Módulos de processamento de texto baseados em regras para sistemas de conversão Texto-Fala em PE

Daniela Braga* e Fernando Gil V. Resende Jr**
*Microsoft Language Development Center
**Universidade Federal do Rio de Janeiro

1. Resumo

Neste artigo, apresentamos os últimos avanços e resultados obtidos no desenvolvimento dos módulos de divisão silábica, de marcação de sílaba tónica e de transcrição grafema-fone que constituem o conversor grafema-fone de um sistema de conversão Texto-Fala ou *Text-to-Speech* (TTS), baseado em *Hidden Markov Models* (HMMs), para o Português Europeu (PE).

Testes efectuados a cada módulo revelaram as seguintes taxas de acerto: 99,06% para o divisor silábico, 99,54 % para o marcador de sílaba tónica e 99,11% para o transcritor grafema-fone. Estes resultados, juntamente com as nossas perspectivas de desenvolvimento futuro, vêm confirmar o papel crucial do conhecimento linguístico para o desenvolvimento de sistemas de conversão Texto-Fala. Estes algoritmos foram já adaptados e implementados a um sistema de conversão Texto-Fala em Português do Brasil (PB), com resultados similares, dada a grande compatibilidade de padrões gráficos, fonéticos e fonológicos existentes entre as duas variedades do Português.

2. Introdução

São cada vez mais comuns as aplicações do nosso quotidiano que integram sistemas de síntese da fala, também designados por sistemas de conversão Texto-Fala, com o objectivo de facilitar o acesso à informação de um número cada vez maior de utilizadores. De entre as principais aplicações, contam-se os sistemas de auxílio à leitura para cegos, os sistemas de auxílio à navegação por GPS (Sistema de Posicionamento Global, do inglês *Global Positioning System*) ou PDA (*Personal Digital Assistant*), as aplicações para telemóveis usando sistemas de pergunta-resposta, as aplicações industriais do tipo de máquinas com comandos mãos-livres, as aplicações médicas de monitorização de doentes, o software de ensino de línguas e de e-learning em geral e as aplicações que facilitem a acessibilidade na internet (leitura de e-mail, leitura de conteúdos de páginas web, etc.), entre muitas outras. A indústria das tecnologias da fala tem evoluído face a uma procura crescente de soluções de voz por parte do mercado. São já várias as empresas que se dedicam ao desenvolvimento de tecnologias de síntese e de reconhecimento de voz, como a italiana Loquendo, a belga Acapela, a catalã Atlas

XXII Encontro Nacional da Associação Portuguesa de Linguística, Lisboa, APL, 2007, pp. 141-155

ou as multinacionais Nuance (anterior Scansoft), IBM ou Microsoft. É nosso objectivo com este trabalho desenvolver uma aplicação que contribua para o incremento de sistemas de síntese da fala em Português Europeu, embora o ensino/aprendizagem de línguas e a linguística clínica possam também ser áreas de aplicação contempladas. Este trabalho tem sido desenvolvido em parceria com o Laboratório de Processamento de Sinais da Universidade Federal do Rio de Janeiro. Desta cooperação resultou já a aplicação destes algoritmos ao Português do Brasil (Silva et al., 2006). A proximidade entre o Português e o Galego, demonstrada em publicações recentes (Braga et al., 2006), perspectiva também a adaptação destes algoritmos a sistemas de Texto-Fala para Galego. O processamento da fala em Português possui pouco mais de uma década de trabalho, o que contrasta com outras línguas como o Inglês ou o Francês. A nível académico, estão descritos vários sintetizadores em PE, como os sintetizadores por formantes DIXI (Oliveira, 1996) e o Multivox (Teixeira, 1998); o sintetizador de base articulatória desenvolvido pela Universidade de Aveiro (Teixeira, 2000) e o sintetizador por concatenação de unidades (Barros, 2001) ou o sintetizador por modelos escondidos de Markov (HMMs) actualmente em desenvolvimento (Barros et al., 2005). Para o PB, são de salientar o sistema de síntese baseado no HTS[1] para o PB[2]; o sintetizador concatenativo aliado a técnicas PSOLA e Híbrida do grupo IEL/LAFAP da Universidade de Campinas (Gomes, 1998), o sistema ORTOPHON, baseado em fonologia articulatória e desenvolvido pelo grupo do LAFAPE (Albano, 1996) e o TTS concatenativo do grupo LINSE da Universidade de Santa Catarina[3]. Porém, nenhuma abordagem parece ter alcançado resultados tão animadores partindo de um sistema baseado essencialmente em regras linguísticas. Além disso, poucos autores revelam os seus algoritmos e ainda menos publicam os resultados de testes de desempenho comparativamente com outros sistemas.

3. Arquitectura do sistema de TTS

Apesar da grande flutuação existente a nível da arquitectura dos modernos sistemas de síntese da fala[4], existem pelo menos três blocos comuns a todos eles: o pré-processamento de texto ou *front-end*, o motor de síntese ou *back-end* e a base de dados de voz ou *voice font* (vide Figura 1). Em todos os casos, o objectivo é a geração de fala sintética resultante da conversão do texto em etiquetas fonéticas, obtida à saída do front-end, etiquetas essas que serão depois interpretadas e transformadas pelo motor de síntese em voz. A base de dados de voz é rigorosamente seleccionada, gravada e

[1] HTS é o mesmo que "HMM-Based Speech Synthesis System" (para mais informações consultar: http://hts.ics.nitech.ac.jp/).

[2] Disponível em http://kt-lab.ics.nitech.ac.jp/~maia/demo.html.

[3] LINSE em http://www.linse.ufsc.br/.

[4] Para uma visão global da arquitectura e técnicas dos sistemas de síntese, veja-se Huang et al. (2001, capítulo IV).

etiquetada, sendo as suas unidades seleccionadas por algoritmos de processamento de sinal que as concatenam e transformam em voz sintética.

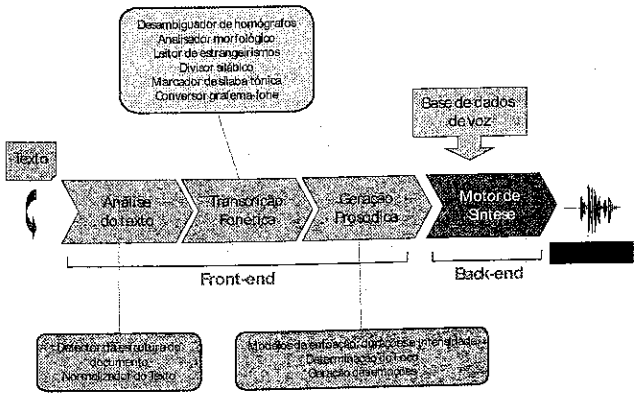

Figura 1: Arquitectura de um sistema de TTS.

O front-end é composto por três componentes: 1) a análise de texto, do qual fazem parte o detector da estrutura do documento (responsável por identificar o tipo de texto, informação que pode ser útil para a geração dos modelos prosódicos) e o normalizador de texto (em que se incluem os sub-módulos de interpretação de pontuação, expansão de abreviaturas, leitura de siglas e de acrónimos e conversão de numerais, datas, números romanos e árabes, quantias em dinheiro, números ordinais); 2) a análise fonética, da qual fazem parte os módulos de leitura de estrangeirismos, de desambiguação de homógrafos heterófonos, de análise morfossintáctica e de conversão grafema-fone, objecto deste trabalho; 3) análise e geração prosódica, módulo que aproveita as informações anteriormente obtidas a partir do texto, como tipo de frase, divisão silábica, marcação de tónica, classificação morfossintáctica, para produzir modelos prosódicos.

O back-end é composto pelo motor de síntese que interpreta a transcrição fonética gerada pelo front-end e a transforma em fala sintética. Neste trabalho, estamos a usar a técnica de síntese por HMMs, alimentada por uma base de dados de 1000 frases foneticamente balanceada, com a duração de cerca de 1 hora e 15 minutos de voz. Esta técnica permite precisamente obter bons resultados com uma base de dados relativamente reduzida.

No presente artigo relata-se a estrutura e funcionamento do módulo de conversão grafema-fone, que se enquadra na componente da transcrição fonética do bloco de pré-processamento de texto, como pode ser observado na Figura 1. Este módulo é central na arquitectura global dos sistemas de síntese e é aquele em que o conhecimento linguístico pode ter maior impacto e eficácia.

4. Módulos do conversor grafema-fone

4.1. Divisor silábico

4.1.1.Estado da arte

O número de trabalhos em que se reportam divisores silábicos para o PE e para o PB permite constatar que se trata de uma questão da maior importância para o desenvolvimento da naturalidade da fala sintética. De entre os principais trabalhos em que se descrevem divisores silábicos aplicáveis a sistemas de TTS, destacamos as propostas de Teixeira, et al. (2000), Seara Jr. et al. (2004) e Oliveira, et al. (2005). Em todos os casos, os autores reconhecem a importância da identificação da unidade silábica, quer para a implementação de algumas regras do conversor grafema-fonema, quer para a modelização da prosódia, ao nível da duração, intensidade e mesmo frequência fundamental. Em todos os casos, seguiu-se uma abordagem linguística, por regras e fizeram-se testes de desempenho dos sistemas.

Símbolo	Significado
V	vogal (a,e,o, á,é,ó, ú, í, ã, õ, â, ê, ô, à)
G	glide (i, u)
C	qualquer consoante (<lh>, <nh>, CO, CF, CL, CN)
CO	consoante oclusiva (p, t, c+a,o,u; qu+e,i, b, d, g+a,o,u, gu+e,i)
CF	Consoante fricativa (f,v, s, c+e,i, ç, z, ss, ch, j, g+e,i, x)
CL	consoante líquida (l, r, rr excepto <lh>)
CN	consoante nasal (m, n)
SP	espaço
^(+1)= C	o primeiro grafema à direita da vogal é uma consoante qualquer
^(+2) = G	o segundo grafema à direita da vogal é uma glide
^(+3) = V	o terceiro grafema à direita da vogal é uma vogal
^(-1) =CO	o primeiro grafema à direita da vogal é uma consoante oclusiva
→	então

Tabela 1: Simbologia usada no algoritmo de divisão silábica.

Caso	Operação
Caso 1	V separa-se do grafema seguinte
Caso 2	V junta-se ao primeiro grafema da direita e separa-se dos grafemas subsequentes
Caso 3	V junta-se ao grafema anterior e separa-se dos seguintes
Caso 4	V junta-se ao grafema anterior e ao seguinte e separa-se dos subsequentes
Caso 5	V junta-se aos dois grafemas seguintes e separa-se do terceiro
Caso 6	V junta-se ao grafema anterior e a todos os grafemas seguintes até final da palavra
Caso 7	V junta-se ao grafema anterior e aos dois seguintes e separa-se dos subsequentes
Caso 8	V junta-se aos dois grafemas anteriores e separa-se do seguinte

Tabela 2: Casos e operações considerados.

À luz das experiências anteriores mencionadas, desenvolvemos um módulo de divisão silábica de base ortográfica, mas com objectivos fonológicos, ou seja, tentando conciliar as modernas teorias decorrentes da Fonologia (Mateus & Andrade, 2000) com as necessidades práticas subjacentes ao desenvolvimento tecnológico dos sistemas de síntese da fala para o Português.

4.1.2. Algoritmos de divisão silábica

1	Se V é início de sílaba e ^(+1)= V → Caso 1	a.eronave, a.inda
2	Se V é início de sílaba e ^(+1) =C e ^(+2) =C e ^(+3) =CO → Caso 5	o.bstar, adstrito
3	Se V é início de sílaba e ^(+1)= G, <s>, <r>, <l>, <x>, <c> CN e ^(+2)= C → Caso 2	am.bos, en.te, as.pas, al.tura, ar.gúcia, eu.ropa, as.tral, ex.por, ei.ra, ai.po
4	Se V é início de sílaba e ^(+1) =C e ^(+2) =C e ^(+3) =V → Caso 1	o.ptar, a.dvogar, a.gnóstico, a.florar, a.fia
5	Se V é início de sílaba e ^(+1)= C e ^(+2)= V, CL → Caso 1	a.rrendar, a-tlas, a.lho, a.mor, a.clamado
6	Se V não é início de sílaba e ^(-1) =C e (^+1)=C e (^+2) =V → Caso 3	ca.lha, ca.la, me.ta, ca.choeira
7	Se V não é início de sílaba e ^(-1) =C e (^+1) =G e ^(+2)= "r" e ^(+3) C → Caso 3	ca-irmos
8	Se V não é início de sílaba e ^(-2) = CO, CF e ^(-1) = CL e (^+1)=C → Caso 8	pro.duto, democra.cia
9	Se V não é início de sílaba e ^(-1) =C e (^+1) =G e (^+2)= <s> e (^+3) =CO→ Caso 7	claus.tro
10	Se V não é início de sílaba e ^(-1) =C e (^+1) = CN e (^+2)= <s> e (^+3)=CO → Caso 7	demons.tra
11	Se V não é início de sílaba e ^(-1) =C, G e (^+1) =G e (^+2)= C → Caso 4	cai.ro, rai.va, quei.xar, cau.sa
12	Se V não é início de sílaba e ^(-1) =C e (^+1) =G e (^+2)= V ou SP → Caso 4	prai.a, mei.a, sei
13	Se V não é início de sílaba e ^(-2) =C e ^(-1) = G e (^+1) =C e (^+2)= V → Caso 3	pia.da, via.gem, sua.da, Sue.ca
14	Se V não é início de sílaba e ^(-1) =C e (^+1) =CL, CN, <s>, <c> e (^+2)= C e (^+3)=V → Caso 4	car.ta, mal.dade, con.tar, spor.ting, hos.pital, pro.jec.to
15	Se V não é início de sílaba e ^(-1) =C e (^+1) =CL, CN, <r> ou (^+2)= <s>, SP → Caso 6	Ama.ral, sóis, se.jam, ima.gens, mais
16	Se V não é início de sílaba e ^(+1) = V igual → Caso 1	ni.i.lismo, re.estruturar
17	Se V não é início de sílaba e ^(-1) =C e ^(-1)=V e ^(+2)=CN → Caso 3	transe.unte, influ.ência
18	Se V não é início de sílaba e ^(-1) = C e ^(+1)= V → Caso 3	lisbo.eta, Lisbo.a, isra.e.lita, pesso.as, cacho.eira
19	Se V não é início de sílaba e ^(-1)=C e ^(+1)= V e ^(+2)=CN e ^(+3) =C → Caso7	influên.cia

20	Se Vogal não é início de sílaba e ^(+1) =C e ^(+2) =C e ^(+3) =V → Caso 3	su.btil, su.blime, de.clarações, ra.pto[5] ma.gno,
21	Se V não é início de sílaba e é = <i> e ^(-1) = C e ^(+1) = <a>, <o> → Caso 3	polí.cia, democra.ci.a, ó.ci.o,
22	Se V não é início de sílaba e é = <i> e ^(-1) = C e ^(+1) = <a>,< o> e ^(+2) = C, <i> → Caso 3	polí.ci.al, polí.ci.ais, Di.as, ri.os
23	Se V não é início de sílaba e é = <á>,<ó> e ^(-1) = C e ^(+1) = <o>,<e> e/ou ^(+2)=<s> → Caso 6	ambi.ção, cora.ções,
24	Se nenhum dos casos anteriores se verificar e a palavra terminar, a Vogal forma sílaba com os grafemas que restarem até ao espaço em branco ou sinal de pontuação ou hífen.	cas.to, des.cer

Tabela 3: Regras de divisão silábica.

Foram consideradas três condições prévias: 1) uma sílaba nunca pode corresponder a apenas uma consoante; 2) separam-se sempre os grafemas <xc>; 3) são tratados como um grafema apenas os dígrafos <nh>, <lh>, <ch>, < rr>, <ss>, <qu> e <gu> quando seguidos de <e> e de <i>.

O algoritmo de divisão silábica assenta numa busca vogal a vogal dentro de cada palavra. Uma vez identificada a vogal, analisa-se a vizinhança grafemática à esquerda e à direita e de acordo com as regras listadas na Tabela 3, aplicam-se os casos ou saídas apresentados na Tabela 2. Na Tabela 1, pode ver-se a simbologia utilizada.

4.2. Marcador de sílaba tónica

4.2.1. Estado da arte

A marcação da sílaba tónica tem impacto em dois módulos do sistema de TTS: ao nível do transcritor grafema-fone, por um lado, na medida em que algumas regras de conversão grafema-fone utilizam a informação da tonicidade da vogal (o símbolo de vogal tónica está contemplado na Tabela 6 que descreve a simbologia utilizada no transcritor grafema-fone), e ao nível do módulo de análise e geração prosódica, por outro, na medida em que a tonicidade está associada ao aumento da frequência fundamental, intensidade e duração da vogal ou sílaba abrangidas. A alternância entre sílabas tónicas e átonas é ainda responsável pelo ritmo e por fenómenos de micro-prosódia.

Contudo, são escassos os trabalhos publicados sobre este assunto em concreto, dos quais se destaca o artigo de Teixeira et al. (1998), em que se descreve o algoritmo de marcação de sílaba tónica, construído com apenas três regras[6], seguidas de uma tabela

[5] Esta divisão é discutível, mas optámos por ela por ser psicolinguisticamente mais lógica, logo, mais adequada a sistemas de síntese da fala.

[6] "A marcação da sílaba tónica obedece às seguintes três regras com prioridade decrescente: 1- o texto escrito que contenha uma das seguintes letras será convertido com a marca de acentuação da palavra (á, é, í, ó, ú, à, è, ì, ò, ù, ã, õ, â, ê, ô). (...) 2- quando não há nenhuma marca de acentuação nas palavras terminadas por uma das seguintes sequências (al, el, il, ol, ul, ar, er, ir, or, ur, az, ez, iz, oz, uz), é colocada uma marca de acentuação na última sílaba. 3- quando não existe ainda na palavra nenhuma marca de acentuação, é seguida a regra geral de acentuação na penúltima sílaba." (*apud* Teixeira et al., 1998).

de excepções. No entanto, não são conhecidos os níveis de desempenho do sistema nem as tabelas de excepções.

^(0)	último grafema de uma palavra
^(1)	penúltimo grafema de uma palavra
^(2)	antepenúltimo grafema de uma palavra
^(3)	terceiro grafema a contar do final da palavra
^(4)	quarto grafema a contar do final da palavra
T	posição ocupada pela vogal tónica
T=1	a tónica corresponde ao penúltimo grafema
→	então
{x}	grafema x

Tabela 4: Simbologia usada no algoritmo de marcação de sílaba tónica.

A constatação da falta de documentação sobre este tema funcionou como motivação para a produção de uma ferramenta de marcação de tonicidade, em colaboração com o Laboratório de Processamento de Sinais, cuja primeira versão pode ser consultada em Silva, et al. (2006). Estas regras (inicialmente 20, agora 24) foram testadas para o PB com um extracto do Cetem-Folha, tendo sido obtidos 98,58% de percentagem de acerto.

Neste artigo, testámos estas regras com o corpus do Cetem-Público para o PE, com vista a analisar a adaptabilidade deste algoritmo a sistemas de conversão Texto-Fala em PE, como veremos no ponto 5.

4.2.2. Algoritmos de marcação de sílaba tónica

O algoritmo de marcação de tonicidade proposto foi refinado e adaptado ao PE[7], funcionando agora com 24 regras construídas a partir da análise das sequências ortográficas e do conjunto de regras de acentuação gráfica vigentes (Estrela et al., 2004; Bergstrom & Neves, 1997).

Na Tabela 4, apresenta-se a descrição da simbologia utilizada. Na Tabela 5, podem ver-se as regras de marcação de tonicidade.

O algoritmo começa por verificar se existem as excepções no texto, ou seja, as palavras átonas.

[7] Retirou-se a regra 11 da versão inicial de Silva et al. (2006) relativa ao funcionamento de <porque> em PB, que é sempre oxítona.

1	Lista de palavras átonas	por, um, se
2	Se existir acento agudo, grave, circunflexo ou til, a vogal marcada é tónica. O acento circunflexo tem precedência sobre o til.[8]	órgão, órgãos, bênção, bênçãos
3	Se a palavra só tiver uma vogal → T= vogal	tem, vem, bem, vi,
4	Se ^(0) = {r, l, z, x} → T = 1	propor, carrossel, rapaz, triplex, juiz
5	Se ^(0) = {m} e ^(1) = {i,o,u} → T = 1	pudim, bombom, comum
6	Se ^(0) = {s} e ^(1) = {n} e ^(2) = {i,o,u} → T = 2	pudins, bombons, comuns
7	Se ^(0) = {i} e ^(1) = {u,} e ^(2) = {q,g} → T = 0	caqui, aqui, sagüi
8	Se ^(0) = {s} e ^(1) = {i} e ^(2) = {u,} e ^(3) = {q,g} → T = 1	caquis, sagüis
9	Se ^(0) = {i,u} e ^(1) é vogal → T = 1	caiu, grau, pneu
10	Se ^(0) = {i,u} e ^(1) não é vogal → T = 0	caju, javali
11	If ^(0) = {s} e ^(1) = {i,u} e ^(2) não é vogal → T = 1	cajus, javalis
12	Se ^(0) = {s} e ^(1) = {i,u} e ^(2) é vogal → T = 2	andais, paqis, graus.
13	Se ^(0) = {e} e ^(1) = {u} e ^(2) = {q} e ^(3) é vogal → T = 3	alambique, Henrique, destaque
14	Se ^(0) = {e} e ^(1) = {u} e ^(2) = {q} e ^(3) = {r} → T = 4	embarque, marque
15	Se ^(0) = {s} e ^(1) = {e} e ^(2) = {u} e ^(3) = {q} e ^(4) é vogal → T = 4	alambiques, Henriques, destaques
16	Se ^(0), ^(1), ^(2) são vogais e se ^(1) = {i,u} → T = 2	meia, seio
17	Se ^(0) e ^(3) são vogais e ^(2) = {i,u} e ^(1) não é vogal e ^(4) ≠{q,g} → T =3	cadeira, queima, louco
18	Se ^(0) = {s}, ^(1), ^(4) são vogais e ^(3) = {i,u} e ^(2) não são vogais → T = 4	cadeiras, queimas, loucos
19	Se ^(0) = {a,e,o} e ^(1) é consoante e ^(2) = {n} e ^(3) = {i,u} e ^(4) é vogal → T = 3	ainda, caindo, fluindo, incluindo
20	Se ^(k) = penúltima vogal e ^(k) = {i,u} e ^(k+1) é vogal e ^(k-1) não é vogal e ^(k+2) não é {q,g} → T = k+1	outro, claustro
21	Se ^(0) = {m} e^(1) = {e} e ^(2) = {u} e ^(3) = {q} → T = 1	quem
22	Se ^(0) = {a,o, e} e ^(1) = {i, u} e ^(2) é consoante ou {u} → T = 1	academia, inicie, assobio, conseguia, continua, rua
23	Se ^(0) = {s, m} e ^(1) = {a,o, e} e ^(2) = {i, u} e ^(3) é consoante ou {u} → T = 2	academias, assobios, conseguias, deveriam, continuam, iniciem
24	Se nenhuma das regras anteriores se verificar → T = penúltima vogal da palavra	casa, homem, guerra

Tabela 5: Algoritmo de marcação da sílaba tónica.

[8] Constituem excepção a esta regra nomes ou adjectivos a que se junta o sufixo <-zinho> (ex: pãezinhos, sotãozinho) ou <-mente> (ex: cristãmente), em que a vogal tónica se desloca para a penúltima sílaba, embora psico-cognitivamente se possa considerar que existem dois acentos fonológicos.

Foram considerados átonos, e portanto, desprovidos de tonicidade, os seguintes vocábulos: 1) os artigos definidos <o, a, os, as> e os indefinidos <um, uns>; 2) os pronomes pessoais oblíquos <me, te, se, o, a, os, as, lo, la, los, las, no, na, nos, nas, lhe, lhes, nos, vos> e suas contracções <mo, ma, mos, mas, to, ta, tos, tas, lho, lha, lhos, lhas, no-lo, no-la, no-los, no-las, vo-lo, vo-la, vo-los, vo-las>; 3) o pronome relativo <que>; 4) as preposições <a, com, de, em, por, sem, sob> e as contracções <do, da, dos, das, ao, à, aos, às, no, na, nos, nas, num, nuns>; 5) e as conjunções <e, mas, nem, ou, que, se>.

4.3. Transcritor grafema-fone

4.3.1. Estado da arte

...	Qualquer grafema
< x >	Grafema ou conjunto de grafemas <x>
/ y /	Fonema ou conjunto de fonemas y
,	Separa opções
{ x_1, x_2, x_3 }	Conjunto de grafemas
< x_1 {x_2, x_3} >	< x_1 x_2 > ou < x_1 x_3 >
< C / y >	Consoante excepto <y>
< C / {w, z} >	Consoante excepto < w> e < z>
V	Qualquer vogal gráfica (e.g. a, e, i, o, u)
C	Quelquer consoante gráfica (e.g. p, t, k, b, d, g...)
Pont	Sinal de pontuação (e.g., . ! ? () –; sp)
Ltr	Caracteres que são letras (e.g. a, b, c, ...)
SP	Espaço entre palavras
Hf	Hifen
<(case) x >	Caso que modifica o grafema <x>
<(C) x>	<x> é uma consoante
<(V) x>	<x> é uma vogal
<(UV) x>	<x> é não vozeado
<(VO) x>	<x> é vozeado
<(US) x>	<x> é vogal átona
<(S) x>	<x> é vogal tónica
<(W_bgn) x>	<x> está em início de palavra
<(Prn_D) x>	O grafema <x> é um demonstrativo (ex. este(s), esse(s), aquele(s))

Tabela 6: Símbolos e convenções de anotação usados no conversor grafema-fone.

A questão da transcrição grafema-fone é uma questão central em síntese da fala, constituindo um problema ainda longe de estar solucionado. Além disso, é o módulo por excelência da análise fonética.

Foram propostos vários quadros teóricos para resolver o problema da conversão grafema-fonema nos sistemas de conversão Texto-Fala, de entre os quais, destacamos os seguintes: árvores de decisão (Lucassen & Mercer, 1984), árvores de decisão treinadas automaticamente (Black et al., 1988), modelos de "Tabela look-up" (Bosh & Daelemans, 1993), abordagens baseadas em dicionários (Coker et al., 1990), abordagens

baseadas em regras linguísticas (Kaplan & Kay, 1994), modelos híbridos (Meng et al., 1994), abordagens por redes neuronais (Sejnowski &. Rosenberg, 1987), por máquinas de estados finitos (Roche & Schabes, 1995; Paulo, 2005), por cadeias escondidas de Marcov (Taylor, 2005) e modelos estatísticos (Chotimongkol & Black, 2000). Em Demper et al. (1998), faz-se uma comparação entre várias técnicas e discutem-se os respectivos resultados.

Uma das técnicas mais utilizadas é a abordagem por dicionário, que consiste numa lista de palavras ou léxico, a que se faz corresponder a respectiva transcrição fonética. Esta técnica tem sido mais aplicada a línguas que não apresentam uma correspondência grafema-fonema unívoca, como é o caso do Inglês ou do Francês. Mas esta abordagem falha drasticamente quando surgem palavras que não constam no dicionário, como neologismos, estrangeirismos, etc.

Outra possibilidade são os sistemas mistos que podem gerar regras linguísticas ou de análise estatística de padrões ortográficos encontrados nas transcrições fonéticas fornecidas pelos dicionários.

Justificamos a nossa abordagem baseada em regras linguísticas apoiando-nos assim em três premissas: primeiro, o Português é uma língua com bastantes regularidades fonológicas; segundo, uma abordagem por regras é mais económica em termos de memória computacional do que uma abordagem por dicionário e, por último, uma abordagem por regras é sempre capaz de "ler" uma palavra nova.

Para a construção dos nossos algoritmos, foram considerados os estudos mais recentes em Fonética e Fonologia do Português (Mateus et al. 1990; Mateus & Andrade, 2000; Rodrigues, 2003).

4.3.2. Algoritmos de transcrição grafema-fone

Na Tabela 6 apresentam-se as convenções de anotação usadas na descrição dos algoritmos propostos para a transcrição grafema-fone.

A nível da anotação fonética, seguimos o alfabeto SAMPA[9] (vide Tabela 7) com uma extensão (a consoante lateral velarizada [l*] que ocorre na articulação da palavra <sal> em PE), por ser o alfabeto mais adequado do ponto de vista do processamento computacional das línguas. A nossa transcrição é fonética, e não fonológica, de forma a traduzir com rigor os fenómenos de sandhi, que não são descritos fonologicamente.

[9] Acerca do SAMPA: "SAMPA (Speech Assessment Methods Phonetic Alphabet) is a machine-readable phonetic alphabet. It was originally developed under the ESPRIT project 1541, SAM (Speech Assessment Methods) in 1987-89 by an international group of phoneticians, and was applied in the first instance to the European Communities languages Danish, Dutch, English, French, German, and Italian (by 1989); later to Norwegian and Swedish (by 1992); and subsequently to Greek, Portuguese, and Spanish (1993)." Além disso, "Where Unicode (ISO 10646) is not available or not appropriate, SAMPA and the proposed X-SAMPA (Extended SAMPA) constitute the best robust international collaborative basis for a standard machine-readable encoding of phonetic notation." Disponível em http://www.phon.ucl.ac.uk/home/sampa/index.htm.

[p], [t], [k]	**Oclusivas orais surdas**
[b], [d], [g]	**Oclusivas orais sonoras**
[m], [n], [J]	**Oclusivas nasais**
[f], [s], [S]	**Fricativas surdas**
[v], [z], [Z]	**Fricativas sonoras**
[l], [L], [l*]	**Laterais**
[r], [R]	**Vibrantes**
[a], [6], [E], [e], [@], [O], [o], [i], [u]	**Vogais orais**
[6~], [e~], [o~], [i~], [u~]	**Vogais nasais**
[j], [w], [j~], [w~]	**Semivogais**

Tabela 7: Alfabeto SAMPA para o Português.

Na Tabela 8 é apresentado, como exemplo, o algoritmo de transcrição para o grafema <a>, proposto para o PE, ilustrado com exemplos.

O conjunto completo dos algoritmos propostos, bem como as suas excepções, pode ser consultado em Braga et al. (2006). Foram considerados todos os padrões gráficos usados no PE, incluindo grafemas estrangeiros, como <k>, <y>, <w>, <ü>, que podem ocorrer em estrangeirismos. Os ditongos crescentes e decrescentes foram igualmente considerados, embora alguns tenham sido incorporados nas regras de transcrição dos grafemas <i> e <u>. Ao conceber estas regras, tentou-se, sempre que possível, reduzir a sua dependência em relação aos módulos anteriores da divisão silábica e da marcação de tónica. Para executar a transcrição dos 26 grafemas passíveis de serem encontrados em textos de PE foram necessárias 149 regras de conversão para 38 fones.

1	...<(Rad_G) a>...	[a]	radioterapia
2	... < á, à >...	[a]	rápido
3	... <ão>...	[6~w~]	chão, leilão
4	...<ã>...	[6~]	romã, irmã
5	....<â {m,n}><C/h>...	[6~]	lâmpada
6	... <â> <C>...	[6]	câmara
7	... <am><Pont> ...	[6~w~]	sejam, andam
8	....<a (m,n)><C/h>...	[6~]	campo, canto
9	...<a><l><C/h>...	[a]	calmo, palco
10	...<a><i,u,o><C,Pont>...	[a]	paisagem, ao
11	...<(S) a> <r><Pont>...	[a]	matar, andar
12	...< (S) a><m,n>...	[6]	ramo, banha
13	...<(S) a>...	[a][10]	cacto, gato
14	... < a >...	[6]	amador

Tabela 8: Regras de transcrição para o grafema <a> .

Neste artigo, incluímos três novas regras de conversão grafema-fone, que provaram melhorar a performance dos nossos algoritmos:

[10] A preposição <para> constitui excepção a esta regra.

1. Regra extra para o grafema <a>:

...<a> <ct, cç, pt, cc> ... => [a] (ex: acção, captura,faccioso, facto)

2. Regras extra para o grafema <e>:

...<C><(US) e> <r> <es, Pont>.... => [E] (ex: repórter, Hélder, Alcácer)
...<(W_bgn) e><u>.... => [e] (ex: europa, eufemismo, eucaristia)

A nível da implementação dos algoritmos tentou-se, sempre que possível, que as regras que convergissem para a saída default não fossem implementadas. O default é considerado a saída mais frequente no total de todas as regras propostas para um dado grafema. Com este procedimento, das 149 regras descritas apenas cerca de 100 foram implementadas, tornando o sistema mais optimizado e rápido a nível do processamento computacional.

5. Discussão de resultados

Os três módulos apresentados foram testados usando 207 frases extraídas automaticamente do corpus do Cetem-Público[11] contendo 1926 palavras e 9434 caracteres sem espaços. Nas Tabelas 9, 10 e 11 apresentam-se os tipos de erros encontrados em cada um dos módulos. Nos testes submetidos a estes módulos não foram consideradas as entidades que são processadas no módulo de normalização do texto, como siglas, acrónimos, numerais, etc, por pertencerem ao nível anterior do processamento de texto.

A observação dos resultados ao nível do divisor silábico permitiu-nos concluir que se obteve 0,94% de erros repartidos entre palavras em que a divisão falhava e estrangeirismos ou nomes próprios estrangeiros. A comparação dos nossos resultados com os valores documentados para sistemas análogos (99,94% em Teixeira et al., 2000 e 99,5% em Oliveira et al., 2005) deixa-nos optimistas, apesar de sabermos que ainda é necessária alguma refinação no algoritmo.

Tipo de erro	% occur.
Estrangeirismos	0,40
Falha na separação	0,54
Total	0,94

Tabela 9: Erros no divisor silábico.

De igual modo, ao nível do marcador de tonicidade, verificam-se dificuldades na identificação da vogal tónica perante palavras estrangeiras (ex: <internet>, <floor>) e alguns nomes próprios também estrangeiros (ex:<Pausini>), num total de 0,46% de erros. Não são conhecidas avaliações de outros sistemas análogos.

[11] Disponível em http://www.linguateca.pt/.

Tipo de erro	% occur.
Nomes próprios	0,21
Palavras estrangeiras	0,25
Total	0,46

Tabela 10: Erros no marcador de tonicidade.

No que respeita aos erros encontrados no transcritor grafema-fone, constata-se que a maior parte deles tem a ver com erros na decisão do timbre vocálico dos grafemas <e> (0,264%) e <o> (0,275%), a par do problema da desambiguação fonética da vogal tónica nos homógrafos heterófonos (0,042%). Uma vez mais, os estrangeirimos e nomes estrangeiros representam uma dificuldade de leitura neste algoritmo, com cerca de 0,094% de erro. Outro problema decorre da dificuldade de previsão do timbre dos grafemas <a> e <o> em advérbios de modo derivados de adjectivos (ex: <obrigatoriamente> [O], <delicadamente> [a]), que graficamente passam a ser átonos, mantendo-se no entanto com o timbre aberto da vogal original. Descrições de outros sistemas similares apresentam resultados menos interessantes (2,08% de erro em Oliveira, 1996; 97,7% de acerto em Oliveira et al., 2004) ou não mencionam avaliação (Teixeira et al., 1998).

Tipo de erro	% occur.
Erros na transcrição de <a>	0,095
Erros na transcrição de <e>	0,264
Erros na transcrição de <o>	0,275
Erros na transcrição de <i>	0,031
Erros na transcrição de <u>	0,021
Homógrafos	0,042
Nomes próprios estrangeiros	0,063
Nomes próprios portugueses	0,031
Palavras estrangeiras	0,031
Topónimos	0,010
Advérbios em -mente	0,021
Total	0,89

Tabela 11: Erros no transcritor grafema-fone.

6. Conclusões e Trabalho Futuro

Neste trabalho, fez-se a descrição da estrutura e funcionamento de três módulos subjacentes à arquitectura de um conversor grafema-fone para o Português Europeu, construídos segundo uma abordagem linguística. As regras propostas foram implementadas e testadas usando 207 frases extraídas aleatoriamente do Cetem-Público. A avaliação do nosso sistema teve como percentagens de acerto 99,06% para o divisor silábico, 99,54% para o marcador de sílaba tónica e 99,11% para o transcritor grafema-fone.

Está em curso já o desenvolvimento do módulo de desambiguação de homógrafos heterófonos, o que permitirá resolver os erros na transcrição de pares como <governo> [e] e <governo> [E], ou <molho>[o] e <molho>[O].

Como trabalho futuro, temos em vista um estudo sobre a forma como é feita a integração fonética de palavras estrangeiras no Português, bem como o desenvolvimento de algoritmos que permitam programar a alternância vocálica ao longo da conjugação verbal (em <meto> [e] vs <metes>, <mete>, <metem> [E]) e que permitam igualmente prever os casos de alternância vocálica sub-morfémica em palavras historicamente atingidas por metafonia (como <sogro> [o] vs <sogra> e <sogros> [O]).

Referências

Albano, E., Moreira, A. (1996) Archisegment-based Letter-to-Phone Conversion for Concatenative Speech Synthesis in Portuguese. *Proceedings of ICSLP 96*, vol.3.

Barros, M.J.; Maia, R.; Tokuda, K.; Resende, F.; Freitas, D. (2005) HMM-based European Portuguese TTS System. *Proceedings of Interspeech 2005*. Lisboa, Portugal.

Barros, Maria João (2001) *Estudo Comparativo e Técnicas de Geração de Sinal para a Síntese da Fala*. Dissertação de Mestrado. Universidade do Porto.

Bergstrom, Magnus; Neves, Reis (1997) *Prontuário Ortográfico e guia da língua portuguesa*. Lisboa: Editorial Notícias, pp. 15-23.

Black, A., Lenzo, K. and Pagel, V. (1988) Issues in Building General Letter to Sound Rules. *3rd ESCA Workshop on Speech Synthesis*. Jenolan Caves, Australia, pp 77-80.

Bosch, A. & Daelemans, W. (1993) Data-oriented methods for grapheme-to-Phoneme conversion. *Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*, Utrecht, Holanda, pp. 45-53.

Braga, Daniela & Coelho, Luís (2006) Letter-to-sound conversion for Galician TTS systems. *IV Jornadas en Tecnologías del Habla*, Zaragoza, Espanha. pp. 171-176.

Braga, Daniela; Coelho, Luís; Resende Jr., Fernando (2006) A Rule-Based Grapheme-to--Phone Converter for TTS Systems in European Portuguese. *VI International Telecommunications Symposium (ITS2006)*, Fortaleza, CE, Brasil.

Chotimongkol, A. & Black, A. (2000) Statistically trained orthographic to sound models for Thai. *Proceedings of ICSLP2000*. Beijing, China.

Coker, Cecil H.; Church, Kenneth W.; Liberman, Mark Y. (1990) Morphology and rhyming: Two powerful alternatives to letter-to-sound rules for speech synthesis. *Proceedings of the ESCA Workshop on Speech Synthesis*. Autrans, France, pp. 83-86.

Damper, R.I.; Marchand Y.; Adamson, M.J.; Gustafson, K. (1998) Comparative Evaluation of Letter-To-Sound Conversion Techniques For English Text-To-Speech Synthesis. *Proceedings of 3rd ESCA/COCOSDA International Workshop on Speech Synthesis*, Jenolan Caves, Australia, pp. 53-58.

Estrela, Edite; Soares, M. Almira; Leitão, M. José (2004) *Saber Escrever, Saber Falar*. Lisboa: Dom Quixote, pp. 37-41.

Gomes, L.C.T. (1998) *Sistema de conversão texto-fala para a língua portuguesa utilizando a abordagem de síntese por regras*. Dissertação de Mestrado. Campinas: Unicamp.

Huang, Xuedong; Acero, Alex; Hon, Hsisa-Wuen (2001) *Spoken Language Processing*. New Jersey: Prentice may PTR.

Kaplan, R. M. & Kay, M. (1994) Regular models of phonological rule systems. *Computational Linguistics 20(3)*, pp. 331-378.

Lucassen, J. M. & Mercer, R. L. (1984) Discovering Phonemic based forms automatically: an information theoretic approach. *IEEE International Conference on Acoustics, Speech and Signal Processing*. 42.5.1-42.5.4.

Mateus, M. & Andrade, E. (2000) *The Phonology of Portuguese*. Oxford: Oxford University Press.

Mateus, M. H. M., Andrade, A., Viana, M. C., Villalva, A. (1990) *Fonética, Fonologia e Morfologia do Português*, Lisboa: Universidade Aberta.

Meng, H. M.; Seneff, S.; Zue, V. (1994) Phonological parsing for bi-directional letter-to--sound/sound-to-letter generation. *ARPA Human Language Technology Workshop*. Princeton, USA.

Oliveira, Luís Caldas (1996) *Síntese de Fala a Partir de Texto*. Dissertação de Doutoramento. Universidade Técnica de Lisboa.

Paulo, S. G.; Oliveira, Luis C. (2005) Generation of Word Alternative Pronunciations Using Weighted Finite State Transducers. *Proceedings of Interspeech 2005*. Lisboa, Portugal.

Roche E. & Schabes, Y. (1995) *Exact Generalization of Finite-State Transductions: Application to Grapheme-to-Phoneme Transcription*. Technical Report TR-95-08, Mitsubishi Electric Research Laboratories, Cambridge, USA.

Rodrigues, Maria Celeste (2003) *Lisboa e Braga: Fonologia e Variação*. Lisboa: FCT/MCT.

Sejnowski, T.J. & Rosenberg, C. R. (1987) Parallel networks that learn to pronounce English Text. *Complex Systems, 1*, pp. 145-168.

Silva, D.; Lima, A.; Maia, R.; Braga, D.; Moraes, J.F.; Moraes, J.A.; Resende Jr., F.G.V. (2006) A rule-based grapheme-phone converter and stress determination for Brazilian Portuguese natural. *VI International Telecommunications Symposium (ITS2006)*, Fortaleza, CE, Brasil.

Taylor, Paul (2005) Grapheme-to- Phoneme conversion using Hidden Markov models. *Proceedings of Interspeech 2005*. Lisboa, Portugal.

Teixeira, António (2000) *Síntese Articulatória das vogais nasais do Português Europeu*. Dissertação de Doutoramento. Universidade de Aveiro.

Teixeira, J. P., Freitas, D., Gouveia, P. Olaszy, G., Németh, G. (1998) Multivox: Conversor Texto-Fala para Português. *Proceedings of PROPOR'98*. Novembro de 1998. Porto Alegre, Brazil.

Teixeira, J.P.; Gouveia, P., Freitas, D. (2000) Divisão silábica automática do texto escrito e falado. *Proceedings of PROPOR'2000*. Novembro de 2000. Atibaia, SP. Brasil.

Seara Jr., R.; Kafka, S.; Seara, I.; Pacheco, F.; Klein, S.; Seara, R. (2004) Parâmetros Lingüísticos Utilizados para a Geração Automática de Prosódia em Sistemas de Síntese de Fala. XXI *Simpósio Brasileiro de Telecomunicações – SBT 2004*. pp. 1-6, Belém, PA, Brasil.

Oliveira, C.; Moutinho, L.; Teixeira, A. (2004) Um novo sistema de conversão grafema-fone para PE baseado em transdutores. *Actas do II Congresso Internacional de Fonética e Fonologia*, Maranhão, Brasil.

Oliveira, C.; Moutinho, L.; Teixeira, A. (2005) On European Portuguese Automatic Syllabification. *Proceedings of Interspeech 2005*. Lisboa. Portugal.