

Nuno J. Mamede Jorge Baptista  
Isabel Trancoso  
Maria das Graças Volpe Nunes (Eds.)

# Computational Processing of the Portuguese Language

6th International Workshop, PROPOR 2003  
Faro, Portugal, June 26-27, 2003  
Proceedings



Springer

# Table of Contents

## Speech Analysis and Recognition

Devoicing Measures of European Portuguese Fricatives .....	1
<i>Luis M.T. Jesus, Christine H. Shadle</i>	
AUDIMUS.MEDIA: A Broadcast News Speech Recognition System for the European Portuguese Language .....	9
<i>Hugo Meinedo, Diamantino Caseiro, João Neto, Isabel Trancoso</i>	
Pitch Restoration for Robust Speech Recognition .....	18
<i>Carlos Lima, Adriano Tavares, Carlos Silva</i>	

## Speech Synthesis

Grapheme-Phone Transcription Algorithm for a Brazilian Portuguese TTS .....	23
<i>Filipe Barbosa, Guilherme Pinto, Fernando Gil Resende, Carlos Alexandre Gonçalves, Ruth Monserrat, Maria Carlota Rosa</i>	
Improving the Accuracy of the Speech Synthesis Based Phonetic Alignment Using Multiple Acoustic Features .....	31
<i>Sérgio Paulo, Luís C. Oliveira</i>	
Evaluation of a Segmental Durations Model for TTS .....	40
<i>João Paulo Teixeira, Diamantino Freitas</i>	
From Portuguese to Mirandese: Fast Porting of a Letter-to-Sound Module Using FSTs .....	49
<i>Isabel Trancoso, Céu Viana, Manuela Barros, Diamantino Caseiro, Sérgio Paulo</i>	
A Methodology to Analyze Homographs for a Brazilian Portuguese TTS System .....	57
<i>Filipe Barbosa, Lúlian Ferrari, Fernando Gil Resende</i>	
Automatic Discovery of Brazilian Portuguese Letter to Phoneme Conversion Rules through Genetic Programming .....	62
<i>Evandro Franzen, Dante Augusto Couto Barone</i>	
Experimental Phonetics Contributions to the Portuguese Articulatory Synthesizer Development .....	66
<i>António Teixeira, Lurdes Castro Moutinho, Rosa Lúdia Coimbra</i>	

## Pragmatics, Discourse, Semantics, Syntax, and the Lexicon

A Study on the Reliability of Two Discourse Segmentation Models . . . . .	70
<i>Eva Arim, Francisco Costa, Tiago Freitas</i>	
Reusability of Dictionaries in the Compilation of NLP Lexicons . . . . .	78
<i>Bento C. Dias-da-Silva, Mirna F. de Oliveira, Helio R. de Moraes</i>	
Homonymy in Natural Language Processes: A Representation Using Pustejovsky's Qualia Structure and Ontological Information . . . . .	86
<i>Claudia Zavaglia, Juliana Galvani Greggi</i>	
Using Adaptive Formalisms to Describe Context-Dependencies in Natural Language . . . . .	94
<i>João José Neto, Miryam de Moraes</i>	
Some Regularities of Frozen Expressions in Brazilian Portuguese . . . . .	98
<i>Oto Araújo Vale</i>	

## Tools, Resources, and Applications

Selva: A New Syntactic Parser for Portuguese . . . . .	102
<i>Sheila de Almeida, Ariadne Carvalho, Lucien Fantin, Jorge Stolfi</i>	
An Account of the Challenge of Tagging a Reference Corpus for Brazilian Portuguese . . . . .	110
<i>Sandra Aluísio, Jorge Pelizzoni, Ana Raquel Marchi, Lucélia de Oliveira, Regiana Manenti, Vanessa Marquiasável</i>	
Multi-level NER for Portuguese in a CG Framework . . . . .	118
<i>Eckhard Bick</i>	
HMM/MLP Hybrid Speech Recognizer for the Portuguese Telephone SpeechDat Corpus . . . . .	126
<i>Astrid Hagen, João P. Neto</i>	
Managing Linguistic Resources and Tools . . . . .	135
<i>David M. de Matos, Joana L. Paulo, Nuno J. Mamede</i>	
Using Morphosyntactic Information in TTS Systems: Comparing Strategies for European Portuguese . . . . .	143
<i>Ricardo Ribeiro, Luís Oliveira, Isabel Trancoso</i>	
Timber! Issues in Treebank Building and Use . . . . .	151
<i>Diana Santos</i>	
A Lexicon-Based Stemming Procedure . . . . .	159
<i>Gilberto Silva, Claudia Oliveira</i>	

Contractions: Breaking the Tokenization-Tagging Circularity . . . . .	167
<i>António Horta Branco, João Ricardo Silva</i>	
A Linguistic Approach Proposal for Mechanical Design Using Natural Language Processing . . . . .	171
<i>João Carlos Linhares, Altamir Dias</i>	
Identification of Direct/Indirect Discourse in Children's Stories . . . . .	175
<i>Nuno J. Mamede</i>	
Curupira: A Functional Parser for Brazilian Portuguese . . . . .	179
<i>Ronaldo Martins, Graça Nunes, Ricardo Hasegawa</i>	
ANELL: A Web System for Portuguese Corpora Annotation . . . . .	184
<i>Cristina Mota, Pedro Moura</i>	
Email2Vmail – An Email Reader . . . . .	189
<i>Hugo Pereira, Pedro Teixeira, Luís C. Oliveira</i>	
A Large Speech Database for Brazilian Portuguese Spoken Language Research . . . . .	193
<i>Carlos Alberto Ynoguti, Plínio Almeida Barbosa, Fábio Violaro</i>	
<b>Dialogue Systems</b>	
Interpretations and Discourse Obligations in a Dialog System . . . . .	197
<i>Márcio Mourão, Pedro Madeira, Nuno J. Mamede</i>	
Using Dialogues to Access Semantic Knowledge in a Web IR System . . . . .	201
<i>Paulo Quaresma, Irene Rodrigues</i>	
Managing Dialog and Access Control in Natural Language Querying . . . . .	206
<i>Luís Quintano, Irene Rodrigues</i>	
<b>Summarization and Information Extraction</b>	
GistSumm: A Summarization Tool Based on a New Extractive Method . . . . .	210
<i>Thiago Alexandre Salgueiro Pardo, Lucia Helena Machado Rino, Maria das Graças Volpe Nunes</i>	
Topic Indexing of TV Broadcast News Programs . . . . .	219
<i>Rui Amaral, Isabel Trancoso</i>	
<b>Evaluation</b>	
An Initial Proposal for Cooperative Evaluation on Information Retrieval in Portuguese . . . . .	227
<i>Rachel Aires, Sandra Aluísio, Paulo Quaresma, Diana Santos, Mário J. Silva</i>	

Evaluation of Finite-State Lexical Transducers of Temporal Adverbs for Lexical Analysis of Portuguese Texts .....	235
<i>Jorge Baptista</i>	
Evaluating Automatically Computed Word Similarity .....	243
<i>Caroline Varaschin Gasperin, Vera Lúcia Strube de Lima</i>	
Evaluation of a Thesaurus-Based Query Expansion Technique .....	251
<i>Luiz Augusto Sangoi Pizzato, Vera Lúcia Strube de Lima</i>	
Cooperatively Evaluating Portuguese Morphology .....	259
<i>Diana Santos, Luís Costa, Paulo Rocha</i>	
<b>Author Index</b> .....	267