

Lecture Notes in Artificial Intelligence 3960

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Renata Vieira Paulo Quaresma
Maria das Graças Volpe Nunes
Nuno J. Mamede Cláudia Oliveira
Maria Carmelita Dias (Eds.)

Computational Processing of the Portuguese Language

7th International Workshop, PROPOR 2006
Itatiaia, Brazil, May 13-17, 2006
Proceedings



Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Renata Vieira

Universidade do Vale de Rio de Sinos, UNISINOS, São Leopoldo, Brazil
E-mail: renata@unisinós.br

Paulo Quaresma

Universidade de Évora, Portugal
E-mail: pq@di.uevora.pt

Maria das Graças Volpe Nunes

Universidade de São Paulo, USP, São Carlos, SP, Brazil
E-mail: gracan@icmc.usp.br

Nuno J. Mamede

Universidade Técnica de Lisboa, L2F, INESC-ID, Lisboa, Portugal
E-mail: Nuno.Mamede@inesc-id.pt

Cláudia Oliveira

Instituto Militar de Engenharia, Rio de Janeiro, Brazil
E-mail: cmaria@de9.ime.eb.br

Maria Carmelita Dias

Pontifícia Universidade Católica do Rio de Janeiro, PUC-Rio, Rio de Janeiro, Brazil
E-mail: mcdias@let.puc-rio.br

Library of Congress Control Number: 2006925088

CR Subject Classification (1998): I.2.7, F.4.2-3, I.2, H.3, I.7

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-540-34045-9 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-34045-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2006

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11751984 06/3142 5 4 3 2 1 0

Preface

Since 1993, PROPOR Workshops have become an important forum for researchers involved in the Computational Processing of Portuguese, both written and spoken. This PROPOR Workshop follows previous workshops held in 1993 (Lisbon, Portugal), 1996 (Curitiba, Brazil), 1998 (Porto Alegre, Brazil), 1999 (Évora, Portugal), 2000 (Atibaia, Brazil) and 2003 (Faro, Portugal). The workshop has increasingly contributed to bringing together researchers and industry partners from both sides of the Atlantic.

The constitution of an international Program Committee and the adoption of high-standard referee procedures demonstrate the steady development of the field and of its scientific community.

In 2006 PROPOR received 56 paper submissions from 11 different countries: Brazil, Portugal, Spain, Norway, USA, Italy, Japan, France, Canada, Denmark and the UK, from which 9 are represented in the accepted papers.

Each submitted paper underwent a careful, triple-blind review by the Program Committee. All those who contributed are mentioned in the following pages. The reviewing process led to the selection of 20 regular papers for oral presentation and 17 short papers for poster sections, which are published in this volume.

The workshop and this book were structured around the following main topics, seven for full papers: (i) automatic summarization; (ii) resources; (iii) automatic translation; (iv) named entity recognition; (v) tools and frameworks; (vi) systems and models; and another five topics for short papers; (vii) information extraction; (viii) speech processing; (ix) lexicon; (x) morpho-syntactic studies; (xi) web, corpus and evaluation.

We are especially grateful to our invited speakers, Adam Kilgarriff (Visiting Research Fellow at the Department of Informatics, University of Sussex) and Marcelo Finger (IME, University of São Paulo), for their invaluable contribution, which undoubtedly increased the interest in the workshop and its quality.

We are indebted to all members of our Technical Program Committee and additional reviewers, as mentioned in the following pages.

May 2006

Renata Vieira
Paulo Quaresma
Maria das Graças Volpe Nunes
Nuno J. Mamede
Claudia Oliveira
Maria Carmelita Dias

Organization

Conference Chairs

Claudia Oliveira (Instituto Militar de Engenharia, Brazil)

Maria Carmelita Dias (Pontifícia Universidade Católica do Rio de Janeiro, Brazil)

Program Co-chair

Renata Vieira (Universidade do Vale do Rio dos Sinos, Brazil)

Paulo Quaresma (Universidade de Évora, Portugal)

Publication Chairs

Renata Vieira (Universidade do Vale do Rio dos Sinos, Brazil)

Paulo Quaresma (Universidade de Évora, Portugal)

Maria das Graças Volpe Nunes (Universidade de São Paulo, Brazil)

Nuno Mamede (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento, Portugal)

Program Committee

Aldebaro Klautau (Universidade Federal do Pará, Brazil)

Aline Villavicencio (Universidade Federal do Rio Grande do Sul, Brazil)

Ana Frankenberg-Garcia (Instituto Superior de Línguas e Administração, Portugal)

António Branco (Universidade de Lisboa, Portugal)

António Ribeiro (European Commission, Joint Research Centre, Italy)

António Teixeira (Universidade do Aveiro, Portugal)

Ariadne Carvalho (Universidade de Campinas, Brazil)

Belinda Maia (Universidade do Porto, Portugal)

Bento Carlos Dias da Silva (Universidade Estadual Paulista, Brazil)

Berthold Crysmann (German Research Centre for Artificial Intelligence, Germany)

Carlos Augusto Prolo (Pontifícia Universidade Católica do Rio Grande do Sul, Brazil)

Caroline Hagège (Xerox Research Centre, France)

Claudia Oliveira (Instituto Militar de Engenharia, Brazil)

Diana Santos (The Foundation for Scientific and Industrial Research at the Norwegian Institute of Technology, Norway)

Diamantino Freitas (Universidade do Porto, Portugal)

Eckhard Bick (Southern Denmark University, Denmark)

Egídio Terra (Pontifícia Universidade Católica do Rio Grande do Sul, Brazil)

Elisabete Ranchhod (Universidade de Lisboa, Portugal)

Eric Laporte (University of Marne-la-Vallée, France)

Gabriel Pereira Lopes (Universidade Nova de Lisboa, Portugal)
Horacio Saggion (University of Sheffield, UK)
Irene Rodrigues (Universidade de Évora, Portugal)
Isabel Trancoso (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento, Portugal)
Jason Baldrige (University of Texas at Austin, USA)
João Paulo Neto (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento, Portugal)
Jorge Baptista (Universidade do Algarve, Portugal)
Louisa Sadler (University of Essex, UK)
Lúcia Rino (Universidade Federal de São Carlos, Brazil)
Luís Caldas Oliveira (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento, Portugal)
Marcelo Finger (Universidade de São Paulo, Brazil)
Maria Carmelita Dias (Pontifícia Universidade Católica do Rio de Janeiro, Brazil)
Maria das Graças Volpe Nunes (Universidade de São Paulo, Brazil)
Massimo Poesio (University of Essex, UK)
Michel Gagnon (École Polytechnique Montréal, Canada)
Mikel L. Forcada (Universitat d'Alacant, Spain)
Nuno Mamede (Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento, Portugal)
Pablo Gamallo (Universidade de Compostela, Spain)
Palmira Marrafa (Universidade de Lisboa, Portugal)
Paulo Quaresma (Universidade de Évora, Portugal)
Renata Vieira (Universidade do Vale do Rio dos Sinos, Brazil)
Rove Chishman (Universidade do Vale do Rio dos Sinos, Brazil)
Sandra Aluisio (Universidade de São Paulo, Brazil)
Saturnino Luz (University of Dublin, Ireland)
Tracy Holloway King (Palo Alto Research Center, USA)
Vera Lúcia Strube de Lima (Pontifícia Universidade Católica do Rio Grande do Sul, Brazil)
Violeta Quental (Pontifícia Universidade Católica do Rio de Janeiro, Brazil)

Additional Reviewers

Daniel Muller (Universidade Federal do Rio Grande do Sul, Brazil)
Geert-Jan Kruijff (Deutsche Forschungszentrum für Künstliche Intelligenz, Germany)
José Carlos Mombach (Universidade do Vale do Rio dos Sinos, Brazil)
Maria Claudia de Freitas (Pontifícia Universidade Católica do Rio de Janeiro, Brazil)
Milena Garrão (Pontifícia Universidade Católica do Rio de Janeiro, Brazil)
Ronaldo Teixeira Martins (University of Mackenzie, Brazil)
Silvia Moraes (Pontifícia Universidade Católica do Rio Grande do Sul, Brazil)
Thiago Pardo (Universidade de São Paulo, Brazil)

Table of Contents

Summarization

Modeling and Evaluating Summaries Using Complex Networks <i>Thiago Alexandre Salgueiro Pardo, Lucas Antigueira, Maria das Graças Volpe Nunes, Oswaldo N. Oliveira Jr., Luciano da Fontoura Costa</i>	1
SABIO: An Automatic Portuguese Text Summarizer Through Artificial Neural Networks in a More Biologically Plausible Model <i>Télvio Orrú, João Luís Garcia Rosa, Márcio Luiz de Andrade Netto</i>	11

Resources

Building a Dictionary of Anthroponyms <i>Jorge Baptista, Fernando Batista, Nuno Mamede</i>	21
REPENTINO – A Wide-Scope Gazetteer for Entity Recognition in Portuguese <i>Luís Sarmiento, Ana Sofia Pinto, Luís Cabral</i>	31

Translation

Using Natural Alignment to Extract Translation Equivalents <i>Pablo Gamallo Otero</i>	41
Open-Source Portuguese–Spanish Machine Translation <i>Carme Armentano-Oller, Rafael C. Carrasco, Antonio M. Corbí- Bellot, Mikel L. Forcada, Mireia Ginestí-Rosell, Sergio Ortiz- Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, Miriam A. Scalco</i>	50
Weighted Finite-State Transducer Inference for Limited-Domain Speech-to-Speech Translation <i>Diamantino Caseiro, Isabel Trancoso</i>	60

Named Entity Recognition

A Golden Resource for Named Entity Recognition in Portuguese <i>Diana Santos, Nuno Cardoso</i>	69
---	----

Functional Aspects in Portuguese NER
Eckhard Bick 80

SIEMÊS - A Named-Entity Recognizer for Portuguese Relying on
 Similarity Rules
Luís Sarmento 90

Tools and Frameworks

Tools for Nominalization: An Alternative for Lexical Normalization
*Marco Antonio Insaurriaga Gonzalez, Vera Lúcia Strube de Lima,
 José Valdeni de Lima* 100

A Framework for Integrating Natural Language Tools
João Graça, Nuno J. Mamede, João D. Pereira 110

Methods and Tools for Encoding the WordNet.Br Sentences, Concept
 Glosses, and Conceptual-Semantic Relations
Bento C. Dias-da-Silva, Ariani Di Felippo, Ricardo Hasegawa 120

A Multi-agent Approach to Question Answering
*Cássia Trojahn dos Santos, Paulo Quaresma, Irene Rodrigues,
 Renata Vieira* 131

Systems and Models

Adaptation of Data and Models for Probabilistic Parsing of Portuguese
Benjamin Wing, Jason Baldridge 140

A Set of NP-Extraction Rules for Portuguese: Defining, Learning and
 Pruning
*Claudia Oliveira, Maria Claudia Freitas, Violeta Quental,
 Cícero Nogueira dos Santos, Renato Paes Leme, Lucas Souza* 150

Resolving Portuguese Nominal Anaphora
*Jorge C.B. Coelho, Vinicius M. Muller, Sandra Collovini,
 Renata Vieira, Lucia H.M. Rino* 160

Design of a Multimodal Input Interface for a Dialogue System
João P. Neto, Renato Cassaca, Márcio Viveiros, Márcio Mourão 170

Review and Evaluation of DiZer – An Automatic Discourse Analyzer
 for Brazilian Portuguese
Thiago Alexandre Salgueiro Pardo, Maria das Graças Volpe Nunes . . . 180

Classroom Lecture Recognition <i>Isabel Trancoso, Ricardo Nunes, Luís Neves</i>	190
--	-----

Information Extraction

Semi-supervised Learning for Portuguese Noun Phrase Extraction <i>Ruy Milidiú, Cicero Santos, Julio Duarte, Raúl Rentería</i>	200
--	-----

Automatic Extraction of Keywords for the Portuguese Language <i>Maria Abadia Lacerda Dias, Marcelo de Gomensoro Malheiros</i>	204
--	-----

Semi-automatically Building Ontological Structures from Portuguese Written Texts <i>Túlio Lima Baségio, Vera Lúcia Strube de Lima</i>	208
--	-----

Speech Processing

On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme-Phone Conversion <i>António Teixeira, Catarina Oliveira, Lurdes Moutinho</i>	212
---	-----

A Model to Computational Speech Understanding <i>Daniel Nehme Müller, Mozart Lemos de Siqueira, Philippe O.A. Navaux</i>	216
---	-----

Phonetic Sequence to Graphemes Conversion Based on DTW and One-Stage Algorithms <i>Rafael Teruszkin, Fernando Gil Vianna Resende Jr.</i>	220
---	-----

Lexicon

Very Strict Selectional Restrictions: A Comparison Between Portuguese and French <i>Éric Laporte, Christian Leclère, Maria Carmelita Dias</i>	225
--	-----

Towards a Formalization of Tense and Aspect for the Generation of Portuguese Sentences <i>Michel Gagnon, Eliana de Mattos Pinto Coelho, Roger Antonio Finger</i>	229
---	-----

The Need for Application-Dependent WSD Strategies: A Case Study in MT <i>Lucia Specia, Gabriela Castelo Branco Ribeiro, Maria das Graças Volpe Nunes, Mark Stevenson</i>	233
A Lexical Database of Portuguese Multiword Expressions <i>Sandra Antunes, Maria Fernanda Bacelar do Nascimento, João Miguel Casteleiro, Amália Mendes, Luísa Pereira, Tiago Sá</i>	238
Morpho-syntactic Studies	
Dedicated Nominal Featurization of Portuguese <i>António Branco, João Ricardo Silva</i>	244
Part-of-Speech Tagging of Portuguese Based on Variable Length Markov Chains <i>Fábio Natanael Kepler, Marcelo Finger</i>	248
From Syntactical Analysis to Textual Segmentation <i>Ana Luísa Leal, Paulo Quaresma, Rove Chishman</i>	252
Syntactical Annotation of COMPARA: Workflow and First Results <i>Susana Inácio, Diana Santos</i>	256
Web, Corpus and Evaluation	
A Complex Evaluation Architecture for HAREM <i>Nuno Seco, Diana Santos, Nuno Cardoso, Rui Vilela</i>	260
What Kinds of Geographical Information Are There in the Portuguese Web? <i>Marcirio Silveira Chaves, Diana Santos</i>	264
Corpus-Based Compositionality <i>Milena Garrão, Claudia Oliveira, Maria Claudia de Freitas, Maria Carmelita Dias</i>	268
Author Index	273