

# Lecture Notes in Artificial Intelligence

12037

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

Founding Editor

Jörg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*


More information about this series at <http://www.springer.com/series/1244>


Paulo Quaresma · Renata Vieira ·  
Sandra Aluísio · Helena Moniz ·  
Fernando Batista · Teresa Gonçalves (Eds.)


# Computational Processing of the Portuguese Language


14th International Conference, PROPOR 2020  
Evora, Portugal, March 2–4, 2020  
Proceedings


### *Editors*


Paulo Quaresma   
University of Évora  
Evora, Portugal

Sandra Aluísio   
University of São Paulo  
São Carlos, Brazil

Fernando Batista   
INESC-ID/ISCTE-IUL  
Lisbon, Portugal

Renata Vieira   
University of Évora  
Evora, Portugal

Helena Moniz   
University of Lisbon  
Lisbon, Portugal

Teresa Gonçalves   
University of Évora  
Evora, Portugal

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Artificial Intelligence

ISBN 978-3-030-41504-4

ISBN 978-3-030-41505-1 (eBook)

<https://doi.org/10.1007/978-3-030-41505-1>

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2020

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 14th edition of the International Conference on the Computational Processing of Portuguese (PROPOR 2020). The conference was held at the Colégio do Espírito Santo, Universidade de Évora, in Évora, Portugal, during March 2–4, 2020. Although the conference is in its 14th edition, it is celebrating the maturity of 27 years of existence, since it has been predominantly a biannual event. In that route it returns to Évora after 21 years.

PROPOR is the main scientific meeting in the area 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. It is also a meeting for the reception of new members to the community, represented mainly by the students in the field. A total of 70 submissions were received for the main event, involving 233 authors from many countries worldwide, such as Portugal, Brazil, Spain, Macau, the UK, Norway, Nepal, India, Germany, and China.

This volume presents 36 full and 5 short papers, selected from 70 submissions, which reflects an acceptance rate of 51% for full papers and 58% overall. Each paper was reviewed by 3 reviewers from a total of 48 academic institutions and companies. The papers are organized thematically and include the most recent developments in speech processing, resources and evaluation, semantics, natural language processing tasks, multilinguality, and applications. This year, reflecting the technological presence of the area in everyday life, the topic with the largest number of submissions and acceptance was natural language applications.

For 2020 our selected team of invited speakers represented the worldwide importance of this area, not only in the academy but also in very well-established companies. They are:

- João Graça, Co-founder and CTO of Unbabel
- Fernando Pereira, VP and Engineering Fellow at Google
- Cícero dos Santos, Research Scientist at Amazon
- Isabel Trancoso, Senior Researcher at Spoken Language Systems Lab, INESC-ID

In this edition we also had four workshops, regarding the following topics: relations of natural language with the digital humanities, technologies for language acquisition, language corpora availability, and tools and resources for paraphrasing. The proceedings of the workshops are not included in this volume, as the organizers of each workshop arranged them separately. The event also includes a contest for best dissertation and thesis, a tutorial, and a session for systems demonstration.

Our sincere thanks 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, the invited speakers, and the general organization staff. We are also grateful to the agencies and organizations that supported and promoted the event.

January 2020

Paulo Quaresma  
Renata Vieira  
Sandra Aluísio  
Helena Moniz  
Fernando Batista  
Teresa Gonçalves

# Organization

## General Chairs

Paulo Quaresma	Universidade de Évora, Portugal
Renata Vieira	Universidade de Évora, Portugal

## Program Chairs

Sandra Aluísio	University of São Paulo, Brazil
Helena Moniz	INESC-ID/FLUL, Universidade de Lisboa, Portugal

## Editorial Chairs

Fernando Batista	INESC-ID and ISCTE-IUL, Portugal
Teresa Gonçalves	Universidade de Évora, Portugal

## Demo Chairs

José Saias	Universidade de Évora, Portugal
Vlória Pinheiro	University of Fortaleza, Brazil

## Best MSc/MA and PhD Dissertation Award Chair

Pablo Gamallo	University of Santiago de Compostela, Spain
---------------	---

## Workshops and Tutorials Chairs

André Adami	Universidade de Caxias do Sul, Brazil
Irene Rodrigues	Universidade de Évora, Portugal

## Local Organization Committee

Madhu Agrawal	Universidade de Évora, Portugal
Roy Bayot	Universidade de Évora, Portugal
José Duarte	Universidade de Évora, Portugal
Nuno Miranda	Universidade de Évora, Portugal
Prakash Poudyal	Universidade de Évora, Portugal
João Sequeira	Universidade de Évora, Portugal
Hua Yang	Universidade de Évora, Portugal

## Steering Committee

André Adami	Universidade de Caxias do Sul, Brazil
Amália Mendes	Centro de Linguística da Universidade de Lisboa, Portugal
Paulo Quaresma	Universidade de Évora, Portugal
João Silva	Universidade de Lisboa, Portugal
Aline Villavicencio	Institute of Informatics, Federal University of Rio Grande do Sul, Brazil, and The University of Sheffield, UK

## Program Committee

Alberto Abad	IST/INESC-ID, Portugal
André Adami	Universidade de Caxias do Sul, Brazil
Sandra Aluisio	University of São Paulo, Brazil
João Balsa	Universidade de Lisboa, Portugal
Jorge Baptista	University of Algarve, and L2F-Spoken Language Lab, INESC-ID Lisboa, Portugal
Plínio Barbosa	University of Campinas, Brazil
Anabela Barreiro	INESC-ID, Portugal
Fernando Batista	INESC-ID and ISCTE-IUL, Portugal
Luciana Benotti	Universidad Nacional de Cordoba, Argentina
Maria Bocorny Finatto	UFRGS, Brazil
António Branco	Universidade de Lisboa, Portugal
Nuno C. Marques	DI – FCT/UNL, Portugal
Vera Cabarrão	FLUL/CLUL, Universidade de Lisboa, and INESC-ID, Portugal
Sara Candeias	Microsoft, Portugal
Helena Caseli	Federal University of São Carlos, Brazil
Luísa Coheur	IST/INESC-ID Lisboa, Portugal
Berthold Cysmann	CNRS, France
Valeria de Paiva	Samsung Research America, USA, and University of Birmingham, UK
José de Souza	eBay Inc., USA
Ariani Di Felippo	Universidade Federal de São Carlos, Brazil
Gaël Dias	Normandie University, France
Brett Drury	Scicrop, Brazil
Magali Duran	University of São Paulo, Brazil
Isabel Falé	Universidade Aberta/CLUL, Portugal
Valeria Feltrim	Universidade Estadual de Maringá, Brazil
Marcelo Finger	University of São Paulo, Brazil
Erick Fonseca	Instituto de Telecomunicações, Portugal
Claudia Freitas	PUC-Rio, Brazil
Erick Galani Maziero	Universidade de Lavras, Brazil
Pablo Gamallo	University of Santiago de Compostela, Spain



Teresa Gonçalves	Universidade de Évora, Portugal
Leandro Henrique	Embrapa, Brazil
Mário J. Silva	IST/INESC-ID, Universidade de Lisboa, Portugal
Arnaldo Cândido Junior	UTFPR, Brazil
Fabio Kepler	Unbabel, Portugal
Eric Laporte	Université Paris-Est Marne-la-Vallée, France
Ana Leal	University of Macau, Macau
Nuno Mamede	IST/INESC-ID Lisboa, Portugal
Palmira Marrafa	Universidade de Lisboa, Portugal
David Martins de Matos	Universidade de Lisboa, Portugal
Ana Isabel Mata	Universidade de Lisboa, Portugal
Amália Mendes	Centro de Linguística da Universidade de Lisboa, Portugal
Helena Moniz	L2F/INESC-ID, Portugal
Nelson Neto	Universidade Federal do Pará, Brazil
Hugo Gonçalves Oliveira	CISUC, University of Coimbra, Portugal
Ivandre Paraboni	University of São Paulo, Brazil
Thiago Pardo	University of São Paulo, Brazil
Carla Parra Escartín	Unbabel, Portugal
Fernando Perdígao	IT, Portugal
Prakash Poudyal	Kathmandu University, Nepal
Carlos Prolo	UFRN, Brazil
Paulo Quaresma	Universidade de Évora, Portugal
Violeta Quental	PUC-Rio, Brazil
Alexandre Rademaker	IBM Research Brazil and EMap/FGV, Brazil
Carlos Ramisch	Aix Marseille University, France
Eraldo Rezende Fernandes	FACOM/UFMS, Brazil
Ricardo Ribeiro	INESC ID Lisboa/ISCTE-IUL, Portugal
Vitor Rocio	Universidade Aberta, Portugal
Irene Rodrigues	Universidade de Évora, Portugal
Norton Roman	USP, Brazil
José Saias	Universidade de Évora, Portugal
Diana Santos	University of Oslo, Norway
Carolina Scarton	The University of Sheffield, UK
Evandro Seron Ruiz	University of São Paulo, Brazil
Antonio Serralheiro	INESC-ID and Academia Militar, Portugal
Christopher Shulby	DefinedCrowd, Portugal
João Silva	Universidade de Lisboa, Portugal
Alberto Simões	2AI Lab-IPCA, Portugal
Augusto Soares da Silva	Universidade Católica Portuguesa, Portugal
Rubén Solera-Ureña	Spoken Language Systems Laboratory, INESC-ID Lisboa, Portugal
Marcos Treviso	Instituto de Telecomunicações/Instituto Superior Técnico, Portugal
Luis Felipe Uebel	Samsung Instituto de Desenvolvimento para a Informática da Amazônia (SIDIA), Brazil

Oto Vale	Universidade Federal de São Carlos, Brazil
Renata Vieira	PUCRS, Brazil
Aline Villavicencio	Institute of Informatics, Federal University of Rio Grande do Sul, Brazil, and The University of Sheffield, UK
Maria Volpe Nunes	USP, Brazil
Rodrigo Wilkens	University of Strasbourg, France
Marcos Zampieri	Rochester Institute of Technology, USA
Leonardo Zilio	CTS, University of Surrey, UK

## **Additional Reviewers**

Danielle Caled  
Nathan Hartmann  
João Rodrigues

# Contents

## Speech Processing

Towards Automatic Determination of Critical Gestures for European Portuguese Sounds . . . . .	3
<i>Samuel Silva, Conceição Cunha, António Teixeira, Arun Joseph, and Jens Frahm</i>	
Comparison of Heterogeneous Feature Sets for Intonation Verification. . . . .	13
<i>Mariana Julião, Alberto Abad, and Helena Moniz</i>	
The BioVisualSpeech European Portuguese Sibilants Corpus . . . . .	23
<i>Margarida Grilo, Isabel Guimarães, Mariana Ascensão, Alberto Abad, Ivo Anjos, João Magalhães, and Sofia Cavaco</i>	
Evaluation of Deep Learning Approaches to Text-to-Speech Systems for European Portuguese . . . . .	34
<i>Sebastião Quintas and Isabel Trancoso</i>	
Evaluation and Extensions of an Automatic Speech Therapy Platform . . . . .	43
<i>Anna Pompili, Alberto Abad, Isabel Trancoso, José Fonseca, and Isabel P. Martins</i>	

## Resources and Evaluation

A Dataset for the Evaluation of Lexical Simplification in Portuguese for Children . . . . .	55
<i>Nathan S. Hartmann, Gustavo H. Paetzold, and Sandra M. Aluísio</i>	
Situational Irony in Farcical News Headlines . . . . .	65
<i>Paula Carvalho, Bruno Martins, Hugo Rosa, Silvio Amir, Jorge Baptista, and Mário J. Silva</i>	
Inferring the Source of Official Texts: Can SVM Beat ULMFiT? . . . . .	76
<i>Pedro Henrique Luz de Araujo, Teófilo Emidio de Campos, and Marcelo Magalhães Silva de Sousa</i>	
Native Language Identification on L2 Portuguese . . . . .	87
<i>Iria del Río</i>	
Aligning IATE Criminal Terminology to SUMO. . . . .	98
<i>Daniela Schmidt, Avner Dal Bosco, Cássia Trojahn, Renata Vieira, and Paulo Quaresma</i>	

The Construction of a Corpus from the Brazilian Historical-Biographical Dictionary. . . . .	109
<i>Lucas Ribeiro, Jaqueline P. Zulini, and Alexandre Rademaker</i>	
<b>Natural Language Processing Applications</b>	
Making the Most of Synthetic Parallel Texts: Portuguese-Chinese Neural Machine Translation Enhanced with Back-Translation . . . . .	121
<i>Rodrigo Santos, João Silva, and António Branco</i>	
Leveraging on Semantic Textual Similarity for Developing a Portuguese Dialogue System. . . . .	131
<i>José Santos, Ana Alves, and Hugo Gonçalo Oliveira</i>	
Fake News Detection on Fake.Br Using Hierarchical Attention Networks. . . .	143
<i>Emerson Yoshiaki Okano, Zebin Liu, Donghong Ji, and Evandro Eduardo Seron Ruiz</i>	
Screening of Email Box in Portuguese with SVM at Banco do Brasil . . . . .	153
<i>Rafael Faria de Azevedo, Rafael Rodrigues Pereira de Araujo, Rodrigo Guimarães Araújo, Régis Moreira Bittencourt, Rafael Ferreira Alves da Silva, Gabriel de Melo Vaz Nogueira, Thiago Marques Franca, Jair Otharan Nunes, Klailton Ralff da Silva, and Emmanuelle Regiane Cunha de Oliveira</i>	
Back to the Feature, in Entailment Detection and Similarity Measurement for Portuguese. . . . .	164
<i>Pedro Fialho, Luísa Coheur, and Paulo Quaresma</i>	
Emoji Prediction for Portuguese . . . . .	174
<i>Luis Duarte, Luís Macedo, and Hugo Gonçalo Oliveira</i>	
Vitality Analysis of the Linguistic Atlas of Brazil on Twitter . . . . .	184
<i>Arley Prates M. Nunes, Luis Emanuel N. de Jesus, Daniela Barreiro Claro, Jacyra Mota, Silvana Ribeiro, Marcela Paim, and Josane Oliveira</i>	
Fact-Checking for Portuguese: Knowledge Graph and Google Search-Based Methods. . . . .	195
<i>Roney Lira de Sales Santos and Thiago Alexandre Salgueiro Pardo</i>	
Extraction and Use of Structured and Unstructured Features for the Recommendation of Urban Resources . . . . .	206
<i>Brenda Salenave Santana and Leandro Krug Wives</i>	

## Semantics

A Portuguese Dataset for Evaluation of Semantic Question Answering . . . . .	217
<i>Denis Andrei de Araujo, Sandro José Rigo, Paulo Quaresma, and João Henrique Muniz</i>	
Exploring the Potentiality of Semantic Features for Paraphrase Detection . . . .	228
<i>Rafael Torres Anchiêta and Thiago Alexandre Salgueiro Pardo</i>	
Portuguese Language Models and Word Embeddings: Evaluating on Semantic Similarity Tasks . . . . .	239
<i>Ruan Chaves Rodrigues, Jéssica Rodrigues, Pedro Vitor Quinta de Castro, Nádia Felix Felipe da Silva, and Anderson Soares</i>	
Relation Extraction for Competitive Intelligence . . . . .	249
<i>Sandra Collovini, Patricia Nunes Gonçalves, Guilherme Cavalheiro, Joaquim Santos, and Renata Vieira</i>	

## Natural Language Processing Tasks

Evaluating Methods of Different Paradigms for Subjectivity Classification in Portuguese . . . . .	261
<i>Luana Balador Belisário, Luiz Gabriel Ferreira, and Thiago Alexandre Salgueiro Pardo</i>	
Sentence Compression for Portuguese . . . . .	270
<i>Fernando A. A. Nóbrega, Alipio M. Jorge, Pavel Brazdil, and Thiago A. S. Pardo</i>	
An Investigation of Pre-trained Embeddings in Dependency Parsing . . . . .	281
<i>Juliana C. Carvalho de Araújo, Cláudia Freitas, Marco Aurélio C. Pacheco, and Leonardo A. Forero-Mendoza</i>	
Segmentation of Words Written in the Latin Alphabet: A Systematic Review . . . . .	291
<i>Marcelo A. Inuzuka, Aquila S. Rocha, and Hugo A. D. Nascimento</i>	
A Deep Learning Model of Common Sense Knowledge for Augmenting Natural Language Processing Tasks in Portuguese Language . . . . .	303
<i>Cecília Silvestre Carvalho, Vlândia C. Pinheiro, and Lívio Freire</i>	
Linguistic Analysis Model for Monitoring User Reaction on Satirical News for Brazilian Portuguese . . . . .	313
<i>Gabriela Wick-Pedro, Roney L. S. Santos, Oto A. Vale, Thiago A. S. Pardo, Kalina Bontcheva, and Carolina Scarton</i>	

## Multilinguality

Word Embeddings at Post-Editing. . . . .	323
<i>Marcio Lima Inácio and Helena de Medeiros Caseli</i>	
Argument Identification in a Language Without Labeled Data . . . . .	335
<i>João Rodrigues and António Branco</i>	
Natural Language Inference for Portuguese Using BERT and Multilingual Information . . . . .	346
<i>Marco Antonio Sobrevilla Cabezedo, Marcio Inácio, Ana Carolina Rodrigues, Edresson Casanova, and Rogério Figueredo de Sousa</i>	
Exploiting Siamese Neural Networks on Short Text Similarity Tasks for Multiple Domains and Languages . . . . .	357
<i>João Vitor Andrioli de Souza, Lucas Emanuel Silva E Oliveira, Yohan Boneski Gumiel, Deborah Ribeiro Carvalho, and Claudia Maria Cabral Moro</i>	
CrossOIE: Cross-Lingual Classifier for Open Information Extraction . . . . .	368
<i>Bruno Souza Cabral, Rafael Glauber, Marlo Souza, and Daniela Barreiro Claro</i>	
One Book, Two Language Varieties . . . . .	379
<i>Anabela Barreiro, Ida Rebelo-Arnold, Fernando Batista, Isabel Garcez, and Tanara Zingano Kuhn</i>	

## Short Papers

Speech Breathing and Expressivity: An Experimental Study in Reading and Singing Styles . . . . .	393
<i>Plínio A. Barbosa, Sandra Madureira, Mario A. S. Fontes, and Paulo Menegon</i>	
Exploring Portuguese Word Embeddings for Discovering Lexical-Semantic Relations. . . . .	399
<i>Tiago Sousa, Ana Alves, and Hugo Gonçalves Oliveira</i>	
The ASSIN 2 Shared Task: A Quick Overview . . . . .	406
<i>Livy Real, Erick Fonseca, and Hugo Gonçalves Oliveira</i>	
A Multiplayer Voice-Enabled Game Platform for the Elderly . . . . .	413
<i>Francisco Oliveira, Alberto Abad, and Isabel Trancoso</i>	
Towards a Conversational Agent with “Character” . . . . .	420
<i>Gonçalo Melo and Luisa Coheur</i>	
Author Index . . . . .	425