

2º CICLO DE ESTUDOS
MESTRADO EM LINGUÍSTICA

**Desenvolvimento de um sistema de conversão
grafema-fone[ma] e de translineação para o
Vocabulário Fundamental da Academia das
Ciências de Lisboa**

Leonor de Oliveira Martins

Junho

2025



Leonor de Oliveira Martins

Desenvolvimento de um sistema de conversão grafema-fone[ma] e de translineação para o *Vocabulário Fundamental* da Academia das Ciências de Lisboa

Dissertação realizada no âmbito do Mestrado em Linguística, orientada pela Professora Doutora Ana Maria de Castro Faria Salgado, pelo Professor Doutor Carlos Rogério Sousa e Silva e pela Professora Doutora Maria de Fátima Henriques da Silva

Faculdade de Letras da Universidade do Porto

2025



Este trabalho está licenciado ao abrigo de uma licença CC BY 4.0

“Chega mais perto e contempla as palavras

Cada uma

tem mil faces secretas sob a face neutra

e te pergunta, sem interesse pela resposta,

pobre ou terrível, que lhe deres:

Trouxeste a chave?”

Carlos Drummond de Andrade

Procura da Poesia, 2016

Sumário

Declaração de honra	7
Agradecimentos	8
Resumo.....	10
Abstract	11
Índice de figuras	12
Índice de tabelas	13
Lista de abreviaturas e siglas.....	16
Introdução.....	17
1. Fonologia, fonética e ortografia	20
1.1. Níveis de representação fonológicos e fonéticos	20
1.2. Fonologia do português europeu.....	23
1.2.1. Segmentos consonânticos	25
1.2.1.1. Alternâncias fonético-fonológicas	29
1.2.1.2. Variações na norma-padrão.....	30
1.2.2. Segmentos vocálicos	33
1.2.2.1. Processos fonológicos e variação.....	37
1.2.3. Sílabas	40
1.2.4. Acento	45
1.3. Interface fonologia-ortografia.....	48
1.3.1. Relação entre grafema e fone[ma]	48
1.3.2. Sistemas de representação	50
2. Tratamento lexicográfico da pronúncia nos dicionários de língua portuguesa.....	52
2.1. Dicionários impressos de língua portuguesa	52
2.2. Dicionários digitais de língua portuguesa	57
2.3. O <i>Dicionário da Língua Portuguesa</i> da Academia das Ciências de Lisboa	59
3. Sistemas de conversão	61
3.1. <i>Grafone</i>	62
3.2. <i>Convert Text to IPA Transcription</i>	63
4. Metodologia para o desenvolvimento do sistema de conversão	64
4.1. Elaboração dos critérios de conversão	65

4.1.1. Translineação	65
4.1.2. Marcação da sílaba tónica.....	66
4.1.3. Transcrição fonética.....	66
4.2. Desenvolvimento do sistema de conversão	70
4.2.1. Constituição do <i>corpus</i>	70
4.2.2. Instrumentos.....	71
4.2.3. Procedimentos	73
4.2.3.1. Translineação	75
4.2.3.2. Marcação da sílaba tónica	82
4.2.3.2.1. Palavras acentuadas lexicalmente	82
4.2.3.2.2. Palavras oxítonas	85
4.2.3.2.3. Palavras paroxítonas	90
4.2.3.3. Transcrição fonética	93
4.2.3.4. Importação para o DLP-ACL (2025).....	108
5. Análise e discussão dos resultados	109
5.1. Translineação	109
5.2. Marcação da sílaba tónica.....	111
5.3. Transcrição fonética	112
5.3.1. Consoantes.....	112
5.3.2. Vogais átonas	115
5.3.2.1. Vogais pretónicas	115
5.3.2.2. Vogais em início absoluto de palavra.....	117
5.3.3. Vogais tónicas	118
5.3.3.1. Dissimilação das palatais.....	118
5.3.3.2. Exceções.....	119
5.4. Integração no DLP-ACL (2025).....	119
6. Conclusões.....	123
Referências bibliográficas	128
Anexos.....	141
Anexo 1 – Dicionários impressos de língua portuguesa representativos dos sécs. XVIII e XIX.	142
Anexo 2 – Dicionários impressos de língua portuguesa representativos do séc. XX.....	143
Anexo 3 – Dicionários impressos de língua portuguesa representativos do séc. XXI.....	144
Anexo 4 – Dicionários digitais de língua portuguesa	145

Anexo 5 – Edições do <i>Dicionário da Língua Portuguesa</i> da Academia das Ciências de Lisboa.	146
Anexo 6 – Resultados das <i>regex</i> de translineação, marcação da sílaba tónica e transcrição fonética após correção manual.....	147
Anexo 7 – Critérios da translineação no guia de uso	166
Anexo 8 – Critérios da transcrição fonética no guia de uso.....	167
Anexo 9 – Inventário de caracteres consonânticos no guia de uso	168
Anexo 10 – Inventário de caracteres vocálicos orais no guia de uso.....	171
Anexo 11 – Inventário de caracteres vocálicos nasais no guia de uso.....	173
Anexo 12 – Inventário de ditongos orais no guia de uso	174
Anexo 13 – Inventário de ditongos nasais	175

Declaração de honra

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

Declaro, ainda, que não utilizei ferramentas de inteligência artificial generativa (*chatbots* baseados em grandes modelos de linguagem) para realização de parte(s) da presente dissertação, encontrando-se todas as interações (*prompts* e respostas) transcritas em anexo.

Porto, 30 de junho de 2025

Leonor de Oliveira Martins

Agradecimentos

Agradeço, primeiramente, a quem me ajudou a tornar este trabalho realidade.

À Professora Ana, exemplo de perseverança e bondade, a quem devo tanto. Muito obrigada, professora, por me ter deixado emergir neste mundo tão bonito que é a lexicografia, por acreditar que consigo sempre mais e melhor e, acima de tudo, por estar sempre lá.

Ao Professor Carlos, pelas palavras de alento nos momentos mais difíceis. Muito obrigada por todo o cuidado, atenção e disponibilidade – mesmo às horas mais inaceitáveis do dia!

À professora Fátima, por ter plantado em mim a semente da Linguística e por constantemente regar a flor que dela nasceu. Muito obrigada pela confiança, pelo encorajamento, por tudo.

Agradeço igualmente à Academia das Ciências de Lisboa, instituição que me acolheu. Nela aprendi que se não é útil o que fazemos, vã é a glória.

Agradeço, principalmente, à minha família.

À minha mãe, que sempre me encorajou a ser quem eu quisesse ser. Ao longo destes vinte e três anos, enquanto me guiava, não sei se alguma vez se apercebeu de que a pessoa em quem eu mais me queria tornar era ela. Obrigada, mãe, és o meu farol.

Ao meu pai, que me deu a cor dos seus olhos e que festeja as minhas conquistas mais fervorosamente do que toda a gente. Espero um dia ter metade da bondade dele. Obrigada, pai, és a jangada que me leva a bom porto.

Agradeço ao André, par de vida e de viagem, pelo apoio incansável. Agradeço todas as horas gastas em frente ao ecrã, todos os conselhos nos momentos de maior indecisão, todos os abraços quando mais necessitei. Estar-te-ei eternamente grata pela ajuda, paciência, atenção e cumplicidade. Obrigada, André, és a mão que se estende e me ampara.

Agradeço também à minha colega e amiga, Maria Leonor Reis, por escutar as minhas lamúrias e por partilhar comigo as suas. Por vezes, o que importa não é o destino, mas sim as pessoas que encontramos pelo caminho.

Agradeço igualmente à minha querida amiga, Ana Luísa Fernandes, companheira de percurso e quem admiro muito. Muito obrigada, amiga, por me ajudares sempre a encontrar o meu rumo.

Agradeço ainda às minhas amigas, Leonor Silva e Inês Santos, que, durante uma conversa de café, me encorajaram a enveredar no Mestrado em Linguística. Acreditaram em mim muito antes de eu o fazer. Decidi ouvi-las e cá estou eu.

Agradeço, por fim, a todos os familiares, professores, amigos e colegas que fizeram parte desta jornada. Sem vocês, nada disto teria sido possível.

Resumo

A presente investigação tem como principal objetivo o desenvolvimento e aplicação de um sistema de conversão que permita automatizar a translineação, a marcação da sílaba tónica e a transcrição fonética em português europeu (PE) dos lemas anotados do nível 1 (Muito Fácil) do *Vocabulário Fundamental* do *Dicionário da Língua Portuguesa* da Academia das Ciências de Lisboa (DLP-ACL). Esta investigação insere-se no âmbito de uma parceria com o Instituto de Lexicologia e Lexicografia da Língua Portuguesa (ILLLP), da mesma instituição.

Atualmente, o DLP-ACL carece de critérios fonético-fonológicos uniformes e transversalmente aplicáveis, quer aos lemas existentes, quer a lemas que venham a ser incorporados. Simultaneamente, os fonólogos do português carecem ainda de poucos recursos de acesso aberto, com qualidade e cobertura lexical alargada.

Neste contexto, a primeira etapa do trabalho consistiu na revisão e reformulação dos critérios adotados no *Dicionário da Língua Portuguesa Contemporânea* (DLPC) para a transcrição fonética e a marcação da sílaba tónica, bem como na definição de novos critérios para a translineação. A conversão foi realizada a partir de expressões regulares (*regex*), selecionadas pela sua ampla aplicabilidade e por permitirem identificar padrões baseados em regras predefinidas, e de um programa em *Python*, de modo a aplicar as *regex* ao nosso *corpus*.

De um modo geral, os resultados obtidos foram bastante satisfatórios, revelando maior eficácia na automatização da translineação e da marcação da sílaba tónica do que na transcrição fonética, a qual exigiu intervenção manual em casos linguísticos excecionais. Esta investigação representa um primeiro passo na construção de um sistema de conversão abrangente, prevendo-se a sua futura expansão aos mais de cem mil lemas do DLP-ACL, e constitui um contributo significativo para o desenvolvimento de recursos computacionais aplicados à língua portuguesa.

Palavras-chave: conversão grafema-fonema – processamento de linguagem natural – transcrição fonética automática – dicionário de língua geral

Abstract

This present research's primary aim is the development and application of a conversion system that enables the automatic processing of syllabification, stress marking, and phonetic transcription in European Portuguese (EP) for annotated lemmas belonging to level 1 (Very Easy) of the *Vocabulário Fundamental* of the *Dicionário da Língua Portuguesa* by the Academy of Sciences of Lisbon (DLP-ACL). This study is part of a partnership with the Institute of Lexicology and Lexicography of the Portuguese Language (ILLLP), from the same institution. Currently, the DLP-ACL lacks a thorough audit and the establishment of uniform, broadly applicable phonetic and phonological criteria for both existing entries and future additions. At the same time, Portuguese phonologists still lack open-access resources that offer both quality and broad lexical coverage.

In this context, the first stage of the work involved reviewing and reformulating the criteria adopted in the *Dicionário da Língua Portuguesa Contemporânea* (DLPC) for phonetic transcription and stress marking, as well as defining new criteria for syllabification. The conversion process was carried out using regular expressions (regex), chosen for their broad applicability and ability to identify rule-based patterns designed for specific purposes, and implemented in a Python program to apply the regex rules to our corpus.

Overall, the results were highly satisfactory and point to a more effective automatic treatment of syllabification and stress marking than of phonetic transcription, the latter requiring manual intervention in certain exceptional cases in the language. This research marks the first step towards the development of a conversion system, with future plans to expand it to the more than one hundred thousand lemmas of the DLP-ACL, and represents a significant contribution to the development of computational resources applied to the Portuguese language.

Key-words: grapheme-to-phoneme conversion – natural language processing – automatic phonetic transcription – general language dictionary

Índice de figuras

FIGURA 1 – REPRESENTAÇÕES FONOLÓGICAS POR CONTEXTO E POR AUTOR PARA OS RÓTICOS NO PORTUGUÊS, EM QUE /R/ REPRESENTA UM ARQUIFONEMA, /r/ REPRESENTA O R-FORTE (<i>TRILL</i>) E /r/ REPRESENTA O R- FRACO (<i>TAP</i>)	28
FIGURA 2 – ESTRUTURA INTERNA DA SÍLABA.....	41
FIGURA 3 – DISTRIBUIÇÃO DAS CONSOANTES DO PE EM ATAQUE NÃO RAMIFICADO SIMPLES NO INÍCIO E NO INTERIOR DE PALAVRA.....	41
FIGURA 4 – ESTRUTURAS SILÁBICAS DAS PALAVRAS <i>CRU</i> E <i>PNEU</i>	43
FIGURA 5 – ESTRUTURAS SILÁBICAS DAS PALAVRAS <i>PÉ</i> E <i>PAI</i>	43
FIGURA 6 – ESTRUTURA SILÁBICA DA PALAVRA <i>DIA</i>	44
FIGURA 7 – ESTRUTURA SILÁBICA DA PALAVRA <i>QUANDO</i>	44
FIGURA 8 – LEMA <i>ABACATE</i> DO DICIONÁRIO DE AULETE (1881)	54
FIGURA 9 – <i>CORPUS</i> E <i>OUTPUTS</i>	74
FIGURA 10 – EXEMPLO DO PROCESSO DE TRANSLINEAÇÃO DA PALAVRA <i>CAIR</i>	75
FIGURA 11 – ESTRUTURA DA <i>REGEX</i> DE TRANSLINEAÇÃO DA PALAVRA <i>ALTO</i>	76
FIGURA 12 – ESTRUTURA DA <i>REGEX</i> DE MARCAÇÃO DA SÍLABA TÓNICA DA PALAVRA <i>PÁ</i>	82
FIGURA 13 – ESTRUTURA DAS <i>REGEX</i> DE TRANSCRIÇÃO FONÉTICA DA PALAVRA <i>ASSIM</i>	93
FIGURA 14 – <i>POP-UP</i> DA TRANSLINEAÇÃO NA ENTRADA <i>FINAL</i> NO DLP-ACL (2025)	121
FIGURA 15 – <i>POP-UP</i> DA TRANSCRIÇÃO FONÉTICA NA ENTRADA <i>FINAL</i> NO DLP-ACL (2025)	121
FIGURA 16 – ENTRADA <i>FERVER</i> NO DLP-ACL (2025).....	121
FIGURA 17 – ENTRADA <i>MULHER</i> NO DLP-ACL (2025)	122
FIGURA 18 – ENTRADA <i>FORTE</i> NO DLP-ACL (2025)	122
FIGURA 19 – ENTRADA <i>AMOR</i> NO DLP-ACL (2025)	122

Índice de tabelas

TABELA 1 – CLASSIFICAÇÃO ARTICULATÓRIA DOS SEGMENTOS CONSONÂNTICOS DO PE	24
TABELA 2 – CLASSIFICAÇÃO ARTICULATÓRIA DOS SEGMENTOS VOCÁLICOS ORAIS DO PE	24
TABELA 3 – DISTRIBUIÇÃO DAS CONSOANTES FONÉTICAS DO PE	25
TABELA 4 – PERCENTAGENS DE DIFERENTES REALIZAÇÕES FONÉTICAS DO R-FORTE EM ADULTOS DA REGIÃO DE LISBOA.....	32
TABELA 5 – DISTRIBUIÇÃO DAS VOGAIS FONÉTICAS DO PE	33
TABELA 6 – DISTRIBUIÇÃO DAS VOGAIS NASAIS DO PE	35
TABELA 7 – DISTRIBUIÇÃO DAS GLIDES DO PE.....	36
TABELA 8 – SEQUÊNCIAS CONSONÂNTICAS (C_1C_2) QUE PODEM PREENCHER UM ATAQUE SILÁBICO EM POSIÇÃO INICIAL E NÃO INICIAL NA PALAVRA. C_1 É OBSTRUINTE E C_2 LÍQUIDA ($/r/$, À ESQUERDA, E $/l/$, À DIREITA). 42	42
TABELA 9 – CARACTERÍSTICAS DAS FERRAMENTAS DE CONVERSÃO GRAFEMA-FONEMA DO PE	61
TABELA 10 – CRITÉRIOS DE TRANSCRIÇÃO DAS FERRAMENTAS DE CONVERSÃO GRAFEMA-FONEMA DO PE	62
TABELA 11 – CRITÉRIOS DE TRANSLINEAÇÃO	66
TABELA 12 – DIFERENÇAS ENTRE OS CRITÉRIOS DE TRANSCRIÇÃO DO DLPC (2001) E DO DLP-ACL (2025)...	67
TABELA 13 – COMPARAÇÃO ENTRE O INVENTÁRIO DE CARACTERES DO DLPC (2001) E DO DLP-ACL (2025)	69
TABELA 14 – SINTAXE DOS CARACTERES UTILIZADOS NAS <i>REGEX</i>	72
TABELA 15 – FRONTEIRAS SILÁBICAS DO PORTUGUÊS.....	76
TABELA 16 – <i>REGEX</i> S1	77
TABELA 17 – <i>REGEX</i> S2	77
TABELA 18 – <i>REGEX</i> S3 E S4	78
TABELA 19 – <i>REGEX</i> S5 E S6	79
TABELA 20 – <i>REGEX</i> S7	79
TABELA 21 – <i>REGEX</i> S8	80
TABELA 22 – <i>REGEX</i> S9 E S10	81
TABELA 23 – <i>REGEX</i> S11	81
TABELA 24 – <i>REGEX</i> A1	83
TABELA 25 – <i>REGEX</i> A2 E A3	84
TABELA 26 – <i>REGEX</i> A4.....	84
TABELA 27 – <i>REGEX</i> A5.....	84
TABELA 28 – <i>REGEX</i> A6.....	85
TABELA 29 – <i>REGEX</i> A7.....	85
TABELA 30 – <i>REGEX</i> A8, A9 E A10	86

TABELA 31 – <i>REGEX</i> A11.....	87
TABELA 32 – <i>REGEX</i> A12 E A13	87
TABELA 33 – <i>REGEX</i> A14 E A15	88
TABELA 34 – <i>REGEX</i> A16, A17 E A18	88
TABELA 35 – <i>REGEX</i> A19.....	89
TABELA 36 – <i>REGEX</i> A20 E A21	89
TABELA 37 – <i>REGEX</i> A22.....	90
TABELA 38 – <i>REGEX</i> A23.....	90
TABELA 39 – CATEGORIZAÇÃO DAS <i>REGEX</i> DE MARCAÇÃO DA SÍLABA TÓNICA DAS PALAVRAS PAROXÍTONAS	91
TABELA 40 – <i>REGEX</i> A45 E A57	92
TABELA 41 – <i>REGEX</i> T1 E T2	94
TABELA 42 – <i>REGEX</i> T3 E T4	95
TABELA 43 – <i>REGEX</i> T5, T6 E T7.....	95
TABELA 44 – <i>REGEX</i> T8, T9 E T10	96
TABELA 45 – <i>REGEX</i> T11-T16.....	96
TABELA 46 – <i>REGEX</i> T17-T25	98
TABELA 47 – <i>REGEX</i> T26 E T27	98
TABELA 48 – <i>REGEX</i> T28-T38.....	99
TABELA 49 – <i>REGEX</i> T39-43.....	100
TABELA 50 – <i>REGEX</i> T44-T46.....	101
TABELA 51 – <i>REGEX</i> T47-T50.....	102
TABELA 52 – <i>REGEX</i> T51	102
TABELA 53 – <i>REGEX</i> T52-T58.....	104
TABELA 54 – <i>REGEX</i> T59-T62.....	105
TABELA 55 – <i>REGEX</i> T63-T66.....	105
TABELA 56 – <i>REGEX</i> T67-T71	106
TABELA 57 – <i>REGEX</i> T72-T84.....	107
TABELA 58 – <i>REGEX</i> T85-T92	108
TABELA 59 – AMOSTRA DOS RESULTADOS DE TRANSLINEAÇÃO	110
TABELA 60 – CORREÇÃO MANUAL DAS PALAVRAS ACENTUALMENTE IRREGULARES.....	111
TABELA 61 – AMOSTRA DOS RESULTADOS DA MARCAÇÃO DA SÍLABA TÓNICA.....	112
TABELA 62 – AMOSTRA DE RESULTADOS DA TRANSCRIÇÃO FONÉTICA DOS SEGMENTOS CONSONÂNTICOS.....	113
TABELA 63 – CORREÇÃO MANUAL DA PALAVRA <i>EXEMPLO</i>	114

TABELA 64 – CORREÇÃO MANUAL DOS CASOS EXCECIONAIS EM CONTEXTO PRETÓNICO	116
TABELA 65 – CORREÇÃO MANUAL DE ADVÉRBIOS EM <i>-MENTE</i> E DE ADJETIVOS COM SUFIXO <i>-Z</i> AVALIATIVO... 117	117
TABELA 66 – CORREÇÃO MANUAL DOS CASOS EXCECIONAIS EM INÍCIO ABSOLUTO DE PALAVRA.....	117
TABELA 67 – CORREÇÃO MANUAL DE PALAVRAS NÃO ACENTUADAS	118
TABELA 68 – CORREÇÃO MANUAL DAS VOGAIS SEGUIDAS DE CONSOANTE PALATAL	119
TABELA 69 – CORREÇÃO MANUAL DE EXCEÇÕES	119
TABELA 70 – REPRESENTAÇÃO DAS FORMAS FEMININAS E DAS VARIANTES ORTOGRÁFICAS.....	120
TABELA 71 – DESAMBIGUAÇÃO DAS PALAVRAS HOMÓGRAFAS NÃO HOMÓFONAS	122

Lista de abreviaturas e siglas

ACL	ACADEMIA DAS CIÊNCIAS DE LISBOA
AULETE	DICIONÁRIOS DE CALDAS AULETE
AURÉLIO	DICIONÁRIOS DE AURÉLIO BUARQUE DE HOLANDA
AFI	ALFABETO FONÉTICO INTERNACIONAL
A090	<i>ACORDO ORTOGRÁFICO DA LÍNGUA PORTUGUESA (1990)</i>
BACELAR	<i>DICCIONARIO DA LINGUA PORTUGUEZA</i> DE BERNARDO DE LIMA E MELO BACELAR (1783)
DICIO	DICIONÁRIO <i>ONLINE</i> DE PORTUGUÊS (2025)
DLP-ABL	<i>DICIONÁRIO DA LÍNGUA PORTUGUESA</i> DA ACADEMIA BRASILEIRA DE LETRAS (2025)
DLP-ACL	<i>DICIONÁRIO DA LÍNGUA PORTUGUESA</i> DA ACADEMIA DAS CIÊNCIAS DE LISBOA (2025)
DLPC	<i>DICIONÁRIO DA LÍNGUA PORTUGUESA CONTEMPORÂNEA (ACL)</i>
DP	DICIONÁRIO DE PORTUGUÊS (2025)
ESTRAVIZ	DICIONÁRIO <i>ESTRAVIZ</i> (2025)
FARIA	<i>NOVO DICCIONARIO DA LINGUA PORTUGUEZA</i> DE EDUARDO FARIA (1849)
FIGUEIREDO	DICIONÁRIOS DE CÂNDIDO DE FIGUEIREDO
HOUAISS	DICIONÁRIOS DE ANTÔNIO HOUAISS
INFORMAL	DICIONÁRIO <i>INFORMAL</i> (2025)
IPA	INTERNATIONAL PHONETIC ASSOCIATION
ILLLP	INSTITUTO DE LEXICOLOGIA E LEXICOGRAFIA DA LÍNGUA PORTUGUESA (ACL)
LÉXICO	DICIONÁRIO DE PORTUGUÊS <i>ONLINE</i> (2025)
MD	MEU DICIONÁRIO (2025)
MICHAELIS	MICHAELIS <i>ONLINE</i> (2025)
MORAIS SILVA	DICIONÁRIOS DE ANTÔNIO DE MORAIS SILVA
ODS	<i>OPENDOCUMENT SPREADSHEET</i>
PE	PORTUGUÊS EUROPEU
PORTO EDITORA	DICIONÁRIOS DA PORTO EDITORA
PRIBERAM	DICIONÁRIO <i>PRIBERAM</i> DA LÍNGUA PORTUGUESA (2025)

Introdução

O registo da pronúncia nos dicionários monolíngues de língua portuguesa tem sido, historicamente, subvalorizado, revelando a necessidade de repensar o tratamento lexicográfico deste componente. Torna-se, por isso, imperativo desenvolver recursos rigorosos, atualizados e acessíveis, capazes de responder tanto às necessidades do público em geral como às exigências da investigação académica.

A presente dissertação apresenta o desenvolvimento e aplicação de um sistema de conversão, baseado em expressões regulares, para a translineação e transcrição fonética de 703 lemas anotados pertencentes ao nível 1 (Muito Fácil) do *Vocabulário Fundamental* (Reis, 2025), no âmbito do projeto *iRead4Skills* (Wilkens *et al.*, 2024)¹.

O *Dicionário da Língua Portuguesa* da Academia das Ciências de Lisboa (DLP-ACL) é um dicionário digital de acesso público, disponibilizado em abril de 2023, que atualmente inclui mais de cem mil entradas lexicais. Este recurso resulta da revisão e atualização (Salgado *et al.*, 2023) do *Dicionário da Língua Portuguesa Contemporânea* (DLPC) – a edição impressa em dois volumes e publicada em 2001 pela ACL e a Editorial Verbo.

Embora, na transição para o formato digital, tenha havido um esforço para incluir transcrições fonéticas dos lemas², a presença dos caracteres não-ASCII gerou diversos erros durante a retrodigitalização e conversão do PDF para HTML (Simões *et al.*, 2016). Por esta razão, a CSS (*Cascading Style Sheets*) foi configurada para ocultar temporariamente as transcrições fonéticas. Simultaneamente, a transcrição fonética do DLPC, realizada “segundo a norma culta, aproximada, de Lisboa e Centro do país” (DLPC, 2001: 18), tem sido alvo de críticas, nomeadamente pela omissão de ditongos e vogais, pelo uso de símbolos não pertencentes ao Alfabeto Fonético Internacional (AFI), pela inclusão de ditongos crescentes e pela presença de gralhas diversas (Emiliano, 2009).

Assim, o DLP-ACL carece atualmente de uma auditoria e da definição de critérios fonético-fonológicos uniformes e aplicáveis transversalmente, quer aos lemas

¹ Para saber mais sobre o projeto *iRead4Skills*, consultar: <https://doi.org/10.5281/zenodo.10889986>.

² Por lema, entende-se a cabeça do artigo lexicográfico.

existentes, quer àqueles que venham a ser incorporados. Em paralelo, os fonólogos do português continuam a dispor de escassos recursos de acesso aberto com cobertura lexical alargada, que permitam testar hipóteses fonológicas envolvendo, por exemplo, a frequência de encontros consonânticos problemáticos ou a identificação de lemas cuja tonicidade depende do peso silábico (Trigo & Silva, 2022).

Os principais objetivos desta investigação são:

- a revisão e a reestruturação dos critérios fonético-fonológicos definidos no DLPC (2001);
- a aplicação desses critérios no desenvolvimento do sistema de conversão grafema-fone[ma] e de translineação;
- a identificação de erros resultantes da atribuição automática de caracteres;
- a avaliação das vantagens e limitações das expressões regulares na representação do PE em *corpora* lexicográficos;
- a reflexão e planeamento do emprego do sistema de conversão no DLP-ACL para a exploração fonológica e a sua utilização na expansão do mesmo.

A dissertação organiza-se em seis capítulos.

No Capítulo 1, estabelece-se a base teórica para o desenvolvimento do sistema de conversão, com uma reflexão sobre os níveis e unidades de representação fonológica e fonética (secção 1.1.), a caracterização do inventário fonológico consonântico e vocálico do PE, e a descrição da estrutura silábica e acentual da língua (secção 1.2.). A secção 1.3. analisa a interface entre fonologia e ortografia, com particular atenção aos sistemas de representação.

O Capítulo 2 apresenta uma análise diacrónica do tratamento da pronúncia nos dicionários monolingues de língua portuguesa, desde o século XVIII até à atualidade, com foco nas diferentes edições do dicionário da Academia (secção 2.1.). O Capítulo 3 sintetiza os sistemas de conversão grafema-fonema disponíveis para o PE, com especial destaque para o *Grafone* (secção 3.1.) e o *Convert Text to IPA Transcription* (secção 3.2.).

No Capítulo 4, descrevem-se os passos metodológicos adotados na construção do sistema de conversão. A secção 4.1. analisa os critérios de transcrição do DLPC (2001) e propõe novas diretrizes. A secção 4.2. descreve o tratamento do *corpus*, os

instrumentos utilizados e o procedimento de implementação das expressões regulares, culminando na aplicação do sistema ao DLP-ACL.

O Capítulo 5 discute os dados resultantes da aplicação do sistema: translineação (secção 5.1.), marcação da sílaba tónica (secção 5.2.) e transcrição fonética (secção 5.3.), destacando as vantagens e limitações do modelo adotado, bem como os principais obstáculos enfrentados. A secção 5.4. apresenta os resultados da integração no DLP-ACL (2025) e as estratégias de disponibilização da informação fonética.

Por fim, o Capítulo 6 reúne as considerações finais, recuperando os objetivos e contributos principais da investigação, identificando os seus desafios e propondo linhas futuras para a expansão e consolidação do sistema desenvolvido.

1. Fonologia, fonética e ortografia

Neste capítulo, estabelecemos a fundamentação teórica necessária ao desenvolvimento do sistema de conversão, começando pelo esclarecimento da relação entre a fonologia e a fonética e a distinção dos níveis e unidades de representação destas disciplinas (secção 1.1.). Posteriormente, procedemos à identificação e classificação do inventário fonológico, da estrutura silábica e dos padrões acentuais do PE (secção 1.2.). Encerrámos este capítulo com uma breve reflexão sobre a transparência fonológica desta variedade e os diferentes sistemas de representação (secção 1.3.).

1.1. Níveis de representação fonológicos e fonéticos

A explicitação da dicotomia fonética-fonologia constitui um ponto de partida indispensável à compreensão de qualquer investigação que tenha no seu cerne a representação dos sons de uma língua. Não obstante o estatuto que possuem enquanto campos distintos de ciência, a fonologia e a fonética encontram-se intrinsecamente conectadas pelo domínio no qual operam, ocupando-se da componente sonora das línguas (Andrade, 2020: 3242), ainda que os seus objetos e objetivos de estudo difiram.

É consensualmente aceite que, enquanto subdisciplina linguística, a fonologia se ocupa dos sistemas sonoros da língua, mais especificamente da função, do comportamento e da organização dos sons enquanto unidades linguísticas. Cabe ao fonólogo o estudo do conhecimento fonológico interiorizado e dos respetivos processos e estruturas, sendo-lhe atribuído, por conseguinte, o papel da descrição do inventário fonológico de uma língua, da descrição das regras de combinação dos sons pertencentes a esse inventário, da exploração das variações ocorrentes na pronúncia desses sons, bem como das relações estabelecidas entre eles e do modo como estes se influenciam (Lass, 1984: 1; Davenport & Hannahs, 2005: 2-3; Odden, 2005: 2-4), entre outros.

Por outro lado, a fonética ocupa-se das propriedades físicas, acústicas e articulatórias dos sons emitidos pelo aparelho fonador, sendo o seu principal objeto “a manifestação sensorial das expressões linguísticas, mediante os mecanismos biológicos humanos de produção e perceção da fala” (Andrade, 2020: 3242), inserindo-se no domínio das ciências naturais e tirando proveito de técnicas e métodos provenientes

dessas ciências. Podemos dividir a fonética em três ramos distintos: (i) a fonética acústica, que se centra na observação dos parâmetros acústicos dos sons (*e.g.*, a amplitude, a duração, a frequência, etc.); (ii) a fonética articulatória, responsável pela descrição dos processos articulatórios e das estruturas anatómicas envolvidos na produção dos sons; e (iii) a fonética perceptiva, que trata da forma como os sons são percebidos pelo ouvinte (Malmberg, 1954: 10; Clark & Yallop, 1990: 1; Davenport & Hannahs, 1998:2; Odden, 2005: 2).

Considerando o exposto, o principal nível de distinção entre estes dois domínios estabelece-se a partir dos seus objetos de estudo – os fonemas, objetos de análise da fonologia, e os fones, objetos de análise da fonética – e de três parâmetros que os caracterizam: a materialidade, a distintividade e a variabilidade.

Considerando o primeiro dos parâmetros referidos, os fonemas posicionam-se no domínio imaterial, correspondendo a “entidades puramente teóricas, armazenadas no conhecimento da língua dos falantes, em número finito e reduzido de cada língua” (Veloso, 2015: 409) que representam as variadas realizações fonéticas possíveis. Por outro lado, os fones compreendem essas realizações materiais, abrangendo qualquer som físico emitido pelo aparelho fonador utilizado para transmitir significado. Por outras palavras, um fonema poderá conceber diferentes realizações fonéticas, mas um fone contará apenas com uma representação fonológica.

De acordo com o estruturalismo europeu, a distintividade/contrastividade (ou a ausência desta) constitui o requisito sublime para a separação das classes fonémica e fonética (Veloso, 1996: 412). Assim sendo, um fonema será identificado enquanto fonema se e apenas se for capaz de distinguir ou contrastar significados, sendo esta oposição distintiva o mecanismo responsável pela sua integração no inventário fonológico de uma determinada língua (Trubetskoy, 1939; Jakobson, 1967).

Aos exemplos contrastivos que expressam este tipo de oposição atribuímos o nome de par mínimo, um par de palavras do léxico coincidentes contextualmente em todos os segmentos sonoros exceto num e cuja comutação provoca uma alteração de significado. Atentemos, como exemplo, ao par de palavras *queixo* ['kejʃu] e *queijo* ['kejʒu] (Veloso, 2015: 4), em que todos os segmentos se encontram em conformidade

à exceção de [j] e [ʒ]³, o que desencadeia a diferenciação semântica entre as duas palavras. É a partir deste tipo de observação que nos é possível identificar os fonemas de uma determinada língua (desta forma, considerando o par anterior, /j/ e /ʒ/ constituem fonemas do português), tendo em conta os contextos partilhados entre dois sons.

Uma vez estabelecidas as unidades mínimas distintivas integrantes de um sistema fonológico – os fonemas –, será de igual relevância determinar as relações estabelecidas entre elas, expressadas pelos contextos nos quais se inserem (Gonçalves, 2008: 91). Posto isto, existem certas situações em que a oposição estabelecida entre as unidades lexicais não é distintiva, mas antes o produto de um contraste fonético, como é possível observar através dos exemplos *lado* ['ladu] e *saísa* ['saɪsɐ] (Mateus *et. al*, 2005: 161), em que o mesmo fonema (/l/) apresenta diferentes realizações ([l] e [ɫ]), influenciadas pelo contexto fonético em concreto – neste caso, [l] surge em ataque silábico e [ɫ] em coda silábica.

Às realizações fonéticas provenientes do mesmo fonema que se exibem em determinados contextos de ocorrência atribuímos o nome de alofones (Abaurre, 2006: 47). Apesar das suas diferenças articulatórias e acústicas, os alofones são aceites enquanto variantes do mesmo som pelo falante/ouvinte sem treino fonético, resultando na dificuldade de diferenciação entre eles. Ao compararmos o contexto de ocorrência destas unidades, é possível verificar que se encontram em distribuição complementar, ou seja, ocupam posições mutuamente exclusivas, sendo que cada um deles ocorre num conjunto fixo de contextos no qual o outro não pode ocorrer (Gleason, 1985: 279).

Outro dos fenómenos observáveis entre determinados fonemas é a neutralização das oposições, na qual diferentes fonemas podem coincidir ao nível fonético por serem condicionados foneticamente pelos sons que os rodeiam (Mateus *et. al*, 2005: 162), ocorrendo uma neutralização da distinção da sonoridade. Tomemos como exemplo *aś horas* [ez'ɔrɐ], *aś casas* [ɐj'kazɐ] e *aś meninas* [ɐʒ mɨ'nɪnɐ], em que o morfema de plural -s é realizado respetivamente pelos fonemas /z/, /ʃ/ e /ʒ/. Neste

³ Seguimos a convenção mais geralmente empregada na literatura, representando os fonemas entre barras oblíquas (/ /) e os fones entre parênteses retos ([]).

caso, não ocorre qualquer tipo de mutação semântica entre os determinantes, o que nos leva a verificar que -s perde a sua função distintiva devido à oposição entre dois pontos de articulação (alveolar e alveopalatal) que se manifesta no final de sílaba entre estes três fonemas. Considera-se, assim, que existe um segmento abstrato que reúne os três fonemas, /S/, que designámos de arquifonema e que representa a perda de distintividade fonológica entre /z/, /ʃ/ e /ʒ/ (Silva, 2020).

Finalmente, os fones e os fonemas diferem na medida em que os primeiros estão sujeitos a uma grande diversidade, potencialmente infinita – quer a nível articulatorio, quer a nível acústico –, motivada por diversos fatores contextuais ou idioletais. Pelo contrário, os fonemas são unidades estáticas, invariáveis e finitas, gravadas no sistema fonológico interiorizado do falante e, por conseguinte, independentes de qualquer parâmetro exterior a esse mesmo sistema (Gussenhoven & Jacobs, 1998: 57-58; Veloso, 2015: 14).

1.2. Fonologia do português europeu

A descrição do inventário fonológico aqui realizada cinge-se à norma-padrão do português europeu (PE), a variedade linguística atribuída às regiões compreendidas entre as cidades de Lisboa e Coimbra (Emiliano, 2009: 4). Considerando a sua notoriedade enquanto unidade suprarregional e que o foco desta investigação é um dicionário de língua portuguesa, é a esta variedade que nos referimos quando invocámos o termo PE e que constituirá a base desta análise.

A tradição clássica, assente nas convenções de notação contempladas no AFI da International Phonetic Association (IPA)⁴, sugere a agregação dos segmentos do PE em classes de acordo com as suas características articulatorias (Cruz-Ferreira, 1999: 126; Barroso, 1999: 104; Veloso, 1999a: 26; Emiliano, 2009: 21-31). A Tabela 1 representa uma das várias classificações atribuídas às consoantes do PE quanto ao seu modo de articulação – oclusivas, nasais, vibrantes, fricativas ou laterais –, e quanto ao seu ponto de articulação – bilabiais, labiodentais, apicoalveolares, pré-dorsoalveolares, pré-dorsopalatais, dorsopalatais, pós-dorsovelares e pós-dorsouvulares.

⁴ Ver <https://www.internationalphoneticassociation.org/>.

Tabela 1 – Classificação articulatória dos segmentos consonânticos do PE

	bi labiais	labio dentaís	apico alveolares	pré- dorso alveolares	pré- dorso pre palataís	dorso palataís	pós- dorso velares	pós- dorso uvulares
oclusivas	p b		t d				k g	
nasais	m		n			ɲ		
vibrante múltipla								ʀ
vibrante simples			r					
fricativas		f v		s z	ʃ ʒ			
laterais			l			ʎ		

Fonte: adaptado de Veloso (1999a: 26)

Por sua vez, a Tabela 2 apresenta os segmentos vocálicos orais do PE, produzidos sem qualquer tipo de restrição ou constrição na cavidade oral. As vogais orais do PE podem ser classificadas como anteriores, centrais ou recuadas, tendo em conta a posição longitudinal da língua em relação ao palato e enquanto fechadas, semifechadas, semiabertas e abertas, considerando a abertura dos maxilares na sua produção⁵.

Tabela 2 – Classificação articulatória dos segmentos vocálicos orais do PE

	anteriores	centrais	recuadas
fechadas	i	ɨ	u
semifechadas	e		o
semiabertas	ɛ	ə	ɔ
abertas	a		

Fonte: adaptado de Veloso (1999a: 27)

⁵ De acordo com outras classificações (Emiliano, 2009), é ainda possível distinguir estes segmentos a partir da elevação da língua (altas, semialtas, semibaixas, baixas) e do arredondamento dos lábios (arredondadas e não arredondadas).

1.2.1. Segmentos consonânticos

Um segmento consonântico distingue-se dos demais por ser um segmento com, pelo menos, uma constrição significativa ao fluxo de ar expiratório. A distribuição das consoantes fonéticas do PE ao nível da palavra encontra-se representada na Tabela 3.

Tabela 3 – Distribuição das consoantes fonéticas do PE

Iniciais	Mediais	Finais
[p] pala	[p] ripa	—
[b] bala	[b] riba	—
[t] tom	[t] lato	—
[d] dom	[d] lado	—
[k] calo	[k] rasca	—
[g] galo	[g] rasga	—
[f] fala	[f] estafa	—
[v] vala	[v] estava	—
[s] selo	[s] caça	—
[z] zelo	[z] casa	—
[ʃ] chá	[ʃ] acha	[ʃ] más
[ʒ] já	[ʒ] haja	—
[m] mata	[m] gama	—
[n] nata	[n] gana	[n] abdómen ⁶
—	[ɲ] sanha	—
[l] lato	[l] mala	—
—	—	[ɫ] mal
—	[ʎ] malha	—
[ʀ] rato	[ʀ] carro	—

⁶ As formas terminadas em [n] são excepcionais, pertencentes na sua maioria a um vocabulário especializado e de importação recente (Andrade & D'Andrade, 2020: 3386)

—	[r] caro	[r] mar
—	[ks] táxi	—

Fonte: adaptado de Mateus *et al.* (2003: 992); Andrade & D'Andrade (2020: 3344-3345)

Identificadas as consoantes fonéticas do PE, partimos para a descrição do seu nível fonológico subjacente, formulado a partir de hipóteses baseadas na distribuição e processos que afetam estas mesmas unidades. As propostas de Mateus & D'Andrade (2000) e Mateus *et al.* (2003; 2005) vão ao encontro da generalidade da literatura e identificam um total de 19 consoantes fonológicas para o PE:

- as consoantes oclusivas orais /p/, /b/, /t/, /k/, /g/;
- as consoantes oclusivas nasais /m/, /n/, /ɲ/;
- as consoantes fricativas /f/, /v/, /s/, /z/, /ʃ/, /ʒ/;
- as consoantes líquidas /l/, /ʎ/, /r/ e /R/.

Primeiramente, Mateus *et al.* (2005: 177) partem do pressuposto de que todos os segmentos que correspondem a uma única consoante fonética deverão ser representados por essa consoante, como ocorre com /p/, /b/, /t/, /k/, /g/, /m/, /n/, /ɲ/, /f/, /v/, /ʎ/, /r/ e /R/. No entanto, destaca o problema associado à perda de oposição distintiva entre as consoantes [z, ʃ, ʒ] em certos contextos e às realizações fonéticas [l] e [ʎ], exploradas adiante em 1.2.1.1., e ao segmento mais adequado para as representar.

Por outro lado, Rodrigues (2020) distingue um total de 21 consoantes fonológicas para o PE, acrescentando os segmentos /kʷ/, /gʷ/ e /ks/ à proposta de Mateus *et al.* (2005). A proposta da existência de dois segmentos fonológicos subjacentes, as oclusivas velares /kʷ/ e /gʷ/, em palavras como quanto e guarda, e que diferem daquelas encontradas em palavras como cume e gato, correspondentes às oclusivas orais /k/ e /g/, encontra-se em conformidade com as análises de D'Andrade & Viana (1993) e Freitas (2001). A pressuposição da existência destes segmentos em português relaciona-se com a possibilidade de formarmos pares mínimos como quartel [kʷer'tetʃ] / cartel [ker'tetʃ] e com a impossibilidade de substituímos estas consoantes por outras sem comprometer a recuperação do seu sentido na variedade-padrão (e.g., guarda ['gʷarde] vs. *[gu'arde]). Adicionalmente, Rodrigues (2020) contempla ainda a

existência da africada /ks/ no inventário fonológico do PE, surgindo em palavras como óxido ['ɔksiðu] e cuja pronúncia, tal como as anteriores, não poderia ser substituída por [k] ou [s] sem dificultar a sua inteligibilidade.

Este raciocínio reflete-se, de igual forma, na vibrante /r/ cujas realizações fonéticas [r] e [ʀ] permitem a comutação semântica a partir de pares mínimos (e.g., carro ['karu] / caro ['karu]), mas cuja ocorrência se encontra condicionada pelo contexto fonológico envolvente. Assim sendo, por questões de economia e simplicidade, Rodrigues (2020: 3346) exclui a variante correspondente ao R-forte, mantendo exclusivamente /r/ como o fonema representante destas realizações.

Ao compararmos as propostas de Mateus *et al.* (2003; 2005) e as propostas de Rodrigues (2003; 2020), levanta-se a questão do estatuto fonológico das consoantes vibrantes do PE, a vibrante simples [r] e a vibrante múltipla [ʀ]. Estas vibrantes constituem segmentos alofónicos e, por isso, encontram-se em distribuição complementar – a vibrante simples não ocorre em posição inicial de palavra e a vibrante múltipla não ocorre em posição final de palavra –, tal como podemos observar:

1. Contextos de ocorrência de [r]
 - segunda posição de ataque silábico ramificado, precedida de consoante oclusiva ou fricativa (e.g., prato ['pratu], fraco ['fraku]);
 - coda silábica medial ou final (e.g., parto ['partu], ler ['ler]).
2. Contextos de ocorrência de [ʀ]
 - ataque silábico inicial ou início de radical em certas palavras prefixadas (e.g., rato ['ratu], desratizar [diʃreti'zar]);
 - ataque silábico medial, precedida de /s/ ou /l/ em coda silábica ou de autossegmento nasal (e.g., israelense [iʃʀe'eləsi], melro ['mɛʃru], tenro ['tẽru]);
 - posição inicial de palavra fonológica (e.g. vice-rei [visi'ʀej]).

Contudo, estas vibrantes partilham um contexto de ocorrência, o ataque silábico medial (intervocálico), a partir do qual é possível formar pares mínimos:

caro ['karu] / carro ['karu]

Mateus *et al.* (2003; 2005) inserem-se no grupo dos autores que defendem a existência de duas consoantes vibrantes no inventário fonológico do PE: /ʀ/ e /r/

(Barbosa, 1983; Câmara, 1977; Mateus, 1982a; Miranda, 1996; Bonet & Mascaró, 1997; Pereira, 2020) ou /r/ e /ʁ/ (Cruz-Ferreira, 1999). Por outro lado, Rodrigues (2003, 2020) privilegia a existência de somente uma consoante fonológica vibrante – /r/ –, tal como Monaretto (1997), Mateus & D’Andrade (2000), Vigário (2003), Veloso (2019) e ainda Câmara (1953).

As diferentes propostas diferem, por vezes, devido aos modelos teóricos que as subentendem, contudo, a admissão de dois fonemas (/r/ e /ʁ/) é a proposta mais geralmente aceite e mais representada na literatura, como podemos observar a partir da Figura 1.

Figura 1 – Representações fonológicas por contexto e por autor para os róticos no português, em que /R/ representa um arquifonema, /ʁ/ representa o R-forte (*trill*) e /r/ representa o r-fraco (*tap*)

Autores Contexto Fonético	Mattoso Câmara (1953)	Mattoso Câmara (1977)	Morais Barbosa (1983)	Bonnet & Mascarò (1997)	Monaretto (1997); Mateus & Andrade (2000)	Mateus (1982, 1984), Mateus et al. (2003) e Mateus, Falé e Freitas (2005)
<i>rato</i> ['ratu] ou <i>melro</i> ['meɫru]	/R/	/R/	/R/	/R/	/ɾ/	/R/
<i>carro</i> ['karu]	/R/	/R/	/R/	/R/	/ɾ.ɾ/	/R/
<i>caro</i> ['karu]	/R/	/ɾ/	/ɾ/	/ɾ/	/ɾ/	/ɾ/
<i>prato</i> ['pratu]	/R/	/ɾ/	/R/	/R/	/ɾ/	/ɾ/
<i>mar</i> ['mar]	/R/	/R/	/R/	/R/	/ɾ/	/ɾ/

Fonte: Pereira (2020: 22)

A existência de um segmento fonológico /R/ ou /r/ para além de /r/ na variedade-padrão do PE é uma questão ainda não consensual e bastante debatida. Perante este cenário, Pereira (2020: 148) esclarece que, se considerarmos a existência do segmento /R/, teremos de ter em mente que o seu contraste é mais limitado do que o dos outros segmentos por não ocorrer em posição intervocálica e que a menor marcação de /r/ se relaciona com a sua simplicidade e elevada frequência. Por outro lado, ao adotarmos a proposta que apenas contempla a existência do segmento /r/, estaremos a adotar uma representação ao nível superficial, por intermédio de [r] e [ʁ].

1.2.1.1. Alternâncias fonético-fonológicas

Os fenómenos de alternância fonético-fonológica⁷ correspondem a processos regulares segundo os quais uma unidade fonológica apresenta mais do que uma realização fonética num determinado contexto (Rodrigues, 2020: 3352). Neste caso, abordámos: (i) as alternâncias associadas a /S/ em coda silábica e em posição intervocálica; e (ii) as alternâncias associadas a /l/ em coda silábica, cuja ocorrência se pode considerar relativamente uniforme na variedade-padrão do PE.

Tal como referido em 1.1.2., /S/ constitui um arquifonema representante de múltiplas realizações fonéticas em fim de palavra, que assumem diferentes pontos de articulação e vozeamento e cuja alternância depende dos segmentos fonológicos subsequentes. O primeiro fenómeno de alternância fonético-fonológica trata de /S/ e das suas respetivas realizações [ʃ, ʒ, z]:

1. Contextos de ocorrência das realizações fonéticas de /S/⁸

- [ʃ]: coda silábica medial e final, seguida de consoante surda /p/, /t/, /k/, /f/ ou /s/ (e.g., paʃta ['paʃte], casaʃ claraʃ ['kazeʃ klareʃ])⁹;
- [ʒ]: coda silábica medial e final, seguida de consoante sonora /b/, /d/, /g/, /m/, /n/, /ɲ/, /r/, /v/, /z/, /l/ ou /ʎ/ (e.g., meʒmo ['meʒmu], casaʒ bonitaʒ ['kazeʒ buniteʒ]);
- [z]: ataque silábico intervocálico (e.g., aza ['aze]).

Tal como é explicitado por Mateus *et al.* (2003: 1016), este fenómeno deve-se à assimilação por parte de /S/ em posição de coda silábica do traço [vozeado] do segmento seguinte, refletindo-se na sua pronúncia como [ʃ] quando seguida de consoante surda e como [ʒ] quando seguida de consoante sonora. Por outro lado, quando se junta a uma vogal, para além de assimilar o seu vozeamento, a especificação

⁷ Uma vez que todos os lemas verbais do dicionário se encontram na sua forma canónica infinitiva, as alternâncias morfofonológicas não são aqui consideradas.

⁸ Apesar da diferente origem etimológica, a consoante final comporta-se igualmente como /S/ nas palavras *capaz* e *tenaz*, sendo realizada como uma das suas variantes [ʃ] ou [ʒ] por se encontrar em coda silábica.

⁹ D'Andrade & Rodrigues (1999) admitem a ocorrência de /ʃ/ em coda silábica inicial em palavras como *esquecida*.

do articulador de /s/ é [+anterior] por não se encontrar em coda, pronunciando-se então como [z].

O segundo destes fenómenos relaciona-se com os diferentes níveis de velarização que a consoante /l/ pode materializar, dependendo do contexto em que se encontra inserida. De acordo com a generalidade dos autores, esta consoante associa-se a dois alofones, condicionados pela sua posição na sílaba e em distribuição complementar (Andrade, 1998; 1999; Mateus *et al.*, 2003: 995; Mateus *et al.*, 2005: 161; Rodrigues, 2020: 3357), sendo que [ɫ] se distingue de [l] do ponto de vista articulatorio por possuir um ponto de articulação secundário (a velarização):

1. Contextos de ocorrência de [l]
 - ataque silábico inicial e medial (*e.g.*, lapa ['lape], mala ['male]);
2. Contexto de ocorrência de [ɫ]
 - coda silábica medial e final (*e.g.*, saltar [saɫ'tar], sal ['saɫ]).

A eleição de /l/ como o segmento fonológico subjacente às realizações fonéticas [l] e [ɫ] justifica-se pela maior frequência de [l] comparativamente a [ɫ] devido aos seus contextos de ocorrência e pelo facto de o ponto de articulação secundário de [ɫ] a transformar numa consoante mais rara comparativamente a [l], muito mais vulgar nas línguas do mundo (Mateus *et al.*, 2005: 178).

1.2.1.2. Variações na norma-padrão

Para além dos processos de alternância fonético-fonológica, a variedade-padrão do PE regista também fenómenos de variação consonântica. Nesta secção, abordámos: (i) a variação associada a /b, d, g/ em posição intervocálica; e (ii) a variação associada ao R-forte.

A variação associada a /b, d, g/ diz respeito às realizações contínuas fricativas não estridentes [β, ð, ɣ] a que as consoantes /b, d, g/ se encontram sujeitas no interior da palavra, nomeadamente em contexto intervocálico, tendo em conta que a posição silábica inicial preserva as oclusivas simples [b, d, g] (Velo, 1997: 846). Estes dois pares de consoantes são alofónicos e encontram-se em distribuição complementar, como podemos constatar a partir dos seus típicos contextos de ocorrência:

1. Contextos de ocorrência das fricativas [β, ð, ɣ]

– ataque silábico medial intervocálico (*e.g.*, baba ['baβe], dada ['daðe], gaga ['gaye]).

2. Contextos de ocorrência das oclusivas [b, d, g]¹⁰

– ataque silábico inicial (*e.g.*, baba ['baβe], dada ['daðe], gaga ['gaye]).

No entanto, foram já registadas limitações contextuais a esta distribuição como, por exemplo, a não fricatização se a vogal precedente for nasal (*e.g.*, ambos ['ẽbu], lindo ['lĩdu], ganga ['gẽge]), a maior resistência de /d/ à fricatização relativamente a /b/ e /g/ e a fricatização quer em posição inicial de palavra, quer em posição não intervocálica medial (Rodrigues, 2020: 3365). Assim sendo, podemos depreender que /b, d, g/ se encontram sujeitas a restrições de natureza discursiva e societal ainda pouco estudadas. Dado que os ouvintes do português não reconhecem auditivamente qualquer tipo de diferença entre as fricativas e as oclusivas, não existindo, portanto, qualquer tipo de contraste fonológico entre estas (Veloso, 1997: 849; Veloso, 1999b: 229), surge uma oscilação na sua produção ao nível idioletal.

Conforme indicado, o PE dispõe de duas consoantes fonéticas vibrantes: uma vibrante simples, ou r-fraco, e uma vibrante múltipla, ou R-forte. A primeira é representada na generalidade da literatura através da realização fonética [r], não se registando grande variação na sua produção, à exceção de [ɾ], [ɽ], [ɹ] e [ɹ̥] (Jesus & Shadle, 2005; Rodrigues, 2003; Veloso, 2015), que ocorrem em contextos e regiões dialetais muito específicos. É, contudo, sob a representação da segunda vibrante que recai a nossa atenção, tendo constituído tema de discussão significativo entre os fonólogos do PE devido à variação livre a que se encontra sujeita.

Admitem-se para o R-forte diferentes realizações fonéticas, desde as vibrantes [R] e [r] (Mateus & D'Andrade, 2000: 8; Rennie & Martins, 2013: 513), as fricativas [ʁ], [χ], [x] (Barbosa, 1983: 192; Barroso, 1999; Cruz-Ferreira, 1999; Rodrigues, 2015), e mais recentemente as fricativas [ɹ̥], [ɹ̥̥], [ɹ̥̥̥] e a glotal [h] (Rodrigues, 2016; Pereira, 2020: 147). Esta elevada oscilação constitui um problema para a representação fonética, especialmente quando se pretende fazer a representação de uma variedade específica

¹⁰ A oclusiva [b] pode ainda surgir em posição não intervocálica medial (*e.g.*, abdómen) e no final de palavra fonológica (*e.g.*, ab-rogar).

da língua como a variedade-padrão. Deste modo, detenhamo-nos nos dados referentes à região de Lisboa¹¹, apresentados na Tabela 4.

Tabela 4 – Percentagens de diferentes realizações fonéticas do R-forte em adultos da região de Lisboa

Realizações fonéticas do R-forte	Rodrigues (2003)	Pereira (2020)
[ʀ]	83,6%	47%
[ʁ]		16%
[χ]		20%
[x]		-
[r]	16,5%	-
[ɾ]	-	17%

Apesar da escassa literatura acerca das realizações do R-forte na variedade-padrão do PE, Mateus & D’Andrade (2000) e Vigário (2003) defendem a presença da alveolar [r] em dialetos portugueses que não o de Lisboa, hipótese posteriormente refutada por Rodrigues (2003), como é possível observar na Tabela 4. Rodrigues (2003) apenas considera a posição inicial de palavra e, mesmo assim, deixa clara a predominância da vibrante múltipla [ʀ], da uvular [ʁ] e das fricativas [χ] e [x], em detrimento da alveolar [r]. Numa outra análise, Jesus & Schadle (2005) descrevem os dados de quatro falantes naturais de Aveiro, Braga, Sintra e Lisboa, sendo que o falante de Lisboa realiza a fricativa [χ] dez vezes, não apresentando mais nenhuma das realizações possíveis em contexto de R-forte.

O estudo de Pereira (2020) veio contribuir para uma descrição mais minuciosa das realizações associadas a R-forte na região de Lisboa. Apesar de a alveolar [r] se destacar como a vibrante predominante no território nacional (Pereira, 2020: 157), o que vai ao encontro do que a generalidade da literatura reporta, esta vibrante não se encontra na variedade-padrão do PE. Em vez disso, Pereira (2020: 119) regista uma

¹¹ Dado que não se encontram dados da região de Coimbra, cingimo-nos a Lisboa.

prevalência da vibrante múltipla [ʀ], ocorrente tantas vezes em ataque inicial como medial, seguida pela fricativa [χ], pela uvular [ʁ] e pela fricativa [ʀ].

1.2.2. Segmentos vocálicos

Contrariamente às consoantes, os segmentos vocálicos são produzidos sem qualquer tipo de constrição à passagem do ar pela cavidade oral. No PE, as vogais distribuem-se ao nível acentual conforme ilustrado na Tabela 5.

Tabela 5 – Distribuição das vogais fonéticas do PE

Vogais orais			
Átonas pretónicas	Tónicas	Átonas postónicas não finais	Átonas finais
[i] mirar	[i] silo	[i] súbito	[i] júri
—	[e] selo	—	—
—	[ɛ] selo	—	—
[u] morar	[u] bula	[u] cómoda	[u] juro
—	[o] bola	—	—
—	[ɔ] bola	—	—
—	[a] talha	—	—
[e] pagar	[e] telha	[e] sábado	[e] jura
[ɨ] pegar	—	[ɨ] vértebra	[ɨ] jure

Fonte: adaptado de Mateus *et al.* (2003: 991-992)

Uma vez determinadas as vogais fonéticas do PE, podemos identificar 6 vogais fonológicas às quais corresponde uma única realização fonética: /i/, /e/, /ɛ/, /o/, /ɔ/ e /u/. Contudo, tal como acontece para as consoantes vibrantes, também o estatuto fonológico das vogais [a], [e] e [ɨ] gera alguma controvérsia.

Quanto ao estatuto fonológico de [a] e [e], a generalidade da literatura encontra-se alinhada com as propostas de Mateus & D'Andrade (2000) e Mateus *et al.* (2003; 2005), segundo as quais estamos perante uma vogal fonológica /a/ com duas realizações

fonéticas subjacentes. Esta conclusão parte da análise dos contextos de ocorrência de [e] em sílaba tónica pois é o único contexto partilhado entre estas duas vogais:

1. Contextos de ocorrência de [e] tónico¹²

- antes de consoante palatal /ɲ/, /ʃ/, /ʎ/, /ʒ/ (e.g., *telha* ['teʎe], *fecho* ['feʃu])¹³;
- antes de glide palatal [j] (e.g., *lei* ['lej]);
- antes de consoante nasal /m/, /n/ (e.g., *cama* ['keme], *cana* ['kene])¹⁴.

No caso de <e> seguido de consoante palatal, tanto Mateus & D'Andrade (2000: 17) como Mateus *et al.* (2005: 174) afirmam que a ocorrência de [e] é inexistente na região de Lisboa, realizando-se sempre [ɛ]. Por outro lado, regista-se a ocorrência de [e] na região de Coimbra antes das palatais não nasais /ʃ/, /ʎ/ ou /ʒ/ (e.g., *telha* ['teʎe]).

O mesmo ocorre para a vogal [i] que, de acordo com a maioria dos autores, não pertence ao inventário fonológico do português por apenas ocorrer em contexto átono e por ser sempre resultante de /e/ ou /ɛ/ (e.g., *selo* ['selu] vs. *selar* [sɨ'lar]) (Mateus *et al.*, 2005: 174). Por outro lado, certos autores defendem o estatuto fonológico desta vogal, como Veloso (2003b; 2005b; 2007) que, entre outros argumentos, se apoia: (i) na inexistência de pares em que este constituinte ocorre foneticamente ora como [e] ou [ɛ] tónico, ora como átono [i] com as formas clíticas *que*, *te*, *lhe* e nomes de tema em [i]; (ii) na contrastividade que estabelece em pares como *parte* ['partɨ] e *parto* ['partu]; e (iii) na possibilidade de estabelecer oposições gramaticais em pares como *infante* ['ĩ fẽtɨ] e *infanta* ['ĩ fẽtɛ].

Para além das vogais orais, o PE dispõe ainda de cinco vogais nasais que se encontram distribuídas em termos de tonicidade (Tabela 6).

¹² Correspondem a contextos onde [e] e [ɛ] podem alternar, exceto na região de Lisboa.

¹³ O mesmo não se verifica para palavras como *velho* ou *mecha*, realizadas com [ɛ].

¹⁴ Constituem exceções a estas ocorrências a forma da primeira pessoa do plural do pretérito perfeito dos verbos da primeira conjugação (e.g., *apanhámos*) e casos raros como o verbo *ganhar*.

Tabela 6 – Distribuição das vogais nasais do PE¹⁵

Vogais nasais	
Tónicas	Átonas pretónicas
[ĩ] cinto	[ĩ] cintar
[ẽ] sento	[ẽ] sentar
[õ] mondo	[õ] mondar
[ũ] mundo	[ũ] mundial
[ẽ] mando	[ẽ] mandar

Fonte: adaptado de Mateus *et al.* (2003: 992)

Também o estatuto fonológico das vogais nasais tem sido objeto de discussão. Embora certos autores proponham que a nasalidade vocálica é distintiva, integrando as vogais nasais no inventário fonológico do PE (Câmara, 1953; Leite, 1974; Abaurre, 1983), a generalidade dos autores defende que esta é condicionada, inserindo as vogais nasais somente no nível fonético do PE (Barbosa, 1983; Parkinson, 1983; D'Andrade & Khim, 1988; Mateus & D'Andrade, 2000; Mateus *et al.*, 2003; 2005; Bisol & Veloso, 2016; Andrade, 2020).

As vogais podem ainda surgir agrupadas a uma outra grande classe fonética, as semivogais ou glides, sons produzidos sem uma constrição significativa, mas com um grau de abertura do trato vocal menor do que o dos sons vocálicos. No PE, encontramos as glides orais [j] e [w] e as glides nasais [j̃] e [w̃], que se caracterizam por serem idênticas a [i] e [u], respetivamente, mas que se distinguem dos últimos por serem mais breves, não serem acentuáveis, nem poderem constituir núcleo de sílaba (Mateus *et al.*, 2005: 170), como veremos adiante em 1.2.3. Quando agrupados à direita de uma vogal, estes segmentos constituem ditongos decrescentes (orais e nasais) e encontram-se distribuídos ao nível da passagem do ar pelas cavidades supraglotais, conforme exemplificado na Tabela 7.

¹⁵ Para diversos falantes do PE, [ɣ̃] ocorre excecionalmente no início da palavra *ontem* (Andrade, 2020: 3275).

Tabela 7 – Distribuição das glides do PE

Glides fonéticas			
Orais		Nasais	
[j]	[w]	[ɰ]	[w̃]
—	[iw] riu	—	—
— ¹⁶	[ew] meu	—	—
[ɛj] papéis	[ɛw] véu	—	—
[oj] noite	— ¹⁷	[õ̃] põe	—
[ɔj] rói	—	—	—
[uj] cuida	—	[ũ̃] muito ¹⁸	—
[ej] lei	[ɛw] saudade	[ẽ̃] mãe	[ẽw̃] mão
[aj] pai	[aw] pau	—	—

Fonte: adaptado de Mateus *et al.* (2003: 994)

Segundo Mateus *et al.* (2003: 993), as glides apenas surgem no nível fonético do PE, pois não existem pares mínimos que ponham em paralelo ditongos e sequências de duas vogais (*e.g.*, pai [paj] vs. *[pai]), não havendo contraste fonológico entre estas. Por outro lado, Mateus *et al.* (2005: 177) defende a existência de glides no nível fonológico, apoiando-se num domínio superior ao segmento: a sílaba. Segundo esta proposta, ao pedir-se a um falante que pronuncie pausadamente palavras com ditongos decrescentes como *leite* e *paixão*, não obtemos uma divisão silábica entre vogal e semivogal como *le.i.te ou *pa.i.xão. O contrário ocorre, no entanto, quando estamos perante palavras como *piar* ou *petróleo*, em que se espera que as divisões correspondam a pi.ar e pe.tró.le.o, respetivamente. Deste modo, admite-se que, nas sequências do primeiro tipo (ditongos decrescentes), estamos perante uma vogal e uma semivogal ao nível

¹⁶ O ditongo [ej] apenas ocorre em certas variedades regionais do PE.

¹⁷ O ditongo [ow] apenas ocorre em certas variedades regionais do PE, sofrendo de monotongação ([o]) na variedade-padrão.

¹⁸ O ditongo nasal [ũ̃] apenas se encontra nas palavras *muito* e *muitíssimo* o PE.

fonológico (VG), ao passo que, nas sequências do segundo tipo, estamos perante duas vogais ao nível fonológico (VV).

Importa frisar, contudo, que estas sequências podem, em determinados contextos, resultar no enfraquecimento de um segmento e na conseqüente formação de ditongos crescentes (GV), como em *piar* [pjar] e *petróleo* [pitʁolju]. De acordo com Andrade (2020: 3274), a pressuposição da existência de ditongos crescentes no PE numa fala mais rápida e coloquial não é só plausível como provável; contudo, trata-se de um fenómeno dependente de fatores extralinguísticos, nomeadamente segmentais, prosódicos, lexicais e morfológicos, situando-se somente no nível fonético.

Considerando o exposto, existe um grupo específico de palavras como *quando*, *quatro*, *quarto*, *língua*, *antiguidade* e *água* que contêm o segmento [w] antes de vogal e que parecem contrariar o pressuposto da ausência de ditongos crescentes em português. Surge então a questão do estatuto de [w], que, de acordo com as descrições mais tradicionais do português, corresponde a uma semivogal e forma, excepcionalmente um ditongo crescente. Por outro lado, vários autores defendem a sua realização como uma articulação secundária pertencente às oclusivas velares labializadas /kʷ/ e /gʷ/ (Bisol, 1989: 215, D'Andrade & Viana, 1993: 39; Vigário & Falé, 1994: 476; Freitas, 2001: 220; Bisol, 1992: 121; Rodrigues, 2020: 3345; Andrade & D'Andrade, 2020: 3382), apoiando-se no facto de a comutação entre o segmento e uma vogal não ser possível (e.g., *quarto* ['kwartu] vs. *[ku'artu]), da consoante que precede o segmento ser obrigatoriamente uma oclusiva velar (/k/ ou /g/) e de apenas o segmento [w] ocorrer nestes contextos e nunca [j]. Esta última proposta tem algumas implicações no sistema consonântico fonológico do português, pelo que admite o acréscimo destas duas consoantes, conforme verificámos anteriormente em 1.2.1.

1.2.2.1. Processos fonológicos e variação

O sistema vocálico do PE é mais suscetível a determinados processos fonológicos do que o sistema consonântico, entre os quais destacámos a redução do vocalismo átono. Como pudemos observar anteriormente na Tabela 3, o inventário de vogais ocorrentes em sílaba pretónica e postónica ([i, ɨ, e, u]) é significativamente mais reduzido do que o inventário de vogais ocorrentes em sílaba tónica ([i, e, ε, o, ɔ, u, a]).

Isto deve-se, do ponto de vista estrutural, a um processo de elevação de um ou dois graus de altura (/a/ [+baixa] → [e] [-baixa]) e ao recuo (/e, ε/ [-recuadas] → [i] [+recuada]) das vogais fonológicas átonas e não altas (Mateus *et al.* 2003: 1012), resultando nas seguintes modificações do sistema vocálico:

- /a/ realiza-se como [e] quando átona;
- /e, ε/ realizam-se como [i] quando átonas;
- /o, ɔ/ realizam-se como [u] quando átonas;
- /i, u/ não se alteram.

Apesar de se tratar de um processo estável do PE, a redução do vocalismo átono não ocorre linearmente em todos os contextos e dado que podemos terminar alguns dos contextos que impedem a sua aplicação, encontramos-nos perante exceções regulares que não sofrem redução vocálica:

- vogais nasais ([ẽ, ẽ, ĩ, õ, ũ]) (*e.g.*, dançarina [dẽse'rine]);
- vogais em ditongos decrescentes ([aj, ej, oj, uj, aw, ew]) (*e.g.*, bairrista [baj'rifte]);
- vogais átonas não altas ([e¹⁹, ε, o, ɔ, a]) seguidas de /l/ em coda silábica (*e.g.*, saltar [saɫ'tar]);
- vogais átonas não altas ([a, ε, ɔ]) com consoante diferente de /s/ em coda silábica final (*e.g.*, ímpar [ĩ'par], pólen [p'ɔɫɛn], fórceps [f'ɔɾseps]);
- vogais átonas não altas coronais ou labiais com uma sequência de segmentos consonânticos heretossilábicos à sua direita realizam-se como baixas ([ε, ɔ, a]) (*e.g.*, prínceps [p'ɾĩseps], cognição [kɔgni'sẽw]), embora este fenómeno esteja em declínio no que respeita a /a/ (*e.g.*, magnésio [ma'gneziu] vs. diagnóstico [die'gnɔʃtiku]);
- vogal coronal não alta átona ([ε, e]) seguida de outra vogal em início de sílaba, realizando-se como [i] (*e.g.*, leao [li'ẽw]), exceto quando a vogal seguinte é [i], realizando-se como [e] (*e.g.*, veiculo [ve'ikulu]);
- vogais acentuadas de base que passam a átonas em palavras que integram os sufixos -avaliativos e o sufixo -mente (*e.g.*, devagarzinho [devegar'ziju], facilmente [fasit'mẽti]);

¹⁹ Ocorre em casos raros como *fedspato*, *feltro* ou *felpo*.

– vogais de ligação de certos compostos morfológicos (e.g., *sonoplastia* [sɔnɔplej'tie]).

Para além destas exceções, existem determinados contextos em que a redução vocálica parece ser variável de acordo com a palavra e o falante, tratando-se de um fenómeno instável da fonologia do PE. Um destes contextos é a sílaba iniciada por vogal, em que se observa uma tendência para a realização da vogal subjacente /e/ como [ɛ] (e.g., *bienal* [biɛ'nal]) ou [e] (e.g., *sociedade* [susie'dadɨ]) quando a vogal anterior é /i/, para a sua redução para [i] quando a vogal anterior é /a/ e para [ɨ] quando a vogal anterior é /o/ ou /i/ (e.g., *roedor* [ruɨ'dɔr]). Outro dos contextos é o início absoluto de palavra, em que se verifica uma tendência para a redução de /a/ para [e] em contexto átono, à exceção de um conjunto específico de palavras (e.g., *armazam*, *arma*, *alem*, *aquem*, etc.). Para além disso, as vogais coronais e labiais não altas átonas não registam redução para [ɨ] ou [u], pelo que se observa uma preferência da produção de [i] e [o] para falantes mais conservadores (e.g., *erguer* [ir'ger]; *olhar* [o'lar]) e da produção de [e] e [ɔ] para falantes menos conservadores (e.g., *ergeer* [er'ger]; *oloar* [ɔ'lar]) (Andrade, 2020: 3327-28). Por outro lado, surgem determinadas palavras cuja vogal inicial é baixa ([ɛ, ɔ]) para a generalidade dos falantes, como é o caso de *hotel*, *hospital*, *oculista*, *ervanária*, *herbácea*, etc.

Ainda em fronteira inicial de palavra, importa explorar o que ocorre nas sequências do tipo #sC e a possibilidade de ocorrência (ou não) de vogal à esquerda da fricativa /j/. A proposta de D'Andrade & Rodrigues (1999: 118) faz uma distinção de quatro estruturas que podem surgir neste contexto: (i) *escola*, *estar*; (ii) *stress*, *snob*; (iii) *isqueiro*, *islamismo*; e (iv) *experiência*, *extrair*. Os padrões observados neste estudo vão ao encontro do que é observado em Freitas & Rodrigues (2003: 62) registando-se, para a variedade-padrão do PE, a predominância da produção sem vogal para palavras do grupo (i) (e.g., *escola* [j'kɔle]); a produção exclusiva da fricativa [s] para palavras do grupo (ii) (e.g., *snob* [j'snɔb]); a produção maioritária de [i] para palavras do grupo (iii) (e.g., *isqueiro* [ij'kejru]) e a produção predominante sem vogal, mas também [ejj] para palavras do grupo (iv) (e.g., *extrair* [jtre'ir] / [ejjtre'ir]).

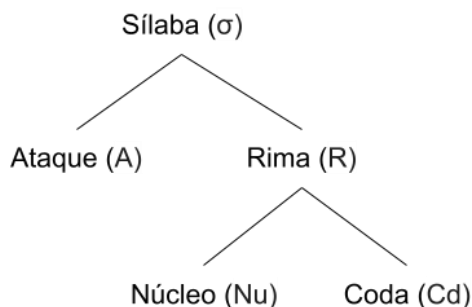
Por outro lado, Andrade (2020: 3310) defende a existência da pronúncia da vogal [i] no início de palavras pertencentes ao grupo (i) (*e.g.*, *espontâneo*, *estudo*) e ao grupo (ii) (*e.g.* *exposição*, *extensivo*). Por se tratar de um fenómeno de variação livre, não obrigatória e não sistemática, a maioria dos estudos incide sobre a estrutura silábica e respetivo processamento destas estruturas (Delgado-Martins, 1996; Freitas, 1998; Mateus & D'Andrade, 2000: 44; Veloso, 2002; Fikkert & Freitas, 2004; Goad, 2012; Henriques, 2012a; 2012b; Valada, 2017), pelo que carecemos de estudos atuais que esclareçam a variação aqui registada, limitando-nos, portanto, à identificação de tendências.

Por fim, surgem ainda casos em que as vogais não obedecem às regras da redução vocálica, mas também não se encontram condicionadas pelo contexto. Estes casos são verdadeiras exceções que apenas se encontram marcadas no léxico mental dos falantes nativos do PE e que dão lugar à neutralização de altura entre as vogais não altas a favor das baixas [ɛ, ɔ, a] (Andrade, 2020: 3314), como nas palavras *correção* [kʊɛ'ẽw̃], *normal* [nɔɾ'maɫ] e *invasão* [ĩva'zẽw̃]. Estas exceções encontram-se ainda em empréstimos relativamente recentes (*e.g.*, *raquete* [ɾa'keti], *boné* [bɔ'nɛ]) e em posição final de formas resultantes do truncamento de compostos (*e.g.*, *euro* ['ewɾɔ], *foto* ['fɔtɔ], *tele* ['tɛɛ]).

1.2.3. Sílabas

A sílaba é uma unidade linguística mínima de organização segmental “criada no espírito do ouvinte, com propriedades específicas que não decorrem da simples segmentação fonética das sequências de segmentos” (Mateus *et al.* 2003: 1038), possuindo uma estrutura interna composta por quatro constituintes silábicos que se organizam de forma binária, tal como ilustrado pela Figura 2.

Figura 2 – Estrutura interna da sílaba



Fonte: adaptado de Pike & Pike (1947) e adotado por Selkirk (1984); Blevins (1995: 213); Freitas (1997: 31); Bisol (1999: 702); Freitas & Santos (2001: 23); Veloso (2003a: 92); Mateus (1994); Mateus *et al.* (2003: 1038)

O ataque é um constituinte silábico facultativo em PE, corresponde ao(s) segmento(s) consonântico(s) localizado(s) à esquerda do núcleo e pode ser: (i) não ramificado vazio (*e.g.*, Øolá); (ii) não ramificado simples (*e.g.*, pão); e (iii) ramificado (*e.g.*, preparar). Em ataque não ramificado simples, todas as consoantes fonológicas do PE podem ocorrer no início e no interior de palavra, à exceção de /r/, conforme podemos observar na Figura 3. As consoantes palatais /ʎ/ e /ɲ/ surgem apenas palavras de origem estrangeira (*e.g.*, lhana) ou obscura (*e.g.*, nhonhas) (Freitas, 1997; Andrade & D’Andrade, 2020: 3377; Trigo & Silva, 2022: 358).

Figura 3 – Distribuição das consoantes do PE em ataque não ramificado simples no início e no interior de palavra

(3)	pa-pa	[p-p]	fo-fa	[f-f]	li-lás	[l-l]	ca-ro	[-r]
	ba-ba	[b-b]	vi-ve	[v-v]	ta-lha	[-ʎ]	ca-rro	[-R]
	te-ta	[t-t]	sen-so	[s-s]			ra-ro	[R-]
	de-do	[d-d]	zon-za	[z-z]	ma-mar	[m-m]		
	co-ca	[k-k]	cho-cho	[ʃ-ʃ]	ni-nar	[n-n]		
	ga-go	[g-g]	je-jum	[ʒ-ʒ]	se-nha	[-ɲ]		

Fonte: Mateus *et al.* (2003: 1039)

Por outro lado, os ataques ramificados em PE possuem um maior número de restrições, admitindo um número limitado de grupos consonânticos, também conhecidos como grupos próprios, apresentados na Tabela 8.

Tabela 8 – Sequências consonânticas (C₁C₂) que podem preencher um ataque silábico em posição inicial e não inicial na palavra. C₁ é obstruiente e C₂ líquida (/r/, à esquerda, e /l/, à direita)

C ₁ C ₂	Inicial	Não inicial	C ₁ C ₂	Inicial	Não inicial
/pr/	prisma	lampreia	/pl/	plana	réplica
/tr/	trevo	mastro	/tl/	tlaspiáceas	atlas
/kr/	cru	ocre	/kl/	clima	reclame
/br/	brasa	abril	/bl/	blusa	emblema
/dr/	dreno	quadro	/dl/	—	—
/gr/	graça	regra	/gl/	glosa	sigla
/fr/	fruta	cifra	/fl/	flor	reflexão
/vr/	vrancelhas ²⁰	lavra	/vl/	—	—

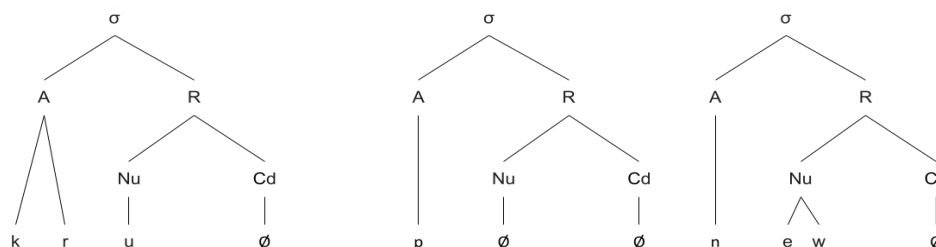
Fonte: Andrade & D’Andrade (2020: 3377)

Como a Tabela 8 comprova, apenas as oclusivas /p, b, t, d, k, g/ e as fricativas /f, v/ podem preencher a posição C₁ de um ataque ramificado com líquida em C₂. Pelo contrário, os grupos consonânticos que não respeitam estes princípios denominam-se grupos impróprios, como é o caso dos grupos formados por oclusiva + oclusiva (*e.g.*, [pt] ptério, [bd] bdélio), oclusiva + fricativa (*e.g.*, [ps] psicologia, [bs] absurdo) (Mateus & D’Andrade, 2000: 42-43), entre outros. Posto isto, se ambos os elementos de uma sequência possuírem o mesmo grau de sonoridade (*e.g.*, [mn]) e, se o segundo elemento não for uma consoante líquida (*e.g.*, [pt]), os dois elementos pertencentes a essa sequência não podem pertencer a um mesmo constituinte silábico. Quer isto dizer que o primeiro elemento, independentemente de ocorrer em ataque no início ou no interior

²⁰ À semelhança das palatais /ʎ/ e /ɲ/, também a ocorrência de /vr/ em ataque silábico inicial é bastante rara.

da palavra, faz-se acompanhar de um núcleo vazio ($\emptyset$)²¹. Comparemos as estruturas silábicas da palavra *cru*, que contém um grupo próprio, e da palavra *pneu*, que contém um grupo impróprio, representadas na Figura 4.

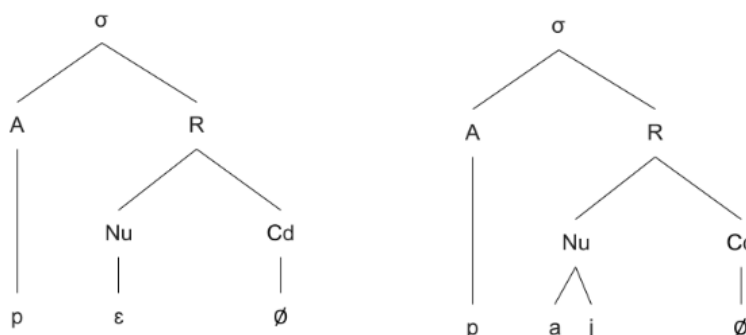
Figura 4 – Estruturas silábicas das palavras *cru* e *pneu*



A rima é um constituinte silábico constituído pelo núcleo e pela coda e pode ser: (i) não ramificada/simples (*e.g.*, pé); ou (ii) ramificada/complexa (*e.g.*, ter).

De igual modo, o núcleo é um constituinte silábico obrigatório e também ele pode ser: (i) não ramificado/simples (*e.g.*, pé); ou (ii) ramificado/completo (*e.g.*, pai). Em posição de núcleo, podem ocorrer todas as vogais orais e nasais do PE, como se comprova a partir das palavras *vi*, *lê*, *pé*, *má*, *da*, *de*, *mó*, *avô*, *tu*, *fim*, *pente*, *irmã*, *ponte* e *fundo* (Mateus *et al.*, 2003: 1044). No entanto, as glides [j] e [w] não podem constituir núcleo de sílaba, ocupando obrigatoriamente a posição de segundo elemento do núcleo ramificado à direita das vogais, com as quais formam ditongos decrescentes:

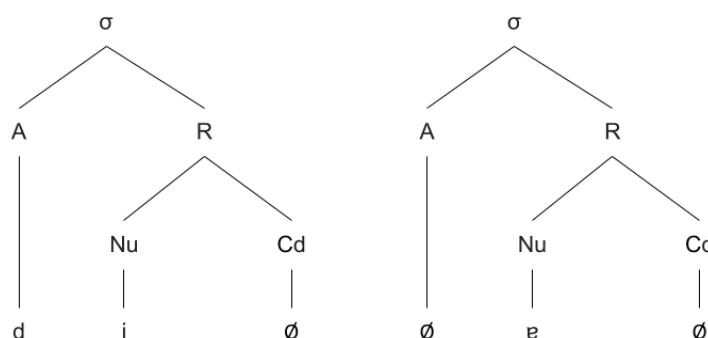
Figura 5 – Estruturas silábicas das palavras *pé* e *pai*



²¹ Atente-se, ainda assim, que uma sílaba com núcleo vazio pode ter a coda preenchida, nomeadamente pela fricativa [ʃ], de que são exemplo as palavras *obstar*, *subscrever* e *feldspato* (Andrade & D'Andrade, 2020: 3394).

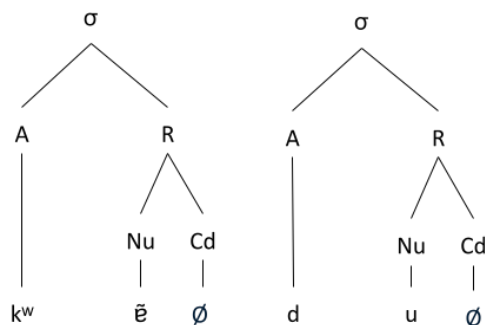
Por sua vez, uma sequência com dois elementos vocálicos em que o primeiro é alto e átono e o segundo não é alto corresponde sempre a dois núcleos, logo a duas sílabas independentes (Andrade & D’Andrade, 2020: 3378), ou seja, o PE não possui ditongos “crescentes” ao nível fonológico, como podemos observar a partir da estrutura silábica da palavra *dia* ilustrada na Figura 6.

Figura 6 – Estrutura silábica da palavra *dia*



Se, todavia, admitirmos a existência de sequências GV ao nível fonético do PE, é possível admitirmos a realização fonética [dje], na qual a glide fará sempre parte do ataque e nunca do núcleo (Mateus *et al.*, 2003: 1045). Se, contudo, assumirmos a existência de uma articulação secundária pertence às oclusivas velares labializadas /k^w/ e /g^w/ em palavras como *quando* e *água*, não estaremos perante um ataque ramificado (D’Andrade & Viana, 1993: 39; Freitas, 2001: 215; Andrade & D’Andrade, 2020: 3382), conforme podemos verificar na Figura 7.

Figura 7 – Estrutura silábica da palavra *quando*



No caso de ambas as vogais serem altas, a segunda poderá integrar o núcleo da sílaba antecedente, como ocorre nas palavras *viu* ['viw], *fui* ['fuj] e *cuidado* [kuj'dadu]. Contudo, nem todas as sequências cujo segundo elemento é [+alto] correspondem a um ditongo, pois pertencem a sílabas diferentes todas as sequências vocálicas (V.V), nas quais /i/ e /u/ são antecedidas de outra vogal, mas não são marcadas lexicalmente como não-acentuáveis e recebem o acento da palavra, como exemplificam as sequências [iu], [ui], [ei] (e.g., *miúdo*, *fluído*, *rainha*).

Para além do ataque e do núcleo, surge o constituinte silábico com mais restrições, a coda, correspondente à consoante à direita da vogal. Em coda silábica no interior de palavra, o PE admite as consoantes [r], [ʔ], [ʃ] e [ʒ] (e.g., *corpo* ['korpu], *volta* [vo'tar], *isto*, ['iftu]), *mesmo* ['mezmu]) e, adicionalmente a estas, admite ainda [n] (e.g., *abdómen* [eb'dɔmɛn]) e [s] em final de palavra, o último como segundo elemento das sequências fonéticas [ks] ou [ps], ainda que em casos muito excepcionais, provenientes de vocabulário muito especializado (e.g., *tórax* ['tɔraks], *bíceps* ['bisɛps]) (Andrade & D'Andrade, 2020: 3386).

1.2.4. Acento

O acento é uma propriedade das palavras de um conjunto significativo de línguas (línguas acentuais)²², entre as quais o português, que consiste na maior saliência de uma sílaba (tónica) relativamente às outras (átonas), resultante da conjugação das propriedades de duração, intensidade e frequência do som vocálico (Mateus, 1982b; D'Andrade & Bernard, 1992; Bisol, 1992; Mateus *et al.*, 1990; Pereira, 1999; Mateus & D'Andrade, 2000; Mateus *et al.*, 2003; 2005; Pereira, 2020). Para além disso, o acento é culminativo, no sentido em que cada palavra fonológica²³ tem um e apenas um acento tónico; é delimitativo, no sentido em que há uma tendência para se localizar nas extremidades da palavra e é também distintivo, por permitir distinguir significados entre as palavras (e.g., *explícito* / *explicito*).

Em português, o acento posiciona-se numa “janela de três sílabas” (Pereira, 1999: 98), i.e., recai sobre as últimas três sílabas da palavra. Consoante a sua localização,

²² As línguas acentuais, como a generalidade das línguas europeias, diferenciam-se das línguas tonais, como o tailandês ou o vietnamita.

²³ Sobre a palavra fonológica, ver Vigário (2003).

podemos dividir as palavras em três grupos²⁴: (i) oxítonas, quando o acento se localiza na última sílaba (*e.g.*, café, olá, etc.); (ii) paroxítonas, quando o acento se localiza na penúltima sílaba (*e.g.*, amarelo, amigo, etc.); (iii) proparoxítonas, quando o acento se localiza na antepenúltima sílaba (*e.g.*, sábado, abóbora, etc.).

Existem na literatura duas propostas alternativas acerca da natureza da distribuição do acento no português. Por um lado, certos autores defendem o valor fonológico do peso silábico na distribuição acentual, apoiando-se na Teoria Métrica²⁵ (Bisol, 1992; 1999; 2013; Wetzels, 1991; Veloso, 2017) e, por outro, a generalidade da literatura defende a existência de dois sistemas acentuais, determinados pela sua respetiva morfologia (Mateus, 1982b; D'Andrade & Bernard, 1992; Mateus *et al.*, 1990; 2003; 2005; Pereira, 1999; 2020; Mateus & D'Andrade, 2000)²⁶.

Partindo da afirmação de Bisol (1992: 26) que 78% das palavras terminadas em consoante são oxítonas e de Mateus & D'Andrade (2000:109), que referem que 70% das palavras não-verbais terminadas em vogal oral são paroxítonas, a segunda proposta considera relativamente regulares os seguintes padrões acentuais:

- as palavras que terminam em consoante, vogal nasal ou ditongo tendem a ser oxítonas (*e.g.*, chafariz, bacalhau, benjamim, etc.);
- as palavras cujas vogais finais, grafadas <a>, <o> e <e>, são realizadas como [e], [u] e [i], respetivamente, tendem a ser paroxítonas (*e.g.*, escada, sapato, parede).

Importa, contudo, salientar que, nesta perspetiva, tais padrões estão condicionados pela estrutura morfológica dos nomes e adjetivos, conduzindo à conclusão de que o acento regular incide sobre a vogal final do radical (Mateus & D'Andrade, 2000; Mateus *et al.*, 2003: 1050):

(1) paroxítonas: mesa MES]_{RADa}]_{IT}, lindo LIND]_{RNO}]_{IT}, leite LEIT]_{RADE}]_{IT}

(2) oxítonas: rapaz RAPAZ]_{RAD_}]_{IT}, rapazes RAPAZ]_{RADE}]_{ITS}]_{MP}

²⁴ Os termos oxítóna, paroxítóna e proparoxítóna equivalem respetivamente aos termos tradicionais aguda, grave e esdrúxula.

²⁵ Sobre a Teoria Métrica, ver Hayes (1981).

²⁶ Por questões de escopo, apenas iremos explicitar aqui os padrões acentuais das palavras não-verbais de radical simples, que possuem uma maior regularidade do que as palavras verbais (a forma infinitiva dos verbos insere-se nas palavras não-verbais).

Ainda que determinadas classes de palavras não-verbais, como pronomes, preposições, conectores, etc., não possuam a mesma estrutura morfológica dos nomes e dos adjetivos, a incidência do acento parece seguir os mesmos padrões acima referidos, pois as palavras terminadas em consoante, vogal nasal, ditongo ou vogal tendem a ser oxítonas (e.g., pois, assim, meu) e as palavras cujas vogais finais, grafadas <a>, <o> e <e> tendem a ser paroxítonas (e.g., porventura, contudo, bastante) (Pereira, 2020: 3410).

Apesar de estas regras ditarem o panorama acentual geral das palavras não-verbais do português, existem determinadas estruturas que escapam a estes padrões, as quais denominámos de acentualmente irregulares (Mateus & D'Andrade, 2000). Estas irregularidades permitem concluir que a localização do acento não é exclusivamente morfológica, podendo também depender da informação lexical (Pereira, 1999: 123). Por serem acentuadas na penúltima vogal do radical, estas palavras podem ser paroxítonas ou proparoxítonas e, apesar de serem tipicamente grafadas com acento, como podemos observar nos exemplos (3) – (4), nem sempre o são (Pereira, 2020: 3407):

(3) paroxítonas: útil, caráter, ónix, sótão, táxi

(4) proparoxítonas: sílaba, plástico, índole

Dado que o sistema de conversão aqui desenvolvido parte da ortografia, é relevante considerarmos a acentuação gráfica das palavras não-verbais acentualmente irregulares. Partindo das Bases VIII, IX, X e XI do *Acordo Ortográfico da Língua Portuguesa* (AO90)²⁷, são grafadas com acento agudo: (i) oxítonas terminadas nas vogais tónicas abertas grafadas -a, -e ou -o, seguidas ou não de -s (e.g., bebé); (ii) oxítonas com mais de uma sílaba terminadas no ditongo nasal grafado -em (e.g., ninguém); (iii) oxítonas com ditongos abertos grafados -éi, -éu ou -ói, podendo os dois últimos ser seguidos ou não de -s (e.g., céu); (iv) paroxítonas cuja vogal tónica é uma vogal aberta grafada a, e, o e ainda i ou u e que terminam em -l, -n, -r, -x e -ps (e.g., açúcar) e que terminam em -ã(s), -ão(s), -ei(s), -i(s), -um, -uns ou -us (e.g., órfão); (v) as vogais tónicas grafadas i e u das palavras oxítonas e paroxítonas quando antecedidas de uma vogal com a qual não foram ditongo (e.g., país); (vi) as vogais tónicas grafadas i e u quando

²⁷ Ver <https://vocabulario.acad-ciencias.pt/index.php/ortografia/texto-integral-do-ao90>.

precedidas de ditongos (*e.g.*, *teiú*); (vii) as proparoxítonas cuja vogal tónica é uma vogal aberta grafada *a*, *e*, *o* e ainda *i* ou *u* (*e.g.*, *pássaro*) ou ditongo oral (*e.g.*, *cáustico*) e que terminam em sequências vocálicas (*e.g.*, *área*); e (viii) as proparoxítonas cujas vogais tónicas grafadas *e* ou *o* ocorrem em final de sílaba seguidas das consoantes *m* ou *n* (*e.g.*, *académico*).

Adicionalmente, são grafadas com acento circunflexo: (i) oxítonas terminadas nas vogais tónicas fechadas que se grafam *-e* ou *-o*, seguidas ou não de *-s* (*e.g.*, *avô*); (ii) paroxítonas cuja vogal tónica é uma vogal fechada grafada *a*, *e*, *o* e que terminam em *-l*, *-n*, *-r*, *-x* e *-ps* (*e.g.*, *têxtil*) e que terminam em *-ão(s)*, *-eis*, *-i(s)* ou *-us* (*e.g.*, *bênção*); (iii) as proparoxítonas cuja sílaba tónica apresenta vogal fechada ou ditongo com vogal fechada (*e.g.*, *sôfrego*) e que terminam em sequências vocálicas (*e.g.*, *amêndoa*).

1.3. Interface fonologia-ortografia

1.3.1. Relação entre grafema e fone[ma]

Nas vertentes mais tradicionais da linguística, nomeadamente nas escolas estruturalistas (Saussure, 1916; Gleason, 1985, Martinet, 1960, entre outros), há a tendência para adotar uma visão fonocêntrica (Veloso, 2003a; 2005a; 2019; 2020), que separa a ortografia e a fonologia enquanto entidades ontológicas e epistemológicas completamente distintas, assumindo a escrita como uma representação meramente secundária e acessória aos objetos linguísticos. Entre outros aspetos, a visão fonocêntrica assenta no facto de a escrita poder (i) espelhar, de forma mais ou menos transparente, o sistema fonológico da língua, (ii) preservar certas relações morfofonológicas estabelecidas dentro do léxico e presentes na forma subjacente das palavras e (iii) preservar certos aspetos etimológicos sem representação fonética sincrónica nas palavras (Veloso, 2003a: 140-146; Veloso, 2005a: 3-4).

Em contrapartida, surge uma outra perspetiva mais relacionada com a psicolinguística, que, apesar de não negar a dissimilitude da relação entre as formas linguísticas e as representações gráficas, prevê uma maior integração entre estas, admitindo até que o conhecimento ortográfico poderá condicionar em certos aspetos o conhecimento fonológico (Coulmas, 2003; Daniels, 2010, entre outros). Por outras palavras, esta proposta admite a inclusão das representações ortográficas no conjunto

de objetos da linguística através da identificação de relações mais ou menos sistemáticas entre a ortografia e, por exemplo, a fonologia (Veloso, 2022: 68).

Independentemente da proposta adotada, é relevante fazer uma distinção, nas línguas que recorrem à escrita alfabética, entre línguas opacas e línguas fonemicamente transparentes. De acordo com Veloso (2005a: 8), as primeiras distinguem-se por não possuírem uma correspondência direta e isomórfica entre os grafemas e os fone[mas] da língua (*e.g.*, o inglês), enquanto as segundas se caracterizam por uma correspondência regular, sistemática e biunívoca entre os segmentos fonológicos e os símbolos gráficos discretos. O PE aproxima-se deste último grupo, acondicionando um alinhamento mais estreito entre a fonologia/fonética e a ortografia, especialmente nas consoantes (Veloso, 2022: 70):

<p> → [p] (pá ['pa])	<z> → [z] (gozo ['gozu])
 → [b] (boa ['boe])	<z> → [ʃ] (paʒ ['paʃ])
<t> → [t] (tu ['tu])	<ch> → [ʃ] (chá ['ʃa])
<d> → [d] (dar ['dar])	<x> → [ʃ] (xá ['ʃa])
<c ^{a,o,u} > → [k] (casa ['kaze])	<x> → [s] (sintaxe [sĩ'tasi])
<qu ^{e,i} > → [k] (quero ['keru])	<x> → [z] (êxodo ['ezudu])
<qu ^{a,o} > → [kʷ] (quarto ['kʷartu])	<x> → [ks] (táxi ['taksi])
<g ^{a,o,u} > → [g] (gato ['gatu])	<g ^{e,i} > → [ʒ] (gingar [ʒĩ'gar])
<gu ^{e,i} > → [g] (guita ['gite])	<j> → [ʒ] (jato ['ʒatu])
<gu ^{a,e,i,o} > → [gʷ] (guardar [gʷer'dar])	<l> → [l] (maia ['male])
<ss> → [s] (passo ['pasu])	<l> → [t] (maɫ ['mat])
<ss> → [ç] (paço ['pasu])	<lh> → [ʎ] (alho ['aʎu])
<c ^{e,i} > → [s] (cego ['segu])	<r> → [r] (caro ['karu])
<s> → [s] (saber [se'ber])	<r> → [r] (rato ['ratu])
<s> → [z] (casa ['kaze])	<rr> → [ʀ] (carro ['karu])
<s> → [ʃ] (pás ['paʃ])	

Por sua vez, a correspondência entre os grafemas e os segmentos vocálicos orais do PE é a seguinte (adaptado de Mateus & D'Andrade, 2000: 6-7):

<a> → [a] (caso ['kazu])	<e> → [ɛ] (seco ['seku])
<a> → [e] (café [ka'fe])	<e> → [e] (telha ['teʎe])
<e> → [ɛ] (beio ['beiu])	<e> → [i] (dever [di'ver])

<e> → [i] (cear [si'ar])

<i> → [i] (tirar [ti'rar])

<i> → [j] (pai ['paj])

<o> → [ɔ] (bola ['bɔle])

<o> → [o] (força ['forse])

<o> → [u] (pato ['patu])

<u> → [u] (tudo ['tudu])

<u> → [w] (pau ['paw])

Em ataque silábico, os grafemas <m>, <n> e <nh> correspondem a consoantes nasais; no entanto, quando surgem em coda silábica (seguidos de vogal, antes de consoante ou em fronteira final de palavra) correspondem a marcadores de nasalização:

<am> → [ẽ] (campina [kẽ'pine])

<am> → [ẽw̃] (falam ['falẽw̃])

 → [ẽ] (embora [ẽ'bɔre])

 → [ẽj̃] (viagem [vi'aʒẽj̃])

<om> → [õ] (pompom [põ'põ])

<um> → [ũ] (um ['ũ])

<im> → [ĩ] (assim [e'sĩ])

<an> → [ẽ] (andar [ẽ'dar])

<en> → [ẽ] (pensar [pẽ'sar])

<in> → [ĩ] (pingar [pĩ'gar])

<on> → [õ] (contar [kõ'tar])

<un> → [ũ] (untar [ũ'tar])

Como podemos inferir, a transparência ou opacidade de uma língua deve ser percecionada num *continuum* e, apesar de o PE ser considerado fonemicamente transparente, possui determinados aspetos de cariz mais opaco como a motivação contextual de certas correspondências (*e.g.*, <c> com o valor de [s]); a existência de dígrafos, em que dois grafemas correspondem a um segmento fonológico (*e.g.*, <lh> /ʎ/); a redução átona e a harmonização vocálica a que estão sujeitas as vogais (Veloso, 2005a: 14; Mateus, 2006: 117; Veloso, 2022: 72), o que evidencia a complexidade da relação entre a fonologia e a ortografia.

1.3.2. Sistemas de representação

A ferramenta de notação responsável pela conversão dos sinais de fala contínuos em unidades mínimas, discretas e mensuráveis é a transcrição. Este é um mecanismo de representação da componente sonora das línguas e que subentende uma relação biunívoca entre os segmentos sonoros e os respetivos símbolos (Emiliano, 2009: 6). Dos vários métodos de representação da fala humana, destacámos os sistemas monotípicos, que utilizam um conjunto limitado de símbolos que se podem assemelhar aos grafemas e que pretendem representar todos os sons de todas as línguas existentes no mundo (Barroso, 1999: 105). Neste tipo de sistema, insere-se o AFI, um conjunto esquematizado de símbolos inspirados no alfabeto romano, com caracteres e diacríticos especiais,

destinado a representar de forma simples, sistemática e biunívoca os sons das línguas (IPA, 1999: 3) e que se distingue dos restantes sistemas de representação pela sua universalidade e inventário relativamente reduzido e pela facilidade do seu processamento computacional. Para além disto, possui diferentes aplicações e propósitos, nomeadamente a descrição do inventário fonológico de determinado sistema linguístico e das respetivas variações fonéticas, a constituição da base para um sistema de escrita de uma língua, a anotação de características acústicas de uma pronúncia com um alto nível de precisão.

A transcrição pode ser seccionada em três categorias: (i) a transcrição fonética geral; (ii) a transcrição alofónica; e (iii) a transcrição fonémica (Abercrombie, 1964: 35-36; Laver, 1994: 549-561; Spencer, 1996: 9; IPA, 1999: 28-30; Heselwood, 2013: 144-157). As duas primeiras inserem-se ao nível da transcrição estreita, e a última, ao nível da transcrição larga.

A transcrição fonética geral é tipicamente empregada nas primeiras fases de uma análise fonológica de uma língua desconhecida pelo transcritor (IPA, 1999: 28) e o elevado nível de meticulosidade a que se encontra sujeita situa-a no extremo do espectro das transcrições fonéticas estreitas. Por sua vez, a transcrição alofónica possibilita a representação das diferentes realizações fonéticas de um fonema nos variados ambientes contextuais e estruturais. Todavia, é possível (e costume) ser-se seletivo em relação à informação a incluir neste tipo de transcrição (IPA, 1999: 28), i.e., o foco de interesse da transcrição é selecionado e expresso por meio de um conjunto de convenções que, por sua vez, determinam o seu grau de estreiteza ou largura (*e.g.*, podemos escolher incluir ou não a fricativização das oclusivas).

Por fim, a transcrição fonémica caracteriza-se pelo seu maior nível de abstração, ao fornecer uma representação única para cada unidade linguística de uma língua (Laver, 1994: 551), não atendendo à sua variação contextual. Convencionalmente, um fonema é representado entre barras oblíquas // e corresponde à sua realização fonética mais generalizada (Veloso, 2015: 10). São estas formas canónicas que se observam tipicamente nos lemas de um dicionário, uma vez que se trata de palavras individuais,

independentes de fronteiras lexicais (à exceção de compostos) e em contexto formal (Laver, 1994: 551).

Concluída a análise da fonologia do português europeu, passamos agora a observar de que forma esta se reflete na representação lexicográfica da língua portuguesa.

2. Tratamento lexicográfico da pronúncia nos dicionários de língua portuguesa

Este capítulo apresenta uma análise diacrónica do tratamento lexicográfico da pronúncia nos dicionários monolingues de língua portuguesa, tanto em formato impresso (secção 2.1.) como digital (secção 2.2.), com especial destaque para o DLP-ACL (secção 2.3.). A abordagem adotada visa compreender até que ponto a dimensão fonética tem sido valorizada ou secundarizada na tradição lexicográfica portuguesa.

Embora os dicionários monolingues tenham desempenhado um papel essencial na fixação e documentação da língua portuguesa, a representação da pronúncia revelou-se, ao longo do tempo, irregular e frequentemente marginal. Desde as primeiras obras modernas – como as de BACELAR (1783), MORAIS SILVA (1789) e da própria ACL (1793) – até aos recursos lexicográficos contemporâneos (DLP-ACL, 2025; PORTO EDITORA, 2025; PRIBERAM, 2025, entre outros), a informação fonética surge de forma esporádica, muitas vezes restrita a casos considerados problemáticos, sem nunca se afirmar como uma prática sistemática — mesmo perante os avanços tecnológicos e editoriais do século XXI.

2.1. Dicionários impressos de língua portuguesa

A análise histórica do tratamento da pronúncia nos dicionários impressos de língua portuguesa permite distinguir dois grandes momentos: o período anterior ao século XX e o que se lhe segue. O Anexo 1 sintetiza os dicionários impressos representativos dos séculos XVIII e XIX, destacando os elementos fonéticos considerados.

Os primeiros dicionários monolingues modernos do português datam do final do séc. XVIII e, embora tenham representado um marco significativo na lexicografia da

língua portuguesa, a única informação relativa à pronúncia que incluem é uma prática já adotada no *Vocabulário Português e Latino* (BLUTEAU, 1712), que consiste na inclusão de indicações sobre a pronúncia de cada grafema no seu lema principal. Este parâmetro parece ter-se integrado na tradição lexicográfica do português, estendendo-se a alguns dicionários do séc. XIX e do séc. XX, nomeadamente ao dicionário de MORAIS SILVA (1813; 1949-1959), ao de Caldas Aulete (AULETE, 1881) e, de forma não sistemática, ao dicionário de Faria (FARIA, 1849), que apenas o faz em determinados verbetes como, por exemplo, a letra C:

C, s.m. terceira letra do alphabeto na lingua portugueza e em todas as outras filhas da latina, e uma das consoantes. O *c*, antes de *a*, *o*, *u* sã como *k*; antes de *e* ou *i* como *s*. Quando se lhe quer dar som de *s* antes de *a*, *o* ou *u* é preciso ajuntar-lhe uma cedilha, como se vê na palavra *cabeça* (FARIA, 1849: 607)

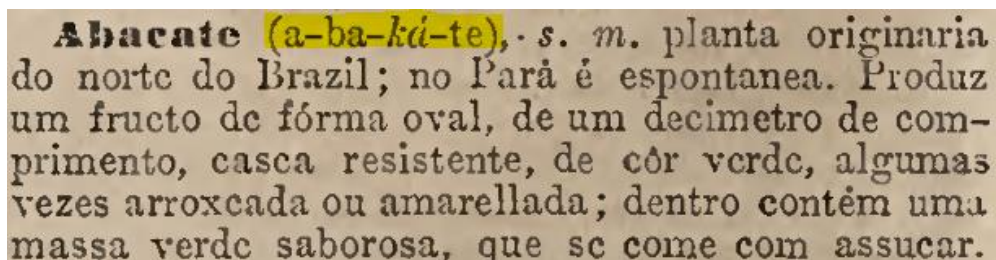
A segunda metade do séc. XIX representou um momento de viragem na produção lexicográfica da língua portuguesa com a compilação e publicação de grandes dicionários motivados pelo “espaço editorial, científico e nomeadamente modesto investimento filológico no âmbito dos estudos da língua portuguesa” (Verdelho, 2002: 31); contudo, o tratamento lexicográfico da pronúncia não acompanhou essa evolução.

Dos principais dicionários de língua portuguesa do séc. XIX, destaca-se o *Diccionario Contemporaneo da Lingua Portuguesa* de Caldas Aulete (AULETE, 1881), não só pela sua abundante combinatória lexical, fraseologia, exemplos e abonações recolhidos (Iriarte Sanróman, 2016: 4), mas também pelo estatuto concedido à “pronúncia” como uma das quatro componentes essenciais do verbe, a par da formação, ortografia e significação.

O prólogo da obra inclui uma secção específica sobre a pronúncia, na qual se expõem os critérios adotados, nomeadamente a marcação dos acentos tónicos e o tratamento de vocábulos científicos. Estas tarefas foram confiadas ao “saber dos distintos glotologos, os srs. Antonio J. Viale e Gonçalves Viana” (AULETE, 1881: 21). Conforme ilustrado na Tabela 9, este dicionário apresenta uma tentativa pioneira de transcrição fonética aproximada por meio de grafemas e sinais diacríticos, assinalando a tonicidade (ver *itálico* na Figura 11), nasalidade (*e.g.*, ampliado (an-pli-á-du)) e abertura vocálica (*e.g.*, abandono (a-ban-dô-nu)), assim como os diferentes valores das

consoantes (e.g., abastecer (a-bas-te-ssêr), agacho (a-gha-xo), agenciador (a-jên-ssi-a-dôrr)) e ditongos (e.g., afinagem (a-fi-ná-jan-e)). Adicionalmente, compreende a divisão silábica, feita através de hífenes²⁸, como é possível observar na Figura 8.

Figura 8 – Lema *abacate* do dicionário de AULETE (1881)



Abacate (a-ba-ká-te), . s. m. planta originaria do norte do Brazil; no Pará é espontanea. Produz um fructo de fôrma oval, de um decimetro de comprimento, casca resistente, de côr verde, algumas vezes arroxcada ou amarellada; dentro contém uma massa verde saborosa, que se come com assucar.

Fonte: AULETE (1881: 2)

Por fim, esta obra fornece ainda um compêndio de esclarecimentos acerca desta transcrição, dedicando uma secção do prólogo aos grafemas e diacríticos utilizados na representação dos sons e aos critérios empregados na sua configuração.

Embora nenhum dicionário do Anexo 1 apresente um tratamento tão completo da pronúncia como o de AULETE (1881), todos os dicionários dedicam uma determinada secção à pronúncia, à exceção do dicionário de Faria (FARIA, 1849), cujo prólogo apenas contém um breve parágrafo sobre “maneira de partir as palavras” (FARIA, 1849: 39), apesar de não apresentar a divisão silábica de nenhum dos lemas. Por sua vez, o prefácio do *Grande Diccionario Portuguez* de Domingos Vieira (VIEIRA, 1871-1874) conta com um longo capítulo sobre a pronúncia do português, redigido por Adolfo Coelho, à semelhança do que ocorre no dicionário de FIGUEIREDO (1899), que inclui adicionalmente uma secção sobre as variedades dialetais portuguesas.

Apesar destes avanços pontuais, a prática de registar sistematicamente a pronúncia nos dicionários portugueses não se consolidou durante o século XX. Verdelho (2002: 36) sublinha que a anotação fonética dos lemas permaneceu residual, sendo retomada de forma mais consistente apenas com a publicação dos *Vocabulários Ortográficos* da ACL (ACL, 1940; 47; 70) e da Academia Brasileira de Letras (ABL, 1943) e do *Vocabulário da Língua Portuguesa* de Rebelo Gonçalves (1966). Estas obras,

²⁸ Atente-se que se trata da divisão silábica e não da translineação dos lemas, como podemos inferir a partir de palavras como *abairrar* (a-bai-rrár).

influenciadas pelo trabalho ortoépico²⁹ de Gonçalves Viana (1912), abriram caminho à introdução de indicações ortoépicas em vários dicionários posteriores, ainda que de forma limitada e não sistemática.

Na generalidade dos dicionários registados no Anexo 2, a ortoépia encontra-se encerrada entre parênteses curvos e cinge-se aos casos cuja pronúncia suscita dúvidas como:

- vogais tónicas ou pretónicas médias não acentuadas com diacrítico (*e.g.*, verbete (ê); abonador(ô));
- som da letra x (*e.g.*, xícara (ch), anexo (cs), exame (z), aproximar (ss), contexto (s));
- palavras homónimas homógrafas não homófonas (*e.g.*, bordo (ô) vs. bordo (ó));
- palavras com pronúncias flutuantes (*e.g.*, grumete (ê ou é));
- dígrafos *gu* e *qu*, seguidos de *e* e *i*, nos quais *u* é proferido (*e.g.*, aguentar (gü));
- encontros vocálicos heterossilábicos (*e.g.*, abaular (ba-u)).

Certos dicionários como o de AULETE (1987) e o de FIGUEIREDO (1996) passaram ainda a incluir aproximações fonéticas de estrangeirismos, transcritos de acordo com a pronúncia aportuguesada ou adaptada e através de grafemas e, em alguns casos no dicionário de AULETE (1987), acompanhados da divisão silábica (*e.g.*, aal (a-ál)).

Para além da inclusão de notas ortoépicas, a atenção dedicada à pronúncia nos prólogos destas obras aumentou significativamente. Dos dicionários analisados do séc. XX, o que mais se destaca neste aspeto é o dicionário AURÉLIO (1986), com uma longa introdução acerca das consoantes e vogais da língua portuguesa, da sua organização, do acento e da estrutura silábica, apresentando ainda uma tabela do AFI.

O séc. XXI assinalou o aprimoramento dos parâmetros adotados nos dicionários do séc. XX através da crescente integração de recursos fonéticos mais precisos e padronizados, embora não se tenha verificado unanimemente em todos os dicionários, conforme podemos observar no Anexo 3. Esta evolução manifestou-se, por exemplo, na inédita utilização do AFI na transcrição fonética dos estrangeirismos nos dicionários

²⁹ Por ortoépia entende-se o conjunto das regras que definem a pronúncia aproximada da variedade-padrão.

AURÉLIO (1999) (*e.g.*, *res nullius* [res 'nullɐs]) e HOUAISS (2002-2003) (*e.g.*, *dé-jà vu* [de'ʒa vy]), adaptados à língua de origem e incluindo ainda o acento de intensidade (acento tónico), o acento melódico, o tom lexical e a duração, quando aplicáveis.

A maior inovação registou-se, todavia, no DLPC (2001) e no da PORTO EDITORA (2004), que passam a incluir a transcrição fonética de todos os seus lemas. Uma vez que os critérios do primeiro se encontram analisados em 3.1., passamos à observação dos parâmetros de transcrição do segundo. Tal como é indicado nas suas páginas iniciais, os símbolos utilizados na transcrição fonética correspondem aos do AFI, à exceção da vogal átona, que ocorre no final de palavras como *doce* ou *contente*, representada com [ə] (*e.g.*, *doce* ['dos(ə)]) em vez de [ɨ], sendo seguida a “pronúncia neutra da língua portuguesa não marcada diatopicamente e socialmente bem aceite pela comunidade linguística” (PORTO EDITORA, 2004: 7). Esta transcrição é feita entre parênteses retos [] e inclui a marcação da sílaba tónica através do símbolo '. Entre outros parâmetros, destacámos:

- a exclusão da fricativização das oclusivas (*e.g.*, *aba* ['abe]; *agora* [e'gɔrɐ]; *dado* ['dadu]);
- a manutenção da velarização do <l> em coda silábica (*e.g.*, *mal* ['matɫ]);
- a manutenção de ditongos crescentes quando /i/ e /u/ são átonas e se encontram seguidas de vogal (*e.g.*, *família* [fe'milje] vs. *alegria* [elə'gɾie]);
- a inclusão do ditongo crescente [we], no caso de *qu* e *gu* seguidos de /a/ átono (*e.g.*, *água* ['agwe]).

Adicionalmente, este dicionário opta por encerrar entre parênteses retos a vogal átona [ə] em fronteira final de palavra (*e.g.*, *tarde* ['tard(ə)]), alvo de uma elisão contextual que ocorre de forma generalizada, e a semivogal [w] pertencente ao ditongo oral [ow] (*e.g.*, *pouco* ['po(w)ku]), cuja realização é tipicamente circunscrita aos dialetos centro-meridionais e insulares do PE. Por fim, estes critérios são ainda aplicados à transcrição dos estrangeirismos, que se encontram adaptados aos sons do português.

2.2. Dicionários digitais de língua portuguesa

Com o advento da era digital e a modernização das ferramentas lexicográficas, seria de esperar uma evolução significativa no tratamento fonético dos lemas. No entanto, constata-se um cenário algo paradoxal: a informação fonética encontra-se muitas vezes ausente ou é disponibilizada de forma fragmentária e não sistemática. Atualmente, apenas dois dos dicionários digitais de língua portuguesa – Infopédia (PORTO EDITORA, 2025) e INFORMAL (2025) – disponibilizam a transcrição fonética dos lemas, como podemos observar a partir do Anexo 4, que apresenta o panorama geral do estado da pronúncia nos dicionários digitais de língua portuguesa que se encontram disponíveis publicamente até à data para consulta.

De momento, a Infopédia (PORTO EDITORA, 2025) é o único dicionário digital do português que oferece a transcrição fonética dos lemas em PE. Embora não haja até à data qualquer tipo de informação aberta sobre os critérios adotados nesta transcrição, é possível constatar uma clara continuidade em relação ao modelo do *Grande Dicionário da Língua Portuguesa* (PORTO EDITORA, 2004). Contudo, é possível observar algumas diferenças na transcrição atual, nomeadamente: (i) a omissão dos parênteses retos; (ii) a representação da vibrante que surge em palavras como *caro* e *flor* como vibrante múltipla [r] (e.g., *caro* ['karu], *flor* ['flor]) e não como vibrante simples [r], utilizada na versão de 2004.

Por sua vez, o INFORMAL (2025) é o único dicionário de língua portuguesa que disponibiliza a transcrição fonética dos lemas em português do Brasil, incluindo ainda a marcação da sílaba tónica e a divisão silábica. Tal como na Infopédia (PORTO EDITORA, 2025), não são apresentados os critérios e caracteres utilizados, sendo possível apenas inferir algumas características da transcrição a partir da observação dos lemas. No entanto, por se centrar exclusivamente na variedade brasileira, este recurso não é integrado na presente investigação.

Para além da Infopédia (PORTO EDITORA, 2025), também o INFORMAL (2025), o HOUAISS (2025) e o MICHAELIS (2025) apresentam a transcrição fonética dos estrangeirismos, sendo que este último se faz acompanhar de uma secção detalhada sobre os critérios e caracteres fonéticos adotados na sua transcrição, nomeadamente o

encerramento da transcrição entre parênteses retos [] e a adaptação da pronúncia ao modelo estrangeiro, sem ter em conta as variedades regionais e a utilização do API (2005).

Outro dos parâmetros relativos à pronúncia que podemos observar nos dicionários digitais de língua portuguesa é o registo da ortoépia que se encontra presente, de forma mais ou menos homogénea, nos dicionários AULETE (2025), DLP-ABL (2025), HOUAISS (2025), PORTO EDITORA (2025) e PRIBERAM (2025). Como pudemos verificar anteriormente, esta característica parece ter sido adotada dos *Vocabulários Ortográficos* pois tem tendência a recair sobre os mesmos casos problemáticos da pronúncia do português, designadamente as vogais médias tónicas /o/ e /ɔ/ e /e/ e /ɛ/ e as várias realizações fonéticas do grafema <x>, /s/, /ʃ/, /z/, etc. No caso da Infopédia (PORTO EDITORA, 2025), por exemplo, as indicações ortoépicas das vogais ocorrem maioritariamente nas palavras homógrafas, embora a sua representação não seja uniforme (e.g., corte /ô/ vs. corte /ó/, sede /sé/ vs. sede /sê/), e nas palavras com vogais tónicas semifechadas /o/ e /e/, deixando de lado as palavras com vogais tónicas semiabertas /ɔ/ e /ɛ/ (cf. flor /ô/ e flora, mesa /ê/ e belo). Além disso, as indicações ortoépicas encontram-se representadas de diferentes formas nos variados dicionários:

- entre barras oblíquas / / (PORTO EDITORA, 2025);
- entre parênteses retos [] (AULETE, 2025);
- entre parênteses curvos () (DLP-ABL, 2025);
- entre barras verticais | | (PRIBERAM, 2025);
- sem qualquer delimitação (HOUAISS, 2025).

Excecionalmente, encontra-se ainda o registo da ortoépia na palavra *amor*, no INFORMAL, e na palavra *táxi*, no DP.

A marcação da sílaba tónica ocorre em quatro dos doze dicionários disponíveis, podendo ser assinalada pela própria transcrição fonética (PORTO EDITORA, 2025, INFORMAL, 2025), pelo itálico (AULETE, 2025; DLP-ABL, 2025) ou pelo sublinhado (LÉXICO, 2025). O tratamento silábico dos lemas parece constituir o parâmetro mais presente na maioria dos dicionários. Apesar do DICIO (2025) e do DP (2025) afirmarem que possuem uma “separação silábica”, a divisão do lema *carro* como car.ro indica que

estamos perante uma translineação e não propriamente uma divisão silábica dos lemas, sendo que o único dicionário digital português que aparenta fazer uma verdadeira divisão silábica é o INFORMAL (2025).

Por fim, a Infopédia (PORTO EDITORA, 2025) é o único dicionário que fornece a gravação sonora de determinados lemas em PE (*e.g.*, azul, claro, frio, gato), embora esta funcionalidade não se encontre disponível para todos os lemas. Este dicionário é, deste modo, o único dicionário de língua portuguesa que dispõe simultaneamente da transcrição fonética e da gravação sonora dos lemas, já que os restantes dicionários que contemplam a gravação sonora na variedade do PB (DICIO, 2025; HOUAISS, 2025)) não contemplam a transcrição dos lemas.

Como é possível observar, a prática de incluir informação sobre a pronúncia foi bastante esporádica na tradição lexicográfica portuguesa e, paradoxalmente, a transição digital não se traduziu numa evolução significativa, revelando uma subvalorização histórica deste registo nos dicionários monolingues de língua portuguesa.

2.3. O Dicionário da Língua Portuguesa da Academia das Ciências de Lisboa

O *Dicionário da Língua Portuguesa* da Academia das Ciências de Lisboa (DLP-ACL) constitui o foco desta investigação, pelo que é acrescidamente importante analisar a evolução do registo da pronúncia nesta obra. Posto isto, o Anexo 5 apresenta todas as edições anteriores do DLP-ACL.

Embora incompleto, abrangendo apenas a letra A, o primeiro volume do *Dicionário da Língua Portuguesa* foi publicado em 1793, destacando-se enquanto primeiro dicionário académico português e marcando o início de um esforço sistemático de compilação e estudo da língua portuguesa num contexto académico. Na sua composição, estiveram envolvidos Pedro José da Fonseca, responsável pelos textos introdutórios, Agostinho José de Costa de Macedo, responsável pelo catálogo dos autores, e Bartolomeu Inácio Jorge (Verdelho & Silvestre, 2007: 25). Apesar de se apresentar como uma obra lexicográfica moderna, incluindo informação gramatical, variantes ortográficas e abonações nos seus verbetes, este dicionário, no que respeita à

transcrição fonética, é muito similar aos dicionários de língua portuguesa que foram surgindo no séc. XVIII. Na sua introdução, é possível encontrar uma secção dedicada aos critérios ortográficos utilizados, com base na pronúncia da língua portuguesa:

Aquellas porém onde a duplicação das consoantes, ou a troca das vogaes não produz hum tal effeito como succede em muitas vozes, que se escrevião por *e* e *o*, em lugar de *i* e *u*, e ao contrario; ou por *f* e *am* em lugar de *ph* e *ão* (ACL, 1793: 15).

No entanto, não é dada qualquer atenção à pronúncia dos lemas em si, excluindo qualquer tipo de transcrição fonética, marcação de sílaba tónica ou divisão silábica/translineação. A meados e finais do séc. XX, deu-se a publicação de três *Vocabulários Ortográficos da Língua Portuguesa* da Academia de Ciências de Lisboa (1940; 47; 90), que, conforme referido anteriormente, introduziram informações ortoépicas. Ainda que estas obras tenham influenciado alguns dicionários a prestar maior atenção à pronúncia, a edição do dicionário da ACL (1976), coordenada por Jacinto Prado Coelho, não foi mais longe do que a primeira em todos os aspetos referidos, ocupando-se apenas da letra A. Contudo, é no ano de 2001 que, sob a alçada de João Malaca Casteleiro, o DLP recebe uma edição completa intitulada *Dicionário da Língua Portuguesa Contemporânea* (DLPC), dividida em dois volumes, o primeiro de A a F e o segundo de G a Z. Para além de manter toda a informação previamente incluída nas entradas lexicais, o DLPC acrescenta-lhes a sua transcrição fonética “segundo a norma culta, aproximada, de Lisboa e Centro do país” (DLPC, 2001: 18), cujos critérios e caracteres se encontram discriminados no prefácio desta edição.

A atualização e revisão da edição de 2001 resultou no DLP-ACL, um dicionário digital de acesso público, disponibilizado em abril de 2023 e composto atualmente por mais de 103 000 entradas lexicais. Apesar de inicialmente ter sido empreendido um esforço em prol da inclusão da transcrição fonética dos lemas na versão *online* do dicionário, adaptando o que fora elaborado nos dois volumes impressos de 2001, esta não se encontra atualmente disponível.

Concluído o levantamento do tratamento lexicográfico da pronúncia nos dicionários monolingues de língua portuguesa, debruçamo-nos agora sobre os sistemas de conversão que procuram representar essa dimensão fonológica/fonética de forma sistemática.

3. Sistemas de conversão

Esta secção reúne uma síntese crítica dos principais sistemas de conversão grafema-fonema disponíveis para o PE, nomeadamente o *Grafone* (secção 3.1.) e o *Convert Text to IPA Transcription* (secção 3.2.), contextualizando o desenvolvimento do sistema proposto nesta investigação.

Os sistemas de conversão grafema-fonema ou, como são geralmente conhecidos, os sistemas G2P (do inglês *grapheme to phoneme*), destinam-se à transcrição automática de texto escrito para uma sequência de símbolos fonéticos.

Os esforços delegados à elaboração destes sistemas para a língua portuguesa têm resultado no desenvolvimento de variados algoritmos e modelos, suportados em diferentes abordagens teóricas, entre as quais se destacam, entre outros, os modelos baseados em regras linguísticas (Oliveira *et al.*, 1992; Caseiro *et al.*, 2002; Teixeira, 2004; Braga *et al.*, 2006), em dados estatísticos (Trancoso *et al.*, 1994; Oliveira *et al.*, 2001; Barros & Weiss, 2006) e numa junção híbrida dos dois (Candeias & Perdigão, 2011). No entanto, os únicos recursos de conversão grafema-fonema de língua portuguesa que se encontram atualmente disponíveis para acesso público são o *Grafone* e o *Convert Text to IPA Transcription*, para o PE, e o *Petrus*, para o PB.³⁰

As Tabelas 9 e 10 apresentam uma síntese comparativa das principais ferramentas de conversão grafema-fonema disponíveis para o português europeu, destacando as respetivas características gerais e os critérios fonético-fonológicos adotados.

Tabela 9 – Características das ferramentas de conversão grafema-fonema do PE

	Informação sobre a conversão	Palavras homógrafas	AO90	Divisão silábica	Acento
<i>Grafone</i>	Sim*	Sim	Sim	Não	Sim

³⁰ Para o PB, surge ainda o *Potígrafone* (Costa Carvalho, 2017), destinado à conversão grafema-fonema da variedade potiguar; contudo, foi excluído por não se encontrar de momento publicamente disponível.

<i>Convert Text to IPA Transcription</i>	Sim	Não	Não	Sim	Sim
--	-----	-----	-----	-----	-----

Tabela 10 – Critérios de transcrição das ferramentas de conversão grafema-fonema do PE

	Velarização do [l]	Oclusivas fricativizadas	Modificação do [s]	Semivogais crescentes	Vogais médias tónicas	Vogais pretónicas não reduzidas
<i>Grafone</i>	Não	Não	Não	Não	Sim	Não
<i>Convert Text to IPA Transcription</i>	Sim	Não	Sim	Não	Não	Não

3.1. *Grafone*

O *Grafone* constitui uma ferramenta de conversão grafema-fonema para o PE que recorre a uma “abordagem híbrida na qual concorrem modelos estatísticos/probabilísticos imbuídos de informação linguística” (Candeias & Perdigão, 2013: 191). Uma das maiores vantagens desta ferramenta de conversão relaciona-se com o facto de apresentar a opção da transcrição dos grafemas em duas linguagens distintas: o AFI e o Alfabeto Fonético dos Métodos de Avaliação de Fala (SAMPA). O SAMPA é uma linguagem resultante do mapeamento de “symbols of the International Phonetic Alphabet onto ASCII codes” (Wells, 1997), i.e., constitui o resultado da adaptação dos fonemas fornecidos pela IPA num conjunto de símbolos mais facilmente reconhecidos pelos sistemas informáticos, de modo a facilitar a sua integração em programas de conversão automática.

Para além de incluir as variantes de pronúncia resultantes de palavras homógrafas (e.g., gosto /g'oʃtu/, /g'ɔʃtu/), esta ferramenta oferece a opção de transcrição das grafias anteriores e posteriores ao AO90 e possibilita a transcrição de uma quantidade abundante de texto, não se restringindo à conversão de unidades lexicais isoladas, ainda que não conceba determinadas variações fonéticas que ocorrem na fronteira destas. Para além disso, o *Grafone* exclui a divisão silábica dos vocábulos e

opta pela marcação do acento na vogal tónica em vez da sílaba tónica (e.g., boneco /bun'eku/), não se estendendo esta marcação, no entanto, aos monossílabos (e.g., ser /ser/). Adicionalmente, integra os acentos secundários, posicionando-os ao lado esquerdo da vogal, tipicamente em advérbios como *rapidamente* (/ɾ,apidem'ëtə/).

Comparativamente ao *Convert Text to IPA Transcription*, é possível considerar que o *Grafone* faz um tratamento mais fonológico da transcrição, tal como é constatado em Candeias & Perdigão (2013), o que se traduz na exclusão (i) de semivogais crescentes (e.g., frio /fr'iu/); (ii) da lateral alveolar velarizada [ɭ] (e.g., volta /v'olte/); (iii) das oclusivas fricativizadas intervocálicas [β], [ð] e [ɣ] (e.g., agora /eg'ore/); (iv) de algumas alternâncias associadas à consoante fricativa [s], como a modificação para [ʒ] seguida de unidade lexical iniciada por consoante sonora (e.g., casas douradas /k'azeʃ dor'adeʃ/) ou para [z] quando seguida de unidade lexical iniciada por vogal (e.g., casas azuis /k'azeʃ ez'uj/).

Apesar desta ferramenta apresentar informação sobre a maioria dos parâmetros acima mencionados, não fornece informação acerca dos caracteres utilizados na transcrição (à exceção das vibrantes /r/ e /ɾ/), e.g., a representação da vogal central [ɨ] enquanto /ə/ (e.g., tarde /t'ardə/). Este último parâmetro permitiu a deteção de incorreções na transcrição de certas vogais pretónicas que não sofrem redução vocálica, de que é exemplo a palavra bebé */bəb'ɛ/. Não obstante, o *Grafone* apresenta resultados positivos na conversão das vogais tónicas, suportando a variação entre as vogais médias [e] e [ɛ] (e.g., cabeça /keb'ese/, amarelo /emer'elu/) e [o] e [ɔ] (e.g., ovo /'ovu/, forte /f'ɔrtə/), o que não se verifica no *Convert Text to IPA Transcription*.

3.2. Convert Text to IPA Transcription

Tal como o *Grafone*, o *Convert Text to IPA Transcription* destina-se à conversão grafema-fone[ma] do PE e permite a transcrição de uma quantidade elevada de texto. Esta ferramenta destaca-se da anterior por fornecer um conjunto extenso de informação acerca dos critérios de transcrição utilizados, oferecendo uma descrição extensiva das vogais e consoantes do PE com a indicação dos respetivos símbolos do IPA e da sua gravação sonora. Adicionalmente, apresenta secções dedicadas às consoantes em coda, as variações associadas ao grafema <x>, o acento tónico, as vogais nasais, os

ditongos, entre outros fenómenos, permitindo uma experiência de transcrição mais personalizada. O utilizador pode ativar ou desativar determinadas opções, como o separador silábico, o marcador do acento tónico e a dissimilação do <i>.

Das duas ferramentas apresentadas, o *Convert Text to IPA Transcription* é a única que fornece a divisão silábica dos vocábulos, apesar de não incluir as variações associadas a palavras homógrafas e de não ser sensível à ortografia anterior ao AO90. Comparativamente ao *Grafone*, é possível verificar que este conversor faz um tratamento da transcrição mais fonético, o que se exemplifica através da inclusão da lateral alveolar velarizada [ɫ] (e.g., tal ['taɫ]) e da inclusão de algumas alternâncias associadas à consoante fricativa [s], como a modificação para [ʒ] seguida de unidade lexical iniciada por consoante sonora (e.g., casas douradas ['ka.zeʒ do.'ra.deʃ]) ou para [z] quando seguida de unidade lexical iniciada por vogal (e.g., casas azuis ['ka.zez e.'zuʃ]).

Contudo, são excluídas as semivogais crescentes e as oclusivas fricativizadas. Na mesma linha do que ocorre com o *Grafone*, determinadas palavras cujas vogais em posição pretónica não sofrem redução vocálica apresentam erros (bebé *[bi.'be], normal *[nur.'maɫ], amanhã *[e.me.'ɲẽ]), o que poderá apontar para um possível desafio geral à conversão grafema-fonema do PE. Adicionalmente, é possível observar algumas incongruências na transcrição das vogais tónicas médias não grafadas com diacrítico (e.g., amarelo *[e.me.'re.ɫu], forte *['for.ti])³¹ e na transcrição do ditongo [ɔj] (e.g., comboio *[kõ.'boj.u], boi ['boj]).

Encerrada a análise crítica dos sistemas de conversão fonética disponíveis para o PE, partimos para a descrição do sistema desenvolvido no presente trabalho.

4. Metodologia para o desenvolvimento do sistema de conversão

Este capítulo apresenta a revisão e uma proposta de reestruturação dos critérios definidos no prefácio do DLPC (secção 4.1.). Posteriormente, são descritos todos os passos metodológicos tomados no desenvolvimento desse sistema, desde o tratamento

³¹ Cf. pó ['pɔ], pé ['pe].

do *corpus*, os instrumentos utilizados, o procedimento da translineação, da marcação da sílaba tónica, da transcrição fonética e da importação dos dados no DLP-ACL (secção 4.2.).

O processo de adaptação digital do DLPC, realizado em 2023, deu origem a um recurso lexicográfico online profundamente reestruturado, com mais de cem mil entradas e de acesso público. Como referido na introdução desta investigação, apesar de, num momento inicial, terem sido empreendidos esforços em prol da inclusão da transcrição fonética nesta ferramenta, surgiram variados problemas associados aos caracteres não-ASCII durante o processo de retrodigitalização e conversão do PDF para HTML (Simões *et al.*, 2016: 14:2). Por esta razão, a CSS (*Cascading Style Sheets*) encontra-se atualmente configurada para ocultar as transcrições fonéticas. O desenvolvimento de um sistema de conversão grafema-fone[ma] automático e a sua consequente aplicação ao *Vocabulário Fundamental* do DLP-ACL (Reis, 2025) visa resolver este problema, bem como avaliar a aplicabilidade das expressões regulares na representação do PE.

4.1. Elaboração dos critérios de conversão

4.1.1. Translineação

Dado que o DLPC não apresenta a divisão silábica e/ou a translineação dos lemas, foi necessária a elaboração de critérios. A decisão de incluirmos a translineação deveu-se à importância que este componente tem para a ortografia e a revisão de texto, tendo optado por incluí-la num componente distinto do da transcrição fonética, uma vez que ilustram diferentes propósitos. Partimos, assim, dos critérios estabelecidos pela Base XX do AO90, assente em critérios fonológicos silábicos, e estabelecemos os parâmetros enunciados na Tabela 11.

Tabela 11 – Critérios de translineação

DLP-ACL (2025)
Utilização do divisor silábico do AFI (.)
Silabificação tendo em conta os princípios silábicos do ataque, do núcleo e da coda do PE
Divisão das sequências vocálicas que não constituem ditongos decrescentes (<i>e.g.</i> , frio → fri.o; área → á.re.a)
Indivisibilidade de grupos consonânticos próprios (<i>e.g.</i> , sobrar → so.brar)
Divisão de sequências que não constituem grupos consonânticos próprios (<i>e.g.</i> , carro → car.ro; assado → as.sa.do)
Conservação do hífen nas palavras compostas (<i>e.g.</i> , boa tarde → bo.a-tar.de)

4.1.2. Marcação da sílaba tónica

No DLPC, o acento principal da palavra encontra-se marcado com um diacrítico equivalente a um acento agudo (*e.g.* posterior [puʃtɨ'riór]), o que, para além de não corresponder ao que o AFI prevê, dificulta o processamento computacional e tem como consequência que as palavras com ditongos ou vogais nasais tónicas não possuam qualquer marcação (*e.g.*, mão [mẽw]). Desta forma, optámos pela eliminação do diacrítico da vogal tónica e pela utilização do acento principal fornecido pelo AFI (') na marcação da sílaba tónica (*e.g.* posterior [puʃtɨri'or]). Decidimos igualmente não incluir acentos secundários por uma questão de legibilidade e por não existirem, até ao momento, evidências que atestem a sua existência em português.

4.1.3. Transcrição fonética

O DLPC inclui a transcrição fonética larga e segundo a variedade-padrão do PE de todas as entradas lexicais, à exceção de abreviaturas (*e.g.*, NB), elementos de formação (*e.g.*, -logia), locuções latinas (*e.g.*, ad hoc) e símbolos (*e.g.*, km), e fornece o conjunto de critérios e caracteres adotados nesta mesma transcrição, confrontados adiante com os que foram estabelecidos para o presente sistema de conversão. Contudo, apesar da sua extensiva criteriação, a transcrição do DLPC foi alvo de fortes críticas por parte de alguns autores, nomeadamente Emiliano (2009: 285), que ressalta a incompletude do

inventário de caracteres, nomeadamente vogais³² e ditongos (*e.g.*, [ej], [uj], [ũj]); a utilização dos símbolos ‘R’, ‘Y’ e ‘λ’ em vez dos respetivos símbolos do AFI ‘R’, ‘Y’ e ‘λ’; a utilização de ‘r’ em vez do símbolo ‘r’ para representar a vibrante simples; a contemplação de ditongos crescentes orais e nasais e a presença de diversas gralhas nas transcrições.

Dos critérios estabelecidos para o DLPC, foram mantidos os seguintes parâmetros: (i) a transcrição encerrada entre parênteses retos, larga e segundo a variedade-padrão do PE; (ii) a velarização do /l/ em coda silábica; (iii) a transcrição das sequências do tipo #sC iniciadas por <es> como [ɨ] e por <is> como [i]. No entanto, foram adotados alguns critérios no DLP-ACL que diferem dos presentes no DLPC, encontrando-se essas diferenças resumidas na Tabela 12.

Tabela 12 – Diferenças entre os critérios de transcrição do DLPC (2001) e do DLP-ACL (2025)

DLPC (2001)	DLP-ACL (2025)
Marcação da nasalidade dos ditongos apenas na vogal (<i>e.g.</i> , mão [mẽw])	Marcação da nasalidade dos ditongos na vogal e na glide (<i>e.g.</i> , mão [ˈmẽw̃])
Adoção da pronúncia menos conservadora [ɔ] para <o> em fronteira inicial de palavra (<i>e.g.</i> , oferecer [ɔfiriˈsér])	Apresentação da pronúncia menos conservadora [ɔ] e mais conservadora [o] para <o> em fronteira inicial de palavra (<i>e.g.</i> , oferecer [ɔfiriˈser] / [ofiriˈser]), à exceção de determinadas palavras cuja vogal inicial é [ɔ] para a generalidade dos falantes (<i>e.g.</i> , olá, hospital, hotel)
Adoção da pronúncia menos conservadora [e] para <e> em fronteira inicial de palavra (<i>e.g.</i> , equipa [ekípe])	Apresentação da pronúncia menos conservadora [e] e mais conservadora [i] para <e> em fronteira inicial de palavra (<i>e.g.</i> , equipa [eˈkipe] / [iˈkipe]), à exceção de determinadas palavras cuja vogal inicial é [e] para a generalidade dos falantes (<i>e.g.</i> , ervanária, herbácea, etc.)
Inclusão de ditongos crescentes, por se revelar pouco económica em termos de espaço a adoção de duas ou mais transcrições	Exclusão de ditongos crescentes e a sua transcrição como sequências vocálicas heterossilábicas (<i>e.g.</i> , iogurte [ioˈgurti])

³² Apesar de Emiliano (2009: 285) afirmar que o inventário das vogais do DLPC (2001) está incompleto, encontram-se registadas as nove vogais orais ([a, e, ɛ, o, ɔ, i, ɨ, u]), as duas glides orais ([j, w]) e as cinco vogais nasais ([ẽ, ẽ̃, ỹ, õ, õ̃]) pertencentes ao PE. Os únicos segmentos em falta são as glides nasais [j̃, w̃] que não se verificam, pois a nasalidade é marcada somente na vogal do ditongo (*e.g.*, mão [mẽw]).

(e.g., iogurte [jɔgúrtɨ])	
Inclusão da transcrição do morfema de género feminino -a, se ocorrente, separado da forma masculina por uma vírgula (e.g., professor, a [prufisór, -e])	Inclusão da transcrição completa da forma feminina do lema, separado da forma masculina por uma vírgula (e.g., professor, professora [prufɨ'sor], [prufɨ'sore])
Adaptação dos estrangeirismos aos sons do português	

Apesar de o DLPC apenas marcar a nasalidade na vogal pertencente ao ditongo nasal, esta proposta encontra-se em desacordo com a generalidade das descrições existentes para o PE, exploradas anteriormente em 1.2.2., segundo as quais existem duas glides nasais ([ĩ, ã]) no inventário vocálico desta língua. Posto isto, optámos por assinalar a nasalidade na vogal e na glide que lhe é adjacente (e.g., mão ['mẽw̃]) de modo a representar com maior precisão a realização fonética dos ditongos nasais.

No que diz respeito às vogais <e> e <o> em fronteira inicial de palavra, optámos pela representação simultânea das pronúncias mais e menos conservadoras ([i, o] e [e, ɔ], respetivamente), separadas por uma barra oblíqua, contrariamente ao que se observa no DLPC, que apenas apresenta as realizações fonéticas menos conservadoras. A decisão de incluirmos ambas as pronúncias justifica-se por se tratar ainda de um fenómeno de variação livre e instável no PE e porque, tendo em conta o carácter de uma obra como o dicionário, esta opção torna a transcrição mais abrangente e representativa quanto à realidade fonética do português. Os únicos casos que não apresentam ambas as pronúncias constituem palavras cuja vogal inicial é baixa ([ɛ, ɔ]) para a generalidade dos falantes, como é o caso das palavras *olá*, *hotel* e *hospital*, presentes no nosso *corpus*.

Para além disso, ao contrário do que ocorre no DLPC, excluímos a representação dos ditongos crescentes e optámos pela sua transcrição enquanto sequências vocálicas heterossilábicas. Ainda que a existência de ditongos crescentes no PE numa fala mais coloquial e rápida seja provável, a representação destas sequências entraria não só em conflito com a translineação (e.g. fri.o *['frju]), que as assume como heterossilábicas, como se demonstraria incompatível com a transcrição larga, representativa de uma fala formal e pausada, a que a nossa transcrição pretende descrever.

Importa salientar que, dado que esta transcrição está concebida para um ambiente digital, a economia de espaço não constituiu um critério relevante na decisão de incluir os ditongos crescentes, ao contrário do que sucedeu no DLPC, por se tratar de um dicionário impresso. Este argumento estendeu-se de igual forma à inclusão da transcrição das formas femininas na sua íntegra.

Adicionalmente, comparámos o inventário de caracteres utilizado no DLPC aos caracteres definidos para o sistema de conversão do DLP-ACL na Tabela 13.

Tabela 13 – Comparação entre o inventário de caracteres do DLPC (2001) e do DLP-ACL (2025)

	DLPC (2001)	DLP-ACL (2025)
Consoantes	[p], [b], [t], [d], [k], [g], [m], [n], [ɲ], [ʀ], [r], [f], [v], [s], [z], [ʃ], [ʒ], [l], [ʎ], [λ]	[p], [b], [β], [t], [d], [ð], [k], [kʷ], [g], [ɣ], [gʷ], [m], [n], [ɲ], [ʀ], [r], [f], [v], [s], [z], [ʃ], [ʒ], [l], [ʎ], [λ]
Vogais orais	[a], [e], [ɛ], [o], [ɔ], [u], [i]	[a], [e], [ɛ], [o], [ɔ], [u], [i]
Glides orais	[j], [w]	[j], [w]
Ditongos orais (descrescentes)	[aj], [ej], [oj], [ɔj], [aw], [ew], [ɛw]	[aj], [ej], [ɛj], [oj], [ɔj], [uj], [aw], [ew], [ɛw], [iw]
Vogais nasais	[ẽ], [ẽ̃], [ĩ], [õ], [ũ]	[ẽ], [ẽ̃], [ĩ], [õ], [ũ]
Glides nasais		[j̃], [w̃]
Ditongos nasais (descrescentes)	[ẽj], [õj], [ẽw]	[ẽj̃], [õj̃], [ũj̃], [ẽw̃]

Conforme podemos observar, o inventário consonântico por nós selecionado é mais extenso do que o utilizado no DLPC. Esta extensão acrescida deve-se à inclusão da fricativação das oclusivas /b/ [β], /d/ [ð] e /g/ [ɣ], que, apesar de constituir um fenómeno com restrições de natureza discursiva e societal pouco estudadas, é observado com uma frequência relevante na variedade-padrão. Optámos ainda pela incorporação das oclusivas velares labializadas /kʷ/ e /gʷ/ na transcrição de palavras como *quarto* ([ˈkʷartu]) e *guardar* ([ˈgʷerder]), em detrimento da inclusão de ditongos crescentes como [wa] ou [we], adotados na transcrição do DLPC (e.g., guardar [gwárde]), devido à emergência de estudos mais recentes que apontam para o surgimento destes

segmentos do sistema consonântico do PE (Rodrigues, 2003; 2020; Andrade & D'Andrade, 2020). Além disso, todos os símbolos por nós selecionados pertencem ao AFI, ao contrário do que ocorre no DLPC que, tal como mencionado por Emiliano (2009: 285), emprega os símbolos [R], [ɾ] e [λ] em vez de [ʀ], [ɽ] e [ʌ], respetivamente. Por fim, o DLPC apresenta a vibrante [r] para a representação da vibrante simples do PE que apenas constitui uma realização possível de R-forte e nunca (ou muito raramente) de R-fraco. Posto isto, adotámos a realização [r] na representação da vibrante simples por se tratar da realização mais comum desta vibrante e por se encontrar de acordo com a generalidade da literatura. Optámos de igual forma pela utilização de [ʀ] na representação do R-forte por razões já citadas em 1.2.1.2., mas que voltamos a referir. Apesar de [r] se destacar como a vibrante predominante no território nacional, a maioria dos estudos aponta para a sua inexistência na variedade-padrão do PE (Mateus & D'Andrade, 2000; Vigário, 2003; Pereira, 2020) ou a sua realização em pequena percentagem (Rodrigues, 2003). Assim, para além de [ʀ] constituir a consoante mais comumente empregada na literatura do PE, também corresponde, segundo o estudo de Pereira (2020: 119), à realização prevalente do R-forte na região de Lisboa.

De igual forma, o inventário de ditongos orais e nasais decrescentes adotado na presente transcrição corresponde ao inventário completo do PE, sendo, por isso, também mais abrangente do que o apresentado no DLPC, ao qual foram adicionados os ditongos orais [ɛj] (e.g., papéis [pe'pej]), [uj] (e.g., cuidar [kuj'ðar]) e [ew] (e.g., saudade [sew'ðaði]) e o ditongo nasal [ũj] (e.g., muito ['mũjtu]).

4.2. Desenvolvimento do sistema de conversão

4.2.1. Constituição do *corpus*

Tendo em conta a ampla dimensão lexicográfica do DLP-ACL, o desenvolvimento do sistema de conversão exigiu, numa fase inicial, a seleção de uma amostra reduzida que permitisse testar o funcionamento do sistema, identificar eventuais erros e implementar as respetivas correções. Considerando a necessidade de delimitar o escopo da análise, optou-se por analisar os itens lexicais de nível 1 (Muito Fácil) que integram o

Vocabulário Fundamental (Reis, 2025), tendo por base o projeto *iRead4Skills* (Wilkens *et al.*, 2024)³³.

Os vocabulários fundamentais consistem na recolha e listagem das palavras mais frequentes da língua, tendo por base critérios teóricos e estatísticos. Contudo, existem algumas limitações relacionadas com estas listas, nomeadamente a inexistência da associação dos itens lexicais a significados concretos, o que se traduz na exclusão dos diferentes níveis de proficiência (Reis, 2025). O projeto *iRead4Skills* (Wilkens *et al.*, 2024) veio colmatar esta lacuna através da extração das listas de léxico por níveis de complexidade, entre os quais destacámos o Nível 1, que aglomera os vocábulos mais frequentes e compreensíveis por falantes nativos com níveis muito baixos de escolaridade (i.e., que não terminaram o ensino básico ~ 6.º ano) (Monteiro *et al.*, 2023). De modo a garantir um alinhamento correto desta lista aos lemas do DLP-ACL, procedemos à exclusão de itens lexicais repetidos e vocábulos não correspondentes a lemas como antropónimos (*e.g.*, Rita, Pedro, etc.), topónimos (*e.g.*, Portugal, Porto, etc.), locuções (*e.g.*, quanto baste, pelo menos, etc.) e combinatórias (*e.g.*, colher de sopa), o que resultou numa redução de 839 para 730 vocábulos.

4.2.2. Instrumentos

As expressões regulares, ou simplesmente *regex*, constituem um tipo de padrão de pesquisa que possibilita operações de procura e/ou substituição versáteis. Utilizando padrões e não variáveis, as expressões regulares revelam-se extremamente eficazes na identificação ou modificação de dados, baseando-se na deteção de padrões de caracteres em vez da correspondência literal de sequências fixas (Goyvaerts, 2007; Goyvaerts & Levithan, 2009). A decisão de utilizarmos expressões regulares como ferramenta primária para o desenvolvimento deste sistema de conversão justificou-se, assim, pela sua ampla aplicabilidade a qualquer domínio textual, por permitirem identificar padrões baseados em regras definidas de acordo com um propósito específico e por serem altamente interoperáveis e estáveis ao nível dos *softwares*.

³³ Para saber mais sobre o projeto *iRead4Skills*, consultar: <https://doi.org/10.5281/zenodo.10889986>.

As expressões regulares elaboradas³⁴ partiram de um conjunto de caracteres especiais, apresentados na Tabela 14, que desempenham uma função específica e que, quando combinados com outros caracteres (*e.g.*, grafemas), possibilitam a seleção de um determinado contexto ou sequência.

Tabela 14 – Sintaxe dos caracteres utilizados nas *regex*

Carácter	Descrição	Exemplo
Qualquer carácter exceto [<code>\^\$. ?*+(){}]</code>	Qualquer carácter corresponde a si mesmo.	<code>a</code> → a
<code>.</code>	Corresponde a qualquer carácter.	<code>.</code> → a
<code>*</code>	O carácter anterior pode aparecer qualquer número de vezes (0, 1, 2, etc.).	<code>a*</code> → <code>a*</code> → a <code>a*</code> → aaa
<code>[]</code> (barras oblíquas)	Inicia e termina uma classe de caracteres. Uma classe de caracteres seleciona individualmente cada carácter das possibilidades oferecidas dentro dessa mesma classe.	<code>[eio]</code> → respeito; preto
<code>^</code> depois de <code>[</code> (<i>caret</i> após barra oblíqua)	Nega a classe de caracteres. Seleciona qualquer carácter desde que não pertença à classe de caracteres.	<code>[^ai]</code> → pai
<code>^</code> (<i>caret</i>)	Seleciona a posição inicial da expressão à qual o padrão está a ser aplicado.	<code>^g</code> → ganhar
<code>\$</code> (dólar)	Seleciona a posição final da expressão à qual o padrão está a ser aplicado.	<code>á\$</code> → cá

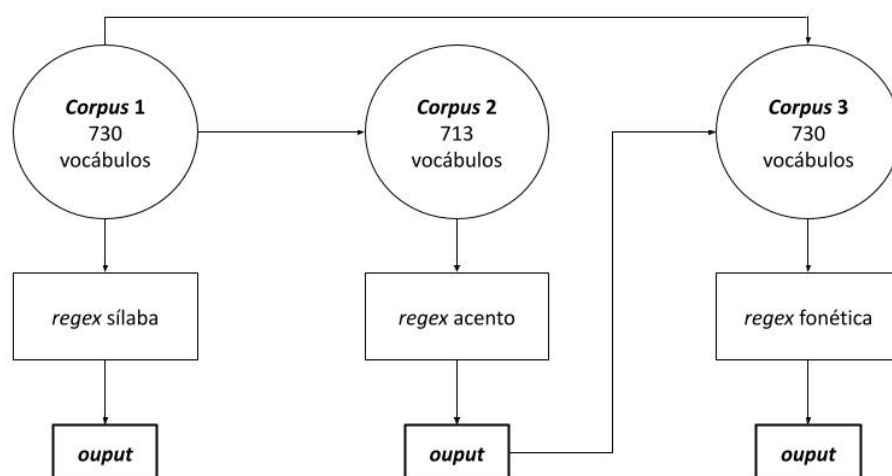
³⁴ Disponíveis em <https://github.com/LeonorMartins/G2P-DLP/tree/main/Regex>

() (<i>pipe</i> entre parênteses)	Seleciona o(s) caractere(s) à esquerda ou o(s) caractere(s) à direita. Útil para quando queremos incluir ditongos e dígrafos numa classe de caracteres.	(m n lh) → malha ; mana; mala ≠ [mnlh] → malha ; mana ; mala
{n} (parênteses curvos)	Seleciona uma sequência de n caracteres com o mesmo carácter.	r{2} → sorrir
(?=) (<i>positive lookahead</i>)	Seleciona o carácter quando é seguido pelo que se encontra após =	p(?=ão) → pão ; pã p(?=[ão]) → pão ; pã p(?=(ão é)) → pão ; pé
(?!) (<i>negative lookahead</i>)	Seleciona o carácter quando não é seguido pelo que se encontra após !	p(?!e) → perto; preto ; porta p(?![er]) → perto; preto; porta p(?!(er re or)) → perto; preto; porta
(?<=) (<i>positive lookbehind</i>)	Seleciona o carácter quando é antecedido pelo que se encontra após =	(?<=o)l → voltar; possível (?<=[eo])l → voltar; possível (?!(vo ve))l → voltar; possível
(?<!) (<i>negative lookbehind</i>)	Seleciona o carácter quando não é antecedido pelo que se encontra após !	(?<!g)ui → seguir, aquilo (?<![gq])ui → seguir, aquilo (?<!(eg aq)) → seguir, aquilo

4.2.3. Procedimentos

Uma vez que a divisão silábica e a transcrição fonética foram registadas em dois componentes distintos, apenas os *outputs* das suas respetivas *regex* foram incluídos integralmente no DLP-ACL, pois o *output* das *regex* da marcação da sílaba tónica representou somente uma fase transitória, integrante do *output* das *regex* de transcrição fonética, como podemos observar a partir da Figura 9.

Figura 9 – Corpus e outputs



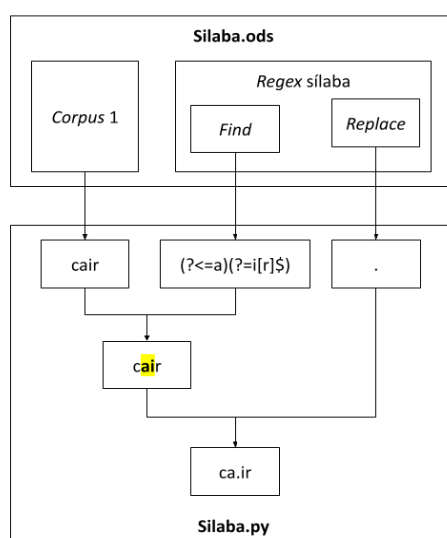
Como referido anteriormente, a exclusão de vocábulos não correspondentes a lemas do DLP resultou num *corpus* com um total de 730 palavras. Este conjunto (*corpus* 1) constituiu o ponto de partida para a implementação das *regex* da sílaba, sendo que não houve necessidade de se excluírem vocábulos, pois os 77 monossílabos pertencentes ao *corpus* permaneceram inalterados (*e.g.*, bom -> bom). Por outro lado, surgiu a posterior necessidade de remoção de todas as palavras inacentuadas do *corpus* 1 para a aplicação das *regex* do acento, como os pronomes pessoais nas formas de complemento direto (*e.g.*, o, a); artigos definidos e indefinidos (*e.g.*, um), algumas preposições (*e.g.*, de, em, para, etc.) e algumas conjunções (*e.g.*, e, mas, ou, porque, etc.), para que não se dessem marcações incorretas como *'que* ou *'se*. Com a exclusão dos 17 vocábulos inacentuados, o *corpus* 2 passou a incluir 713 dos 730 vocábulos iniciais.

Posteriormente, para que pudéssemos realizar a conversão grafema-fonema do *corpus* inicial na sua totalidade, mantendo os marcadores silábicos inseridos através das *regex* do acento, procedemos à construção do *corpus* 3. A elaboração do *corpus* 3 baseou-se na utilização dos 713 vocábulos acentuados do *output* do *corpus* 2 e na devolução dos 17 vocábulos inacentuados do *corpus* 1 previamente excluídos. A decisão de partirmos do *output* das *regex* do acento para as *regex* da transcrição fonética e não vice-versa deveu-se à utilidade que o marcador silábico impôs nos contextos de redução

vocálica pretónicos e postónicos, que, caso não fosse essa a metodologia usada, não seriam possíveis de captar, tal como iremos explorar no ponto 4.1.

Para alcançar os resultados pretendidos, optámos por armazenar as expressões regulares relativas à sílaba, ao acento e à fonética em três folhas de cálculo no formato ODS (Silaba.ods, Acento.ods, Fonetica.ods)³⁵. Para aplicar essas expressões ao *corpus*, desenvolvemos um programa em *Python*³⁶, uma escolha justificada pela flexibilidade e eficiência no processamento de texto que a linguagem oferece, bem como pela vasta disponibilidade de bibliotecas especializadas. Em particular, utilizámos a biblioteca *pandas* (McKinney, 2010) para a leitura e manipulação dos dados nos ficheiros ODS e a biblioteca *regex* (Aho, 1991) para o tratamento avançado de expressões regulares, o que permitiu, por exemplo, o uso de *look-behinds* de largura variável, algo não suportado pelo módulo padrão *re* do *Python*. A Figura 10 ilustra este processo.

Figura 10 – Exemplo do processo de translineação da palavra *cair*



4.2.3.1. Translineação

Dado que a intenção das *regex* de translineação não consiste na substituição de nenhum carácter do *corpus*, mas sim na inserção de um divisor silábico (.) entre caracteres, a nossa atenção recaiu especialmente sobre as fronteiras silábicas

³⁵ Disponíveis em https://github.com/LeonorMartins/G2P-DLP/tree/main/Ficheiros_ods

³⁶ Para saber mais, consultar <https://github.com/LeonorMartins/G2P-DLP/tree/main/Notebooks> e https://github.com/LeonorMartins/G2P-DLP/tree/main/Ficheiros_python

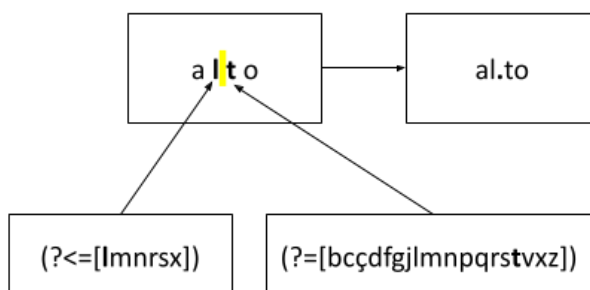
ocorrentes entre os diferentes tipos de sílaba listados do português, exemplificados na Tabela 15.

Tabela 15 – Fronteiras silábicas do português³⁷

Fronteira silábica	Exemplo
C.C (consoante em coda + consoante em ataque)	castanho → cas.ta.nho abdicar → ab.di.car
C(V).C (consoante em ataque e núcleo vazio + consoante em ataque)	pneu → pneu
V.C (vogal + consoante em ataque)	viver → vi.ver
G.C (glide + consoante em ataque)	deitar → dei.tar
G.V (glide + vogal)	meio → mei.o
V.V (vogal + vogal)	peão → pe.ão notícia → no.tí.ci.a

Estas fronteiras constituíram o ponto de partida na composição das *regex* de marcação silábica. As *regex* que permitem captar uma fronteira entre duas classes de caracteres diferenciam-se das restantes, na medida em que são constituídas por uma *positive lookbehind* com a classe de caracteres que surgem à esquerda da fronteira e por uma *positive lookahead* com a classe de caracteres que surgem à direita da fronteira, tal como a Figura 11 exemplifica.

Figura 11 – Estrutura da *regex* de translineação da palavra *alto*



³⁷ As células assinaladas a cinzento indicam contextos não contemplados no nosso *corpus*.

Importa igualmente ressaltar que, no processo de elaboração destas expressões, a hierarquia na qual se encontravam dispostas não teve particular relevância, partindo do pressuposto de que não devem captar duas vezes os mesmos caracteres, operando a diferentes níveis da palavra, salvo casos excepcionais que não integram o nosso *corpus*, como grupos consonânticos impróprios³⁸ (e.g., <gn>, <pt>, etc.).

Iniciámos a elaboração das *regex* da translineação na fronteira que contempla mais restrições no português, a fronteira entre consoante em coda e consoante em ataque (C.C), representada na Tabela 16. No PE, ocorrem em coda silábica no interior de palavra [l], [r], [s] e [j], representadas no nosso *corpus* pelos grafemas <l>, <r>, <s> e <x>. Podem ainda surgir nesta posição os grafemas <m> e <n>, ainda que não possam ser considerados ao nível fonológico, uma vez que não representam um segmento em coda, marcando simplesmente a nasalidade da vogal anterior (Andrade & Andrade, 2020: 3386).

Tabela 16 – Regex S1

ID	Fronteira	Find	Replace	Exemplo
S1	C.C	(?<=[lmnrstx])(?=[bcçdfgjlmnpqrstvxz])	.	partir -> par.tir

Uma vez delimitada a fronteira entre C.C, passou-se à identificação das fronteiras entre vogal e consoante (V.C) e glide e consoante (G.C), representadas na Tabela 17, em que foi necessário incluir uma *negative lookahead* entre as duas classes com as consoantes ocorrentes em coda no interior de palavra e coda final <l>, <r>, <s>, <x>, <z>, bem como as consoantes sinalizadoras de nasalidade <m> e <n>, para que não ocorram erros do tipo alto → *a.l.to ou limpar → *li.m.par e para que esta *regex* fosse aplicada a consoantes que surgem exclusivamente em ataque.

Tabela 17 – Regex S2

ID	Fronteira	Find	Replace	Exemplo
----	-----------	------	---------	---------

³⁸ Por oposição aos grupos consonânticos próprios, os grupos consonânticos impróprios possuem núcleo vazio.

S2	V.C G.C	(?<=[aáãääãẽêëèèîïïóóòôõúúûû])(?![lmnrscxz])(?=[bcçdfgjpqrvtv])	.	tratar -> tra.tar baixo -> bai.xo
----	------------	---	---	--------------------------------------

Dada a exclusão das consoantes <l>, <r>, <s>, <x>, <z>, <m> e <n> da *regex* anterior, foi necessário criar as expressões representadas na Tabela 18, que nos permitiram captar as fronteiras:

– V.CV e G.CV, entre vogal ou glide e consoante <l>, <m>, <n>, <r>, <s>, <x> <z> em ataque e seguida de vogal, de modo a não se confundir com consoante em coda;

– V.CC, entre vogal e os dígrafos <lh> e <nh>, sendo que o dígrafo <ch> e os encontros consonânticos próprios <pr>,
, etc. já se encontravam captados por S2.

Tabela 18 – *Regex* S3 e S4

ID	Fronteira	Find	Replace	Exemplo
S3	V.CV G.CV	(?<=[aáãääãẽêëèèîïïóóòôõúúûû])(?=[lmnrscxz][aáãääãẽêëèèîïïóóòôõúúûû])	.	galo -> ga.lo deixar -> dei.xar
S4	V.CC	(?<=[aáãääãẽêëèèîïïóóòôõúúûû])(?=(lh nh))	.	sonhar -> so.nhar

Uma vez captadas todas as fronteiras iniciadas por consoante à direita, restava-nos a inclusão das fronteiras:

- VG.V, entre ditongo decrescente e vogal;
- V.ŨG, entre vogal e ditongo nasal;
- V.V ou V.Ũ, entre duas vogais.

A primeira destas fronteiras apenas se observa com ditongos orais, dado que o português não dispõe de sequências VG.V com ditongos nasais. Posto isto, criámos a *regex* S5, apresentada na Tabela 19, seleccionando a fronteira entre a classe de ditongos orais presentes no nosso *corpus* (<ai>, <ei>, <oi>, <ui>, <au>, <eu>, <iu>, <ou>³⁹) e a classe de vogais. Optou-se por incluir os ditongos na sua integridade na primeira classe desta *regex* de modo a evitarmos a sua incorreta divisão, como prevíamos que fosse acontecer em palavras como praia → *pra.i.a, se apenas fosse seleccionada a fronteira entre glide e vogal em início de sílaba. Esta lógica foi estendida à elaboração de S6,

³⁹ Atente-se que este ditongo sofre monotongação para [o] na variedade padrão do PE.

também apresentada na Tabela 19, a partir da comutação de posições entre a classe de ditongos e a classe de vogais e da substituição dos ditongos orais por ditongos nasais, visto que no nosso *corpus* apenas constam sequências do tipo V.VG com estes ditongos⁴⁰, nomeadamente com o ditongo [ẽũ], representado graficamente por <ão>.

Tabela 19 – Regex S5 e S6

ID	Fronteira	Find	Replace	Exemplo
S5	VG.V	(?<=[ai ei oi ui au eu iu ou])(?=[aáâãäéèêîïïoóóôôõuúúûû])	.	meia -> mei.a
S6	V.ŨG	(?<=[aáâãäéèêîïïoóóôôõuúúûû])(?=(ão))	.	leão -> le.ão

Por fim, passámos às sequências vocálicas heterossilábicas V.V., compostas por uma vogal alta e uma vogal não alta, duas vogais altas ou duas vogais não altas. Começámos por separar a classe dos grafemas correspondentes às vogais altas /i/ (<i>, <e>) e /u/ (<o>) e as vogais não altas <a>, <e> e a vogal <o>, conforme podemos observar na Tabela 20.

Tabela 20 – Regex S7

ID	Fronteira	Find	Replace	Exemplo
S7	V.V V.Ũ	(?<=[ieo])(?=[aeo])	.	voar -> vo.ar cliente -> cli.en.te

Decidimos excluir <u> da primeira classe de vogais para que não fossem erradamente divididas sequências como quanto → *qu.anto e querer → *qu.erer. De igual forma, decidimos não incluir este grafema na segunda classe, de modo a não dividirmos ditongos, como <eu> (e.g., euro → *e.uro) e <ou> (e.g., ouvir → *o.uvir), bem como o grafema <i>, para não obtermos divisões erróneas dos ditongos <ei> (e.g., deixar → *de.ixar) e <oi> (e.g., oito → *o.ito). Por outro lado, decidimos manter o grafema <o> na primeira classe de vogais de modo a captarmos sequências como [ue] <oe> (e.g., coelho → co.elho), [ua] <oa> (e.g., voar → vo.ar) e [oe] (e.g., pessoa → pesso.a), bem

⁴⁰ Em português, são escassas as palavras de sequência VG.V com ditongos orais (e.g., miau, piau).

como na segunda classe, de modo a dividirmos a sequência [iu], composta pelas duas vogais altas e grafada <io> ou <eo> (e.g., ar.má.ri.o; vídeo → víde.o).

Para além das sequências heterossilábicas V.V, esta *regex* também nos permitiu captar sequências V.Û, como [iẽ] <ian> (e.g., criança → cri.ança), [iẽ] <ien> (e.g., cliente → cli.ente) e [uẽ] <oen> (e.g., doença → do.en.ça).

Com a exclusão de <u>, surgiu a necessidade de elaborarmos uma expressão adicional que captasse exclusivamente as sequências vocálicas heterossilábicas [ua] (e.g., continuar → continu.a.r) e [ue] (e.g., rua → ru.a), as únicas sequências iniciadas por <u> do nosso *corpus*. Cingimo-nos então a este mesmo contexto, alargando o escopo da *regex* à consoante que antecede <u> e excluindo <g> e <q>, para que não fossem divididas incorretamente sequências como *quanto → qu.an.to ou *língua → língu.a, nos quais o grafema <u> se encontra antecedido por <g> ou <q>, não se comportando como uma glide, mas antes como um elemento vocálico associado à consoante em ataque (Andrade & Andrade, 2020: 3382), conforme representado na Tabela 21.

Tabela 21 – *Regex* S8

ID	Fronteira	Find	Replace	Exemplo
S8	V.V	(?<=[bcçdfhjlmnpstrvxz]u)(?=a)	.	rua -> ru.a

Vejamos, adicionalmente, as sequências em que /i/ e /u/ são precedidos por outra vogal, mas não estão lexicalmente marcados como não acentuáveis, recebendo, por isso, o acento da palavra. A título exemplificativo, temos a sequência <ui>, constituída por duas vogais altas que tanto podem formar um núcleo complexo, i.e., um ditongo decrescente ([uj]), correspondendo à sequência VG (e.g., cuidar → cui.dar), como podem corresponder a uma sequência vocálica V.V ([ui]) (e.g., fluidez → flu.i.dez) e ainda a [i], quando precedida de <g> ou <q>, correspondendo foneticamente a V (e.g., seguir → se.guir; aquilo → a.qui.lo). A imprevisibilidade de [uj] e [ui] no contexto gráfico de <ui> poderá representar uma limitação destas *regex* numa eventual extensão de amostra e a potencial necessidade de uma recolha probabilística dos contextos de ocorrência destas sequências, de modo a viabilizar a sua correta divisão. Todavia, apesar

de prevermos este problema, o nosso *corpus* apenas integra a sequência VG (e.g. cuidado → cui.da.do) e V (e.g., aqui → a.quei), cujas fronteiras se encontravam já detetadas pelas *regex* anteriores, não se levantando, portanto, qualquer problema na divisão de <ui>.

Outra destas sequências corresponde graficamente a <ai>, que poderá corresponder ao ditongo decrescente [aj] (e.g., baixo → bai.xo) ou à sequência vocálica V.V [ei] (e.g., rainha → ra.inha). No nosso *corpus*, a sequência vocálica [ei] ocorre em dois contextos distintos:

- em palavras oxítonas terminadas em consoante, nomeadamente formas verbais infinitivas pertencentes à terceira conjugação (e.g., sair → sa.ir);
- em sequência heterossilábica V.Ñ (e.g., ainda → a.in.da).

Tendo em conta estes contextos, procedemos à elaboração das *regex* S9 e S10, representadas na Tabela 22. Importa ressaltar, no entanto, que, tal como ocorre para <ui>, esta sequência de grafemas poderá levantar problemas num futuro alargamento do *corpus*.

Tabela 22 – *Regex* S9 e S10

ID	Fronteira	Find	Replace	Exemplo
S9	V.VC	(?<=a)(?=i[r]§)	.	sair -> sa.ir
S10	V.Ñ	(?<=a)(?=in)	.	ainda -> a.in.da

Finalmente, surgem as sequências vocálicas V.V acentuadas [ei] (e.g., aí → a.í, país → pa.ís) e [eu] (e.g., saúde → sa.úde), às quais associámos a expressão S11 (Tabela 23):

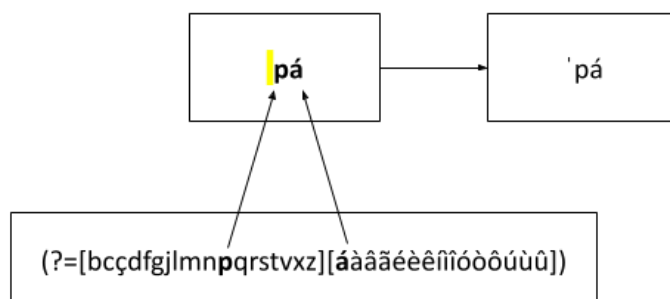
Tabela 23 – *Regex* S11

ID	Fronteira	Find	Replace	Exemplo
S11	V.V	(?<=a)(?=[íú])	.	país -> pa.ís

4.2.3.2. Marcação da sílaba tónica

As *regex* de marcação da sílaba tónica são menos restritas do que as anteriores, dado que, para além de poderem seleccionar fronteiras entre dois elementos, seleccionam também posições, dependendo (ou não) do que vier antes da classe designada. Para seleccionarmos uma posição anterior a determinada sequência, apenas necessitámos de uma *positive lookahead* que contenha a(s) classe(s) de caracteres integrantes de determinada sequência silábica, tal como demonstra a Figura 12.

Figura 12 – Estrutura da *regex* de marcação da sílaba tónica da palavra *pá*



Considerando que apenas seleccionámos a posição anterior às sílabas tónicas, partimos das sequências silábicas que podem ocorrer nesta mesma posição⁴¹: V (e.g., é), VG (e.g., ei), VC (e.g., ar), VGC (e.g., eis), CV (e.g., pé), CCV (próprio), CË (e.g., mamã), CCË (e.g., pronto), CVG (e.g., céu), CCVG (e.g., praia), CËG (e.g., coração), CVC (e.g., dor), CCVC (e.g., cumprir), CVGC (e.g., cais), CCVGC (e.g., trais), CCËC (trans), CËGC (e.g., pães).

4.2.3.2.1. Palavras acentuadas lexicalmente

Visto que esta conversão é baseada na ortografia, iniciámos estas *regex* pelas sílabas cuja tonicidade é marcada no léxico através de acento agudo, grave, circunflexo e til (oxítonas terminadas em vogal, ditongo nasal e oral). Primeiramente, decidimos captar as sequências tónicas com vogal acentuada com diacrítico iniciadas por consoante, nomeadamente CV (e.g., pé → 'pé), CË (e.g., mamã → ma'mã), CVG (e.g., céu → 'céu), CËG (e.g., coração → cora'ção) e CCV (e.g., próprio → 'próprio). Para tal,

⁴¹ Nesta listagem, V corresponde a vogal, C corresponde a consoante e G corresponde a glide.

procedemos à construção da *regex* representada na Tabela 24, que engloba a classe das consoantes, dos dígrafos e dos encontros consonânticos ocorrentes em ataque silábico neste tipo de sequências.

Tabela 24 – *Regex* A1

ID	Posição	Find	Replace	Exemplo
A1	'CV 'CŨ 'CVG 'CŨG 'CCV	(?=(b c ç d f gu g ch nh lh j l m n p q qu rr ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)[áâãäåêëêîîôòóùû])	'	pé -> 'pé

Decidimos excluir o grafema <h> desta classe, posto que, apesar de ocorrer em ataque silábico inicial em palavras como *haver* e *horta*, não pertence a nenhuma sílaba tónica cuja vogal é acentuada com diacrítico e poderia levantar problemas na marcação dos dígrafos <ch>, <nh> e <lh> (e.g., *amanhã* → *ama'n'hã). De igual forma, procedemos à remoção de <r> e <s> da classe de consoantes para que o marcador silábico não fosse duplamente aplicado (i) às sequências <rr> ou <ss> (e.g., *possível* → *po's'sível) e (ii) aos grupos consonânticos próprios (e.g., *três* → *'t'rês)⁴². Foi ainda necessário introduzirmos as sequências <gu> e <qu>, para que casos como *ninguém* → nin'guém pudessem ser captados.

Dada a exclusão de <r> e <s> da classe de consoantes, surgiu naturalmente a necessidade de detetarmos as sílabas tónicas nas quais <r> ocorre em ataque silábico seguido de vogal acentuada por diacrítico, não precedido de <s> ou de consoante com a qual forma encontro consonântico⁴³ (<p>, , <t>, <d>, <c>, <g>, <f>, <v>), e <s> ocorre em ataque silábico seguido de vogal acentuada por diacrítico, não precedido de <s>. Para tal, adicionámos duas *negative lookbehind* com as condições listadas e procedemos à elaboração das expressões ilustradas na Tabela 25.

⁴² No nosso *corpus*, não ocorre nenhuma sílaba tónica com grupo consonântico com lateral seguido de vogal acentuada com diacrítico (e.g., *plátano*).

⁴³ Atentemos que, no nosso *corpus*, o dígrafo <rr> não surge seguido de vogal acentuada com diacrítico (e.g., *terrível*), daí não ser necessário incluímos o grafema <r> na respetiva *negative lookbehind*.

Tabela 25 – Regex A2 e A3

ID	Posição	Find	Replace	Exemplo
A2	'CV 'C [~] V	(?<![pbtdcgfvr])(?=r[áâãäéèêîíóôùú])	'	rádio -> 'rádio
A3		(?<![s])(?=s[áâãäéèêîíóôùú])	'	sítio -> 'sítio

Para além das sílabas iniciadas por consoante, o português dispõe ainda de sílabas acentuadas com diacrítico iniciadas por vogal, com coda vazia ou preenchida. No nosso *corpus*, alguns destes casos surgem em início de palavra, pelo que procedemos à marcação da posição anterior a estas vogais através da *regex* A4 (Tabela 26), que permitiu a identificação quer de V, quer da sequência VC:

Tabela 26 – Regex A4

ID	Fronteira	Find	Replace	Exemplo
A4	^'V ^'VC	(?=^[áâãäéèêîíóôùú])	'	área -> 'área árvore -> 'árvore

Posteriormente, surgem as sequências V.ŨG nos vocábulos *avião*, *leão* e *região* que, ao contrário dos restantes casos, necessitaram da deteção da fronteira V.ŨG e não somente da posição anterior a ŨG, sendo que tal resultaria na deficiente marcação de vocábulos com sequência CŨG, como, por exemplo, *razão* → *raz'ão. Posto isto, foi necessário alargarmos o escopo da *regex* à vogal anterior, de modo a captarmos somente a sequência V.ŨG, tal como procedemos anteriormente para S6 e conforme podemos observar na Tabela 27.

Tabela 27 – Regex A5

ID	Fronteira	Find	Replace	Exemplo
A5	V.'ŨG	(?<=[aeiou])(?=(ão))	'	leão -> le'ão

Por fim, estendemos esta mesma lógica às sequências vocálicas heterossilábicas <aí> [ei] e [eu] <aú>, cuja segunda vogal é acentuada com diacrítico e pode surgir acompanhada de vogal em coda ou isolada, reutilizando a *regex* S11 (Tabela 28):

Tabela 28 – *Regex A6*

ID	Fronteira	Find	Replace	Exemplo
A6	V.'VC	(?<=a)(?=[íú])	'	país -> pa'ís saúde -> sa'úde

4.2.3.2.2. Palavras oxítonas

Embora no português a tonicidade de algumas palavras seja marcada lexicalmente, a maioria não possui qualquer diacrítico. Para estes casos, é possível delimitar determinados padrões de tonicidade, apresentados anteriormente em 1.2.3. O primeiro destes padrões corresponde às palavras oxítonas que tendem a terminar:

– em consoante <l>, <r>, <s>, <x> ou <z> (e.g., feliz → fe'liz)⁴⁴;

– em ditongo oral (e.g., bacalhau → baca'lhau);

– em vogal nasal (e.g., mamã → ma'mã);

– em ditongo nasal e vogal oral plena, que, por serem tipicamente marcados graficamente, já se encontram maioritariamente captados nas *regex* anteriores (e.g., televisão → televi'são, cá → 'cá), à exceção do ditongo [ẽj], grado (e.g., bem → 'bem) e da vogal [i] (e.g., aqui → a'qui).

Começámos por captar as palavras terminadas em consoante, considerando os tipos silábicos que podem ocorrer em final de palavra neste mesmo contexto: CVC (e.g., mar → 'mar), CVGC (e.g., pois → 'pois), CCVC (e.g., flor → 'flor) e VC (e.g., sair → sa'ir). De modo a captarmos as primeiras três sequências formulámos a *regex* representada na Tabela 29, na qual a vogal final corresponde sempre a <l>, <r>, <s>, <z>, excluindo-se desta classe o grafema <x> por não ocorrer em coda final no nosso *corpus*.

Tabela 29 – *Regex A7*

ID	Posição	Find	Replace	Exemplo
A7	'CVC 'CVGC 'CCVC	(?<!'.*)(?=(b c ç d f gu g ch nh lh j m n p qu rr ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)(a e i o u ei oi ui ao au eu iu)[rlsz]\$)	'	natal -> na'tal

⁴⁴ Note-se que o grafema <n> apenas ocorre em empréstimos recentes, tipicamente grafados com acento (e.g., abdómen, hífen, etc.)

Decidimos ainda excluir da classe das consoantes ocorrentes em ataque silábico os grafemas <r> e <s> pelas mesmas razões anteriormente mencionadas: o grafema <h>, para que não interferisse com a divisão dos dígrafos de que é membro (*e.g.*, <ch>, <lh>, <nh>) e por não surgir em ataque silábico em interior de palavra ou em ataque inicial de monossílabo; e o grafema <l>, para que não interferisse com a divisão dos encontros consonânticos de que é membro (*e.g.*, <pl>, <bl>, etc). Para além disto, introduzimos uma *negative lookbehind* nesta e nas restantes expressões para que nenhuma sequência já com marcador silábico fosse duplamente marcada (*e.g.*, di'fícil → *di'fí'cil).

Tendo retirado os grafemas <l>, <r> e <s> da classe de consoantes em ataque, ficou em falta este mesmo contexto. Posto isto, seguimos o mesmo procedimento adotado para A2 e A3 e compusemos três expressões individuais (Tabela 30):

- com <r> em ataque silábico, exceto quando se encontra precedido por <r> e pelas consoantes com as quais forma encontro consonântico
- com <l> em ataque silábico, exceto quando se encontra precedido pelas consoantes com as quais forma encontro consonântico;
- com <s> em ataque silábico, exceto quando se encontra precedido por <s>.

Tabela 30 – Regex A8, A9 e A10

ID	Posição	Find	Replace	Exemplo
A8	'CVC	(?<!'.*)(?<![pbtdcgfvr])(?=r(a e i o u ei oi ui ao au eu iu ou)[rlsz]\$)	'	adorar -> ado'rar
A9		(?<!'.*)(?<![pbtdcgfv])(?=l(a e i o u ei oi ui ao au eu iu ou)[rlsz]\$)	'	falar -> fa'lar
A10		(?<!'.*)(?<!s)(?=s(a e i o u ei oi ui ao au eu iu ou)[rlsz]\$)	'	pensar -> pen'sar

Note-se que a inclusão destes três contextos numa só expressão, com uma sequência do tipo (?<![pbtdcgfvr])(?=[lrs]) marcaria corretamente casos como mostrar → mos'trar ou passar → pa'ssar, contudo, palavras como *conversar* não seriam captadas pela *regex* devido à sucessão de <r> em coda e <s> em ataque.

Uma vez estabelecidos todos os contextos silábicos tónicos iniciados por consoante, partimos para as sequências iniciadas por vogal, nomeadamente para os monossílabos como *ar* e *ir*, facilmente captados pela *regex* da Tabela 31.

Tabela 31 – *Regex* A11

ID	Posição	Find	Replace	Exemplo
A11	^'VC	(?<!'.*)(?=[aeiou][rlsz]\$)	'	ar -> 'ar

À semelhança do que ocorreu para a divisão silábica, casos em que <ai> corresponde à sequência vocálica V.V [ei] e não ao ditongo decrescente GV [aj] provocaram erros ao serem selecionados pela *regex* A7 (e.g., *cair* → *'cair). Para solucionarmos este problema, foi necessário, num primeiro instante, averiguarmos os contextos de ocorrência de [aj] e [ei]:

- [aj]: mais
- [ei]: cair, sair

Sendo que os contextos não coincidem, decidimos retirar a sequência <ai> da classe de vogais das *regex* anteriores (A7, A8, A9 e A10) e criar duas expressões distintas, apresentadas na Tabela 32, que contemplam as sequências 'CVVC e V'.VC.

Tabela 32 – *Regex* A12 e A13

ID	Posição/ Fronteira	Find	Replace	Exemplo
A12	'CVVC	(?<!'.*)(?=(b c ç d f gu g ch nh lh j m n p qu rr ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)ais\$)	'	mais -> 'mais
A13	V'.VC	(?<!'.*)(?<=a)(?=i[r]\$)	'	cair -> ca'ir

Contudo, é importante ressaltar que, conforme se observou anteriormente, um alargamento de *corpus* poderá levantar problemas quer para <ai> quer para <ui> numa eventual extensão de amostra, o que significa que estes casos terão de ser sempre assinalados para uma posterior intervenção manual.

Seguindo a mesma ordem das *regex* de translineação e a lógica das *regex* S7 e S8, partimos para a marcação das restantes sequências heterossilábicas terminadas em consoante, que nos permitiram marcar palavras como material → materi'al, passear → passe'ar e continuar → continu'ar, conforme representado na Tabela 33.

Tabela 33 – *Regex* A14 e A15

ID	Fronteira	Find	Replace	Exemplo
A14	V.'VC	(?<!.*)(?<=[ieo])(?=[aeo][rszl]§)	'	voar -> vo'ar
A15	CV.'VC	(?<!.*)(?<=[bcçdfhjklmnpstrvxwz]u)(?=[rsz l]§)	'	continuar -> continu'ar

Uma vez marcadas todas as palavras oxítonas terminadas em consoante, partimos para a marcação das palavras oxítonas terminadas em ditongo oral através da seleção da posição anterior a CVG (e.g., mau → 'mau) e a CCVG com a *regex* A16, de onde foram novamente excluídos os grafemas <h>, <l>, <r> e <s> da classe de consoantes em ataque. Tal como procedemos anteriormente, para A8, A9 e A10, compusemos também duas *regex* para <r> e <s> em ataque silábico (Tabela 34), sendo que <h> e <l> não ocorrem no nosso *corpus* neste contexto.

Tabela 34 – *Regex* A16, A17 e A18

ID	Posição	Find	Replace	Exemplo
A16	'CVG 'CCVG	(?<!.*)(?=(b c ç d f gu g ch nh lh j m n p qu rr ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)(ai ei oi ui ao au eu iu ou)§)	'	bacalhau -> baca'lhau
A17	'CVG	(?<!.*)(?<![pbtcdgfvrr])(?=(ai ei oi ui ao au eu iu ou)§)	'	rei -> 'rei
A18		(?<!.*)(?<![s])(?=(ai ei oi ui ao au eu iu ou)§)	'	museu -> mu'seu

Posteriormente, procedeu-se à identificação das sequências monossilábicas do tipo VG, correspondentes a ditongos orais decrescentes através da *regex* A19, apresentada na Tabela 35.

Tabela 35 – Regex A19

ID	Fronteira	Find	Replace	Exemplo
A19	^'VG	(?<'!.*)(?=(ai ei oi ui au eu iu ou)\$)	'	eu -> 'eu

Para além de terminarem em consoante e ditongo oral, as palavras oxítonas também podem terminar em vogal nasal, pelo que procedemos à seleção da posição anterior à sequência C $\tilde{V}$, à exceção de <h>, <l>, <r> e <s> em ataque silábico, terminada nas vogais nasais [ẽ], [ĩ], [õ] ou [ũ], grafadas respetivamente <am>, <im>, <om> ou <um>, através da *regex* A20, representada na Tabela 36. Adicionalmente, dada à inexistência de sequências C $\tilde{V}$ não monossilábicas com <h>, <l>, <r> ou <s> em ataque silábico no nosso *corpus*, procedemos à sua identificação através da *regex* A21, também apresentada na Tabela 36.

Tabela 36 – Regex A20 e A21

ID	Posição	Find	Replace	Exemplo
A20	'C $\tilde{V}$	(?<'!.*)(?=(b c ç d f gu g ch nh lh j m n p qu rr ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)[aiou][m]\$)	'	assim -> a'ssim
A21	^'C $\tilde{V}$	(?<'!.*)(?=(hlrs)[aiou][m]\$)	'	sim -> 'sim

Posteriormente, seleccionámos a posição anterior às sílabas terminadas em ditongo oral [ẽĩ], grafado , sendo o único ditongo do PE que não possui acentuação gráfica. No momento de elaboração destas expressões, representadas na Tabela 37, a acentuação gráfica imposta pelo AO90 revelou-se essencial, pois permitiu a desambiguação entre as palavras oxítonas com mais de uma sílaba terminadas no ditongo nasal grafado -em que recebem acento gráfico agudo (*e.g.*, ninguém) e as palavras paroxítonas terminadas no ditongo nasal grafado -em que não recebem qualquer acentuação gráfica (*e.g.*, viagem). Deste modo, procedemos à identificação dos monossílabos oxítonos terminados em ditongo nasal [ẽĩ], à exceção de <h>, <l> e <r> em ataque silábico, através das *regex* A22. Atente-se que não foi necessária a exclusão do grafema <s>, pois a sequência <ss> não ocorre em ataque silábico inicial,

bem como as sequências <nh>, <lh> ou <rr>, nem a inclusão de uma *regex* que contemplasse <h>, <l> ou <r> em ataque silábico em sequência monossilábica 'CŨĜ, pois não ocorrem no nosso *corpus*.

Tabela 37 – *Regex* A22

ID	Posição	Find	Replace	Exemplo
A22	^'CŨĜ	(?<!'.*)(?=(b c ç d f gu g ch j m n p qu r s t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)(em)\$)	'	sem -> 'sem

Por fim, passámos às sílabas tónicas CV terminadas em vogal [i], grafada <i>, que identificámos a partir da expressão contemplada na Tabela 38.

Tabela 38 – *Regex* A23

ID	Posição	Find	Replace	Exemplo
A23	'CV	(?<!'.*)(?<=(?=(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)[i])\$)	'	ali -> a'li

4.2.3.2.3. Palavras paroxítonas

O segundo padrão de acento regular do português corresponde às palavras terminadas em vogal <a>, <o> ou <e>, realizadas fonética e respetivamente como [e], [u] e [i], que tendem a ser paroxítonas. Importa ressaltar que, na construção destas *regex*, lidámos com uma sequência de duas sílabas, pois partimos do final da palavra para o início da sílaba tónica. Por este motivo, os padrões são significativamente mais variados e extensos, o que lhes confere uma camada adicional de complexidade. Dada a vasta quantidade de expressões elaboradas para a identificação das palavras paroxítonas e uma vez que o raciocínio subjacente à sua formulação não difere do que foi até ao momento exposto, optámos por apresentar a sua categorização na Tabela 39, evitando uma descrição exaustiva.

Tabela 39 – Categorização das *regex* de marcação da sílaba tônica das palavras paroxítonas

Sequências iniciadas por consoante			
Sequência	Padrões silábicos	Expressões	Exemplos
Sequências simples	'CV.CV, 'CV.CCV, 'CCV.CV, 'CCV.CCV, 'CVG.CV, 'CVG.CCV, CCVG.CV	A24-A28	'boca 'negro com'prido de'pressa
Sequências com coda preenchida	CVC.CV, 'CCVC.CV	A29-A33	'perna 'conversa
Sequências com vogal nasal	'CŨ.CV, 'CŨ.CCV, 'CCŨ.CV	A35-A37	impor'tante du'rante
Sequências com ditongo + vogal em fronteira silábica inicial	'CVG.V, 'CCVG.V	A38	al'deia com'boio 'maio 'praia
Sequências com estruturas vocálicas heterossilábicas	'CV.V, 'CCV.V	A39-A42	'dia pe'ssoa ale'gria
Sequências iniciadas por vogal			
Sequências simples	'V.CV, VG.CV, 'V.CCV, VG.CCV	A43	'ano 'oito
Sequências com coda preenchida	'VC.CV	A44	'algo 'urso
Sequências com vogal nasal em fronteira inicial de palavra	'Ũ.CV, 'Ũ.CCV	A45	'entre 'onde
Sequências iniciadas pela segunda vogal de uma estrutura vocálica heterossilábica	V'V.CV, V'V.CCV	A46	demasi'ado co'elho
Sequências iniciadas por vogal nasal, precedidas por vogal	V. 'Ũ.CV	A47	cli'ente do'ença a'inda

Sequências terminadas em ditongo nasal [ẽj], grafado 			
Sequências iniciadas por consoante	'CV.CṼ	A48-A50	ima'gem 'jovem
Sequências iniciadas por vogal	V.'V.CṼ	A51	vi'agem
Interjeições			
Sequências com vogal + <h>	'VC	A52	'ah 'oh

Como referido previamente, as *regex* de marcação da sílaba tónica não só selecionam posições como também captam fronteiras. Observemos, a título exemplificativo, as *regex* A45 e A47, representadas na Tabela 40.

Tabela 40 – *Regex* A45 e A57

ID	Posição	Find	Replace	Exemplo
A45	^'Ṽ.CV ^'Ṽ.CCV	(?<!'.*)(?=[aeiou][mn](b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)[aoe]š)	'	onde -> 'onde
ID	Fronteira	Find	Replace	Exemplo
A47	'Ṽ.CV	(?<!'.*)(?=[aeio])(?=[aeiou][mn](b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)[aoe]š)	'	ainda -> a'inda

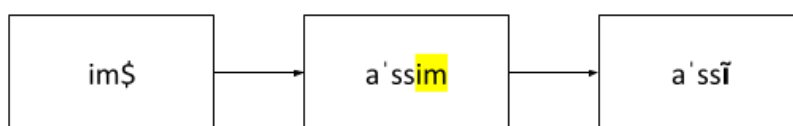
Uma vez que as vogais nasais apenas ocorrem com ataque vazio em dois contextos distintos, ou seja, em fronteira inicial de palavra ou em início de sílaba precedidas de vogal, surgiu a necessidade de compormos duas expressões que as captassem. Deste modo, a *regex* A45 seleciona a posição anterior às palavras paroxítonas iniciadas por vogal nasal (*e.g.*, onde → 'onde, entre → 'entre) e a *regex* A57 capta a fronteira entre vogal e as sequências iniciadas por vogal nasal (*e.g.*, ainda → a'inda, cliente → cli'ente). Note-se que a inclusão destes contextos numa só *regex* resultaria em marcações incorretas como quando → 'qu'ando ou tempo → 't'empo, o

que levou à necessidade de alargarmos o contexto da *regex* A57 e de captarmos uma fronteira silábica em vez de uma posição.

4.2.3.3. Transcrição fonética

O desenvolvimento das *regex* da transcrição fonética diferiu das restantes por três razões: (i) ao contrário do que ocorre na translineação e na marcação da sílaba tónica, não procurámos inserir qualquer carácter, mas sim substituir os que fazem parte do nosso *corpus* com um conjunto delimitado de caracteres fonéticos; (ii) certos contextos captados nestas *regex* são condicionados pelas *regex* da marcação da sílaba tónica; e (iii) a hierarquia em que se encontram dispostas é fundamental para a obtenção de bons resultados. Adicionalmente, tratam-se das expressões menos restritas deste sistema de conversão, pelo que não existe um padrão de pesquisa específico como nas anteriores, conforme podemos observar na Figura 13, potenciando a utilização de toda a sintaxe das *regex* previamente apresentada.

Figura 13 – Estrutura das *regex* de transcrição fonética da palavra *assim*



A inclusão de <p> [p], <t> [t], <f> [f] e <v> [v] nas *regex* não foi necessária pelo que a realização fonética destas consoantes corresponde à sua forma ortográfica canónica, não sofrendo qualquer alternância fonológica. De igual forma, foram excluídos certos contextos como em ataque silábico não intervocálico ([b]), <d> em ataque silábico não intervocálico ([d]), <g> em ataque silábico não intervocálico ([g]), <m> em ataque silábico intervocálico ([m]), <n> em ataque silábico intervocálico ([n]), <z> em ataque silábico ([z]), <s> em ataque silábico ([s]) e <l> em ataque silábico ([l]).

Num primeiro momento, procedeu-se à substituição de todos os hífenes dos compostos por um espaço em branco, utilizando a expressão T1, e à eliminação de todos os <h>, exceto quando precedidos de <c>, <l> ou <n>, através da expressão T2 (Tabela 41). A necessidade de aplicarmos estas expressões numa fase inicial deveu-se, por um lado, ao facto de o hífen (-) ser entendido como um carácter especial nas expressões

regulares (e.g., para definir um intervalo de caracteres [0-10]), o que dificultaria a identificação adequada dos compostos; e, por outro lado, ao facto de o grafema <h> funcionar como obstáculo à deteção de certos contextos, nomeadamente a fronteira inicial de palavra.

Tabela 41 – Regex T1 e T2

ID	Find	Replace	Exemplo
T1	–		bom-'dia -> bom 'dia
T2	(?<![cln])h		'hora -> 'ora

Tal como nas *regex* anteriores, seguimos a lógica de abordar primeiramente os casos que influenciam o maior número de expressões (as vogais) e só depois seguimos para as mais simples (consoantes). Visto que o *corpus* das *regex* de transcrição é constituído pelo *output* resultante das *regex* de marcação da sílaba tónica, começámos pelas vogais afetadas pelo marcador silábico: as vogais pretónicas e postónicas. Contudo, à medida que fomos desenvolvendo o sistema, revelou-se necessária a contemplação *a priori* da neutralização de altura de todos os <e> seguidos de sílaba iniciada por vogal pois este tipo de sequências era captado pelas *regex* pretónicas e postónicas resultando na redução errónea em palavras como le'ão → *lɨ'ão. Note-se que a adição de <o> à *negative lookbehind* não produziu melhores resultados, pois a conversão de palavras como 'vídeo → *'vídeu resultava numa posterior coincidência ortográfica entre o ditongo [ew] e a sequência vocálica heterossilábica [iu], resultando na conversão *'vídeu → *'vídew. Posto isto, as *regex* T3 e T4, apresentadas na Tabela 42, seleccionam todos os <e> seguidos das vogais [a, e, ě, e, ε, o, ɔ, õ]⁴⁵ em início de sílaba, tónica ou não.

⁴⁵ Note-se que, quando seguido de [i] em início de sílaba, realiza-se como [e] (e.g., veículo [ve'ikulu]) e que apenas ocorre seguido de [u] na mesma sílaba, formando o ditongo [ew].

Tabela 42 – Regex T3 e T4

ID	Find	Replace	Exemplo
T3	e(?=[áâãäēéêóôõö])	i	passe'ar -> passi'ar
T4	e(?=[áâãäēéêóôõö])	i	'área -> 'ária

As *regex* T5, T6 e T7 da Tabela 43 captam todos os contextos pretónicos existentes no nosso *corpus*, contudo, foi necessário ter em conta todos os contextos listados em 1.2.2.1. nos quais a redução vocálica não ocorre. Posto isto, as *regex* mencionadas apenas captam as vogais que (i) não ocorrem em fronteira inicial de palavra; (ii) não se encontram seguidas de marcador de nasalidade <m> ou <n> ou <l> em coda silábica (seguidos de consoante, marcador silábico ou espaço em branco); (iii) não se encontram seguidas de /l/ em coda silábica e (iv) não se encontram seguidos dos grafemas <i>, <o> ou <u> correspondentes às glides [j] e [w] com as quais formam ditongos decrescentes. Note-se que a utilização da *positive lookahead* (?=.*') permitiu que qualquer vogal anterior a marcador silábico e respeitando as condições acima listadas, independentemente da sua posição na frase, fosse detetada.

Tabela 43 – Regex T5, T6 e T7

ID	Find	Replace	Exemplo
T5	(?<!^)(?!([mnliu]['bcçdfhgjlmnpqrstvxz]))(?=.*')	u	to'mar -> tu'mar
T6	(?<!^)(?!([mnliou]['bcçdfhgjlmnpqrstvxz]))(?=.*')	e	sa'ber -> se'ber
T7	(?<!^)(?!([mnliu]['bcçdfhgjlmnpqrstvxz]))(?=.*')	ï	se'car -> si'car

Uma vez detetados todos os contextos pretónicos, seguimos para as vogais postónicas, começando pela margem direita de palavra. Tal como ocorreu para as vogais pretónicas, foi necessário impor algumas restrições, nomeadamente para <o> e <e> finais cujas *regex* T8 e T10, apresentadas na Tabela 44, foram configuradas para não captar ocorrências de <o> ou <e> finais antecidos do grafema <ã>, a fim de evitar a identificação indevida de ditongos decrescentes <ão> ou <õe> (*e.g.*, m – → *'mãu).

Tabela 44 – Regex T8, T9 e T10

ID	Find	Replace	Exemplo
T8	(?<![ã'])o\$	u	'golo -> 'golu
T9	(?<!['])a\$	e	'gota -> 'gote
T10	(?<![ã])e\$	í	'bife -> 'bifi

Além das vogais postônicas em fronteira final de palavra, surge a necessidade de detetarmos as vogais postônicas no interior de palavras proparoxítonas como *abóbora*, *pássaro*, *número* e *férias*, bem como as vogais postônicas em sílaba final de palavras proparoxítonas acentuadamente irregulares terminadas em consoante como *menos*, *apenas* e *simples*. Estes contextos encontram-se descritos pelas *regex* T11-T16 da Tabela 45, cuja elaboração partiu das sequências silábicas (tônicas) que ocorrem no nosso *corpus* antes de vogal postônica ('VC, 'CV, 'CCV, 'CVG, 'CCVG e 'CVCV) e dos contextos que constituem restrições à redução vocálica, excluindo-se todos os <o>, <a> e <e> postônicos seguidos de glide (<i>, <u>) e de <l> ou <r> em coda silábica final e marcador de nasalidade <m> em fronteira final de palavra. Atentemos que, no nosso *corpus*, apenas <a> ocorre depois de sequência silábica 'CVCV (*e.g.*, férias), detetada por T14, apenas <e> ocorre depois de sequência silábica 'C $\tilde{V}$ (*e.g.*, simples), detetada por T15, e apenas <o> ocorre depois de sequência silábica tónica 'VC (*e.g.*, árvore), detetada por T16.

Tabela 45 – Regex T11-T16

ID	Find	Replace	Exemplo
T11	(?<='(b c ç d f g gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)(a á â ã e é ê í î o ó ô õ u ú û ai ei oi ui ao au eu iu ou)(b c ç d f g gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))o(?![eiulrm])\$)	u	a'bóbore -> a'bóbure
T12	(?<='(b c ç d f g gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl)(a á â ã e é ê í î o ó ô õ u ú û ai ei oi ui ao au eu iu ou)(b c ç d f g gu g ch nh lh	e	'sábadu -> 'sábedu

	j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))a(?![eiulrm]\$)		
T13	(?<='(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))(a á â ã e é ê i í î o ó ô õ u ú û ai ei oi ui ao au eu iu ou)(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))e(?![iulrm]\$)	ï	'númeru -> 'númiru
T14	(?<='(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))(a á â ã e é ê i í î o ó ô õ u ú û ai ei oi ui ao au eu iu ou)(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))a(?![iulrm]\$)	e	'férias -> 'féries
T15	(?<='(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))[aeiou][mn](b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))e(?![iulrm]\$)	ï	'simples -> 'simplis
T16	(?<='(a á â ã e é ê i í î o ó ô õ u ú û ai ei oi ui ao au eu iu ou))[lrsx](b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))o(?![eiulrm]\$)	u	'árvori -> 'árvuri

Para além de integrarmos os contextos sobre os quais a redução vocálica incide, foi igualmente necessário captar os que representam exceções regulares a este mesmo processo fonológico. Para além da neutralização da altura de [e, ε] seguidos de sílaba iniciada por vogal, já acima detetada por T3 e T4, outra das exceções à redução vocálica são as vogais nasais [ẽ, ẽ̃, ĩ̃, õ̃, ã̃] constituídas por vogal e marcador de nasalidade <m> ou <n> seguidos de consoante (*e.g.*, limpu → 'lĩpu), de marcador silábico (*e.g.*, man'ter → mẽ'ter) ou de espaço em branco (*e.g.*, bom 'die → bõ 'die), ou acentuados graficamente com um til, como é o caso de <ã> (ma'mã → me'mẽ), que se encontram representadas na Tabela 46.

Tabela 46 – *Regex* T17-T25

ID	Find	Replace	Exemplo
T17	(im in ín)(?=['bcçdgjlmnpqrstvxz])	ĩ	'limpu -> 'lĩpu
T18	(om on ôn)(?=['bcçdgjlmnpqrstvxz])	õ	'longu -> 'lõgu
T19	(am âm an ân)(?=['bcçdgjlmnpqrstvxz])	ẽ	man'ter -> mẽ'ter
T20	(em en ên)(?=['bcçdgjlmnpqrstvxz])	ẽ	ven'der -> vẽ'der
T21	(um un)(?=['bcçdgjlmnpqrstvxz])	ũ	jun'tar -> jũ'tar
T22	im\$	ĩ	jer'dim -> jer'dĩ
T23	om\$	õ	'bom -> 'bõ
T24	ã(?!o)\$	ẽ	me'nhã -> me'nhẽ
T25	um\$	ũ	al'gum -> al'gũ

Posteriormente, passámos para as sequências vocálicas que formam ditongos decrescentes e que também não evidenciam redução vocálica. Todavia, foi necessário substituímos primeiramente todos os grafemas <j> correspondentes à consoante [ʒ] (Tabela 47), para que este não se confundisse com a semivogal [j], de modo a evitarmos erros como 'meio → 'meʒo ou 'quem → 'kẽʒ. Posto isto, procedemos à contemplação de [ʒ] em ataque silábico, precedido e seguido de vogal (*e.g.*, 'loje → 'loʒe) ou precedido de marcador silábico e seguido de vogal (*e.g.*, 'junhu → 'ʒunhu) através de T26. Apesar desta *regex* captar a maioria das ocorrências de [ʒ], faltava ainda captá-lo em sílaba átona e em fronteira inicial de palavra, detetada pela *regex* T27.

Tabela 47 – *Regex* T26 e T27

ID	Find	Replace	Exemplo
T26	(?<=['aáeěééēīĩōóōuúũ])j(?=['aáeěééēīĩōóōuúũ])	ʒ	'loje -> 'loʒe
T27	^j	ʒ	jer'dĩ -> ʒer'dĩ

Seguimos para a conversão das sequências vocálicas que formam ditongos decrescentes nasais (Tabela 48), dos quais se excluiu [õĩ], grafado <õe>, por não ocorrer

no nosso *corpus*, e ditongos decrescentes orais, dos quais se excluiu [iw], que surge tipicamente em verbos flexionados (*e.g.*, agiu), [ɛj], que surge tipicamente em formas do plural (*e.g.*, papéis) e é variável com [ej] e [ew] (*e.g.*, saudade [sɛw'ðãði]), por não ocorrerem no nosso *corpus*. Adicionalmente, foi necessária a adição da restrição da ocorrência de <g> e <q> antes de <ui> na *regex* T34, para que não obtivéssemos resultados desajustados em palavras como a'qui → *e'kuj.

Importa ainda ressaltar que, no nosso *corpus*, os grafemas <oi> correspondem a sete realizações de [oj] (*e.g.*, oito, depois, pois, boi, coisa, noite, dois) e apenas a uma realização de [ɔj] na palavra *comboio* que, tal como foi feito para *muito*, é detetado pela *regex* T34.

Tabela 48 – *Regex* T28-T38

ID	<i>Find</i>	<i>Replace</i>	Exemplo
T28	(ém em)\$	ẽ̃ ⁴⁶	'zovem -> 'zovẽ̃
T29	ão	ẽw	fei'zão -> fei'zẽw
T30	'muitu	'mũjtu	'muitu -> 'mũjtu
T31	ai	aj	'mais -> 'majs
T32	ei	ej	'feire -> 'fejre
T33	oi	oj	'noiti -> 'nojti
T34	kõ'boiu	kõ'bɔju	kõ'boiu -> kõ'bɔju
T35	(?<![gq])ui	uj	cui'dar -> cuj'dar
T36	(au ao)	aw	'mau -> 'maw
T37	eu	ew	'deus -> 'dews
T38	éu	ɛw	'céu -> 'cɛw

As vogais acentuadas graficamente integram geralmente a sílaba tónica da palavra, pelo que não sofrem naturalmente redução vocálica. Posto isto, procedemos à

⁴⁶ No nosso *corpus*, o ditongo [ẽ̃] não se encontra representado graficamente pela sequência <ãe>.

conversão das vogais tónicas acentuadas graficamente presentes no nosso *corpus* (Tabela 49), à exceção de <ê> e <ô>.

Tabela 49 – Regex T39-43

ID	Find	Replace	Exemplo
T39	í	i	i'níciu -> i'niciu
T40	é	ɛ	'pé -> 'pɛ
T41	ú	u	'únicu -> 'unicu
T42	ó	ɔ	'só -> 'sɔ
T43	á	a	'zá -> 'za

Por fim, abordámos os contextos das vogais átonas não altas seguidas de /l/ em coda silábica (Tabela 50), pelo que o nosso *corpus* não evidencia vogais átonas não altas com consoante diferente de /S/ em coda silábica final nem vogais átonas não altas coronais ou labiais com uma sequência de segmentos consonânticos heterossilábicos à sua direita. Tal como para [ʒ], foi necessário detetarmos primeiramente /l/ em coda silábica no interior de palavra precedido de vogal e seguido de consoante (e.g., 'altu → 'aɫtu) ou de marcador silábico (e.g., tal'vez → taɫ'vez), através de T44, e /l/ em coda silábica final (e.g., lo'cal → lu'caɫ) através de T45, de modo a simplificar as restantes expressões.

Importa notar que, neste contexto, a forma canónica gráfica de <a>, que coincidentemente corresponde ao símbolo fonético da vogal [a], fica intacto devido às restrições impostas pelas expressões das vogais pretónicas e postónicas, não havendo necessidade de uma *regex* que capte a persistência do seu valor baixo nesta circunstância. O mesmo ocorre para <o> e, embora [ɔ] átono possa ser admitido em determinados contextos (e.g., moldar), há uma tendência para a sua realização enquanto [o], conforme verificámos em 1.2.2.1. Considerando isto, o único contexto na qual a vogal não corresponde à sua forma gráfica canónica é o grafema <e> que, conforme referido previamente, apenas se converte em [e] em palavras como *feldspato*, *feltro* ou *felpo*, realizando-se [ɛ] nos restantes casos, detetados pela *regex* T46.

Tabela 50 – *Regex* T44-T46

ID	Find	Replace	Exemplo
T44	(?<=[aeêêiioôuww])(?=['bcçdfgzklmnpqrstvxz])	†	'altu -> 'a†tu
T45	\$l	†	lo'cal -> lu'ca†
T46	(?<!(b c ç d f gu g ch nh lh j l m n p qu r rr s ss t v x z pr br tr dr cr gr fr vr pl bl cl gl fl))e(?=†)	ε	pu'ssive† -> pu'ssive†

A fronteira inicial de palavra também constitui um dos contextos no qual a redução vocálica não se verifica, embora se registe uma tendência da redução de /a/ para [e] em contexto átono, à exceção de um conjunto específico de palavras, pelo que adotámos esta mesma tendência na *regex* A45. Por sua vez, as vogais médias registam alguma variação, nomeadamente entre falantes mais e menos conservadores, que materializam uma preferência para os pares [i, o] e [e, ɔ], respetivamente. Como referido anteriormente em 3.1.3., optou-se pela representação de ambas as pronúncias, separadas por uma barra oblíqua, através de T48 e T49 (Tabela 51), excluindo-se todos os <o> e <e> iniciais tónicos ou seguidos de semivogal e /l/ em coda silábica. Note-se que, em T48, foi ainda necessário incluir o carácter isolado [~], uma vez que <ẽ> não constitui uma combinação orgânica em português, ao contrário de <õ> ou <ã>, e, portanto, não ser reconhecida pelo sistema como um carácter, mas sim como a junção dos caracteres [e] + [~]. Surgiu ainda a necessidade de adicionarmos o grafema <s> à *negative lookahead* desta *regex* para que não fossem incluídas as sequências do tipo #sC ocorrentes em palavras como *estudar*, *escolher* e *esperar*, posteriormente captadas por T50 da Tabela 51. Importa ainda ressaltar que nas palavras *olá*, *hotel* e *hospital*, cuja pronúncia para a generalidade dos falantes é [ɔ], apenas foi incluída esta mesma realização.

Tabela 51 – Regex T47-T50

ID	Find	Replace	Exemplo
T47	(?<!')^a(?![jwɥ])	e	a'mar -> e'mar
T48	(?<!')^o(?![juɥ])	ɔ / o	o'lhar -> ɔ'lhar / o'lhar
T49	(?<!')^e(?![juɥʃ])	e / i	e'quipe -> e'quipe / i'quipe
T50	(?<!')^es(?![aaeeêëioôuwwẽĩõũ])	ɨj	es'pere -> ɨj'pere

Uma vez contemplados todos os contextos vocálicos átonos, passámos para as variações ocorridas em contexto tónico através de T51 da Tabela 52, nomeadamente /a/, que tende a realizar-se [e] antes de consoante nasal.

Tabela 52 – Regex T51

ID	Find	Replace	Exemplo
T51	a(?=(m n nh))	e	si'mane -> si'mene

Dado que /i/ e /u/ não se alteram independentemente da sua tonicidade, prosseguimos para a integração do contexto mais problemático do *corpus*, – as vogais médias /e, ɛ/ e /o, ɔ/ –, cuja imprevisibilidade representa um desafio para o sistema de conversão. Esta dificuldade relaciona-se com o facto de estes dois pares de vogais se encontrarem, em contexto tónico e sem qualquer acentuação gráfica, numa espécie de distribuição livre. Quer isto dizer que o conhecimento de que os grafemas <e> e <o> se realizam respetivamente como [e] e [o] nas palavras *mesa* e *zona* e como [ɛ] e [ɔ] nas palavras *janela* e *bola* advém exclusivamente do conhecimento linguístico intrínseco do falante nativo, não estando dependente de um contexto fonológico previsível.

O nosso *corpus* dispõe um total de 107 ocorrências de <e> tónico, das quais 76 são realizadas como [e] (*e.g.*, pena, parede, etc.) e as restantes 32 como [ɛ] (*e.g.*, amarelo, belo, etc.) e de um total de 76 ocorrências de <o> tónico, das quais 48 são realizadas como [o] (*e.g.*, doçe, novo, etc.) e as restantes 28 como [ɔ] (*e.g.*, enorme, forte, etc.)⁴⁷. Como podemos verificar, as vogais semifechadas [e] e [o] correspondem,

⁴⁷ Excluimos desta quantificação as palavras homógrafas não homófonas (*e.g.*, este, colher e forma), cuja realização da vogal tónica média pode ser descodificada a partir do seu significado.

respetivamente, a 71.03% e a 63.16% das ocorrências de <e> e <o> tónicos, surgindo em mais de metade das palavras que contêm este contexto. Ainda assim, a previsão sistemática destas ocorrências revelou-se inviável e a transcrição manual destes 184 casos não só se revelaria ineficiente, como contrariaria o principal intuito desta investigação, isto é, a automatização da transcrição fonética para o PE a partir da sua representação ortográfica.

Considerando o exposto, decidimos recorrer a uma representação mais fonológica, a partir de dois arquifonemas – /E/ e /O/ – que representassem respetivamente as vogais médias [e, ɛ] e [o, ɔ], salvaguardando este contexto no sistema de conversão. Como podemos verificar na Tabela 53, as *regex* T52 e T53 permitiram a captação dos <o> tónicos precedidos de consoante ou sequência consonântica (*e.g.*, 'docɪ → 'd/O/cɪ) e em fronteira inicial de palavra ('oʒɪ → '/O/ʒɪ), tal como T55 e T56 para os <e> tónicos.

A inserção das expressões regulares T54 e T58, responsáveis pela transcrição dos grafemas <ô> e <ê>, teve de ocorrer após as *regex* T52, T53, T56 e T57. Neste caso particular, a hierarquia revelou-se fundamental uma vez que, caso as expressões T54 e T58 tivessem sido aplicadas antes, a sua transcrição enquanto [e] e [o] levaria a uma incorreta e desnecessária aplicação de T52, T53, T56 e T57 (*e.g.*, pur'quê → pur'que → * pur'qu/E/). O mesmo se verificou para a fusão das labiais correspondente à sequência de grafemas <ou>, representada pela *regex* T55.

Tabela 53 – Regex T52-T58

ID	Find	Replace	Exemplo
T52	(?<='(b c ch d f g z l lh m n nh p q r rr ss s t v x z pr br tr dr kr gr fr vr pl bl kl gl fl))o(?![ju])	/O/	'doc <i>i</i> -> 'd/O/c <i>i</i>
T53	(?<=')o(?![ju])	/O/	'o3 <i>i</i> -> ' /O/3 <i>i</i>
T54	ô	o	'pô <i>r</i> -> 'por
T55	ou	o	ou'vir -> o'vir
T56	(?<='(b c ch d f g gu z l lh m n nh p q qu r rr ss s t v x z pr br tr dr kr gr fr vr pl bl kl gl fl))e(?![w])	/E/	'vez -> 'v/E/z
T57	(?<=')e(?![w])	/E/	'est <i>i</i> -> ' /E/st <i>i</i>
T58	ê	e	pur'quê -> pur'que

Concluída a deteção de todos os contextos vocálicos do *corpus*, seguimos para a conversão das consoantes, começando pelas consoantes sem quaisquer restrições contextuais, i.e., as consoantes cujo(s) grafema(s) correspondem a um e a um só som independentemente dos segmentos que os rodeiam, através das *regex* T59, T60 e T61 da Tabela 54. Por sua vez, o grafema <x> apresenta duas realizações possíveis no *corpus*, realizando-se [ʃ] seis vezes (*e.g.*, puxar, mexer, deixar, texto, peixe, baixo) e [z] apenas na palavra *exame*. Tal como procedemos para os grafemas <oi>, optámos igualmente por atribuir a <x> o valor de [ʃ], corrigido manualmente a palavra *exame*. Importa, no entanto, sublinhar que esta generalização poderá constituir um entrave em futuros alargamentos da amostra a aplicar no sistema de conversão.

Tabela 54 – Regex T59-T62

ID	Find	Replace	Exemplo
T59	ch	ʃ	'chẽw̃ -> 'jẽw̃
T60	nh	ɲ	'banhu -> 'baɲu
T61	lh	ɭ	o'lhar -> o'ɭar
T62	x	ʃ	pu'xar -> pu'ʃar

Posteriormente, contemplámos todas as consoantes que sofrem alternâncias fonológicas ou variações fonéticas, começando pela fricativação das oclusivas /b/ e /d/ átonas e tónicas representadas respetivamente por T63, T64 e T65, T66 da Tabela 55. A necessidade de uma expressão que captasse estes grafemas em posição intervocálica tónica justificou-se pelo facto de a fricativação geralmente não ocorrer se a vogal ou glide precedente for nasal.

Tabela 55 – Regex T63-T66

ID	Find	Replace	Exemplo
T63	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u w))b(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w ẽ ẽ ĩ õ ũ j w̃))	β	'sabedu -> 'saβedu
T64	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u j w))'b(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w ẽ ẽ ĩ õ ũ j w̃))	β	se'bor -> se'βor
T65	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u w))d(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w ẽ ẽ ĩ õ ũ j w̃))	ð	'modu -> 'moðu
T66	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u j w))'d(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w ẽ ẽ ĩ õ ũ j w̃))	ð	e'duʈtu -> e'ðuʈtu

Por sua vez, o grafema <g> exigiu a definição de um conjunto mais alargado de restrições, devido à variação fonética que evidencia, realizando-se como: [ʒ] quando seguido dos grafemas <e> ou <i> (T67); [g] quando seguido de <u> e dos grafemas <e> ou <i> (T68); [gʷ] quando, tipicamente, seguido de vogal <a> (T69); e [ɣ], em posição intervocálica (T70, T71). Note-se que em T67 e T68, representados na Tabela 56, são incluídas todas as vogais que podem corresponder aos grafemas <e> e <i>, daí a inclusão

A hierarquia em que as expressões correspondentes aos grafemas <s> e <c> se encontram distribuídas, conforme podemos verificar na tabela abaixo, revelou-se particularmente relevante, uma vez que os contextos associados a estes grafemas se influenciam mutuamente, o que pode conduzir à ocorrência de erros na conversão. A título exemplificativo, temos a expressão T74 cuja inserção antes de T72 e T73 resultaria na conversão de <ç> em [s], que, ao encontrar-se em posição intervocálica, transformar-se-ia erradamente em [z], resultando em sequências de transcrição incorretas como *dê'çar* → *dê'sar* → **dê'zar*. De igual forma, todos os contextos em que [s] ocorre teriam de estar contemplados antes das *regex* T78 e T79, caso contrário, palavras como *nascer*, em que <c> se realiza [s] e consequentemente <s> se realiza [ʃ], não seriam captados.

Tabela 57 – Regex T72-T84

ID	Find	Replace	Exemplo
T72	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u w))s(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w))	z	'ase -> 'aze
T73	(?<=(a e e ε /E/ ĩ i j o ɔ /O/ u w))'s(?=(a e e ε /E/ ĩ i j o ɔ /O/ u w))	z	ce'sar -> ce'zar
T74	ç	s	'forçe -> 'forse
T75	c(?!(ej ẽ e ε /E/ ẽ i ĩ ı))	k	'marce -> 'marke
T76	c(?=(ej ẽ e ε /E/ ẽ i ĩ ı))	s	'certu -> 'sertu
T77	s{2}	s	'issu -> 'isu
T78	s(?=[ptkfs])	ʃ	'istu -> 'iftu
T79	s(?='[ptkfs])	ʃ	nes'ser -> neʃ'ser
T80	s(?=[bdgmnjrvzlʌ])	ʒ	'mesmu -> 'meʒmu
T81	s\$	ʃ	'majs -> 'majʃ
T82	z\$	ʃ	e'rroz -> e'rroʃ
T83	qu(?=(ej ẽ e ε /E/ ẽ i ĩ ı))	k	e'qui -> e'ki
T84	qu(?=(a e ẽ))	k ^w	'quał -> 'k ^w ał

O último grafema consonântico a representar é o grafema <r> (Tabela 58), pelo que foi necessário compormos expressões para os diferentes contextos de ocorrência das duas realizações: [R] e [r]. Dado que no nosso *corpus* <r> não ocorre precedido de /S/ ou /l/ em coda silábica, precedido de autossegmento nasal ou em posição inicial de palavra fonológica, apenas foi necessária a captação de [R] em ataque silábico inicial, através de T85 e T86, e em posição intervocálica, grafado <rr>, através da *regex* T87. Por sua vez, contemplámos [r] em coda silábica no interior e no final de palavra, através das expressões T88 e T89, em posição intervocálica átona e tónica, através das expressões T90 e T91, e em ataque silábico ramificado, precedida de consoante oclusiva ou fricativa, através da expressão T92.

Tabela 58 – Regex T85-T92

ID	Find	Replace	Exemplo
T85	$\wedge r$	R	re'paz -> re'paz
T86	$\wedge(?(<=['])r$	R	'rej -> rej
T87	$r\{2\}$	R	'terre -> 'tere
T88	$(?(<=(a e e \epsilon /E /i i j o o /O /u w))r(?(=[\text{'bdfgjn}\Lambda\text{3klmnpstvz}]))$	r	'ursu -> 'ursu
T89	$r\$$	r	e'mor -> e'mor
T90	$(?(<=(a e e \epsilon /E /i i j o o /O /u w))r(?(=[a e e \epsilon /E /i i j o o /O /u w \tilde{e} \tilde{e} \tilde{i} \tilde{o} \tilde{u} \tilde{j} \tilde{w}))$	r	'karu -> 'karu
T91	$(?(<=(a e e \epsilon /E /i i j o o /O /u w))'r(?(=[a e e \epsilon /E /i i j o o /O /u w \tilde{e} \tilde{e} \tilde{i} \tilde{o} \tilde{u} \tilde{j} \tilde{w}))$	r	netu'rał -> netu'rał
T92	$(?(<=[\text{pbtdkgfv}])r$	r	'friú -> 'friú

4.2.3.4. Importação para o DLP-ACL (2025)

Concluída a aplicação do sistema de conversão, os *outputs* obtidos a partir das expressões regulares da translineação e da transcrição fonética foram integrados automaticamente no DLP-ACL (2025), recorrendo ao editor de dicionários *LeXmart*⁴⁸

⁴⁸ <https://lexmart.eu/>

(Simões & Salgado, 2022), por meio de dois ficheiros ODS⁴⁹. Considerando que o DLP-ACL (2025) adota o formato TEI Lex-0⁵⁰ (Salgado *et al.*, 2019 a codificação desses elementos foi realizada através das etiquetas <syll>⁵¹ e <pron>⁵², respetivamente, à translineação e à transcrição fonética.

Terminada a descrição do sistema de conversão, procede-se agora à análise dos resultados obtidos com a sua aplicação ao *corpus* selecionado.

5. Análise e discussão dos resultados

Neste capítulo, procedemos à apresentação e análise dos resultados obtidos através da aplicação do sistema de conversão, respetivamente, na translineação (secção 4.1.), na marcação da sílaba tónica (secção 4.2.) e na transcrição fonética (secção 4.3.), à luz dos objetivos e dos critérios previamente delineados. Adicionalmente, discutimos os resultados da integração dos dados no DLP-ACL (2025) (secção 4.4.).

A totalidade dos resultados obtidos pelo sistema de conversão desenvolvido encontram-se representados no Anexo 6, pelo que a coluna A corresponde ao *output* das expressões regulares de translineação, a coluna B ao *output* das expressões regulares da marcação da sílaba tónica e a coluna C ao *output* das expressões regulares da transcrição fonética. As células assinaladas a cinzento representam os casos que necessitaram de ajustes manuais ou que não foram incluídos em determinado *corpus*, como é o caso das palavras não acentuadas na coluna B.

5.1. Translineação

As *regex* da translineação permitiram a correta translineação dos 730 lemas integrantes do *corpus*, não se revelando necessária a intervenção manual nos resultados. Este resultado reflete a eficácia das expressões regulares e dos critérios silábicos aplicados na definição das regras de translineação, o que poderá ser explicado pela natureza relativamente sistemática e estável da silabificação do PE e pela raridade

⁴⁹ Disponíveis em <https://github.com/LeonorMartins/G2P-DLP/tree/main/ODS>

⁵⁰ <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

⁵¹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-syll.html>

⁵² <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-pron.html>

de ocorrência de grupos consonânticos que violem o princípio de sonoridade e a condição de dissemelhança.

Com este sistema, foram captadas palavras de diferentes classes gramaticais (substantivos, adjetivos, advérbios, pronomes, entre outros) e de diferentes estruturas fonológicas e silábicas, incluindo vogais nasais, ditongos decrescentes, sequências vocálicas heterossilábicas, grupos consonânticos próprios, monossílabos e compostos hifenizados. Adicionalmente, foram detetadas todas as fronteiras silábicas existentes em PE (C.C, V.C, G.C, G.V, V.V), conforme podemos observar na Tabela 59, à exceção de grupos consonânticos impróprios, como <gn>, <pt>, <nm>, entre outros, cuja integração no sistema de conversão será objeto de exploração futura.

Tabela 59 – Amostra dos resultados de translineação

Palavra	Translineação
antigo	an.ti.go
baixo	bai.xo
vídeo	ví.de.o
branco	bran.co
bom	bom
boa-tarde	bo.a-tar.de

O desenvolvimento das expressões regulares de translineação permitiu ainda destacar determinados contextos silábicos do PE que poderão representar uma futura limitação aquando da expansão do *corpus* deste sistema de conversão, nomeadamente as sequências de grafemas como <ui> e <ai>, que podem corresponder simultaneamente a ditongos decrescentes e a sequências vocálicas (*e.g.*, fluidez → flu.i.dez vs. cuidar → cui.dar; rainha → ra.i.nha vs. baixo → bai.xo).

Em suma, a translineação pode ser tratada computacionalmente com elevado grau de precisão. A previsibilidade das estruturas silábicas da língua, aliada à adequação das expressões regulares desenvolvidas, contribuiu para um desempenho sólido e consistente nesta fase do sistema de conversão.

5.2. Marcação da sílaba tónica

À semelhança do que ocorreu na translineação, todas as sílabas tónicas do *corpus* 2 foram marcadas corretamente, à exceção das sílabas tónicas pertencentes às palavras simples → *sim'ples, apenas → *ape'nas e menos → *me'nos. Estes casos correspondem a verdadeiras exceções que escapam ao paradigma geral acentual não-verbal do português por se tratar de palavras paroxítonas terminadas em consoante. A criação de expressões regulares que captassem estes casos não seria naturalmente plausível, pelo que interferiria com a generalidade das palavras do *corpus* (99,58%) que seguem o padrão regular de acentuação da língua portuguesa. Posto isto, a correção da sua acentuação teve de ser feita manualmente, resultando na conversão apresentada na Tabela 60.

Tabela 60 – Correção manual das palavras acentualmente irregulares

Antes	Depois
sim'ples	'simples
ape'nas	a'penas
me'nos	'menos

Note-se que estes ajustes provêm do conhecimento de falante nativo e tiveram de ser feitos antes da sua integração no *corpus* da transcrição fonética, de modo a evitarmos que fossem erroneamente captados, por exemplo, pelas expressões dedicadas às vogais pretónicas (*e.g.*, *ape'nas → *[apɨ'nas]).

Importa também ressaltar que as palavras paroxítonas terminadas em ditongo nasal [ẽj], grafado , como vi'agem, 'homem ou o'rigem constituem igualmente casos de palavras acentualmente irregulares. Todavia, foi possível prever estes contextos por meio de expressões regulares, uma vez que a própria ortografia permite desambiguar a acentuação: as palavras oxítonas com mais de uma sílaba terminadas em apresentam acento agudo (*e.g.*, ninguém); assim, todas as palavras com mais de uma sílaba terminadas em e sem qualquer acento agudo são paroxítonas.

Mesmo perante as três palavras excepcionais de acento irregular, as expressões regulares desenvolvidas permitiram a correta marcação silábica dos 710 vocábulos restantes, o que corresponde a uma taxa de acerto de 0,42%. Adicionalmente, permitiu a deteção de todos os padrões acentuais do PE: palavras proparoxítonas; palavras paroxítonas terminadas em vogal [e], [u] e [i], grafadas respetivamente <a>, <o> e <e>, e palavras oxítonas terminadas em consoante, em ditongo oral, em vogal oral plena, em vogal nasal e em ditongo nasal, exemplificadas na Tabela 61.

Tabela 61 – Amostra dos resultados da marcação da sílaba tónica

Palavra	Marcação da sílaba tónica
simpático	sim'pático
temperatura	tempera'tura
trabalho	tra'balho
durante	du'rante
feliz	fe'liz
bacalhau	baça'lhau
café	ca'fé
manhã	ma'nhã
região	regi'ão

5.3. Transcrição fonética

5.3.1. Consoantes

Os resultados positivos obtidos pelo sistema de conversão na translineação e na marcação da sílaba tónica estenderam-se de igual forma à transcrição fonética das consoantes, pelo que as expressões elaboradas permitiram captar todas as correspondências consonânticas grafema-fone[ma] do *corpus*, conforme podemos observar a partir da Tabela 62.

Tabela 62 – Amostra de resultados da transcrição fonética dos segmentos consonânticos

Grafema	Fone[ma]	Exemplo
<p>	[p]	palavra -> [pe'lavre]
	[b]	barulho -> [be'ruʎu]
<a,e,i,o,u b a,e,i,o,u>	[β]	trabalho -> [tre'βaʎu]
<t>	[t]	três -> ['treʃ]
<d>	[d]	domingo -> [du'mĩgu]
<a,e,i,o,u d a,e,i,o,u>	[ð]	nada -> ['naðe]
<c ^{a,o,u} >	[k]	cozinha -> [ku'ziɲe]
<qu ^{e,i} >		queijo -> ['kejʒu]
<qu ^{a,o,u} >	[kʷ]	quarto -> ['kʷartu]
<g ^{a,e,o} >	[g]	gostar -> [gu'tar]
<gu ^{e,i} >		alguém -> [aʎ'gẽ]
<a,e,i,o,u g a,e,i,o,u>	[ɣ]	ligar -> [li'ɣar]
<gu ^{a,e,o} >	[gʷ]	guardar -> [gʷer'dar]
<m>	[m]	macaco -> [me'kaku]
<n>	[n]	nenhum -> [ni'ɲũ]
<nh>	[ɲ]	vinho -> ['viɲu]
<r>	[R]	rei -> ['Rej]
<rr>		sorrir -> [su'Rir]
<r>	[r]	marca -> ['marke]
<f>	[f]	futuro -> [fu'turu]
<v>	[v]	verão -> [vi'rẽw]

<s>	[s]	sapato -> [se'patu]
<ss>		passo-> ['pasu]
<c ^{e,i} >		fácil -> ['fasiɫ]
<ç>		braço -> ['brasu]
<a,e,i,o,u_s a,e,i,o,u>	[z]	usar -> [u'zar]
<z>		cozinha -> [ku'ziɲe]
<x>		exemplo -> [e'zẽplu] / [i'zẽplu]*
<s>	[ʃ]	seis -> ['sejʃ]
<z>		rapaz -> [re'paʃ]
<ch>		chuva -> ['fuve]
<x>		puxar -> [pu'ʃar]
<g ^{e,i} >	[ʒ]	gente -> ['ʒẽti]
<j>		jantar -> [ʒẽ'tar]
<l>	[l]	local -> [lu'kaɫ]
	[ɫ]	local -> [lu'kaɫ]
<lh>	[ʎ]	alho -> ['aʎu]

O único contexto consonântico que não foi captado pelas expressões regulares relativas às consoantes foi o grafema <x>, pertencente à palavra *exemplo*, cuja realização [z] optámos por ajustar manualmente, conforme a Tabela 63.

Tabela 63 – Correção manual da palavra *exemplo*

Antes	Depois
[e'ʃẽplu] / [i'ʃẽplu]	[e'zẽplu] / [i'zẽplu]

O facto de o grafema <x> não ser um grafema originário do português, surgindo maioritariamente em empréstimos e palavras eruditas, e de possuir diferentes realizações ([ʃ], [s], [z], [ks]), das quais apenas alguns contextos são previsíveis – não se

incluindo entre eles o caso aqui apresentado –, representa uma limitação deste sistema de conversão.

Durante o desenvolvimento destas expressões, foi possível ainda destacar um outro contexto potencialmente problemático num eventual alargamento do *corpus* – a realização das sequências <gue>/<gui> como [g^w] e de <que>/<qui> como [k^w] em palavras como *linguiça* e *equidade* e que tipicamente se realizam como [g] e [k] em palavras como *seguir* e *querer* –, devido à ambiguidade na correspondência entre grafema-fone[ma]. É de salientar, contudo, que no *corpus* atual estas ocorrências não suscitaram qualquer incorreção, tendo sido corretamente tratadas pelas expressões regulares implementadas.

5.3.2. Vogais átonas

Ao contrário do que se verificou no tratamento dos segmentos consonânticos, os segmentos vocálicos revelaram-se substancialmente mais problemáticos devido à sua maior sensibilidade ao contexto morfológico, à variabilidade das suas realizações e à presença de casos excecionais. Esta instabilidade traduziu-se naturalmente numa acrescida dificuldade na criação de expressões regulares, originando uma maior margem de erro e exigindo, em certos casos, um tratamento mais específico ou excecional.

5.3.2.1. Vogais pretónicas

Entre os principais obstáculos encontrados na transcrição fonética das vogais átonas, destaca-se o tratamento das vogais pretónicas pertencentes às palavras *bebé*, *normal*, *procurar*, *sobretudo*, *esqueceer* e ao composto *boa-tarde*, que, por serem pretónicas, foram captadas pelas *regex* T5 e T7 e sofreram redução para [u] e [i], resultando na sua incorreta transcrição como *[bɨ'βɛ], *[nur'maɪ], *[pruku'rar], *[subrɨ'tuðu], *[eski's/E/r] e *[bue 'tardɨ], respetivamente. Adicionalmente, regista-se ainda a forma verbal *ganhar*, na qual a realização da vogal átona /a/ varia entre [a] e [ɐ]⁵³. Estas ocorrências constituem casos que não obedecem às regras de redução vocálica, mas também não se encontram condicionadas pelo contexto, integrando verdadeiras exceções que apenas se encontram marcadas no léxico mental dos falantes

⁵³ Ver Andrade (2020: 3290).

nativos e que dão lugar à neutralização de altura entre as vogais não altas a favor das baixas [ɛ, ɔ, a]. Posto isto, procedemos à sua correção manual, conforme a Tabela 64 ilustra.

Tabela 64 – Correção manual dos casos excecionais em contexto pretónico

Antes	Depois
[bi'βɛ]	[bɛ'βɛ]
[nur'mat]	[nɔr'mat]
[pruku'rar]	[pɔku'rar]
[subri'tuðu]	[sobri'tuðu]
[i]kɪ's/E/r]	[i]kɛ's/E/r]
[bue'tardi]	[boe'tardi]
[ge'nar]	[ge'nar] / [ga'nar]

O mesmo ocorreu para as vogais pretónicas dos advérbios *normalmente* e *rapidamente*, dado que as vogais acentuadas de base ([ɔ] e [a]) não sofrem redução vocálica nas palavras que integram sufixos z-avaliativos e o sufixo *-mente*. No caso de *normal*, como vimos, estamos perante um caso excecional em que a vogal pretónica de base não sofre redução vocálica (e.g., [nur'mat]). Por sua vez, *rápida* trata-se de uma palavra proparoxítone, cuja vogal se encontra acentuada graficamente por um diacrítico (<á>) que cai quando lhe é acrescentado o sufixo *-mente*, o que dificulta a sua desambiguação por parte do sistema de conversão. Este cenário estendeu-se de igual forma à vogal pretónica do adjetivo terminado em sufixo -z avaliativo *sozinho*, cuja vogal acentuada de base [ɔ], pertencente à palavra de base *só* e acentuada graficamente <ó>, perde o seu acento, o que dificulta a sua deteção. Considerando o exposto, procedemos também à correção manual destes casos, representados na Tabela 65.

Tabela 65 – Correção manual de advérbios em *-mente* e de adjetivos com sufixo *-z* avaliativo

Antes	Depois
[nurmaɫ' mēti]	[nɔɾmaɫ' mēti]
[ɾepiðe' mēti]	[ɾapiðe' mēti]
[su' ziɲu]	[sɔ' ziɲu]

Note-se que os advérbios *finalmente* e *geralmente* obtiveram resultados corretos na transcrição ([finaɫ' mēti], [ʒiɾaɫ' mēti]) pelo facto de as suas vogais acentuadas de base ([a]) se encontrarem seguidas de /l/ em coda silábica, não sofrendo consequentemente qualquer redução vocálica, apesar de se encontrarem em posição pretónica nas formas adverbiais.

5.3.2.2. Vogais em início absoluto de palavra

Para além das vogais pretónicas, registaram-se igualmente alguns casos excecionais nos quais /a/ átono inicial é realizado [a] e que escapam ao paradigma geral da redução vocálica em início absoluto de palavra. Trata-se das palavras *ação*, *amanhã* e *ator*, cuja ausência de redução não é, tanto quanto sabemos, contextualmente condicionada, pelo que pertencem a um conjunto limitado de palavras excecionais como *Aveiro*, *além*, *arma*, entre outras. Por não se tratar de uma variação contextual, e por dependerem da gramática implícita do falante nativo, procedemos à sua correção manual (Tabela 66).

Tabela 66 – Correção manual dos casos excecionais em início absoluto de palavra

Antes	Depois
[e' sēw̃]	[a' sēw̃]
[eme' ɲē]	[ame' ɲē]
[e' t/O/r]	[a' t/O/r]

5.3.2.3. Palavras não acentuadas

Por fim, ainda em contexto átono, obtivemos erros na conversão de um pequeno conjunto de palavras não acentuadas, nomeadamente *e*, *mas*, *para*, *por* e *porque*, que,

por não possuírem qualquer marcador silábico, não foram captadas pelas expressões regulares que tratam dos contextos pré- e postónicos das vogais no interior de pala –, daí procedermos ao ajuste manual das suas transcrições, conforme a Tabela 67 evidencia.

Tabela 67 – Correção manual de palavras não acentuadas

Antes	Depois
[i]	[i]
[ma]	[me]
[pare]	[pere]
[por]	[pur]
[porki]	[purki]

Importa ressaltar, contudo, que as restantes palavras não acentuadas (*e.g.*, o, um, ao, como, ou, que, se, sem, a, com, de, em) não registaram qualquer erro de transcrição por corresponderem a ditongos nasais, orais ou pelo facto de a única vogal que possuem ser final e, portanto, ser captada pelas respetivas *regex*. Atentemos que, ao contrário do que ocorre com os advérbios em *-mente* e os nomes/adjetivos com sufixo *-z* avaliativo, não podemos considerar que estas ocorrências se tratam de uma limitação à aplicação do sistema a uma futura amostra de maior escala, dado que se tratam de um conjunto limitado de palavras da língua e que, uma vez captadas, não o voltarão a ser. Podemos, todavia, afirmar que se trata de uma restrição do sistema de conversão em si, por não ser capaz de captar estes casos.

5.3.3. Vogais tónicas

5.3.3.1. Dissimilação das palatais

Outro dos contextos que mereceu especial atenção foi a variação que o grafema <e> se encontra sujeito antes de consoante palatal /ɲ/, /ʃ/, /ʎ/ ou /ʒ/ e num grupo restrito de palavras, do qual fazem parte as palavras *vermelho*, *coelho* e *texto*. Tal como a variação ocorrente em fronteira inicial de palavra, também este fenómeno se trata de um processo instável no português, dependente de fatores lexicais, individuais e

extralinguísticos. Posto isto, procedemos à contemplação das diferentes realizações ([e], [ej] e [e]), como ilustra a Tabela 68, que tendem a ocorrer particularmente antes da palatal /ʎ/ e de [ej] antes da palatal /j/, de modo a aproximarmo-nos o mais possível da variedade-padrão.

Tabela 68 – Correção manual das vogais seguidas de consoante palatal

Antes	Depois
[vɨr'm/E/ʎu]	[vɨr'me(j)ʎu] / [vɨr'meʎu]
[ku'E/ʎu]	[ku'e(j)ʎu] / [ku'eʎu]
['t/E/ftu]	['te(j)ftu]

5.3.3.2. Exceções

Finalmente, procedemos ao ajuste manual de certos casos pontuais que constituem verdadeiras exceções às regras fonológicas da língua portuguesa ou que não puderam ser captadas pelo sistema de conversão (Tabela 69). Pertence ao primeiro grupo a forma resultante do truncamento de um composto *euro*, cuja vogal final não obedece ao paradigma de redução vocálica do português, realizando-se [ɔ]. Por sua vez, pertence ao segundo grupo a palavra *comboio*, cujo ditongo, como vimos anteriormente, se realiza [ɔj] e não [oj], contrariando as restantes ocorrências da sequência <oi> do *corpus*, representando um obstáculo do presente sistema de conversão e uma barreira ao potencial alargamento do *corpus*.

Tabela 69 – Correção manual de exceções

Antes	Depois
['ewru]	['ewɔ]
[kõ'boju]	[kõ'boju]

5.4. Integração no DLP-ACL (2025)

Tal como verificado na análise dos resultados, a incorporação dos dados gerados pelo sistema de conversão no DLP-ACL implicou um conjunto de operações

complementares, de natureza predominantemente manual, com o objetivo de assegurar a coerência e completude da informação fonológica e silábica associada a cada lema.

No caso específico da translineação, detetaram-se falhas na leitura de certos lemas devido à presença de espaços em branco no início ou no final da cadeia gráfica (por exemplo, *cenoura* ou *bebé*), os quais inviabilizavam a correta correspondência entre os dados do ficheiro ODS e a base de dados do dicionário. Este tipo de erro, que escapava ao sistema automático de correspondência por *regex*, foi corrigido por meio de uma limpeza manual das ocorrências afetadas, seguida de nova exportação e importação dos ficheiros corrigidos.

Paralelamente, observou-se que os lemas que apresentam formas variantes — nomeadamente formas femininas, variantes ortográficas anteriores ao AO90, ou grafias alternativas válidas — não eram contemplados automaticamente pelo sistema, que apenas aplicava a translineação e a transcrição fonética à forma principal. Para colmatar essa limitação, optou-se por uma estratégia diferenciada de integração: as formas femininas foram inseridas no mesmo componente <syll> da forma masculina, separadas por vírgula, enquanto as variantes ortográficas foram inseridas num componente distinto, delimitado por um espaço e identificado na interface com a conjunção “ou”. Este procedimento teve como objetivo não apenas garantir a exaustividade dos dados fonológicos, mas também preservar a inteligibilidade e a ergonomia da apresentação digital, como se observa na Tabela 70.

Tabela 70 – Representação das formas femininas e das variantes ortográficas

Lema	Translineação	Transcrição fonética
alto, alta	al.to, al.ta	[ˈaɫtu], [ˈaɫte]
cenoura, cenoira	ce.nou.ra ou ce.noi.ra	[sɨˈnɔre] ou [sɨˈnoɣre]

Com esta organização, cada um dos 730 lemas do *corpus* passou a disponibilizar, além da forma gráfica, a sua correspondente segmentação silábica e transcrição fonética, acessíveis ao utilizador por meio de *pop-ups* interativas, criadas para facilitar a descodificação da informação linguística especializada. Estas janelas, integradas na

interface digital do DLP-ACL, oferecem uma visualização intuitiva e pedagógica da estrutura fonológica das palavras, conforme exemplificado nas Figuras 14 e 15.

Figura 14 – Pop-up da translineação na entrada *final* no DLP-ACL (2025)

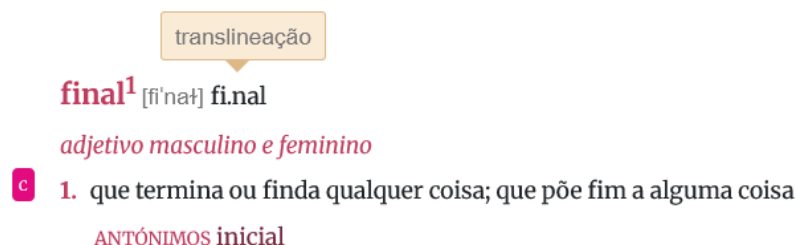
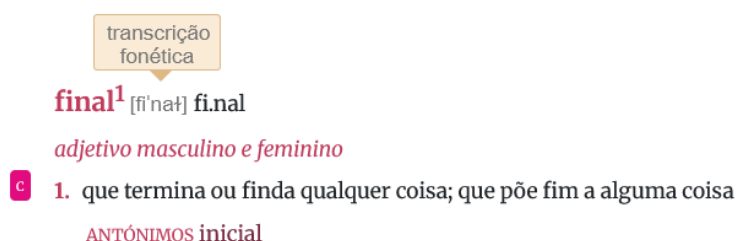


Figura 15 – Pop-up da transcrição fonética na entrada *final* no DLP-ACL (2025)



Uma das principais dificuldades enfrentadas no desenvolvimento deste sistema de conversão prendeu-se com a imprevisibilidade da ocorrência das vogais tónicas médias /e, ε/ e /o, ɔ/. Como referido previamente, a solução encontrada para a contemplação deste contexto passou pela utilização dos arquifonemas /E/ e /O/, que representam respetivamente os pares enquanto variantes mutuamente exclusivas em contexto tónico. Contudo, para que esta informação pudesse ser decodificada pelo utilizador, procedemos à sua decodificação manual, incluindo duas *pop-ups* como as ilustradas nas Figuras 15, 16, 17 e 18.

Figura 16 – Entrada *ferver* no DLP-ACL (2025)

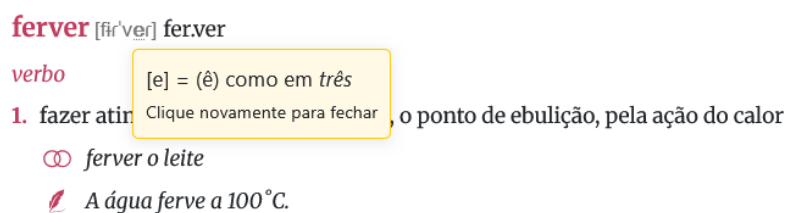


Figura 17 – Entrada *mulher* no DLP-ACL (2025)

mulher¹ [mu'ʎɛr] mu.lher
nome feminino [ɛ] = (é) como em *pé*
 1. ser humano do sexo feminino
 2. pessoa do sexo feminino em idade adulta

Figura 18 – Entrada *forte* no DLP-ACL (2025)

forte¹ ['fɔrti] for.te
adjetivo masculino [ɔ] = (ó) como em *pó*
 1. que tem muita força
 SINÓNIMOS **enérgico; potente; robusto; vigoroso**
 ANTÓNIMOS **fraco; débil**
 homem forte; mulher forte

Figura 19 – Entrada *amor* no DLP-ACL (2025)

amor [e'mor] a.mor
nome masculino [o] = (ó) como em *pôr*
 1. predisposição para o amor
 2. afeição profunda de uma pessoa por outra, de carácter passional e que, geralmente, implica atração sexual
 SINÓNIMOS **paixão**
 ANTÓNIMOS **ódio**
 casamento de amor; declaração de amor; desgosto de amor

Este procedimento revelou-se particularmente importante no caso de lemas homógrafos não homófonos, em que a variação fonológica tem implicações semânticas diretas. Nestes casos, a transcrição fonética foi cuidadosamente desambiguada com base nas diferentes aceções associadas a cada entrada. A Tabela 71 apresenta exemplos de como essa distinção foi incorporada no sistema, garantindo a fidelidade linguística e a clareza informativa.

Tabela 71 – Desambiguação das palavras homógrafas não homófonas

Entrada lexical (DLP-ACL, 2025)	Aceção	Transcrição
colher ¹	(n. f.) utensílio de mesa ou cozinha [...]	[ku'ʎ/E/r] -> [ku'ʎɛr]

colher ²	(v.) separar da haste, do caule [...]	[ku'Λ/E/r] -> [ku'Λer]
este ¹	(adj. m. e fem.) relativo ao lado onde nasce o sol	['/E/fti] -> ['εfti]
este ²	(n. m.) ponto cardeal situado a nascente	['/E/fti] -> ['εfti]
este, esta ³	(pronome)	['/E/fti] -> ['εfti], ['εfte]
forma ¹	(n. f.) conjunto de contornos [...]	['f/O/rme] -> ['forme]
forma ²	(n. f.) molde [...]	['f/O/rme] -> ['forme]

Para assegurar a transparência metodológica e a reprodutibilidade do sistema, foram elaborados os Anexos 7 a 13, que documentam exaustivamente os critérios adotados para a translineação e transcrição fonética, bem como a codificação dos caracteres utilizados. Estes documentos, atualmente em fase de integração no Guia de Uso da plataforma digital do DLP-ACL, destinam-se tanto a utilizadores especializados como a desenvolvedores de recursos linguísticos, oferecendo uma base para a compreensão do sistema e possibilitando, eventualmente, a sua adaptação a outros níveis de entrada lexical ou a diferentes variedades do português.

A integração destes dados representa um avanço significativo no processo de modernização fonológica do DLP-ACL, contribuindo simultaneamente para a melhoria da sua acessibilidade, para o reforço da sua dimensão científica e para a disponibilização de ferramentas relevantes para a investigação linguística, o ensino do português e o desenvolvimento de novos recursos computacionais baseados em dados lexicográficos estruturados.

Finalizada a análise e discussão dos resultados, sintetizámos agora os principais contributos, possibilidades de aprofundamento futuro.

6. Conclusões

A presente investigação teve no seu cerne o desenvolvimento e a aplicação de um sistema de translineação e de conversão grafema-fone[ma], baseado em expressões regulares para os 730 lemas pertencentes ao Nível 1 do *Vocabulário Fundamental* do

DLP-ACL (2025). Esta secção retoma os principais objetivos estabelecidos e as etapas envolvidas na construção deste sistema, apresentando em seguida as principais conclusões, vantagens, limitações e possíveis direções para uma futura expansão do mesmo.

Esta dissertação procurou rever e reestruturar os critérios fonético-fonológicos propostos no DLPC (2001), aplicando-os ao desenvolvimento do sistema de conversão. Foram ainda analisados os erros associados à atribuição automática de caracteres, bem como as potencialidades e limitações do uso de expressões regulares na representação do PE em corpora lexicográficos. Por fim, refletiu-se sobre a integração deste sistema no DLP-ACL, tanto no que respeita à exploração fonológica como à sua futura expansão.

Num primeiro momento, e de forma a definir critérios teóricos sólidos para a construção do sistema, recorreremos a uma revisão do inventário fonológico, dos padrões acentuais e da estrutura silábica do PE, bem como ao alinhamento entre as suas representações gráficas e as suas representações fonémicas, uma vez que este sistema tem por base a ortografia. Posteriormente, procedeu-se a uma análise do tratamento da pronúncia ao longo da história no panorama lexicográfico português, a partir da qual foi possível concluir que a inclusão de informação fonético-fonológica nos dicionários monolingues do português foi bastante esporádica e que, paradoxalmente, a digitalização destes recursos não se traduziu em avanços significativos, revelando uma subvalorização prolongada e uma falta de atenção especializada deste registo. De igual modo, também os sistemas de conversão grafema-fone[ma] de acesso aberto se revelaram limitados uma vez que, de momento, apenas se encontram disponíveis para o PE o *Grafone* e o *Convert Text to IPA Transcription*.

O desenvolvimento concreto deste sistema iniciou-se pela composição de critérios para a translineação e a revisão e reestruturação dos critérios de marcação da sílaba tónica e de transcrição fonética do DLPC (2001), que, entre outros aspetos, apresentava um inventário de caracteres incompleto e, por vezes, inconsistente com aquele estabelecido pelo AFI. Em seguida, procedeu-se à formulação, aplicação e ajustamento de expressões regulares para a translineação, a marcação da sílaba tónica e a transcrição fonética do *corpus* selecionado. A opção de utilizarmos expressões

regulares justificou-se pela sua elevada versatilidade e aplicabilidade, bem como pela sua capacidade de identificar padrões com base em regras previamente definidas.

Os resultados obtidos apontam para um tratamento automático mais eficaz na translineação e na marcação da sílaba tónica do que na transcrição fonética. No caso da translineação, todos os 730 lemas foram processados corretamente, não se revelando necessária a intervenção manual nas conversões. Os resultados da marcação da sílaba tónica revelaram-se fortemente positivos, o que se traduziu na correta marcação acentual de todas as palavras do *corpus*, salvo três casos excepcionais (*e.g.*, apenas, menos, simples), as quais escapam ao paradigma geral acentual não-verbal do português por se tratar de palavras paroxítonas terminadas em consoante. Em contraste, os resultados obtidos na transcrição fonética, particularmente nos segmentos vocálicos, revelaram-se menos satisfatórios e necessitaram de alguns ajustes manuais ou de expressões *ad hoc*.

Este desempenho das expressões regulares relaciona-se com os diferentes níveis de estabilidade de cada patamar de representação (sílaba, acento, fonética), pelo que o carácter relativamente sistemático e constante da sílaba e do acento do PE se traduz naturalmente num tratamento computacional mais eficiente. Esta conclusão comprova-se, adicionalmente, ao compararmos os resultados alcançados na transcrição fonética das consoantes e na transcrição fonética das vogais, pelo que este último grupo, bastante mais instável e variável, suscitou maiores desafios na sua automatização.

Os resultados permitiram concluir que os principais obstáculos ao sistema desenvolvido relacionam-se com casos excepcionais da língua como: (i) palavras que não seguem o padrão acentual geral do português (*e.g.*, apenas, menos, simples); (ii) vogais pretónicas que não sofrem redução vocálica (*e.g.*, bebé, esquecer, normal); (iii) /a/ átonos em fronteira inicial de palavra que não sofrem redução vocálica (*e.g.*, ação, amanhã, ator); e (iv) vogais finais das formas resultantes do truncamento de compostos que não sofrem redução vocálica (*e.g.*, euro). Estes casos excepcionais constituem, *per se*, bases de estudo importantes para a fonologia do português. Estas barreiras correspondem a casos cuja descodificação depende exclusivamente do léxico mental do

falante nativo e que, por essa razão, permanecem opacos a qualquer processo de automatização linguística.

Adicionalmente, contextos fonologicamente opacos como: (i) as vogais tónicas médias; (ii) as vogais pretónicas dos advérbios em *-mente* e dos nomes e adjetivos com sufixo *-z* avaliativo; (iv) o grafema <x>; e (v) a sequência de grafemas <oi> dificultaram a aplicação do sistema de conversão, dado que não é possível fazer a sua desambiguação a partir da ortografia. Tais dificuldades não comprometem a qualidade geral do sistema, pelo contrário, o reconhecimento destas limitações reforça o potencial de melhoria do sistema e abre caminho para um futuro aperfeiçoamento que nos permita lidar de forma mais eficiente com essas particularidades da língua.

O presente estudo propõe, deste modo, o primeiro sistema de translineação e conversão grafema-fone[ma] automático para o PE concebido e integrado num recurso lexicográfico. De uma forma geral, os resultados obtidos na translineação, na marcação da sílaba tónica e na transcrição fonética foram bastante satisfatórios e, apesar de necessitar da intervenção manual em certos casos pontuais, revelou-se eficaz na automatização dos 730 lemas. Para além disto, trata-se de uma ferramenta desenvolvida de raiz, de acesso público e de fácil adaptação, características que constituem vantagens significativas para uma futura expansão aos mais de cem mil lemas do DLP-ACL (2025).

Considerando o exposto, a presente investigação representa o passo inicial para o desenvolvimento de um sistema de conversão mais robusto e abrangente. Com base nos resultados obtidos, uma futura expansão deverá passar pela ampliação do *corpus*, pela inclusão de sequências que escaparam a esta fase inicial (*e.g.*, encontros consonânticos impróprios), pela automatização de plurais e formas femininas que impliquem alterações fonéticas e do afinamento progressivo das regras implementadas. A consolidação do sistema poderá igualmente beneficiar de estudos mais aprofundados sobre a frequência contextual de determinados grafemas, pela sistematização das exceções da língua portuguesa e pela integração de alternâncias morfofonológicas que afetem diretamente a transcrição fonética.

Consideramos que os resultados deste trabalho contribuem para a valorização da componente fonético-fonológica nos dicionários através da disponibilização de uma ferramenta de acesso livre e código aberto e com repercussões diretas no ambiente digital lexicográfico português. Para além do seu potencial no auxílio da realização de estudos fonológicos, este sistema constitui um contributo relevante para o desenvolvimento de recursos computacionais aplicados à língua portuguesa, abrindo novas possibilidades para a fonologia e a lexicografia computacionais.

Referências bibliográficas

Dicionários digitais

AULETE. 2025. *Aulete Digital*. Lexikon Editora Digital. Disponível em <https://aulete.com.br/>

DICIO. 2025. *Dicionário Online de Português*. Porto: 7Gaus. Disponível em <https://www.dicio.com.br/>

DLP-ABL. 2025. Academia Brasileira de Letras. *Dicionário da Língua Portuguesa*. Disponível em <https://servbib.academia.org.br/dlp/>

DLP-ACL. 2025. Academia das Ciências de Lisboa. *Dicionário da Língua Portuguesa*. Salgado, A. (Coord.). Lisboa: Academia das Ciências de Lisboa. Disponível em <https://dicionario.acad-ciencias.pt/>

DP. 2025. *Dicionário de Português*. Disponível em <https://www.dicionario.info/>

ESTRAVIZ. 2025. *Dicionário Estraviz*. Através Editora. Disponível em <https://www.meudicionario.org/>

HOUAISS. 2025. *Dicionário Houaiss*. UOL. Disponível em https://houaiss.online/houaission/apps/uol_www/v7-0/html/index.php#0

PORTO EDITORA. 2025. *Dicionário Infopédia da Língua Portuguesa*. Porto Editora. Disponível em <https://www.infopedia.pt/>

INFORMAL. 2025. *Dicionário inFormal*. Disponível em <https://www.dicionarioinformal.com.br/>

LÉXICO. 2025. *Dicionário de Português Online*. Porto: 7Gaus. Disponível em <https://www.lexico.pt/>

MD. 2025. *Meu Dicionário*. Disponível em <https://www.meudicionario.org/>

MICHAELIS. 2025. *Michaelis On-line*. Editora Melhoramentos Ltda. Disponível em <https://michaelis.uol.com.br/>

PRIBERAM. 2025. *Dicionário Priberam da Língua Portuguesa*. Retirado de <https://dicionario.priberam.org/>

Dicionários impressos

ACL. 1793. *Diccionario da Lingoa Portuguesa*, Tomo primeiro A.

_____. 1976. *Dicionário da Língua Portuguesa*: [vol. 1: A-Azuverte]. Lisboa, Imprensa Nacional.

AULETE, F. J. C. 1881. *Diccionario Contemporaneo da Lingua Portuguesa feito sobre um plano inteiramente novo*, 2 vol., Lisboa, Imprensa Nacional.

_____. 1987. *Dicionário Contemporâneo da Língua Portuguesa Caldas Aulete: A – inquinamento, Volume 1*, por H. de Garcia (ed.) & A. Nascentes (colab.), Ed. Delta.

AURÉLIO, B. F. H. 1986. *Novo dicionário Aurélio da língua portuguesa*. 2. ed. rev. e aument. Rio de Janeiro, Nova Fronteira.

_____. 1999. *Novo Aurélio, Dicionário de Língua Portuguesa Séc. XXI*. Rio de Janeiro, Editora Nova Fronteira.

BLUTEAU, R. 1712. *Vocabulario portuguez e latino, aulico, anatomico, architectonico, bellico, botanico, brasilico, comico, critico, chimico, dogmatico, dialectico, dendrologico, ecclesiastico, etymologico, economico, florifero, forense, fructifero... autorizado com exemplos dos melhores escritores portugueses, e latinos... / pelo Padre D. Raphael Bluteau*. Coimbra: no Collegio das Artes da Companhia de Jesus.

DLPC. 2001. ACADEMIA DAS CIÊNCIAS DE LISBOA. *Dicionário da Língua Portuguesa Contemporânea*, Casteleiro, J. M. (Coord.). Lisboa: Academia das Ciências de Lisboa & Editorial Verbo.

BACELAR, B. L. M. 1783. *Diccionario da Lingua Portuguesa*, Lisboa, Aquino Bulhoens.

EDITORIAL VERBO. 2006 *Dicionário Verbo: Língua Portuguesa*. Lisboa, Editorial Verbo.

FARIA, E. 1849. *Novo Diccionario da Lingua Portuguesa, seguido de um Diccionario de Sinônimos*, Lisboa.

FIGUEIREDO, A. C. 1899. *Nôvo Diccionário da Língua Portuguêsa*, 2 vol., Lisboa, Tavares Cardoso & Irmão.

_____. 1996. *Grande Dicionário de Língua Portuguesa*, por R. Guedes, T. Lino & R. Costa (eds.), Bertrand Editora.

HOUAISS, A. 2002-2003. *Dicionário Houaiss da Língua Portuguesa*. Lisboa, Círculo de Leitores.

MACHADO, J. P. (coord.). 1981. *Grande Dicionário da Língua Portuguesa*, 12 vols, Lisboa, Sociedade de Língua Portuguesa / Amigos do Livro Editores.

MORAIS SILVA, A. 1789. *Diccionario da Lingua Portuguesa, composto pelo Padre D. Rafael Bluteau, reformado e acrescentado*, Lisboa, Ferreira (1813: repr. facsimilada Rio de Janeiro, Fluminense).

_____. 1813. *Diccionario da lingua portugueza recopilado dos vocabularios impressos até agora, e nesta segunda edição novamente emendado, e muito acrescentado...* Lisboa: Na Typographia Lacerdina.

_____. 1949-1959. *Grande Dicionário da Língua Portuguesa*, por A. Moreno, J. F. Cardoso Júnior e J. P. Machado. Lisboa, Tavares Confluência. [10.^a edição revista, corrigida, muito aumentada e actualizada; 12 vols.]

PORTO EDITORA. 2004. *Grande Dicionário da Língua Portuguesa*, Teixeira, G. (coord.) Porto, Porto Editora.

TEXTO EDITORES. 2007. *Novo Dicionário da Língua Portuguesa conforme Acordo Ortográfico*, Espinha, M. P. (coord.), Lisboa.

VIEIRA, Fr. D. 1871-1874. *Grande Dicionario Portuguez ou Theouro da Lingua Portuguesa*, 5 vol., Porto, Ernesto Chardron e Bartholomeu H. de Moraes.

Vocabulários impressos

ABL. 1943. *Vocabulário Ortográfico da Língua Portuguesa*.

ACL. 1940. *Vocabulário Ortográfico da Língua Portuguesa*. Lisboa, Imprensa Nacional.

ACL. 1947. *Vocabulário Ortográfico Resumido da Língua Portuguesa*. Lisboa, Imprensa Nacional.

ACL. 1970. *Vocabulário Ortográfico Resumido da Língua Portuguesa*. Lisboa, Imprensa Nacional.

GONÇALVES VIANA, A. R. 1912. *Vocabulário Ortográfico e Remissivo da Língua Portuguesa*. Paris, Aillaud, Lisboa, Livraria Bertrand.

REBELO GONÇALVES, F. L. 1966. *Vocabulário da Língua Portuguesa*. Coimbra, Coimbra Editora.

Literatura

ABAURRE, M. B. M. 1983. *Nasality in Portuguese: a critical consideration of a proposed analysis for nasal diphthongs*.

_____. 2006. Fonologia e Fonética. Em E. Guimarães & M. Zoppi-Fontana (orgs), *Introdução às Ciências da Linguagem: A Palavra e a Frase*, pp. 39-69. Campinas, SP: Pontes Editores.

ABERCROMBIE, D. 1964. *English Phonetics Texts*. London: Faber and Faber.

ANDRADE, A. 1998. Variação fonética do /l/ em ataque silábico em Português Europeu. *Atas do IX Encontro Nacional da Associação Portuguesa de Linguística*, pp. 55–76.

_____. 1999. On /l/ velarization in European Portuguese. Em J. Ohala; Y. Hasegawa; M. Ohala; D. Granville; A. Bailey (eds.). *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 543–546.

_____. 2020. Vocalismo. Em Raposo, E., M.F. B. Nascimento, M. A. Mota, L. Segura, A. Mendes & A. Andrade (orgs). *Gramática do Português*. Vol. III. Lisboa: Fundação Calouste Gulbenkian.

ANDRADE, A; D'ANDRADE, E. 2020. Sílabas. Em Raposo, E., M.F. B. Nascimento, M. A. Mota, L. Segura, A. Mendes & A. Andrade (orgs). *Gramática do Português*. Vol. III. Lisboa: Fundação Calouste Gulbenkian.

BARBOSA, J. M. 1983. *Études de phonologie portugaise*. (2ª. ed.). Universidade de Évora.

BARROS, M. J. & WEISS, C. 2006. Maximum entropy motivated grapheme-to-phoneme, stress and syllable boundary prediction for Portuguese text-to-speech. Em *IV Jornadas en Tecnologia del Habl*, 177-182.

BARROSO, H. 1999. *Forma e Substância da Expressão da Língua Portuguesa*. Lisboa: Edições Almedina, S.A. <https://hdl.handle.net/1822/24827>

BISOL, L. 1992. O acento e o pé métrico binário. Em *Cadernos de Estudos Linguísticos*, Campinas, SP, 22.

_____. 1989. O ditongo na perspectiva da fonologia atual. Em *D.E.L.T.A*, 5(2), pp. 185-224.

_____. 1999. A Sílabas e os seus Constituintes. In Neves, M. H. M.(Org.). Em *Gramática do Português Falado*. São Paulo: Humanitas.

_____. 2013. *O acento: duas alternativas de análise*. Organon: Porto Alegre.

- BISOL, L. & VELOSO, J. 2016. Phonological Processes Affecting Vowels: Neutralization, Harmony and Nasalization. Em Wetzels, W. L.; Sergio Menuzzi; João Costa (Eds.). *The Handbook of Portuguese Linguistics*. Malden Ma/Oxford. Wiley/Blackwell: 69.84.
- BLEVINS, J. 1995. The Syllable in Phonological Theory. Em Goldsmith, J. A. (ed.), *The Handbook of Phonology*. Cambridge: Blackwell.
- BONET, E. & MASCARÓ, J. 1997. On the representation of contrasting rhotics. Em F. Martínéz-Gil & Morales-Front (eds.), *Issues in the Phonology and Morphology of the Major Iberian Languages*. Washington: Georgetown University Press.
- BRAGA, D., COELHO, L., RESENDE Jr, F. G. 2006. A Rule-Based Grapheme-to-Phone Converter for TTS Systems in European Portuguese. Em *VI International Telecommunications Symposium*, 328-333.
- CÂMARA, J. M. C. 1953. *Para o estatuto da fonêmica portuguesa*. Rio de Janeiro: Organização Simões.
- _____. 1977. *Para o estudo da fonêmica portuguesa*. (2.^a ed. do original de 1953). Rio de Janeiro: Livraria Editora.
- CANDEIAS, S. & PERDIGÃO, F. 2011. Integração linguística em sistemas de conversão de grafema para fone[ma]. Em A. R. Luís (Ed.), *Estudos de linguística*, 1, 181-194. doi: [10.14195/978-989-26-0231-8_12](https://doi.org/10.14195/978-989-26-0231-8_12)
- _____. 2013. Grafone: uma ferramenta de/com recurso à informação linguística. Em *Textos Seleccionados, XXVIII Encontro Nacional da Associação Portuguesa de Linguística*, 189-203. doi: <http://doi.org/10.1109/its.2006.4433293>
- CASEIRO, D. A., TRANCOSO, I., OLIVEIRA, L. & VIANA, C. 2002. Grapheme-to-phone using finite-state transducers. Em *Proceedings of the IEEE Workshop on Speech Synthesis*, 215-218. doi: [10.1109/WSS.2002.1224412](https://doi.org/10.1109/WSS.2002.1224412)
- CLARK, J. & YALLOP, C. 1990. *An Introduction to Phonetics and Phonology*. Blackwell.
- COSTA CARVALHO, C. I. 2017. Conversor de transcrição fonética automática para as formas linguísticas da variedade linguística potiguar. Em *Domínios de Linguagem*, 734-752. doi: <https://doi.org/10.14393/DL30-v11n3a2017-13>
- COULMAS, F. 2003. *Writing Systems*. Cambridge: Cambridge University Press.

- CRUZ-FERREIRA, M. 1999. Illustrations of the IPA. Portuguese (European). Em *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*, pp. 126-130. Cambridge: Cambridge University Press.
- D'ANDRADE, E. & BERNARD, L. 1992. Na crista da onda: o acento da palavra em português. Em *Actas do VII Encontro Nacional da APL*, Lisboa, 1992: 15-26
- D'ANDRADE, E. & VIANA, M. C. 1993. Sinérese, diérese e estrutura silábica. Em *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, 31-42.
- D'ANDRADE, E. & RODRIGUES, C. 1999. Das Escolas e das Culturas: História de Uma Sequência Consonântica. Em A. M. Lopes & C. Martins (org.), *Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística*. Braga: Edições Colibri, pp. 117-133.
- D'ANDRADE, E.; KHIM, A. 1988. Fonologia auto-segmental e nasais em português. Em *Actas do III Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa. APL: 3-15.
- DANIELS, P. T. 2010. Writing Systems. Em M. Aronoff & J. Reesmiller (eds.), *The Handbook of Linguistics*, pp. 43-80. Oxford: Blackwell.
- DAVENPORT, M. & HANNAHS, S. J. 1998. *Introducing Phonetics and Phonology*. (2.^a ed.). New York: Hodder Arnold.
- DELGADO-MARTINS, M. R. 1996. Relação fonética/fonologia: A propósito do sistema vocálico do português. Em I. Duarte & I. Leiria (eds.), *Congresso Internacional sobre o Português*, Volume I, Lisboa, Edições Colibri/APL, pp. 311-325.
- EMILIANO, A. 2009. *Fonética do Português Europeu*. Guimarães Editores.
- FIKKERT, P. & FREITAS, M. J. 2004. The role of language-specific phonotactics in the acquisition of onset clusters. Em *Linguistics in the Netherlands*, Vol. 21, pp 58–68.
- FREITAS, M. J. 1997. *Aquisição da estrutura silábica em português europeu*. Dissertação de Doutoramento. Universidade de Lisboa.
- _____. 1998. O grupo consonântico s+C em início de palavra na aquisição do português europeu. Em *Actas do XV Encontro Nacional da APL*, pp. 499-512.
- _____. 2001. Os ping[w]ins são diferentes dos c[w]elhos? Questões sobre oclusivas velares, semivogais e arredondamentos na aquisição do português europeu. Em *Actas*

- do XVI Encontro Nacional da Associação Portuguesa de Linguística. Lisboa: APL/Colibri, 213-225.
- FREITAS, M. J. & SANTOS, A. L. 2001. *Contar (Histórias de) Sílabas, Descrição e Implicações para o Ensino de Português como Língua Materna*. Lisboa: Colibri / Associação de Professores de Português.
- FREITAS, M. J. & RODRIGUES, C. 2003. On the nature of sC- clusters in European Portuguese. Em *Journal of Portuguese Linguistics*, Vol. 2, pp. 55-85.
- GLEASON, H. A. JR. 1985. *Introdução à Linguística Descritiva*. Pinguelo, J. (trad.). Lisboa: Fundação Calouste Gulbenkian.
- GOAD, H. 2012. sC Clusters are (almost always) coda-initial. Em *The Linguistic Review*, 29(3), pp. 335-373. doi: doi.org/10.1515/tlr-2012-0013
- GONÇALVES, M. 2008. *Fonética e Fonologia do Português*. Braga: Alêtheia.
- GOYVAERTS, J. 2007. *Regular Expressions: The Complete Tutorial*. LuluPress, Incorporated.
- GOYVAERTS, J. & LEVITHAN, S. 2009. *Regular Expressions Cookbook*. O'Reilly Media, Inc.
- GUSSENHOVEN, C. & JACOBS, H. 1998. *Understanding Phonology*. London: Harnold.
- HAYES, B. 1981. *A Metrical Theory of Stress Rules*. Dissertação de doutoramento. Indiana University of Linguistics. Club, Bloomington, Indiana.
- HENRIQUES, I. 2012a. The Production and the Syllabic Nature of Word Initial sC- Clusters in European Portuguese Speakers. Em *Corela*, doi: <https://doi.org/10.4000/corela.2660>
- HENRIQUES, I. 2012b. *A Fricativa Coronal /S/ em /#(Ø)SC/ em Português Europeu*. Dissertação de doutoramento. Faculdade de Letras da Universidade do Porto.
- HESELWOOD, B. 2013. *Phonetic Transcription in Theory and Practice*. Edinburgh: Edinburgh University Press.
- IPA – INTERNATIONAL PHONETIC ASSOCIATION. (n.d.). *International Phonetic Alphabet*. Retirado de <https://www.internationalphoneticassociation.org/>.
- _____. 1999. *Handbook of the IPA*. Cambridge: Cambridge University Press.
- IRIARTE SANRÓMAN, A. 2016. *Dicionários Portugueses*. [Publicação Pedagógica]. Universidade do Minho. <https://hdl.handle.net/1822/78348>

- JAKOBSON, R. 1967. *Fonema e Fonologia*. Câmara, M. J. (trad.). Rio de Janeiro: Livraria Acadêmica.
- JESUS, L. M. T., & SHADLE, C. H. 2005. Acoustic Analysis of European Portuguese Uvular [χ, ʁ] and Voiceless Tapped Alveolar [ɾ] Fricatives. Em *Journal of the International Phonetic Association*, 35(1), pp. 27–44.
- LASS, R. 1984. *Phonology. An Introduction to basic concepts*. Cambridge University Press.
- LAVER, J. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.
- LEITE, Y. F. 1974. *Portuguese stress and related rules*. Tese de Doutorado, University of Texas.
- MALMBERG, B. 1954. *A Fonética*. Tradução de Oliveira Figueiredo. Lisboa: Edição «Livros do Brasil».
- MARTINET, A. 1960. *Éléments de Linguistique Générale*. Paris: Armand Colin.
- MATEUS, M. H. M.; ANDRADE, A.; VILLALVA, A.; 1990. *Fonética, Fonologia e Morfologia do Português*. Lisboa: Universidade Aberta.
- MATEUS, M. H. M. 1982a. *Aspectos da Fonologia do Português*. Lisboa: INIC.
- _____. 1982b. O acento da palavra em português: uma nova proposta. Em *Boletim de Filologia*, 28, 211-229.
- _____. 1994. A silabificação de base em português. Em *Actas do X Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, pp. 289-300.
- _____. 2006. Sobre a Natureza Fonológica da Ortografia Portuguesa. Em *Estudos da Língua(gem)*, 3, 159-180.
- MATEUS, M. H. M.; D'ANDRADE, E. 2000. *The Phonology of Portuguese*. Oxford: University Press.
- MATEUS, M. H.; BRITO, A. M.; DUARTE, I.; FARIA, I.; FROTA, S.; MATOS, G.; OLIVEIRA, F.; VIGÁRIO, M.; VILLALVA, A. 2003. *Gramática da Língua Portuguesa* (5.ª edição). Lisboa: Caminho
- MATEUS, M. H., FALÉ, I. & FREITAS, M. J. 2005. *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta.
- MCKINNEY, W. 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, Vol. 445, pp. 51-56.

- MIRANDA, A. R. 1996. *A aquisição do 'r': uma contribuição à discussão sobre seu status fonológico*. Dissertação de Mestrado. Porto Alegre: Pontifícia Universidade Católica.
- MONARETTO, V. N. O. 1997. *Um reestudo da vibrante: análise variacionista e fonológica*. Tese de Doutorado. Porto Alegre: UFRGS.
- MONTEIRO, R., AMARO, R., CORREIA, S., PINTARD, A., GAUCHOLA, R., MOUTINHO, M., & BLANCO ESCODA, X. 2023. *iRead4Skills – Complexity Levels*. Zenodo. <https://doi.org/10.5281/zenodo.10459090>
- ODDEN, D. 2005. *Introducing Phonology*. Cambridge University Press.
- OLIVEIRA, L. C., VIANA, M. & TRANCOSO, I. 1992. A rule-based text-to-speech system for portuguese. Em *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2, 72-76. Imprensa da Universidade de Coimbra. doi: [10.1109/ICASSP.1992.226117](https://doi.org/10.1109/ICASSP.1992.226117)
- _____. 2001. Progress report of project dixi+: A portuguese text-to-speech synthesizer for alternative and augmentative communication. Em *Relatório técnico, FCT*.
- PARKINSON, S. R. 1983. Portuguese nasal vowels as phonological diphthongs. Em *Lingua*, Vol. 61, N.º 2-3, pp. 157-177.
- PEREIRA, M. I. P. 1999. *O acento de palavra em português: uma análise métrica*. Dissertação de Doutorado. Faculdade de Letras da Universidade de Coimbra.
- _____. 2020. O acento. Em Raposo, E., M.F. B. Nascimento, M. A. Mota, L. Segura, A. Mendes & A. Andrade (orgs). *Gramática do Português*. Vol. III. Lisboa: Fundação Calouste Gulbenkian.
- PEREIRA, R. 2020. *O R-forte em Português Europeu: análise fonológica de dados dialetais*. Tese de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- PIKE, K. & PIKE, E. 1947. 'Immediate constituents of Mazatec syllables. Em *International Journal of American Linguistics*, 13, pp. 8-91.
- REIS, M. L. 2025. *Vocabulário Fundamental da Academia das Ciências de Lisboa: seleção lexical, alinhamento de sentidos e codificação*. Tese de Mestrado. Faculdade de Ciências Sociais e Humanas da Universidade NOVA de Lisboa.
- RENNICKE, I. & MARTINS, P. T. 2013. As realizações fonéticas de /R/ em português europeu: análise de um corpus dialetal e implicações no sistema fonológico. Em F. Silva,

- I. Falé & I. Pereira (eds.), *Textos Seleccionados do XXVIII Encontro Nacional da Associação Portuguesa de Linguística*, pp. 509-523. Coimbra: Associação Portuguesa de Linguística.
- RODRIGUES, C. 2003. Lisboa e Braga: Fonologia e Variação. Lisboa: Lisboa: FCG / FCT.
- _____. 2016. Variação sociolinguística. Em A. M. Martins & E. Carrilho (eds.), *Manual de linguística portuguesa*. De Gruyter.
- _____. 2020. Consonantismo. Em Raposo, E., M.F. B. Nascimento, M. A. Mota, L. Segura, A. Mendes & A. Andrade (orgs). *Gramática do Português*. Vol. III. Lisboa: Fundação Calouste Gulbenkian.
- RODRIGUES, S. 2015. *Caracterização acústica das consoantes líquidas do Português Europeu*. Tese de doutoramento. Universidade de Lisboa.
- SALGADO, A., SIMÕES, A., IRIARTE, A., VIEIRA, R., FERREIRA, M., CARMO, R., PINHEIRO, C. 2023. *Dicionário da Língua Portuguesa: a new lexicographic resource of Academia das Ciências de Lisboa*. *eLex 2023 Conference*, 27th to 29th June 2023. Brno: República Checa.
- SALGADO, A., COSTA, R., TASOVAC, T., & SIMÕES, A. (2019). [*TEI Lex-0 In Action: Improving the encoding of the Dictionary of the Academia das Ciências de Lisboa*](#). In I. Kosem et al. (Eds.), *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference* (pp. 417–433). Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.
- SAUSSURRE, F. 1916. *Cours de Linguistique Générale*. Paris: Payot.
- SELKIRK, E. 1984. On the Major Class Features and Syllable Theory. In Aronoff, M., Oehrle, R. T. (Eds.), *Language Sound Structure*. Cambridge, Mass: MIT Press, pp. 107-136.
- SILVA, C. 2020. *Um problema plural, uma proposta singular: a representação fonológica do morfema de plural no português*. Tese de Mestrado. Faculdade de Letras da Universidade do Porto.
- SIMÕES, A., ALMEIDA, J. J. & SALGADO, A. 2016. Building a Dictionary Using XML Technology. In *5th Symposium on Languages, Applications and Technologies (SLATE'16)*. Open Access Series in Informatics (OASICS), Vol. 51, pp. 14:1-14:8, Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/OASICS.SLATE.2016.14>

- SIMÕES, A., SALGADO, A. (2022). *Smart Dictionary Editing with LeXmart*. In Klosa-Kückelhaus, A., Engelberg, Stefan, Möhrs, Christine & Storjohann, Petra (eds.). *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*. Mannheim: IDS-Verlag, pp. 423-434. ISBN 978-3-937241-87-6.
- SPENCER, A. 1996. *Phonology: theory and description*. Oxford & Cambridge, Mass.: Blackwell Publishers.
- TEIXEIRA, J. 2004. *A Prosody Model to TTS Systems*. [Tese de Doutorado, Faculdade de Engenharia da Universidade do Porto]. <http://hdl.handle.net/10198/1496>
- TRANCOSO, I., VIANA, M. C., SILVA, F. M., MARQUES, G. C. & OLIVEIRA, L. C. 1994. Rule-based vs neural network-based approaches to letter-to-phone conversion for Portuguese common and proper names. Em *3rd International Conference on Spoken Language Processing (ICSLP 1994)*, 1767-1770. doi: [10.21437/ICSLP.1994-197](https://doi.org/10.21437/ICSLP.1994-197)
- TRIGO, L., & SILVA, C. 2022. Comparing lexical and usage frequencies of palatal segments in Portuguese. In *Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, Proceedings*, pages 353–362, Berlin. Springer.
- TRUBETZKOY, N. 1939. *Grundzüge der Phonologie*. (Traduzido por Christiane Baltaxe (1969) *Principles of Phonology*). Berkeley: University of California Press.
- VALADA, F. M. 2017. *Portuguese words that begin with the written form “es+consonant”: do they have a prothetic vowel?*. Dissertação de Mestrado, Faculdade de Letras e Wijsbegeerte.
- VELOSO, J. 1996. Elementos para uma reavaliação da importância da distintividade como conceito linguístico. Em *Revista da Faculdade de Letras – Línguas e Literaturas*, Porto, pp. 407-433.
- _____. 1997. As diferentes denominações das “oclusivas fricativizadas” do português: implicações linguísticas da questão. Em A. M. Brito, F. Oliveira, I. P. Lima & R. M. Martelo (eds.), *Sentido que a vida faz. Estudos para Óscar Lopes*, pp. 845-854. Porto: Campo das Letras.
- _____. 1999a. *Na Ponta da Língua*. Porto: Granito, Editores e Livreiros, Lda.

____ 1999b. The effects of two higher-level variables on the processing of an allophonic contrast in European Portuguese. Em *A psicolinguística no limiar do ano 2000*, pp. 255-231. Porto: Faculdade de Letras da Universidade do Porto.

____ 2002. Do fricative + plosive onsets exist word-initially in European Portuguese? Gathering preliminary data and suggestive evidence from a small group of children in the phonological explicit task of syllable segmentation and examining some theoretical implications for the relationship between phonology and psycholinguistics. Em K. Petrova, A. Eftimova & R. Shopov (eds), *Litora Psycholinguistica. Сборникът е посветен на юбилея на доц. д-р Пенка Илиева-Балтова*. Sófia, Sema RSH, pp. 49-56.

____ 2003a. *Da influência do conhecimento ortográfico sobre o conhecimento fonológico. Estudo longitudinal de um grupo de crianças falantes nativas do português europeu*. Dissertação de Doutoramento. Faculdade de Letras da Universidade do Porto.

____ 2003b. A distinção entre palavras terminadas em consoante e palavras terminadas na sequência ortográfica «consoante + “e”» num grupo de crianças falantes do português europeu em idade pré-escolar» in *Língua Portuguesa: Estruturas, Usos e Contrastes*. Em *Volume Comemorativo dos 25 Anos do Centro de Linguística da Universidade do Porto*, Centro de Linguística da Universidade do Porto, pp. 259-288.

____ 2005a. A língua na escrita e a escrita na língua. Algumas considerações gerais sobre a transparência e opacidade fonémicas na escrita do português e outras questões. Em *Da Investigação às Práticas. Estudos de Natureza Educacional*. Escola Superior de Educação de Lisboa, Centro Interdisciplinar de Estudos Educacionais. Vol. VI, N.º 1, pp. 49-69.

____ 2005b. Considerações sobre o estatuto fonológico de [i] em português. Em *Revista da Faculdade de Letras – Línguas e Literaturas*, II Série, Vol. XXII, Porto, pp. 621-632.

____ 2007. Schwa in European Portuguese: The Phonological Status of [i]. Em *5^{èmes} Journées d’Études Linguistiques*, Nantes, France, pp. 55-60.

____ 2015. *Introdução à Fonologia. Nível fonético e fonológico*. Porto: FLUP.

____ 2017. Monossílabos CV do português: leves e degenerados? Sonoridade vocálica e iteração de elementos na atribuição de peso e na preservação da minimalidade em

português. Em *Revista de Estudos Linguísticos da Universidade do Porto*, Vol. 12, pp. 201-226.

_____. 2019. Phonology and writing: can we look at written productions to “see the unseeable” in phonology? Em *Loquen*, Vol. 6, pp. 1-12.

_____. 2020. Conhecimento ortográfico e representações fonológicas em português. Em S. N. SALOMÃO (Coord.), *Temas da Língua Portuguesa: Do pluricentrismo à didática*, pp. 91-103. Nuova Cultura.

_____. 2022. Fonologia e Ortografia do Português Europeu. Em R. A. Alves & I. Leite (Eds.), *Ensino da Leitura e da Escrita Baseado em Evidências*, 65-81. Fundação Belmiro de Azevedo.

VERDELHO, T. 2002. Dicionários portugueses: Breve história. Em J. Horta Nunes & M. Petter (Orgs.), *História do saber lexical e constituição de um léxico brasileiro*, pp. 15-64. USP: Pontes.

VERDELHO, T. & SILVESTRE, J. P. 2007. *Dicionarística Portuguesa. Inventariação e Estudo do Património Lexicográfico*. Aveiro: Universidade de Aveiro.

VIGÁRIO, M.; FALÉ, I. 1994. A sílaba do português fundamental: uma descrição e algumas considerações de ordem teórica. Em *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística (Coimbra 1993)*. Lisboa, Colibri: 675-687

VIGÁRIO, M. 2003. *The Prosodic Word in European Portuguese*. Dissertação de Doutoramento. Universidade de Lisboa.

WELLS, J. C. 1997. SAMPA computer readable phonetic alphabet. Em D. Gibbon, R. Moore & R. Whisky (Eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter.

WETZELS, L. 1991. *A non-linear analysis of vowel alternations in the Brazilian Portuguese verb system*. Texto não publicado. Free University of Amsterdam/University of Nijmegen.

WILKENS, R., PINTARD, A., FRANÇOIS, T., BARBOSA, S., REIS, M. L., AMARO, R., RIBEIRO, E., MAMEDE, N., BAPTISTA, J., BLANCO, X., CATENA, A., GAUCHOLA, R., & MU, K. 2024. *iRead4Skills – Basic Lexicons per Complexity Level (v1.0)* [Data set]. Zenodo. <https://doi.org/10 – 81/zenodo.10889986>

Anexos

Anexo 1 – Dicionários impressos de língua portuguesa representativos dos sécs. XVIII e XIX

Dicionário	Autor(es)/ Editor(es), Ano	Info. sobre a pronúncia no lema inicial	Transcrição fonética	Info. transcrição	Sílaba tónica	Divisão silábica
<i>Diccionario da Lingua Portuguesa</i>	Bacelar, B. (1783)	Sim	Não	-	Não	Não
<i>Diccionario da Lingua Portuguesa</i>	Morais Silva, A. (1789)	Sim	Não	-	Não	Não
<i>Diccionario da Lingua Portuguesa</i>	Morais Silva, A. (1813)	Sim	Não	-	Não	Não
<i>Novo Diccionario da Lingua Portuguesa</i>	Faria, E. (1849)	Sim	Não	-	Não	Não
<i>Grande Diccionario Portuguez ou Thesouro da Lingua Portuguesa</i>	Vieira, F. D. (1871-1874)	Não	Não	-	Não	Não
<i>Diccionario Contemporane o da Lingua Portuguesa</i>	Caldas Aulete, F. (1881)	Sim	Sim	Sim	Sim	Sim
<i>Nôvo Dicionário da Língua Portuguêsa</i>	Figueiredo, C. (1899)	Não	Não	-	Não	Não

Anexo 2 – Dicionários impressos de língua portuguesa representativos do séc. XX

Dicionário	Autor(es)/ Editor(es), Ano	Info. sobre a pronúncia no lema inicial	Ortoépia	Transcrição fonética	Sílaba tónica	Divisão silábica
<i>Grande Dicionário da Língua Portuguesa</i>	Morais Silva, A. (1949-1959)	Sim	Sim	Não	Não	Não
<i>Grande Dicionário da Língua Portuguesa</i>	Machado, J. P. (1981)	Sim	Sim	Não	Não	Não
<i>Novo Dicionário Aurélio da Língua Portuguesa</i>	Aurélio, F. B. H. (1986)	Não	Sim	Não	Não	Não
<i>Dicionário Contemporâneo da Língua Portuguesa</i> <i>Caldas Aulete</i>	Caldas Aulete, J. (1987)	Sim	Sim	Sim*	Não	Sim*
<i>Grande Dicionário da Língua Portuguesa</i> <i>de Cândido de Figueiredo</i>	Figueiredo, C. (1996)	Não	Sim	Sim*	Não	Não

*Parâmetro apenas observado em empréstimos.

Anexo 3 – Dicionários impressos de língua portuguesa representativos do séc. XXI

Dicionário	Autor(es)/ Editor(es), Ano	Info. sobre a pronúncia no lema inicial	Ortoépia	Transcrição fonética/ Info.	Sílaba tónica	Divisão silábica
<i>Novo Aurélio Séc. XXI</i>	Aurélio, F. B. H. (1999)	Não	Sim	Sim*	Sim*	Não
<i>Dicionário Houaiss da Língua Portuguesa</i>	Houaiss, A. (2002-2003)	Não	Sim	Sim*	Sim*	Não
<i>Grande Dicionário da Língua Portuguesa</i>	Porto Editora (2004)	Não	Não	Sim	Sim	Não
<i>Dicionário Verbo: Língua Portuguesa</i>	Editorial Verbo (2006)	Não	Não	Não	Não	Não
<i>Novo Grande Dicionário Língua Portuguesa</i>	Texto Editores (2007)	Não	Não	Não	Não	Não

Anexo 4 – Dicionários digitais de língua portuguesa

Dicionário	Info. sobre a transcrição fonética	Transcrição fonética	Ortoépia	Sílaba tónica	Divisão silábica/ Translineação	Áudio
AULETE	Não	Não	Sim	Sim	Translineação	Não
DICIO	Não	Não	Não	Não	Translineação	Sim
DLP-ABL	Não	Não	Sim	Sim	Translineação	Não
DP	Não	Não	Não	Não	Translineação	Não
ESTRA VIZ	Não	Não	Não	Não	Não	Não
HOUAISS	Não	Sim*	Sim	Não	Não	Sim
INFO PÉDIA	Não	Sim	Sim	Sim	Translineação	Sim
INFOR MAL	Não	Sim	Não	Sim	Divisão silábica	Não
LÉXICO	Não	Não	Não	Sim	Translineação	Não
MD	Não	Não	Não	Não	Translineação	Não
MICHAELIS	Não	Sim*	Não	Não	Translineação	Não
PRIBERAM	Não	Não	Sim	Não	Translineação	Não

Anexo 5 – Edições do *Dicionário da Língua Portuguesa* da Academia das Ciências de Lisboa

Dicionário	Ano	Tipo	Info. sobre a conversão fonética	Transcrição fonética	Sílaba tónica	Divisão silábica	Áudio
<i>Diccionario da Lingoa Portuguesa</i>	1793	Impresso	Não	Não	Não	Não	—
<i>Dicionário da Língua Portuguesa da Academia</i> [A-Azuverte]	1976	Impresso	Não	Não	Não	Não	—
DLPC	2001	Impresso	Sim	Sim	Sim	Não	—
DLP-ACL	2023–presente	Digital	Não	Não	Não	Não	Não

Anexo 6 – Resultados das *regex* de translineação, marcação da sílaba tónica e transcrição fonética após correção manual

Translineação	Marcação da sílaba tónica	Transcrição fonética
alto -> al.to	alto -> 'alto	'alto -> ['aɫtu]
a -> a		a -> [e]
abóbora -> a.bó.bo.ra	abóbora -> a'bóbora	a'bóbora -> [e'βɔβure]
abraço -> a.bra.ço	abraço -> a'braço	a'braço -> [e'brasu]
abril -> a.bril	abril -> a'bril	a'bril -> [e'brit]
abrir -> a.brir	abrir -> a'brir	a'brir -> [e'brir]
acabar -> a.ca.bar	acabar -> aca'bar	aca'bar -> [eke'βar]
ação -> a.ção	ação -> a'ção	a'ção -> [a'sẽw]
achar -> a.char	achar -> a'char	a'char -> [e'ʃar]
acidente -> a.ci.den.te	acidente -> aci'dente	aci'dente -> [esi'ðeti]
acontecer -> a.con.te.cer	acontecer -> aconte'cer	aconte'cer -> [ekõti'ser]
acordar -> a.cor.dar	acordar -> acor'dar	acor'dar -> [ekur'dar]
acreditar -> a.cre.di.tar	acreditar -> credi'tar	acredi'tar -> [ekriði'tar]
açúcar -> a.çú.car	açúcar -> a'çúcar	a'çúcar -> e'sukar
adorar -> a.do.rar	adorar -> ado'rar	ado'rar -> [eðu'rar]
adulto -> a.dul.to	adulto -> a'dulto	a'dulto -> e'ðuɫtu
agora -> a.go.ra	agora -> a'gora	a'gora -> [e'ɣɔre]
agosto -> a.gos.to	agosto -> a'gosto	a'gosto -> e'ɣoftu
agradecer -> a.gra.de.cer	agradecer -> agrade'cer	agrade'cer -> [egreði'ser]
água -> á.gua	água -> 'água	'água -> ['agwe]
ah -> ah	ah -> 'ah	'ah -> ['a]
aí -> aí	aí -> a'í	a'í -> [e'i]
ainda -> a.in.da	ainda -> a'inda	a'inda -> [e'ĩde]
ajuda -> a.ju.da	ajuda -> a'juda	a'juda -> [e'ʒuðe]
ajudar -> a.ju.dar	ajudar -> aju'dar	aju'dar -> [eʒu'dar]
aldeia -> al.dei.a	aldeia -> al'deia	al'deia -> [aɫ'deje]
alegria -> a.le.gri.a	alegria -> ale'gria	ale'gria -> [eli'grie]
algo -> al.go	algo -> 'algo	'algo -> ['aɫgu]
alguém -> al.guém	alguém -> al'guém	al'guém -> [aɫ'gẽ]
algum -> al.gum	algum -> al'gum	al'gum -> [aɫ'gũ]
alho -> a.lho	alho -> 'alho	'alho -> ['aɫu]
ali -> a.li	ali -> a'li	a'li -> [e'li]
almoçar -> al.mo.çar	almoçar -> almo'çar	almo'çar -> [aɫmu'sar]
almoço -> al.mo.ço	almoço -> al'moço	al'moço -> [aɫ'mosu]

altura -> al.tu.ra	altura -> al'tura	al'tura -> [at'ture]
aluno -> a.lu.no	aluno -> a'luno	a'luno -> [e'lunu]
amanhã -> a.ma.nhã	amanhã -> ama'nhã	ama'nhã -> [ame'ɲẽ]
amar -> a.mar	amar -> a'mar	a'mar -> [e'mar]
amarelo -> a.ma.re.lo	amarelo -> ama'relo	ama'relo -> [eme'relu]
amigo -> a.mi.go	amigo -> a'migo	a'migo -> [e'miɣu]
amizade -> a.mi.za.de	amizade -> ami'zade	ami'zade -> [emi'zaði]
amor -> a.mor	amor -> a'mor	a'mor -> [e'mor]
andar -> an.dar	andar -> an'dar	an'dar -> [ẽ'dar]
animal -> a.ni.mal	animal -> ani'mal	ani'mal -> [eni'mat]
ano -> a.no	ano -> 'ano	'ano -> ['enu]
antigo -> an.ti.go	antigo -> an'tigo	an'tigo -> [ẽ'tiɣu]
ao -> ao		ao -> [aw]
apanhar -> a.pa.nhar	apanhar -> apa'nhar	apa'nhar -> [epe'ɲar]
aparecer -> a.pa.re.cer	aparecer -> apare'cer	apare'cer -> [eperi'ser]
apartamento -> a.par.ta.men.to	apartamento -> aparta'mento	aparta'mento -> [eperte'mẽtu]
apenas -> a.pe.nas	apenas -> a'penas	a'penas -> [e'peneɲ]
aprender -> a.pren.der	aprender -> apren'der	apren'der -> [eprẽ'der]
aquele -> a.que.le	aquele -> a'quele	a'quele -> [e'keli]
aqui -> a.qui	aqui -> a'qui	a'qui -> [e'ki]
aquilo -> a.qui.lo	aquilo -> a'quilo	a'quilo -> [e'kilu]
ar -> ar	ar -> 'ar	'ar -> ['ar]
área -> á.re.a	área -> 'área	'área -> ['arie]
armário -> ar.má.ri.o	armário -> ar'mário	ar'mário -> [er'mariu]
arranjar -> ar.ran.jar	arranjar -> arran'jar	arran'jar -> [erẽ'zar]
arroz -> ar.roz	arroz -> a'rroz	a'rroz -> [e'Roɲ]
arrumar -> ar.ru.mar	arrumar -> arru'mar	arru'mar -> [eru'mar]
artigo -> ar.ti.go	artigo -> ar'tigo	ar'tigo -> [er'tiɣu]
árvore -> ár.vo.re	árvore -> 'árvore	'árvore -> ['arvuri]
asa -> a.sa	asa -> 'asa	'asa -> ['aze]
assado -> as.sa.do	assado -> a'ssado	a'ssado -> [e'saðu]
assim -> as.sim	assim -> a'ssim	a'ssim -> [e'sĩ]
até -> a.té	até -> a'té	a'té -> [e'te]
atenção -> a.ten.ção	atenção -> aten'ção	aten'ção -> [etẽ'sẽw]
atirar -> a.ti.rar	atirar -> ati'rar	ati'rar -> [eti'rar]
atividade -> a.ti.vi.da.de	atividade -> ativi'dade	ativi'dade -> [etivi'ðaði]
ator -> a.tor	ator -> a'tor	a'tor -> [a'tor]
aula -> au.la	aula -> 'aula	'aula -> ['awle]
autor -> au.tor	autor -> au'tor	au'tor -> [aw'tor]

ave -> a.ve	ave -> 'ave	'ave -> ['avi]
avenida -> a.ve.ni.da	avenida -> ave'nida	ave'nida -> [evɨ'niðe]
avião -> a.vi.ão	avião -> avi'ão	avi'ão -> [evɨ'ẽw]
avô -> a.vô	avô -> a'vô	a'vô -> [e'vo]
azeite -> a.zei.te	azeite -> a'zeite	a'zeite -> [e'zejti]
azul -> a.zul	azul -> a'zul	a'zul -> [e'zuʔ]
bacalhau -> ba.ca.lhau	bacalhau -> baca'lhau	baca'lhau -> [beke'ʎaw]
bairro -> bair.ro	bairro -> 'bairro	'bairro -> ['bajru]
baixo -> bai.xo	baixo -> 'baixo	'baixo -> ['bajfu]
banho -> ba.nho	banho -> 'banho	'banho -> ['beɲu]
barba -> bar.ba	barba -> 'barba	'barba -> ['barbe]
barco -> bar.co	barco -> 'barco	'barco -> ['barku]
barulho -> ba.ru.lho	barulho -> ba'ulho	ba'ulho -> [be'ruʎu]
bastante -> bas.tan.te	bastante -> bas'tante	bas'tante -> [beʃ'tẽti]
batata -> ba.ta.ta	batata -> ba'tata	ba'tata -> [be'tate]
bater -> ba.ter	bater -> ba'ter	ba'ter -> [be'ter]
bebé -> be.bé	bebé -> be'bé	be'bé -> [be'βe]
beber -> be.ber	beber -> be'ber	be'ber -> [bi'βer]
bebida -> be.bi.da	bebida -> be'bida	be'bida -> [bi'βiðe]
beijinho -> bei.ji.nho	beijinho -> bei'jinho	bei'jinho -> [bej'ziɲu]
belo -> be.lo	belo -> 'belo	'belo -> ['belu]
bem -> bem	bem -> 'bem	'bem -> ['bẽ]
bicicleta -> bi.ci.cle.ta	bicicleta -> bici'cleta	bici'cleta -> [bisi'klete]
bife -> bi.fe	bife -> 'bife	'bife -> ['bifi]
bilhete -> bi.lhe.te	bilhete -> bi'lhete	bi'lhete -> [bi'ʎeti]
boa-tarde -> bo.a tar.de	boa-tarde -> boa 'tarde	boa 'tarde -> [boe 'tardi]
boca -> bo.ca	boca -> 'boca	'boca -> ['boke]
boi -> boi	boi -> 'boi	'boi -> ['boj]
bola -> bo.la	bola -> 'bola	'bola -> ['boʎe]
bolo -> bo.lo	bolo -> 'bolo	'bolo -> 'bolu
bom -> bom	bom -> 'bom	'bom -> ['bõ]
bom-dia -> bom di.a	bom-dia -> bom-'dia	bom-'dia -> [bõ 'die]
boneca -> bo.ne.ca	boneca -> bo'neca	bo'neca -> [bu'neke]
boneco -> bo.ne.co	boneco -> bo'neco	bo'neco -> [bu'neku]
bonito -> bo.ni.to	bonito -> bo'nito	bo'nito -> [bu'nitu]
braço -> bra.ço	braço -> 'braço	'braço -> ['brasu]
branco -> bran.co	branco -> 'branco	'branco -> ['brẽku]
brincar -> brin.car	brincar -> brin'car	brin'car -> [brĩ'kar]
brinquedo -> brin.que.do	brinquedo -> brin'quedo	brin'quedo -> [brĩ'keðu]

buraco -> bu.ra.co	buraco -> bu'racó	bu'racó -> [bu'ra.ku]
cá -> cá	cá -> 'cá	'cá -> ['ka]
cabeça -> ca.be.ça	cabeça -> ca'beça	ca'beça -> [ke'βese]
cabelo -> ca.be.lo	cabelo -> ca'belo	ca'belo -> [ke'βelu]
cada -> ca.da	cada -> 'cada	'cada -> ['keðe]
caderno -> ca.der.no	caderno -> ca'derno	ca'derno -> [ke'ðɛnu]
café -> ca.fé	café -> ca'fé	ca'fé -> [ke'fe]
cair -> ca.ir	cair -> ca'ir	ca'ir -> [ke'ir]
calor -> ca.lor	calor -> ca'lor	ca'lor -> [ke'lor]
cama -> ca.ma	cama -> 'cama	'cama -> ['keme]
caminho -> ca.mi.nho	caminho -> ca'minho	ca'minho -> [ke'miɲu]
camisa -> ca.mi.sa	camisa -> ca'misa	ca'misa -> [ke'mize]
campo -> cam.po	campo -> 'campo	'campo -> ['kẽpu]
canção -> can.ção	canção -> can'ção	can'ção -> [kẽ'sẽw]
canela -> ca.ne.la	canela -> ca'nela	ca'nela -> [ke'nele]
cantar -> can.tar	cantar -> can'tar	can'tar -> [kẽ'tar]
canto -> can.to	canto -> 'canto	'canto -> ['kẽtu]
cão -> cão	cão -> 'cão	'cão -> ['kẽw]
capaz -> ca.paz	capaz -> ca'paz	ca'paz -> [ke'paʃ]
cara -> ca.ra	cara -> 'cara	'cara -> ['kare]
carnaval -> car.na.val	carnaval -> carna'val	carna'val -> [kerne'vaʔ]
carne -> car.ne	carne -> 'carne	'carne -> ['karni]
caro -> ca.ro	caro -> 'caro	'caro -> ['karu]
carro -> car.ro	carro -> 'carro	'carro -> ['karu]
carta -> car.ta	carta -> 'carta	'carta -> ['karte]
casa -> ca.sa	casa -> 'casa	'casa -> ['kaze]
casar -> ca.sar	casar -> ca'sar	ca'sar -> [ke'zar]
casca -> cas.ca	casca -> 'casca	'casca -> ['ka[ke]
caso -> ca.so	caso -> 'caso	'caso -> ['kazu]
castanha -> cas.ta.nha	castanha -> cas'tanha	cas'tanha -> [kes'teɲe]
castanho -> cas.ta.nho	castanho -> cas'tanho	cas'tanho -> [kes'tɛɲu]
cavalo -> ca.va.lo	cavalo -> ca'valo	ca'valo -> [ke'valu]
cebola -> ce.bo.la	cebola -> ce'bola	ce'bola -> [si'βole]
cedo -> ce.do	cedo -> 'cedo	'cedo -> ['seðu]
cenoura -> ce.nou.ra	cenoura -> ce'noura	ce'noura -> [si'nore]
centro -> cen.tro	centro -> 'centro	'centro -> ['sẽtru]
certeza -> cer.te.za	certeza -> cer'teza	cer'teza -> [sir'teze]
certo -> cer.to	certo -> 'certo	'certo -> ['sẽrtu]
céu -> céu	céu -> 'céu	'céu -> ['sɛw]
chamar -> cha.mar	chamar -> cha'mar	cha'mar -> [ʃe'mar]

chão -> chão	chão -> 'chão	'chão -> ['jẽw]
chave -> cha.ve	chave -> 'chave	'chave -> ['javɨ]
chegar -> che.gar	chegar -> che'gar	che'gar -> [ʃɨ'ɣar]
cheio -> chei.o	cheio -> 'cheio	'cheio -> ['jeju]
cheiro -> chei.ro	cheiro -> 'cheiro	'cheiro -> ['fejru]
chinês -> chi.nês	chinês -> chi'nês	chi'nês -> [ʃi'neʃ]
chocolate -> cho.co.la.te	chocolate -> choco'late	choco'late -> [ʃuku'latɨ]
chorar -> cho.rar	chorar -> cho'rar	cho'rar -> [ʃu'rar]
chouriço -> chou.ri.ço	chouriço -> chou'riço	chou'riço -> [ʃo'risu]
chover -> cho.ver	chover -> cho'ver	cho'ver -> [ʃu'ver]
chuva -> chu.va	chuva -> 'chuva	'chuva -> ['juve]
cidade -> ci.da.de	cidade -> ci'dade	ci'dade -> [si'ðaðɨ]
cigarra -> ci.gar.ra	cigarra -> ci'garra	ci'garra -> [si'ɣare]
cinco -> cin.co	cinco -> 'cinco	'cinco -> ['sĩku]
cinema -> ci.ne.ma	cinema -> ci'nema	ci'nema -> [si'neme]
claro -> cla.ro	claro -> 'claro	'claro -> ['klaru]
cliente -> cli.en.te	cliente -> cli'ente	cli'ente -> [kli'ẽti]
clube -> clu.be	clube -> 'clube	'clube -> ['kluβɨ]
coelho -> co.e.lho	coelho -> co'elho	co'elho -> [ku'e(j)ʎu] / [ku'eʎu]
coisa -> coi.sa	coisa -> 'coisa	'coisa -> ['kojze]
colega -> co.le.ga	colega -> co'lega	co'lega -> [ku'ʎɐ]
colher -> co.lher	colher -> co'lher	co'lher -> [ku'ʎer] / [ku'ʎɛr]
colocar -> co.lo.car	colocar -> colo'car	colo'car -> [kulu'kar]
com -> com		com -> [kõ]
comboio -> com.boi.o	comboio -> com'boio	com'boio -> [kõ'bɔju]
começar -> co.me.çar	começar -> come'çar	come'çar -> [kumi'sar]
comer -> co.mer	comer -> co'mer	co'mer -> [ku'mer]
comida -> co.mi.da	comida -> co'mida	co'mida -> [ku'miðe]
como -> co.mo		como -> [komu]
compra -> com.pra	compra -> 'compra	'compra -> ['kõpre]
comprar -> com.prar	comprar -> com'prar	com'prar -> [kõ'prar]
comprido -> com.pri.do	comprido -> com'prido	com'prido -> [kõ'priðu]
curso -> con.cur.so	curso -> con'curso	con'curso -> [kõ'kursu]
conhecer -> co.nhe.cer	conhecer -> conhe'cer	conhe'cer -> [koɲi'ser]
consulta -> con.sul.ta	consulta -> con'sulta	con'sulta -> [kõ'suʎte]
conta -> con.ta	conta -> 'conta	'conta -> ['kõte]
contar -> con.tar	contar -> con'tar	con'tar -> [kõ'tar]
contente -> con.ten.te	contente -> con'tente	con'tente -> [kõ'tẽti]
continuar -> con.ti.nu.ar	continuar -> continu'ar	continu'ar -> [kõtinu'ar]

contra -> con.tra	contra -> 'contra	'contra -> ['kõtre]
conversa -> con.ver.sa	conversa -> con'versa	con'versa -> [kõ'verse]
conversar -> con.ver.sar	conversar -> conver'sar	conver'sar -> [kõvir'sar]
copo -> co.po	copo -> 'copo	'copo -> ['kõpu]
cor -> cor	cor -> 'cor	'cor -> ['kor]
coração -> co.ra.ção	coração -> cora'ção	cora'ção -> [kure'sẽw]
corda -> cor.da	corda -> 'corda	'corda -> ['kõrde]
corpo -> cor.po	corpo -> 'corpo	'corpo -> ['korpu]
correr -> cor.rer	correr -> co'rrer	co'rrer -> [ku'rer]
cortar -> cor.tar	cortar -> cor'tar	cor'tar -> [kur'tar]
costa -> cos.ta	costa -> 'costa	'costa -> ['kõjte]
costumar -> cos.tu.mar	costumar -> costu'mar	costu'mar -> [kufu'mar]
cozinha -> co.zi.nha	cozinha -> co'zinha	co'zinha -> [ku'zipe]
cozinhar -> co.zi.nhar	cozinhar -> cozi'nhar	cozi'nhar -> [kuzi'nar]
creme -> cre.me	creme -> 'creme	'creme -> ['kreĩ]
crescer -> cres.cer	crescer -> cres'cer	cres'cer -> [kri'fer]
criança -> cri.an.ça	criança -> cri'ança	cri'ança -> [kri'ẽse]
cuidado -> cui.da.do	cuidado -> cui'dado	cui'dado -> [kuj'ðaðu]
cuidar -> cui.dar	cuidar -> cui'dar	cui'dar -> [kuj'ðar]
curso -> cur.so	curso -> 'curso	'curso -> ['kursu]
custar -> cus.tar	custar -> cus'tar	cus'tar -> [kuf'tar]
dançar -> dan.çar	dançar -> dan'çar	dan'çar -> [dẽ'sar]
dar -> dar	dar -> 'dar	'dar -> ['dar]
data -> da.ta	data -> 'data	'data -> ['date]
de -> de		de -> [di]
decidir -> de.ci.dir	decidir -> deci'dir	deci'dir -> [disi'ðir]
dedo -> de.do	dedo -> 'dedo	'dedo -> ['deðu]
deitar -> dei.tar	deitar -> dei'tar	dei'tar -> [dej'tar]
deixar -> dei.xar	deixar -> dei'xar	dei'xar -> [dej'jar]
demasiado -> de.ma.si.a.do	demasiado -> demasi'ado	demasi'ado -> [dimezi'aðu]
demorar -> de.mo.rar	demorar -> demo'rar	demo'rar -> [ðimu'rar]
dente -> den.te	dente -> 'dente	'dente -> ['dẽti]
depois -> de.pois	depois -> de'pois	de'pois -> [di'poj]
depressa -> de.pres.sa	depressa -> de'pressa	de'pressa -> [di'prese]
descer -> des.cer	descer -> des'cer	des'cer -> [di'fer]
desculpa -> des.cul.pa	desculpa -> des'culpa	des'culpa -> [dif'kufpe]
desculpar -> des.cul.par	desculpar -> descul'par	descul'par -> [difku'tpar]
desde -> des.de	desde -> 'desde	'desde -> ['dezdĩ]
desporto -> des.por.to	desporto -> des'porto	des'porto -> [dif'portu]

deus -> deus	deus -> 'deus	'deus -> ['dew]
dever -> de.ver	dever -> de'ver	de'ver -> [dɨ'ver]
dez -> dez	dez -> 'dez	'dez -> ['dɛ]
dezembro -> de.zem.bro	dezembro -> de'zembro	de'zembro -> [dɨ'zẽbru]
dia -> di.a	dia -> 'dia	'dia -> ['diɛ]
diferente -> di.fe.ren.te	diferente -> dife'rente	dife'rente -> [difɨ'rẽti]
difícil -> di.fí.cil	difícil -> di'fícil	di'fícil -> [di'fisiɫ]
dinheiro -> di.nhei.ro	dinheiro -> di'nheiro	di'nheiro -> [di'nejru]
direito -> di.rei.to	direito -> di'reito	di'reito -> [di'rejtu]
dizer -> di.zer	dizer -> di'zer	di'zer -> [di'zer]
doce -> do.ce	doce -> 'doce	'doce -> ['dosɨ]
doença -> do.en.ça	doença -> do'ença	do'ença -> [du'ẽsɛ]
dois -> dois	dois -> 'dois	'dois -> ['doj]
domingo -> do.min.go	domingo -> do'mingo	do'mingo -> [du'mĩgu]
dono -> do.no	dono -> 'dono	'dono -> ['donu]
dormir -> dor.mir	dormir -> dor'mir	dor'mir -> [dur'mir]
doutor -> dou.tor	doutor -> dou'tor	dou'tor -> [do'tor]
doze -> do.ze	doze -> 'doze	'doze -> ['dozi]
durante -> du.ran.te	durante -> du'rante	du'rante -> [du'rẽti]
e -> e		e -> [i]
em -> em		em -> [ẽ]
embora -> em.bo.ra	embora -> em'bora	em'bora -> [ẽ'bɔrɛ]
empresa -> em.pre.sa	empresa -> em'presa	em'presa -> [ẽ'prezɛ]
encher -> en.cher	encher -> en'cher	en'cher -> [ẽ'ʃɛr]
encontrar -> en.con.trar	encontrar -> encon'trar	encon'trar -> [ẽkõ'trar]
enganar -> en.ga.nar	enganar -> enga'nar	enga'nar -> [ẽgɛ'nar]
enorme -> e.nor.me	enorme -> e'norme	e'norme -> [e'nɔrmɨ] / [i'nɔrmɨ]
enquanto -> en.quan.to	enquanto -> en'quanto	en'quanto -> [ẽ'kwẽtu]
então -> en.tão	então -> en'tão	en'tão -> [ẽ'tẽw]
entrada -> en.tra.da	entrada -> en'trada	en'trada -> [ẽ'traðɛ]
entrar -> en.trar	entrar -> en'trar	en'trar -> [ẽ'trar]
entre -> en.tre	entre -> 'entre	'entre -> ['ẽtri]
entregar -> en.tre.gar	entregar -> entre'gar	entre'gar -> [ẽtri'ɣar]
entretanto -> en.tre.tan.to	entretanto -> entre'tanto	entre'tanto -> [ẽtri'tẽtu]
enviar -> en.vi.ar	enviar -> envi'ar	envi'ar -> [ẽvi'ar]
equipa -> e.qui.pa	equipa -> e'quipa	e'quipa -> [e'kipe] / [i'kipe]
escola -> es.co.la	escola -> es'cola	es'cola -> [ɨf'kole]
escolher -> es.co.lher	escolher -> esco'lher	esco'lher -> [ɨku'ler]
esconder -> es.con.der	esconder -> escon'der	escon'der -> [ɨkõ'der]

escuro -> es.cu.ro	escuro -> es'curo	es'curo -> [iʃ'kuru]
espaço -> es.pa.ço	espaço -> es'paço	es'paço -> [iʃ'pasu]
espera -> es.pe.ra	espera -> es'pera	es'pera -> [iʃ'pɛrɐ]
esperar -> es.pe.rar	esperar -> espe'rar	espe'rar -> [iʃpi'rar]
esquecer -> es.que.cer	esquecer -> esque'cer	esque'cer -> [iʃkɛ'ser]
esse -> es.se	esse -> 'esse	'esse -> ['esi]
estar -> es.tar	estar -> es'tar	es'tar -> [iʃ'tar]
este -> es.te	este -> 'este	'este -> ['efti]
estranho -> es.tra.nho	estranho -> es'tranho	es'tranho -> [iʃ'trɛnu]
estudante -> es.tu.dan.te	estudante -> estu'dante	estu'dante -> [iʃtu'ðɛti]
estudar -> es.tu.dar	estudar -> estu'dar	estu'dar -> [iʃtu'ðar]
estudo -> es.tu.do	estudo -> es'tudo	es'tudo -> [iʃ'tuðu]
eu -> eu	eu -> 'eu	'eu -> ['ew]
euro -> eu.ro	euro -> 'euro	'euro -> ['ewrɔ]
exemplo -> e.xem.plo	exemplo -> e'xemplo	e'xemplo -> [e'zɛplu] / [i'zɛplu]
fácil -> fá.cil	fácil -> 'fácil	'fácil -> ['fasit]
fada -> fa.da	fada -> 'fada	'fada -> ['faðɐ]
falar -> fa.lar	falar -> fa'lar	fa'lar -> [fɛ'lar]
faltar -> fal.tar	faltar -> fal'tar	fal'tar -> [fa'tar]
família -> fa.mí.li.a	família -> fa'mília	fa'mília -> [fɛ'milie]
famoso -> fa.mo.so	famoso -> fa'moso	fa'moso -> [fɛ'mozu]
farinha -> fa.ri.nha	farinha -> fa'rinha	fa'rinha -> [fɛ'riɲɐ]
fazer -> fa.zer	fazer -> fa'zer	fa'zer -> [fɛ'zer]
fechar -> fe.char	fechar -> fe'char	fe'char -> [fɛ'ʃar]
feijão -> fei.jão	feijão -> fei'jão	fei'jão -> [fɛj'ʒɛw]
feira -> fei.ra	feira -> 'feira	'feira -> ['fɛjɐ]
feliz -> fe.liz	feliz -> fe'liz	fe'liz -> [fɛ'liʃ]
férias -> fé.ri.as	férias -> 'férias	'férias -> ['fɛrie]
ferver -> fer.ver	ferver -> fer'ver	fer'ver -> [fir'ver]
festa -> fes.ta	festa -> 'festa	'festa -> ['fɛʃtɐ]
ficar -> fi.car	ficar -> fi'car	fi'car -> [fi'kar]
filho -> fi.lho	filho -> 'filho	'filho -> ['fiʎu]
filme -> fil.me	filme -> 'filme	'filme -> ['fiʎmɛ]
fim -> fim	fim -> 'fim	'fim -> ['fi]
final -> fi.nal	final -> fi'nal	fi'nal -> [fi'nat]
finalmente -> fi.nal.men.te	finalmente -> final'mente	final'mente -> [fina'tmɛti]
fino -> fi.no	fino -> 'fino	'fino -> ['finu]
flor -> flor	flor -> 'flor	'flor -> ['flor]
folha -> fo.lha	folha -> 'folha	'folha -> ['foʎɐ]

força -> for.ça	força -> 'força	'força -> ['fɔrsɐ]
forma -> for.ma	forma -> 'forma	'forma -> ['fɔrmɐ] / ['fɔrmɐ]
formiga -> for.mi.ga	formiga -> for'miga	for'miga -> [fur'miɣɐ]
forno -> for.no	forno -> 'forno	'forno -> ['fɔrnɐ]
forte -> for.te	forte -> 'forte	'forte -> ['fɔrtɨ]
frase -> fra.se	frase -> 'frase	'frase -> ['frazɨ]
frente -> fren.te	frente -> 'frente	'frente -> ['frɛtɨ]
fresco -> fres.co	fresco -> 'fresco	'fresco -> ['frɛʃku]
frio -> fri.o	frio -> 'frio	'frio -> ['friɐ]
fruta -> fru.ta	fruta -> 'fruta	'fruta -> ['frutɐ]
fruto -> fru.to	fruto -> 'fruto	'fruto -> ['frutu]
fugir -> fu.gir	fugir -> fu'gir	fu'gir -> [fu'ʒir]
fundo -> fun.do	fundo -> 'fundo	'fundo -> ['fũdu]
futebol -> fu.te.bol	futebol -> fute'bol	fute'bol -> [futi'βɔt]
futuro -> fu.tu.ro	futuro -> fu'turo	fu'turo -> [fu'turu]
galo -> ga.lo	galo -> 'galo	'galo -> ['galu]
ganhar -> ga.nhar	ganhar -> ga'nhar	ga'nhar -> [gɐ'nar] / [ga'nar]
gastar -> gas.tar	gastar -> gas'tar	gas'tar -> [gɛ'tar]
gato -> ga.to	gato -> 'gato	'gato -> ['gatu]
gema -> ge.ma	gema -> 'gema	'gema -> ['ʒemɐ]
gente -> gen.te	gente -> 'gente	'gente -> ['ʒɛtɨ]
geralmente -> ge.ral.men.te	geralmente -> geral'mente	geral'mente -> [ʒɨrat'mɛtɨ]
golo -> go.lo	golo -> 'golo	'golo -> ['golu]
gostar -> gos.tar	gostar -> gos'tar	gos'tar -> [gu'tar]
gosto -> gos.to	gosto -> 'gosto	'gosto -> ['goʃtu]
gota -> go.ta	gota -> 'gota	'gota -> ['gotɐ]
grande -> gran.de	grande -> 'grande	'grande -> ['grɛdɨ]
gritar -> gri.tar	gritar -> gri'tar	gri'tar -> [gri'tar]
grupo -> gru.po	grupo -> 'grupo	'grupo -> ['grupu]
guardar -> guar.dar	guardar -> guar'dar	guar'dar -> [g'wɛr'dar]
guerra -> guer.ra	guerra -> 'guerra	'guerra -> ['gɛrɐ]
haver -> ha.ver	haver -> ha'ver	ha'ver -> [ɐ'ver]
história -> his.tó.ri.a	história -> his'tória	his'tória -> [iʃ'tɔrie]
hoje -> ho.je	hoje -> 'hoje	['oʒɨ]
homem -> ho.mem	homem -> 'homem	'homem -> ['ɔmɛ]
hora -> ho.ra	hora -> 'hora	'hora -> ['ɔrɐ]
horário -> ho.rá.ri.o	horário -> ho'rário	ho'rário -> [ɔ'rariu] / [o'rariu]
horta -> hor.ta	horta -> 'horta	'horta -> ['ɔrtɐ]
hospital -> hos.pi.tal	hospital -> hospi'tal	hospi'tal -> [ɔʃpi'tat]
hotel -> ho.tel	hotel -> ho'tel	ho'tel -> [ɔ'tɛt]

idade -> i.da.de	idade -> i'dade	i'dade -> [i'ðaðɨ]
ideia -> i.dei.a	ideia -> i'deia	i'deia -> [i'ðeje]
igual -> i.gual	igual -> i'gual	i'gual -> [i'gʷat]
ilha -> i.lha	ilha -> 'ilha	'ilha -> ['ilɐ]
imagem -> i.ma.gem	imagem -> i'magem	i'magem -> [i'mazẽ]
imenso -> i.men.so	imenso -> i'menso	i'menso -> [i'mẽsu]
importante -> im.por.tan.te	importante -> impor'tante	impor'tante -> [ĩpur'tɛtɨ]
início -> i.ní.ci.o	início -> i'nício	i'nício -> [i'nisiu]
inventar -> in.ven.tar	inventar -> inven'tar	inven'tar -> [ĩvẽ'tar]
inverno -> in.ver.no	inverno -> in'verno	in'verno -> [ĩ'vernu]
ir -> ir	ir -> 'ir	'ir -> ['ir]
irmão -> ir.mão	irmão -> ir'mão	ir'mão -> [ir'mẽw]
isso -> is.so	isso -> 'isso	'isso -> ['isu]
isto -> is.to	isto -> 'isto	'isto -> ['iftu]
já -> já	já -> 'já	'já -> ['ʒa]
janeiro -> ja.nei.ro	janeiro -> ja'neiro	ja'neiro -> [ʒe'nejru]
janela -> ja.ne.la	janela -> ja'nela	ja'nela -> [ʒe'nele]
jantar -> jan.tar	jantar -> jan'tar	jan'tar -> [ʒẽ'tar]
jardim -> jar.dim	jardim -> jar'dim	jar'dim -> [ʒer'dĩ]
jogador -> jo.ga.dor	jogador -> joga'dor	joga'dor -> [ʒuɣe'dor]
jogar -> jo.gar	jogar -> jo'gar	jo'gar -> [ʒu'ɣar]
jogo -> jo.go	jogo -> 'jogo	'jogo -> ['ʒoɣu]
jovem -> jo.vem	jovem -> 'jovem	'jovem -> ['ʒovẽ]
julho -> ju.lho	julho -> 'julho	'julho -> ['ʒulu]
junho -> ju.nho	junho -> 'junho	'junho -> ['ʒunu]
juntar -> jun.tar	juntar -> jun'tar	jun'tar -> [ʒũ'tar]
junto -> jun.to	junto -> 'junto	'junto -> ['ʒũtu]
lá -> lá	lá -> 'lá	'lá -> ['la]
lado -> la.do	lado -> 'lado	'lado -> ['laðu]
lavar -> la.var	lavar -> la'var	la'var -> [le'var]
leão -> le.ão	leão -> le'ão	le'ão -> [li'ẽw]
legume -> le.gu.me	legume -> le'gume	le'gume -> [li'ɣumi]
leite -> lei.te	leite -> 'leite	'leite -> ['lejtɨ]
lembrar -> lem.brar	lembrar -> lem'brar	lem'brar -> [lẽ'brar]
lenda -> len.da	lenda -> 'lenda	'lenda -> ['lẽde]
ler -> ler	ler -> 'ler	'ler -> ['ler]
letra -> le.tra	letra -> 'letra	'letra -> ['letre]
levantar -> le.van.tar	levantar -> levan'tar	levan'tar -> [livẽ'tar]
levar -> le.var	levar -> le'var	le'var -> [li'var]

leve -> le.ve	leve -> 'leve	'leve -> ['levi]
ligar -> li.gar	ligar -> li'gar	li'gar -> [li'ɣar]
limpar -> lim.par	limpar -> lim'par	lim'par -> [lĩ'par]
limpo -> lim.po	limpo -> 'limpo	'limpo -> ['lĩpu]
lindo -> lin.do	lindo -> 'lindo	'lindo -> ['lĩdu]
língua -> lín.gua	língua -> 'língua	'língua -> ['lĩg ^w e]
linha -> li.nha	linha -> 'linha	'linha -> ['lĩne]
lista -> lis.ta	lista -> 'lista	'lista -> ['lĩfte]
livre -> li.vre	livre -> 'livre	'livre -> ['livri]
livro -> li.vro	livro -> 'livro	'livro -> ['livru]
local -> lo.cal	local -> lo'cal	lo'cal -> [lu'kaɫ]
logo -> lo.go	logo -> 'logo	'logo -> ['lɔɣu]
loja -> lo.ja	loja -> 'loja	'loja -> ['lɔʒe]
longe -> lon.ge	longe -> 'longe	'longe -> ['lõʒi]
longo -> lon.go	longo -> 'longo	'longo -> ['lõgu]
lua -> lu.a	lua -> 'lua	'lua -> ['lue]
lugar -> lu.gar	lugar -> lu'gar	lu'gar -> [lu'ɣar]
lume -> lu.me	lume -> 'lume	'lume -> ['lumi]
lutar -> lu.tar	lutar -> lu'tar	lu'tar -> [lu'tar]
luz -> luz	luz -> 'luz	'luz -> ['lu]
macaco -> ma.ca.co	macaco -> ma'caco	ma'caco -> [me'kaku]
madeira -> ma.dei.ra	madeira -> ma'deira	ma'deira -> [me'ðejre]
maio -> mai.o	maio -> 'maio	'maio -> ['maju]
mais -> mais	mais -> 'mais	'mais -> ['maj]
mal -> mal	mal -> 'mal	'mal -> ['maɫ]
mala -> ma.la	mala -> 'mala	'mala -> ['male]
mamã -> ma.mã	mamã -> ma'mã	ma'mã -> [me'mẽ]
mandar -> man.dar	mandar -> man'dar	man'dar -> [mẽ'dar]
maneira -> ma.nei.ra	maneira -> ma'neira	ma'neira -> [me'nejre]
manhã -> ma.nhã	manhã -> ma'nhã	ma'nhã -> [me'nẽ]
manteiga -> man.tei.ga	manteiga -> man'teiga	man'teiga -> [mẽ'tejye]
manter -> man.ter	manter -> man'ter	man'ter -> [mẽ'ter]
mão -> mão	mão -> 'mão	'mão -> ['mẽw]
máquina -> má.qui.na	máquina -> 'máquina	'máquina -> ['makine]
mar -> mar	mar -> 'mar	'mar -> ['mar]
marca -> mar.ca	marca -> 'marca	'marca -> ['marke]
março -> mar.ço	março -> 'março	'março -> ['marsu]
margarina -> mar.ga.ri.na	margarina -> marga'rina	marga'rina -> [merge'rine]
marido -> ma.ri.do	marido -> ma'rido	ma'rido -> [me'riðu]
mas -> mas		mas -> [me]

mata -> ma.ta	mata -> 'mata	'mata -> ['mate]
material -> ma.te.ri.al	material -> materi'al	materi'al -> [metiri'aʔ]
mau -> mau	mau -> 'mau	'mau -> ['maw]
médico -> mé.di.co	médico -> 'médico	'médico -> ['mɛðiku]
medo -> me.do	medo -> 'medo	'medo -> ['meðu]
meia -> mei.a	meia -> 'meia	'meia -> ['meje]
meio -> mei.o	meio -> 'meio	'meio -> ['meju]
mel -> mel	mel -> 'mel	'mel -> ['mɛʔ]
menino -> me.ni.no	menino -> me'nino	me'nino -> [mi'ninu]
menos -> me.nos	menos -> 'menos	'menos -> ['menu]
mês -> mês	mês -> 'mês	'mês -> ['mej]
mesa -> me.sa	mesa -> 'mesa	'mesa -> ['meze]
mesmo -> mes.mo	mesmo -> 'mesmo	'mesmo -> ['mezmu]
meter -> me.ter	meter -> me'ter	me'ter -> [mi'ter]
metro -> me.tro	metro -> 'metro	'metro -> ['mɛtru]
mexer -> me.xer	mexer -> me'xer	me'xer -> [mi'fer]
minuto -> mi.nu.to	minuto -> mi'nuto	mi'nuto -> [mi'nutu]
modo -> mo.do	modo -> 'modo	'modo -> ['mɔðu]
momento -> mo.men.to	momento -> mo'mento	mo'mento -> [mu'mɛtu]
montanha -> mon.ta.nha	montanha -> mon'tanha	mon'tanha -> [mõ'tɛpe]
morar -> mo.rar	morar -> mo'rar	mo'rar -> [mu'rar]
morrer -> mor.rer	morrer -> mo'rrer	mo'rrer -> [mu'rer]
morte -> mor.te	morte -> 'morte	'morte -> ['mɔrti]
mosca -> mos.ca	mosca -> 'mosca	'mosca -> ['moʃke]
mostrar -> mos.trar	mostrar -> mos'trar	mos'trar -> [muʃ'trar]
mudar -> mu.dar	mudar -> mu'dar	mu'dar -> [mu'ðar]
muito -> mui.to	muito -> 'muito	'muito -> ['mũtu]
mulher -> mu.lher	mulher -> mu'lher	mu'lher -> [mu'ʎer]
mundo -> mun.do	mundo -> 'mundo	'mundo -> ['mũdu]
museu -> mu.seu	museu -> mu'seu	mu'seu -> [mu'zew]
música -> mú.si.ca	música -> 'música	'música -> ['muzike]
nada -> na.da	nada -> 'nada	'nada -> ['naðe]
nadar -> na.dar	nadar -> na'dar	na'dar -> [ne'ðar]
não -> não	não -> 'não	'não -> ['nẽw]
nascer -> nas.cer	nascer -> nas'cer	nas'cer -> [neʃ'ser]
natal -> na.tal	natal -> na'tal	na'tal -> [ne'taʔ]
natural -> na.tu.ral	natural -> natu'ral	natu'ral -> [netu'raʔ]
natureza -> na.tu.re.za	natureza -> natu'reza	natu'reza -> [netu'reze]
negro -> ne.gro	negro -> 'negro	'negro -> ['negru]
nem -> nem	nem -> 'nem	'nem -> ['nẽ]

nenhum -> ne.nhum	nenhum -> ne'nhum	ne'nhum -> [nɛ'ɲũ]
neve -> ne.ve	neve -> 'neve	'neve -> ['nɛvɛ]
ninguém -> nin.gué[m]	ninguém -> nin'guém	nin'guém -> [nĩ'gẽ]
noite -> noi.te	noite -> 'noite	'noite -> ['nojti]
nome -> no.me	nome -> 'nome	'nome -> ['nomɛ]
normal -> nor.mal	normal -> nor'mal	nor'mal -> [nɔr'maɫ]
normalmente -> nor.mal.men.te	normalmente -> normal'mente	normal'mente -> [nɔrmaɫ'mẽti]
notícia -> no.tí.ci.a	notícia -> no'tícia	no'tícia -> [nu'tisɛ]
novembro -> no.vem.bro	novembro -> no'vembro	no'vembro -> [nu'vẽbru]
novo -> no.vo	novo -> 'novo	'novo -> ['novu]
noz -> noz	noz -> 'noz	'noz -> ['nɔ]
número -> nú.me.ro	número -> 'número	'número -> ['numɛru]
nunca -> nun.ca	nunca -> 'nunca	'nunca -> ['nũke]
nuvem -> nu.vem	nuvem -> 'nuvem	'nuvem -> ['nuvẽ]
o -> o		o -> [u]
ó -> ó	ó -> 'ó	'ó -> ['ɔ]
obra -> o.bra	obra -> 'obra	'obra -> ['ɔbre]
obrigado -> o.bri.ga.do	obrigado -> obri'gado	obri'gado -> [ɔbri'ɣaðu] / [obri'ɣaðu]
oferecer -> o.fe.re.cer	oferecer -> ofere'cer	ofere'cer -> [ɔfɛri'sɛr] / [ofɛri'sɛr]
oh -> oh	oh -> 'oh	'oh -> 'u
oito -> oi.to	oito -> 'oito	'oito -> ['ojtu]
olá -> o.lá	olá -> o'lá	o'lá -> [ɔ'la]
olhar -> o.lhar	olhar -> o'lhar	o'lhar -> [ɔ'ɫar] / [o'ɫar]
olho -> o.lho	olho -> 'olho	'olho -> ['oɫu]
onde -> on.de	onde -> 'onde	'onde -> ['õdi]
ora -> o.ra	ora -> 'ora	'ora -> ['ɔre]
origem -> o.ri.gem	origem -> o'rigem	o'rigem -> [ɔ'riʒẽ] / [o'riʒẽ]
osso -> os.so	osso -> 'osso	'osso -> ['osu]
ótimo -> ó.ti.mo	ótimo -> 'ótimo	'ótimo -> ['ɔtimu]
ou -> ou		ou -> [o]
outono -> ou.to.no	outono -> ou'tono	ou'tono -> [o'tonu]
outro -> ou.tro	outro -> 'outro	'outro -> ['otru]
outubro -> ou.tu.bro	outubro -> ou'tubro	ou'tubro -> [o'tubru]
ouvir -> ou.vir	ouvir -> ou'vir	ou'vir -> [o'vir]
ovo -> o.vo	ovo -> 'ovo	'ovo -> ['ovu]
padre -> pa.dre	padre -> 'padre	'padre -> ['padri]
pagar -> pa.gar	pagar -> pa'gar	pa'gar -> [pe'ɣar]
pai -> pai	pai -> 'pai	'pai -> ['paj]

país -> pa.ís	país -> pa'ís	pa'ís -> [pe'i]
palavra -> pa.la.vra	palavra -> pa'lavra	pa'lavra -> [pe'lavre]
panela -> pa.ne.la	panela -> pa'nela	pa'nela -> [pe'nele]
pão -> pão	pão -> 'pão	'pão -> ['pẽw]
papel -> pa.pel	papel -> pa'pel	pa'pel -> [pe'pɛt]
para -> pa.ra		para -> [pere]
parar -> pa.rar	parar -> pa'rar	pa'rar -> [pe'rar]
parecer -> pa.re.cer	parecer -> pare'cer	pare'cer -> [peɾ'ser]
parede -> pa.re.de	parede -> pa'rede	pa'rede -> [pe'reði]
parte -> par.te	parte -> 'parte	'parte -> ['parti]
partir -> par.tir	partir -> par'tir	par'tir -> [peɾ'tir]
passado -> pas.sa.do	passado -> pa'ssado	pa'ssado -> [pe'saðu]
passar -> pas.sar	passar -> pa'ssar	pa'ssar -> [pe'sar]
pássaro -> pás.sa.ro	pássaro -> 'pássaro	'pássaro -> ['paseru]
passear -> pas.se.ar	passear -> passe'ar	passe'ar -> [pesi'ar]
passo -> pas.so	passo -> 'passo	'passo -> ['pasu]
pata -> pa.ta	pata -> 'pata	'pata -> ['pate]
pau -> pau	pau -> 'pau	'pau -> ['paw]
paz -> paz	paz -> 'paz	'paz -> ['pa]
pé -> pé	pé -> 'pé	'pé -> ['pɛ]
pedir -> pe.dir	pedir -> pe'dir	pe'dir -> [pi'dir]
pedra -> pe.dra	pedra -> 'pedra	'pedra -> ['pedre]
pegar -> pe.gar	pegar -> pe'gar	pe'gar -> [pi'ɣar]
peixe -> pei.xe	peixe -> 'peixe	'peixe -> ['pejfi]
pele -> pe.le	pele -> 'pele	'pele -> ['peli]
pena -> pe.na	pena -> 'pena	'pena -> ['pene]
pensar -> pen.sar	pensar -> pen'sar	pen'sar -> [pẽ'sar]
pequeno -> pe.que.no	pequeno -> pe'queno	pe'queno -> [pi'kenu]
perceber -> per.ce.ber	perceber -> perce'ber	perce'ber -> [piɾi'βer]
perder -> per.der	perder -> per'der	per'der -> [piɾ'der]
pergunta -> per.gun.ta	pergunta -> per'gunta	per'gunta -> [piɾ'gũte]
perguntar -> per.gun.tar	perguntar -> pergun'tar	pergun'tar -> [piɾgũ'tar]
perna -> per.na	perna -> 'perna	'perna -> ['perne]
peessoa -> pes.so.a	peessoa -> pe'ssoa	pe'ssoa -> [pi'soe]
planeta -> pla.ne.ta	planeta -> pla'neta	pla'neta -> [ple'nete]
pó -> pó	pó -> 'pó	'pó -> ['pɔ]
pobre -> po.bre	pobre -> 'pobre	'pobre -> ['pɔbri]
poder -> po.der	poder -> po'der	po'der -> [pu'der]
pois -> pois	pois -> 'pois	'pois -> ['poj]

polícia -> po.lí.ci.a	polícia -> po'lícia	po'lícia -> [pu'lisie]
ponto -> pon.to	ponto -> 'ponto	'ponto -> ['põtu]
por -> por		por -> [pur]
pôr -> pôr	pôr -> 'pôr	'pôr -> ['por]
porque -> por.que		porque -> [purki]
porquê -> por.quê	porquê -> por'quê	por'quê -> [pur'ke]
porta -> por.ta	porta -> 'porta	'porta -> ['põrte]
portanto -> por.tan.to	portanto -> por'tanto	por'tanto -> [pur'tẽtu]
português -> por.tu.guês	português -> portu'guês	portu'guês -> [purtu'ye]
possível -> pos.sí.vel	possível -> po'ssível	po'ssível -> [pu'sivẽt]
pouco -> pou.co	pouco -> 'pouco	'pouco -> ['poku]
povo -> po.vo	povo -> 'povo	'povo -> ['povu]
praia -> prai.a	praia -> 'praia	'praia -> ['praje]
prato -> pra.to	prato -> 'prato	'prato -> ['pratu]
precisar -> pre.ci.sar	precisar -> preci'sar	preci'sar -> [pri'si'zar]
preciso -> pre.ci.so	preciso -> pre'ciso	pre'ciso -> [pri'sizu]
preço -> pre.ço	preço -> 'preço	'preço -> ['presu]
preparar -> pre.pa.rar	preparar -> prepa'rar	prepa'rar -> [pripe'rar]
presente -> pre.sen.te	presente -> pre'sente	pre'sente -> [pri'zẽti]
preto -> pre.to	preto -> 'preto	'preto -> ['pretu]
primavera -> pri.ma.ve.ra	primavera -> prima'vera	prima'vera -> [prime'vere]
primeiro -> pri.mei.ro	primeiro -> pri'meiro	pri'meiro -> [pri'mejru]
primo -> pri.mo	primo -> 'primo	'primo -> ['primu]
principal -> prin.ci.pal	principal -> princi'pal	princi'pal -> [pri'si'pat]
problema -> pro.ble.ma	problema -> pro'blema	pro'blema -> [pru'bleme]
procurar -> pro.cu.rar	procurar -> procu'rar	procu'rar -> [prõku'rar]
produto -> pro.du.to	produto -> pro'duto	pro'duto -> [pru'ðutu]
professor -> pro.fes.sor	professor -> profe'ssor	profe'ssor -> [prufi'sor]
programa -> pro.gra.ma	programa -> pro'grama	pro'grama -> [pru'greme]
pronto -> pron.to	pronto -> 'pronto	'pronto -> ['prõtu]
próprio -> pró.pri.o	próprio -> 'próprio	'próprio -> ['prõpriu]
puxar -> pu.xar	puxar -> pu'xar	pu'xar -> [pu'far]
qual -> qual	qual -> 'qual	'qual -> ['kʷat]
qualquer -> qual.quer	qualquer -> qual'quer	qual'quer -> [kʷat'kẽr]
quando -> quan.do	quando -> 'quando	'quando -> ['kʷẽdu]
quanto -> quan.to	quanto -> 'quanto	'quanto -> ['kʷẽtu]
quarto -> quar.to	quarto -> 'quarto	'quarto -> ['kʷartu]
quase -> qua.se	quase -> 'quase	'quase -> ['kʷazi]
quatro -> qua.tro	quatro -> 'quatro	'quatro -> ['kʷatru]
que -> que		que -> [ki]

queijo -> quei.jo	queijo -> 'queijo	'queijo -> ['kejʒu]
quem -> quem	quem -> 'quem	'quem -> ['kẽ]
quente -> quen.te	quente -> 'quente	'quente -> ['kẽti]
querer -> que.rer	querer -> que'rer	que'rer -> [ki'rer]
quinta -> quin.ta	quinta -> 'quinta	'quinta -> ['kĩte]
rádio -> rá.di.o	rádio -> 'rádio	'rádio -> ['rađiu]
ramo -> ra.mo	ramo -> 'ramo	'ramo -> ['remu]
rapariga -> ra.pa.ri.ga	rapariga -> rapa'riga	rapa'riga -> [repe'riye]
rapaz -> ra.paz	rapaz -> ra'paz	ra'paz -> [re'pa]
rapidamente -> ra.pi.da.men.te	rapidamente -> rapida'mente	rapida'mente -> [repiðe'mẽti]
raposa -> ra.po.sa	raposa -> ra'posa	ra'posa -> [re'poze]
razão -> ra.zão	razão -> ra'zão	ra'zão -> [re'zẽw]
receber -> re.ce.ber	receber -> rece'ber	rece'ber -> [risi'βer]
receita -> re.cei.ta	receita -> re'ceita	re'ceita -> [ri'sejte]
região -> re.gião	região -> regi'ão	regi'ão -> [riʒi'ẽw]
rei -> rei	rei -> 'rei	'rei -> ['rej]
repetir -> re.pe.tir	repetir -> repe'tir	repe'tir -> [ripi'tir]
respeito -> res.pei.to	respeito -> res'peito	res'peito -> [ri'pejtu]
respirar -> res.pi.rar	respirar -> respi'rar	respi'rar -> [ri'pi'rar]
responder -> res.pon.der	responder -> respon'der	respon'der -> [ri'põ'der]
resposta -> res.pos.ta	resposta -> res'posta	res'posta -> [ri'põfte]
restaurante -> res.tau.ran.te	restaurante -> restau'rante	restau'rante -> [ri'taw'rẽti]
revista -> re.vis.ta	revista -> re'vista	re'vista -> [ri'viʃte]
rico -> ri.co	rico -> 'rico	'rico -> ['riku]
rio -> ri.o	rio -> 'rio	'rio -> ['riu]
rir -> rir	rir -> 'rir	'rir -> ['rir]
roupa -> rou.pa	roupa -> 'roupa	'roupa -> ['rope]
rua -> ru.a	rua -> 'rua	'rua -> ['rue]
sábado -> sá.ba.do	sábado -> 'sábado	'sábado -> ['saβeðu]
saber -> sa.ber	saber -> sa'ber	sa'ber -> [se'βer]
sabor -> sa.bor	sabor -> sa'bor	sa'bor -> [se'βor]
saco -> sa.co	saco -> 'saco	'saco -> ['saku]
sair -> sa.ir	sair -> sa'ir	sa'ir -> [se'ir]
sal -> sal	sal -> 'sal	'sal -> ['sa]
sala -> sa.la	sala -> 'sala	'sala -> ['sale]
saltar -> sal.tar	saltar -> sal'tar	sal'tar -> [sa'tar]
santo -> san.to	santo -> 'santo	'santo -> ['sẽtu]
sapato -> sa.pa.to	sapato -> sa'pato	sa'pato -> [se'patu]
sardinha -> sar.di.nha	sardinha -> sar'dinha	sar'dinha -> [ser'dine]

saúde -> sa.ú.de	saúde -> sa'úde	sa'úde -> [se'uðĩ]
se -> se		se -> [sɨ]
secar -> se.car	secar -> se'car	se'car -> [sɨ'kar]
seco -> se.co	seco -> 'seco	'seco -> ['seku]
seguir -> se.guir	seguir -> se'guir	se'guir -> [sɨ'ɣir]
segundo -> se.gun.do	segundo -> se'gundo	se'gundo -> [sɨ'ɣũdu]
seis -> seis	seis -> 'seis	'seis -> ['sejʃ]
sem -> sem		sem -> [sẽ]
semana -> se.ma.na	semana -> se'mana	se'mana -> [sɨ'mene]
semente -> se.men.te	semente -> se'mente	se'mente -> [sɨ'mẽti]
sempre -> sem.pre	sempre -> 'sempre	'sempre -> ['sẽpri]
senhor -> se.nhor	senhor -> se'nhor	se'nhor -> [sɨ'nɔr]
sentar -> sen.tar	sentar -> sen'tar	sen'tar -> [sẽ'tar]
sentir -> sen.tir	sentir -> sen'tir	sen'tir -> [sẽ'tir]
ser -> ser	ser -> 'ser	'ser -> ['ser]
serviço -> ser.vi.ço	serviço -> ser'viço	ser'viço -> [sɨ'visu]
servir -> ser.vir	servir -> ser'vir	ser'vir -> [sɨ'vir]
sete -> se.te	sete -> 'sete	'sete -> ['seti]
setembro -> se.tem.bro	setembro -> se'tembro	se'tembro -> [sɨ'tẽbru]
sim -> sim	sim -> 'sim	'sim -> ['sĩ]
simpático -> sim.pá.ti.co	simpático -> sim'pático	sim'pático -> [sĩ'patiku]
simples -> sim.ples	simples -> 'simples	'simples -> ['sĩplɨ]
sítio -> sí.ti.o	sítio -> 'sítio	'sítio -> ['sitiu]
só -> só	só -> 'só	'só -> ['so]
sobrar -> so.brar	sobrar -> so'brar	so'brar -> [su'brar]
sobre -> so.bre	sobre -> 'sobre	'sobre -> ['sobri]
sobretudo -> so.bre.tu.do	sobretudo -> sobre'tudo	sobre'tudo -> [sobri'tuðu]
sofrer -> so.frer	sofrer -> so'frer	so'frer -> [su'frer]
sol -> sol	sol -> 'sol	'sol -> ['soɫ]
soldado -> sol.da.do	soldado -> sol'dado	sol'dado -> [soɫ'daðu]
som -> som	som -> 'som	'som -> ['sõ]
sonhar -> so.nhar	sonhar -> so'nhar	so'nhar -> [su'nar]
sonho -> so.nho	sonho -> 'sonho	'sonho -> ['soɲu]
sopa -> so.pa	sopa -> 'sopa	'sopa -> ['sope]
sorrir -> sor.rir	sorrir -> so'rrir	so'rrir -> [su'rir]
sorte -> sor.te	sorte -> 'sorte	'sorte -> ['sorti]
sozinho -> so.zi.nho	sozinho -> so'zinho	so'zinho -> [so'zɨnu]
subir -> su.bir	subir -> su'bir	su'bir -> [su'βir]
sul -> sul	sul -> 'sul	'sul -> ['suɫ]
sumo -> su.mo	sumo -> 'sumo	'sumo -> ['sumu]

tal -> tal	tal -> 'tal	'tal -> ['taɫ]
talvez -> tal.vez	talvez -> tal'vez	tal'vez -> [taɫ'veʃ]
também -> tam.bém	também -> tam'bém	tam'bém -> [tẽ'bẽ]
tanto -> tan.to	tanto -> 'tanto	'tanto -> ['tẽtu]
tanto -> tan.to	tanto -> 'tanto	'tanto -> ['tẽtu]
tão -> tão	tão -> 'tão	'tão -> ['tẽw]
tapete -> ta.pe.te	tapete -> ta'pete	ta'pete -> [te'peti]
tarde -> tar.de	tarde -> 'tarde	'tarde -> ['tardi]
tarde -> tar.de	tarde -> 'tarde	'tarde -> ['tardi]
televisão -> te.le.vi.são	televisão -> televi'são	televi'são -> [tilivi'zẽw]
temperatura -> tem.pe.ra.tu.ra	temperatura -> tempera'tura	tempera'tura -> [tẽpire'ture]
tempo -> tem.po	tempo -> 'tempo	'tempo -> ['tẽpu]
tentar -> ten.tar	tentar -> ten'tar	ten'tar -> [tẽ'tar]
ter -> ter	ter -> 'ter	'ter -> ['ter]
terceiro -> ter.cei.ro	terceiro -> ter'ceiro	ter'ceiro -> [tĩr'sejru]
terra -> ter.ra	terra -> 'terra	'terra -> ['tẽrẽ]
texto -> tex.to	texto -> 'texto	texto -> ['te(j)tu]
tio -> ti.o	tio -> 'tio	'tio -> ['tiu]
tipo -> ti.po	tipo -> 'tipo	'tipo -> ['tipu]
tirar -> ti.rar	tirar -> ti'rar	ti'rar -> [ti'rar]
tocar -> to.car	tocar -> to'car	to'car -> [tu'kar]
todo -> to.do	todo -> 'todo	'todo -> ['toðu]
tomar -> to.mar	tomar -> to'mar	to'mar -> [tu'mar]
tornar -> tor.nar	tornar -> tor'nar	tor'nar -> [tur'nar]
trabalhar -> tra.ba.lhar	trabalhar -> traba'lhar	traba'lhar -> [treβe'λar]
trabalho -> tra.ba.lho	trabalho -> tra'balho	tra'balho -> [tre'βaɫu]
tratar -> tra.tar	tratar -> tra'tar	tra'tar -> [tre'tar]
trazer -> tra.zer	trazer -> tra'zer	tra'zer -> [tre'zer]
três -> três	três -> 'três	'três -> ['treʃ]
triste -> tris.te	triste -> 'triste	'triste -> ['trifti]
tronco -> tron.co	tronco -> 'tronco	'tronco -> ['trõku]
tudo -> tu.do	tudo -> 'tudo	'tudo -> ['tuðu]
turma -> tur.ma	turma -> 'turma	'turma -> ['turme]
último -> úl.ti.mo	último -> 'último	'último -> ['uɫtimu]
um -> um		um -> [ũ]
único -> ú.ni.co	único -> 'único	'único -> ['uniku]
urso -> ur.so	urso -> 'urso	'urso -> ['ursu]
usar -> u.sar	usar -> u'sar	u'sar -> [u'zar]
uva -> u.va	uva -> 'uva	'uva -> ['uve]

valor -> va.lor	valor -> va'lor	va'lor -> [ve'lor]
velho -> ve.lho	velho -> 'velho	'velho -> ['veʎu]
venda -> ven.da	venda -> 'venda	'venda -> ['vẽde]
vendedor -> ven.de.dor	vendedor -> vende'dor	vende'dor -> [vẽdi'ðor]
vender -> ven.der	vender -> ven'der	ven'der -> [vẽder]
vento -> ven.to	vento -> 'vento	'vento -> ['vẽtu]
ver -> ver	ver -> 'ver	'ver -> ['ver]
verão -> ve.rão	verão -> ve'rão	ve'rão -> [vi'rẽw]
verdade -> ver.da.de	verdade -> ver'dade	ver'dade -> [vir'daði]
vermelho -> ver.me.lho	vermelho -> ver'melho	[vir'me(j)ʎu] / [vir'meʎu]
vestir -> ves.tir	vestir -> ves'tir	ves'tir -> [viʃ'tir]
vez -> vez	vez -> 'vez	'vez -> ['veʃ]
viagem -> vi.a.gem	viagem -> via'gem	vi'agem -> [vi'aʒẽ]
vida -> vi.da	vida -> 'vida	'vida -> ['viðe]
vídeo -> ví.de.o	vídeo -> 'vídeo	'vídeo -> ['viði.u]
vinho -> vi.nho	vinho -> 'vinho	'vinho -> ['viɲu]
vir -> vir	vir -> 'vir	'vir -> ['vir]
visitar -> vi.si.tar	visitar -> visi'tar	visi'tar -> [vizi'tar]
viver -> vi.ver	viver -> vi'ver	vi'ver -> [vi'ver]
vivo -> vi.vo	vivo -> 'vivo	'vivo -> ['vivu]
vizinho -> vi.zi.nho	vizinho -> vi'zinho	vi'zinho -> [vi'ziɲu]
voar -> vo.ar	voar -> vo'ar	vo'ar -> [vu'ar]
volta -> vol.ta	volta -> 'volta	'volta -> ['vɔʎte]
voltar -> vol.tar	voltar -> vol'tar	vol'tar -> [voʎ'tar]
vontade -> von.ta.de	vontade -> von'tade	von'tade -> [võ'taði]
voz -> voz	voz -> 'voz	'voz -> ['vɔʃ]
zona -> zo.na	zona -> 'zona	'zona -> ['zone]

Anexo 7 – Critérios da translineação no guia de uso

Critério	Exemplo
Separam-se as vogais que não formam ditongo.	frio → fri.o sair → sa.ir
Separam-se as consoantes dobradas <i>rr</i> e <i>ss</i> .	carro → car.ro passar → pas.sar
Não se separam os ditongos.	baixo → bai.xo oito → oi.to
Não se separam os grupos consonânticos que iniciam sílaba.	prato → pra.to sobrar → so.brar
Não se separam os dígrafos <i>ch</i> , <i>lh</i> , <i>nh</i> .	fechar → fe.char velho → ve.lho sonhar → so.nhar
Não se separam as sequências <i>gu</i> nem <i>qu</i> .	água → á.gua seguir → se.guir qual → qual aqui → a.qui
Nos compostos, se o hífen coincidir com o fim da linha, deve repetir-se na linha seguinte.	boa-tarde → bo.a-tar.de bom-dia → bom-di.a

Anexo 8 – Critérios da transcrição fonética no guia de uso

Critério	Exemplo
O acento principal da palavra é marcado antes da sílaba tónica através do símbolo '.	chocolate [ʃuku'latɨ]
A transcrição da forma feminina é incluída após a masculina, separada por uma vírgula.	alto, alta [ˈaɫtu], [ˈaɫte]
As transcrições das variantes ortográficas surgem separadas por um espaço em branco.	cenoura ou cenoira [sɨ'noɾe] ou [sɨ'nojɾe]
Para a vogal <o> em início de palavra, incluem-se as variantes [ɔ] e [o], exceto em palavras cuja vogal inicial é [ɔ] para a generalidade dos falantes (e.g., olá, hospital, hotel).	origem [ɔ'ɾizẽ] / [o'ɾizẽ]
Para a vogal <e> em início de palavra, incluem-se as variantes [e] e [i], exceto em palavras cuja vogal inicial é [ɛ] para a generalidade dos falantes (e.g., erva, herbácea, etc.).	equipa [e'kipe] / [i'kipe]
Os ditongos crescentes, são excluídos, optando-se pela transcrição de sequências vocálicas heterossilábicas.	área [ˈarie]
Nos ditongos nasais, a nasalidade é marcada na vogal e na glide.	mão [ˈmẽw̃]
As oclusivas fricativizadas [β], [ð] e [ɣ] são incluídas para as consoantes , <d> e <g> intervocálicas.	trabalhar [treβe'ɫar]
A velarização de /l/ em final de sílaba é incluída.	sal [ˈsaɫ]

Anexo 9 – Inventário de caracteres consonânticos no guia de uso

Grafema(s)	Fone[ma](s)	Exemplo
<p>	[p]	p alavra [pe'lavre]
	[b]	b arulho [be'ruʎu]
<a,e,i,o,u b a,e,i,o,u>	[β]	tr ab alho [tre'βaʎu]
<t>	[t]	tr ê s ['treʃ]
<d>	[d]	d omingo [du'mĩgu]
<a,e,i,o,u d a,e,i,o,u>	[ð]	n ada ['naðe]
<c ^{a,o,u} >	[k]	c ozinha [ku'zine]
<qu ^{e,i} >		q ueijo ['kejʒu]
<qu ^{a,o,u} >	[kʷ]	q uarto ['kʷartu]
<g ^{a,e,o} >	[g]	g ostar [guʃ'tar]
<gu ^{e,i} >		algu ém [aʔ'gẽ]
<a,e,i,o,u g a,e,i,o,u>	[ɣ]	l igar [li'ɣar]
<gu ^{a,e,o} >	[gʷ]	g uardar [gʷer'dar]
<m>	[m]	m acaco [me'kaku]
<n>	[n]	n enhum [nɨ'ɲũ]

<nh>	[n]	vinho [ˈviɲu]
<r>	[ʀ]	rei [ˈʀej]
<rr>		sorrir [suˈʀir]
<r>	[r]	marca [ˈmarke]
<f>	[f]	futuro [fuˈturu]
<v>	[v]	verão [viˈrẽw̃]
<s>	[s]	sapato [seˈpatu]
<ss>		passo [ˈpasu]
<c ^{e,i} >		fácil [ˈfasiɫ]
<ç>		braço [ˈbrasu]
<x>		sintaxe [sĩˈtasi]
<x>	[ks̃]	táxi [ˈtaksi]
<a,e,i,o,u,ç a,e,i,o,u>	[z]	usar [uˈzar]
<z>		cozinha [kuˈziɲe]
<x>		exemplo [eˈzẽplu] / [iˈzẽplu]
<s>	[ʃ]	seis [ˈsejʃ]
<z>		rapaz [reˈpaʃ]
<ch>		chuva

		['juve]
<x>		puxar [pu'ʃar]
<g ^{e,i} >	[ʒ]	gente [ʒẽti]
<j>		jantar [ʒẽ'tar]
<ɫ>	[l]	local [lu'kaɫ]
	[ɫ]	local [lu'kaɫ]
<lh>	[ʎ]	alho [aʎu]

Anexo 10 – Inventário de caracteres vocálicos orais no guia de USO

Grafema	Fone[ma]	Exemplo
<á>	[a]	pássaro [ˈpasɐɾu]
<a>		casa [ˈkaze]
<â>	[e]	câmara [ˈkɐmɐɾɐ]
<a>		casa [ˈkaze]
<é>	[ɛ]	pé [ˈpɛ]
<ê>	[e]	três [ˈtɾɛʃ]
<e>	[e]	letra [ˈlɛtɾɐ]
	[ɛ]	janela [ʒɐˈnɛlɐ]
	[ɨ]	carne [ˈkarnɨ]
	[i]	área [ˈaɾiɐ]
	[ɐ(j)]	vermelho [vɨɾˈmɐ(j)ʎu]
<í>	[i]	polícia [puˈlisɪɐ]
<i>		bonito [buˈnɪtu]
<ó>	[ɔ]	pó [ˈpɔ]

<ô>	[o]	avô [e'vo]
<o>	[o]	amor [e'mor]
	[ɔ]	bola ['bole]
	[u]	carro ['karu]
<ú>	[u]	música ['muzike]
<u>		azul [e'zuɫ]

Anexo 11 – Inventário de caracteres vocálicos nasais no guia de uso

Grafema(s)	Fone[ma](s)	Exemplo
<am>	[ẽ]	campo [ˈkẽpu]
<an>		branco [ˈbrẽku]
<âm>		âmbar [ˈẽbar]
<ân>		âncora [ˈẽkure]
<ã>		mamã [meˈmẽ]
	[ẽ]	tempo [ˈtẽpu]
<en>		contente [kõˈtẽti]
<ên>		bênção [ˈbẽsẽw]
<im>	[ĩ]	fim [ˈfi]
<in>		quinta [ˈkĩte]
<om>	[õ]	compra [ˈkõpre]
<on>		tronco [ˈtrõku]
<um>	[ũ]	algum [aɫˈgũ]
<un>		perguntar [pĩrgũˈtar]

Anexo 12 – Inventário de ditongos orais no guia de uso

Grafema(s)	Fone[ma](s)	Exemplo
<ai>	[aj]	baixo [ˈbajʃu]
<ei>	[ej]	queijo [ˈkejʒu]
<éi>	[ɛj]	papéis [peˈpejʃ]
<oi>	[oj]	coisa [ˈkojzɐ]
	[ɔj]	comboio [kõˈboju]
<ói>	[ɔj]	herói [eˈɔj] / [iˈɔj]
<au>	[aw]	pau [ˈpaw]
	[ɛw]	saudade [sawˈðaðɨ]
<eu>	[ew]	museu [muˈzew]
<éu>	[ɛw]	céu [ˈsɛw]
<iu>	[iw]	piu [ˈpiw]

Anexo 13 – Inventário de ditongos nasais

Grafema(s)	Fone[ma](s)	Exemplo
<ãe>	[ẽ̃]	mãe [ˈmẽ̃]
		viagem [viˈaʒẽ̃]
<õe>	[õ̃]	leões [liˈõ̃]
<ui>	[ũ̃]	muito [ˈmũ̃tu]
<ão>	[ẽ̃w̃]	verão [viˈrẽ̃w̃]