

Um Transcritor Fonético Pedagógico para as Variações Vocálicas do Português Brasileiro

Computer on the Beach

Wagner Ferreira Lima
Universidade Estadual de Londrina
Londrina, Paraná, Brasil
wflima@uel.br

Cinthyan R. S. C. de Barbosa
Universidade Estadual de Londrina
Londrina, Paraná, Brasil
cinthyan@uel.br

Dircel A. Kailer
Universidade Estadual de Londrina
Londrina, Paraná, Brasil
dikailer@uel.br

ABSTRACT

This paper addresses the challenge of teaching Brazilian Portuguese (BP) phonetics, a difficulty compounded by the lack of transcription tools that adequately account for dialectal variations. The proposed solution was to develop an automatic phonetic transcriber for pedagogical purposes. To validate it, three conditions for effectiveness were established: to symbolically represent speech sounds, to show the phonetic environment of vowels, and to indicate the corresponding dialect. These conditions were met by the results, which had a 97.29% accuracy. The study concludes that the transcriber is more suitable for teaching than its online counterparts, as it provides more detailed transcriptions that are more faithful to specific BP pronunciations, thereby surpassing existing generalist tools.

KEYWORDS

Phonetic transcription, Phonetics teaching, Brazilian Portuguese.

1 INTRODUÇÃO

A linguagem verbal consiste em um fenômeno dinâmico, variável e social. Isso impõe uma condição a todos quanto que se dedicam a seu estudo: conhecer as diferentes variedades linguísticas faladas por um povo. No caso do Brasil, existem muitas variedades que caracterizam o Português Brasileiro (PB). Nesta apresentação, o foco é a pronúncia típica das variedades linguísticas de origem geográfica, ou simplesmente dos dialetos. O termo “dialeto” implica variações de um código comum. Nesse sentido, pode-se falar, no Brasil, de dialeto carioca, sulista, nordestino etc. [1].

Na academia há muitas disciplinas que lidam com a variação linguística. Aqui destaca-se a fonética articulatória, que se ocupa de revelar os aspectos articulatórios do plano de expressão do idioma, ou seja, as características e distribuição dos sons da fala.

Esses sons são conhecidos também por fones ou segmentos sonoros (consoantes, vogais e semivogais).

A fonética articulatória descreve, por exemplo, a variação da pronúncia sibilante em final de sílaba em relação ao dialeto em que ela é encontrada. Assim, ela mostra que o sibilante encontrado em *casas* é alveopalatal [ʃ]¹, [ˈka.zaʃ], no dialeto carioca; e alveolar [s], [ˈka.zas], em outros dialetos [2]. Com isso, a fonética articulatória captura os nuances sonoros e explicita as diferentes marcas dialetais.

A fonética articulatória é, dessa maneira, relevante aos estudantes de Letras e afins, uma vez que ela mostra como a linguagem é perpassada por fatores sociais. Contudo, ele requer práticas de análise tão técnicas e especializadas quanto descrever, classificar e transcrever os sons da fala [2]. Isso coloca grandes desafios a esses estudantes, futuros professores de Português.

Questiona-se, diante disso, se as tecnologias digitais poderiam auxiliar esses estudantes nas práticas de análise fonética. Assume-se que Agentes de Inteligência Artificial (AIA) podem cumprir esse papel, porém desde que suplementados, eles mesmos, com ferramentas externas que realizem, senão todas as práticas, pelo menos a de transcrição fonética, que consiste em converter segmentos sonoros percebidos pelos ouvidos em símbolos fonéticos correspondentes [3]. Em geral, adota-se, por razões de convenção e uniformidade, os símbolos do Alfabeto Fonético Internacional (AFI) [2].

Trabalhos há os que oferecem transcrições fonéticas prontas, mas esses são inadequados ao propósito pedagógico de representar as pronúncias típicas dos dialetos. Como se verá adiante, trata-se de aplicativos online com forte apelo comercial, de modo a priorizar transcrições generalistas sem compromisso com as peculiaridades dialetais [4, 5, 6, 7]. Também por isso a existência da referida ferramenta se faz necessária e urgente.

Em resposta a isso, uma função computacional foi criada para converter palavras escritas em representações fonéticas, de acordo com o dialeto especificado. Essa função é um transcritor fonético

¹ “Alveopalatal” e “alveolar” são gestos articulatórios produzidos com a parte anterior da língua em relação a pontos distintos da abóboda da cavidade oral.

automático – daqui por diante chamado simplesmente de “transcritor”. Por motivo de agilidade e controle, adotou-se a abordagem computacional clássica de declaração de regras, ao invés de Aprendizagem de Máquina baseada em Redes Neurais Artificiais.

O objetivo geral deste texto é expor a construção desse transcritor. Contudo, um transcritor como esse coloca o desafio de superar o problema de restrição de regras, ou seja, o fato de que essas declaradas são, por razões óbvias, incapazes de capturar a ampla gama de variações. A fim de mitigar esse problema, o escopo do transcritor foi readequado. A motivação da proposta sendo pedagógica, o transcritor foi construído então para lidar com uma amostra específica de pronúncias que tivessem relevância para o ensino. Optou-se, nesse sentido, por um *corpus* com variações vocálicas que representam alguns processos inerentes ao vocalismo do PB, a saber, a abertura, a redução e a nasalidade das vogais; e também indicam alguns dialetos do Brasil, vale dizer, o caipira, o carioca, o nordestino e o sulista.

Desse modo, o objetivo específico desta apresentação é elucidar como o transcritor representa essas pronúncias e, por extensão, contribui com o ensino. Assim sendo, essa ferramenta será validada não com respeito aos aplicativos correlatos, mas sim com as condições de efetividade que serão criadas para isso, as quais possuem um caráter pedagógico.

Esta exposição está organizada como se segue. Em “Conhecimento fonético”, apresentam-se as condições que o transcritor deve satisfazer (condições de efetividade) para que seja considerado adequado ao referido propósito. Já em “Vocalismo do PB”, é apresentada a teoria fonética sobre as vogais que orienta a proposta do transcritor. Em “Trabalhos correlatos”, uma breve avaliação é realizada sobre os aplicativos *online* de transcrição fonética do Português. “Metodologia” apresenta os passos de construção da ferramenta. “Resultados” exhibe uma estatística intuitiva das transcrições geradas pelo transcritor. Em “Discussão geral”, as condições de efetividade da ferramenta são avaliadas à luz da proposta. E, finalmente, “Considerações gerais” apresenta os limites da ferramenta criada e os encaminhamentos futuros.

2 CONHECIMENTO FONÉTICO

A implementação do referido transcritor coloca a questão fundamental sobre como ele pode favorecer o ensino da fonética e se, portanto, ele é adequado à prática analítica proposta. Essa questão só pode ser respondida à medida que se considera o conhecimento exigido para a correta consecução de análise fonética. Dado o seu propósito pedagógico, esse conhecimento não precisa ser exaustivo. O que se busca com ele é esclarecer os passos

analíticos que subjazem à transcrição fonética do *corpus* e que podem ser replicados em situações inéditas.

Assim sendo, para além das métricas de acurácia², um transcritor será adequado à prática proposta desde que suas saídas atendam a pelos menos três condições de efetividade: (a) representar simbolicamente os sons da fala, de modo a permitir a inferência de suas características articulatórias; (b) mostrar o ambiente fonético do segmento de interesse, tal que se possa deduzir a influência das adjacências sobre ele; (c) indicar o dialeto a que pertence a pronúncia transcrita.

A condição (a) está relacionada à capacidade de definir e classificar os fones. Por exemplo, dizer que o “é” encontrado em *pé*, [pe], é uma vogal “médio-baixa, anterior, não-arredondada” (cf. seção 3), e que ele pertence à classe das vogais anteriores, como “ê” e “i” em *ipê*, [i.ˈpe], e *ti*, [ˈti], respectivamente, é uma expressão desse conhecimento [2]. Isso é representado pelos símbolos fonéticos [ɛ], [e] e [i].

Todavia, embora essa capacidade seja necessária, ela não é suficiente. Porque variação sonora é condicionada pelo contexto, esse conhecimento deve abrigar também a capacidade de descrever como os fones se distribuem no interior da palavra. Essa é uma exigência da condição (b). Por exemplo, a vogal ortográfica “a” pode ser pronunciada como [a] em *pá*, [ˈpa], e como [ə] em *bola*, [ˈbo.lə]. A variação nesse caso se justifica pelo ambiente fonético (cf. seção 3), ou seja, o grau de tonicidade da sílaba em que o fone ocorre. Em sílaba tônica, o “a” é forte e deve ser indicado pelo símbolo [a]; enquanto em sílaba postônica final, ele é fraco e como tal deve ser simbolizado por [ə].

Como se verá adiante, esse fenômeno de variação é conhecido por redução da vogal, pois o [ə] é pronunciado com a boca mais fechada que a vogal baixa [a] [2]. Esse caso de redução é em geral independente de região. Contudo, casos há em que a variação é específica de dialeto. A abertura de vogais ilustra isso, pois diz respeito ao processo pelo qual uma vogal fechada se torna aberta em virtude do dialeto ([e] > [ɛ]; [o] > [ɔ]³). A palavra *colega*, por exemplo, é pronunciada em geral como [ko.ˈle.gə]; mas particularmente como [ko.ˈle.gə], pronúncia aberta, no dialeto nordestino. Essas transcrições atendem à exigência de (c), pois sinalizam o dialeto ao qual as pronúncias em questão pertencem.

Em suma, transcrever foneticamente uma expressão é representar os sons da fala em sequência de símbolos fonéticos. Essa representação envolve mais que a mera conversão de fones identificados para os símbolos correspondentes, senão também um entendimento sobre o condicionamento contextual deles (quer esse seja interno ou externo à linguagem); conforme se viu nas três condições definidas.

² Definidas em termos da capacidade do transcritor de gerar dados adequados para o ensino da fonética do PB, tais condições devem ser vistas como distintas das métricas de validação existentes, como a acurácia; as quais não deixam de ser, elas mesmas, condições de efetividade.

³ O símbolo “>” significa neste contexto “passa para...” ou “transforma-se em...”.

3 VOCALISMO DO PB

O vocalismo do PB coloca algumas dificuldades, porquanto o emprego das vogais não é uniforme. Há alguns processos de variação que podem ser encontrados quando o comportamento das vogais é observado [8]. Para efeito de ensino, destacam-se especificamente estes três: a redução, a abertura e a nasalidade [2]. O entendimento de tais processos requer o conhecimento tanto dos sons em si, quanto do contexto em que eles ocorrem. Esse, por seu turno, pode envolver fatores internos e/ou externos. No primeiro caso, a pronúncia é influenciada pelo ambiente fonético, ou seja, os segmentos adjacentes ao fone de interesse; no segundo, pelos chamados fatores sociolinguísticos, como é o caso do dialeto.

As teorias sobre o vocalismo não são consensuais, tal que a abordagem do tema é fruto de uma escolha teórica. Um transcritor fonético automático deve, nesse sentido, apoiar-se em uma teoria alinhada com os presentes objetivos. Assim, a teoria selecionada é aquela encontrada em Silva [2], cujas razões se esclarecerão adiante.

A fonética levada a efeito por essa pesquisadora trata da descrição, classificação e transcrição de padrões de pronúncia do Português. Os fones considerados são descritos e classificados em termos de suas características articulatórias. As vogais são sons orais produzidos sem obstrução da passagem da corrente de ar pelo trato bucal. A produção delas supõe apenas uma constrição desse trato, do que decorrem os seguintes parâmetros de definição: Parâmetro “altura da língua”, pelo qual se observa se essa se encontra “baixa”, “médio-baixa”, “médio alta” ou “alta” na boca; parâmetro “posição da língua”, por meio de que se nota se ela está na região “anterior”, “central” ou “posterior” da cavidade oral; e, por fim, “arredondamento dos lábios”, se os lábios estão “arredondados” ou “não-arredondados”.

	Anterior		Central		Posterior	
	Arred.	Não-arred.	Arred.	Não-arred.	Arred.	Não-arred.
Alta		i				u
Médio-alta		e				o
Médio-baixa		ɛ				ɔ
Baixa				a		

Figura 1. Quadro das vogais tônicas orais do PB. Adaptado de Silva (2002)

A título de exemplificação, a vogal tônica em *pá* [a] é descrita como “baixa, central” (Figura 1). Já as vogais tônicas encontradas em *pé* [ɛ], *ipê* [e] e *ti* [i] são definidas como “médio-baixa, anterior, não-arredondada”; “médio-alta, anterior, não-arredondada”; e “alta, anterior, não-arredondada”, respectivamente. Finalmente, as vogais tônicas em *avó* [ɔ], *avô* [o] e *anu* [u] são identificadas como “médio-baixa, posterior, arredondada”; “médio-alta, posterior, arredondada”; e “alta, posterior, arredondada”, nessa ordem.

O quadro da Figura 1 representa as vogais orais que aparecem na sílaba tônica e que são quase uniformes no território nacional. Esse quadro se altera, contudo, quando se consideram o ambiente fonético das vogais – especificamente as sílabas átonas –, e também as condições geográficas de proferimento – os dialetos que elas pertencem. As sílabas átonas podem ser pretônicas (anteriores à tônica) ou postônicas (posteriores à tônica). Essas, por seu turno, são frequentemente subdivididas em mediais ou finais. O “o” em *colega* é uma vogal pretônica, [ko.'lɛ.gə]; enquanto o “e” e “o” em *número* são postônicos medial e final, nessa ordem, [ˈnu.me.ru].

Redução e abertura são dois processos de variação que acometem a pronúncia das vogais nesse ambiente fonético. Redução ou alçamento é o processo articulatorio que decorre da elevação da língua durante a produção de uma vogal que, em sílaba tônica, seria pronunciada com a língua mais baixa. Por exemplo, o “a” final em *colega*; ou o “o” final em *número*. Nesses casos, ambas as vogais ocorrem em sílaba postônica final. Já a abertura é um processo inverso. A vogal que seria pronunciada mais alta se estivesse em sílaba tônica é abaixada durante a sua articulação. É o caso do “e”, que de vogal “anterior médio-alta”, [e], passa para “anterior médio-baixa” [ɛ], como na pronúncia acima da palavra *colega*.

Finalmente, a nasalidade é um processo por assimilação de traço articulatorio adjacente ou questões dialetais. Trata-se de uma vogal oral que incorpora o traço nasal da consoante imediatamente subsequente a ela, seja “n” [n], “m” [m] ou “nh” [ɲ]. Isso independente do grau de tonicidade da sílaba, podendo ocorrer em sílaba tônica e também átona; por exemplo, o “a” em *manha*, [ˈmã.nə], e *amanhã*, [ã.mã.ˈɲã], respectivamente.

A diversificação ora explanada é um processo intrinsecamente fonético; porém, esse é condicionado por fatores sociolinguísticos, especialmente o dialeto e o estilo de fala, como será esclarecido na próxima subseção, acerca do condicionamento extralinguístico.

3.1 Condicionamento sociolinguístico

O fato de as vogais variarem em sílabas átonas certamente explica em grande medida o vocalismo do PB. Contudo, essa explicação não estará completa sem o esclarecimento dos fatores sociolinguísticos que consolidam essas variações. A seguir, verificar-se-á o papel dos dialetos “carioca”, “nordestino”, “sulista” e “caipira”.

Vale esclarecer, antes de tudo, que os dialetos a serem referenciados são rótulos convencionais usados para facilitar o ensino de algumas pronúncias prototípicas do PB. Nesse sentido, o mapeamento da variedade de falares que eles denotam é limitado. Por exemplo, o dialeto “nordestino” é usado para se referir àquelas formas clássicas de expressão que, baseados nas escutas de exemplos fornecidos por Silva [2], sinalizam a fala de alguns falantes nordestinos, mas não necessariamente de todos.

O mesmo acontece com o rótulo de “carioca”, que diz respeito aos falares estereotipados dos moradores do Rio de Janeiro; e assim

por diante. Além desses dois dialetos, serão mencionados o “sulista” e o “caipira”. Nesse último caso, o rótulo é meramente descritivo de um falar típico do interior de São Paulo, e mesmo norte do Paraná, sem qualquer conotação pejorativa. Seja como for, o fato é que os dialetos são condicionantes das variações vocálicas. Considerem-se alguns exemplos.

Em sílabas pretônicas, por exemplo, pode ocorrer o processo de abertura das vogais médio-altas no diletto nordestino. Assim, as palavras *beleza* e *bolota* são pronunciadas respectivamente como [be.'le.zə] e [bo.'lo.tə]; ao invés de [be.'le.zə] e [bo.'lo.tə], como acontece nos demais dialetos. Já em sílabas postônicas podem ocorrer os processos de abertura e também de oclusão. Nesse último, a vogal é produzida com a língua mais elevada: [a] para [ə]; [e] para [i]; [o] para [u] etc.

Em postônicas mediais, no dialeto nordestino, pode acontecer a abertura também: a palavra *número*, por exemplo, pode ser pronunciada como [ˈnu.me.ɾu]. Já em sílabas postônicas finais, é o processo de redução que pode ocorrer. Isso é mais típico em situação de estilo de fala informal, quaisquer que sejam os dialetos.

Nas palavras *mata* e *mato*, por exemplo, verifica-se a oclusão: [a] para [ə] e [o] para [u]; respectivamente [ˈma.tə] e [ˈma.tu]. A exemplo de acentamento em postônicas finais, a nasalidade é atestada também em muitos dialetos. Inclusive, ela aparece como uma pronúncia possível nos referidos dialetos. Isso significa que, à princípio, ela não sinaliza características regionais, como o faz a abertura vocálica.

Em suma, esses processos de abertura e redução são de natureza fonética, porém condicionados por fatores sociolinguísticos. O entendimento deles requer o conhecimento tanto da distribuição das vogais quanto da correlação entre elas e os dialetos. A transcrição fonética, nesse sentido, deve capturar as variações e, por conseguinte, representar esse conhecimento.

Do exposto pode-se concluir que Silva [2] oferece uma teoria adequada ao propósito de ensino/aprendizagem da fonética. Ela não apenas mostra como fazer a descrição, classificação e transcrição dos fones, como também enfatiza a correlação da variação fonética com os processos extralinguísticos. O transcritor em questão deve assim ser capaz de retornar transcrições fidedignas com esses processos de variação. A seguir, serão relatadas algumas ferramentas de transcrição fonética automática, disponíveis *online*.

4 TRABALHOS CORRELATOS

Há aplicativos que fazem transcrições fonéticas do Português Brasileiro de forma automática. Tais ferramentas estão disponíveis na web; e muitas das quais empregam recursos de Inteligência Artificial (IA). Em uma busca na web, por meio da palavra-chave “transcrição fonética”, foi possível identificar algumas dessas ferramentas, são elas: *Petrus (Phonetic Transcriber for User Support)* [4]; *Musely (Transcritor Fonético Online Gratuito)* [5];

LogicBalls (AI Conversor AI IPA) [6]; *EasyPronunciation.com (Pronúncia em português - transcrição fonética de palavras (Versão Beta))* [7].

Com exceção do *Petrus*, as demais ferramentas empregam IA e, portanto, são de natureza comercial, ou seja, o usuário precisa pagar pelo serviço, mesmo quando esse é voltado para o ensino de idiomas, como é o caso do *EasyPronunciation.com*. O corolário disso é uma generalização das transcrições tal que, em muitos casos, lembra a transcrição fonológica, por não oferecer representações fonéticas dos dialetos. Isso acontece particularmente no *EasyPronunciation.com* e no *LogicBalls*, as quais oferecem uma única opção de pronúncia; mas não no *Musely*, onde essa opção está disponível.

Um breve teste revela que o enunciado *Meu colega é legal* é transcrito como se segue: no *LogicBalls* como [mew ko'legə 'ε le'gaw]; no *EasyPronunciation.com* como [ˈmew ko.'le.gə 'ε le.'gaw]; e finalmente no *Musely* como [mɛw ku'legə ε le'gaw], no dialeto padrão, e [ˈmew ko'legə 'ε le'gaw], no dialeto nordestino, quando especificado.

Note-se ainda que em todas as transcrições *meu* é transcrito com o símbolo [w], que é índice de semiconsoante; e não com o símbolo [ɥ], que é índice de semivogal. Assim, a transcrição encontrada nesses aplicativos supõe a interpretação de que o ditongo em *meu* seria a expressão da sequência silábica CVC (consoante-vogal-consoante), visão que expressa um raciocínio fonológico e não fonético⁴. Portanto, a referida representação está em desarmonia com a interpretação da fonética do PB [2].

Quanto ao *Petrus*, ele não oferece opções de pronúncia de acordo com o dialeto. Além disso, algumas transcrições não são contundentes com a pronúncia real. Por exemplo, a mesma expressão acima é transcrita de maneira indiscriminada como [ˈmeu ko'lega 'ε le'gau]. Note-se que essa transcrição é mais adequada a uma interpretação fonética do PB, porquanto a semivogal é representada como [u]. Contudo, há importantes erros de transcrição como o emprego de [ε] em *colega* onde é para ser [e].

Em suma, nenhuma dessas ferramentas é adequada para cumprir o propósito pedagógico que se vislumbra para o transcritor. De um lado, os aplicativos disponíveis, além de envolverem custos financeiros com o pagamento dos serviços, apresentam transcrições generalizadas, o que pode dificultar mais que auxiliar o ensino; de outro, eles podem gerar transcrições equivocadas das pronúncias do PB, algo indesejável se o objetivo é ensinar as pronúncias existentes no país. Por tudo isso, assume-se que um transcritor mais especializado é necessário, cujo processo de construção será relatado em “Metodologia”.

⁴ Em comparação com a fonética, que estuda os sons da linguagem verbal em si mesmos, a fonologia estuda a estrutura formal desses sons.

5 METODOLOGIA

O transcritor é uma função escrita em Linguagem de Programação Python [9], denominada “transcrever”. Trata-se, na verdade, de uma função principal contendo funções especializadas, aqui denominadas de “funções-componente”; tais como “converter”, “dividir”, “acentuar” e “normalizar”. Cada função dessas foi escrita para cumprir uma tarefa específica de transcrição e, por conseguinte, satisfazer as referidas condições de efetividade. A Figura 2, abaixo, apresenta tais funções encadeadas no interior da função principal “transcrever”.

Nessa codificação, o uso de expressões regulares (*regex*) foi fundamental. Por meio da variável “pattern” do *regex*, foi possível encontrar as construções desejadas e substituí-las pelos símbolos fonéticos correspondentes. Essa codificação, conquanto simples à primeira vista, foi na realidade muito laboriosa. Foram necessárias algumas centenas de linhas de código para gerar as transcrições referentes aos dialetos considerados.

5.1 Padrão de transcrição adotado

O padrão de transcrição adotado é o da Silva [2], com o incremento de “ponto final” para sinalizar a divisão silábica. Um exemplo é a transcrição de *estacionamento*, ou seja: [es.ta.sjõ.nã.mẽi.tu]. Essa representação resulta da aplicação de diferentes sistemas de notação, correspondentes a cada função-componente e encadeados no interior da função principal.

```
def transcrever(palavra: str, dialeto: str)-> str:
    '''Transcreve foneticamente as palavras de entrada, de acordo com o dialeto.
    arg: palavra (str), dialeto (str)
    ret: palavra (str)
    ...
    palavra = palavra.lower()
    palavra = converter(palavra, dialeto)
    palavra = dividir(palavra)
    palavra = acentuar(palavra)
    palavra = normalizar(palavra, dialeto)
    return f'[{palavra}]'
```

Figura 2. Código da função principal.

Assim, esse padrão de transcrição passa três tipos de informação, a saber, o símbolo fonético correspondente ao fone observado; o arranjo silábico dos fones presentes; e a sílaba tônica da palavra transcrita (“normalizar” não gera informações, pois é uma correção das saídas das funções anteriores). Em vista desse padrão de transcrição, as funções-componente foram escritas para aplicar, cada qual, um sistema de notações específico (Figura 2).

5.2 Funções-componente

A função “converter” mapeia palavras proferidas em palavras fonéticas (Figura 2). Mais especificamente, ela aplica os símbolos correspondentes aos fones atestados, tendo em vista o proferimento da palavra de entrada: “e” -> [e], “s” -> [s], “t” -> [t], “a” -> [a], “c”

-> [s]... etc. Essa função emprega os símbolos do Alfabeto Fonético Internacional (IPA), o quais podem ser primários – os caracteres que representam os fones –, quanto secundários – os diacríticos que incrementam a descrição. Nesse último caso, note-se que, no dialeto local, o segmento ortográfico “-men-” de *estacionamento* é foneticamente pronunciado como um ditongo nasal [mẽi]. Assim, o diacrítico subscrito [̃] indica a semivogal do ditongo, enquanto o diacrítico sobrescrito [˘] indica a nasalidade da vogal do ditongo.

A função-componente “separar” demarca unidades silábicas da palavra fonética, por meio da notação do ponto final [·]: [esta.sjõ.nã.mẽi.tu] -> [es.ta.sjõ.nã.mẽi.tu] (Figura 2). A função “acentuar”, por sua vez, indica a sílaba forte da palavra, aplicando o diacrítico [˘] (Figura 2). Note-se o diacrítico sobrescrito antes da sílaba tônica do exemplo, que é o segmento “-men-”: [es.ta.sjõ.nã.mẽi.tu] -> [es.ta.sjõ.nã.˘mẽi.tu].

Por fim, a função “normalizar” foi escrita para corrigir eventuais erros de transcrição (Figura 2). Como se trata de um encadeamento de funções, a saída de uma sendo a entrada da próxima, alguns erros surgiram que foram eliminados por essa função. Os erros comuns foram a presença de dois diacríticos de tonicidade na mesma palavra; ausência de marcados silábicos; e a ocorrência de vogais abertas no lugar de fechadas, e vice-versa.

Vale dizer que esses erros decorrem de um processo de superposição de regras. Na ausência de critérios declarados de exclusão mútua, duas ou mais regras diferentes podem se aplicar à mesma saída em pontos distintos do fluxo de processamento. O resultado é, como mencionado, transcrições com notações duplicadas, apagadas ou equivocadas.

Para encerrar, a função “transcrever” recebe dois argumentos, a “palavra” a ser transcrita e o “dialeto” para o qual ela será transcrita. Isso é assim pelas razões acima colocadas: o dialeto restringe a representação fonética da palavra. Note-se, nesse sentido, a pronúncia de “s” em *estacionamento*, que pode ser alveopalatal, no carioca, [eʃ.ta.sjõ.nã.˘mẽi.tu]; ou alveolar, nos demais dialetos, [es.ta.sjõ.nã.˘mẽi.tu].

5.3 Corpus

Corpus é constituído por palavras extraídas de Silva [2], referentes ao vocalismo do PB. Ao todo são cento e noventa e quatro palavras ($n = 194$), as quais foram selecionadas por permitirem representar a variação dos segmentos vocálicos. Assim, cada palavra do *corpus* possibilita a produção das pronúncias típicas dos dialetos considerados. Essas palavras pertencem, em sua maioria, às classes de substantivo, adjetivo e advérbio; já a classe verbal requerendo assim um tratamento futuro.

Além disso, um gabarito foi criado com as transcrições corretas dessas palavras que serviu de linha de base para comparação e validação estatística do transcritor. As palavras foram anotadas manualmente por especialistas, tendo por base a teoria fonética de Silva [2].

Em suma, são essas pronúncias específicas de dialetos que vão permitir a explanação pedagógica das regras de produção fonética do vernáculo. Esse *corpus*, conquanto numericamente reduzido, é representativo do conhecimento fonético do PB; tal que ele se ajusta ao escopo desta proposta, que é o ensino fonético, e não a descrição exaustiva.

5.4 Métrica de validação

Duas métricas são empregadas, a Acurácia e a Taxa de Erro [10]. A acurácia mede a proporção de transcrições corretas, produzidas pelo transcritor, em relação ao número total de transcrições previstas pelo gabarito. Em termos simples, ela responde à pergunta quanto à porcentagem de vezes que o algoritmo acertou.

A fórmula para a acurácia é: $Acurácia = (\text{Número de Transcrições Corretas}) / (\text{Total de Transcrições})$. Para expressar como uma porcentagem, multiplica-se o resultado por 100. Já a Taxa de Erro quantifica diretamente a frequência com que a função falha em produzir a transcrição correta. É uma métrica muito comum na área de Processamento em Linguagem Natural (PLN), porque foca diretamente nas falhas, cujo custo pode ser mais relevante dependendo da aplicação. A fórmula para a Taxa de Erro é: $Taxa de Erro = (\text{Número de Transcrições Erradas}) / (\text{Total de Transcrições})$.

Assim como a Acurácia, para expressar como uma porcentagem, multiplica-se o resultado por 100. A Acurácia e a Taxa de Erro são as métricas mais intuitivas para avaliar um sistema de classificação; no presente caso, transcrições fonéticas. Contudo, para o propósito deste projeto de transcritor, a validação das condições de efetividade é certamente mais assertiva. Em suma, uma vez esclarecida a metodologia da pesquisa, vejamos os resultados do emprego do transcritor sobre esse *corpus*.

6 RESULTADOS

O transcritor processou as 194 palavras do *corpus* e gerou 776 transcrições (4 vezes 194), distribuídas igualmente em quatro dialetos: caipira, carioca, nordestino e sulista. Para avaliar a acurácia desses dados, empregou-se o gabarito descrito acima, que serviu como linha de base de comparação. Assim, foram identificadas nos dados 21 transcrições erradas e 755 transcrições corretas.

Os erros identificados foram os seguintes: 6 erros de acentuação (p.ex., *útero* [u.'te.ru], correto [u.te.ru]); 12 erros relacionados à ausência de abertura (p.ex., *peteca* [pe.'te.kə], correto [pe.'te.kə]); 2 erros relativos à abertura indevida (p.ex., *sonho* ['sɔ.nu], correto ['sɔ.nu]); e 1 erro por supressão de segmento (p.ex., *parênteses* [a.'rɛ̃i.'te.zis], correto [pa.'rɛ̃i.te.zis]). Vale lembrar que as regras responsáveis por tais transcrições podem (e devem) ser corrigidas para ajustarem os resultados à explanação de Silva [2].

Por outro lado, o algoritmo gerou 755 transcrições corretas, isto é, de acordo com o gabarito usado para comparação. Segue uma amostra dessas transcrições na Tabela 1.

palavras	carioca	nordestino
<i>legal</i>	[le.'gaw]	[le.'gaw]
<i>remoçar</i>	[Xe.mo.'saX]	[he.mo.'sah]
<i>terrinha</i>	[te.'Xi.nə]	[te.'hi.nə]
<i>sôfrego</i>	[so.fre.gɔ]	[so.fre.gɔ]
<i>mole</i>	['mɔ.li]	['mɔ.li]
<i>tônico</i>	['tõ.ni.ko]	['tõ.ni.ko]
<i>centrifuga</i>	[sɛ̃i.'tri.fu.gə]	[sɛ̃i.'tri.fu.gə]
...
palavras	sulista	caipira
<i>legal</i>	[le.'gal]	[le.'gaw]
<i>remoçar</i>	[re.mo.'sar]	[Xe.mo.'sa]
<i>terrinha</i>	[te.'ri.nə]	[te.'Xi.nə]
<i>sôfrego</i>	[so.fre.gɔ]	[so.fre.gɔ]
<i>mole</i>	['mɔ.li]	['mɔ.li]
<i>tônico</i>	['tõ.ni.ko]	['tõ.ni.ko]
<i>centrifuga</i>	[sɛ̃i.'tri.fu.gə]	[sɛ̃i.'tri.fu.gə]
...

Tabela 1. Amostra das transcrições geradas

A Tabela 1 apresenta algumas palavras cujas transcrições representam os processos de abertura, redução e nasalidade das vogais. Note-se que as transcrições estão distribuídas por dialetos, como deve ser.

Aplicaram-se, então, as métricas da Acurácia e da Taxa de Erro para se estimar o significado estatístico disso. A Acurácia dos resultados foi de 97,29%, um valor alto que, à primeira vista, indica um excelente desempenho. Este número significa que, para o conjunto de dados avaliado, o transcritor foi capaz de transcrever corretamente mais de 97 em cada 100 palavras.

Já a Taxa de Erro foi de 2,71%, um valor baixo que reforça a conclusão de que o algoritmo é eficiente. Esse valor indica que, em média, espera-se que a função cometa um erro a cada 37 transcrições ($1 / 0,0271 \approx 37$). Note-se que a soma da Acurácia e da Taxa de Erro é 100% (97,29% + 2,71%), pois são métricas complementares.

Os valores conseguidos indicam que o transcritor foi bem-sucedido, pois a quantidade de acertos/erros é estatisticamente significativa. Contudo, os critérios de validação mais relevantes advêm mesmo das condições de efetividade acima declaradas. E a avaliação de tais critérios depende de uma análise qualitativa. Desse modo, para analisar a satisfação de tais condições pelo transcritor, os dados foram divididos de acordo com os processos de variação das vogais.

7 DISCUSSÃO GERAL

Como dito, o transcrito será adequado se satisfizer as três condições de efetividade, (a) representar simbolicamente os sons da fala, (b) mostrar o ambiente fonético das vogais e (c) indicar o dialeto a que a pronúncia representa pertence. A seguir, serão considerados alguns processos de variação vocálica que as transcrições deveriam

representar e que são relevantes ao ensino do vocalismo português, a saber, a redução, a abertura e a nasalidade. Tais processos vão ser tratados à luz das referidas condições de efetividade.

7.1 Representar simbolicamente os fones

Abertura de vogais acontece no PB especialmente em sílabas pretônicas e indica tipicamente o dialeto nordestino: [e] > [ɛ] e [o] > [ɔ]. Note-se que os símbolos fonéticos são contudentes com esse processo, diferindo dos casos em que esse não ocorre. A palavra *beleza* e *pedal*, por exemplo, são transcritas com a postônica aberta [ɛ] ou fechada [e] a depender do dialeto: [be.'le.zə] e [pe.'daw] (nordestino); [be.'le.zə] e [pe.'daw]/[pe.'da] (demais dialetos). (Vale dizer que a lateral alveolar velarizada [ɮ] que aparece em *pedal* é uma pronúncia tipicamente sulista.)

A redução, por outro lado, é o processo inverso. Vogais mais baixas são alçadas a uma posição mais alta durante a articulação: [a] > [æ]; [e] > [i]; e [o] > [u]. É possível notar que, de modo geral, esse processo acontece indiscriminadamente no PB. Alguns exemplos ilustram os dados obtidos: *moleza* [mo.'le.zə]; *mole* ['mɔ.li]; e *pulo* ['pu.lu].

Finalmente, a nasalidade é o processo pelo qual uma vogal oral assimila a articulação nasal da consoante subsequente a ela. Assim, ela se torna nasal em virtude do ambiente fonético. Esse processo foi também devidamente gerado pelo transcritor. As transcrições de *bolinha* e *pezinho* são exemplos: [bo.'lĩ.nə] e [pe.'zi.nu], respectivamente. Note-se a nasalidade do “i” em ambos os casos. Em suma, o transcritor consegue satisfazer a condição (a), o que sugere a adequação dele à prática analítica.

7.2 Mostrar o ambiente fonético

Observando-se a Tabela 1, nota-se que o ambiente fonético é adequadamente mostrado. O “ponto final” separando as sílabas é um diacrítico significativo não apenas para indicar o limite de sílabas, como também informar a quantidade de sílabas na palavra. Vejam-se, ademais, as seguintes transcrições obtidas: *vi* ['vi], uma sílaba; *doce* ['do.sɪ], duas sílabas; *safári* [sa.'fa.ri], três sílabas; *seriamente* [se.ri.ã.'mẽi.tɪ], quatro sílabas etc. Vale dizer que esse diacrítico delimitador de sílabas foi um incremento no padrão de transcrição encontrado em Silva [2].

Junte-se a isso a presença do marcador de sílaba tônica, o diacrítico aplicado antes da sílaba. Ele é pedagogicamente muito útil também, pois, ao marcar a tônica, facilita a apreensão das átonas, em posições pretônica ou postônica (final e medial): variações em *povoar* [pɔ.vo.'ah] (pretônicas) e *número* ['nu.mɛ.ru] (postônica medial). Trata-se, portanto, de um recurso simbólico com implicações para o entendimento do vocalismo do PB.

Assim, sem uma marcação de sílaba tônica, seria muito difícil esclarecer a posição das pretônicas ou postônicas e dizer quais processos de variação estão envolvidos. Tampouco seria suave informar que a nasalidade da vogal pode advir da influência da

consoante nasal na sílaba subsequente; por exemplo, [m] e [n] em *câmera* ['kã.me.rə] e *anêmona* [ã.'nẽ.mõ.nə], nessa ordem. Nesse sentido, o transcritor consegue também satisfazer mais uma exigência pedagógica.

7.3 Indicar o dialeto da pronúncia transcrita

É esperado que as transcrições representem pronúncias que são típicas de dialetos. Se isso for o caso para o presente transcritor, então ele é adequado ao ensino. Ao se analisarem os processos, verifica-se que efetivamente as transcrições geradas sinalizam variações dialetais.

No caso de abertura, isso aconteceu sistematicamente no dialeto nordestino, onde essa variação é esperada, e não nos demais dialetos, onde dificilmente vai ocorrer. No tocante à redução, por outro lado, esse processo foi mostrado como uniforme em todas as variedades. A redução acontece dessa maneira porquanto, em estilo informal de fala, vogais postônicas finais tendem à realização fechada no PB.

Por fim, o processo de nasalidade é menos indicativo de dialeto, pois, no PB, apenas em poucas comunidades a nasalidade é marca de identidade linguística [11]. Por outro lado, no dialeto nordestino, essa nasalidade pode não ocorrer e, no lugar, verificar-se uma vogal aberta: p.ex., *tônico* ['tõ.ni.ku] no dialeto nordestino, e ['tõ.ni.ku] nos demais dialetos (Tabela 1). Uma vez que essas marcações são verificadas nos dados, o transcritor atende assim à exigência em (c).

Esses resultados sugerem que o transcritor proposto é adequado à prática analítica empregada no ensino. Com efeito, as transcrições obtidas satisfazem a todas as três condições de efetividade e, por extensão, permitem abordar o conhecimento fonético que é exigido de futuros professores do vernáculo.

7.4 O transcritor e seus congêneres

O fato de o referido algoritmo ser adequado ao ensino dirigido de fonética nada informa sobre sua viabilidade enquanto recurso pedagógico em comparação com os aplicativos disponíveis *online*. Nesse sentido, pode-se afirmar que, em relação àquilo para o qual ele foi concebido, o transcritor é superior a seus congêneres.

Algumas razões sugerem que isso é caso. Em primeiro lugar, trata-se de uma aplicação sem custos financeiros; o que não acontece, em geral, com os aplicativos avaliados acima, como o *Musely*, para se mencionar apenas um. Em segundo lugar, mesmo seu escopo sendo um *corpus* específico de processos de variação vocálica em PB, o transcritor fornece transcrições mais pormenorizadas e fidedignas que seus correlatos, como o *Petrus*, o *Musely*, o *EasyPronunciation.com* e o *LogicBalls*. Como se viu antes, os aplicativos baseados em IA são generalistas demais, uma vez que lidam com diferentes idiomas e, por conseguinte, não focam nas particularidades das línguas consideradas. Isso é notável quando se considera a neutralização da diferença entre o ditongo “eu”, [ɛu], e a consoante lateral alveolar vocalizada “al”, [aw];

distinção que passa a ser representada sumariamente como [w], em palavras como *meu* e *legal*: [ˈmew] e [le.ˈgaw], respectivamente.

Por fim, o transcritor se ocupa especificamente das pronúncias que são específicas de dialetos brasileiros. Comparado com seu congêneres nacional, o aplicativo *Petrus*, ele acerta as transcrições que esse último erra. Verifica-se isso quando a transcrição da palavra *colega* é solicitada ao aplicativo. Com efeito, *Petrus* retorna [koˈlega], e não [ko.ˈle.gə], como era de se esperar. Trata-se de um erro inaceitável, uma vez que em sílaba tônica, como é o caso do exemplo, a pronúncia das vogais não sofre variação de redução [ɛ] > [e].

Assim sendo, em que pese a especialidade de seu escopo, o transcritor pode cumprir o desempenho pedagógico ora defendido em bases mais realistas e precisas que seus congêneres, notadamente aqueles ancorados em IA.

8 CONSIDERAÇÕES FINAIS

Na Introdução foi questionado se as tecnologias digitais poderiam facilitar o ensino linguístico na Academia, diante dos desafios colocados pelas práticas analíticas, como a transcrição fonética. A hipótese aventada foi a de que um AIA poderia cumprir esse papel contanto que suplementado por ferramentas externas, como uma função computacional dedicada à transcrição fonética. Na ausência dessa ferramenta, propôs-se, nesse sentido, então, um transcritor fonético automático (transcritor) que fosse adequado a essa tarefa, a qual deveria ser acima de tudo pedagógica.

Os objetivos foram, assim, descrever a implementação do transcritor e demonstrar como ele poderia contribuir com o ensino. Depois de tudo, os objetivos foram alcançados. Com efeito, evidenciou-se que, como base em um *corpus*, o transcritor é capaz de representar o conhecimento que subjaz o ensino do vocalismo do PB. Isso foi validado não apenas pela métrica de contagem de dados, mas especialmente pelas chamadas condições de efetividade. Como foi possível mostrar, ele satisfaz a todas essas condições.

Além disso, em comparação com os aplicativos *online* de transcrição fonética, a ferramenta criada mostra ser mais contundente com os fatos fonéticos do PB. Ela consegue, com efeito, não apenas capturar nuances sonoras, decorrentes de mudanças do ambiente fonético, como também ser, por isso mesmo, mais precisa que os aplicativos avaliados. Assim sendo, o transcritor pode ser empregado na disciplina de fonética e fonologia como um suporte pedagógico.

Em termos de tecnologias disruptivas, ele poderia ser integrado a um AIA como uma ferramenta externa, o modelo de linguagem sendo mobilizado para refinar as entradas e saídas de dados. Esse modelo ofereceria explicações detalhadas sobre processos representados nas transcrições geradas. Nesse sentido, um agente de IA contendo o transcritor pode funcionar como um assistente virtual do docente nas aulas de fonética articulatória.

Por fim, o transcritor apresenta limitações que devem ser mitigadas em trabalhos futuros. Destaquem-se duas: a exiguidade do *corpus* e os erros de transcrição obtidos. Quanto à exiguidade, embora o transcritor tenha o potencial de transcrever dados não vistos, essa capacidade ainda não foi testada. As 196 palavras testadas referem-se, como dito, ao *corpus* extraído das explanações de Silva [2]. Para os propósitos pedagógicos em questão, eles são suficientes e, por isso mesmo, não precisam ser escalados.

Com relação aos erros, esses podem ser corrigidos em versões melhoradas do algoritmo. Tal melhoria consiste em especificar mais as regras que geraram as transcrições erradas.

Seja como for, desde que as regras de produção fonéticas sejam explicitadas a partir de dados representativos, como é o caso em questão, a explicação da variação é suficiente para promover a compreensão e generalização do vocalismo no PB. Contudo, vale lembrar que a total efetividade do transcritor como ferramenta pedagógica requer também uma avaliação com os estudantes expostos a ele. Assim, em trabalhos avaliativos futuros pretende-se ampliar o escopo da efetividade do transcritor.

REFERÊNCIAS

- [1] Djiby Mané. 2012. As concepções de língua e dialeto e o preconceito sociolinguístico. *ViaLitterae*. 4, 1 (Jan/Jun, 2012), 39-51.
- [2] Thaís C. Silva. 2002. *Fonética e fonologia do Português* (6. ed.). São Paulo: Contexto.
- [3] Iara M. Telles e Adelaide H. P. Silva. Algumas considerações sobre a transcrição fonética nos Atlas Linguísticos do Brasil. *Signum: Estudos da Linguagem*. 11, 2 (Dez, 2008), 277-287. DOI: <https://doi.org/10.5433/2237-4876.2008v11n2p277>
- [4] Serrani, V. M., Bokan, A. e Zavaglia, C. 2014. PETRUS: A rule-based grapheme-to-phone converter for Brazilian Portuguese. In: *Proceedings of the International Conference on Computational Processing of Portuguese (PROPOR'14)*. Springer, Heidelberg, 154-158. DOI: https://doi.org/10.1007/978-3-319-09761-9_17.
- [5] Musely. 2025. *Transcritor Fonético Online Gratuito*. URL: <https://musely.ai>.
- [6] LogicBalls. 2025. *AI Conversor AI IPA*. URL: <https://logicballs.com>.
- [7] Baytukalov, T. 2017. Pronúncia em Português - transcrição fonética de palavras (Versão Beta). *EasyPronunciation.com*. URL: <https://easypronunciation.com/pt/>.
- [8] Bismarck Z. de Moutra. 2020. Vocalismo átono do Português do Brasil. *Revista Philologus*. 26, 78 Supl. (Set/Dez, 2020), p. 3715- 3730.
- [9] Python Software Foundation. 2025. *Python Language Reference* (Version 3.14.0). URL: <https://www.python.org>
- [10] Jurafsky, D. and Martin, J. H. 2023. *Speech and Language Processing* (3rd ed. draft). Prentice Hall, Englewood Cliffs, New Jersey. URL: <https://web.stanford.edu/~jurafsky/slp3/>
- [11] Regina Cruz. 2012. Vogais na Amazônia Paraense. *Alfa*. 56, 3 (dez. 2012). 945-972. URL: <https://repositorio.ufpa.br/server/api/core/bitstreams/415aa83d-14db-4778-b6a7-78d1ef0201ad/content>