

MRL 2026

**The 6th Workshop on Multilingual Representation Learning
(MRL 2026)**

Proceedings of the Workshop

October 28, 2026

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

Capital One



©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 979-8-89176-469-9

Organizing Committee

Workshop Organizers

David Ifeoluwa Adelani
Jesujoba Alabi
Catherine Arnett
Tyler Chang
C.M. Downey
Omer Goldman
Senyu Li
Lester James Validad Miranda
Jirui Qi
David Stap
Genta Indra Winata
Xinyu Crystina Zhang

Program Committee

Program Chairs

David Ifeoluwa Adelani
Jesujoba Alabi
Catherine Arnett
Tyler Chang
C.M. Downey
Omer Goldman
Senyu Li
Lester James Validad Miranda
Jirui Qi
David Stap
Genta Indra Winata
Xinyu Crystina Zhang

Reviewers

Naman Ahuja, Victor Olalekan Akinode, Maro Akpobi, Mehdi Ali, Pulkit Arya, Joseph Attieh, Muhammad Falensi Azmi

Kartik Balasubramaniam, Travis M. Bartley, Ali Basirat, Vishal Bhalla, Terra Blevins, Philipp Borchert, Laurie Burchell

Jiale Chen, Luis Chiruzzo, Olutosoye Taiwo Christian

Paresh Dashore, Dan DeGenaro

Catherine Nana Nyaah Essuman

Malik Hammad Faisal, Aloka Fernando

Félix Gaschi, Somraj Gautam, Jiahui Ge, Gideon George, Luke Gessler, Faeze Ghorbanpour, Javier García Gilabert, Nora Graichen, Ximena Gutierrez

Asfaw Gedamu Haileslasie, Coleman Haley, Pingjun Hong, Marisa Hudspeth, Kathy Hämmerl

Gaganpreet Jhaggi, Jiby Mariya Jose, Haeji Jung

Sulaiman Kagumire, Olga Kellert, Jonghwi Kim, Sai Tulasi Kolapudi, Ross Deans Kristensen-McLachlan, Olga Kriukova, Hongzhi Kuai, Shreyas Kulkarni, Artur Kulmizev, Anuj Kumar

Tomasz Limisiewicz, Dianqing Lin, Kaiying Kevin Lin, Danni Liu, Yihong Liu

Anish Mahishi, Evans Moitui Makori, Nimshi Venkat Meripo, Shafie Abdi Mohamed

Usman Nawaz, Idriss Nguepi Nguefack, Ercong Nie, Mitodru Niyogi

Esther Odunayo Oduntan, Toyib Ogunremi, Shu Okabe, Theophilus Lincoln Owiti, Abraham

Toluwase Owodunni

Nilo Pedrazzini, Rhitabrat Pokharel

Osvaldo Quinjica

Falia Ramanantsoa, Vipul Kumar Rathore, Frederick Riemenschneider

Jiu Sha, Avinash Kumar Sharma, Sona Elza Simon, Pratik Rakesh Singh, Pedro Ortiz Suarez, Janet Yunchen Sung, Eduardo Sánchez

Chihiro Taguchi, Wenjia Tan, Akasha Tigalappanavara, Cagri Toraman, Nemika Tyagi

Ana Sabina Uban

Praneetha Vaddamanu, Jannis Vamvas, Sahil Verma

Sahil Walia, Xiao Wang, Zeqiang Wang, Michelle Wastl, Song-Li Wu

Cheng Xu

Ivory Yang, Xiulin Yang, Zekai Ye, Boya Yu

Chen Zhang

Table of Contents

<i>Testing the Assumptions of Active Learning for Translation Tasks with Few Samples</i> Lorenzo Jaime Yu Flores, Cesare Spinoso-Di Piano, Ori Ernst, David Ifeoluwa Adelani and Jackie Chi Kit Cheung	1
<i>NEDOQwen: Diagnosing and Repairing a Turkish-Centric Small Language Model for Multilingual Evaluation</i> Ahmet Rifat Öztürk, Yağız Ekrem Dalar, Nedim Mutlu Sezer, Ömer Faruk Aksoy and Feyzi Arda Salihoğlu	19
<i>One Image, No Tokens: A Controlled Study of Glyph-Based Chinese Language Modeling</i> Shuyang Xiang and Hao Guan	27
<i>Subword-Based Comparative Linguistics across 242 Languages Using Wikipedia Glottosets</i> Iaroslav Chelombitko, Mika Hämmäläinen and Aleksey Komissarov	37
<i>Language Coverage Is Not Variety Coverage: Lexical Underrepresentation of Classical Chinese in Multilingual Tokenizers</i> Lifan Deng	53
<i>More Capable, Less Faithful: A Multilingual Analysis of Mathematical (Un)Solvability Detection in LLMs</i> Maria-Eleni Zoumpoulidi, Nikolaos Xiros and Georgios Paraskevopoulos	58
<i>Leaving the English Basin: A Logit-Lens View of Cross-Lingual Emotion Transfer in Multilingual LLMs</i> Zarina Uvalieva	68
<i>mAceReason-Math: A Dataset of High-quality Multilingual Math Problems Ready for RLVR</i> Konstantin Dobler, Simon Lehnerer, Federico Scozzafava, Jonathan Janke and Mohamed Ali ..	91
<i>Paradigms over Surface Cues: Inflection Class Drives Hindi Gender Assignment in Expert-Annotated Noun Paradigms</i> John Alderete, Yogya Agrawal, Tra My Nguyen, Puja Shah and Aanchan Mohan	102
<i>Why Better Cross-Lingual Alignment Fails for Better Cross-Lingual Transfer: Case of Encoders</i> Yana Veitsman, Yihong Liu and Hinrich Schuetze	119
<i>What Improves Akkadian Machine Translation, and What Only Seems To?</i> Francesco la Rosa, Jaehyuk Choi and King Shun Hanson Siu	135
<i>Wisdom in Unity: The Role of Multilingual Training in Figurative Language Identification in Proverbs</i> Rama Khalid Alomair, Remas Hezam Alsubaie, Walaa Saif Aleslam Mohammed, Rima Khalid Alsonbul, Mona Saleh Alnajjar, Razan Saeed Aldosari, Haya Al-Ibrahim and Abeer Aldayel	151
<i>Cross-Lingual Gaps in Visual STEM Reasoning: Auditing VLMs Across Six Indian Languages</i> Swastik R	167
<i>Testing Language Model Sensitivity to Object-Anaphoric Accessibility: Persian ye(k) and English in-definite this</i> Nikta Yekrang Safakar and Sebastian Padó	183
<i>Loss-Guided Vocabulary Discovery for Multilingual Static Encoders</i> Boris Malashenko, Ivan Jarsky, Sviatoslav A. Stumpf and Valeria Efimova	196

<i>VaaniCLIP: Cross-Lingual Speech-Text Retrieval for Indic Languages</i>	
Vamsi Adari	206
<i>Natural, Synthetic, or Mixed Supervision for Open-Weight Arabic Grammatical Error Correction</i>	
Mohammed Alnuwaiser, Mohammed Albalawi and Thikra Mansoor Alghamdi	221
<i>Geometric Representations of African Languages: A Regional Semantic Hub and Cultural Steering</i>	
Muhammad Abdullahi Said and Jonathan P. Shock	230
<i>Does Safety Live in a Shared Multilingual Subspace? Cross-Lingual (Non-)Transfer of the Refusal Direction</i>	
Ruiyang Zhang	245
<i>Is TunisianMMLU Actually Tunisian? A Native-Speaker Audit of a Machine-Translated Dialect Benchmark</i>	
Fatma Ezzahra Jlali and Marek Šuppa	257
<i>Two Emojis of Difference: What Multilingual Affective Generation Benchmarks Actually Measure</i>	
Fardeen Sadab and Adib Sakhawat	270
<i>State Tracking vs. Attention in a Morphologically Rich Low-Resource Language: A Controlled Bangla Case Study with Mamba-3</i>	
Sahil al Farib, Momota Ahsana Meem, Sheikh Redwanul Islam, Khan Md. Shams Arefin and Azizur Rahman Anik	280
<i>Beyond Aggregate Loss: Training Dynamics and Language-Specific Scaling Laws for Multilingual Language Models</i>	
Renren Jin and Deyi Xiong	289
<i>Evaluating and Debiasing Multilingual Backpack Language Model Representations</i>	
Jennifer Jane Ries and Kavya Venkatesh	300
<i>When Cross-Lingual Consistency Is Not Enough: Diagnostics for Low-Resource Political Actor Portrayal</i>	
Abdul Razaq Vahidi and Sebastian Padó	310
<i>What Do We Talk About When We Talk About Tokenization?</i>	
Gianluca Vico and Jindřich Libovický	320
<i>Where Are Language-Specific Features Represented in Multilingual LLMs? Everywhere</i>	
Mikaela Martins and Jonathan Dunn	334
<i>Typological Features are Multiple-Choice Questions, not Yes/No Questions: Lessons from Typological Feature-Value Embeddings</i>	
Tung-Le Pan and Shu-Kai Hsieh	349
<i>One Meaning, Many Romanizations: Selective Stability of Multilingual Representations under Informal Indic Romanization</i>	
Divyajot Singh	366
<i>When Multilingual Embeddings Collapse: Representation Failure and Metric (Un)reliability for Hausa and Fongbe Machine Translation</i>	
Mahounan Pericles Adjovi, Roald Eiselen and Prasenjit Mitra	377
<i>Multilingual Datasets for Augmentative and Alternative Communication</i>	
Keith Vertanen	386

<i>Whisper-Style Speech Encoders Align Semantically Beyond Phonetic Cues</i>	
Ryan Soh-Eun Shim, Domenico de Cristofaro, Chengzhi Martin Hu, Alessandro Vietti and Barbara Plank	394
<i>Cross-Lingual Transfer and Shared Geometry of Deception Probes in LLMs</i>	
Matheus Prado Miranda, Marcelo Sartori Locatelli, Virgilio Almeida and Wagner Meira Jr.	404
<i>Does CoT-Pass@k Really Check the CoT? A Multilingual Mathematical Audit</i>	
Tarik Tuna Taşaltı, Burcu Hudaverdi and David Semedo	422
<i>Cross-Lingual Transfer of Evil Vectors Is Directionally Asymmetric</i>	
Juann Kim and Kang MinKu	445
<i>Interpretable Text-Based Prosodic Boundary Prediction for Hindi Children’s Stories</i>	
Sakshi Jadhav, Mansi Dhamne and Preeti Rao	455
<i>Impact of Bilingual Corpus on Pretraining Large Language Models: An Analysis for Turkish</i>	
Elif Özge Yılmaz, Melike Uysal and Cagri Toraman	467
<i>RunyaNER: Auxiliary Language Selection for Runyankore NER</i>	
Prosper Arineitwe Asiimwe, Francois Meyer and Jan Buys	480
<i>The Cost of Asking in Turkish: Prompt Language Effects on Code LLMs</i>	
Mert İncidelen, Melike Kaya, Elif İşbilir and Yunus Santur	492
<i>How Does Continued African Adaptation Reshape Multilingual Representations?</i>	
Oluwadarasimi Olowookere-Ajayi and Jesse-Paul Osemeke	497
<i>Pag-refuse: Safety Alignment Transfers Unevenly to Tagalog and Taglish</i>	
Albrave Raphael Albayda and Angelica Anne Araneta Naguio	510
<i>When Whisper Italianizes: Measuring and Mitigating Dialect Substitution Bias in Neapolitan ASR</i>	
Michael Cacioli	519
<i>A False Positive in Cross-Lingual Unlearning Evaluation</i>	
Mohamed Waseem and Mohammed Fawaz	533
<i>Revisiting the Effectiveness of Word-level Tokenization in Language Model Pretraining</i>	
Takahiro Shimizu, Tianqi Wang and Jun Suzuki	542
<i>MultiInpaintCOCO: Fine-Grained Visual Understanding in Multiple Languages</i>	
Andrei-Alexandru Manea, Jindřich Libovický and Philipp J. Rösch	560
<i>Language Distances are Practical for Equitable Cross-Lingual Transfer</i>	
York Hay Ng, Razan Ahsan Rifandi, Aditya Khan and En-Shiun Annie Lee	586
<i>KMMU: Evaluation of Massive Multi-discipline Multimodal Understanding in Korean Language and Context</i>	
Nahyun Lee, Guijin Son, Hyunwoo Ko, Chanyoung Kim, JunYoung An, Kyubeen Han, Youngjae Yu and Il-Youp Kwak	609
<i>UzLitAttr: A Benchmark for Closed-Set Authorship Attribution in Uzbek Literary Texts</i>	
Oybek Eraliev Maripjon Ugli, Azizbek Muhsinov and Jumabek Alikhanov	635
<i>MGhana-ST: A Low-Resource Speech Translation Dataset for Ghanaian Languages and an Analysis of Multilingual Training Trade-offs</i>	
Frank Lawrence Nii Adoquaye Acquaye, Eric George Parakal, Jesse Johnson, Kishankumar Bhimani and Jochebed Afua Basil	653

A Research Compass for the Acquisition of Thousands of Low-Resource Languages by AI Models via Form-Focused Study and Usage-Based Linguistic Immersion

Juan Antonio Pérez-Ortiz 665

PAINT: A Framework for Multilingual Pain Representation from Medical Licensing Exams to Real-World Clinical Texts

Boya Zhang, Yuanchao Ma, Jiageng Wu, Krittapas Chaisutyakorn, Rodrigo Gameiro, Ahran Han, Chien-Yu Huang, Nikhil Jaiswal, Marianna Leite, Miriam Mora-Mau, Jaime Alonso Pineda Moreno, Siena Kathleen V. Placino, Ryohei Yamamoto, Yugang Jia, Jie Yang, Leo Anthony Celi and Douglas Teodoro 684