para efetuar os roubos sem que fossem apanhados.
9. O assalto ocorreu em plena madrugada neste estabelecimento, onde foram retirados 48 telefones.
10. Este jovem que comprou dois telefones também foi apanhado em flagrante com alguns objetos.
11. Parte deste gás é transformado em LPG e é disponibilizado ao mercado.
12. O plano de desenvolvimento nacional 2018-2020 foi lavrado na conjuntura dos anos 2017 e 2018 durante a sua elaboração previa-se já uma revisão intercalar em 2020 porque se avistavam mudanças na economia frutos diferentes choques externos que resultaram numa acentuada redução dos recursos financeiros.
13. Em cumprimento à demanda da revisão intercalar do PDN, foi feita uma adequação do plano à atual conjuntura macroeconómica nacional, tendo sido alterados alguns programas nos eixos do plano de desenvolvimento nacional.
14. O eixo dedicado ao desenvolvimento humano e bem-estar foram priorizados 23 programas dos 27 inicialmente previstos.
15. No eixo 2 dedicado ao desenvolvimento económico sustentável diversificado inclusivo, foram priorizados 20 programas dos 24 previstos inicialmente.
16. No eixo 5, o eixo das infraestruturas nacionais para o desenvolvimento, dos 11 programas inicialmente previstos foram priorizados apenas 10.
17. O eixo da garantia da estabilidade integridade territorial de Angola de reforço do seu papel no seu contexto internacional e regional será dado particular impacto aos 3 programas inicialmente previsto.
18. Uma outra referência a ser já implementada tem a ver com o desenvolvimento de uma aplicação informática capaz de recolher, tratar e sistematizar a informação para gerar relatórios do Balanço de execução dos projetos de atividades previstas no PDN.
19. Temos fundamentalmente a destacar o trabalho que está a ser feito, para que se consolide em Angola verdadeiramente um estado de direito.
20. [...] que foram estabelecidos.

21. Para o ministro de estado para coordenação económica é mais um órgão de apoio ao presidente da República que acaba de ser criado para dar resposta aos vários projetos que visam a melhoria da situação social e económica do país.
22. Temos de ser capazes de analisarmos a atuação que nos é colocada a nós, mas também a atuação dos investidores e a atuação de todos aqueles que intervêm de forma Apex no processo de captação e implementação dos projetos tenham que garantir que assim aconteça a luz dos acordos existentes.
23. Edgar Cunha era daqueles profissionais que, junto das pessoas mais novas, deixava os seus ensinamentos e, como profissional, emprestava a sua voz dos seus conhecimentos para que a informação fosse levada ao mais alto nível em todo o território.
24. [...] Somos a informar que até às 16:00 do dia 6 de agosto, foram notificados ao centro de processamento de dados de vigilância epidemiológica do Ministério da Saúde 88 casos com idades compreendidas entre os 2 anos e os 39 anos de idade; 45 do sexo masculino e 43 do sexo feminino, perfazendo um total cumulativo de 1483 casos. Desde o dia 21 de março temos a destacar hoje a proveniência de um caso do município do Libolo e outro município do Cubal. Os restantes ocorreram todos nos municípios de Luanda. Dos 899 casos ativos, 2 encontram-se em estado crítico. Infelizmente, hoje também temos a informar a ocorrência de 2 óbitos.
25. [...] Que medidas é que passa a ser tomadas a partir de agora? [...]
26. Eu fui a uma sessão depois na província do Kwanza Norte de Luanda através de caminhos fiotes no rio kwanza então toda precaução estão sendo tomadas para que essa situação não alarme na comunidade nem um pouco ao aqueles contatos dos contatos que ainda continua aqui contactos com outras pessoas.
27. [...] Significa que ela será transformado em unidade de internamento em tratamento.
28. [...] Profissionais da saúde concluem que o local era viável para ser transformado em centro do internamento em tratamento dos doentes infetados pelo novo coronavírus.
29. [...] Depois de terem sido cadastradas, as primeiras famílias da localidade já começaram a receber o seu dinheiro no âmbito do projeto de transferências sociais humanitárias.
30. São mais de 150 m. A obra vai ser realizada em toda sua extensão. O empreiteiro garante que seis meses é um tempo suficiente para conclusão da obra [...].
31. Nesse tipo de trabalho, tem sido retirado.
32. E agora uma nota da televisão pública de Angola que comunica que não será emitido o especial informação.
33. É bom que os angolanos saibam o que está sendo feito.
34. Está sendo feito um trabalho que também já foi partilhado com a comunicação social no sentido aumentar a nossa capacidade não só em Luanda, mas noutras províncias.
35. Por outro lado, também foi aqui questionado quantas pessoas vão entrar em quarentena domiciliar.
36. Já foram feitos vários voos humanitários.
37. Por outro lado, foi aqui colocada a questão das vacinas.
38. Também dizer que este acompanhamento vai ser feito a nível local, a nível do município, profissionais de saúde, os adecos, as comissões de moradores a administração municipal – todas estas entidades terão responsabilidades no seguimento desses casos[...].
39. É um doente o que é aceita com muita facilidade, que percebe as orientações que são dadas, se é cumpridores, se não é.
40. Você pode continuar a seguir as novas medidas no âmbito da situação de calamidade no canal 2 da televisão pública de Angola e, ainda neste telejornal, vamos recuperar um anúncio feito pelo Ministro de Estado e Chefes da Casa Civil do Presidente da República, Adão de Almeida, que o início desta noite deu a conhecer as novas regras que deverão ser cumpridas a partir de segunda-feira 10 de agosto.
41. Questionada pelo Público, Fátima Date disse que foi apanhada num turbilhão que não tem nada a esconder.
42. Um cidadão que se fazia passar por técnico da ENDE está a ser acusado de vandalizar os armários elétricos para fazer ligações anárquicas na centralidade 5 de abril na cidade de Moçâmedes.
43. Este cidadão, que se fazia passar por técnico da ENDE, é acusado de arrombar os quadros de instalação elétrica para fazer ligações anárquicas, solicitando dos moradores o pagamento à vista para repor a energia.
44. Cerca de três lojas na centralidade do horizonte do Cuíto província do Bié foram vandalizadas.

45. Os empreendimentos sociais também já funcionam, mas na zona comercial do projeto habitacional das mais de 100 lojas erguidas apenas dez foram adquiridas.
46. O falso técnico da ENDE será apresentado ao Ministério Público nos próximos dias.
47. As demais continuam a ser comercializadas no sistema de pronto pagamento com um custo de mais de 4000000 de kwanzas cada.
48. Algumas lojas começaram a ser vandalizadas.
49. Lhe foi agora confiada a tarefa de dirigir o partido e o governo na província de Luanda.
50. Toda atenção do nosso partido na província tem que ser dirigida aos nossos comitês de ação.
51. Quanto à quarentena, há um conjunto de alterações que mudam bastante o modo como a abordagem tem sido feita até agora.
52. Da avaliação feita em função de um conjunto de elementos e, naturalmente, sempre com recomendação e parecer positivo das autoridades sanitárias, entendeu-se também fazer uma alteração do modo de abordagem, quer dizer que os cidadão que tenham testado positivo em testes aos SASCOV-2 e que não manifestem sintomas passa a ser feita também uma abordagem domiciliar.
53. A comissão provincial de prevenção e combate à pandemia garante o controlo da situação, uma vez que o cidadão já foi encaminhado para o centro de tratamento.
54. Louvar a atitude que o cidadão teve no sentido de entrar em contato com as equipas de resposta rápida, para que fosse testado muito antes de entrar na sua residência.
55. Foi testado muito antes inclusive de descer da viatura e tomaram-se todas as medidas de prevenção da desinfestação da viatura e, neste momento, os outros dois cidadãos encontram-se em quarentena institucional.
56. Nesta primeira fase, estão indicados para gerir os casos suspeitos os cidadãos que violem a cerca sanitária e alguns profissionais da saúde que são expostos assim como os doentes com internados no hospital com doença respiratória aguda grave ou crónica que têm critério de realizar estes testes.
57. É reforçado o apelo no sentido de respeitarem as medidas de prevenção no interior de cada carruagem.
58. Já fomos atendidos.
59. Estamos a ser atendidos.
60. Mas, há aqueles a quem é levantada a suspeição por COVID-19.
61. A quem é levantada a suspeição, é feita a testagem, obviamente, editado internamento.
62. Aqui são realizados os testes RTPCR e serológico da COVID-19.
63. Um legado que não será desperdiçado.
64. Além de familiares e amigos, a cerimônia foi ainda presenciada por diplomatas da embaixada de Angola em Portugal e pela atual administradora da TPA para a área de marketing, Nadir Ferreira.
65. Mais de 40000000 de kwanzas foram disponibilizados para a construção de duas escolas do município do Mucari, em Malanje.
66. As condições de ensino também serão melhoradas.
67. As eleições na Federação Angolana de Basquetebol são realizadas no dia 22 de novembro.
68. Nós aprovamos calendário eleitoral, as eleições da Federação Angolana de Basquetebol serão realizadas no dia 22 de novembro de 2020.
69. Também foi aprovado o relatório de atividade de contas, depois de uma longa discussão com dezanove votos a favor e um voto contra.
70. Até que isso aconteça, só temos uma atitude a tomar: tudo fazer para evitar sermos contaminados ou contaminarmos os que nos rodeiam, numa palavra, prevenirmos.
71. Depois das mais de 197 pessoas terem sido testadas no bairro Estufa.
72. Os casos que deram reativos foram encaminhados para uma área de isolamento, visando uma nova testagem.
73. E a cerca sanitária dos bairros Kandjonge e Kaisaka no município do Bocoio em Benguela foi levantada após os exames conclusivos de duas pessoas identificadas como reativas que resultaram em negativo.
74. Hoje, foram confirmados 131.838 casos em todo mundo.
75. As novas regras dizem que é proibida a venda ambulante e nos mercados, mas o que vimos em Cacuo é um autêntico desrespeito às normas e um perigo para a saúde pública.

76. As vendedoras sabem que é proibida a venda ambulante às segundas, quartas e sextas, mas ainda assim não cumprem.
77. Quem tem obrigação de fazer cumprir as normas, muitas vezes, é acusado de furtar e apropriar-se dos bens de quem comercializa na via pública.
78. A seguir, os apartamentos da centralidade do Capari, no Bengo, continuam a ser invadidos.
79. para que seja dado um tratamento processual que é exigido por lei.
80. Os apartamentos do bloco 8 da centralidade do Capari continuam a ser invadidos e os cidadãos que já lá se encontram dizem que não saem porque não têm para onde ir.
81. Porque se o processo ou quando o processo chegar à fase inicial em que vão ser convidados a desocupar os imóveis, se as pessoas assim não procederem, vão incorrer num outro crime que é o crime de desobediência.
82. Naquela altura, a PGR havia dado 72 horas para que os imóveis fossem desocupados, mas a verdade é que a onda de invasão tem estado a crescer.
83. Dos dez apartamentos, sete continuam ocupados, enquanto os outros três já foram desocupados.
84. Desde quinta-feira última, sete pessoas acusadas de invasão à propriedade alheia e detidas na urbanização Vida Pacífica estão a ser julgadas na 17ª secção do tribunal provincial de Luanda.
85. A sentença será lida na próxima terça-feira às 10h.
86. Esses elementos estão a ser feita a perícia de como as pessoas tiveram acesso, mas também começou já ontem no tribunal de Viana o julgamento, elementos que foram encontrados no interior das residências no Zango Zero.
87. Vimos que muitas dessas habitações estão a ser vendidas no café.
88. Estão a ser vendida em ruas, becos, na internet.
89. É o bloco 8 que não foi comercializado por situações ainda de infraestruturas.
90. Em Bengala foi apreendido estas pessoas que invadiram o escritório da Imogestin e subtraíram algumas chaves e comercializavam essas chaves.
91. Porque as pessoas não têm legitimidade; essas pessoas acabam por também incorrer e passam a ser arguidas dos processos que estão a ser colocados nos tribunais.
92. Tudo fazer para evitar sermos contaminados ou contaminar os que nos rodeiam.
93. A TAAG lembra que esse mesmo comportamento já tinha sido observado aquando do voo humanitário de repatriamento dos 50 cidadãos angolanos mantidos na Turquia.
94. Duas fábricas de detergente foram encerradas hoje porque estavam funcionar sem alvará industrial.
95. Entretanto, a produção foi suspensa e a fábrica encerrada temporariamente.
96. Na zona económica especial, foram encerradas duas fábricas por várias irregularidades.
97. As propostas deverão ser apresentadas ao IGAP até agosto deste ano.
98. Há anos que o país aguarda pelo bum destas três indústrias têxteis onde foi delineado um investimento técnico tecnológico com denominadores comuns e que agora aguçam os novos investidores.
99. Os dois projetos foram arrolados e consequentemente devolvidos ao estado no âmbito da recuperação de empreendimentos construídos com fundos públicos.
100. O ato de devolução foi concluído no passado dia 30 de junho mediante a assinatura de um termo de entrega a favor da empresa de produção e eletricidade PRODEL EP.
101. Recorde-se que os empreendimentos em causa foram objeto da outorga para concessão por dia das resoluções números 108 e 69/9 do Conselho de Ministros de 23 de novembro de 2009 a favor da empresa Canazul Electric AS a quem foi atribuído o direito para reabilitar, explorar e expandir o negócio por um período de vinte anos.
102. O que é preciso é que as estradas sejam abertas, que as pessoas estiquem mais.
103. Um cidadão foi condenado a três meses de prisão efetiva por ter ficado provado ter sido responsável por queimadas no município do Libolo, que resultaram em danos, até agora não calculados, a diversas propriedades agrícolas e residências naquela região.
104. Fruto da ação da polícia, a sua motorizada foi devolvida.
105. Aqueles indivíduos que ainda teimarem em incomodar as populações, alterarem a ordem e a tranquilidade pública vão ser tratados na medida grau de perigo, ou seja, a medida da sua ameaça.
106. Também foram entregues uma picareta.

107. Os deputados da sétima comissão da assembleia nacional estão satisfeitos com o facto do governo não se envolver nos problemas internos na IURD. Os parlamentares defendem que estes problemas não devem ser levados para o nível político e diplomático por se tratarem de assuntos de fórum judicial, devendo serem resolvidos na esfera jurídica.
108. Vão ser remetidos.
109. Reforçam ainda que situações criminais devem ser resolvidas na esfera jurídica.
110. Helena Xan, porta-voz do centro comercial Cidade da China, garante que os salários estão salvaguardados e afirma que estão a ser cumpridas a rigor as medidas de prevenção contra a COVID-19.
111. Uma cidadã de Benguela violou a cerca sanitária em Luanda e foi testada positivo com coronavírus no Sumbe, capital do Kwanza-Sul.
112. Neste caso, apesar de ter sido detetado no Sumbe, não coloca a província do Kwanza-Sul na lista amarela da COVID-19.
113. Do ponto de vista laboratorial e por RTPCR foram testados 447 amostras das quais temos 18 positivas e 429 negativas.
114. Vamos olhar para os números de hoje, números confirmados 16.536.862. Só hoje foram confirmados 125.902 casos.
115. Ainda neste telejornal, o General Kundy Pahama será sepultado na quinta-feira no Kipungo.
116. Na reunião de hoje da comissão económica do conselho de ministros, foi também aprovada a estratégia de transição da economia informal para formal e que o 13º começa a ser pago já no próximo mês.
117. Paralelamente a esse trabalho, está sendo feito no domínio da importação porque nós ainda não produzimos os fertilizantes aqui dentro.
118. Também está sendo executado um trabalho junto, digamos, dos diferentes investidores, sejam eles nacionais como estrangeiros, convidando-os no sentido a montarmos cá dentro de Angola fábricas que possam produzir os fertilizantes.
119. Temos tido reuniões entre as duas partes para podermos, porque, como sabe, essa é uma zona de comunicação, uma zona em que fluência de povo de uma parte e de outra e essa comunicação precisa ser feita de forma segura.
120. Fomos obrigado também a aliviar algumas medidas por causa da economia que estava a ser sufocada, como exemplo, também para os países, mas estamos a criar medidas de criar os procedimentos para podermos tanto o alívio como a condição da vida das pessoas pudesse funcionar, mas guardando sempre aquilo que são as recomendações em termos gerais a nível das autoridades sanitárias.
121. Para materializar as políticas do executivo em matéria de direitos humanos de acordo com o programa de governação para 2017-2022 e o plano de desenvolvimento nacional 2018-2022 foi aprovada a estratégia nacional de direitos humanos que conta com a ampla participação da sociedade civil.
122. Com esse instrumento, pretende-se que Angola no gozo integral da sua soberania passe a exercer legitimamente e com rigor recomendado a avaliação periódica das atividades ligadas aos direitos humanos e é também no quadro dessa estratégia que os direitos humanos em Angola foram elevados à categoria da questão de segurança nacional.
123. Na quinta-feira, depois do culto e do velório no salão protocolar do governo da Huíla da cidade do Lubango, o corpo de Kundy Pahama segue para o município do Kipungo onde será sepultado na tarde do mesmo dia.
124. O valor deve agora ser adicionado à adoção do Ministério da Saúde no projeto de combate à malária.
125. A avenida Lueji Nkondi no distrito do Sambizanga voltou a ser iluminada vinte anos depois de muita escuridão.
126. Essas são imagens de pacientes que estão a ser ventilados na clínica Girassol.
127. Baleia que veio à costa da praia morena já foi removida.
128. Uma brigada do patrulhamento composta por efetivos da polícia nacional e das empresas de segurança instaladas na região de Capanda será criada nos próximos dias com vista a prevenção e o combate às queimadas anárquicas nas áreas de produção da companhia de biocombustíveis de Angola, Biocom.
129. Só no ano passado foram registadas mais de 900000 queimadas anárquicas no país afetando a flora e a fauna.
130. O grande problema é que essa cana é queimada fora de toda a programação da nossa produção.

131. Tiram os cabos de aterramento das torres, porque são feitos de cobre.
132. Este da Lunda Norte violou a cerca sanitária de Luanda e foi encontrado quando menos esperava naquela província.
133. O combate deve ser encerrado em todos os locais.
134. É nas unidades de cuidados intensivos onde são internados os doentes graves com o novo coronavírus.
135. A paramentação que nós utilizamos aqui no centro é do nível 3, sendo que a permanência dentro desses fatos é recomendada até uma hora.
136. Diariamente, é feita a visita médica em 3 tempos.
137. São medicados com terapia específica e ficam sobre nossa vigilância diária.
138. Só hoje, foram confirmados 161924 casos.
139. Daí que todas as questões estão a ser levadas em conta.
140. As unidades sanitárias principal da província há um espaço reservado com 12 ventiladores para eventuais casos positivos e informações recolhidas junto da comissão técnica multisectorial dão conta de que este paciente pela das condições que apresenta não será então necessário que seja transferido para Luanda, porque os técnicos estão preparados para dar resposta necessária ao seu tratamento.
141. Este contrato visa potencialmente apoiar as micro, pequenas e médias empresas nos setores que já foram referenciados nomeadamente o setor de agronegócio dentro da sua cadeia desde a produção de consumo até a transformação do produto final.
142. Das atividades de atenção estão desenhadas no contrato de modo a... e os critérios de elegibilidade vão ser desenhados de acordo com cada atividade.
143. A nova nota do kwanza é que vai ser afirmada aqui no mercado.
144. Apesar do assunto do momento ser a entrada em circulação da nota de 200 kwanzas, nos mercados em dias de vendas, a aglomeração de pessoas continua a ser uma preocupação por que as medidas de prevenção a COVID-19 em muitos casos passam despercebidas.
145. Também esclarecemos aos deputados que em sede da aprovação do código penal do código de processo penal, digo, estão asseguradas as condições necessárias para que em Angola o exercício do... ou os crimes praticados por via das telecomunicações e tecnologia de informação possam ser resolvidos porque este é o código penal e código de processo penal abordam aspetos relacionados com a matéria.
146. Nos últimos dois anos, foram registados em Angola mais de cem casos de tentativa de tráfico de seres humanos dos quais vinte pessoas foram julgadas e condenadas.
147. Em termos do tráfico, nós somos um país de origem, quer dizer que há angolanos que saem daqui e são traficados fora de Angola.
148. A maior parte é traficada para exploração sexual, seguindo o trabalho forçado e remoção de órgãos.
149. É concedida aqui por duas cidadãs, duas voluntárias.
150. As duas conhecem pessoas na SIC, são pessoas muito bem relacionadas e que foram convidadas pela SIC para apurar efetivamente a informação que foi veiculada nos órgãos de informação.
151. E já foi removida a baleia de 7 m de comprimentos que apareceu ontem morta na orla marítima da praia morena em Benguela.

Estruturas passivas estativas

1. Essas pessoas ocuparam residências que já estão comercializadas pelas Imogestin e que os beneficiados chegam até nós reclamando essas habitações.
2. Mal porque as autoridades sanitárias dizem que está controlado esse caso positivo que foi anunciado por mais uma atualização.
3. Mesmo sem estar autorizada em Portugal, Fátima Date já tinha um representante em Angola.
4. Em relação aos óbitos, tratam-se de dois indivíduos do sexo feminino, com idade de 58 e 64 anos, com cormobilidade descompensadas e que estavam internadas na clínica girassol em estado crítico.
5. E agora que está confirmada praticamente a transmissão comunitária do vírus em Luanda e se você gosta de se juntar a grupos.
6. Saiba que uma das pessoas no grupo, quando você estiver, pode estar infetada com o novo coronavírus, depois o vírus passa para si quando chegar à casa transmite a sua família.
7. Estavam armados com uma star.
8. Durante as discussões do orçamento geral do estado na assembleia nacional, para o projeto estavam reservados os cerca de 745 milhões de kwanzas.
9. São agendado para aí acho que sexta-feira a direção de informação da TV a lamentar os possíveis constrangimentos.
10. O falso técnico da ENDE já está detido.
11. Agora está levantada a cerca sanitária na mina do Tchinguvo.
12. É possível de facto ver por estas imagens que Luanda ainda não está completamente recolhida pelo menos nesta zona de Cacucaco.
13. A LIMA está reunida aqui em Luanda.
14. Então, neste preciso momento, as nossas equipas estão deslocadas no município do Libolo no sentido de se criar a cerca sanitária a este indivíduo.
15. Estão conservados os cerca de 745 milhões de kwanzas.

Anexo IV – Estruturas passivas do *Telejornal de Portugal*

Passivas eventivas longas

1. Está a ser combatido por mais de 500 bombeiros e vários meios aéreos, aviões e helicópteros.
2. Isso significa que esta frente está agora a lavar sem controle, apesar de ter sido combatida por muitos meios terrestres, mas também por muitos meios aéreos sobretudo meios aéreos pesados – dois aviões cannondale e quatro feibox, que, durante a tarde, várias horas despejaram toneladas de água em cima desta floresta.
3. O receio de que tal pudesse acontecer tinha sido assumido pela polícia que acredita no envolvimento de Brueckner no desaparecimento de Madeleine e noutros crimes, mas que diz não ter evidências inequívocas.
4. Resulta dos processos que as crianças viveram com a mãe numa instituição e, desde que deixaram de lá viver devido a incapacidade da mãe, passaram a viver em sucessivas AMAS após o que ingressam no lar universal, sendo que o ingresso no lar universal foi determinado pela segurança social ao contrário do que é afirmado na reportagem.
5. Os pais eram toxicodependentes, viviam numa barraca e usavam as crianças para a mendicidade, tendo sido retiradas aos pais pela própria segurança social.
6. A informação foi confirmada à TVI pela Procuradoria Geral da República.
7. Há vários meses que a rede estava a ser investigada pelo Departamento Central de Investigação e Ação Penal.
8. Fomos literalmente proibidos pela Guarda Nacional Republicana.
9. O da venda do banco e também o que foi assinado pelo fundo de resolução em 2017.
10. A revelação foi feita pela câmara de Arouca.
11. Depois de ter sido considerada pela European Best.
12. A última vez foi no Porto, onde logo a seguir na inauguração da feira do livro, Marcelo Rebelo de Sousa foi confrontado por uma cidadã muito descontente com a situação do país.
13. Está em causa uma carteira de ativos no valor de 250 milhões de euros, cuja venda terá sido suspensa por determinação do Ministério das Finanças.
14. A Alantra é representada por Ana Rita Barosa ex-assessora de Ricardo Salgado, visada no processo ex-entrepresa conhecida como o saco azul do BES.
15. Foram resgatadas pelo navio pago pelo artista anónimo mais famoso do mundo.
16. A situação tem tudo para ser mediática já que esta embarcação é financiada pelo famoso artista de rua britânico Banksy, cuja identidade continua por revelar.
17. Há, neste momento, dois grandes incêndios a serem combatidos por cerca de 600 bombeiros.
18. De recordar que a estação de São Bento já foi considerada uma das mais bonitas do mundo por várias revistas e organismos internacionais.
19. Ele pertence ao Museu de Lisboa e foi dado pelo rei Luís Filipe que era padrinho do nosso rei D. Luís.
20. O trabalho é feito por historiadores investigadores de história contemporânea e deverá estar concluído até ao fim do ano.
21. Estão agora a ser biografados pela Assembleia da República.
22. A vindima é feita à mão e escrupulosamente acompanhada por Ryan e pelo adegueiro João Alabaça, assegurando a qualidade das uvas antes de chegarem aos lagares de pedra, tudo feito à moda antiga e no tempo certo que as coisas têm.
23. Um novo relatório confirma as falhas, que foram tratadas pela ordem dos Médicos no lar de Reguengos de Monsaraz onde morreram 18 idosos.
24. Não havia equipas distintas de cuidadores para infetados e os não infetados, graves problemas que já tinham sido denunciados no dia anterior pelos médicos de família do centro de saúde de Reguengos.

25. Costa voltou ainda a defender que a ministra do trabalho e solidariedade não desvalorizou a situação em Reguengos, o que é demonstrado pela forma como atuaram.
26. Era o desejo da família e foi atendido pelos médicos do hospital na Sibéria.
27. Várias regiões do hemisfério sul estão a ser assoladas por tempestades de neve.
28. O acidente vai ser agora investigado pelo gabinete de acidentes com aeronaves e acidentes ferroviários.
29. A campanha é suportada pela câmara municipal que aos restaurantes dão o apoio de 7 euros e meio por criança.

Passivas eventivas curtas

1. A chamusca chegaram a ser dominadas durante a manhã.
2. Está a ser equacionada a hipótese de evacuar o Couto da Panasqueira e Cambões.
3. O incêndio foi dominado nas primeiras horas da manhã.
4. No passado dia 21 de dezembro de 2017, no telejornal das 20h da TVI e da TVI24 e no debate que lhe seguiu, foram feitas referências a vários processos de adoção que mais uma vez foram apelidados de ilegais.
5. Se de manhã um incêndio em Sobral de São Miguel foi dado combinado, ao início da tarde vários reacendimentos fizeram temer o pior.
6. O bombeiro dos voluntários de Cuba sofreu queimaduras graves e foi transportado, neste mesmo dia de helicóptero para o hospital.
7. Trump argumenta que o sistema de voto por correio, que está a ser planeado como medida de segurança devido a pandemia, irá conduzir a uma fraude em massa.
8. O voto por correspondência está a ser ponderado como forma de evitar mais contágio de COVID-19.
9. Nas últimas horas, foi ultrapassada a marca negra de 150.000 mortos, mais de 4.400.000 infetados.
10. Foram ultrapassadas as 90.000 mortes.
11. O sindicato dos médicos estima que mais de 1800 profissionais tenham sido afetados.
12. Nesta quinta-feira, foram confirmados seis novos casos no bairro, coisa pouca no dia em que a Índia bate um novo recorde.
13. As medidas vão sendo reforçadas de dia para dia.
14. O ministro britânico dos transportes estava de férias e acabou por ser apanhado durante a implementação da quarentena obrigatória no regresso à casa.
15. Em alguns hospitais, foram novamente canceladas as consultas e cirurgias para se prepararem para este novo aumento de casos.
16. Qualquer cidadão português que chegue ao país por via aérea é obrigado a fazer teste a COVID-19.
17. Se for estrangeiro e não quiser fazer o teste, será obrigado a regressar ao país de onde veio.
18. O teste deve ser feito na origem.
19. Quem tiver 38° ou mais será direcionado para estruturas de apoio sanitário no aeroporto.
20. Essa decisão é tomada principalmente, tendo em conta uma evolução positiva que se tem verificado no último mês relativamente a este território com uma redução na última semana de cerca de 30% dos casos em toda Lisboa e Vale do Tejo.
21. Passam a ser permitidos, por exemplo, ajuntamentos até dez pessoas.
22. É levantado o dever cívico de confinamento.
23. Mas, onde é também clarificado que o encerramento desses mesmos espaços tem que se verificar a uma da manhã.
24. De situação de calamidade para estado de contingência, mas as normas têm de ser respeitadas.
25. Acho que as mudanças foram mal feitas.
26. Era sabido que idade é importante.

27. A produção de hidrogénio da próxima década representa investimento na ordem dos sete mil a nove mil milhões de euros e poderá ser acompanhada de uma redução da metade das importações de gás natural.
28. No passado dia 21 de dezembro de 2017, no telejornal das 20h da TVI e da TVI24 e no debate que lhe seguiu, foram feitas referências a vários processos de adoção que mais uma vez foram apelidados de ilegais.
29. Mais uma vez, é afirmado que os processos de adoção integram o esquema de adoções ilegais que a IURDE, Igreja Universal do Reino de Deus, tem vindo a ser acusada.
30. Na referida reportagem, é apresentado o caso de adoção das menores Cristela Daniela Reis e Daniela Cristela Reis, sendo mesmo qualificado como um processo de adoção ilegal.
31. Ao contrário do que é afirmado na reportagem emitida, as crianças começaram apenas a frequentar a casa de uma família meses depois de ingressarem no lar, tendo por decisão judicial e mesmo após ponderado o pedido de guarda da avó paterna sendo a sua guarda confiada à família com o qual já tinham relação de afeto.
32. A decisão proferida foi anulada, tenho prosseguido o processo de promoção e proteção das menores que por decisão do Tribunal de Família e Menores de Lisboa foi remetido para o Tribunal do Porto e, posteriormente, para o Tribunal de Vila Nova de Gaia, culminando com a decisão judicial que decretou a adoção das menores.
33. A propósito deste processo, é afirmado que os pais biológicos das crianças foram enganados e que, quando quiseram recuperar os filhos, já não foi possível e ainda que o processo teve por base relatórios falsos.
34. Está feito para ser usufruído também pelos invisuais.
35. No caso destas crianças e dado o tempo prolongado de institucionalização, foi o próprio tribunal que determinou que deveria ser instaurada a respetiva ação de confiança judicial com vista a futura adoção, tendo esta ação sido instaurada na sequência de tal determinação e os menores confiados ao lar.
36. Os pais de Matilde criaram em junho de 2019 uma conta solidária, onde foram depositados dois milhões e meio de euros.
37. Através das redes sociais, os pais disseram que a verba não seria devolvida para ajudar outras crianças com problemas de saúde e, considerando que o dinheiro pertencia à filha, pretendiam usá-lo para todos os tratamentos, terapias e equipamentos necessários para o desenvolvimento da bebé.
38. A polícia judiciária apreendeu agora perto de 7 t de haxixe numa operação em que foram detidos nove homens, todos estrangeiros.
39. Uma parte dos factos estava no armazém outra acondicionada em viaturas para ser transportada para fora do país.
40. Além do haxixe, também foram apreendidas cinco viaturas.
41. O dinheiro tinha acabado de ser levantado do banco, o que indicia que os assaltantes conheciam a rotina.
42. A polícia foi rapidamente chamada ao local e colocou meios no terreno, mas não conseguiu localizar os responsáveis.
43. De facto, não foi aberto o local porque, na altura, não havia já incêndio no local.
44. Aqui houve meio milhão de visitantes em 2017, o que é um número absolutamente extraordinário sobretudo se pensarmos que foi construído no meio do nada.
45. A estrutura que vê durou 150 anos a ser construída.
46. Qualquer grande dignatário, embaixador obrigatoriamente tinha que passar pela Batalha para ser vislumbrado o grande empreendimento que a coroa tinha em mãos.
47. Por que é que isso não foi acabado?
48. Realmente, não foi acabado.
49. Recursos foram canalizados.
50. Se tivesse sido terminado, falta aqui a abóbada.

51. Desenvolviam-se porque os mosteiros cistercienses recebiam senhoras de casas senhoriais e elas levavam as receitas, quando eram colhidas.
52. É desenvolvida nos conventos.
53. Vendido a particular, o que não caiu foi saqueado.
54. Se não forem feitos no tacho de cobre, não é a mesma coisa, nem sai da mesma maneira.
55. Numa altura em que COVID-19 continua a condicionar a vida de qualquer pessoa, o secretário geral das Nações Unidas, António Guterres, defende que a vacina, quando surgir, tem que ser considerado um bem público.
56. Sempre defendi que qualquer tratamento ou qualquer vacina que seja descoberta para o COVID-19 tem que ser considerada ou considerado um bem público mundial.
57. O tribunal de contas acredita que o ministério das finanças não esclarece estas questões porque ele próprio também não sabe a resposta e é isso mesmo que escreve neste relatório que foi conhecido na quarta-feira à noite.
58. O Tribunal de Contas quis saber também se estão a ser cumpridos os acordos.
59. Não foi entregue toda a documentação sobre a operação de venda de 75% do capital social do Novo Banco constata-se, assim, que o Ministério das Finanças não dispõe de toda a informação sobre essa superação.
60. Passando por Agar do Oriente, Moscavide e Portela será ainda criado um corredor em toda a extensão da freguesia do Parque das Nações.
61. O Vitória de Setúbal e o Desportivo das Aves foram despromovidos para o campeonato de Portugal, por não cumprirem os requisitos de inscrição nas provas profissionais.
62. A decisão não será aplicada até que haja a decisão final do Conselho de Justiça sobre esta matéria.
63. À custa do Vitória de Setúbal, o emblema algarvio pode manter a posição, apesar de ter sido despromovido.
64. Acho que, se for feita justiça, e assim deverá ser o Vitória de Setúbal sempre em incumprimento.
65. E é com perseverança que este veículo robótico vai recolher amostra de solo e rochas de Marte, que serão enviadas para a terra.
66. São 482 milhões de quilómetros, que irão ser percorridos em sete meses.
67. Apenas deveria ser mantido o distanciamento social.
68. As mesas com distanciamento social são o primeiro passo do caminho de regresso à escola, numa altura que o governo não consegue chegar a um consenso sobre as medidas a serem implementadas nas salas de aula.
69. O primeiro caso foi identificado num restaurante e o vírus aparenta ter alastrado na comunidade.
70. É natural e é provável que nos próximos dias, em função do número de testes que está a ser realizado, estes números possam ainda ter alguma evolução relativamente rápida.
71. Ao primeiro sintoma contactem as autoridades de saúde e façam uma pequena quarentena até serem vigiados e que sobretudo continuem a reforçar o uso das máscaras.
72. Será analisada em função do momento e da necessidade.
73. A segunda cadeia de contágio foi detetada nos bombeiros voluntários de Arouca que tem nove casos confirmados de COVID-19 e trinta pessoas em quarentena.
74. A praia da Comporta foi considerada uma das 10 mais seguras da Europa.
75. Em resposta às imagens polémicas que têm sido partilhada e que mostram uma fila de dezenas de pessoas no aeroporto, o SEF garante que essa semana reforçou o efetivo com mais doze inspetores e serão mais cinco a partir de setembro.
76. Na entrevista ao jornal expresso, que só foi concluída na edição de hoje, António Costa extrema o discurso e avisa a esquerda que se não houver acordo para o orçamento de Estado, então o desfecho será mesmo uma crise política e eleições antecipadas.
77. Uma semana depois do episódio entre o Costa e a Ordem dos Médicos, esta semana o vídeo da entrevista não foi divulgado.
78. Pede a palavra e, quando raramente lhe é concedida, fala de escolhas.
79. A gestão dos bens foi entregue à empresa Parva Lourent, criada para o efeito.

80. Foi comprado por cerca de 12 milhões de euros em leilão.
81. Tinha sido condenada a 13 anos de prisão acusada de múltiplos crimes ligados a uma organização terrorista marxista.
82. Foi condenada a partir de uma testemunha anónima que viria depois a pedir anulação das suas declarações devido ao seu estado de perturbação mental.
83. Este é um ano diferente como mostrou os turistas, mas por aqui foram adotadas medidas que pretendem ser um exemplo global de responsabilidade cívica.
84. As modalidades são divididas por grau de perigo devido à possível infeção.
85. Nas atividades desportivas, o que se percebe é que, quando surge um foco, ele tem tendência para alastrar se não for contrariado com medidas eficazes e rápidas.
86. Foi testado e o resultado positivo ficou conhecido.
87. Inclusive a esposa de um atleta trabalhava no mar e também já foi testada.
88. A origem do foco ainda é desconhecida, mas o clube garante que os reforços da equipa, que começaram a treinar no passado dia sete, foram todos testados.
89. Geraram-se muitas dúvidas após terem sido conhecidas as orientações da Direção Geral da Saúde, para o regresso aos treinos e às competições.
90. Os testes só serão realizados de forma aleatória, quando estejam envolvidas equipas de zonas com transmissão comunitária ativa.
91. Para além das normas básicas de desinfeção, a testagem começou a ser feita há dois meses.
92. Muito antes de este desporto ser considerado de alto risco.
93. O hóquei em patins foi colocado no risco médio.
94. Tal como em tantas modalidades, o andebol é enquadrado no risco médio.
95. Foi oferecido e foi montado no Palácio das Necessidades.
96. E, quando a monarquia caiu, foram transformados para o serviço do Chefe de Estado.
97. Foram escolhidas onze personalidades.
98. Foram escolhidos onze de quatro forças partidárias.
99. A tarefa foi entregue a historiadores e apenas um nome dificuldade em conseguir encontrar quem o quisesse biografar, Emílio Guerreiro.
100. Eles são presos com os filhos que vão com as mulheres que vão para a cadeia também.
101. A vida do poeta do desassossego é contada agora de uma forma diferente, para que o visitante procure compreender o peculiar processo criativo de um dos maiores escritores portugueses de sempre.
102. Já me foram relatadas também todas as situações.
103. É sobre um homem que aceitou ser preso por um crime que não cometeu.
104. Eu sou testada todas as semanas.
105. Como tem sido feito.
106. Desde logo, nas viagens com a utilização de máscaras, que também é mantida nas mais pequenas.
107. A adega foi construída em 1901 e, praticamente, está tudo igual desde então.
108. Foi engarrafado apenas o ano passado.
109. São pisadas durante algum tempo até que ocorra a forma alcoólica.
110. Estes dez ou onze anos foram realmente muito difíceis para o Mouchão no sentido de que as vinhas foram abandonadas, literalmente abandonadas.
111. Quando foi devolvido ao meu avô em 86, praticamente não se podia abrir as portas.
112. Sejam bem vindos à cidade de Tomar, nas margens do rio Nabão enquadrada pelo convento de cristo, património mundial, cuja forma octogonal inspirou o escritor Umberto Eco no seu “Pêndulo de Foucault”, ponto de passagem incontornável para os apaixonados da lenda dos templários, cujo tesouro terá sido escondido dentro daquelas muralhas.
113. O jornal Público teve acesso ao documento onde a enfermeira encarregue pelo delegado de saúde do Alentejo alerta que, quando visita a instituição, os idosos infetados estavam ainda a ser separados dos que não tinham a doença.

114. Em entrevista ao “Jornal Expresso”, o primeiro ministro diz que não há dúvidas que houve falhas no lar privado, mas que o estado, quando foi alertado, reagiu de imediato.
115. Portanto, aguardemos serenamente que o inquérito por parte do ministério público seja feito para ver se tem relevância criminal.
116. O surto na vila de Moura foi detetado a 9 agosto.
117. Costa que apurasse responsabilidades políticas pelos erros que foram cometidos e que apontasse a porta da saída à sua ministra e que, no mínimo, pedisse desculpas às famílias que perderam entes queridos por falhas grosseiras do Estado.
118. Os bloquistas exigem também um ponto final do dinheiro que é dado ao Novo Banco.
119. Não é possível nós começarmos a negociar o mais próximo orçamento de Estado, sem o Partido Socialista cumprir, se quer, o que já foi acordado para o último orçamento de Estado.
120. O secretário geral do PCP garante que vão ser tomadas todas as medidas de proteção necessárias na festa do Avante.
121. Ainda é cedo para considerar que a normalidade no fluxo de turistas neste aeroporto foi restabelecida.
122. Foi ultrapassada a barreira dos 800.000 mortos por COVID-19 no mundo inteiro.
123. Marcharam pelos mortos que acreditam poderiam ter sido evitados.
124. No país, mais de 70.000 profissionais de saúde foram infetados.
125. Os descobrimentos obviamente foram pensados aqui em Tomar.
126. As chamas e as espessas nuvens de fumo podem ser vistas a muitos quilómetros de distância.
127. Estão a ser usados meios aéreos no combate aos incêndios.
128. Já chegou a Berlim o político russo Alexei Navalny, que suspeita ter sido envenenado.
129. Mas, pelo menos agora foi dado o primeiro passo."
130. Opositor e crítico do regime de Vladimir Putin, Alexei Navalny é também um ativista anticorrupção e foi detido várias vezes por organizar protestos e investigações incómodas para o presidente russo.
131. Ainda assim, foi encaminhada para o hospital.
132. Doze drones foram comprados por ordem do Governo por 4 milhões e meio de euros com o objetivo de serem utilizados durante a época dos incêndios, mas ainda não estão a funcionar.
133. Segundo o jornal Público, há dois anos e meio que as autoridades estão a avaliar se devem ou não ser reparados.
134. Na ilha de São Miguel nos Açores, foi lançada uma campanha para atrair novos clientes aos restaurantes do concelho da Ribeira Grande.
135. Isto poderia ser alargado ao dia, tanto ao almoço e o jantar.
136. Neste momento, já foram atribuídos cerca de 700 vales.
137. Quanto ao concelho da Ribeira Grande, vai ser estendida até ao final de setembro.
138. Desenhado pela arquiteta Madalena Caiado, o projeto onde serão investidos dois milhões de euros já saiu do papel.
139. Recuperado recentemente no momento em que foi obrigado a encerrar por causa da atual pandemia.
140. Estás mesmo a levar com o sino em cima aqui junto a beira do Nabão, nesta parte da cidade que foi construída no tempo do infante D. Henrique na altura em que ele preparava os Descobrimentos.
141. É sabido e é público.
142. E, depois, a resposta da senhora diretora geral de saúde, que já está entre nós a ser muito fotografada, e bem.
143. As gravações passam a ser feitas todas em Portugal.
144. A quarta temporada está a ser gravada em Portugal.
145. Para ser entregue para ajuda de mãe e cá estamos.
146. Eu, em 1949, fui convidado para vir trabalhar com ela.
147. Vinhos que são cada vez mais reconhecidos e que vão ano após ano conquistando lugares entre os melhores.

148. Acreditamos que normalmente que situações semelhantes entram em Portugal por via marítima e depois são distribuídas pelo resto da Europa.
149. A instituição foi desinfetada este sábado e procedeu-se também a separação dos utentes infetados e não infetados.
150. Dois dias depois de internamento, foi dada luz verde para que dissidente russo, Alexei Navalny, fosse transferido para fora da Rússia.
151. Perdeu os sentidos depois de ter bebido um chá que se suspeita ter sido envenenado.

Passivas estativas

1. Comunicamos à saúde a situação e penso que a situação está controlada.
2. Eu diria que é impossível porque efetivamente já está assumido.
3. Só que estão muito escondidos aqui dentro.
4. Quem nos visita pode ver que elas estão organizadas por ordem cronológica.
5. Há pessoas que tinham escolhido Espanha e a Croácia, por exemplo, que passaram a estar condicionadas no regresso e que agora num desvio de fluxo que já aconteceu no sentido contrário, quando nós éramos o país que estava condicionado e que estão a correr ao Algarve.
6. Simplesmente estão apresentados de outra forma que é a moda da nossa casa da taberna da talha.
7. Depois disso, o lagar está aberto.
8. A Herdade do Mouchão esteve ocupada durante dez anos.