Geração de Voz com Sotaque

José João Almeida
Alberto Manuel Simões
Projecto Natura
Departamento de Informática
Universidade do Minho

Como é sabido os sotaques podem estar ligados a uma zona geográfica, a um grupo social, podem até ser uma característica pessoal. O seu estudo e descrição têm interessado muitos investigadores embora normalmente esse estudo tenha sido feito de modo pouco formal.

No trabalho que aqui se relata, tentou-se descrever formalmente sotaques e disfunções através de criação de regras a integrar como variantes num gerador de voz. Deste modo, pretendeu-se criar um ambiente de experimentação dos modelos construídos para descrever algumas características de certos sotaques ou certas disfunções, de modo a permitir a sua validação.

Constatou-se que se consegue obter certas disfunções e certos sotaques com facilidade por simples acrescento de regras opcionais em certas fases da geração da voz. Outros aparentam ser de maior dificuldade, ou por não conhecermos suficientemente bem os fenómenos neles envolvidos ou envolverem maior complexidade prosódica.

Introdução

O nosso trabalho está baseado num sintetizador de Voz [1] escrito com base em regras de reescrita, denominado `Lingua::PT::speaker` que foi desenvolvido com os seguintes objectivos:

- estudo de transcrições fonéticas – embora muitos dicionários contenham transcrições fonéticas, nunca sabemos qual a sua correcção. Ao usar um conjunto de caracteres para representar sons, os lexicógrafos não têm qualquer forma sistemática para verificar a correcção das suas transcrições. O ser possível escrevê-las e ouvi-las torna este processo mais simples. Depois de algumas experiências com profissionais da área, ao tentar transcrever a palavra "*elefante*" para fonemas chegou-se à transcrição "*@lfant@*" como sendo a que mais agradava aos ouvidos humanos em contraste com a transcrição mais comum "*@l@fant@*".
- estudo de regras fonéticas – forma de testar sistematicamente regras fonéticas de gramáticas e de outras fontes, para que se possam experimentar em texto real e ouvir o seu resultado validando-o;

***Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 2002, pp. 47-53.**

- estudo de variações e sotaques – embora vivamos num pequeno país, é verdade que nem sempre nos entendemos com facilidade. O estudo de sotaques e a sua modelação por regras permite que possamos analisar quão sistemáticas elas são.

O `Lingua::PT::speaker` foi desenvolvido por um conjunto de regras de reescrita que afectam a geração de fonemas a diferentes níveis:

- tradução de partes não textuais para textuais;
- separação em frases e em palavras;
- separação em sílabas e detecção da tónica;
- transformação das palavra em fonemas;
- transformação de palavras adjacentes;
- aplicação de regras prosódicas.

A figura mostra de forma gráfica o fluxo das palavras e sua transformação para sons. A secção "*Geração de Voz*" apresenta-as de forma mais pormenorizada embora se aconselhe a leitura do artigo "*Text-to-speech – A rewriting system approach*" para melhor entender o seu funcionamento.

Estando este processo de geração estruturado em camadas, é possível aplicar regras a vários níveis. O nível de aplicação das regras é dependente do tipo de sotaque desejado. Serão apresentados vários exemplos de geração de sotaques na secção "*Geração de Voz*".

A formalização destas regras permite não só estudar a geração de voz com sotaque, mas também a compreensão da fonética da língua e o estudo das limitações do processo.

Geração de Voz

Como foi referido na introdução, o processo de geração de voz é efectuado por fases, que funcionam em cadeia baseados num sistema de reescrita.

O sistema de reescrita não é mais do que um programa que dado um conjunto de regras as aplica no texto até que não seja possível aplicar nenhuma dessas regras.

Estas regras podem ter o seguinte aspecto:

- `lado esquerdo ==> lado direito` – esta é a regra típica, sendo aplicada sempre que o lado esquerdo exista no texto. Nesse caso, este texto é substituído pelo conteúdo do lado direito;
- `lado esquerdo =e=> lado direito` – idêntica à regra básica. No entanto, o texto substituído contém código Perl que é substituído pelo resultado da sua avaliação;
- `lado esquerdo ==> lado direito !! condição` – esta regra é aplicada sempre que o lado esquerdo exista no texto mas só e só se a condição especificada também for verdadeira.

Não texto para texto

Qualquer texto que tentemos ler utilizando o sintetizador de voz está repleto de partes não textuais. Entendemos que são partes não textuais todo o texto que não é lido da forma convencional. Exemplos são os números (que não se apresentem por extenso), e-mails, endereços de Internet e outros.

A forma mais simples de lidar com este problema consiste em substituir cada um destes elementos pela forma de leitura respectiva. Embora pareça simples, este processo envolve um conjunto complexo de regras.

A conversão de números é um dos processos mais estudados. No entanto, alguns exemplos de números são lidos de forma ligeiramente diferente. Lemos "1230" como "*mil, duzentos e trinta*" e lemos "1003" como "*mil e três*". Este pormenor da colocação da conjunção torna a conversão menos trivial. O exemplo seguinte mostra oito regras das utilizadas para a conversão de números.

```
100==>cem                          900==>novecentos
1(\d\d)==>cento e $1                (\d)(\d\d)==>${1}00 e $2
0(\d\d)==>$1                       10==>dez
200==>duzentos                     1==>onze
```

Em relação a URLs e a e-mails temos problemas idênticos. Por exemplo, lemos o endereço electrónico "*jj@di.uminho.pt*" como "*jota jota arroba dê i ponto uminho ponto pê tê*", enquanto que lemos "*alberto@dominio.com*" como "*alberto arroba dominio ponto come*". Isto obriga à definição de um conjunto de heurísticas que indiquem quando as palavras devem ser soletradas ou lidas "à portuguesa".

Divisão em frases e palavras

A divisão do texto em frases e em palavras não é tão simples como se possa pensar. Abreviaturas, siglas e diversidade de pontuação fazem com que esta tarefa se possa tornar complexa. Para realizar esta tarefa estamos a utilizar um segmentador (*tokenizador*) de um módulo Perl para processamento de linguagem natural (`Lingua::PT::pln`).

De palavras a fonemas

Depois de termos palavras temos de as transformar em fonemas. Estes fonemas passam pelo gerador de voz que, dada uma sequência e uma base de fonemas, extraem os sons necessários de uma base de dados de fonemas e os juntam de forma a construir as palavras.

Como base de fonemas e respectivo programa de junção usamos o MBrola [2], que é de domínio público e portanto, gratuito. A linguagem fonética é bastante gráfica (caracteres gregos), existe uma convenção denominada SAMPA [3] que mapeia estes caracteres gregos em caracteres ASCII.

Este processo, mais uma vez, é realizado por um conjunto de regras, das quais apresentamos um pequeno extracto:

```
rr==>R                  ass==>6ss
^r==>R                  ss==>ç
([nls])r==>$1R          ^h==>
```

Tratamento de palavras adjacentes

Qualquer leitor acaba por, por vezes, juntar palavras. Ou seja, unir fins de algumas palavras com o começo de outras. Exemplo disso é a frase "*este elefante*" que muita gente iria acabar por ler "*estelefante*". Existem muitos casos destes e que, para permitir melhor entendimento, devem ser tratados:

```
(e|a)/\1==>/$1
6/6(?!~)==>/a
6/a==>/a
S/([a\@eA6iouOE])==>z/$1
\@/([\@eaui6])==>/$1
```

Neste exemplo a barra "*/*" denota os espaços entre palavras.

Regras prosódicas

Qualquer língua tem a sua música. Em particular, o Português tem algum tipo de modelação que torna algumas palavras ou formas frásicas mais fáceis de reconhecer.

Qualquer tipo de melodia que se coloque na sintetização torna-a mais audível. A diferença entre ouvir o texto sem qualquer melodia e ouvi-lo com uma melodia estranha é idêntica a não perceber nada e passar a perceber alguma coisa.

No entanto, definiram-se regras prosódicas que tentam simular a forma habitual de um leitor Português.

Geração de Sotaque

As diferentes pronúncias, sotaques, etc, quando são sistemáticos, correspondem a diferentes modos de transformar o texto escrito em som e como tal correspondem a processos de geração diferente.

Quando se usa um gerador de voz que seja definido por regras, essas regras definem o processo de transformação que partindo de um texto e de uma base de fonemas produzem o som.

Por vezes é muito trabalhoso ou mesmo impossível obter o som com a pronúncia/sotaque desejados a partir da base de fonemas de que dispomos. Essas dificul-

dades resultam essencialmente de que certas variantes dependem de prosódia mais complexas, ou dependem de certos sons que diferem bastante dos fonemas existentes na base de fonemas.

No entanto há um conjunto de casos que podem ser obtidos por simples alteração de algumas das regras gerais. Quando tal acontece, essas diferenças nas regras podem ser forçadas utilizando pequenos sistemas de reescrita que são inseridos na sequência de processamento.

Deste modo diferentes sotaques e pronúncias podem ser obtidos com conjuntos diferentes de regras.

A geração de sotaque pode ser introduzida em vários níveis deste processo. Alguns tornam-se mais simples de introduzir em determinados níveis do que outros, e ainda há aqueles que são praticamente impossíveis de reproduzir.

Nos exemplos que se seguem não é apresentado nenhum sotaque completo mas apenas regras que permitem descrever fenómenos parciais que existem em acentos de vários locais. Um sotaque seria descrito através da junção de um conjunto de fenómenos parciais.

Dado que é nossa intenção poder aplicar estas variantes em protótipos animados, apresentaremos também regras ligadas a dislexias como por exemplo aquela que caracteriza o personagem Cebolinha da banda desenhada Mônica.

Trocar os V pelos B

O fenómeno de trocar os Vs pelos Bs, bastante habitual pelo menos na zona norte do país,pode ser descrito por uma simples regra de alteração textual:

```
v==>b
```

Cebolinha

Embora não seja propriamente um sotaque mas uma dislexia, o facto de se conseguir introduzir esta modelação no conjunto de regras geral provou que a introdução de sotaques no gerador já existente era realizável.

As regras cingem-se a:

```
RULES cebolinha

rr==>r
r==>l
```

Sotaque de *Bijeu*

O normalmente designado por sotaque de Viseu caracteriza-se pelas seguintes regras:

```
MRULES viseu

v==>b
s==>S
z==>Z
S==>Z
```

ou seja por:
- trocar V por B;
- transformar o som ss em x;
- transformar z e x em j;

Na realidade em Viseu não se fala assim. Um dos aspectos mais característicos corresponde a reforçar os z:

```
MRULES viseu2
z==>zz
```

Interface do `Lingua::PT::speaker`

O módulo `Lingua::PT::speaker` apresenta não só um conjunto de funções para interface em programas Perl mas também uma aplicação que recebe um ficheiro de texto e o lê. Esta aplicação, denominada "fala", permite ainda que se lhe passe opções como sejam o sotaque escolhido:

-cebolinha	dislexia do Cebolinha;
-porto	fala do "nuerte";
-lamego	talvez com nome errado, este substitui os "che" por "tche";
-viseu	fala de "bijeu";
-spain	tentativa de estudo de português a falar espanhol (ou vice-versa).

Resultados e Trabalho Futuro

A possibilidade de definir de forma coerente quer as regras de geração de fonemas quer as utilizadas para a geração de sotaques torna este assunto discutível com bases formais.

Constatou-se que se consegue obter certas disfunções (exemplo: o personagem Cebolinha da banda desenhada que troca "r"s por "l"s) e certos sotaques (exemplo: um subconjunto do sotaque normalmente designado como de Viseu) com facilidade por simples acrescento de regras opcionais em certos blocos.

É no entanto bastante mais complexo a tarefa de descrever sotaques que envolvem alterações prosódicas significativas (como exemplo, o sotaque de Tia de

Cascais) – do mesmo modo que as regras prosódicas são mais complexas que outras regras de transcrição fonética.

Por outro lado, a existência de um gerador de voz configurável e programável torna simples a sua inclusão em projectos maiores como seja um "*browser*" ou outras aplicações para invisuais.

Referências

[1] José João Almeida and Alberto Manuel Simões. Text-to-speech – "A rewriting system approach". In *Sociedade Espanhola de Processamento de Linguagem Natural*, 2001

[2] *The MBrola Project: Twoards a Freely Available Multilingual Speech Synthesizer.* http://tcts.fpms.ac.be/synthesis/mbrola.html

[3] *SAMPA: computer readable phonetic alphabet.* http://www.phon.ucl.ac.uk/home/sampa/home.htm