

LowResNLP 2025

**Proceedings
of the 1st Workshop
on Advancing NLP for Low-Resource Languages**

associated with
**The 15th International Conference on
Recent Advances in Natural Language Processing
RANLP'2025**

Edited by Ernesto Luis Estevanell-Valladares, Alicia Picazo-Izquierdo, Tharindu Ranasinghe,
Besik Mikaberidze, Simon Ostermann, Daniil Gurgurov, Philipp Müller,
Kurt Micallef, Claudia Borg, Michal Gregor and Marián Šimko

13 September, 2025
Varna, Bulgaria

The 1st Workshop
on Advancing NLP for Low-Resource Languages
Associated with the International Conference
Recent Advances in Natural Language Processing
RANLP'2025

PROCEEDINGS

Varna, Bulgaria
13 September 2025

Online ISBN 978-954-452-100-4

Designed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

LowResNLP is a workshop dedicated to advancing NLP research for low-resource languages, fostering methods, resources, and evaluation practices that address the unique challenges faced by underrepresented languages.

This year we received **24 submissions**, of which **18 were accepted**: 9 papers for **oral + poster presentation** and 9 for **poster-only presentation**. We are excited by the variety and quality of contributions, which highlight new directions in cross-lingual transfer, data augmentation, model efficiency, and evaluation for low-resource settings.

We thank all authors for submitting, and we congratulate the accepted teams. We also express our gratitude to the program committee and our supporters for their valuable efforts and support.

The LowResNLP 2025 Organizers

Organizing Committee:

Ernesto Luis Estevanell-Valladares (University of Alicante, Spain; University of Havana, Cuba)

Alicia Picazo-Izquierdo (University of Alicante, Spain)

Tharindu Ranasinghe (Lancaster University, UK)

Besik Mikaberidze (Muskhelishvili Institute of Computational Mathematics, Georgian Technical University, Georgia)

Simon Ostermann (German Research Center for Artificial Intelligence, Germany)

Daniil Gurgurov (German Research Center for Artificial Intelligence, Germany)

Philipp Müller (German Research Center for Artificial Intelligence, Germany)

Kurt Micallef (University of Malta, Malta)

Claudia Borg (University of Malta, Malta)

Michal Gregor (KINIT, Slovakia)

Marián Šimko (KINIT, Slovakia)

Invited Speakers:

Jesujoba Oluwadara Alabi

Programme Committee:

Nora Aranberri (University of Basque Country)

Sudhansu Bala Das (School of Languages, Literatures & Cultures and Insight SFI Research Centre for Data Analytics, University of Galway, Ireland)

Ana-Maria Bucur (University of Bucharest)

Annie Lee En-Shiun (Ontario Tech University and University of Toronto)

Sofía García González (imaxin software, University of the Basque Country)

Albert Gatt (Utrecht University)

Teresa Lynn (Mohamed bin Zayed University of Artificial Intelligence)

Basab Nath (Assam University)

Patrizia Paggio (University of Malta)

Dhrubajyoti Pathak (National Forensic Sciences University)

Fabian Schmidt (University of Würzburg)

Marijn Schraagen (Utrecht University)

A. Seza Doğruöz (University of Ghent)

Marc Tanti (University of Malta)

Sunita Warjri (University of South Bohemia)

Table of Contents

<i>Bridging the Gap: Leveraging Cherokee to Improve Language Identification for Endangered Iroquoian Languages</i>	
Liam Enzo Eggleston, Michael P. Cacioli, Jatin Sarabu, Ivory Yang and Kevin Zhu	1
<i>Building a Lightweight Classifier to Distinguish Closely Related Language Varieties with Limited Supervision: The Case of Catalan vs Valencian</i>	
Raúl García-Cerdá, María Miró Maestre and Miquel Canal	7
<i>A thresholding method for Improving translation Quality for Indic MT task</i>	
Sudhansu Bala Das, Leo Raphael Rodrigues, Tapas Kumar Mishra and Bidyut Ku Patra	12
<i>A Multi-Task Learning Approach to Dialectal Arabic Identification and Translation to Modern Standard Arabic</i>	
Abdullah Khered, Youcef Benkhedda and Riza Batista-Navarro	21
<i>Low-Resource Machine Translation for Moroccan Arabic</i>	
Alexei Rosca, Abderrahmane Issam and Gerasimos Spanakis	32
<i>Efficient Architectures For Low-Resource Machine Translation</i>	
Edoardo Signoroni, Pavel Rychly and Ruggero Signoroni	39
<i>IfGPT: A Dataset in Bulgarian for Large Language Models</i>	
Svetla Peneva Koeva, Ivelina Stoyanova and Jordan Konstantinov Krlev	65
<i>Modular Training of Deep Neural Networks for Text Classification in Guarani</i>	
Jose Luis Vazquez, Carlos Ulises Valdez, Marvin Matías Agüero-Torales, Julio César Mello-Román, Jose Domingo Colbes and Sebastian Alberto Grillo	76
<i>Roman Urdu as a Low-Resource Language: Building the First IR Dataset and Baseline</i>	
Muhammad Umer Tariq Butt, Stalin Varanasi and Guenter Neumann	82
<i>The Brittle Compass: Navigating LLM Prompt Sensitivity in Slovak Migration Media Discourse</i>	
Jaroslav Kopčan, Samuel Harvan and Marek Suppa	88
<i>Explicit Edge Length Coding to Improve Long Sentence Parsing Performance</i>	
Khensa Daoudi, Mathieu Dehouck, Rayan Ziane and Natasha Romanova	102
<i>Evaluating LLM Capabilities in Low-Resource Contexts: A Case Study of Persian Linguistic and Cultural Tasks</i>	
Jasmin Heierli, Rebecca Bahar Ganjineh and Elena Gavagnin	111
<i>A Benchmark for Evaluating Logical Reasoning in Georgian For Large Language Models</i>	
Irakli Koberidze, Archil Elizbarashvili and Magda Tsintsadze	121
<i>Slur and Emoji Aware Models for Hate and Sentiment Detection in Roman Urdu Transgender Discourse</i>	
Muhammad Owais Raza, Aqsa Umar and Mehrub Awan	131
<i>Automatic Fact-checking in English and Telugu</i>	
Ravi Kiran Chikkala, Tatiana Anikina, Natalia Skachkova, Ivan Vykopal, Rodrigo Agerri and Josef van Genabith	140

<i>Synthetic Voice Data for Automatic Speech Recognition in African Languages</i>	
Brian DeRenzi, Anna Dixon, Mohamed Aymane Farhi and Christian Resch	152
<i>ADOR: Dataset for Arabic Dialects in Hotel Reviews: A Human Benchmark for Sentiment Analysis</i>	
Maram I. Alharbi, Saad Ezzini, Hansi Hettiarachchi, Tharindu Ranasinghe and Ruslan Mitkov	187
<i>Towards Creating a Bulgarian Readability Index</i>	
Dimitar Kazakov, Stefan Minkov, Ruslana Margova, Irina Temnikova and Ivo Emauilov	192