



LREC 2026

**LANLP: Bridging Ibero and Latin American NLP
Communities**

Networking Symposium Proceedings

Editors

**Luis Chiruzzo, Pablo Gamallo, Rafael Muñoz Guillena
and German Rigau Claramunt**

16 May 2026

©ELRA Language Resources Association (ELRA), 2026

These proceedings are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-95-1

Preface

It is with great pleasure that we present the proceedings of the **Networking Symposium on Natural Language Processing in Ibero and Latin America (LANLP)**, co-located with **LREC 2026** in Palma de Mallorca. This inaugural symposium was established to fill a critical historical gap: providing a dedicated forum for researchers across the Iberian Peninsula and Latin America to share initiatives, corpora, and tools.

The Ibero-American region represents a vast and diverse linguistic landscape, encompassing major world languages such as Spanish (558M speakers) and Portuguese (267M speakers), alongside tens of millions of speakers of indigenous languages including Quechua, Guaraní, Nahuatl, and Aymara. Despite this cultural richness, over 88% of the world's languages remain unsupported by modern language technologies. This networking event directly addresses that disparity by fostering community-driven resource development and evaluation for both major and minoritized languages.

The emergence of Large Language Models (LLMs) represents a pivotal shift for our community, moving the region from the passive consumption of Northern-centric models to the active development of localized foundation models. While this era presents challenges—such as digital scarcity, algorithmic colonialism, and "language domination"—it also offers unprecedented opportunities for innovation:

- **Technological Sovereignty:** Initiatives like **ILENIA** and **ALIA** projects demonstrate the feasibility of building regional foundation models that prioritize regional specificity and language inclusion.
- **Efficiency and Equity:** Our research agenda emphasizes Small Language Models (SLMs) and synthetic data to mitigate linguistic inequalities.
- **Trustworthy AI:** The community is prioritizing privacy, accountability, and cultural alignment, ensuring that deployment in high-stakes environments like healthcare is both robust and ethically grounded.

These proceedings reflect a research agenda where linguistic diversity and technical robustness are inseparable. The success of this symposium is built upon the synergy of established ecosystems, including **SEPLN**, **PROPOR**, **AmericasNLP** and **CLARIAH-ES**. By weaving together these varied perspectives—from computer science to the digital humanities—we aim to amplify the potential of our current research capabilities on a global stage.

We hope these proceedings serve as a catalyst for long-term engagement, converting short-term networking into durable cross-regional collaboration and ensuring that Ibero-American languages thrive in the digital age.

The LANLP 2026 Organizing Committee

Eugenio Martínez Cámara, Luis Chiruzzo, Pablo Gamallo, Rafael Muñoz, and German Rigau

Organizing Committee

Organizers

Pablo Gamallo (PROPOR and Universidade de Santiago de Compostela, Spain)
Eugenio Martínez Cámara (Universidad de Jaén, Spain)
Rafael Muñoz Guillena (SEPLN and Universidad de Alicante, Spain)
Luis Chiruzzo (AmericasNLP and Universidad de la República, Uruguay)
German Rigau Claramunt (CLARIA-ES and Universidad del País Vasco, Spain)

Scientific Committee

Itziar Aldabe (Universidad del País Vasco, Spain)
Elvis Alves de Souza (Universidade de São Paulo, Brazil)
Xabier Arregi (Universidad del País Vasco, Spain)
Elena Battaner (Universidad Rey Juan Carlos, Spain)
Alba Bonet (Universidad de Alicante, Spain)
Marília Costa Rosendo Silva (Universidade de São Paulo, Brazil)
Viviana Cotik (Universidad de Buenos Aires, Argentina)
Fermín Cruz (Universidad de Sevilla, Spain)
Iria de-Dios-Flores (Universitat Pompeu Fabra, Spain)
Cristina España-Bonet (BSC, Spain)
Ainara Estarrona (Universidad del País Vasco, Spain)
Aritz Farwell (Universidad del País Vasco, Spain)
Joaquim Ferreira da Silva (Universidade Nova de Lisboa, Portugal)
Marcos Garcia (Universidade de Santiago de Compostela, Spain)
José Manuel Gómez Pérez (Expert.ai, Spain)
José ángel González, (Symanto Research, Spain/Germany)
María Grandury (EPFL, Switzerland)
Inma Hernaez (Universidad del País Vasco, Spain)
Elena Lloret (Universidad de Alicante, Spain)
Manuel Mager (Johannes Gutenberg University of Mainz, Germany)
María Teresa Martín Valdivia (Universidad de Jaén, Spain)
Paloma Martínez Fernández (Universidad Carlos III de Madrid, Spain)
Arturo Montejo (Universidad de Jaén, Spain)
Andrés Montoyo Guijarro (Universidad de Alicante, Spain)
Renato Moraes Silva (Universidade de São Paulo, Brazil)
Maite Oronoz (Universidad del País Vasco, Spain)
Manuel Palomar Sanz (Universidad de Alicante, Spain)
Thiago Pardo (Universidade de São Paulo, Brazil)
Paulo Quaresma (Universidade de Évora, Portugal)
Livy Real (Universidade de São Paulo, Brazil)
Aiala Rosá (Universidad de la República, Uruguay)
Robiert Sepúlveda Torres (Universidad de Alicante, Spain)
Susana Sotelo (Universidade de Santiago de Compostela, Spain)
Luis Alfonso Ureña López (Universidad de Jaén, Spain)
Rafael Valencia García (Universidad de Murcia, Spain)

Table of Contents

<i>Bridging Ibero and Latin American NLP communities</i> Eugenio Martínez Cámara, Luis Chiruzzo, Pablo Gamallo, Rafael Muñoz Guillena and German Rigau	1
<i>An Oral-first Interactive Agentic System for Guaraní Speakers</i> Samantha Adorno, Akshata Kishore Moharir and Ratna Kandala	9
<i>AI-TraLow: AI-Driven Translation for Low-Resource Languages and Cultures</i> Antoni Oliver, Maite Melero, Felipe Sanchez-Martinez and Víctor M. Sánchez-Cartagena	15
<i>OpenCor: Latin American and Iberian Languages Open Corpora Forum</i> Livy Real and Valeria de Paiva	21
<i>MedicaLLM: LLM-Driven Speech and Language Solutions for Healthcare</i> Ronghao Pan, Pedro José Vivancos-Vicente, Juan Salvador Castejón-Garrido, Tomás Bernal-Beltrán and Rafael Valencia-García	29
<i>mCS-LM: Multimodal Customer Service and Incident Management Systems Based on Large Language Models</i> Carlos Díaz-Morales, Marcos Checa-Rubio, Tomás Bernal-Beltrán, Ronghao Pan, David Barbáchano, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde and Rafael Valencia-García	38
<i>SAFEWORDS: un marco reproducible para anonimización conforme al RGPD y evaluación de generación en lenguas cooficiales</i> Rafael Muñoz Guillena, Manuel Palomar, Elena Lloret and Nuria Fernández	46
<i>Exploration of Sentence Representations in Spanish BERT-like Models</i> Gonzalo Herrera, Aiala Rosá and Luis Chiruzzo	55

Networking Symposium Program

9:00–9:30	Welcome and Opening Remarks
9:30–10:30	Panel Session I - About the LANLP supporting organizations
9:30–10:30	<i>Bridging Ibero and Latin American NLP communities</i> Eugenio Martínez Cámara, Luis Chiruzzo, Pablo Gamallo, Rafael Muñoz Guillena and German Rigau
11:00–12:00	<i>Keynote: Opportunities & Challenges of LLMs for Ibero-Latin Languages</i> Antonio Branco
12:00–13:00	Paper Session I
12:00–12:20	<i>An Oral-first Interactive Agentic System for Guaraní Speakers</i> Samantha Adorno, Akshata Kishore Moharir and Ratna Kandala
12:20–12:40	<i>AI-TraLow: AI-Driven Translation for Low-Resource Languages and Cultures</i> Antoni Oliver, Maite Melero, Felipe Sanchez-Martinez and Víctor M. Sánchez-Cartagena
12:40–13:00	<i>OpenCor: Latin American and Iberian Languages Open Corpora Forum</i> Livy Real and Valeria de Paiva
14:00–15:20	Paper Session II
14:00–14:20	<i>MedicalLM: LLM-Driven Speech and Language Solutions for Healthcare</i> Ronghao Pan, Pedro José Vivancos-Vicente, Juan Salvador Castejón-Garrido, Tomás Bernal-Beltrán and Rafael Valencia-Garcia
14:20–14:40	<i>mCS-LM: Multimodal Customer Service and Incident Management Systems Based on Large Language Models</i> Carlos Díaz-Morales, Marcos Checa-Rubio, Tomás Bernal-Beltrán, Ronghao Pan, David Barbáchano, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde and Rafael Valencia-Garcia
14:40–15:00	<i>SAFEGWORDS: un marco reproducible para anonimización conforme al RGPD y evaluación de generación en lenguas cooficiales</i> Rafael Muñoz Guillena, Manuel Palomar, Elena Lloret and Nuria Fernández
15:00–15:20	<i>Exploration of Sentence Representations in Spanish BERT-like Models</i> Gonzalo Herrera, Aiala Rosá and Luis Chiruzzo

15:20–16:00 *Panel Session II: Strategic Network RutaMdL*

16:30–17:15 *Hands-on LANLP*

17:15–17:45 *Group Reports & Discussion*

17:45–18:00 *Concluding Remarks & Next Steps*

