LoResLM 2026

The Second Workshop on Language Models for Low-Resource Languages

Proceedings of the Workshop

March 29, 2026

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

In Collaboration With

©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 979-8-89176-377-7

Preface

We are pleased to present the proceedings of the second Workshop on Language Models for Low-Resource Languages (LoResLM 2026), co-located at the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026) in Rabat, Morocco.

Natural language processing (NLP) has experienced rapid progress in recent years, driven largely by advances in neural language models, including transformer-based architectures and large language models, which have achieved state-of-the-art performance across a wide range of tasks with diverse emerging capabilities. Despite these advances, since the effectiveness of language models (LMs) is primarily determined by the characteristics of their pre-trained language corpora, these models tend to be more focused on high-resource languages. As a result, many models often struggle to adequately support low-resource languages, which are estimated to be around 7,000. Despite their worldwide usage, these languages generally receive little research attention and lack sufficient digital data and resources to support NLP tasks. Recognising the growing need to address this imbalance and promote more inclusive language technologies, the research community has increasingly focused on developing and adapting LMs for low-resource languages. In response to this momentum, we initiated LoResLM to provide a dedicated forum for researchers to share insights, resources, and emerging work in this area. Following a successful inaugural edition, LoResLM returned this year for its second iteration, continuing to foster collaboration and advance research toward more equitable and linguistically diverse NLP.

Primarily focusing on developing and evaluating neural language models for low-resource languages, LoResLM 2026 invited submissions on a broad range of topics, including creating corpora, developing benchmarks, building or adapting LMs, and exploring LM applications for low-resource languages. In total, we received 79 submissions. Among these, we accepted 55 papers, including 44 long papers and 11 short papers, to appear in the workshop proceedings following the peer-review process.

The accepted papers reflect substantial linguistic and research diversity, covering low-resource languages from 13 distinct language families. The majority representation is from the Indo-European family, with contributions across nine of its branches. Overall, the studies address 82 distinct languages, highlighting the breadth of linguistic contexts explored within the workshop. In addition to this linguistic diversity, the papers span 11 research areas, with NLP and LLM Applications, Language Modelling, and Machine Translation and Translation Aids emerging as the most represented topics. We are pleased to present such a broad range of contributions, reflecting the growing momentum of research and offering promising directions for future work on low-resource languages.

LoResLM 2026 would not be successful without several wonderful people who joined this initiative. First of all, we would like to thank the authors who submitted their work to the workshop, encouraging research in many low-resource languages that span diverse research areas. We are very grateful for the programme committee members who played a crucial role towards this workshop's success with their timely engagement with the review process, providing constructive feedback to help authors improve the quality of their papers to meet the general standards. We are also particularly thankful to Prof Barbara Plank for accepting our invitation to serve as the keynote speaker, sharing her knowledge and experience, and providing valuable insights to the NLP community. Our sincere appreciation also goes to the Artificial Intelligence Journal (AIJ) for sponsoring the workshop. We are very grateful to everybody for supporting us to make LoResLM 2026 successful.

Hansi Hettiarachchi, Tharindu Ranasinghe, Alistair Plum, Paul Rayson, Ruslan Mitkov, Mohamed Gaber, Damith Premasiri, Fiona Anting Tan, and Lasitha Uyangodage
(LoResLM 2026 Organisers)
https://loreslm.github.io/home

Organizing Committee

Advisors

Paul Rayson, Lancaster University, UK
Ruslan Mitkov, Lancaster University, UK
Mohamed Gaber, Queensland University of Technology, Australia

Organising Committee

Hansi Hettiarachchi, Lancaster University, UK
Tharindu Ranasinghe, Lancaster University, UK
Alistair Plum, University of Luxembourg, Luxembourg
Damith Premasiri, Lancaster University, UK
Fiona Anting Tan, National University of Singapore, Singapore
Lasitha Uyangodage, University of Münster, Germany

Program Committee

Program Chairs

Hansi Hettiarachchi, Lancaster University
Alistair Plum, University of Luxembourg
Tharindu Ranasinghe, Lancaster University

Reviewers

David Ifeoluwa Adelani, Godfred Agyapong, Wafa Aissa, Sergi Alvarez-Vidal

Valerio Basile, Laura Bernardy, Claudia Borg, Pierrette Bouillon, Ivana Brač, Eleftheria Briakou, Ana-Maria Bucur, Necva Bölücü

Tommaso Caselli, Salmane Chafik, Charibeth Ko Cheng, Cagri Coltekin, André Coneglian, Claudiu Creanga

Sudhansu Bala Das, Daryna Dementieva, Sourabh Deoghare, Alphaeus Dmonte, Kaja Dobrovoljc, Bonaventure F. P. Dossou, Ona de Gibert

Ignatius Ezeani

Francisco Javier Lima Florido, Francesca Frontini, Anna Beatriz Dimas Furtado

Voula Giouli

Le An Ha, Gilberta Hadaj, Amal Htait

Sylvain Kahane, Olha Kanishcheva, Diptesh Kanojia, Tracy Holloway King, Kristina Kocijan, Svetla Peneva Koeva, Jaroslav Kopčan

Sheng Li, Constantine Lignos, Veronika Lipp, Irina Lobzhanidze, Cedric Lothritz, Jason S Lucas, Anne-Marie Lutgen

Michał Marcińczuk, Vukosi Marivate, Eugenio Martínez-Cámara, Iván Martínez-Murillo, Pruthwik Mishra, Maria Mitrofan, Abiodun Modupe, Agnieszka Mykowiecka, Simon Münker

Isuri Nanomi Arachchige, Tadashi Nomoto

Fred Philippy, Scott Piao, Alistair Plum, Damith Premasiri

Nishat Raihan, Tharindu Ranasinghe, Paul Rayson, Mariana Romanyshyn

Abdulfattah Safa, Kengatharaiyer Sarveswaran, Guokan Shang, Ravi Shekhar, Archchana Sindhujan, Milena Slavcheva, Rui Sousa-Silva, Ranka Stanković, Ivelina Stoyanova, Sara Stymne, Tianda Sun

Pattabhi R K Rao T, Hristo Tanev, Uthayasanker Thayasivam

Adaku Uchendu

Francielle Vargas, Justin Vasselli, Veronika Vincze, Raúl Vázquez, Daan van Esch

Taro Watanabe, Philip Weber

Zheng Xin Yong

Arturs Znotins

Table of Contents

NE-BERT: A Multilingual Language Model for Nine Northeast Indian Languages
Badal Nyalang 1

Do Tokenizers Fail on Informal Hindi Expressions? Evidence from Static, Downstream, and Robustness Analyses
Manikandan Ravikiran, Tanmay Tiwari, Vibhu Gupta, Rakesh Prakash, Rohit Saluja and Shayan Mohanty 13

Competence Collapse in Code-Mixed Generation: Spectral Evidence and Mechanistic Recovery via Cross-Lingual Activation Steering
Tanushree Ravindra Pratap Yadav 29

When Multilingual Evaluation Assumptions Fail: Tokenization Effects Across Scripts
Manodyna K H and Luc De Nardi 41

Making Large Language Models Speak Tulu: Structured Prompting for an Extremely Low-Resource Language
Prathamesh Devadiga and Paras Chopra 50

To make someone do something: mining alert-style directives in Bulgarian social media for low-resource language modelling
Ruslana Margova and Stanislav Penkov 62

BanglaLlama: LLaMA for Bangla Language
Abdullah Khan Zehady, Shubhashis Roy Dipta, Naymul Islam, Safi Al Mamun and Santu Karmaker 73

Evaluating Retrieval-Augmented Generation for Medication Question Answering on Nigerian Drug Labels in Yorùbá
Zainab Tairu and Aramide Adebesin 90

Grammatical Error Correction for Low-Resource Languages: The Case of Zarma
Mamadou K. Keita, Marcos Zampieri, Christopher M Homan, Adwoa Asantewaa Bremang, Dennis Asamoah Owusu and Huy Le 98

Quantifying Cross-Lingual Interference: Algorithmic Standardization of Kamtapuri in Large Language Models
Roumak Das 110

SinhaLegal: A Benchmark Corpus for Information Extraction and Analysis in Sinhala Legislative Texts
Minduli Lasandi and Nevidu Jayatilleke 114

BanglaIPA: Towards Robust Text-to-IPA Transcription with Contextual Rewriting in Bengali
Jakir Hasan, Shrestha Datta, Md Saiful Islam, Shubhashis Roy Dipta and Ameya Debnath . . . 132

Improving Romanian LLM Pretraining Data using Diversity and Quality Filtering
Vlad-Andrei Negoiță, Mihai Masala and Traian Rebedea 140

Tone in Yoruba ASR: Evaluating the Impact of Tone Recognition on Transformer-Based ASR Models
Joy Olusanya 149

LLM-as-a-Judge for Low-Resource Languages: Adapting Ragas and Comparative Ranking for Romanian
Claudiu Creanga and Liviu P Dinu 157

QARI: Neural Architecture for Urdu Extractive Machine Reading Comprehension
Samreen Kazi and Shakeel Ahmed Khoja 168

When LLMs Annotate: Reliability Challenges in Low-Resource NLI
Solmaz Panahi, John Kelleher and Vasudevan Nedumpozhimana 178

Qomhrá: A Bilingual Irish and English Large Language Model
Joseph McInerney, Khanh-Tung Tran, Liam Lonergan, Neasa Ní Chiaráin, Ailbhe Ni Chasaide and Barry Devereux 189

Anchoring the Judge: Curriculum-Based Adaptation and Reference-Anchored MQM for LLM-Based Machine Translation of an Unseen Low-Resource Language - A Case of Nupe
Umar Baba Umar, Sulaimon Adebayo Bashir and Abdulmalik Danlami Mohammed 200

TeluguEval: A Comprehensive Benchmark for Evaluating LLM Capabilities in Telugu
Revanth Kumar Gundam and Radhika Mamidi 212

Cross-Lingual Emotion Recognition in Balinese Text using Multilingual-LLMs under Peer-Collaborations Settings
Putu Kussa Laksana Utama, Tsegaye Misikir Tashu and Jilles Steeve Dibangoye 225

Enabling Structured Reasoning in Sindhi with Culturally Grounded Instruction Tuning
Mehak Mehak, Kamyar Zeinalipour, Pireh Soomro, Cristiano Chesi, Marco Gori and Marco Maggini 239

Targeted Syntactic Evaluation of Language Models on Georgian Case Alignment
Daniel Gallagher and Gerhard Heyer 259

So, How Much Do LLMs Hallucinate on Low-Resource Languages?A Quantitative and Qualitative Analysis
Kushal Trivedi, Murtuza Shaikh and Sriyansh Sharma 271

Learning from Scarcity: Building and Benchmarking Speech Technology for Sukuma.
Macton Mgonzo, Kezia Oketch, Naome A Etori, Winnie Mang'eni, Elizabeth Fabian Nyaki and Michael Samwel Mollel 288

We Are (Language) Family": Adapting Transformer models to related minority languages with linguistic data
Miguel López-Otal and Jorge Gracia 297

A Comprehensive Evaluation of Chain-of-Thought Faithfulness in Persian Classification Tasks
Shakib Yazdani, Cristina España-Bonet, Eleftherios Avramidis, Yasser Hamidullah and Josef Van Genabith 311

Under-resourced studies of under-resourced languages: lemmatization and POS-tagging with LLM annotators for historical Armenian, Georgian, Greek and Syriac
Chahan Vidal-Gorène, Bastien Kindt and Florian Cafiero 324

Large Language Models for Mental Health: A Multilingual Evaluation
Nishat Raihan, Sadiya Sayara Chowdhury Puspo, Ana-Maria Bucur, Stevie Chancellor and Marcos Zampieri 335

Serbian SuperGLUE: Towards an Evaluation Benchmark for South Slavic Language Models
Mitar Perovic and Teodora Mihajlov 347

Less is More: Adapting Text Embeddings for Low-Resource Languages with Small Scale Noisy Synthetic Data
Zaruhi Navasardyan, bagratuni@metric.am bagratuni@metric.am, Spartak Bughdaryan and Hrant Davtyan 362

Evaluating Large Language Models on Lithuanian Grammatical Cases
Urtė Jakubauskaitė and Raquel G. Alhama 371

Tracking the evolution of LLM capabilities for Belarusian with OpenAI Evals
Vladislav Poritski, Oksana Volchek, Maksim Aparovich, Volha Harytskaya and Pavel Smrz . 378

Beyond Many-Shot Translation: Scaling In-Context Demonstrations For Low-Resource Machine Translation
Luis Frentzen Salim, Esteban Carlin, Alexandre Morinvil, Xi Ai and Lun-Wei Ku 388

Bootstrapping Embeddings for Low Resource Languages
Merve Basoz, Andrew Horne and Mattia Opper 408

The Indonesian Religiolect Corpus: Data Curation for Muslim, Protestant, and Catholic Language Varieties
Dan Sachs 426

Representation-Aware Prompting for Zero-Shot Marathi Text Classification: IPA, Romanization, Repetition
Van-Hien Tran, Huy Hien Vu, Hideki Tanaka and Masao Utiyama 436

MaiBERT: A Pre-training Corpus and Language Model for Low-Resourced Maithili Language
Sumit Yadav, Raju Kumar Yadav, Utsav Maskey, Gautam Siddharth Kashyap, Ganesh Gautam and Usman Naseem 444

KyrText: A Multi-Domain Large-Scale Corpus for Kyrgyz Language
Tilek Chubakov 453

Pretraining and Benchmarking Modern Encoders for Latvian
Arturs Znotins 461

Escaping the Probability Trap: Mitigating Semantic Drift in Cantonese-Mandarin Translation
Yuzhi Liang and Fangqi Chen 471

How multilingual are multilingual LLMs? A case study in Northern Sámi-Finnish Translation
Jonne Sälevä and Constantine Lignos 484

Tokenization and Morphological Fidelity in Uralic NLP: A Cross-Lingual Evaluation
Nuo Xu and Ahrii Kim 493

Hebrew Diacritics Restoration using Visual Representation
Yair Elboher and Yuval Pinter 504

Out-Of-Tune rather than Fine-Tuned: How Pre-training, Fine-tuning and Tokenization Affect Semantic Similarity in a Historical, Non-Standardized Domain
Stella Verkijk and Piek Vossen 515

LuxDiagRC: A Diagnostic Reading Comprehension Corpus for Luxembourgish with Linguistic and Cognitive Annotation Layers
Christophe Friezas Gonçalves, Salima Lamsiyah and Christoph Schommer 532

Cross-Lingual and Cross-Domain Transfer Learning for POS Tagging in Historical Germanic Low-Resource Languages
Irene Miani, Sara Stymne and Gregory R. Darwin 542

MTQE.en-he: Machine Translation Quality Estimation for English-Hebrew
Andy Rosenbaum, Assaf Siani and Ilan Kernerman 559

Tokenization Cost, Retention, and Orthography Robustness for Ladin and Italian Varieties
Alessio Staffini 570

UrHiOdSynth: A Multilingual Synthetic Corpus for Speech-to-Speech Translation in Low-Resource Indic Languages
Jamaluddin, Subhankar Panda, Aditya Narendra, Kamanksha Prasad Dubey and Mohammad Nadeem 584

BanglaSummEval: Reference-Free Factual Consistency Evaluation for Bangla Summarization
Ahmed Rafid, Rumman Adib, Fariya Ahmed, Ajwad Abrar and Mohammed Saidul Islam . . . 595

Parameter-Efficient Quality Estimation via Frozen Recursive Models
Umar Abubacar, Roman Bauer and Diptesh Kanojia 609

Neural Machine Translation for French–Mooré: Adapting Large Language Models to Low-Resource Languages
walkercompaore972@gmail.com walkercompaore972@gmail.com, Maimouna Ouattara, Rodrique Kafando, Tegawendé F. Bissyandé, Abdoul Kader Kabore and aminata.sabane@ujkz.bf aminata.sabane@ujkz.bf 615

Contributing to Speech-to-Speech Translation for African Low-Resource Languages : Study of French-Mooré Pair
Fayçal S. A. Ouedraogo, Maimouna Ouattara, Rodrique Kafando, Abdoul Kader Kabore, aminata.sabane@ujkz.bf aminata.sabane@ujkz.bf and Tegawendé F. Bissyandé 623

Domain-Specific Quality Estimation for Machine Translation in Low-Resource Scenarios
Namrata Bhalchandra Patil Gurav, Akashdeep Ranu, Archchana Sindhujan and Diptesh Kanojia 630

Overview of the Second Workshop on Language Models for Low-Resource Languages (LoResLM 2026)
Hansi Hettiarachchi, Tharindu Ranasinghe, Alistair Plum, Paul Rayson, Ruslan Mitkov, Mohamed Medhat Gaber, Damith Premasiri, Fiona Anting Tan and Lasitha Uyangodage 651

Program

Sunday, March 29, 2026

09:30 - 10:30 *Poster Session I - Linguistic Analysis and Language Structure*

When Multilingual Evaluation Assumptions Fail: Tokenization Effects Across Scripts
Manodyna K H and Luc De Nardi

Learning from Scarcity: Building and Benchmarking Speech Technology for Sukuma.
Macton Mgonzo, Kezia Oketch, Naome A Etori, Winnie Mang'eni, Elizabeth Fabian Nyaki and Michael Samwel Mollel

Evaluating Large Language Models on Lithuanian Grammatical Cases
Urtė Jakubauskaitė and Raquel G. Alhama

The Indonesian Religiolect Corpus: Data Curation for Muslim, Protestant, and Catholic Language Varieties
Dan Sachs

Tokenization and Morphological Fidelity in Uralic NLP: A Cross-Lingual Evaluation
Nuo Xu and Ahrii Kim

Out-Of-Tune rather than Fine-Tuned: How Pre-training, Fine-tuning and Tokenization Affect Semantic Similarity in a Historical, Non-Standardized Domain
Stella Verkijk and Piek Vossen

LuxDiagRC: A Diagnostic Reading Comprehension Corpus for Luxembourgish with Linguistic and Cognitive Annotation Layers
Christophe Friezas Gonçalves, Salima Lamsiyah and Christoph Schommer

11:00 - 12:00 *Poster Session II - Language Modelling and NLP Applications*

Competence Collapse in Code-Mixed Generation: Spectral Evidence and Mechanistic Recovery via Cross-Lingual Activation Steering
Tanushree Ravindra Pratap Yadav

To make someone do something: mining alert-style directives in Bulgarian social media for low-resource language modelling
Ruslana Margova and Stanislav Penkov

Improving Romanian LLM Pretraining Data using Diversity and Quality Filtering
Vlad-Andrei Negoiță, Mihai Masala and Traian Rebedea

LLM-as-a-Judge for Low-Resource Languages: Adapting Ragas and Comparative Ranking for Romanian
Claudiu Creanga and Liviu P Dinu

When LLMs Annotate: Reliability Challenges in Low-Resource NLI
Solmaz Panahi, John Kelleher and Vasudevan Nedumpozhimana

TeluguEval: A Comprehensive Benchmark for Evaluating LLM Capabilities in Telugu
Revanth Kumar Gundam and Radhika Mamidi

Cross-Lingual Emotion Recognition in Balinese Text using Multilingual-LLMs under Peer-Collaborations Settings
Putu Kussa Laksana Utama, Tsegaye Misikir Tashu and Jilles Steeve Dibangoye

We Are (Language) Family": Adapting Transformer models to related minority languages with linguistic data
Miguel López-Otal and Jorge Gracia

Tracking the evolution of LLM capabilities for Belarusian with OpenAI Evals
Vladislav Poritski, Oksana Volchek, Maksim Aparovich, Volha Harytskaya and Pavel Smrz

Representation-Aware Prompting for Zero-Shot Marathi Text Classification: IPA, Romanization, Repetition
Van-Hien Tran, Huy Hien Vu, Hideki Tanaka and Masao Utiyama

MaiBERT: A Pre-training Corpus and Language Model for Low-Resourced Maithili Language
Sumit Yadav, Raju Kumar Yadav, Utsav Maskey, Gautam Siddharth Kashyap, Ganesh Gautam and Usman Naseem

KyrText: A Multi-Domain Large-Scale Corpus for Kyrgyz Language
Tilek Chubakov

Pretraining and Benchmarking Modern Encoders for Latvian
Arturs Znotins

UrHiOdSynth: A Multilingual Synthetic Corpus for Speech-to-Speech Translation in Low-Resource Indic Languages
Jamaluddin, Subhankar Panda, Aditya Narendra, Kamanksha Prasad Dubey and Mohammad Nadeem

Parameter-Efficient Quality Estimation via Frozen Recursive Models
Umar Abubacar, Roman Bauer and Diptesh Kanojia

Neural Machine Translation for French–Mooré: Adapting Large Language Models to Low-Resource Languages
walkercompaore972@gmail.com walkercompaore972@gmail.com, Maimouna Ouattara, Rodrique Kafando, Tegawendé F. Bissyandé, Abdoul Kader Kabore and aminata.sabane@ujkz.bf aminata.sabane@ujkz.bf

Contributing to Speech-to-Speech Translation for African Low-Resource Languages : Study of French-Mooré Pair
Fayçal S. A. Ouedraogo, Maimouna Ouattara, Rodrique Kafando, Abdoul Kader Kabore, aminata.sabane@ujkz.bf aminata.sabane@ujkz.bf and Tegawendé F. Bissyandé

09:30 - 12:00 *Virtual Poster Session*

NE-BERT: A Multilingual Language Model for Nine Northeast Indian Languages
Badal Nyalang

Do Tokenizers Fail on Informal Hindi Expressions? Evidence from Static, Downstream, and Robustness Analyses
Manikandan Ravikiran, Tanmay Tiwari, Vibhu Gupta, Rakesh Prakash, Rohit Saluja and Shayan Mohanty

Making Large Language Models Speak Tulu: Structured Prompting for an Extremely Low-Resource Language
Prathamesh Devadiga and Paras Chopra

BanglaLlama: LLaMA for Bangla Language
Abdullah Khan Zehady, Shubhashis Roy Dipta, Naymul Islam, Safi Al Mamun and Santu Karmaker

Evaluating Retrieval-Augmented Generation for Medication Question Answering on Nigerian Drug Labels in Yorùbá
Zainab Tairu and Aramide Adebesin

Quantifying Cross-Lingual Interference: Algorithmic Standardization of Kamtapuri in Large Language Models
Roumak Das

SinhaLegal: A Benchmark Corpus for Information Extraction and Analysis in Sinhala Legislative Texts
Minduli Lasandi and Nevidu Jayatilleke

BanglaIPA: Towards Robust Text-to-IPA Transcription with Contextual Rewriting in Bengali
Jakir Hasan, Shrestha Datta, Md Saiful Islam, Shubhashis Roy Dipta and Ameya Debnath

Tone in Yoruba ASR: Evaluating the Impact of Tone Recognition on Transformer-Based ASR Models
Joy Olusanya

QARI: Neural Architecture for Urdu Extractive Machine Reading Comprehension
Samreen Kazi and Shakeel Ahmed Khoja

Anchoring the Judge: Curriculum-Based Adaptation and Reference-Anchored MQM for LLM-Based Machine Translation of an Unseen Low-Resource Language - A Case of Nupe
Umar Baba Umar, Sulaimon Adebayo Bashir and Abdulmalik Danlami Mohammed

So, How Much Do LLMs Hallucinate on Low-Resource Languages?A Quantitative and Qualitative Analysis
Kushal Trivedi, Murtuza Shaikh and Sriyansh Sharma

A Comprehensive Evaluation of Chain-of-Thought Faithfulness in Persian Classification Tasks
Shakib Yazdani, Cristina España-Bonet, Eleftherios Avramidis, Yasser Hamidullah and Josef Van Genabith

Under-resourced studies of under-resourced languages: lemmatization and POS-tagging with LLM annotators for historical Armenian, Georgian, Greek and Syriac
Chahan Vidal-Gorène, Bastien Kindt and Florian Cafiero

Large Language Models for Mental Health: A Multilingual Evaluation
Nishat Raihan, Sadiya Sayara Chowdhury Puspo, Ana-Maria Bucur, Stevie Chancellor and Marcos Zampieri

Serbian SuperGLUE: Towards an Evaluation Benchmark for South Slavic Language Models
Mitar Perovic and Teodora Mihajlov

Less is More: Adapting Text Embeddings for Low-Resource Languages with Small Scale Noisy Synthetic Data
Zaruhi Navasardyan, bagratuni@metric.am bagratuni@metric.am, Spartak Bughdaryan and Hrant Davtyan

Beyond Many-Shot Translation: Scaling In-Context Demonstrations For Low-Resource Machine Translation
Luis Frentzen Salim, Esteban Carlin, Alexandre Morinvil, Xi Ai and Lun-Wei Ku

Sunday, March 29, 2026 (continued)

Escaping the Probability Trap: Mitigating Semantic Drift in Cantonese-Mandarin Translation
Yuzhi Liang and Fangqi Chen

How multilingual are multilingual LLMs? A case study in Northern Sámi-Finnish Translation
Jonne Sälevä and Constantine Lignos

Hebrew Diacritics Restoration using Visual Representation
Yair Elboher and Yuval Pinter

MTQE.en-he: Machine Translation Quality Estimation for English-Hebrew
Andy Rosenbaum, Assaf Siani and Ilan Kernerman

Tokenization Cost, Retention, and Orthography Robustness for Ladin and Italian Varieties
Alessio Staffini

BanglaSummEval: Reference-Free Factual Consistency Evaluation for Bangla Summarization
Ahmed Rafid, Rumman Adib, Fariya Ahmed, Ajwad Abrar and Mohammed Saidul Islam

Domain-Specific Quality Estimation for Machine Translation in Low-Resource Scenarios
Namrata Bhalchandra Patil Gurav, Akashdeep Ranu, Archchana Sindhujan and Diptesh Kanojia

12:00 - 14:00 *Lunch Break*

14:00 - 15:00 *Introduction and Keynote*

15:00 - 15:30 *Oral Session I - Syntactic Analysis in the era of Language Models*

Cross-Lingual and Cross-Domain Transfer Learning for POS Tagging in Historical Germanic Low-Resource Languages
Irene Miani, Sara Stymne and Gregory R. Darwin

Targeted Syntactic Evaluation of Language Models on Georgian Case Alignment
Daniel Gallagher and Gerhard Heyer

Sunday, March 29, 2026 (continued)

16:00 - 17:00 *Oral Session II - Language Modelling and Applications*

Qomhrá: A Bilingual Irish and English Large Language Model
Joseph McInerney, Khanh-Tung Tran, Liam Lonergan, Neasa Ní Chiaráin, Ailbhe Ni Chasaide and Barry Devereux

Bootstrapping Embeddings for Low Resource Languages
Merve Basoz, Andrew Horne and Mattia Opper

Enabling Structured Reasoning in Sindhi with Culturally Grounded Instruction Tuning
Mehak Mehak, Kamyar Zeinalipour, Pireh Soomro, Cristiano Chesi, Marco Gori and Marco Maggini

Grammatical Error Correction for Low-Resource Languages: The Case of Zarma
Mamadou K. Keita, Marcos Zampieri, Christopher M Homan, Adwoa Asantewaa Bremang, Dennis Asamoah Owusu and Huy Le

17:00 - 17:15 *Overview*

Overview of the Second Workshop on Language Models for Low-Resource Languages (LoResLM 2026)
Hansi Hettiarachchi, Tharindu Ranasinghe, Alistair Plum, Paul Rayson, Ruslan Mitkov, Mohamed Medhat Gaber, Damith Premasiri, Fiona Anting Tan and Lasitha Uyangodage

17:15 - 17:30 *Awards and Closing*