

Lecture Notes in Artificial Intelligence

13208

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Wolfgang Wahlster

DFKI, Berlin, Germany

Zhi-Hua Zhou

Nanjing University, Nanjing, China

Founding Editor

Jörg Siekmann

DFKI and Saarland University, Saarbrücken, Germany


More information about this subseries at <https://link.springer.com/bookseries/1244>

Vlória Pinheiro · Pablo Gamallo ·
Raquel Amaro · Carolina Scarton ·
Fernando Batista · Diego Silva · Catarina Magro ·
Hugo Pinto (Eds.)


Computational Processing of the Portuguese Language


15th International Conference, PROPOR 2022
Fortaleza, Brazil, March 21–23, 2022
Proceedings


Editors


Vlória Pinheiro 
Universidade de Fortaleza
Fortaleza, Brazil

Raquel Amaro 
Universidade Nova de Lisboa
Lisbon, Portugal

Fernando Batista 
INESC-ID
Lisbon, Portugal

Catarina Magro 
University of Lisbon
Lisbon, Portugal

Pablo Gamallo 
CiTIUS - Universidade de Santiago de
Compostela
Santiago de Compostela, Spain

Carolina Scarton 
University of Sheffield
Sheffield, UK

Diego Silva 
Federal University of São Carlos
São Carlos, Brazil

Hugo Pinto
Sentimonitor
Porto Alegre, Brazil

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-98304-8 ISBN 978-3-030-98305-5 (eBook)
<https://doi.org/10.1007/978-3-030-98305-5>

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume presents the 15th edition of the International Conference on the Computational Processing of Portuguese (PROPOR 2022), held at the University of Fortaleza, in Fortaleza, Ceará, in Brazil, during March 21–23, 2022. Given the situation of the COVID-19 pandemic, PROPOR 2022 was moved to fully online.

Coming into its 30th year of existence, PROPOR continues to be the main scientific meeting in the fields of language and speech technologies for Portuguese and on the basic and applied research issues related to this language. The meeting is a very rich forum for researchers dedicated to the computational processing of Portuguese, promoting the exchange of experiences in the development of methodologies, language resources, tools, applications, and innovative projects. In this edition, a new track division was proposed with a new Linguistics and Resources track and an Industry track, alongside the General track. A total of 88 submissions were received for the main event, of which 33 were to the General track, 29 to the Linguistics and Resources track, and 6 to the Industry track, involving 152 different authors from many countries worldwide, such as Portugal, Brazil, Spain, Norway, France, and Germany.

This volume presents 36 full and 4 short papers, selected from the 88 submissions, which reflects an acceptance rate of 51% for full papers and 45% overall. Each paper was reviewed by 3 reviewers from a total of 59 academic institutions and companies. Full papers are organized thematically and show research and developments in resources and evaluation, natural language processing tasks, natural language processing applications, speech processing and applications, and lexical semantics. In this edition, reflecting the current trends in the field, the topics with the largest number of submissions and acceptance were related to resources and evaluation and natural language tasks and applications. Following the PROPOR tradition of welcoming and introducing new members to the community, there is also a section for the best dissertation and thesis, selected from the 6 candidates presented this year.

For PROPOR 2022, the selected team of invited speakers represents the worldwide importance of the field, not only in academia but also in very well-established companies. They were as follows:

- Maria José Bocorny Finatto, Full Professor at the Federal University of Rio Grande do Sul, Brazil
- Rada Mihalcea, Professor at the University of Michigan, USA
- Sebastião Miranda, Head of Development at Priberam, Portugal.

In parallel, PROPOR 2022 hosted 3 workshops on the following topics: automatic speech recognition for spontaneous and prepared speech and speech emotion recognition in Portuguese; universal dependencies; and digital humanities and natural language processing. The proceedings of the workshops are not included in this volume as the

organizers of each workshop arranged them separately. The event also included a session for systems demonstration and a tutorial, focusing on parallel corpora building and lemmatization criteria and annotation systems.

Our sincere thanks go to every person and institution involved in the complex organization of this event, especially to the members of the Program Committee of the main event, the dissertations contest and the associated workshops chairs, the invited speakers, and the general organization staff. We are also grateful to the agencies and organizations that supported and promoted the event.

January 2022

Vlória Pinheiro
Pablo Gamallo
Raquel Amaro
Carolina Scarton
Fernando Batista
Diego Silva
Catarina Magro
Hugo Pinto

Organization

General Chairs

Vlândia Pinheiro

Pablo Gamallo

University of Fortaleza, Brazil

University of Santiago de Compostela, Spain

Program Chairs

Raquel Amaro

Carolina Scarton

NOVA University Lisbon, Portugal

University of Sheffield, UK

Editorial Chairs

Fernando Batista

Diego Silva

INESC-ID and Iscte, Portugal

Federal University of São Carlos, Brazil

Demo Chairs

Wellington Franco

Juliano Norberto Sales

Luís Trigo

Federal University of Ceará, Brazil

University of Passau, Germany

University of Porto, Portugal

Best MSc/MA and PhD Dissertation Award Chairs

Daniela Claro

Idalete Dias

Federal University of Bahia, Brazil

University of Minho, Portugal

Workshops and Tutorials Chairs

Larissa de Freitas

Marcos Garcia

Federal University of Pelotas, Brazil

University of Santiago de Compostela, Spain

Track Chair Linguistics and Resources

Catarina Magro

University of Lisbon, Portugal

Track Chair Industry

Hugo Pinto

Sentimonitor, Brazil

Local Organization Committee

Vlória Pinheiro

University of Fortaleza, Brazil

Wellington Franco

Federal University of Ceará, Brazil

Steering Committee

Helena Caseli

Federal University of São Carlos, Brazil

Amália Mendes

University of Lisbon, Portugal

Vlória Pinheiro

University of Fortaleza, Brazil

Paulo Quaresma

University of Évora, Portugal

Evandro Seron Ruiz

University of São Paulo, Brazil

Program Committee

Alberto Abad

INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa, Portugal

André Adami

University of Caxias do Sul, Brazil

José João Almeida

University of Minho, Portugal

Sandra Aluísio

University of São Paulo, Brazil

Evelin Amorim

University of Porto, Portugal

João Balsa

University of Lisbon, Portugal

Cláudia Barros

Federal Institute of São Paulo, Brazil

Fernando Batista

INESC-ID and Iscte, Portugal

Daniel Beck

University of Melbourne, Australia

Luciana Benotti

National University of Córdoba, Argentina

Antonio Bonafonte

Amazon, Spain

António Branco

University of Lisbon, Portugal

Vera Cabarrão

Unbabel and INESC-ID, Portugal

Pável Calado

INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa, Portugal

Arnaldo Candido Junior

Federal University of Technology - Paraná, Brazil

Paula Cardoso

Federal University of Lavras, Brazil

Helena Caseli

Federal University of São Carlos, Brazil

Sheila Castilho

Dublin City University, Ireland

Berthold Crysmann

CNRS and University of Paris, France

Ariani Di Felippo

Federal University of São Carlos, Brazil

Gaél Dias

University of Caen Normandy, France

Brett Drury

INESC-TEC, Portugal

| | |
|---------------------------|---|
| Magali Sanches Duran | University of São Paulo, Brazil |
| Aline Evers | Pontifical Catholic University of Rio Grande do Sul, Brazil |
| Isabel Falé | Universidade Aberta and University of Lisbon, Portugal |
| Valeria Feltrim | State University of Maringá, Brazil |
| Eraldo Rezende Fernandes | Leuphana University Lüneburg, Germany |
| Thiago Ferreira | Federal University of Minas Gerais, Brazil |
| Maria José B. Finatto | Federal University of Rio Grande do Sul, Brazil |
| Marcelo Finger | University of São Paulo, Brazil |
| Erick Fonseca | Kaufland, Germany |
| Mikel L. Forcada | University of Alacant, Spain |
| Cláudia Freitas | Pontifical Catholic University of Rio de Janeiro, Brazil |
| Diamantino Freitas | University of Oporto, Portugal |
| Pablo Gamallo | University of Santiago de Compostela, Spain |
| Teresa Gonçalves | University of Évora, Portugal |
| Evandro Gouvea | Interactions, Germany |
| Santiago Pérez Isasi | University of Lisbon, Portugal |
| Eric Laporte | Université Gustave Eiffel, France |
| Ana Luísa V. Leal | University of Macau, Macau |
| Ana R. Luis | University of Coimbra, Portugal |
| Ranniery Maia | Federal University of Santa Catarina, Brazil |
| Nuno Mamede | INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal |
| Palmira Marrafa | University of Lisbon (retired), Portugal |
| Patricia Martin-Rodilla | University of A Coruña, Spain |
| Bruno Martins | INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal |
| Susana Duarte Martins | NOVA University of Lisbon, Portugal |
| Ana Isabel Mata | University of Lisbon, Portugal |
| David Martins de Matos | INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Portugal |
| Erick Galiani Maziero | University of Lavras, Brazil |
| Amália Mendes | University of Lisbon, Portugal |
| Sara Mendes | University of Lisbon, Portugal |
| Helena Moniz | INESC-ID and Universidade de Lisboa, Portugal |
| Cristiane Namiuti | State University of Southwest Bahia, Brazil |
| Nelson Neto | Federal University of Pará, Brazil |
| Maria das Graças V. Nunes | University of São Paulo, Brazil |
| Hugo Gonçalves Oliveira | CISUC and University of Coimbra, Portugal |
| Paula López Otero | Spitch AG, Spain |

| | |
|-------------------------|---|
| Aline Paes | Fluminense Federal University, Brazil |
| Gustavo Paetzold | Federal University of Technology – Paraná, Brazil |
| Adriana Pagano | Federal University of Minas Gerais, Brazil |
| Valeria de Paiva | Topos Institute, USA |
| Ivandr  Paraboni | University of S o Paulo, Brazil |
| Thiago Pardo | University of S o Paulo, Brazil |
| Fernando Perdig o | University of Coimbra, Portugal |
| Mart n Pereira-Fari a | University of Santiago de Compostela, Spain |
| Jos  Ramom Pichel | imaxin software and University of Santiago de Compostela, Spain |
| Prakash Poudyal | Kathmandu University, Nepal |
| Carlos A. Prolo | Federal University of Rio Grande do Norte, Brazil |
| Paulo Quaresma | University of  vora, Portugal |
| Violeta Quental | Pontifical Catholic University of Rio de Janeiro, Brazil |
| Alexandre Rademaker | IBM Research and FGV/EMAp, Brazil |
| Carlos Ramisch | Aix-Marseille University, France |
| Amanda Rassi | Somos Educa  o, Brazil |
| Livy Real | IBM, Brazil |
| Nat lia Resende | Federal University of Santa Catarina, Brazil |
| Ricardo Ribeiro | INESC-ID and Iscte, Portugal |
| Vitor Rocio | Universidade Aberta and INESC-TEC, Portugal |
| Irene Rodrigues | University of  vora, Portugal |
| Ricardo Rodrigues | CISUC and Polytechnic Institute of Coimbra, Portugal |
| Jos  Saias | University of  vora, Portugal |
| Diana Santos | Linguat ca and University of Oslo, Norway |
| Eloize Rossi M. Seno | Federal Institute of S o Paulo, Brazil |
| Christopher Shulby | Defined.ai, USA |
| Augusto Soares da Silva | Catholic University of Portugal, Portugal |
| Jo o Silva | University of Lisbon, Portugal |
| M rio J. Silva | INESC-ID and Instituto Superior T cnico, Universidade de Lisboa, Portugal |
| N dia Silva | University of S o Paulo, Brazil |
| Rub n Solera Ure a | INESC-ID, Portugal |
| Ant nio Teixeira | University of Aveiro, Portugal |
| Isabel Trancoso | INESC-ID and Instituto Superior T cnico, Universidade de Lisboa, Portugal |
| Marcos Treviso | Instituto Superior T cnico, University of Lisbon, Portugal |
| Oto Vale | Federal University of S o Carlos, Brazil |
| Renata Vieira | CIDEHUS and University of  vora, Portugal |

Aline Villavicencio
Derek F. Wong
Marcos Zampieri

University of Sheffield, UK
University of Macau, Macau
Rochester Institute of Technology, USA

Additional Reviewers

Danielle Caled
Iria de Dios Flores
John Ortega
Herbert Pinto
Mário Rodrigues

INESC-ID and Universidade de Lisboa, Portugal
University of Santiago de Compostela, Spain
University of Santiago de Compostela, Spain
Open University, Portugal
University of Aveiro, Portugal

Sponsor

americanas sa

Contents

Resources and Evaluation

| | |
|---|----|
| UlyssesNER-Br: A Corpus of Brazilian Legislative Documents for Named Entity Recognition | 3 |
| <i>Hidemberg O. Albuquerque, Rosimeire Costa, Gabriel Silvestre, Ellen Souza, Nádia F. F. da Silva, Douglas Vitório, Gyovana Moriyama, Lucas Martins, Luíza Soezima, Augusto Nunes, Felipe Siqueira, João P. Tarrega, Joao V. Beinotti, Marcio Dias, Matheus Silva, Miguel Gardini, Vinicius Silva, André C. P. L. F. de Carvalho, and Adriano L. I. Oliveira</i> | |
| A Test Suite for the Evaluation of Portuguese-English Machine Translation | 15 |
| <i>Mariana Avelino, Vivien Macketanz, Eleftherios Avramidis, and Sebastian Möller</i> | |
| MINT - Mainstream and Independent News Text Corpus | 26 |
| <i>Danielle Caled, Paula Carvalho, and Mário J. Silva</i> | |
| Fakepedia Corpus: A Flexible Fake News Corpus in Portuguese | 37 |
| <i>Anderson Cordeiro Charles, Livia Ruback, and Jonice Oliveira</i> | |
| A Targeted Assessment of the Syntactic Abilities of Transformer Models for Galician-Portuguese | 46 |
| <i>Marcos Garcia and Alfredo Crespo-Otero</i> | |
| FakeRecogna: A New Brazilian Corpus for Fake News Detection | 57 |
| <i>Gabriel L. Garcia, Luis C. S. Afonso, and João P. Papa</i> | |
| Implicit Opinion Aspect Clues in Portuguese Texts: Analysis and Categorization | 68 |
| <i>Mateus Tarcinalli Machado, Thiago Alexandre Salgueiro Pardo, Evandro Eduardo Seron Ruiz, Ariani Di Felippo, and Francielle Vargas</i> | |
| CRPC-DB a Discourse Bank for Portuguese | 79 |
| <i>Amália Mendes and Pierre Lejeune</i> | |
| Challenges in Annotating a Treebank of Clinical Narratives in Brazilian Portuguese | 90 |
| <i>Lucas Ferro Antunes de Oliveira, Adriana Pagano, Lucas Emanuel Silva e Oliveira, and Claudia Moro</i> | |

| | |
|--|-----|
| PetroBERT: A Domain Adaptation Language Model for Oil and Gas Applications in Portuguese | 101 |
| <i>Rafael B. M. Rodrigues, Pedro I. M. Privatto, Gustavo José de Sousa, Rafael P. Murari, Luis C. S. Afonso, João P. Papa, Daniel C. G. Pedronette, Ivan R. Guilherme, Stephan R. Perrout, and Aliel F. Riente</i> | |
| SS-PT: A Stance and Sentiment Data Set from Portuguese Quoted Tweets | 110 |
| <i>Miguel Won and Jorge Fernandes</i> | |
| Natural Language Processing Tasks | |
| ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling | 125 |
| <i>Alexandre Alcoforado, Thomas Palmeira Ferraz, Rodrigo Gerber, Enzo Bustos, André Seidel Oliveira, Bruno Miguel Veloso, Fabio Levy Siqueira, and Anna Helena Reali Costa</i> | |
| Banking Regulation Classification in Portuguese | 137 |
| <i>Rafael Faria de Azevedo, Tiago Nunes Silva, Henrique Tibério Brandão Vieira Augusto, Paulo Oliveira Sampaio Reis, Isadora Bastos Chaves, Samara Beatriz Naka de Vasconcellos, Liliany Aparecida dos Anjos Pereira, Mauro Melo de Souza Biccass, André Luiz Monteiro, and Alexandre Rodrigues Duarte</i> | |
| Automatic Information Extraction: A Distant Reading of the Brazilian Historical-Biographical Dictionary | 148 |
| <i>Suemi Higuchi, Claudia Freitas, and Diana Santos</i> | |
| Automatic Recognition of Units of Measurement in Product Descriptions from Tax Invoices Using Neural Networks | 156 |
| <i>Lucas Fernandes Lucena, Telmo de Menezes e Silva Filho, Thaís Gaudencio do Rêgo, and Yuri Malheiros</i> | |
| Entity Extraction from Portuguese Legal Documents Using Distant Supervision | 166 |
| <i>Lucas M. Navarezi, Kenzo Sakiyama, Lucas S. Rodrigues, Caio M. O. Robaldo, Gustavo R. Lobato, Paulo A. Vilela, Edson T. Matsubara, and Eraldo R. Fernandes</i> | |
| Sexist Hate Speech: Identifying Potential Online Verbal Violence Instances | 177 |
| <i>Brenda Salenave Santana, Aline Aver Vanin, and Leandro Krug Wives</i> | |

| | |
|---|------------|
| Book Genre Classification Based on Reviews of Portuguese-Language Literature | 188 |
| <i>Clarisse Scofield, Mariana O. Silva, Luiza de Melo-Gomes, and Mirella M. Moro</i> | |
| Combining Word Embeddings for Portuguese Named Entity Recognition | 198 |
| <i>Messias Gomes da Silva and Hilário Tomaz Alves de Oliveira</i> | |
| BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives | 209 |
| <i>Frederico Dias Souza and João Baptista de Oliveira e Souza Filho</i> | |
| Fostering Judiciary Applications with New Fine-Tuned Models for Legal Named Entity Recognition in Portuguese | 219 |
| <i>Luciano Zanuz and Sandro José Rigo</i> | |
| Natural Language Processing Applications | |
| Using Topic Modeling in Classification of Brazilian Lawsuits | 233 |
| <i>André Aguiar, Raquel Silveira, Vasco Furtado, Vlândia Pinheiro, and João A. Monteiro Neto</i> | |
| PortNOIE: A Neural Framework for Open Information Extraction for the Portuguese Language | 243 |
| <i>Bruno Cabral, Marlo Souza, and Daniela Barreiro Claro</i> | |
| Tracking Environmental Policy Changes in the Brazilian Federal Official Gazette | 256 |
| <i>Flávio Nakasato Cação, Anna Helena Reali Costa, Natalie Unterstell, Liuca Yonaha, Taciana Stec, and Fábio Ishisaki</i> | |
| A Transfer Learning Analysis of Political Leaning Classification in Cross-domain Content | 267 |
| <i>Danielle Caled and Mário J. Silva</i> | |
| Integrating Question Answering and Text-to-SQL in Portuguese | 278 |
| <i>Marcos Menon José, Marcelo Archanjo José, Denis Deratani Mauá, and Fábio Gagliardi Cozman</i> | |
| Named Entity Extractors for New Domains by Transfer Learning with Automatically Annotated Data | 288 |
| <i>Emanuel Matos, Mário Rodrigues, and António Teixeira</i> | |

PTT5-Paraphraser: Diversity and Meaning Fidelity in Automatic
Portuguese Paraphrasing 299
*Lucas Francisco Amaral Orosco Pellicer, Paulo Pirozelli,
Anna Helena Realí Costa, and Alexandre Inoue*

Speech Processing and Applications

A Protocol for Comparing Gesture and Prosodic Boundaries in Multimodal
Corpora 313
Camila Barros and Saulo Santos

Forced Phonetic Alignment in Brazilian Portuguese Using Time-Delay
Neural Networks 323
Cassio Batista and Nelson Neto

Brazilian Portuguese Speech Recognition Using Wav2vec 2.0 333
*Lucas Rafael Stefanel Gris, Edresson Casanova,
Frederico Santos de Oliveira, Anderson da Silva Soares,
and Arnaldo Candido Junior*

A Corpus of Neutral Voice Speech in Brazilian Portuguese 344
*Pedro H. L. Leite, Edmundo Hoyle, Álvaro Antelo, Luiz F. Kruszielski,
and Luiz W. P. Biscainho*

Comparing Lexical and Usage Frequencies of Palatal Segments
in Portuguese 353
Luís Trigo and Carlos Silva

Lexical Semantics

Extracting Valences from a Dependency Treebank for Populating the Verb
Lexicon of a Portuguese HPSG Grammar 365
*Leonel Figueiredo de Alencar, Lucas Ribeiro Coutinho,
Wellington José Leite da Silva, Ana Luíza Nunes,
and Alexandre Rademaker*

CQL Grammars for Lexical and Semantic Information Extraction
for Portuguese and Italian 376
Chiara Barbero

Drilling Lexico-Semantic Knowledge in Portuguese from BERT 387
Hugo Gonalo Oliveira

Short Papers

| | |
|--|-----|
| The Systematic Construction of Multiple Types of Corpora Through the Lapelinc Framework | 401 |
| <i>Bruno Silvério Costa, Jorge Viana Santos, Cristiane Namiuti, and Aline Silva Costa</i> | |
| Revisiting CCNet for Quality Measurements in Galician | 407 |
| <i>John E. Ortega, Iria de-Dios-Flores, José Ramom Pichel, and Pablo Gamallo</i> | |
| Identifying Literary Characters in Portuguese: Challenges of an International Shared Task | 413 |
| <i>Diana Santos, Roberto Willrich, Marcia Langfeldt, Ricardo Gaiotto de Moraes, Cristina Mota, Emanuel Pires, Rebeca Schumacher, and Paulo Silva Pereira</i> | |
| Should I Buy or Should I Pass: E-Commerce Datasets in Portuguese | 420 |
| <i>Henrique Varella Ehrenfried and Eduardo Todt</i> | |

Best MsC/MA and PhD Dissertation

| | |
|---|-----|
| Abstract Meaning Representation Parsing for the Brazilian Portuguese Language | 429 |
| <i>Rafael Torres Anchiêta and Thiago Alexandre Salgueiro Pardo</i> | |
| Enriching Portuguese Word Embeddings with Visual Information | 435 |
| <i>Bernardo Scapini Consoli and Renata Vieira</i> | |
| Author Index | 441 |