

HE[eee]SIT[u]AÇÃO: Modelos de Hesitação para o Português

Sara Candeias⁽¹⁾⁽²⁾, Fernando Perdigão⁽¹⁾⁽³⁾, Arlindo da Veiga⁽¹⁾; Carla Lopes⁽¹⁾⁽⁴⁾

⁽¹⁾ Instituto de Telecomunicações – Pólo de Coimbra

⁽²⁾ ESEC/IPC

⁽³⁾ DEEC-FCTUC

⁽⁴⁾ ESTG/IPL

{saracandeias,fp,calopes,aveiga}@co.it.pt

As **hesitações** tornaram-se alvo de atenção na última década quer com o desenvolvimento dos sistemas de conversão de fala para texto quer com outros esforços para aumentar o conhecimento da fala humana. Embora a tecnologia de fala esteja a ter um desenvolvimento significativo, na verdade, a maioria dos actuais sistemas de reconhecimento carecem de conhecimento pertinente sobre os eventos de hesitação. Esta é a razão pela qual literatura sobre o tema tem vindo a apreciar de forma crescente os fenómenos de hesitação como componentes importantes da linguagem natural.

Representando um desafio interessante para a tecnologia da fala, este projecto pretende analisar e descrever, no Português Europeu falado em contínuo, fenómenos como pausas preenchidas, extensões fonéticas de segmentos, palavras cortadas e repetições sob a designação de hesitações.

Apesar de recorrente, a presença de hesitações no discurso falado tem sido entendida como uma irregularidade ou como um erro de estruturação frásica, e a maioria dos trabalhos linguísticos sobre o português europeu falado que têm sido referenciados na literatura apresentam o seu estudo em contextos particulares e com resultados ainda preliminares. O conhecimento dos fenómenos de hesitação presentes no português manifesta-se ainda titubeante e incertezas sobre qual o padrão de características regulares que definem os eventos ainda permanecem - aspectos que este projecto se propõe investigar e esclarecer.

O objectivo deste projecto é **construir um modelo linguístico da produção de hesitações**, considerando o Português europeu falado de forma espontânea bem como disponibilizar uma base de dados de eventos de hesitação.

Adoptar-se-á uma abordagem que faz uso de bases de dados de fala não preparada nem transcrita (notícias de televisão e outros conteúdos). O método inclui a análise das estruturas linguísticas e de representação das hesitações, sendo os eventos de hesitação detectados automaticamente a partir do sinal de fala. O modelo linguístico que propomos compreenderá a detecção e a análise das características acústico-prosódicas e morfossintácticas pertinentes e regulares na produção dos eventos de hesitação, em linha com estudos realizados para outras línguas. O padrão de hesitação resultará de índices invariáveis quer das características particulares dos segmentos quer do fraseado prosódico como um todo. As relações estabelecidas entre os níveis acústico-prosódico e morfossintáctico serão também exploradas. Para adensar a caracterização dos eventos (em especial das palavras cortadas e das extensões), serão igualmente avaliadas questões relacionadas com a constituição silábica e com o processo de coarticulação entre fones.

O projecto tem igualmente como desafio a detecção automática de eventos de hesitação a partir do sinal de fala, dando continuidade aos interesses de pesquisa de cada membro da equipa, que é multidisciplinar e que encontram nele um objectivo comum. Os recursos resultantes deste projecto (como o é a base de dados) poderão vir a ser ampliados e a permitir a formação crescente da comunidade científica sobre o idioma do Português.

INSTITUIÇÕES ASSOCIADAS:



INSTITUTO
SUPERIOR
TÉCNICO



Faculdade de Ciências
e Tecnologia da
Universidade de Coimbra



universidade
de aveiro



Inovação



SIEMENS
Communications

Aveiro

Instituto de Telecomunicações
Campus Universitário de Santiago
3810-193 Aveiro - Portugal
Tel: +351 234 377 900
Fax: +351 234 377 901
e-mail: info@av.it.pt

Coimbra

Instituto de Telecomunicações
Departamento de Engenharia
Electrotécnica e de Computadores
Universidade de Coimbra - Pólo II
3030-290 Coimbra - Portugal
Tel: + 351 239 796 236
Fax: + 351 239 796 293
e-mail: it@co.it.pt

Lisboa

Instituto de Telecomunicações
Instituto Superior Técnico
Torre Norte, piso 10
Av. Rovisco Pais, 1
1049-001 Lisboa - Portugal
Tel: +351 218 418 454
Fax: +351 218 418 472
e-mail: info@lx.it.pt