

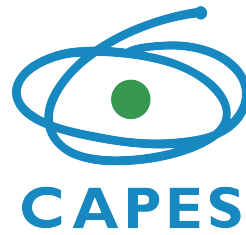
PROPOR 2026

**The 17th International Conference on Computational
Processing of Portuguese (PROPOR 2026)**

**Proceedings - Vol. 2 (Demo and Industry Tracks, Best
Dissertations, Workshops, and Tutorials)**

April 13-16, 2026

The PROPOR organizers gratefully acknowledge the support from the following sponsors.



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-387-6

Introduction

These Proceedings cover the 17th edition of the International Conference on the Computational Processing of Portuguese (PROPOR 2026), held in Salvador, Bahia, Brazil, April 13–16, 2026.

PROPOR is a major forum bringing together researchers, developers, and practitioners dedicated to the computational processing of Portuguese and, since the last edition in 2024, Galician. PROPOR promotes the exchange of experiences in the development of methodologies, language resources, tools, applications, and innovative projects covering the target languages. Established as the main scientific event in the area of natural language processing focused on theoretical and technological issues of written and spoken Portuguese and Galician, PROPOR 2026 also highlights the cultural and linguistic diversity present in Salvador, a place where Portuguese has historically interacted with indigenous and African languages, profoundly influencing Brazilian Portuguese and the surrounding culture.

This year's edition of PROPOR includes three tracks: General, Industry, and Demo, as well as three workshops, one tutorial, best MSc and PhD dissertation awards, and three invited speakers. In this edition, papers could be submitted in English or Portuguese.

The scope of PROPOR 2026 is reflected in the organization of the Proceedings, which are divided into two volumes. The first volume presents the accepted papers of the General track. The present (second) volume is dedicated to the Demo and Industry tracks, the Best Dissertation awards, and the workshops co-located with PROPOR 2026, namely:

- The First Workshop on Language Technologies for Health (Lang4Health);
- The Third Student Research Workshop (SRW);
- The Fourth Workshop on Digital Humanities and Natural Language Processing (DHandNLP).

This edition of PROPOR 2026 also features the following tutorial:

- *From Syntax to Semantics: Introducing UMR for NLP Annotation*
Adriana S. Pagano (UFMG), Magali Sanches Duran (USP), Federica Gamba (CUni)

PROPOR 2026 is structured as a four-day event, commencing with a full day of workshops and tutorials, followed by three days of communications, posters, demos, and community meetings. Our sincere thanks go to every person and institution involved in the complex organization of this event, especially the members of the Program Committee, the Workshop and Tutorial chairs, the chairs of the Demo and Industry tracks, the Best Dissertation chairs and jury, the invited speakers, and the general organization staff. We are also grateful to the agencies and organizations that supported and promoted the conference. Thank you all for participating and contributing to the success of this conference!

Marlo Souza and Iria de-Dios-Flores, General Chairs

Diana Santos and Larissa Freitas, Program Chairs

Jackson Wilke da Cruz Sousa and Eugénio Ribeiro, Editorial Chairs

Organizing Committee

General Chairs

Marlo Souza, Universidade Federal da Bahia, Brazil
Iria de-Dios-Flores, Universitat Pompeu Fabra - Barcelona, Spain

Program Chairs

Diana Santos, Universitetet i Oslo, Norway
Larissa Freitas, Universidade Federal de Pelotas, Brazil

Editorial Chairs

Jackson Wilke da Cruz Souza, Universidade Federal da Bahia, Brazil
Eugénio Ribeiro, Iscte-IUL & INESC-ID Lisboa, Portugal

Demo Chairs

Evandro Fonseca, Blip, Brazil
Susana Sotelo, Universidade de Santiago de Compostela, Spain

Industry Track Chairs

Clarissa Xavier, Banrisul, Brazil
Henrico Brum, Sinch AB, Sweden

Best Dissertation Chairs

Marcos Garcia, Universidade de Santiago de Compostela, Spain
Aline Paes, Universidade Federal Fluminense, Brazil

Workshops and Tutorial Chairs

Roney Lira de Sales Santos, Universidade Federal da Bahia, Brazil
Renata Vieira, Universidade de Évora, Portugal

Workshop on Language Technologies for Health (Lang4Health)

Aline Villavicencio, University of Exeter, UK
Rodrigo Wilkens, University of Exeter, UK
Helena Caseli, Federal University of São Carlos, Brazil
Vânia Neris, Federal University of São Carlos, Brazil

Student Research Workshop (SRW)

Livy Real, Universidade Federal do Amazonas, Brazil
Jackson Wilke da Cruz Souza, Universidade Federal da Bahia, Brazil

Workshop on Digital Humanities and Natural Language Processing (DHandNLP)

Leonardo Zilio, Université Catholique de Louvain, CENTAL, LSTI, Belgium

Helena Freire Cameron, Instituto Politécnico de Portalegre, CIDEHUS, Portugal

Maria José B. Finatto, Universidade Federal do Rio Grande do Sul, CNPq/PPG-LETRAS, Brazil

Renata Vieira, Évora University, CIDEHUS, Portugal

Local Organization

Marlo Souza, Universidade Federal da Bahia, Brazil

Daniela Barreiro Claro, Universidade Federal da Bahia, Brazil

Jackson Wilke da Cruz Souza, Universidade Federal da Bahia, Brazil

Lilian Teixeira, Universidade Federal da Bahia, Brazil

Rerisson Cavalcante, Universidade Federal da Bahia, Brazil

Robespierre Pita, Universidade Federal da Bahia, Brazil

Roney Lira de Sales Santos, Universidade Federal da Bahia, Brazil

Program Committee

Demo Track

Andre Carvalho, Universidade Federal do Amazonas, Brazil
Bruno Souza Cabral, Escavador, Brazil
Daniela Schmidt, Universidade de Évora, Portugal
Jesus M. Benitez Baleato, Universidade de Santiago de Compostela, Spain
José Ramon Pichel, imaxin.software & Universidade de Santiago de Compostela, Spain
Livy Real, Universidade Federal do Amazonas, Brazil
Luis Trigo, Universidade do Porto, Portugal
Luiz Merschmann, Universidade Federal de Lavras, Brazil
Saullo Oliveira, Pontifícia Universidade Católica de Campinas, Brazil
Thiago Pardo, Universidade de São Paulo, Brazil

Industry Track

Beatriz Fagundes, Clio, Canada
Fabio Rezende de Souza, University of São Paulo, Brazil
Marcio Bigolin, Instituto Federal do Rio Grande do Sul, Brazil
Nataly Leopoldina Patti da Silva, SiDi, Brazil
Sidney Evaldo Leal, Venturus, Brazil
Vítor Rodrigues Tonon, Universidade Estadual do Norte do Paraná, Brazil

Best Dissertation

Alberto Abad, Instituto Superior Técnico, Universidade de Lisboa & INESC-ID Lisboa, Portugal
Aline Vanin, Federal University of Health Sciences of Porto Alegre, Brazil
Amália Mendes, Universidade de Lisboa, Portugal
Arnaldo Candido Junior, Universidade Estadual Paulista, Brazil
Daniela Barreiro Claro, Universidade Federal da Bahia, Brazil
David Vilares, Universidade da Coruña, Spain
Helena Caseli, Federal University of São Carlos, Brazil
Ivandre Paraboni, University of São Paulo, Brazil
Jorge Baptista, Universidade do Algarve & INESC-ID Lisboa, Portugal
Luís M. S. Gomes, Faculdade de Ciências da Universidade de Lisboa, Portugal
Maria José Finatto, Universidade Federal do Rio Grande do Sul, Brazil
Magali S. Duran, University of São Paulo, Brazil
Marcelo Finger, University of São Paulo, Brazil
Marcos Fernandez-Pichel, Universidade de Santiago de Compostela, Spain
Maria das Graças Volpe Nunes, University of São Paulo, Brazil
Pablo Gamallo, Universidade de Santiago de Compostela, Spain
Paulo Quaresma, Universidade de Évora, Portugal
Plinio A. Barbosa, Universidade Estadual de Campinas, Brazil
Renata Vieira, Évora University, CIDEHUS, Portugal
Ricardo Ribeiro, Iscte-IUL & INESC-ID Lisboa, Portugal
Vladia C. M. Pinheiro, Universidade de Fortaleza, Brazil

Workshop on Language Technologies for Health (Lang4Health)

Aline Paes, Institute of Computing / Universidade Federal Fluminense, Brazil
Ana Cleide Guimbal de Aquino, Universidade Federal Rural da Amazônia, Brazil
César Sperb, Federal University of Pelotas, Brazil
Claudia Moro, Pontifícia Universidade Católica do Paraná, Brazil
Elisa Terumi Rubel Schneider, Pontifícia Universidade Católica do Paraná, Brazil
Eloize Seno, Federal Institute of São Paulo, Brazil
Emerson Paraiso, Pontifícia Universidade Católica do Paraná, Brazil
Helena Caseli, Federal University of São Carlos, Brazil
Ivandre Paraboni, University of São Paulo, Brazil
João Papa, São Paulo State University, Brazil
Luciana Salgados, Universidade Federal Fluminense, Brazil
Marcelo Finger, University of São Paulo, Brazil
Maria José Finatto, Universidade Federal do Rio Grande do Sul, Brazil
Marília Silveira, Federal University of Pelotas, Brazil
Mateus Monteiro, Brazil
Murilo Vargas da Cunha, Federal University of Pelotas, Brazil
Paula Souza, Universidade Federal de São Carlos, Brazil
Renata Vieira, Évora University, CIDEHUS, Portugal
Renato Silva, University of São Paulo, Brazil
Rodrigo Wilkens, University of Exeter, UK
Sandra Rodrigues, Universidade Federal de Lavras, Brazil
Tiago Torrent, Universidade Federal de Juiz de Fora, Brazil

Student Research Workshop (SRW)

Ana Mata, Universidade de Lisboa, Portugal
António Branco, Universidade de Lisboa, Portugal
Arnaldo Candido Junior, Universidade Estadual Paulista, Brazil
Eulanda M. dos Santos, Universidade Federal do Amazonas, Brazil
Evelin Amorim, INESC-TEC, Portugal
Fábio Lobato, Universidade do Oeste do Pará, Brazil
Fernanda Lopez-Escobedo, Universidad Nacional Autónoma de México, Mexico
Jackson Wilke da Cruz Souza, Universidade Federal da Bahia, Brazil
Jorge Baptista, University of Algarve & INESC-ID Lisboa, Portugal
Leo Sampaio Ferraz Ribeiro, Universidade de São Paulo, Brazil
Leonel Figueiredo de Alencar, Universidade Federal do Ceará, Brazil
Nádia Silva, Universidade Federal de Goiás, Brazil
Raquel Freitag, Universidade Federal de Sergipe, Brazil
Saullo Oliveira, Pontifícia Universidade Católica de Campinas, Brazil
Thiago Pardo, Universidade de São Paulo, Brazil

Workshop on Digital Humanities and Natural Language Processing (DHandNLP)

Álvaro Iriarte, University of Minho, Portugal
Ana Ribeiro, University of Évora, Portugal
Bruno Maroneze, Federal University of Grande Dourados, Brazil
Cassia Trojahn, University of Toulouse, France
Fernanda Olival, University of Évora, Portugal
Idalete da Silva Dias, University of Minho, Portugal
Marcus Dores, Federal University of Bahia, Brazil

Paulo Quaresma, University of Évora, Portugal
Raquel Amaro, NOVA University, Portugal
Sandro Marengo, Federal University of Sergipe, Brazil
Silvana Silva, Federal University of Rio Grande do Sul, Brazil

Table of Contents

<i>ARAMIS: Uma ferramenta web integrada com LLM open-source para apoio à correção de TCCs de estudantes de graduação</i>	
Gustavo Campelo de Sousa, Pablo Kauan Martins Timbó, Luiz Zairo Bastos Viana, Antônio Emerson Barros Tomaz, José Wellington Franco da Silva and Carlos de Oliveira Caminha	1
<i>From Complexity Scores to Readable Texts: iRead4Skills for Adult Literacy in Portuguese</i>	
Jorge Baptista, Eugénio Ribeiro, Nuno Mamede, David Antunes and Raquel Amaro	5
<i>Lexicon-Grammar Web</i>	
Jorge Baptista, David Antunes, Nuno Mamede and Eugénio Ribeiro	8
<i>Bruna: A Real-Time Multimodal Voice Agent with Hybrid Reasoning</i>	
Evandro Fonseca	11
<i>FlowDisco: Interactive Exploration of Dialogue Flows</i>	
Patrícia Ferreira, Carolina Loureiro, Ana Alves, Catarina Silva and Hugo Gonçalo Oliveira . . .	14
<i>AttentionApp: An Interactive Tool for Analyzing Transformer Attention Patterns in Portuguese</i>	
Ricardo G. Oliveira and Daniela Barreiro Claro	18
<i>Sistema Multimodal de Apoio ao Gerenciamento de Riscos de Desastres</i>	
Hosana Iasmin Castro dos Santos Lucena, Gabriel Rocha dos Santos, Jady Lima da Silva, Ricardo José Matos de Carvalho and Patrick Terrematte	21
<i>Lispector: Fine-tuning de Modelos de Linguagem para Revisão Gramatical e Ortográfica em Português Brasileiro</i>	
Andresa Medeiros, Felipe Iszlaji, Claudia Sarmiento-Moreno, Camila Muniz, Larissa Ponciano, Larissa Dejigov, Ronald Monteiro, Pedro Kretikouski and Guilherme Chaves	25
<i>Grounded in Law: A Multi-Stage Anti-Hallucination Pipeline for Legal RAG Systems in Brazilian Portuguese</i>	
Arla Figueiredo, João Lucas, Tatiana Ribeiro, Caio Nery, Alan Rios, Caio Hebert, Luiza Florentino, Arthur Silva, Ícaro Feyerabend, Pedro Vidal and Bruno Cabral	30
<i>Socially Responsible and Explainable Automated Fact-Checking and Hate Speech Detection</i>	
Francielle Vargas, Fabrício Benevenuto and Thiago A. S. Pardo	35
<i>Automated Essay Scoring for Brazilian Portuguese. Evidence from Cross-Prompt Evaluation of ENEM Essays</i>	
André Barbosa and Denis Deratani Mauá	43
<i>Evaluating FrameNet-Based Semantic Modeling for Gender-Based Violence Detection in Clinical Records</i>	
Lívia Dutra, Arthur Lorenzi, Frederico Belcavello, Ely Matos, Marcelo Viridiano, Lorena Larré, Olívia Guaranha, Erick Santos, Sofia Reinach, Pedro de Paula and Tiago Torrent	49
<i>Pretrained Neural Audio Models for Asthma Detection from Voice and Speech</i>	
Leticia Puttlitz Boll, Antonio Oss Boll, Yan Anderson Pires de Oliveira, Victor dos Santos Silva, Mariana Lopes Pestana, Celso Ricardo Fernandes de Carvalho, Marcelo Matheus Gauy and Marcelo Finger	58
<i>A RAG Chatbot with Incremental Context Retrieval based on Local LLMs for Hospital Documents</i>	
Murilo Vargas da Cunha, Marília Rosa Silveira, César Brasil Sperb, Larissa Astrogildo Freitas and Ulisses Brisolará Corrêa	68

<i>A Dataset of Brazilian Portuguese Clinical Notes for Anaphylaxis Detection</i>	
Matheus Machado, Vinícius Vanzin, Dilvan Moreira, Luis Felipe Ensina and Fábio Lario	78
<i>LLM-Based Multi-Agent System with Retrieval-Augmented Generation for Medical Care Planning Generation in Sickle Cell Disease</i>	
Luana Bringel Leite, David Eduardo Pereira, Eyshila Buriti de Araujo Azevedo, Leonardo Mota Meira Filho, Eliane Cristina Araújo, Cláudio E. C. Campelo, Taciana R. O. C. Marques, Leticia B. de Almeida and Herman Martins Gomes	88
<i>Class of LLMs: Benchmarking Large Language Models on the Brazilian National Medical Examination</i>	
João Vitor Mariano Correia, Pedro Henrique Alves de Castro, Gabriel Lino Garcia, Pedro Henrique Paiola and João Paulo Papa	101
<i>Retrieval-Augmented Generation for Clinical Question Answering in Portuguese Drug Leaflets: Benefits and Limitations</i>	
Gabriel Lino Garcia, Pedro Henrique Paiola, João Vitor Mariano Correia, Douglas Rodrigues and João Paulo Papa	112
<i>Annotation Guidelines and Challenges for Automatic Simplification of Portuguese Drug Leaflets</i>	
Arthur Scalercio, Eduarda Bertotto, Silvana Jesus, Maria José Finatto and Aline Paes	121
<i>From Annotated Clinical Narratives to Ontology: Structuring Brazilian Portuguese Clinical Data</i>	
Fernando Henrique Moura de Oliveira and Cleyton Mário de Oliveira Rodrigues	128
<i>The visible and the latent linguistic clues of mental health in Brazilian Portuguese textual posts</i>	
Rodrigo Wilkens, Helena Caseli, Vania Neris and Aline Villavicencio	135
<i>Caracterização lexical e sintática de notícias falsas em português produzidas por humanos e por máquinas</i>	
Pedro Lucas Castro de Andrade, Renato Silva and Thiago Alexandre Salgueiro Pardo	148
<i>Exploração de métodos simbólicos para detecção de emoções para o português</i>	
Stephanie Briere Americo and Thiago Alexandre Salgueiro Pardo	159
<i>Robustness and Diversity Evaluation on ProsSegue-ML: a Free Prosodic Segmentation Tool for Brazilian Portuguese</i>	
Giovana Meloni Craveiro and Sandra Maria Aluísio	170
<i>Combining Semantic Embeddings and Knowledge Graphs for Identifying Decision Patterns in Brazilian Judicial Decisions</i>	
Gustavo Soares Silva, Omar Andres Carmona Cortes, Fábio Manoel França Lobato and Antonio Fernando Lavareda Jacob Junior	181
<i>Development and Evaluation of a Hybrid Information Retrieval System Applied to the Brazilian Legal Domain</i>	
Ana Carolina C. Bessa, Fábio M. F. Lobato and Antonio F. L. J. Junior	186
<i>Viés e Justiça em Modelos de Linguagem: Evidências de Uma Literatura Linguística, Social e Culturalmente Assimétrica</i>	
Vitória P. Firmino, Bruno M. Nogueira and Valéria Q. dos Reis	191
<i>Textual Inference in Portuguese: Comparing Language Models</i>	
Fabiana Avais, Valeria de Paiva and Livy Real	201
<i>Parsing Nheengatu: Performance Gains for a Brazilian Indigenous Universal Dependencies Treebank</i>	
Dominick Maia Alexandre and Leonel Figueiredo de Alencar	210

<i>Bridging Cultural Gaps in Automated Translation of Brazilian Expressions: A Study on Cultural Adaptation</i>	
Maria Luiza Silva de Oliveira, Andressa Andrade Oliveira dos Santos and Leandro Jose Silva Andrade	220
<i>Towards a Universal Dependencies Corpus for Portuguese Epidemiological Reports</i>	
Christian Freitas, Livy Real, Lilian Berton and Valeria de Paiva	228
<i>Gendered Stylistic Variation in Brazilian Portuguese Google Play Reviews: A Large-Scale Study</i>	
Tiago de Melo	238
<i>A Larger Annotated Corpus of Portuguese Coreference</i>	
Evandro Fonseca and Renata Vieira	247
<i>Social-RAG: A Retrieval-Augmented Generation Pipeline for Computational Social Science Research on Telegram</i>	
Leonardo Nascimento, Eric Brasil, Arthur Lima, Gabriel Andrade, Ricardo José Andrade and Tarssio Barreto	255
<i>Comida e bebida nas literaturas portuguesa e brasileira: o projeto ReadingFood</i>	
Diana Santos, Eckhard Bick and Cristina Mota	266
<i>Fauna e Flora setecentista: das Entidades Nomeadas aos problemas de normalização</i>	
Helena Freire Cameron, Fernanda Olival, Daniel Reyes and Renata Vieira	275
<i>Exploring automatic terminology extraction from historical medical data</i>	
Leonardo Zilio and Maria José Bocorny Finatto	282
<i>Marcação semântica de entidades nomeadas em Os Lusíadas</i>	
Adriane Maria de Oliveira Queiroz and Bruno Oliveira Maroneze	293
<i>A elaboração de uma edição digital d’Os Lusíadas</i>	
Bruno Maroneze, Vanessa Martins do Monte, André Bertacchi, Artur Costrino, Alexandre Agnolon and Mário Eduardo Viaro	298
<i>The F1 of Formula One: Applicability of Pre-trained NER Models to Brazilian TV Interview Transcripts</i>	
João Pedro Gonçalves Munhoz, Luiz Felipe Guidorizzi de Oliveira, Isabella Belchior, Evandro Eduardo Seron Ruiz and Oto Araújo Vale	303
<i>From Syntax to Semantics: Introducing UMR for NLP Annotation</i>	
Adriana S. Pagano, Magali Sanches Duran and Federica Gamba	312