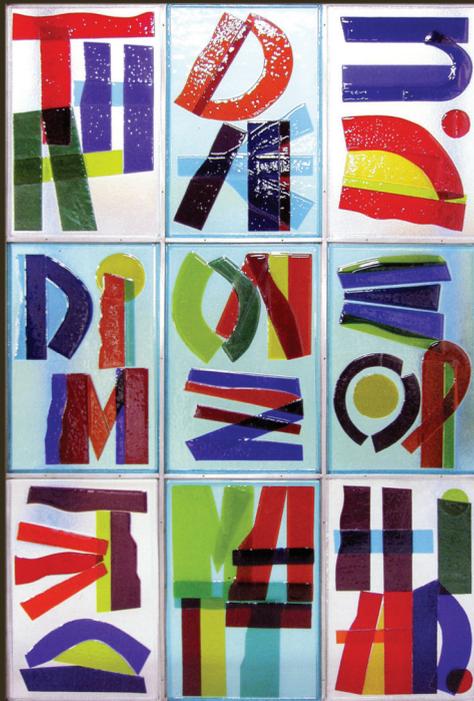


ESTUDOS DE LINGUÍSTICA

VOLUME I

ANA R. LUÍS
COORD.



IMPRESA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

Sara Candeias¹, Fernando Perdigão^{1,2}

¹ Instituto de Telecomunicações, Polo de Coimbra, DEE-UC

² Departamento de Engenharia Eletrotécnica e de Computadores da UC

INTEGRAÇÃO LINGUÍSTICA EM SISTEMAS DE CONVERSÃO DE GRAFEMA PARA FONE(MA)

1. Introdução

Sendo uma das áreas de estudo da linguística teórica, a fonética-fonologia tem como objetivo de estudo a descrição e descodificação da fala de uma língua. O estudo da fala de uma língua implica obrigatoriamente conhecer o conjunto de elementos sonoros que a formam, assim como as possibilidades de combinação desses elementos. É do setor fonético o conhecimento de que, para falar, fazemos uso de consoantes e de vogais; de que as vogais podem ser orais, nasais ou até nasaladas; de que, em português europeu (PE), a lateral <l> a fechar sílaba é pronunciada como velarizada (como em <mel>); de que as vogais posteriores fechadas são reduzidas em posição final de palavra antecidas de consoante bilabial (como em <rumo> ou <campo>); ou de que um simples <e> em contexto inicial de vocábulo e de sílaba fechada por <s> (como em <esperança>) pode ser pronunciado como [ə], como [i], ou até nem ser pronunciado ([∅]) (cf., a propósito desta multirrealização, os estudos de (Andrade 1994) e de (Candeias *et al.* 2011)). É do setor fonológico distinguir as características dos sons que diferenciam significados numa língua, sendo por isso possível afirmar, a título de exemplo, que as vibrantes múltipla e simples são, em PE, entidades fonológicas distintas em contexto intervocálico, uma vez que opõem significados (como é o que acontece no binómio <carro> vs. <caro>).

Ora, uma vez que o módulo de conversão de Grafema em Fone(ma) tem sido, à parte algumas flutuações de estruturas, um bloco essencial da arquitetura geral de sistemas de tecnologias da fala, e que, para o seu

desenvolvimento, aflui naturalmente o conhecimento linguístico, tornou-se premente a necessidade de se estreitar a relação entre a ciência da língua e a engenharia do processamento linguístico. Assim, partir do conhecimento linguístico em direção ao seu processamento e integração em sistemas automáticos quer de reconhecimento (ASR, do inglês *Automatic Speech Recognition*) quer de síntese de voz (TTS, do inglês *Text-to-Speech*)¹⁰⁸ trouxe a necessidade de se estender o estudo da língua a um novo paradigma: o da engenharia do processamento linguístico ou da linguística tecnológica.

Encontramo-nos, obviamente, perante uma área complexa onde saberes clássicos entram em confluência e que tem crescido a par da revolução tecnológica a que temos vindo a assistir. Vivemos numa atualidade que não concebe um quotidiano sem acesso à internet ou sem o recurso ao uso de computador ou de telemóvel. Cada vez com mais frequência, somos atendidos por uma voz artificial que nos dá instruções quando telefonamos para um serviço público ou comandamos com a nossa própria voz aplicações no telemóvel ou no computador. Tem sido a consciência de que as tecnologias linguísticas se destinam a facilitar o diálogo entre Homem-Máquina que tem impulsionado o desenvolvimento nesta área, ampliando soluções tecnológicas com interface de voz adaptadas à língua portuguesa, falada em todo o mundo por 240 827 milhares de pessoas¹⁰⁹.

Este artigo vem no sentido de descrever como o conhecimento fornecido pela linguística teórica tem sido (e pode ser) processado e integrado em sistemas automáticos de processamento do português, nomeadamente no módulo de conversão de Grafema-Fone(ma). Em simultâneo, o problema da conversão é caracterizado e, na secção 2, algumas questões relacionadas com opções terminológicas e formais no que diz respeito à tarefa da conversão são levantadas. A secção 3 descreve as abordagens que têm sido adotadas no desenvolvimento de conversores, incluindo as que resultam de

¹⁰⁸ Um sistema de síntese de fala (TTS) é um modelo computacional que permite gerar fala automaticamente a partir de uma entrada de texto. Ao invés, um sistema de reconhecimento de fala (ASR) identifica as unidades dessa língua a partir de um sinal de voz, construindo estruturas vocabulares.

¹⁰⁹ Número obtido no sítio do Observatório da Língua Portuguesa, <http://observatorio-lp.sapo.pt/pt/dados-estatisticos/falantes-de-portugues>.

opções específicas tomadas pelos autores. Conclusões e desafios para trabalhos futuros são apresentados na última secção, a secção 4.

2. Conversão Grafema-Fone(ma)

O módulo denominado Grafema-Fone(ma) ou G2P (do inglês *Grapheme-to-Phone(me)*¹¹⁰) tem por objetivo interpretar a grafia de uma dada palavra¹¹¹ e convertê-la numa sequência de símbolos sonoros, representativos de como essa palavra é pronunciada. Ainda que com alguma flutuação no que diz respeito às arquiteturas dos sistemas quer de ASR quer de TTS, o processo de conversão tem sido, regra geral, um bloco integrante desses sistemas, fazendo parte de uma tarefa mais ampla, vulgarmente conhecida como pré-processamento de texto (ou *front-end*), e à qual cabe o processo de decodificação e de normalização de elementos presentes nos textos (tais como são os numerais, as siglas, os estrangeirismos, as abreviaturas, os acrónimos e as siglas, entre outros), para além do fracionamento do texto em frases e em vocábulos.

183

Antes de se passar à descrição das abordagens existentes no processo de conversão do grafema na forma simbólica representativa da sua pronúncia, algumas questões devem ser levantadas, as quais se relacionam com as opções terminológicas e formais adotadas. Por merecerem uma ponderação mais cuidada, estas questões são, de seguida, exploradas.

2.1. Fonema ou Fone?

Como é do conhecimento geral no âmbito da linguística teórica, e retomando conceitos já abordados na introdução deste texto, a fonologia de uma língua é a fonética interpretada dessa língua, firmando-se o fonema como uma unidade distintiva que se reconhece por uma relação de contraste.

¹¹⁰ Por vezes, o conversor é também designado por L2S, do inglês *Letter to Sound*.

¹¹¹ Por palavra entende-se, aqui, todo o átomo do léxico português, contendo, pelo menos, um grafema ou dígito.

Dito de outro modo, o fonema de uma língua, sem valor distintivo próprio, é possuidor de um valor distintivo capital dentro da língua. O fonema é assim admitido como a contrapartida linguística, formal e abstrata, da realidade sonora e, podendo ser realizado por um fone ou mais (alofones), é muitas vezes interpretado como uma classe de alofones que se revelam num campo de distribuição específico (sobre a matéria, cf., entre outros, (Barbosa 1994) e (Mateus e Andrade 2000)) e a introdução teórica expressa em (Candeias 2007).

A clarificação terminológica dos termos fone e fonema não tem sido regular nos estudos que têm vindo a ser efetuados no âmbito do G2P, sendo frequente encontrar uma alternância não justificada entre eles (cf. (Caseiro *et al.* 2002) e (Barros e Weiss 2006), como exemplo). Admitindo as razões teóricas ligadas ao binómio fone/fonema, sempre que o resultado do conversor dá conta de fenómenos de pronúncias alternativas, é porque exprime o nível da superfície, atuando no âmbito da fonética (cf. conversor descrito em (Oliveira *et al.* 1992)). Quando o resultado do procedimento de conversão adotado admite valências do contexto mais ou menos alargado no âmbito do vocábulo, opera-se a um nível de maior abstração, atuando no âmbito da fonologia. Nestes casos, mais comuns e geralmente centralizados na pronúncia estandardizada do PE, a unidade para a qual o grafema é convertido é considerada como uma escolha significativa de entre todas as outras unidades que o sistema da língua coloca ao dispor. Estudos como o de (Veiga *et al.* 2011) atendem a essa clareza terminológica, apresentado sistemas que convertem o grafema numa entidade mais abstrata, o fonema. Atuando a este nível, abre-se a possibilidade de poder vir a ser integrada no sistema a capacidade de cada fonema ser realizado por um feixe de alofones, representativos de ambientes acústicos decorrentes de efeitos de coarticulação, de *iod* ou de *sandbi*, e de casos de multipronúnciação. Um caso comum de alofonia em PE é a possibilidade de pronúnciação em [e]¹¹², em [ei]¹¹³ ou em [ɔ], quando a vogal <e> se localiza em posição tónica, seguida de consoante palatal <j> (como em <igreja>) ou <lh> (como em <abelha>). A multirrealização do *schwa* é

¹¹² Notação em SAMPA. Sobre o assunto de notação, cf. secção seguinte.

¹¹³ É opção dos autores não notar glides. Sobre o assunto, cf. 2.3.

outro caso de alofonia a demarcar, em especial porque potencia significativas taxas de erro no reconhecimento de fones (Candeias *et al.* 2011). Ambos os exemplos de alofonia mencionados, evidenciando a coexistência de mais do que um sistema linguístico, tornar-se-iam obrigatoriamente lacunares em termos de representatividade da realidade da língua portuguesa, caso observassem apenas o nível de superfície. Uma perspectiva de representação fonética, nestes casos, anularia a possibilidade da existência de mais do que uma pronúncia. Por outro lado, a multirrealização possível do fonema dependente do contexto é um parâmetro que pode ser automaticamente extraído usando modelos acústicos dessas unidades (Veiga *et al.* 2010).

185

2.2. Alfabeto Fonético

Outra questão subjacente à conversão G2P diz respeito à seleção do sistema de notação, seja ela fonética ou fonológica. Neste ponto, tendo-se já esclarecido o campo de ação relacionado quer com a fonética quer com a fonologia, não é pertinente aqui continuar essa distinção, pelo que o que se afirmar em relação ao alfabeto fonético pode igualmente afirmar-se relativamente ao alfabeto fonológico. No mesmo sentido, a ‘matéria sonora’ compreende metaforicamente segmentos fonéticos e fonológicos.

Ainda que o português seja uma língua cuja ortografia tem uma base fonética-fonológica (Mateus 2006), ocorrências existem em que a correspondência entre grafia e respetiva representação sonora não é invariavelmente unívoca, devida a fatores diversos, muitas vezes advenientes do contexto grafemático (caso dos grafemas <r> e <u> em <português>, <rosa> ou <arguido>) ou dependentes do estatuto morfológico ou sintático (cf. 2.4). São várias as situações nas quais essa correspondência não é unívoca: situações em que a um grafema pode corresponder quer a sons distintos, quer a mais do que um som (Tabelas 1 e 2) ou em que a mais do que um grafema corresponde apenas um som (caso dos dígrafos, por exemplo - Tabela 3). Com vista a uma implementação computacional, todos estes casos têm que ser tratados, seja por forma de regras, seja por anotadores estatísticos (cf. secção 3), numa tentativa de tornar a relação entre o grafema e o som inequívoca.

Grafemas	Fonemas	Exemplos
	ó	cama → /kóma/
a	a	mala → /ma/
	ó~*	canto → /kó-tu/
á		pás → /pa/
à	a	às → /a/
ã	ó~	anão → /no~u~/
â	ó~*	atlântico → /dltíku/
	ó	câmara → /kómó/
e	ó	desenho → /d@só/
	e	aquele → /kél@/
	@	refere → /R@fEr@/
	E	pele → /pEl@/
	i	areal → /órial/
	e~*	vento → /ve~tu/
	i~*	visões → /vizo~is/
	ói*	extra → /óiStró/
	ó~*	votem → /votó~i~/
é	E	café → /kófE/
	ó~i-	contém → /kóntó~i~/
ê	ó	amêijoa → /ómóiZuó/
	e	você → /vóse/
	e~	pêndulo → /pe~dulu/
	ói	êxito → /óizitu/
	ó~i-ó~i~*	têm → /tó~i-ó~i~/
i, í	i	cima → /simó/, saí → /só/
	i~	cinco → /si~ku/, límpida → /li~pidó/
o	o	corpo → /kópu/
	O	copo → /kOpu/
	o~	conto → /ko~tu/
	u~	visão → /vizó~u~/
	u	porque → /pórk@/
ó	O	cópia → /kOpíó/
ô	o	avô → /vó/
	o~	cônsul → /ko~sul/
õ	o~	põe → /po~i~/
	o~i~	põem → /po~i-ó~i~/
u, ú	u	apura → /ópú/
	u~	túnel → /tunEl/
	∅	unto → /u~tu/, anúncio → /ónu~siu/
	∅	quente → /k t@/

Tabela 1 – Possíveis sons (fonemas) associados a cada grafema – caso de vogais. Atendendo a uma relação linear de ‘um para um’ entre grafemas e fonemas, situações em que o fonema é determinado pelo contexto estão indicadas por *. ∅ representa zero, isto é, uma não correspondência fonológica. A coluna à direita mostra exemplos de notação fonológica.

Grafemas	Fonemas	Exemplos
c	k	capa → /kap6/
	s	cedo → /sedu/
	ks	ficcional → /fiksiunal
	∅	actuar → /6tuar/ (<i>pré-AO</i>)
g	g	gatu → /gatu/
	Z	girafa → /Ziraf6/
l	l	lado → /ladu/
	L	malha → /maL6/
m	m	mal → /mal/
	u~*	apelam → /6pEl6~u~/
	i~*	votem → /vOt6~i~/
	∅	campo → /k6~pu/
n	n	ana → /6n6/
	i~*	tens → /t6~i~S/
	J*	manha → /m6J6/
	∅	canto → /k6~tu/
p	p	par → /par/, apto → /aptu/
	∅	óptica → /Otik6/ (<i>pré-AO</i>)
r	r	caro → /karu/
	R	carro → /kaRu/
s	s	massa → /mas6/
	S	às → /aS/
	z	casa → /kaz6/
	Z	abismo → /6biZmu/
x	s	máximo → /masimu/
	S	xadrez → /S6dreS/
	z	exame → /ez6m@/
	Z	ex-bar → /6iZbar/
z	ks	fixo → /fiksu/
	S	arroz → /6RoS/
	z	azar → /6zar/
	Z	felizmente → /f@liZme~t@/

Tabela 2 – Continuação da Tabela 1 – caso de consoantes.
A expressão (*pré-AO*) diz respeito à grafia adotada previamente
ao estabelecimento do Acordo Ortográfico de 1990.

Relativamente ao modo de representação da matéria sonora, no módulo do conversor, o alfabeto SAMPA (*Speech Assessment Methods Phonetic Alphabet*) e a sua extensão X-SAMPA (*The Extended Speech Assessment Methods Phonetic Alphabet*) têm sido usados, com mais frequência, em detrimento do Alfabético Fonético Internacional (AFI ou IPA – *International Phonetic Alphabet*). Esta opção assenta numa atitude de otimização dos recursos, uma vez que o

SAMPA recorre aos caracteres disponíveis no teclado dos computadores (Wells 1997). Neste artigo, o alfabeto usado é o SAMPA, salvo uma exceção de pormenor onde se recorre igualmente ao IPA pela necessidade de comparação e de esclarecimento das representações sonoras. O SAMPA disponibiliza um alfabeto dirigido para a língua portuguesa – *SAMPA for Portuguese*

Dígrafos	Fonemas	Exemplo
cc	s	perfeccionismo → /p@rfEsiuniSmu/ (<i>pré-AO</i>)
	ks	ficcional → /fiksiunal/
cç	s	intersecção → /i~t@rsEs6~w~/ (<i>pré-AO</i>)
	ks	ficção → /fiks6~w~/
rr	R	carro → /kaRu/
ss	s	massa → /mas6/
lh	L	molho → /moLu/
nh	J	unha → /uJ6/
ch	S	chave → /Sav@/
ou	o	dourada → /dorad6/
an, am	6~	canto → /k6~tu/
		angolano → /6~gul6nu/
		campo → /k6~pu/
en, em	e~	sente → /se~t@/
		ensaio → /e~saiu/
		sempre → /se~pr@/
in, im	i~	cinco → /si~ku/
		limbo → /li~bu/
on, om	o~	conto → /ko~tu/
		pomba → /po~b6/
un, um	u~	assunto → /asu~tu/
		tumba → /tu~b6/
ân, âm	6~	pântano → /p6~t6nu/
		iâmbico → /i6~biku/
ên, êm	e~	ênfase → /e~f6z@/
		êmbolo → /e~bulu/
ín, ím	i~	índio → /i~diu/
		límpido → /li~pidu/
ôn, ôm	o~	cônsul → /ko~sul/
		cômputo → /ko~putu/
ún, úm	u~	denúncia → /denu~si6/
		cúmplice → /ku~plis@/

Tabela 3 – Lista de dígrafos associados a fonemas possíveis e exemplos convencionados. Os primeiros 7 símbolos representam consoantes; os restantes 11 representam seqüências específicas de vogal ditongada e de vogais nasais. De observar que as duas primeiras representações de dígrafos podem constituir-se como uma estrutura alternativa do português grafado atendendo às normas listadas pelo Acordo Ortográfico de 1990.

(SAMPA-PT) (disponível em <http://www.phon.ucl.ac.uk/home/sampa/portug.htm>), acerca do qual merece ser tecida uma apreciação. Na verdade, a observação atenta dos alfabetos fonéticos SAMPA-PT e X-SAMPA dá conta de alguma indefinição de regularidade, exemplificada na atribuição de mais do que um símbolo para o mesmo som. A título de exemplo, o símbolo [r] no SAMPA-PT parece ter como correspondente no X-SAMPA o símbolo [4] (IPA: [r]), simbolizando o [r] no X-SAMPA a vibrante alveolar múltipla (IPA: [r]).

189

2.3. Notação

Independentemente do alfabeto usado na conversão, há que ponderar sobre a forma de notação no sentido de otimizar o conversor. Por vezes, poderá ser necessário afinar a notação até um nível que poderá admitir o processo de *iod* e de *sandhi* (exemplificadas em <extra> [6iStr6] e em <a água> [6 i agu6]), numa tentativa de aproximação aos efeitos sonoros provenientes de coarticulação, ou a não consideração da velarização da lateral (presente em <sal>), caso se considere que o contexto fonético é suficiente para categorizar a unidade. Outra situação em termos de notação e que não tem sido consensual é a que se relaciona com a questão dos hiatos, uma vez que as vogais <e>, <i>, <o> e <u> podem ser notadas como parte dos chamados ditongos crescentes (ou falsos ditongos), presentes em sequências como em <Leonor>, <relógio>, <póvoa> e <água>. A decisão da notação de glides, em especial nas sequências que formam os ditos ditongos crescentes, tem vindo a ser ponderada pela necessidade de caracterização do ritmo de fala, e, ainda que a questão tenha sido já levantada na literatura da área (cf., a título de exemplo, o estudo de (Rua *et al.* 2005)), não existe, tanto quanto é do nosso conhecimento, um estudo acústico, sustentado num corpus expressivo do PE falado, que suporte a distinção clara entre as vogais [i] - [j] e [u] - [w]. Uma vez que o contexto fonético é esclarecedor da estrutura da sequência, a nossa opção metodológica recai em não considerar pertinente a identificação de ditongos crescentes nem a notação das vogais ditas decrescentes do ditongo como glides. Admitimos, no entanto, que não é um formato de representação comum, e que pode eventualmente suscitar alguma discussão em termos da teoria fonológica.

2.4. Alternâncias Vocálicas

190 Um dos problemas com o qual o desenvolvimento de um conversor de G2P se tem vindo a deparar diz respeito à situação em que uma vogal, envolvida pelo mesmo contexto, pode apresentar uma alternância de pronúncia. Em PE, referimo-nos especificamente à alternância entre os pares /e/-/E/ e /o/-/O/. Casos de homógrafos heterófonos, nos quais as vogais <e> e <o> podem ser pronunciadas como /e/ e /E/ ou /o/ e /O/, não têm sido de fácil resolução, como já se mostrou no estudo de (Braga e Marques 2007). <selo> ou <namoro>, em estruturas frásicas tais como <Eu **selo** a carta com um **selo** comemorativo.> e <O **namoro** dos dois vem já de longa data, ao passo que eu **namoro** apenas há um mês.>, são disso exemplos. Poder-se-ia implementar um módulo que identificasse a categoria gramatical dos homógrafos, estratégia que desambiguaria ocorrências como <selo> e <namoro> nos exemplos, associando-se à categoria gramatical a pronúncia esperada. No entanto, exemplos como <lobo> em estruturas frásicas tais como <Ele caiu na boca do **lobo**.> e <O **lobo** frontal relaciona-se com a inibição de comportamentos.> mostram que homógrafos com a mesma categoria gramatical (ao caso, a do nome) podem ser igualmente heterófonos (de pronúncias /lobu/ e /lObu/, respetivamente, nos exemplos). Outra situação de alternância vocálica a referir ocorre em vocábulos que, do ponto de vista diacrónico, sofreram metafoia, alterando a pronúncia da vogal tónica por influência da vogal átona final, num processo de assimilação regressiva. Assim, a vogal <o> é pronunciada como /o/ em <miolo> e em <forno> e como /O/ em <miolos> e em <fornos>, independentemente de, na palavra latina primitiva, se pronunciar com o chamado ‘o breve’.

3. Abordagens Teóricas

Várias têm sido as abordagens teóricas adotadas no sentido de se encontrar uma solução para a tarefa da conversão G2P. No que diz respeito ao PE, destacamos as seguintes:

- por regras inferidas a partir dos dados (Teixeira *et al.* 2006);
- por máquinas de estados finitos (Caseiro *et al.* 2002);
- por modelos de máxima entropia (Barros e Weiss, 2006);
- por regras linguísticas (Braga *et al.* 2006), (Oliveira *et al.* 1992), (Teixeira 2004), (Candeias e Perdigão 2008); e
- por modelos híbridos (Veiga *et al.* 2011).

Uma outra técnica que poderá ser mencionada, e que é aplicada especialmente em línguas que não apresentam uma clara correspondência entre grafema e fone(ma), como é o caso do inglês, é a que resulta de uma abordagem por meio de um dicionário com léxico e respetiva notação fonético-fonológica. Funcionando como uma tabela de recurso finita, esta técnica falha sempre que surgem novos vocábulos para conversão.

3.1. Regras Linguísticas

Regras para marcação da tonicidade do vocábulo (Candeias e Perdigão, 2008; Almeida e Simões, 2001; Braga e Resende, 2007; Viana e d'Andrade, 1985) e para a identificação da unidade silábica (Almeida e Simões 2001; Braga e Resende 2007; Teixeira *et al.* 2000; Teixeira *et al.* 2006), foram usadas na implementação de algoritmos para a conversão G2P. A relevância da implementação de regras linguísticas deve-se, em grande parte, à razoável regularidade fonética e fonológica bem como à ortografia de base fonética-fonológica que a língua portuguesa, na sua vertente europeia apresenta, a par das línguas românicas na sua generalidade. Resolvendo-se os casos de falta de univocidade com algumas estratégias, como a da configuração de uma nova entidade, denominada grafonema, identificativa da associação unívoca entre grafema e respetivo símbolo fonético/fonológico (Bisani e Ney 2008), são as características da estrutura da língua que explicam, em parte, o êxito da aplicação de regras linguísticas, facto que os trabalhos aqui referenciados exemplificam.

3.2. Modelos Mistos

192

Contrastando com as abordagens baseadas em regras suportadas por um conhecimento linguístico da língua, que se pretende exaustivo, as abordagens estatísticas baseiam-se no pressuposto de que a pronúncia de um vocábulo pode ser prevista, por analogia, a partir de exemplos de sequências suficientes. Se, por um lado, uma das vantagens apontada pelos modelos estatísticos consiste em não implicar uma verificação constante da interdependência das regras, em especial quando surge uma sequência de grafemas que sai fora das regularidades até então admitidas, por outro, tem-se verificado que a conversão G2P proveniente desses modelos não capta um contexto suficientemente amplo de forma a impedir que a estrutura fonológica do PE seja violada. A título de exemplo, o contexto vizinho do grafema <a> na sequência <vaga>, não é suficiente para determinar a pronúncia a atribuir na medida em que em <vagarosa> ou em <divagar>, a pronúncia é /6/ (/v6g6rOz6/ e /div6gar/, respetivamente, e não /a/ como em /vag6/). Assim, a reunião de abordagens estatísticas e por regras linguísticas foi feita com êxito em (Veiga *et al.* 2011), onde algoritmos baseados em regras para a acentuação vocálica, reconhecendo o núcleo da sílaba tónica de cada vocábulo, foram imbuídos nos algoritmos dos modelos probabilísticos do conversor G2P¹¹⁴.

4. Conclusões e Desafios

Um dos módulos principais que integra a arquitetura geral de sistemas de processamento da fala do português diz respeito à tarefa de encontrar a pronúncia de um dado vocábulo na sua forma escrita. Não sendo uma tarefa de fácil resolução, várias têm sido as abordagens adotadas no sentido de se encontrar uma solução eficaz para o processo de mapear grafemas e

¹¹⁴O sucesso do modelo misto que congrega abordagens estatísticas e por regras fonológicas é possível de ser comprovado pela experimentação do conversor em <http://www.co.it.pt/~labfala/g2p/>.

fone(ma)s. Modelos baseados em regras, ou em probabilidades de sequências de grafonemas são duas estratégias adotadas comuns.

Neste artigo, foi nosso objetivo mostrar como o conhecimento fornecido pela linguística teórica tem sido, e pode ser, integrado em sistemas automáticos de processamento da língua portuguesa, nomeadamente no módulo de conversão de G2P. Nesse sentido, foram apresentadas soluções desenvolvidas no seio do grupo que os autores integram¹¹⁵ e que incluem a identificação de vogal tónica, a divisão silábica e a conversão de grafemas em fone(ma)s.

Simultaneamente, foi igualmente nossa intenção identificar novos rumos de orientação para o estudo da língua portuguesa, apontando-se a engenharia do processamento linguístico ou a linguística tecnológica como uma linha de investigação onde investir.

5. Referências

- Almeida, J. e A. Simões (2001). Text to Speech: a rewriting system approach. XVII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), Jaén: Universidad de Jaén, 247-255.
- Andrade, A. (1994). Reflexões sobre o ‘e mudo’ em Português Europeu. Actas do Congresso Internacional sobre o Português. Lisboa: Colibri, vol II, 303-344.
- Barbosa, J. Morais (1994). Introdução ao Estudo da Fonologia e Morfologia do Português. Coimbra: Almedina.
- Barros, M. e C. Weiss (2006). Maximum Entropy Motivated Grapheme-to-Phoneme, Stress and Syllable Boundary Prediction for Portuguese Text-to-Speech. IV Jornadas en Tecnologías del Habla. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla, 177-182.
- Bisani, M. e H. Ney (2008). Joint-Sequence Models for Grapheme-to-Phoneme Conversion. Speech Communication, Amsterdam: Elsevier, vol. 50(5), 434–451.
- Braga, D. e M. A. Marques (2007). Desambiguação homógrafos para Sistemas de conversão Texto-Fala em Português. Diacrítica 21.1. Braga: CEHUM/Universidade do Minho, 25-50.
- Braga, D., L. Coelho e F. Resende (2006). A Rule-Based Grapheme-to-Phone Converter for TTS Systems in European Portuguese. VI International Telecommunications Symposium. Fortaleza-CE, Brazil: IEEE Signal Processing Society, 328-333.
- Braga, D. e F. Resende (2007). Módulos de Processamento de Texto Baseados em Regras para Sistemas de Conversão Texto-Fala em Português Europeu. XXII Encontro da Associação Portuguesa de Linguística. Lisboa: Colibri, 141-155.

¹¹⁵ Laboratório de Processamento de Sinal (<http://lsi.co.it.pt/spl/>), do Instituto de Telecomunicações, polo de Coimbra.

- Candeias, S. e F. Perdigão (2008). Conversor de grafemas para fones baseado em regras para português. *Perspectivas sobre a Linguateca. Actas do Encontro Linguateca: 10 anos. Linguateca, Lisboa, cap.14.*
- Candeias, S., A. Veiga, C. Lopes e F. Perdigão (2011). A realização do schwa no Português Europeu. Comunicação apresentada no II Workshop On Portuguese Description-JDP (Satellite Workshop of STIL 2011), October 24-26, 2011. Cuiabá, UFMG, Brazil.
- Candeias, S. e F. Perdigão (2011). Investigating new syllables prototypes for the Portuguese language. *Proceedings of the International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, 388-391.* (<http://www.icphs2011.hk/resources/OnlineProceedings/RegularSession/Candeias/Candeias.pdf>)
- Candeias, Sara (2007). Sistema fonológico da Beira Interior e algumas considerações sintáctico-semânticas. PhD Thesis, Universidade de Aveiro (<http://ria.ua.pt/handle/10773/2871>)
- Caseiro, D., I. Trancoso, I., L. Oliveira e C. Viana (2002). Grapheme-to-Phone Using Finite-State Transducers. *Proceedings of the IEEE Workshop on Speech Synthesis. Santa Monica, USA: IEEE Signal Processing Society, 215-218.*
- Mateus, M. e E. d'Andrade, E. (2000). *The Phonology of Portuguese.* Oxford University Press.
- Mateus, M. (2006). Sobre a natureza fonológica da ortografia portuguesa. In: *Questões de Fonética e Fonologia: uma Homenagem a Luís Carlos Cagliari.* Bahia: Universidade Estadual do Sudoeste da Bahia, 159-180.
- Oliveira, L., M. Viana e I. Trancoso (1992). A Rule-Based Text-to-Speech System for Portuguese. *Proceedings of The International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* San Francisco, USA: IEEE Signal Processing Society, vol. 2, 73-76.
- Oliveira, L. (1996). *Síntese de Fala a Partir de Texto.* PhD Thesis, Instituto Superior Técnico – Universidade Técnica de Lisboa.
- Rua, C., A. Teixeira e L. Moutinho (2005). Ditongos Orais no Português Europeu. In: *Actas do XX Encontro da Associação Portuguesa de Linguística.* Lisboa: APL/Colibri, 803-816.
- Teixeira, A., C. Oliveira e L. Moutinho (2006). On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme-Phone Conversion. *Proceedings of The International Conference on Computational Processing of Portuguese (PROPOR).* Springer, 212-215.
- Teixeira, J. P. (2004). *A Prosody Model to TTS Systems.* PhD Thesis, Faculdade de Engenharia da Universidade do Porto.
- Teixeira, J. P., P. Gouveia e D. Freitas (2000). Divisão Silábica Automática do Texto Escrito e Falado. *Proceedings of The International Conference on Computational Processing of Portuguese (PROPOR).* Atibaia, SP, Brasil, vol. 1, 88-98.
- Veiga, A., S. Candeias e F. Perdigão (2011). Generating a Pronunciation Dictionary for European Portuguese Using a Joint-Sequence Model with Embedded Stress Assignment. *Proceedings of the Brazilian Symposium in Information and Human Language Technology (STIL).* Cuiabá, Mato Grosso, Brazil, 144 - 153.
- Veiga, A., S. Candeias, L. Sá, e F. Perdigão (2010). Using Coarticulation Rules in Automatic Phonetic Transcription. *Proceedings of PROPOR 2010.* Porto Alegre. (http://www.inf.pucrs.br/~propor2010/proceedings/regular_papers/VeigaEtAl.pdf)
- Viana, M. C. e E. d'Andrade (1985). *CORSO I: um conversor de texto ortográfico em código fonético para o português.* Relatórios do grupo de fonética e fonologia n. 6. Lisboa: CLUL.