

## **SISTEMA AUTOMÁTICO DE TRANSCRIÇÃO FONOLÓGICA PARA O PORTUGUÊS** **AUTOMATIC PHONOLOGICAL TRANSCRIPTION SYSTEM FOR PORTUGUESE**

Daniel da Silva Santos  
Universidade Federal Rural do Semi-Árido  
[danielsantos7@outlook.com.br](mailto:danielsantos7@outlook.com.br)

Iara Cristina Araújo Nogueira  
Universidade Federal Rural do Semi-Árido  
[yarachristinah@hotmail.com](mailto:yarachristinah@hotmail.com)

Cid Ivan da Costa Carvalho  
Universidade Federal Rural do Semi-Árido  
[cdivanc@gmail.com](mailto:cdivanc@gmail.com)

**RESUMO:** Os sistemas de transcrição automática de grafema para fonema são conhecidos como *Graphem to phoneme* (G2P). Neste trabalho, apresentamos um sistema automático de transcrição fonológica para o português, utilizando a tecnologia de estados finitos. Para o desenvolvimento desse sistema, seguimos os seguintes passos: a compreensão da relação entre as formas gráficas e as formas fonológicas da língua, a construção de um algoritmo, a implementação desse algoritmo numa linguagem de programação, o teste e a avaliação do sistema num *corpus* da língua portuguesa. Após o desenvolvimento, os resultados mostraram que o sistema apresenta nível satisfatório para a maior quantidade de palavras dessa língua; todavia, ainda precisa melhorar em outros aspectos, como a distinção entre o som aberto e o som fechado nas vogais anterior e posterior.

**PALAVRAS-CHAVE:** português; transcrição fonológica automática; forma gráfica; forma fonológica.

**ABSTRACT:** The automatic grapheme transcription systems for phoneme are known as Graphem to phoneme (G2P). In this work, we present an Automatic phonological transcription system for Portuguese, using finite-state technology. For the development of this system, we follow these steps: the understanding of relationship between the graphical form and the phonological form of the language, the building of an algorithm, the implementation of this algorithm in a programming language, the testing and the evaluation of the system in a Portuguese language writing *corpus*. After the development, the results showed that the system presents a satisfactory level for the greatest amount of words of that language; however, it needs to be improved in other aspects, such as the distinction between open and closed sound in the anterior and posterior vowels.

**KEYWORDS:** Portuguese; automatic phonological transcription; graphical form; phonological form.

## 1 Introdução

A Fonética e da Fonologia são ramos da Linguística que estudam a sonoridade, porém, sob aspectos distintos: a primeira estuda a natureza física da produção e da percepção dos sons da fala (fones), sem considerar necessariamente o sentido das palavras, enquanto que a segunda se preocupa com a maneira como os sons se organizam dentro de uma língua, classificando-os em unidades capazes de distinguir significados (fonemas). Essa diferença reflete diretamente na maneira como fazemos a transcrição de formas escritas<sup>1</sup> para uma ou outra forma. Na transcrição fonética, os símbolos representam os sons emitidos por falantes de uma língua, enquanto que, na fonológica, os símbolos representam os sons que distinguem as palavras de uma determinada língua.

Essas duas formas de transcrição apontam para dois tipos de conversores automáticos: o *Grapheme to Phoneme* (doravante G2P) e o *Letter to Sound* (doravante LTS). Como afirma Carvalho (2016), os primeiros enfatizam a conversão de um conjunto de grafemas para uma sequência de símbolos fonológicos de uma determinada língua, enquanto os segundos executam a transcrição da sequência de grafema para os símbolos fonéticos que representam uma variedade linguística, tentando transcrever a palavra mais próxima possível da pronúncia.

A escolha do primeiro ou do segundo tipo de conversor interfere no resultado final, ou seja, na forma de transcrição. Por exemplo, por um lado, o Grafone é um conversor automático que faz a transcrição fonológica da palavra, isto é, a transcrição feita por esse sistema não apresenta símbolos que representam a fala de uma variedade específica, como aponta Veiga, Candeias e Perdigão (2011). Por outro lado, o Potigrafone é um sistema que transcreve a palavra considerando a sua pronúncia da palavra para a variedade linguística potiguar (CARVALHO, 2016, 2017). Então, dizemos que o primeiro conversor se enquadra como um G2P e o segundo como um LTS.

Esses conversores possuem duas utilizações principais: na pesquisa em Fonologia, uma vez que um sistema de transcrição fonológica automático realiza parte dos trabalhos dos profissionais que atuam na área da Fonologia, contribuindo para o estudo das pronúncias e da organização dos fonemas, especificamente, da língua portuguesa do Brasil; na aplicação em ferramentas tecnológicas, ou seja, no pré-processamento dos sistemas da tecnologia de fala e como ferramenta de pesquisa em Linguística. Também fazem parte do pré-processamento para que o sistema de *síntese de fala* e reconhecimento de voz consigam “adivinhar” a pronúncia correta das palavras (TEIXEIRA, OLIVEIRA, MOUTINHO, 2006). A síntese de fala (*Text-To-Speech – TTS*) e o reconhecimento de voz (*Speech-To-Text – STT*) estão presentes em serviços de telecomunicações, na Educação, em controle de equipamentos industriais, em alarmes para situações de risco, em ajuda às pessoas com algum tipo de necessidade especial (especialmente em pessoas com deficiência visual ou auditiva), em controles de lista de espera em hospitais, sinalização de bagagens em transportes públicos, sistemas de navegação GPS, *smartphones*, etc.

1 Destacamos que a expressão “forma gráfica” diz respeito à escrita ortográfica da palavra na língua portuguesa e a “forma fonológica” corresponde à transcrição fonológica da palavra. A primeira é escrita entre colchetes angulares <> e a segunda entre barras inclinadas //. Assim, o símbolo <p> equivale à letra “pê” do alfabeto português e o símbolo /p/ corresponde ao som oclusivo bilabial surdo.

No entanto, são poucos trabalhos que pretendem estudar a conversão automática das formas escritas para a forma fonológica da língua, porque, na maioria das vezes, não existe correspondência direta entre a grafia e a representação fonológica. No entanto, podemos mencionar dois trabalhos muito importantes nessa área: o trabalho de Veiga, Cadeia, Perdigão (2011) que apresentam um *Grapheme to phoneme* (G2P) para o português europeu e o Vasilévski (2008), que desenvolveu um sistema como esse para o estudo estatístico dos fonemas do português brasileiro.

Neste trabalho, apresentamos um sistema automático de transcrição fonológica para o português, considerando especificamente a forma da língua para o português brasileiro. Utilizamos como símbolos de saída o Alfabeto Fonético Internacional (*International Phonetic Alphabet – IPA*) – criado pela Associação Fonética Internacional como uma forma de representação padronizada dos sons das línguas.

Este artigo está dividido em quatro tópicos: no primeiro tópico, apresentamos as abordagens dos conversores grafofônicos na literatura existente; no segundo, distinguimos as formas de transcrição fonológica para o português; no terceiro, mencionamos os processos metodológicos que utilizamos para o desenvolvimento do sistema e, no quarto tópico, apresentamos os resultados e a avaliação do sistema.

## 2 Conversores grafofônicos

Neste tópico, mostraremos e exemplificaremos as abordagens dos conversores desenvolvidos para o português. Alguns sistemas de transcrição grafofônicos automáticos foram desenvolvidos e distribuídos para essa língua, como o Grafone e o Nhenhém. A partir da análise desses sistemas, foi possível distinguir três tipos de técnicas utilizadas na programação desses *softwares*: os sistemas por regras, os sistemas probabilísticos (estatísticos) e os sistemas híbridos. O primeiro tipo de programa utiliza uma série de regras pré-programadas, a partir de padrões presentes em uma língua, para reconhecer e realizar a correta transcrição dos grafemas. O segundo utiliza de métodos estatísticos e/ou probabilísticos para realizar a transcrição. O terceiro utiliza ambas as técnicas já citadas durante o processo.

Desse modo, podemos exemplificar o primeiro tipo por meio dos sistemas de Braga, Coelho e Resende (2006) e Vasilévski (2008). Os primeiros autores construíram um algoritmo que faz as transcrições fonológica e fonética baseadas em regras linguísticas para o português europeu. No entanto, os autores alegam que o algoritmo pode ser muito útil para outras variedades de Português, inclusive para o português brasileiro. Essas regras foram implementadas por meio de padrões fonológicos e utilizando, na transcrição dos grafemas, os símbolos fonéticos do alfabeto fonético SAMPA. Eles implementaram as regras de transcrição agrupando-as de acordo com os caracteres comuns e realizaram vários testes que foram ordenados por frequência dos grafemas na língua e testaram das mais frequentes até a menos frequente. O Nhenhém, programa desenvolvido por Vasilévski (2008), também faz uso da técnica de transcrição por regras linguísticas. Esse programa é usado para o tratamento de textos escritos e são codificados em símbolos fonológicos do AFI, formando um *corpus* para análise estatística.

Um exemplo do segundo tipo é o conversor de Barros e Weiss (2006). Esses autores apresentam um sistema probabilístico baseado em modelos de máxima entropia.

Esse sistema executa a transcrição gráfica e fônica, insere a acentuação e faz a separação silábica das palavras para o português europeu (doravante PE). Segundo esses autores, os sistemas probabilísticos são mais flexíveis quanto ao som sintético natural do discurso contínuo, uma vez que os modelos estatísticos podem ser treinados de tal maneira que são proferidos no ritmo da fala, destinando-se ao uso em determinada aplicação em síntese de fala.

Como exemplo dos sistemas de transcrição fonológica híbridos para o português, podemos mencionar o Grafone – Conversor de grafema para fonema do Português Europeu – que foi desenvolvido pela equipe do Instituto de Telecomunicação, no Laboratório de Processamento de Sinais em Coimbra, Portugal. Esse laboratório desenvolve sistemas de *reconhecimento de fala* e *síntese de fala*, os quais exigem o desenvolvimento de um conversor de grafema para fonema de alto nível.

No que se refere à abordagem desses conversores, os sistemas baseados em regras são muito úteis para gerar aplicação padrão nas transcrições de síntese ou para construir novos *corpora* para sistemas de aprendizado de máquina, como o sistema desenvolvido por Braga, Coelho e Resende (2006) para o PE e o Nhenhém 1.0 de Vasilévski (2008), para o PB. Porém, como bem apontam Barros e Weiss (2006), esses sistemas apresentam alguns inconvenientes, uma vez que a aplicação de uma regra pode ter consequências desastrosas em outras, porque as regras são construídas em forma de cascatas, ou seja, após a execução da primeira regra, o sistema lança como *output* para a segunda e assim sucessivamente. Essa característica serve tanto para os sistemas de regras quanto para o sistema de estados finitos, baseado em regra do padrão gerativo. Outro problema é o fato de serem mais caros do que os sistemas probabilísticos, pois é necessário um linguista perito para configurar todas as regras e exceções com o fim de produzir os bons resultados.

No entanto, como bem aponta Carvalho (2016), os sistemas que têm uma abordagem de estados finitos são mais flexíveis na integração entre as múltiplas fontes de informação linguística – como a integração de um conversor grafofônico com um etiquetador morfossintático – além de integração de conhecimento em métodos baseados em dados dirigidos. No entanto, deve-se tomar cuidado com o uso das regras de reescritas, pois a má utilização pode tornar o programa ineficaz.

Esses sistemas realizam a transcrição fonológica automática de textos e palavras; contudo, esses apresentam distinções as quais se faz necessário apresentar. O *Grafone* apresenta um excelente desempenho, realizando a conversão fonológica de forma correta com a grande maioria dos vocábulos; ademais, o programa permite a conversão para dois tipos de alfabetos de representação, o alfabeto fonético internacional (IPA) e o SAMPA. Apesar desses pontos positivos, o sistema executa a conversão apenas para o português europeu; desse modo, a transcrição não é completamente compatível com os vocábulos brasileiros. O *Nhenhém*, em contrapartida, foi desenvolvido visando ao suporte ao português brasileiro, também possui um ótimo desempenho e ainda permite extrair dados estatísticos durante a transcrição, o que o torna uma excelente ferramenta de suporte a pesquisas fonológicas. Ambos os sistemas realizam a transcrição automática para o português, mas apresentam enfoques diferentes no processo de transcrição.

Feitas essas ressalvas sobre os conversores automáticos, mostraremos as formas de transcrição fonológica automática para o português.

### 3 Transcrição fonológica automática

Utilizaremos o termo *transcrição* para nos referir à representação escrita dos símbolos fonéticos de um texto escrito graficamente. A expressão *transcrição fonológica* se refere à representação dos fonemas da língua, ou seja, é uma representação abstrata do sistema sonoro da língua e se distingue da *transcrição fonética*, que é a representação dos sons emitidos por falantes de uma língua quando produz a fala, como afirmam Seara, Nunes e Volcão (2015). Assim, a expressão *transcrição fonológica automática* remete à representação escrita dos fonemas da língua feita por um sistema computacional.

Quando falamos de transcrição grafofônica, como aponta Carvalho (2016), nos deparamos com o fato incontestável de que a forma escrita da língua portuguesa é diferente da forma como os falantes a pronunciam. Para Bechara (2005, p. 53), três fatores contribuem para que o sistema ortográfico (ou gráfico) do português não possua uma ortografia ideal, ou seja, um sistema em que cada letra corresponde a um som da língua: (1) a adoção do alfabeto latino que nem sempre é capaz de representar os fonemas da língua; (2) as mudanças que ocorreram com os fonemas das línguas neolatinas depois de terem adotado esse alfabeto e (3) a indecisão permanente das convenções ortográficas entre a opção do sistema fônico ou etimológico.

Considerando esses fatores, destacamos que a forma escrita da língua portuguesa apresenta apenas um conjunto de 26 letras para representar na escrita os 33 fonemas existentes na língua. Desse conjunto de letras, cinco representam os sete fonemas vocálicos, dezessete representam os vinte e seis fonemas consonantais, uma letra sem correspondência fonológica – a letra <h> – usada por herança etimológica, e três letras para o uso nas palavras estrangeiras – as letras <k>, <w> e <y>. Por isso, segundo Bechara (2005), o alfabeto absorve diversas variações e extensões, fundindo letras – como ocorre nos dígrafos – e modificando-as com o uso dos diacríticos, atribuindo funções especiais a duplas de letras ou absorvendo letras completamente novas ao alfabeto.

Esses fatores destacam que a relação do grafema para fonema da língua não ocorre de forma direta, ou seja, uma letra nem sempre corresponde a um fonema. A letra vocálica <e>, por exemplo, pode representar o fonema vocálico /e/ ou /ɛ/. Na palavra <teve> (pretérito perfeito do indicativo), a primeira letra corresponde ao fonema /e/ e, na palavra <leve>, a letra corresponde ao fonema /ɛ/. Outros exemplos com as consoantes, como as letras <s>, <x>, <c>, etc., podem ser vistos nos quadros 4 e 5.

Acrescentamos que, além das idiosincrasias da ortografia portuguesa, a *transcrição fonológica automática* se depara com a relação em que as representações gráficas (as formas escritas da língua) são cadeias do nível superior, ou seja, o que o escritor tinha em mente quando escrevia o texto, e as representações fonológicas são cadeias do nível inferior. Em outras palavras, a primeira cadeia no sistema é constituída pelas formas gráficas e a segunda pelas formas fonológicas da língua.

Nessa relação, como mostra a Figura 1, no **tópico 4.5** deste artigo, a regra de reescrita está na sequência intermediária e

consiste em uma sequência ordenada de regras de reescrita que converteram as representações fonológicas abstratas em formas de superfície através de uma série de representações intermediárias. Cada regra de reescrita tem a forma geral

$\alpha \rightarrow \beta / \gamma \delta$ , onde  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$  podem ser arbitrariamente cadeias complexas ou matrizes de traços (BEESLEY e KARTTUNEN, 2002, p. 309, tradução nossa<sup>2</sup>).

As regras são aplicadas em cascata com estágios intermediários, gerando as formas de superfície. A tarefa no desenvolvimento de um transdutor de grafema para pronúncia é criar uma cascata de regras que mapeiam as cadeias ortográficas do português (*lexical side*) para as cadeias fonéticas (*surface side*) que representam a pronúncia da cadeia de entrada.

#### 4 Processo metodológico

A partir dos conhecimentos adquiridos, oriundos da pesquisa e dos estudos na área da Fonologia e da Informática, o *software* de conversão grafofônico foi, então, desenvolvido. Optou-se, a princípio, pelo sistema de regras para realizar as conversões dos grafemas, devido à agilidade desse método. Para isso, o conversor fonológico foi desenvolvido com auxílio do *software Foma*, um compilador de estados finitos, biblioteca e linguagem de programação de uso múltiplo (HULDEN, 2008), desenvolvido pelo programador Mans Hulden. Sua simples sintaxe, rica biblioteca de funções e o suporte a expressões regulares, possibilita o desenvolvimento de poderosas aplicações de processamento com linguagem natural.

O *Foma* é, também, um programa de código livre com licença pública geral (GPL) que permite a livre distribuição e modificação no seu código, todas essas particularidades o torna uma plataforma ideal para o desenvolvimento de analisadores morfológicos/fonológicos, possuindo como um ponto fraco, apenas, o quesito portabilidade, uma vez que é restrita às linguagens de programação C/C++, Perl e JavaScript, não possuindo suporte direto à outras linguagens importantes como o Java e o PHP.

Outra característica importante que destacamos é o uso de expressões regulares. Em Ciência da Computação, são padrões presentes em uma cadeia de caracteres escritos de maneira formal e reconhecidos por um módulo processador (JARGAS, 2006). Várias linguagens de programação, atualmente, reconhecem expressões regulares que possibilitam a manipulação ágil de qualquer padrão de caracteres como textos, frases, palavras e sílabas; por isso, essa poderosa ferramenta foi utilizada no desenvolvimento do programa. Por conveniência, optamos por utilizar as expressões regulares do *Foma* especificamente, uma vez que as funções de manipulação são mais simples, flexíveis e totalmente compatíveis com esses padrões.

2 “consisted of an ordered sequence of REWRITE RULES that converted abstract phonological representations into surface forms through a series of intermediate representations. Such rewrite rules have the general form  $\alpha \rightarrow \beta / \gamma \delta$ , where  $\alpha$ ,  $\beta$  e  $\gamma$  and  $\delta$  can be arbitrarily complex strings or feature-matrices”.

#### 4.1 Algoritmo geral do transcritor

De modo geral, o algoritmo do transcritor é simples, o programa simplesmente solicitará uma palavra de entrada e esta será transcrita fonologicamente, ou seja, terá caracteres substituídos por símbolos fonéticos, de acordo com um conjunto de regras definidas pelas expressões regulares. Essa substituição será feita com auxílio de funções do *Foma*. O programa executa a transcrição de uma palavra por vez. Por isso, no caso de um texto, as palavras devem ser, primeiramente, separadas manualmente ou por meio de um *Tokenizador* – sistema automático que segmenta um texto em *tokens* (palavras e símbolos). Além disso, as regras de substituição necessitam de uma análise de sílabas separadamente; desse modo, também será necessário um separador silábico para que o programa seja executado corretamente. Por isso, desenvolvemos um sistema de separação silábico e um *tokenizador* integrados ao programa *Bag of Tools*, mas que não apresentaremos neste trabalho.

#### 4.2 Estrutura das funções do *Foma*

Segundo Hulden (2008), o *Foma* é um compilador, uma linguagem de programação e uma biblioteca em C para a construção de autômatos e transdutores de estados finitos. Essa biblioteca utiliza caracteres especiais para a manipulação computacional de estruturas mórficas e fonológicas. O conhecimento dos comandos e operadores dessa biblioteca é fundamental para compreender as aplicações das regras chomskianas utilizadas no desenvolvimento do programa.

No Quadro 1, apresentamos alguns operadores utilizados na implementação das expressões regulares em *Foma*.

Quadro 1: Símbolos utilizados nas regras fonológicas do *Foma*.

Símbolo	Nome	Descrição
A_B	Traço	Marca a posição exata em que ocorre o segmento, cujo contexto será caracterizado pelo que o precede e segue.
A*	Kleene star	Corresponde a repetição de zero ou mais vezes.
A+	Kleene plus	Corresponde a repetição de uma ou mais vezes.
A   B	Barra vertical	Une dois conjuntos de caracteres.
A B	Espaço em branco	Concatena dois ou mais elementos do conjunto ou dois ou mais conjuntos.
A .o. B	Composição	Compõe uma rede de estados finitos.
∅	Conjunto vazio	Indica o conjunto vazio; pode ser representado também pelos <i>colchetes vazios</i> ou <i>0 (zero)</i> .
->	Seta à direita	Significa que o elemento à esquerda transforma-se no da direita; é chamada também de substituição direta em contexto.
.#.	Cerquinha	Limite de palavra
	Duas barras verticais	Separa o contexto de aplicação da regra a ser aplicada.
;	Ponto e vírgula	Indica a finalização da linha de código.

Fonte: adaptado de Hulden (2006, p. 87).

O Quadro 1 apresenta os principais operadores utilizados nas regras do *Foma* para executar a relação do *input* e com o *output* do sistema, ou seja, a relação entre grafema e fonema da língua portuguesa, em que cada letra do alfabeto gráfico corresponde a um símbolo do alfabeto fonético. Nesse sentido, ressaltamos que o sistema de transcrição fonológica apresenta apenas uma transcrição possível para cada palavra, ou seja, existe apenas um *output* para um conjunto de grafemas do *input*. Porém, como aponta Carvalho (2016), nem sempre essa relação satisfaz o uso da pronúncia numa variedade linguística da língua portuguesa. Nas transcrições fonéticas, uma palavra escrita pode ter duas ou mais formas de pronunciá-la. Como exemplo, citemos, por exemplo, a palavra <cebola> que pode apresentar duas pronúncias, [si'bole] e [se'bole], numa mesma variedade linguística. Nesse sentido, a primeira forma de transcrição apresenta uma saída igual para todos os dialetos (variedades linguísticas). As particularidades fonéticas de cada variante são expressas na transcrição fonética.

No próximo tópico, compreenderemos a relação entre o *input* e o *output* do sistema com a aplicação das regras, para termos uma visão geral de como ocorre o processo de transcrição fonológica utilizando uma rede estados finitos.

### 4.3 Regras fonológicas

Segundo Chomsky e Halle (1968), o componente fonológico da gramática deve ter uma propriedade recursiva e conter certas regras que possam ser aplicadas indefinidamente com frequência, em novos arranjos e combinações, na geração (especificação) de descrições estruturais das frases em uma língua. Nesse sentido, as regras constituem a gramática de um determinado falante da língua e determinam em detalhes a forma das frases com que o falante produzirá e compreenderá a estrutura linguística. Uma forma direta de efetuar as mudanças em uma gramática é adicionar novas regras ao componente fonológico, atuando na modificação, no acréscimo ou no apagamento dos segmentos das formas de entrada.

Nesse subtópico, enfatizamos a formalização das regras fonológicas para o gerativismo e apresentamos a execução do sistema partindo das representações subjacentes – ou *input* – para as representações de superfície – ou *output*.

Na perspectiva gerativista, Seara, Nunes e Volcão (2015, p. 140) afirmam que "o componente fonológico é formado por um conjunto de representações subjacente e por regras que definem como essas representações emergem na superfície". As regras são aplicadas às representações *subjacentes*, transformando-as e gerando novas formas de representação até o término do processo derivacional, quando se tem a representação *superficial*. Assim, para efeito da transcrição automática de grafema para fonema, consideramos os grafemas como a forma subjacente e os fonemas como a forma superficial. Ilustramos esse processo na Figura 1.



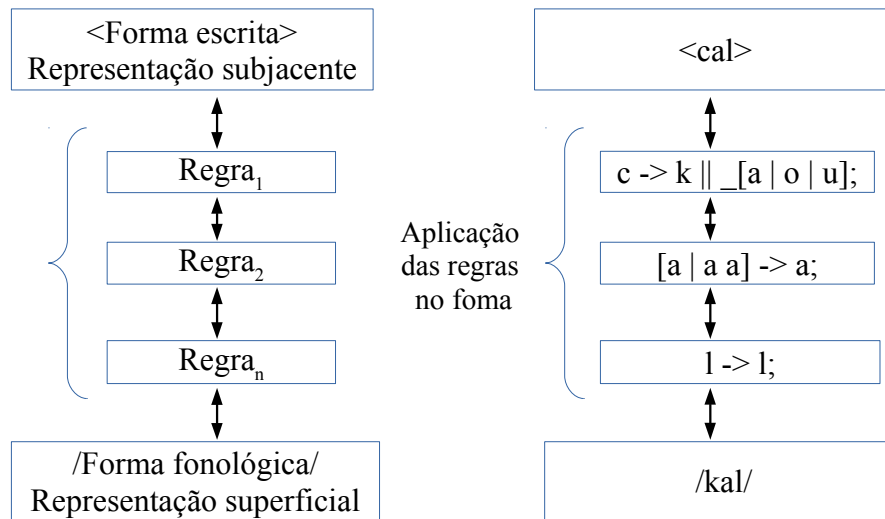


Figura 1: Modelo de análise e geração das representações superficial e subjacente.

Fonte: Elaborada pelos autores.

A partir da análise fonológica da língua portuguesa e de conhecimentos oriundos de trabalhos anteriores, um conjunto de regras foi compilado com o intuito de possibilitar a um computador o reconhecimento dos diferentes fonemas presentes no português brasileiro a partir dos grafemas, dos vocábulos inseridos.

Esse conjunto de regras pode ser classificado da maneira como segue: regras de transformação direta, regras de transcrição para dígrafos consonantais, regras de transcrição de consoantes com regras complexas, regras de transcrição de vogais e regras de transcrição de palavras em casos de neutralização fonêmica.

#### 4.3.1 Regras de transformação direta

Na implementação do sistema, percebemos que alguns grafemas, como mostra o Quadro 2, sempre são transcritos da mesma maneira, ou seja, possuem somente um respectivo fonema em qualquer contexto. Esses grafemas foram implementados por meio de regras de transformação direta, pois não exigiam contextos de mudança sonora. Os grafemas que permaneceram com a relação direta entre as letras e os sons foram: <p>, <b>, <d>, <t>, <v>, <j>, <f>, <ç>, que sempre são transcritos, respectivamente, por: /p/, /b/, /d/, /t/, /v/, /ʒ/, /f/, /s/. Ressaltamos que as consoantes <t> e <d> são representadas, em algumas variedades do português, como [tʃ] e [dʒ] quando vêm precedido da vogal /i/, como nas palavras <tia> [tʃia] e dia [dʒia], respectivamente. No entanto, a transcrição fonológica do sistema não considera essas variações linguísticas desse segmento.

As regras para essas letras podem, então, ser representadas como se segue no Quadro 2.

Quadro 2: Transcrição de grafemas com regras simples.

Regras de transformação	Aplicação no Foma
<p> corresponde ao fonema /p/	p -> p;
<b> corresponde ao fonema /b/.	b -> b;
<d> corresponde ao fonema /d/.	b -> d;
<t> corresponde ao fonema /t/.	t -> t;
<v> corresponde ao fonema /v/.	v -> v;
<j> corresponde ao fonema /ʒ/.	j -> ʒ;
<f> corresponde ao fonema /f/.	f -> f;
<ç> corresponde ao fonema /s/	ç -> s;
<h> corresponde ao zero fonêmico.	h -> 0;

Fonte: Elaborado pelos autores.

### 4.3.2 Regras de transcrição para os dígrafos consonantais

O dígrafo é um fonema representado por duas letras em um vocábulo. A segunda letra existe apenas para auxiliar a primeira em uma determinada pronúncia (PESTANA, 2013). Desse modo, essas duas letras serão representadas apenas por um único símbolo fonético. Pela análise realizada, nota-se que alguns dígrafos consonantais possuem, também, regras simples, como é o caso do <ch>, cujo fonema é sempre representado pelo arquifonema /S/, do <sc>, <ss> e <sç>, ambos representados pelo fonema /s/, e também do <rr>, cuja a transcrição é representada pelo arquifonema /R/ em todos os contextos.

Outros dígrafos consonantais, entretanto, não possuem regras tão simples, uma vez que podem possuir sons diferentes em certas ocasiões, como é o caso do <gu>, <qu>, <lh>, <nh> e <xc>. Salienciamos, com base em Silva (2014), que as sequências tradicionalmente denominadas “tritongos” são transcritas como uma sequência de oclusiva velar-glide seguida de um ditongo decrescente: <quais> /'kwals/.

Para esses, podemos generalizar as regras de modo que contemple corretamente a maioria dos casos. Tais regras foram então resumidas no Quadro 3.

Quadro 3: Transcrição de dígrafos consonantais.

Regras de transformação dos dígrafos	Aplicação no Foma
<ch> corresponde ao fonema /ʃ/.	c h -> ʃ;
<sc>, <ss> e <sç> corresponde ao fonema /s/.	[s c   s s   s ç] -> s;
<rr> corresponde ao arquifonema /R/.	[r r] -> R;
<gu> corresponde ao fonema /g/, quando precedido das letras <e> ou <i>.	[g u] ->    _ [e   i];
<qu> corresponde ao fonema /k/, quando precedido das letras <e> ou <i>.	[q u] -> k    _ [e   i];
<lh> corresponde ao fonema /l/.	[l h] -> l    _ V;
<nh> corresponde ao fonema /ŋ/.	[n h] -> ŋ    _ \ [i a];
<xc> corresponde ao fonema /s/ quando antecede <e> ou <i>.	[x c] -> s    _ [e   i];

Fonte: Elaborado pelos autores.

### 4.3.3 Transcrição de consoantes com regras complexas

Algumas consoantes podem representar vários fonemas na língua portuguesa, como, por exemplo, o grafema <x>, que representa quatro fonemas na nossa língua, tal como ocorre em: <táxi>, <faixa>, <execução> e <próximo>. De fato, nesses casos, a letra <x> pode representar os sons de /ks/, /ʃ/, /z/ e /s/, respectivamente. Para que o programa possa executar, devidamente, a transcrição para esses fonemas, alguns padrões mais complexos foram inseridos. Contudo, para algumas palavras, devido a sua etimologia, mesmo essas regras não se aplicam. De modo genérico, pode-se atribuir as regras do Quadro 4 para a maioria das palavras.

Quadro 4: Transcrição de consoantes com regras complexas.

Regras de transformação das consoantes	Exemplo	Aplicação no Foma
<x> corresponde ao fonema /ʃ/.	<caixa>	x -> ʃ    [e n] _ , .#, [a i] e [i] o u] _ ;
<x> corresponde ao fonema /z/ quando sucede a vogal média <e> em início de palavras e antecede outra vogal.	<exame>	x -> z    [.#, e] _ V;
<x> corresponde ao dífono /ks/ nas outras situações.	<táxi>	x -> ks;
<c> corresponde ao fonema /s/ quando antecede /e/ ou /i/.	<cedo>	c -> s    _ [e   i];
<c> corresponde ao fonema /k/ nas demais situações.	<casa>	c -> k    _ [a   o   u];
<m> e <n> correspondem, respectivamente, aos fonemas /m/ e /n/, quando vem depois de uma vogal.	<mala> <nada>	m -> m, n -> n    _ V;
<r> corresponde ao arquifonema /R/, quando inicia palavra e quando sucede as letras <l>, <n> e <s>.	<retrato>	r -> R    .#, _ , [l   n   s] _ ;
<r> corresponde ao fonema /-r/, quando se situa entre vogais ou quando é a segunda letra nos encontros consonantais prevocálicos.	<prato>	r -> -r    V _ V, [b   c   d   f   g   p   t   v] _ ;
<l> corresponde ao fonema /l/.	<calma>	l -> l;
<s> corresponde ao fonema /s/ quando em início de palavra, ou quando sucede <n>, <l> ou <r>.	<sino>	s -> s    .#, _ , [n   l   r] _ ;
<s> corresponde ao fonema /z/, quando está entre vogais.	<coser>	s -> z    V _ V;
<g> corresponde ao fonema /ʒ/, quando antecede as vogais <e> ou <i>.	<gelo>	g -> ʒ    _ [e   i];
<g> corresponde a fonema /g/ quando antecede <a>, <o>, <r>, <l>.	<galo>	g -> g    _ [a   o   r   l]

Fonte: Elaborado pelos autores.

#### 4.3.4 Transcrição de vogais

Algumas vogais possuem uma transcrição fonológica simples, por exemplo, as vogais <a>, <i> e <u> representam, respectivamente, os fonemas /a/ /i/ e /u/ em qualquer situação. Em contrapartida, os grafemas <e> podem corresponder aos fonemas /e/ ou /ɛ/, assim como a letra <o> pode representar /o/ ou /ɔ/. Infere-se, ainda, que a presença desses fonemas se dá, praticamente, de forma aleatória, de modo que não foi possível definir uma regra eficiente para distingui-los, o que torna a transcrição das vogais <e> e <o> um grande desafio para o propósito do programa. De fato, esse problema apenas poderá ser solucionado com o desenvolvimento de um etiquetador, ou seja, um programa que insere etiquetas morfossintáticas em palavras. Simplificando, pode-se aplicar as regras do Quadro 5 a um certo número de casos.

Quadro 5: Transcrição de vogais.

Regras de transformação para as vogais	Aplicação no Foma
<a> ou <á> corresponde ao fonema /a/.	[a   á] -> a;
<e> ou <ê> corresponde ao fonema /e/.	[e   ê] -> e;
<é> corresponde ao fonema /ɛ/, exceto quando ocorre antes de <m> ou <n>.	é -> ɛ    _\[m] n];
<i> ou <í> corresponde ao fonema /i/.	i -> b;
<o> ou <ô> corresponde ao fonema /o/.	o -> o;
<ó> corresponde ao fonema /ɔ/, exceto quando ocorre antes de <m> ou <n>.	ó -> ɔ    _\[m] n];
<u> ou <ú> corresponde ao fonema <u>.	u -> u;

Fonte: Elaborada pelos autores.

#### 4.3.5 Transcrição de palavras em casos de neutralização fonêmica

Quando há uma perda de contraste fonêmico, isto é, quando dois ou mais fonemas perdem a distinção entre si, temos uma neutralização fonética (SEARA, 2015). Para representar esses casos, representamos esse fenômeno por arquifonemas, tais como o /S/ para representar a neutralização de fonemas como o /s/, /z/, /ʒ/; /N/ para representar a perda de contraste entre fonemas que representam a nasalização pós-vocálica e o /R/, como bem destaca Silva (2014), em todos os dialetos<sup>3</sup> do português falado no Brasil, há distinção fonêmica entre a vibrante simples /r/ e a vibrante múltipla /r/ em posição intervocálica. Os segmentos fonéticos das vibrantes múltiplas podem ocorrer em posição pós-vocálica como uma consoante fricativa [h], [χ], [x], [ħ] ou retroflexa /ɻ/. Por isso, utilizamos o arquifonema /R/ para denotar a neutralização existente no contraste fonético desses segmentos, como destacamos nos Quadros 3 e 4 e retomamos no Quadro 6.

3 Nesse trabalho, os termos *dialeto* e *falar* são sinônimos de *variedade linguística*, ou seja, referem-se ao falar característico de determinado grupo social e/ou regional. (Ver COELHO et al, 2015, p. 15).

Quadro 6: Transcrição de neutralização fonêmica.

Regras de transformação para os arquifonemas	Aplicação no Foma
<m> ou <n> são representados pelo arquifonema /N/, quando está situado depois de vogal.	[m   n] -> N    V_ \V, V_.#.;
<r> ou o dígrafo <rr> são representados pelo arquifonema /R/, quando estão em fim de sílaba ou entre vogais, respectivamente.	r -> R    V_.#., _ . ; [r r] -> R    V_V;

Fonte: Elaborada pelos autores.

Como mostrado em alguns itens dos Quadros 4, 5 e 6, a transcrição fonológica automática do sistema considera o fato de que certos fonemas perdem o contraste fonológico em ambientes específicos. É o caso dos arquifonemas /R/ e // que expressam a neutralização de um ou mais fonemas em um contexto específico. Destacamos que o arquifonema // representa a manifestação escrita da letra <l> como uma consoante lateral alveolar, quando em início de sílaba ou quando precedido por uma consoante na mesma sílaba, como nas palavras <leve> e <planta>, respectivamente. No entanto, quando falada, nesse fonema ocorre uma variação em posição final de sílaba; pode apresentar como uma lateral alveolar velarizada ou como vocalização do fonema (SILVA, 2014). Para todas essas situações envolvendo esse fonema, o sistema de transcrição faz a representação por meio do arquifonema //, como mostramos no Quadro 4. Dessa forma, na transcrição fonológica automática dessa letra, ocorre a neutralização em relação aos falares para essa consoante. A mesma lógica de aplicação dessa regra foi utilizada para a representação da letra <r>, nos casos em que ela manifesta-se como um tepe ou vibrante simples /r/, quando vem em posição intervocálica ou quando está precedida por uma consoante na mesma sílaba. Já nos demais casos essa letra é transcrita por meio do arquifonema /R/.

## 5 Resultados e avaliações

A avaliação de um modelo computacional tem como objetivo principal decidir se o sistema transcreve com precisão um determinada palavra escrita no padrão fonológico e o resultado da avaliação é importante para compreendermos até que ponto o modelo é confiável, para que fins utilizá-lo e, também, para guiar o desenvolvedor na busca de melhorias futuras para o modelo, conforme explicam Bird, Klein e Loper (2009). Pensando assim, apresentamos o processo de avaliação e os resultados obtidos do sistema de transcrição fonológica.

Ressaltamos que esse sistema está integrado em um programa chamando *Bag of Tools*, o qual contém outros sistemas para o processamento de linguagem natural, como *wordlist*, separador silábico e o *Grapheme to Phoneme* que executa a transcrição fonológica, como se pode ver na Figura 2.

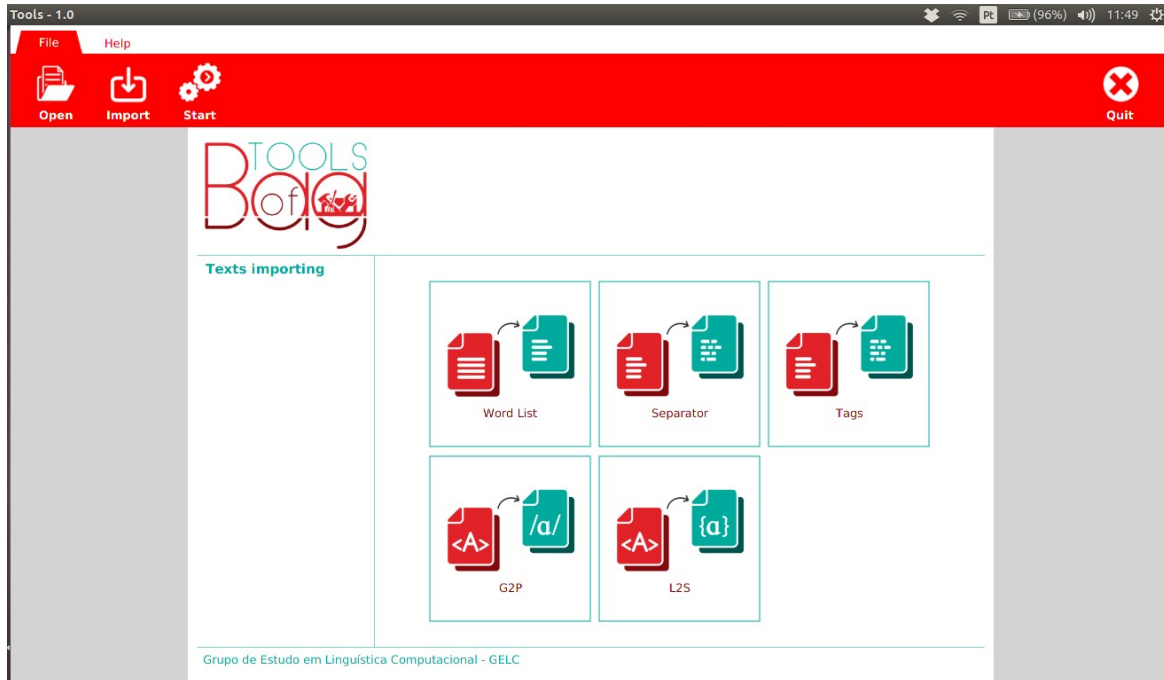


Figura 2: Interface inicial do programa *Bag of Tools*.  
Fonte: Elaborada pelos autores.

Para a avaliação, e conseqüentemente observação dos principais erros do sistema, foi separada uma amostra de mil palavras do *Corpus de Jornais Potiguares (C-Poti)*. Esse *corpus* está sendo desenvolvido junto ao Grupo de Estudo em Linguística Computacional (GELC) da Universidade Federal Rural do Semi-Árido (UFERSA) e é constituído por 78 textos escritos coletados criteriosamente dos jornais *online* do estado Rio Grande do Norte.

Depois disso, o programa foi compilado para um arquivo binário, reconhecido pelo *Foma*, o qual pode ser executado por linha de comando; contudo, pode ser exportado para algumas linguagens de programação e executado por interface gráfica, como podemos ver na Figura 3, que é um modo mais intuitivo e prático.

Utilizando a interface gráfica da *Bag of Tools*, o usuário pode escolher entre a execução do sistema de um único arquivo ou de um conjunto de arquivos. Para isso, basta optar pelo botão *open* na barra de menu, se quiser abrir um arquivo no *input* do sistema ou escolher o botão *import* na mesma barra, para executar o sistema com vários arquivos. Os arquivos ficam dispostos na barra vertical *text imports* no qual o usuário poderá executar uma das ferramentas que se encontram na barra de sistemas.

Após a solicitação de execução, o programa retorna para o usuário o resultado das palavras em duas colunas: na primeira coluna, estão presentes todas palavras, não repetidas, dos textos selecionados para a execução e, na segunda coluna, estão presentes as transcrições das palavras nos símbolos do Alfabeto Fonético Internacional, como mostra a Figura 3.

Feito isso, fizemos a avaliação do sistema utilizando a acurácia – a métrica comumente aplicada para avaliar os sistemas de transcrição fonológica, como podemos ver em Veiga, Candeias e Perdigão (2011) e Barros e Weiss (2006). A acurácia é a métrica mais simples que pode ser usada para avaliar um G2P e medir os percentuais de

entrada no conjunto de teste que o sistema transcreveu corretamente (BIRD, KLEIN e LOPER, 2009). Essa métrica calcula uma pontuação para um modelo, comparando as entradas em um conjunto de teste com os símbolos corretos para essas entradas, ou seja, a tarefa fundamental desse método é a classificação binária dos resultados obtidos.

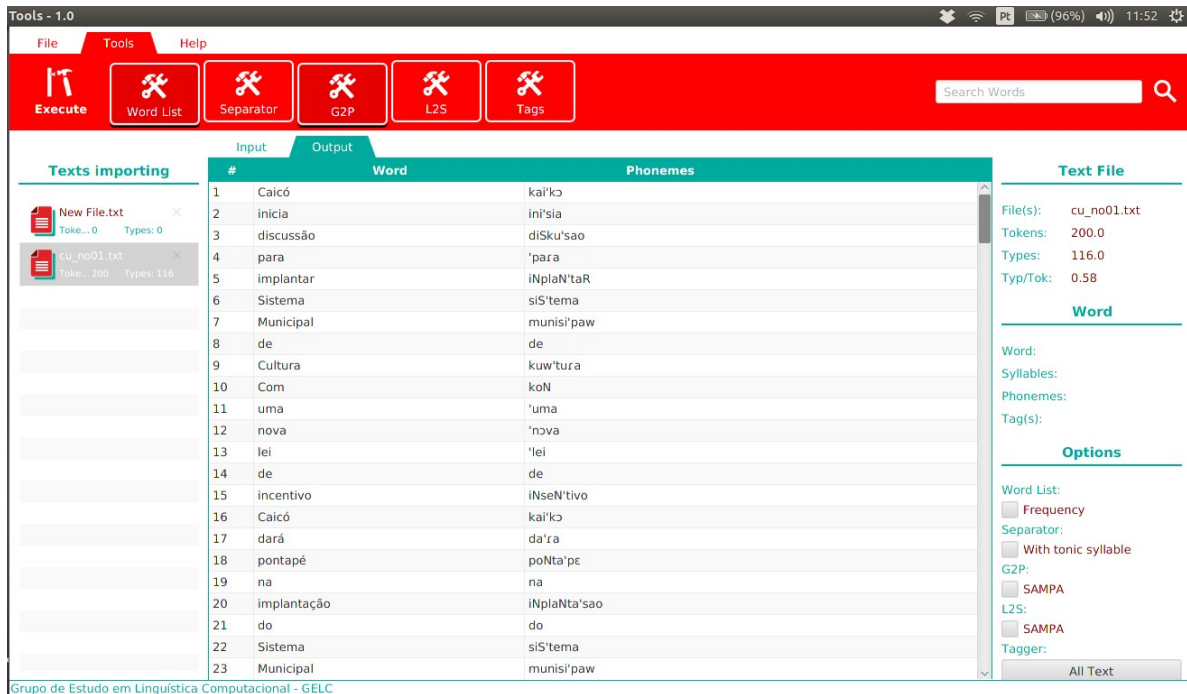


Figura 3: Execução do sistema de transcrição fonológica.  
Fonte: Elaborada pelos autores.

O sistema de transcrição de grafema para fonema da língua portuguesa que apresentamos consegue transcrever muitas palavras da língua portuguesa com eficiência, em testes, cerca de 90% de acurácia nas mil palavras utilizadas que foram transcritas aceitavelmente pelo software. Contudo, de modo inerente, ele comete alguns equívocos com alguns vocábulos.

No Quadro 7, apontamos os principais erros encontrados na transcrição da amostra.

Quadro 7: Principais erros encontrados na transcrição de grafema para fonema do português brasileiro.

Palavras	Transcrição	Tipo de erro
<pelos>	/'peIoS/	Transcrição da vogal <e>
<seja>	/'seʒa/	
<desde>	/'deSde/	
<ter>	/'teR/	
<prender>	/preN'deR/	
<por>	/pɔR/	Transcrição da vogal <o>
<morreu>	/'mɔReu/	
<socorro>	/so'kɔRo/	
<fossem>	/'fɔseN/	

<força>	/fɔRsa/	Transcrição da consoante <x>
<México>	/mɛksiko/	
<aproximando>	/apɔksi'maŋdo/	
<auxiliar>	/auksili'aR/	
<proximidades>	/pɔksimi'dadeS/	
<máximo>	/maksimo/	

Fonte: Elaborado pelos autores.

Como é possível observar no Quadro 7, os principais erros se devem à transcrição das vogais <e>, <o> e da consoante <x>. Desse modo, o sistema de transcrição por regras mostra-se pouco eficiente para o caso desses grafemas, visto que o sistema verbal muitas vezes não utiliza de padrões lógicos na composição de palavras, o que torna a transcrição dessas letras, de certa maneira, imprevisível para um computador. Para esses casos, como já discutido, há a necessidade de outros *softwares* auxiliares que auxiliem na transcrição.

## 6 Considerações finais

A partir dos resultados percebemos que o programa ainda apresenta alguns erros que devem ser reduzidos para que a transcrição fonológica seja mais eficaz. Todavia, o conversor G2P é apenas um dos sistemas que pertence ao programa *Bag of Tools*, no qual serão acrescentadas outros sistemas que auxiliarão no funcionamento completo da transcrição fonológica automática. Além disso, o sistema construído demonstra que é um *software* conveniente de transcrição fonológica, para um grande número de vocábulos do português brasileiro. Entretanto, ainda possui erros, cuja implementação ainda não foi feita, mas será tratada em projetos futuros. Assim, os resultados mostraram que o desempenho do G2P foi de 90% de acurácia. Isso constitui boa performance para um sistema ainda em aprimoramento.

Ressaltamos que, ao final do processo, o programa foi compilado para um arquivo binário, reconhecido pelo *Foma*, o qual pode ser executado por linha de comando. Contudo, pode ser exportado para algumas linguagens de programação e executado por interface gráfica, um modo mais intuitivo e prático.

Destacamos também que esse trabalho contribuirá para o desenvolvimento tecnológico de outras ferramentas que o grupo desenvolverá na área da linguística computacional. Além disso, espera-se, também, que ele contribua no aperfeiçoamento do algoritmo desenvolvido. Como resultado desses estudos, o algoritmo do transcritor fonológico foi, então, desenvolvido e demonstra um desempenho favorável, realizando corretamente a transcrição fonológica da grande maioria dos vocábulos da língua portuguesa.

## Referências

BARROS, M. J.; WEISS, C. Maximum Entropy Motivated Grapheme-To-Phoneme, Stress



and Syllable Boundary Prediction for Portuguese Text-to-Speech, *IV Jornadas en Tecnologías del Habla*, 2006, p. 177-182. Zaragoza, España. Disponível em: <[http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/finals/4jth\\_127.pdf](http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/finals/4jth_127.pdf)>. Acesso em: 09 de agosto 2015.

BECHARA, E. *Moderna gramática portuguesa*. 38. ed. Rio de Janeiro: Lucerna, 2005.

BEESELEY, K. R.; KARTTUNEN, L. *Finite-State Morphology: Xerox Tools and Techniques*, 2002.

BIRD, S.; KLEIN, E.; LOPER, E. Learning to classify text. In: \_\_\_\_\_. *Natural language processing with python*. United States of America: O'Reilly, 2009, p. 221-257. Disponível em: <<http://www.nltk.org/book/>>. Acesso em: mai. 2012.

BRAGA, D.; COELHO, L.; RESENDE Jr., F. G. V. A Rule-Based Grapheme-to-Phone Converter for TTS Systems in European Portuguese, *VI Int. Telecommunications Symposium*, Fortaleza-CE, Brazil, 2006. p. 976-981.

CARVALHO, C. I. C. *Transdutor de estados finitos para conversão de grafema para a pronúncia da variedade linguística potiguar*. 2016. 160 f. Tese (doutorado em Linguística) – Universidade Federal do Ceará, Centro de Humanidades, Departamento de Letras Vernáculas, Fortaleza, 2016.

CARVALHO, C. I. C. Conversor de transcrição fonética automática para as formas linguísticas da variedade linguística potiguar. *Domínios de Linguagem*, [s.l.], v. 11, n. 3, p. 733-752, 30 jun. 2017. EDUFU. <http://dx.doi.org/10.14393/dl30-v11n3a2017-13>. Disponível em: <<http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/37277/20915>>. Acesso em: 10 ago. 2017.

CHOMSKY, N.; HALLE, M. *The sound pattern of english*. New York: Harper e Row, 1968.

HULDEN, M. Finite-State Syllabification. In: HULDEN, M. YLI-JYRÄ, A.; KARTTUNEN, L.; KARHUMÄKI, J. *FSMNLP 2005*, LNAI 4002, 2006, p. 86-96.

HULDEN, M. Foma: a finite-state compiler and library. In: CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 12., 2008, Atenas. *Proceedings...* Atenas: Eacl, p. 29-32, 2008. Disponível em: <[http://dingo.sbs.arizona.edu/~mhulden/hulden\\_foma\\_2009.pdf](http://dingo.sbs.arizona.edu/~mhulden/hulden_foma_2009.pdf)>. Acesso em: 15 ago. 2013.

JARGAS, A. M. *Expressões Regulares: uma abordagem divertida*. Novatec Editora, 2006.

SEARA, I. C.; NUNES, V. G.; LAZZAROTTO-VOLCÃO, C. *Fonética e fonologia do português brasileiro*. Editora Contexta, 2015.

SILVA, T. C. *Fonética e fonologia do português*. 10. ed. São Paulo: Contexto, 2014.

TEIXEIRA, A.; OLIVEIRA, C.; MOUTINHO, L. On the Use of Machine Learning and Syllable Information in European Portuguese GraphemePhone Conversion, *Proc. PROPOR 2006*, 2006. p. 212-215.

VASILÉVSKI, V. *Construção de um sistema computacional para suporte à pesquisa em fonologia do português do Brasil*. 2008. 166f. Tese de doutorado - Pós-graduação em Linguística da Universidade Federal de Santa Catarina, 2008.

VEIGA, A.; CANDEIAS, S.; PERDIGÃO, F. Conversão de Grafemas para Fonemas em Português Europeu – Abordagem Híbrida com Modelos Probabilísticos e Regras Fonológicas. *Linguamática*, v. 3, nº 1, 2, p. 39–51, dez. 2011.

Recebido em 31 de agosto de 2017.  
Aprovado em 18 de dezembro de 2017.