

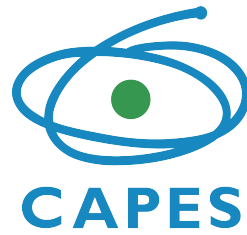
PROPOR 2026

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

**Proceedings - Vol. 1 (Long and Short Papers)**

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](mailto: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.

The scope of PROPOR 2026 is reflected in the organization of the Proceedings, which are divided into two volumes. The present volume presents the accepted papers of the General track. All submissions were peer-reviewed by 82 reviewers from 48 academic and industry institutions across 11 countries. Among the 215 submissions, 94 long papers and 22 short papers were accepted, reflecting an acceptance rate of 54%. In this edition, papers could be submitted in English or Portuguese, with 19 accepted papers written in Portuguese. The accepted contributions span research and developments in resources and evaluation, natural language processing tasks and applications, speech processing tasks and applications, linguistics, multilingual studies, and practical industry applications. The authors are affiliated with institutions in 8 different countries.

The second volume is dedicated to the Demo and Industry tracks, the Best Dissertation awards, the tutorial, 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).

The invited speakers of PROPOR 2026 are internationally recognized experts, reflecting the range, relevance, and applicability of the computational processing of Portuguese:

- Maria das Graças Volpe Nunes (University of São Paulo, Brazil);
- Margarida Petter (University of São Paulo, Brazil);
- Jorge Baptista (University of Algarve & INESC-ID Lisboa, Portugal).

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

## 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

### Program Chairs

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

### Scientific Committee

Adina Vladu, Instituto da Lingua Galega, Universidade de Santiago de Compostela, Spain  
Alberto Simões, Checkmarx & 2Ai, Portugal  
Aline Paes, Universidade Federal Fluminense, Brazil  
Aline Macohin, Pontifícia Universidade Católica do Paraná, Brazil  
Alipio Jorge, Universidade do Porto, Portugal  
Amália Mendes, Universidade de Lisboa, Portugal  
Ana Isabel Mata, Universidade de Lisboa, Portugal  
Anabela Barreiro, Portugal  
António Teixeira, Universidade de Aveiro, Portugal  
António Branco, Universidade de Lisboa, Portugal  
Arnaldo Candido Junior, Universidade Estadual Paulista, Brazil  
Brenda Santana, Federal University of Pelotas, Brazil  
Brett Drury, INESC-TEC, Portugal  
Bruno Martins, IST, Universidade de Lisboa & INESC-ID Lisboa, Portugal  
Carlos H. Ribeiro, Instituto Tecnológico de Aeronáutica, Brazil  
Catarina Oliveira, Universidade de Aveiro, Portugal  
Christopher Shulby, ELSA, USA & Universidade Federal de Goiás, Brazil  
Cláudia Barros, Instituto Federal de São Paulo, Brazil  
Claudia Freitas, Brazil  
Cristina Mota, Linguateca, Portugal  
Daniel Alves, NOVA University, Portugal  
Daniela Barreiro Claro, Universidade Federal da Bahia, Brazil  
David Martins de Matos, IST, Universidade de Lisboa & INESC-ID Lisboa, Portugal  
Dennis Balreira, Federal University of Rio Grande do Sul, Brazil  
Diana Santos, Universitetet i Oslo, Norway  
Eric Laporte, Gustave Eiffel University, France  
Evandro Fonseca, Blip, Brazil  
Evelin Amorim, University of Porto / INESC-TEC, Portugal  
Felipe Meneguzzi, University of Aberdeen, Scotland  
Fernando Batista, Iscte-IUL & INESC-ID Lisboa, Portugal  
Helena Caseli, Federal University of São Carlos, Brazil  
Helena Cameron, Instituto Politécnico de Portalegre, CIDEHUS, Portugal  
Helena Vaz, Universidade Federal de Minas Gerais, Brazil  
Hidelberg Albuquerque, Universidade Federal Rural de Pernambuco, Brazil  
Hugo Gonçalo Oliveira, Universidade de Coimbra, Portugal  
Igor Caetano, University of São Paulo, Brazil  
Iria de Dios Flores, Universitat Pompeu Fabra - Barcelona, Spain  
Ivandré Paraboni, University of São Paulo, Brazil  
Jackson Souza, Universidade Federal da Bahia, Brazil  
João Silva, Universidade de Lisboa, Portugal  
Joel Carbonera, Federal University of Rio Grande do Sul, Brazil

Jorge Baptista, University of Algarve & INESC-ID Lisboa, Portugal  
Larissa Freitas, Federal University of Pelotas, Brazil  
Leandro Wives, Federal University of Rio Grande do Sul, Brazil  
Leonardo Zilio, Université Catholique de Louvain, Belgium  
Lucelene Lopes, University of São Paulo, Brazil  
Marcelo Finger, University of São Paulo, Brazil  
Marcos Garcia, Universidade de Santiago de Compostela, Spain  
Maria Das Graças Volpe Nunes, University of São Paulo, Brazil  
Marlo Souza, Universidade Federal da Bahia, Brazil  
Micaela Aguiar, Universidade do Minho, Portugal  
Nádia Silva, Universidade Federal de Goiás, Brazil  
Oto Vale, Universidade Federal de São Carlos, Brazil  
Pablo Gamallo, Universidade de Santiago de Compostela, Spain  
Pablo Faria, Universidade Estadual de Campinas, Brazil  
Paula Cardoso, Universidade Federal do Pará, Brazil  
Plinio Barbosa, Universidade Estadual de Campinas, Brazil  
Prakash Poudyal, Kathmandu University, Nepal  
Priscila Osório, Brazil  
Purificação Silvano, Universidade do Porto, Portugal  
Rafael Anchieta, Instituto Federal do Maranhão, Brazil  
Raquel Amaro, NOVA University, Portugal  
Renata Vieira, Évora University, CIDEHUS, Portugal  
Ricardo Rodrigues, Universidade de Coimbra, Portugal  
Ricardo Ribeiro, Iscte-IUL & INESC-ID Lisboa, Portugal  
Rozane Rebechi, Federal University of Rio Grande do Sul, Brazil  
Roney Santos, Universidade Federal da Bahia, Brazil  
Rui Sousa-Silva, Universidade do Porto, Portugal  
Sandra Aluisio, University of São Paulo, Brazil  
Sara Mendes, Universidade de Lisboa, Portugal  
Saullo Haniell, Pontifícia Universidade Católica de Campinas, Brazil  
Sebastian Pado, University of Stuttgart, Germany  
Sheila Castilho, ADAPT Centre, Ireland  
Stella Tagnin, University of São Paulo, Brazil  
Suemi Higuchi, Fundação Getulio Vargas, Brazil  
Susana Duarte Martins, NOVA University, Portugal  
Susana Sotelo Docío, Universidade de Santiago de Compostela, Spain  
Thiago Pardo, University of São Paulo, Brazil  
Valeria de Paiva, Topos Institute, USA  
Valeria Feltrim, Universidade Estadual de Maringá, Brazil  
Viviane Moreira, Federal University of Rio Grande do Sul, Brazil  
Vladia Pinheiro, Universidade de Fortaleza, Brazil

### **Invited Speakers**

Maria das Graças Volpe Nunes, University of São Paulo, Brazil  
Jorge Baptista, University of Algarve & INESC-ID Lisboa, Portugal  
Margarida Petter, University of São Paulo, Brazil

## Keynote Talk

# The place of NLP in the age of generative AI

Maria das Graças Volpe Nunes  
University of São Paulo, Brazil



**Abstract:** In this talk, I intend to provoke some reflections on the changes in the field of NLP caused by the boom of generative AI. The widespread use of Large Language Models has shifted the NLP community towards less linguistic and more engineering tasks. The success of chatbots suggests that the goals of NLP have been almost entirely achieved: machines seem to process, understand, and generate natural language as well as, or even better than, humans. How does the NLP community see itself in this scenario? Is the success of LLM our greatest victory, or is it a sign that our particular expertise is becoming obsolete? These and other questions, with almost no answers, will be part of this talk.

**Bio:** Maria das Graças Volpe Nunes is a senior professor at ICMC-USP and associate researcher at NILC-ICMC. She has worked in NLP since her PhD in Computer Science from PUC-Rio de Janeiro (1991). She is one of the founders of NILC (1994), where she coordinated projects on the following topics, always for Brazilian Portuguese: spelling and grammar correction, machine translation, text normalization, automatic summarization, sentiment analysis, parsing, and corpus annotation. She is member of C4AI - Center for Artificial Intelligence (FAPESP, USP, IBM) since 2021, and is member of the Board of the group *BPLN - Brasileiras em PLN* (Brazilian Women in NLP). She is co-editor of the free digital book *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português* (Natural Language Processing: Concepts, Techniques and Applications in Portuguese), an initiative of BPLN.

# Keynote Talk

## Abstract Meaning Representation: New Rosetta Stones for the Babel Entanglement

Jorge Baptista

University of Algarve & INESC-ID Lisboa, Portugal



**Abstract:** The enduring ambition of Natural Language Processing is to move beyond form and toward meaning. Semantic representation frameworks such as Abstract Meaning Representation (AMR) were conceived as a response to this challenge, proposing graph-based encodings of meaning that abstract away from surface syntax and promise a common semantic ground across languages. In this sense, AMR can be seen as a modern attempt at a Rosetta Stone: a shared representation meant to cut through the Babel-like diversity of natural languages. Yet, as AMR has matured, its limitations have become increasingly visible. Designed primarily around English, it struggles to account for the lexical, morphological, and syntactic specificities of other languages, including Portuguese. More fundamentally, open questions remain about the status of AMR-like formalisms: are they true semantic representations or annotation compromises? How much semantic interpretation should be encoded explicitly, and how much can be left to inference, context, or downstream models? These questions acquire renewed urgency in the era of large language models. While LLMs appear to “understand” meaning implicitly, their representations remain opaque, difficult to control, and hard to evaluate. Against this backdrop, explicit semantic representations regain relevance—not as competitors to neural models, but as interfaces, anchors, and points of interpretability between language, meaning, and computation. This talk argues that language-aware adaptations of AMR—grounded in the linguistic structure of languages such as Portuguese—offer a promising path forward. Rather than a single universal Rosetta Stone, we may need a family of interrelated stones: abstract enough to support multilinguality, yet lexicalized enough to respect linguistic diversity. In revisiting semantic representation through this lens, the talk reflects on how explicit meaning models can help disentangle today’s Babel, even in an LLM-dominated landscape.

**Bio:** Jorge Baptista graduated in Languages and Literatures - Portuguese Studies and obtained an MA in Portuguese Linguistics from Faculdade de Letras da Universidade de Lisboa with a thesis on nominal compounding and electronic dictionaries. His PhD thesis on predicative nouns and nominalization of adjectival constructions was presented to the Universidade do Algarve, where he started working in 1992. He is a researcher at INESC-ID Lisboa since 2005. His current research interests in NLP concern Information Extraction (Named Entity Recognition and Relation Extraction), Parsing, and Computer Assisted Language Learning; in Linguistics proper, he’s working on the Lexicon-Syntax description of verbal, adverbial and idiomatic constructions. He is also involved in the REAP.PT and OOBIAN projects.

**Keynote Talk**  
**Portuguese: a language in transit through Europe, Africa,  
South America, and Asia**

**Margarida Petter**  
University of São Paulo, Brazil



**Abstract:** This presentation will consider the Portuguese language that, after leaving Europe—Galicia and Portugal—reached the African, South American, and Asian continents. Given the extensive scope of this journey, only the African and Brazilian varieties resulting from the expansion of the Portuguese language within a colonial context will be addressed. The comparison of these varieties, acquired as second languages in particular linguistic ecologies, reveals similarities and differences arising from the contact between speakers of African language from the Niger-Congo and Afro-Asiatic phyla and speakers of Portuguese. This circumstance, now evaluated as an Afro-Brazilian continuum of Portuguese, can also lead to the creation of new linguistic varieties. It is expected that presenting aspects that bring these varieties closer and further apart will provide elements of interest for research in computational linguistics.

**Bio:** Margarida Petter is a senior professor in the Department of Linguistics of the Faculty of Philosophy, Letters and Human Science at the University of São Paulo. She has a master's degree in African Linguistics from the University of Abidjan, Côte d'Ivoire and a PhD from the University of São Paulo. Her main research areas are African Linguistics and the contact between African languages and Brazilian Portuguese. She conducted an international project in language contact in Brazil and has published an introduction to African linguistics in Portuguese.

## Table of Contents

<i>Retrieval-Augmented Generation and Knowledge Graphs in Portuguese-Language Legal Documents</i> Vinícius Teles de Oliveira, Deivison Oliveira da Silva, Mateus de Almeida Souza, Maurício Rodrigues Lima, Sávio Salvarino Teles de Oliveira and Thierson Couto Rosa . . . . .	1
<i>Gender Identification in Brazilian Portuguese Product Reviews: A Comparative Study of Classical Models, BERT, and LLMs</i> Tiago de Melo and Carlos M. S. Figueiredo . . . . .	11
<i>Leveraging political alignment information for stance detection</i> Matheus Camasmie Pavan and Ivandr� Paraboni . . . . .	20
<i>Dependency Distance Effects on Eye-Tracking Measures in Brazilian Portuguese</i> Diego Alves . . . . .	30
<i>Experimental Evaluation of Topic Modeling Methods for Categorizing Irregularities in Health-related news</i> Alysson Guimarães, Methanias Colaço Junior, Samuel Almeida and Raphael Fontes . . . . .	41
<i>A Multilingual Voice Analytics Module for Contact-Center Hiring</i> Wagner W. �vila Bombardelli, Vanessa Marquiasavel Serrani, Edgard Kuboo and Erica C. Marins Miss�o . . . . .	57
<i>CLARIN-PT-LDB: An Open LLM Leaderboard for Portuguese to assess Language, Culture and Civility</i> Jo�o Ricardo Silva, Lu�s Gomes and Ant�nio Branco . . . . .	67
<i>A Comparison of Methods to Bias Translation Toward Portuguese Variants</i> Catarina Costa and Sebastian Pad� . . . . .	78
<i>Que ao mestre vai mat�-lo? The evolution of prepositional accusatives in Portuguese across time</i> Helena Rodrigues Menezes de Oliveira Vaz . . . . .	93
<i>Topic Modeling in Brazilian Portuguese Documents on Antimicrobial Resistance</i> Enrique Reis Susin and Lilian Berton . . . . .	103
<i>Geological Text Summarization Using Generative Large Language Models</i> Matheus Stein de Aguiar, Rafael Oleques Nunes and Dennis Giovanni Balreira . . . . .	111
<i>Retrieval-Augmented Generation with Small Language Models for Fake News Detection</i> Lucca Baptista Silva Ferraz, Jh�lia de Souza Leal, Anderson Raymundo Avila, Thiago Alexandre Salgueiro Pardo, Fernando Batista and Renato Moraes Silva . . . . .	120
<i>Bridging Citizens and Public Services: Improving Service Association with Retrieval-Augmented Generation (RAG) Labels</i> Ticiania L. Coelho da Silva, Celso Fran�a, Marcos Andr� Gon�alves, Leonardo Rocha, Leonardo Alamy, Fernando Sola Pereira and Eduardo Soares de Paiva . . . . .	131
<i>Negation-Aware Data Augmentation for Portuguese Natural Language Inference</i> Maria Cec�lia M. Corr�a, Felipe S. F. Paula, Matheus Westhelle and Viviane P. Moreira . . . . .	141
<i>To Describe or Not to Describe? Benchmarking Database Representations for Schema Linking in Text-to-SQL</i> Daiane Ucceli Kreitlow and Hil�rio Tomaz Alves de Oliveira . . . . .	151

<i>Evaluating Automated Scoring Models on Official ENEM Essays</i>	
Laís Nuto Rossman, Igor Cataneo Silveira and Denis Deratani Mauá . . . . .	161
<i>Enhancing Brazilian Inflation Forecasts through Sentiment Analysis Using Large Language Models</i>	
Lucas Miranda Mendonça Rezende, Cézio Luiz Ferreira Junior, Mateus Tarcinalli Machado and Evandro Eduardo Seron Ruiz . . . . .	172
<i>NorBERTo: A ModernBERT Model Trained for Portuguese with 331 Billion Tokens Corpus</i>	
Lucas F. A. O. Pellicer and Guilherme Rinaldo . . . . .	183
<i>AspectRAG: Uma Arquitetura de Recuperação e Geração para Análise de Sentimentos Baseada em Aspectos</i>	
Erick R. Ribeiro, André Carvalho and Rhedson Esashika . . . . .	194
<i>Semantic Representation of Relative Clauses in Lexicalized Abstract Meaning Representation</i>	
Jorge Baptista and Sónia Reis . . . . .	203
<i>Portuguese Sentiment Analysis with Open-Source LLMs: Models, Prompts, and Efficient Deployment</i>	
João V R J Lima, Vlândia Pinheiro and Carlos Caminha . . . . .	212
<i>Auditing the Evaluators: How Far Can Automatic Evaluation Go in Assessing Portuguese Financial Texts?</i>	
Marina Ramalhete Masid, Gabriel Assis, Daniela Vianna, Aline Paes and Altigran Soares da Silva	222
<i>A Multimodal Framework for Financial Fake News Detection for Brazilian Portuguese</i>	
José Vitor Souza Cardoso Requena, João Victor Assaoka Ribeiro and Lilian Berton . . . . .	234
<i>A UD Parser to the Rescue: A Method for Bringing a Classical Annotated Corpus to Life Again</i>	
Lucelene Lopes, Magali S. Duran and Thiago A. S. Pardo . . . . .	240
<i>Data Augmentation for Named Entity Recognition in Domain-Specific Scenarios in Portuguese</i>	
Higor Moreira, Patricia Ferreira da Silva, Luciana Bencke and Viviane Moreira . . . . .	250
<i>Token-Level Pun Location Using Multi-Layer BERT with Mixture of Experts</i>	
Rafael Torres Anchiêta, Roney Lira de Sales Santos and Raimundo Santos Moura . . . . .	260
<i>Language Effects in Text-to-SQL Across English and Portuguese</i>	
Lucas Nobre, Suele Sousa, Savio Teles and Anderson Soares . . . . .	270
<i>Specializing a Small Language Model for Closed-Domain Portuguese RAG using Knowledge Graph Supervision</i>	
Josué Caldas, Elvis de Souza, Patrícia Silva and Marco Pacheco . . . . .	281
<i>Modeling Linguistic Violence: An Ontology-Based Framework for the Computational Analysis of Violence Manifested in Language</i>	
Brenda Salenave Santana, Ana Marilza Pernas and Aline A. Vanin . . . . .	291
<i>Software for Automatic Speech Recognition via Whisper models applied to Oral History interviews in the Portuguese language</i>	
Edgleide de Oliveira Clemente da Silva, Fernando Rezende Zagatti, Filipe Loyola Lopes, Anderson Dias Duarte, Rodrigo Bonacin and Angela Maria Alves . . . . .	302
<i>Uma Abordagem Híbrida para Predição de Faixa Etária de Autores de Textos Escritos na Língua Portuguesa</i>	
Alice Rezende Ribeiro and Luiz Henrique de Campos Merschmann . . . . .	311

<i>Enhanced Universal Dependencies in the Wild: Evaluating Portuguese EUD Parsing in Realistic Scenarios</i>	
Elvis A. de Souza and Thiago A. S. Pardo	321
<i>UlyssesLegalNER-Br: from Legislative to Legal, a comprehensive corpus of Brazilian legal documents for Named Entity Recognition</i>	
Hidelberg O. Albuquerque, Ellen Souza, Danilo C. G. Lucena, Héldon J. O. Albuquerque, Nádia F. F. da Silva, Márcio de S. Dias, Rafael O. Nunes, Adriano L. I. Oliveira and André C. P. L. F. de Carvalho	331
<i>The PROPOR Ecosystem: Structure, Roles, and Evolution of Portuguese-Language NLP</i>	
Rafael O. Nunes, Gustavo L. Tamiosso, Pedro L. C. de Andrade, Matheus S. de Aguiar, Rafael P. de Gouveia, Higor Moreira, Bruno Tavares, Laura P. de Gouveia, Felipe S. F. Paula, Andre Spritzer, Hidelberg O. Albuquerque, Nádia F. F. da Silva, Ellen P. R. S. Pereira, Dennis G. Balreira and Joel L. Carbonera	342
<i>Twenty Years of HAREM: A Reproducible Audit and Reassessment of Portuguese Named Entity Recognition</i>	
Rafael O. Nunes, André Spritzer, Carla M. D. S. Freitas and Dennis G. Balreira	351
<i>Analysis of Machine Translators on Sentences Generated by Portuguese Image Captioning Models</i>	
Natan Moura, João Medrado Gondim, Daniela Barreiro Claro and Babacar Mane	360
<i>Analyzing Debate Dynamics in the Portuguese Parliament with Dialogue Action Flows</i>	
Patrícia Ferreira, Ana Alves, Catarina Silva and Hugo Gonçalves Oliveira	369
<i>AMALIA: A Fully Open Large Language Model for European Portuguese</i>	
Afonso Simplício, Gonçalo Vinagre, Miguel Moura Ramos, Diogo Tavares, Rafael Ferreira, Giuseppe Attanasio, Duarte M. Alves, Inês Calvo, Inês Vieira, Rui Guerra, James Furtado, Beatriz Canaverde, Iago Paulo, Vasco Ramos, Diogo Glória-Silva, Miguel Faria, Marcos Treviso, Daniel Gomes, Pedro Gomes, David Semedo, André Martins and João Magalhães	380
<i>LegalSim-PT: Building a Dataset for Legal Document Simplification in Portuguese Leveraging Linguistic Metrics</i>	
Arthur Scalercio, Maria José Finatto and Aline Paes	392
<i>Portho: A Corpus-Based Resource of Orthographic Neighbors in European Portuguese</i>	
Eugénio Ribeiro, David Antunes, Nuno Mamede and Jorge Baptista	406
<i>Discovery of Legal Patterns in Civil Petitions via LLM-Based Fact Extraction and Density Clustering</i>	
Rhedson Esashika, Carlos M. S. Figueiredo and Tiago de Melo	416
<i>Automated Reformulation of Argumentative Essays to Improve Argument Organization and Development</i>	
Naomi James Sutcliffe de Moraes and Denis Deratani Mauá	425
<i>Automatic Question classification in Portuguese: A Large-Scale Dataset and Comparative Evaluation of Classification Strategies</i>	
Murilo Boccardo and Valéria D. Feltrim	436
<i>Exploring Knowledge Graphs for Automatic Fake News Detection in Portuguese</i>	
Lucas dos Santos, Manoel Rodrigues Euclides Santos, Yuri Silva Souza, João Pedro Holanda Sousa and Roney Lira de Sales Santos	446
<i>Anatomy of Data Repositories for the Analysis and Detection of Toxicity in Portuguese</i>	
Lorena Souza Moreira, Paula Teresa M. Gibrim, Leonardo Rocha and Julio C. S. Reis	456

<i>Sintomas Linguísticos: Geração Aumentada por Recuperação e Raciocínio em LLMs sob a Variação Português-Inglês em Contextos Médicos</i>	
Guilherme Vianna de Moura, Gabriel Assis and Aline Paes . . . . .	467
<i>BIPA: Brazilian Portuguese Phonetic Dataset with Dialectal Variations in IPA Standard</i>	
Thiago Monteles de Sousa, Lucas Rafael Gris and Nádia Félix Felipe da Silva . . . . .	478
<i>Avaliação Automática de Redações do Enem: Um Estudo Empírico sobre Representações Linguísticas e Contextuais</i>	
Gabriel Gonçalves de Matos and Valéria D. Feltrim . . . . .	488
<i>Analysing LLMs for spelling normalization of 18th century Portuguese</i>	
Helena Freire Cameron, Aline Paes, Fernanda Olival and Renata Vieira . . . . .	498
<i>Detecting Stuttering with Artificial Intelligence: A Hybrid Method for Brazilian Portuguese</i>	
Rubens Perini Buzzeti, Paula Bianca Meireles de Moura Buzzeti and Roney Lira de Sales Santos	507
<i>Levados em Consideração: Uma Avaliação de Vieses de Estima por Raça, Gênero e Região em Grandes Modelos de Linguagem em Português Brasileiro</i>	
João Lucas Lima de Melo and Marlo Souza . . . . .	516
<i>Global vs. Local Sentence Embeddings for Brazilian Portuguese: Revisiting Monolingual Models in the Age of Foundation Models</i>	
Matheus Peixoto, Guilherme Silva, Giacomo Figueredo, Pedro Silva and Eduardo J. S. Luz . .	529
<i>LexIris-pt and LexBert-pt: Specialized Sentence Embeddings for Legal Similarity in Brazilian Portuguese</i>	
Willgnner Ferreira Santos, João Gabriel Grandotto Viana, Antônio Pires de Castro Júnior, Fernando Ribeiro Trindade and Nádia Félix Felipe da Silva . . . . .	540
<i>The Inadequacy of Automatic Evaluation Metrics in Question Answering: A Case-Study in Portuguese</i>	
Júlia da Rocha Junqueira and Viviane P. Moreira . . . . .	551
<i>Optimizing Efficiency in Multi-Stage Semantic Re-ranking Architectures</i>	
Artur M. A. Novais, Anna P. V. L. B. Moreira, Maria C. X. de Almeida, João P. C. Presa, Fernando M. Federson and Sávio S. T. de Oliveira . . . . .	562
<i>Marcação de correferência para a caracterização de personagens em obras literárias em português</i>	
Diana Santos, Luisa Mara Silva Lima and Emanuel Pires . . . . .	572
<i>Diálogos Tóxicos: Gatilhos e Padrões de Interação no Reddit Brasileiro</i>	
Giovana Piorino, Marco Antônio de Alcântara Machado, Luiz Henrique Quevedo Lima, Adriana Pagano and Ana Paula Couto da Silva . . . . .	581
<i>Contrastive and Adversarial Disentanglement for Speaker Representations in Brazilian Portuguese</i>	
Ariadne Nascimento Matos, Arnaldo Candido Junior and Moacir Antonelli Ponti . . . . .	591
<i>Can I guess where you are from? Modeling dialectal morphosyntactic similarities in Brazilian Portuguese</i>	
Manoel Siqueira and Raquel Freitag . . . . .	601
<i>A Lexicon-Grammar of Brazilian Portuguese Predicative Adjectives</i>	
Ryan Martinez, Jorge Baptista and Oto Araújo Vale . . . . .	611
<i>NLP-based Page Classification for Efficient LLM Extraction from Brazilian Public Tender Documents</i>	
Pedro Campos, Ivo de Medeiros and Adailton de Araújo . . . . .	621

<i>Lexical and Orthographic Variation in Portuguese Financial Tweets: Annotation, Analysis, and Implications for Embedding Models</i>	
Ariani Di Felippo, Norton Trevisan Roman, Bryan K. S. Barbosa, Gabriela Pinheiro de Oliveira and Clarissa Lenina Scandarolli . . . . .	628
<i>Síntese de Voz Emocional Multi-Idioma para Português Brasileiro: Uma Análise Comparativa de Abordagens de Ajuste Fino</i>	
Daniel Oliveira de Brito, Sidney Evaldo Leal and Arnaldo Candido Junior . . . . .	638
<i>CoDEL-BR: An Electoral Debate Corpus from Brazilian Municipal Elections</i>	
Alessandra Gomes, Aline Paes and Helena Caseli . . . . .	647
<i>Causal_QA.PT: A Human–LLM Co-Curated Benchmark for Causal Question Answering in Portuguese Language</i>	
Lia Furtado, Cíntia Araripe, Jocelani Castilhos, Lucas Holanda and Vladia Pinheiro . . . . .	657
<i>ConsumerBR: A Large-Scale Corpus of Consumer Complaints in Brazilian Portuguese</i>	
Luis A. Duarte, Pedro Giacomini, Vitória Bispo, Mariana O. Silva, Adriano C. M. Pereira and Gisele L. Pappa . . . . .	667
<i>Geração de consultas SPARQL a partir de linguagem natural</i>	
Heber Gustavo Xavier de Castro and Clever Ricardo Guareis de Farias . . . . .	676
<i>Retrato_Cantado: Criação e Análise de um Corpus para Representações de Identidade em Letras de Músicas Brasileiras</i>	
Vitória P. Firmino, Janaina N. de S. Lopes, Bruno M. Nogueira and Valéria Q. dos Reis . . . . .	687
<i>ALBA: A European Portuguese Benchmark for Evaluating Language and Linguistic Dimensions in Generative LLMs</i>	
Inês Vieira, Inês Calvo, Iago Paulo, James Furtado, Rafael Ferreira, Diogo Tavares, Diogo Glória-Silva, David Semedo and João Magalhães . . . . .	697
<i>Cartas Indígenas ao Brasil: Classificação Multi-Rótulo</i>	
Caio Almeida, Renata Vieira and Débora Abdalla . . . . .	708
<i>Identificação de notícias falsas em português: um olhar sobre a generalização de modelos</i>	
Raphael Guedes, Bruno Barros and Hugo do Nascimento . . . . .	717
<i>Agent Orchestration - LLM for Legal Metadata Extraction: A Comparative Analysis of Efficiency and Precision</i>	
Luiz Anísio Batitucci, Luciane Inácia Lopes, Rhodie Ferreira and Emerson Cabrera Paraiso . . . . .	727
<i>Avaliação Automática de Redações do Enem: Uma Análise Comparativa entre Engenharia de Características e Transformers</i>	
Pâmela Camilo Chalegre, Vitor da Rocha Machado and Valéria Delisandra Feltrim . . . . .	738
<i>Avaliação End-to-End de um Sistema RAG para Documentos Hospitalares em Português</i>	
Murilo Vargas da Cunha, Marília Rosa Silveira, César Brasil Sperb, Brenda Salenave Santana, Larissa Astrogildo Freitas and Ulisses Brisolará Corrêa . . . . .	749
<i>Extending an Ensemble Baseline with Corpus-Based Graph Features for Portuguese Pun Detection</i>	
Avelar Rodrigues de Sousa, Camilla Soares Sousa, Carlos Henrique Santos Barros and Rafael Torres Anchieta . . . . .	759
<i>RacismoBR: A Manually Annotated Dataset for Racist Discourse Detection in Brazilian Portuguese</i>	
João Vítor Vaz, Fabrício Benevenuto and Marcos André Gonçalves . . . . .	770

<i>Unsupervised Evaluation of Explanations for Hate Speech Classification in Portuguese</i> Isabel Carvalho, Hugo Gonçalo Oliveira and Catarina Silva .....	780
<i>Neuro-symbolic Approaches for Rubric-Based Automatic Essay Evaluation of ENEM Essays</i> Igor Cataneo Silveira and Denis Deratani Mauá .....	790
<i>dialect2vec: Um método baseado em vetores para transcrição dialetal do português a partir de questionários do ALiB</i> Laila Mota, Daniela Barreiro Claro, Eloize R. Marques Seno and Rerisson Cavalcante de Araújo	800
<i>Compression-based Language Complexity under Register Variation in Portuguese</i> Felipe Ribas Serras and Marcelo Finger .....	808
<i>Certas Palavras: A 1980s-90s Brazilian Radio Corpus to Test TTS Models in Noisy Multi-Speaker Dialogues</i> Gustavo Evangelista Araújo, Moacir Ponti, Arnaldo Candido Junior, Sidney Leal, Edresson Casanova, Renato Moraes Silva, Miguel Oliveira Jr., Adriana Barbosa Santos, Gustavo Wadas Lopes and Sandra Aluisio .....	819
<i>Análise de Sentimento Baseada em Aspectos no Domínio de Acomodações Utilizando o modelo BER-Timbau</i> Franco Noronha Pereira, Larissa Astrogildo de Freitas and Ulisses Brisolará Corrêa .....	830
<i>Combining Real and Synthetic Speech for ASR Adaptation in Brazilian Portuguese</i> Daniel R. da Silva, Maria Eduarda S. Borba, Állan C. P. Silva, Maria Carolina S. Barreto, Arthur F. de Moraes, Paulo V. dos Santos, Guilherme C. Dutra, Sávio S. T. de Oliveira and Anderson da S. Soares .....	838
<i>Structured Summaries for Retrieval-Augmented Generation in Portuguese-Language Consumer Complaints</i> Rafael Sant’Ana, Pedro Garcia, Luis A. Duarte, Mariana O. Silva, Adriano C. M. Pereira and Gisele L. Pappa .....	847
<i>Uso de técnicas de Aprendizado de Máquina e Modelos de Língua de Larga Escala para avaliação automática de textos do exame Celpe-Bras</i> Rafael Oleques Nunes, Bernardo Cobalchini Zietolie, Ricardo Zanini De Costa, Rodrigo Brock da Silva, João Victor Piardi Pacheco, Rafaela Dall’Agnol da Rocha, Dennis Giovanni Balreira, Elisa Marchioro Stumpf and Juliana Roquele Schoffen .....	858
<i>Biatron: A Parameter-Efficient Small Language Model for Brazilian Portuguese with Integrated Mathematical Reasoning</i> Daniel Fazzioni, Maria C. X. de Almeida, Anna P. V. L. B. Moreira, Anderson S. Soares, Sávio S. T. de Oliveira and Fernando M. Federson .....	868
<i>Exploring Sentiment Analysis Approaches in a Public Agency Security News Dataset</i> Thiago Ruiz Lobo and Claudia Aparecida Martins .....	878
<i>Viés de gênero na tradução automática: uma avaliação no par linguístico inglês-português</i> Tayane A. Soares, Yohan B. Gumiel, Rafael Junqueira, Tácio Gomes and Adriana Pagano ...	889
<i>Evaluating Reference-Free Summarization Quality Metrics for Portuguese: A Study with Human Judgments in Financial News</i> João Victor Assaoka Ribeiro, Thomas Pires Correia, José Vitor Souza Cardoso Requena and Lilian Berton .....	899

<i>Formalizing the DATASUS RTS: An Ontological Model for a Resource Description Framework Knowledge Graph</i>	
Vitor Pires, Dalvan Griebler and Felipe Meneguzzi . . . . .	908
<i>Efficient Fine-Tuning Methods for Portuguese Question Answering: A Comparative Study of PEFT on BERTimbau and Exploratory Evaluation of Generative LLMs</i>	
Mariela M. Nina, Caio Veloso Costa, Lilian Berton and Didier A. Vega-Oliveros . . . . .	917
<i>Structured Sentiment Analysis in Brazilian Portuguese: An Exploratory Study Using BERTimbau</i>	
Andrew B. Campos, Ulisses B. Corrêa and Larissa A. de Freitas . . . . .	927
<i>JabuticaBERT: Modern Portuguese Encoders from Scratch with RTD and Long-Context Training</i>	
Thiago Porto, Gabriel Gomes, Alexandre Bender, Ulisses Corrêa, Larissa Freitas, William Cruz and Marcellus Amadeus . . . . .	933
<i>Evaluating Small Language Models for English-to-Portuguese Translation: Impact of Model Scale and Quantization</i>	
Gustavo Lopes Tamiosso, Rafael Oleques Nunes and Dennis Giovanni Balreira . . . . .	943
<i>Think Portuguese with Bode Reasoning</i>	
Gabriel Lino Garcia, André da F. Schuck, João R. R. Manesco, Pedro Henrique Paiola, Leandro A. Passos and João Paulo Papa . . . . .	953
<i>Rating–Text Mismatch in Brazilian Portuguese Reviews: How Reliable Are Zero-Shot LLMs?</i>	
Emanuelle Marreira, Carlos M. S. Figueiredo and Tiago de Melo . . . . .	959
<i>Accelerating Portuguese Masked Diffusion Models through Representation Alignment</i>	
Adalberto Ferreira Barbosa Junior, Lucas Lima Neves and Adriano César Santana . . . . .	968
<i>Automatic Speech Recognition for Child Reading: A Phonemic Approach using Isolated Words in Brazilian Portuguese</i>	
Aline N. Rodrigues and Carlos H. C. Ribeiro . . . . .	974
<i>Improving Machine Translation of Idioms: A Spanish–Galician Parallel Dataset and Synthetic Augmentation Approach</i>	
Lúa Santamaría Montesinos, Saúl Buján, Daniel Bardanca and Pablo Gamallo . . . . .	980
<i>Multi-Agent Architecture with RAG and Dynamic Context Windows for Text-to-SQL Optimization</i>	
Willgnner Ferreira Santos, Paulo Victor dos Santos, Marcella Scoczynski Ribeiro Martins, Larissa Freire Lekakis, Frederico Lemes Rosa, Bruno Matheus Costa, Miguel Alves Pereira Filho and Isabella Alves Montalvão . . . . .	988
<i>Automatic Metrical Scansion of Galician Poetry: First Results</i>	
Pablo Ruiz Fabo, Pauline Moreau and Anxo Alonso Pérez . . . . .	994
<i>MATH-PT: A Math Reasoning Benchmark for European and Brazilian Portuguese</i>	
Tiago Teixeira, Ana Carolina Erthal, Juan Belieni, Beatriz Canaverde, Diego Mesquita, Miguel Faria, Eliezer de Souza da Silva and André F. T. Martins . . . . .	1005
<i>Democratizing Legal Analytics: Resource-Efficient Information Extraction for Brazilian Case Law</i>	
Rodrigo Filippi Dornelles . . . . .	1011
<i>Field of Science and Technology Classification of Academic Documents in Portuguese</i>	
Ivo Simões, Hugo Gonçalo Oliveira and João Correia . . . . .	1021
<i>Quando as Máquinas “Pensam”: Antropomorfização no Discurso sobre IA - Riscos Conceptuais e Proposta de Léxico Não Antropomórfico para PLN em Português</i>	
Anabela Barreiro . . . . .	1027

<i>Semantic adapters in text-to-SQL for low-resource languages: the importance of semantic information</i> Anton Bulle Labate and Fabio Gagliardi Cozman .....	1032
<i>CURUPIRA: Clever guard for harm and linguistic prompt mitigation in Brazilian Portuguese</i> Rogério Sousa, William Alberto Cruz-Castañeda, José Roberto Homeli Silva and Marcellus Amadeus .....	1038
<i>Lost in Quantization: Activation Outliers Explain Language-Specific FP8 Sensitivity in Llama-3</i> Guilherme Silva, Pedro Silva, Matheus Peixoto, Gladston Moreira and Eduardo Luz .....	1044
<i>A Multitask Transformer for Offensive Language Detection and Target Identification in HateBR</i> Guilherme Silva, Pedro Silva, Matheus Peixoto, Gladston Moreira and Eduardo Luz .....	1049
<i>LARI Dataset: A Native Portuguese Question Answering Dataset from Brasileiras em PLN</i> Júlia da Rocha Junqueira, Larissa A. de Freitas and Ulisses Brisolara Corrêa .....	1055
<i>Evolução de Padrões Linguísticos na Escrita Científica em Português: Uma Análise com NILC-Matrix</i> Thiago Ruiz Lobo and Claudia Aparecida Martins .....	1062
<i>Libras-UFPeL Corpus: A Parallel Dataset of Brazilian Sign Language and Portuguese for Multimodal Research and Processing</i> Antionielle Martins, Brenda S. Santana, Francielle Martins, Tatiana Lebedeff, Darley Nunes and Luisa Bohm .....	1068
<i>The Superficiality Bias: Community Votes and Answer Utility in Portuguese Health Question Answering</i> Carlos Henrique Santos Barros, Gustavo Figueredo Rodrigues de Sousa and Rogério Figueredo de Sousa .....	1074
<i>FlexQwen: Exploring Hybrid Objectives and Text Originality for Portuguese</i> Miguel de Mello Carpi and Marcelo Finger .....	1079
<i>NormaTex-MapSNOMED: Bridging the Gap Between Brazilian Portuguese Clinical Narratives and SNOMED CT</i> Isabela Araujo, Claudia Moro and Layslla Martinez .....	1085
<i>Prompt Engineering for Named Entity Extraction from Portuguese Legal Documents</i> Giovanni Maffeo, Catarina Silva and Hugo Gonçalo Oliveira .....	1092