THE FIRST INTERNATIONAL CONFERENCE

on

LANGUAGE TECHNOLOGIES FOR LOW-RESOURCE LANGUAGES

LaTeLL'2026

P R O C E E D I N G S

Edited by:

Salima Lamsiyah, Tharindu Ranasinghe, Saad Ezzini,
Ernesto Luis Estevanell-Valladares, Ruslan Mitkov

Fes, Morocco
September 28–30, 2026
www.latell.org/2026/

THE FIRST INTERNATIONAL CONFERENCE
ON LANGUAGE TECHNOLOGIES FOR
LOW-RESOURCE LANGUAGES
(LaTeLL'2026)

PROCEEDINGS

28–30 September, 2026
Fes, Morocco
https://latell.org/2026/

© 2026 The authors of the individual papers, for their respective contributions.

The individual papers in this volume are licensed by their authors under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

LaTeLL'2026 is organised by the University of Alicante (Spain), Lancaster University (UK), King Fahd University of Petroleum and Minerals (Saudi Arabia) and the University of Luxembourg, with the local support of Mohammed VI Polytechnic University (Morocco).

The organisers gratefully acknowledge the support of AutoTrust AI, Diamond Sponsor of LaTeLL'2026.

Online ISBN 978-954-452-112-7
e-book site: www.acl-bg.org, www.latell.org/2026/

Designed by INCOMA Ltd.
Shoumen, BULGARIA

Message from the Conference Chair

Welcome to Fes and to LaTeLL'2026, the First International Conference on Language Technologies for Low-resource Languages.

Most of the world's languages are low-resource languages. They account for at least 85% of the world's linguistic diversity and are often spoken by smaller or marginalised communities, yet the language technology that now shapes how people read, write, search and communicate has been built almost entirely for a few dozen languages with abundant data. For the rest, the basic ingredients are missing: large, well-annotated datasets, standardised tools and reliable linguistic resources. Large language models have narrowed some of these gaps and widened others. They perform impressively in the languages they were trained on and much less well in the languages they were not. Closing this gap is a scientific problem, but it is also a question of fairness: the speakers of low-resource languages should not be the last to benefit from language technology.

Several workshops have addressed natural language processing for low-resource languages over the years and have done valuable work. LaTeLL is the first international conference dedicated specifically to the automatic processing of these languages. Our aim in launching it is to give this research community a venue of its own, where work on any low-resource language, from any part of the world, can be presented, compared and built upon. Fes, home to one of the oldest universities in the world and set in a region where Arabic dialects, Amazigh and other languages are spoken side by side, seemed to us the right place to begin.

The papers in this volume cover the whole range of the field: language resources and corpora; core language technologies such as morphology, syntax and machine translation; applications; multimodal and grounded approaches; multilinguality and cross-lingual methods; human-centred approaches; and the ethical and social questions that arise when technology reaches communities it has previously ignored. The programme also features two special tracks, one on building applications based on large language models for low-resource languages and one on Modern Standard Arabic and Arabic dialects.

Many people made this conference possible. We thank the authors who submitted their work and presented it, in Fes or by video, and the members of the Programme Committee, who reviewed every submission carefully and gave the authors the feedback that improved the final versions of their papers. We are grateful to Professor Nizar Habash of New York University Abu Dhabi for accepting our invitation to give the keynote address. The Programme Committee Chairs, Saad Ezzini, Salima Lamsiyah and Tharindu Ranasinghe, put the programme together, and the quality of the papers in this volume owes much to them. We are also grateful to AutoTrust AI, the Diamond Sponsor of LaTeLL'2026, for its support of the conference.

Finally, we thank the members of the Organising Committee, whose work over many months made LaTeLL'2026 happen: Ismail Berrada, our Local Chair, and Radouane Achamlal, who took care of the local arrangements in Fes; Salmane Chafik, Houdaifa Atou, Milica Ikonić Nešić, Ernesto Luis Estevanell-Valladares and Maram Alharbi; and Nikolai Nikolov, our conference administrator.

Welcome to LaTeLL'2026. We hope you enjoy the conference and your stay in Fes.

Ruslan Mitkov
LaTeLL'2026 Conference Chair

Message from the Programme Committee Chairs

This volume contains the papers accepted for presentation at LaTeLL'2026, the First International Conference on Language Technologies for Low-resource Languages, held in Fes, Morocco, on 28–30 September 2026.

This volume brings together the papers accepted for LaTeLL'2026, the first edition of the conference. Every submission was reviewed by three members of the Programme Committee, who gave their time generously and often under tight deadlines, and the authors responded to the reviews with care, improving their papers as a result. Some papers whose authors could not travel to Fes are presented as recorded videos on the conference website.

The submissions showed how wide the field has become. They deal with languages from Africa, Asia, the Middle East, Europe, the Americas and Oceania, with work on building corpora and annotation schemes, machine translation and cross-lingual transfer, morphology and syntax, speech, and applications that range from the detection of hate speech to the processing of clinical text. The two special tracks, on building applications based on large language models for low-resource languages and on Modern Standard Arabic and Arabic dialects, are both represented in the programme; papers on Arabic and its dialects make up a substantial part of this volume, as befits a conference held in Morocco.

We thank the members of the Programme Committee for their reviews, written to a tight schedule, and in particular those who took on extra reviews at short notice for the late submissions. We are grateful to Nizar Habash for the keynote address, to Ruslan Mitkov for his guidance throughout, to the Organising Committee, and to Ernesto Luis Estevanell-Valladares, who prepared these proceedings.

The proceedings will be included in the ACL Anthology, where the papers will remain freely available. We hope they prove useful to everyone working to bring language technology to the languages that have so far had the least of it.

Saad Ezzini, Salima Lamsiyah and Tharindu Ranasinghe
LaTeLL'2026 Programme Committee Chairs

Organising Committee

Conference Chair
Ruslan Mitkov, Lancaster University, UK and University of Alicante, Spain

Programme Committee Chairs
Saad Ezzini, King Fahd University of Petroleum and Minerals, Saudi Arabia
Salima Lamsiyah, University of Luxembourg, Luxembourg
Tharindu Ranasinghe, Lancaster University, UK

Local Chair
Ismail Berrada, Mohammed VI Polytechnic University, Morocco

Chair of Local Arrangements
Radouane Achamlal

Organising Committee Members
Salmane Chafik, Mohammed VI Polytechnic University, Morocco
Houdaifa Atou, Mohammed VI Polytechnic University, Morocco
Milica Ikonić Nešić, University of Belgrade, Serbia
Ernesto Luis Estevanell-Valladares, University of Alicante, Spain and University of Havana, Cuba
Maram Alharbi, Lancaster University, UK

Programme Committee

Gavin Abercrombie, Heriot-Watt University, UK
Abdelghafour Aboukacem, Mohammed VI Polytechnic University, Morocco
Shadi Abudalfa, King Fahd University of Petroleum and Minerals, Saudi Arabia
Cengiz Acartürk, Jagiellonian University, Poland
Amneh Al Abdi, King Fahd University of Petroleum and Minerals, Saudi Arabia
Hamza Alami, Sidi Mohamed Ben Abdellah University, Morocco
Itziar Aldabe, HiTZ Center, University of the Basque Country UPV/EHU, Spain
Maram Alharbi, Lancaster University, UK
Hassina Aliane
Salim Al Mandhari, Lancaster University, UK
Reem Ali Alzahrani, Imam Abdulrahman Bin Faisal University, Saudi Arabia
Mullosharaf Kurbonovich Arabov, Kazan Federal University, Russia
Ibrahim Aref, Lancaster University, UK
Houdaifa Atou, Mohammed VI Polytechnic University, Morocco
Soran Badawi, Charmo University, Iraq
Timo Baumann, Ostbayerische Technische Hochschule Regensburg, Germany
Abdessamad Benlahbib, Sidi Mohamed Ben Abdellah University, Morocco
Laura Bernardy, University of Luxembourg, Luxembourg
Mariam Biltawi, Al Hussein Technical University, Jordan
Happy Buzaaba, Princeton University, USA
Salmane Chafik, Mohammed VI Polytechnic University, Morocco
Abdelhalim Hafedh Dahou, GESIS – Leibniz Institute for the Social Sciences, Germany
Daniel Dakota, Indiana University, USA
Sudhansu Bala Das, Insight Research Ireland Centre for Data Analytics, University of Galway, Ireland

Tabea De Wille, University of Limerick, Ireland
Liviu Dinu, University of Bucharest, Romania
Damith Dola Mullage, Lancaster University, UK
Omar El Bachyr, University of Luxembourg, Luxembourg
Abdessalam Elhabbash, Lancaster University, UK
Noura El Moussa, Franklin University, USA
Ernesto Luis Estevanell-Valladares, University of Alicante, Spain and University of Havana, Cuba
Saad Ezzini, King Fahd University of Petroleum and Minerals, Saudi Arabia
Christiane Fellbaum, Princeton University, USA
Aadil Gani Ganie, King Fahd University of Petroleum and Minerals, Saudi Arabia
Nadia Ghezaiel Hammouda, University of Ha'il, Saudi Arabia
Yoan Gutiérrez, University of Alicante, Spain
Amal Haddad Haddad, University of Granada, Spain
Hicham Hammouchi, University of Luxembourg, Luxembourg
Ogtay Hasanov, King Fahd University of Petroleum and Minerals, Saudi Arabia
Milica Ikonić Nešić, University of Belgrade, Serbia
Anton Karl Ingason, University of Iceland, Iceland
Danka Jokić, University of Belgrade, Serbia
Nouran Khallaf, University of Leeds, UK
Gražina Korvel, Vilnius University, Lithuania
Sobha Lalitha Devi, AU-KBC Research Centre, Anna University, India
Salima Lamsiyah, University of Luxembourg, Luxembourg
Cedric Lothritz, Luxembourg Institute of Science and Technology, Luxembourg
Aleksandra Marković, Institute for the Serbian Language of SASA, Serbia
Javier Martín Arista, University of La Rioja, Spain
Brigitte Mathiak, GESIS – Leibniz Institute for the Social Sciences, Germany
So Miyagawa, University of Tsukuba, Japan
Simeon Monov, Plovdiv University, Bulgaria
Nikola S. Nikolov, University of Limerick, Ireland
Ivelina Nikolova, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria
Temilola Oloyede
Partha Pakray, National Institute of Technology Silchar, India
Nikolay Pavlov, Plovdiv University, Bulgaria
Scott Piao, Lancaster University, UK
Alistair Plum, University of Luxembourg, Luxembourg
Jihad R'baiti, Mohammed VI Polytechnic University, Morocco
Sara Renji
Branislava Šandrih Todorović, NLB DigIT, Belgrade, Serbia
Nasredine Semmar, CEA, France
Ranka Stanković, University of Belgrade, Serbia
Alina Trubnikova, University of Granada, Spain
Menno van Zaanen, South African Centre for Digital Language Resources, South Africa
Andreas Vlachidis, University College London, UK
Sumit Yadav, IOE Pulchowk Campus, Nepal
Jihad Zahir, Cadi Ayyad University, Morocco
Katerina Zdravkova, Ss. Cyril and Methodius University in Skopje, North Macedonia

Table of Contents

A Kazakh–Russian Corpus of Child-Directed Language for Low-Resource Languages
Albina Mukusheva, Achille Fusco and Cristiano Chesi 1

Data Curation, Annotation Quality, and Error Patterns in Old English Automatic Lemmatisation
Javier Martín Arista 8

Creating the RozMuz Corpus: Applying Ethics and Technology in Language Research
Alicja Helena Derych, Bartłomiej Alberski, Hubert Jankowski and Paweł Dembowski 18

Towards a Universal Dependencies Treebank for Amazigh: A Tarifit Pilot Study
Azzeddine Afrouni, Fadoua Ataa Allah and Jamal Abarnous 24

Arabic Dialect-to-MSA Translation: A Comparative Evaluation of Large Language Models and Neural Machine Translation
Maram I. Alharbi, Jihad R'baiti and Ruslan Mitkov 30

Cross-Lingual Transfer from Portuguese to Nheengatu: Evidence for Contact-Induced Convergence as a Computational Bridge
Rafael Macario Fernandes 40

Retrieval-Augmented Machine Translation for Bohairic Coptic: A Pilot Case Study
So Miyagawa 50

Automating Mwotlap morphology using formal grammar
Fabio Meroni, Alexandre François and Max Silberztein 59

Dependency Parsing Across the Resource Spectrum: Evaluating Architectures on High and Low-Resource Languages
Kevin Guan, Happy Buzaaba and Christiane Fellbaum 68

A finite-state model of Innu verbal inflection
Loïc Daignault-Pichette and François Lareau 80

Morphological Operations in the Persian Verbal System
Marzieh Rabiei and Max Silberztein 89

A Massive Open-Source Corpus for Valencian: Over 4.7 Billion Tokens for Low-Resource Language Modelling
Yoan Gutiérrez, Juan Pablo Consuegra-Ayala, Robiert Sepúlveda-Torres and Rafael Muñoz Guillena 96

LuxIT: A Luxembourgish Instruction Tuning Dataset from Monolingual Seed Data
Julian Valline, Cedric Lothritz, Siwen Guo and Jordi Cabot Sagrera 102

Topic Modeling for Moroccan Darija: A Comparative Study of Classical Machine Learning Approaches
Salma Mekaoui, Ilham Chaker, Arsalane Zarghili and Nikola S. Nikolov 121

Distil and Evolve: Compact Adaptive Document Classification with Continual Learning for Low-Resource Settings
Mehdi Soufiane, Mourad Hammale, Kawtar Zerhouni, Adil Ahidar-Coutrix and Rachid Benouini 131

Do Large Language Models Style-Shift? Register-Conditioned Stylistic Fronting in AI-Generated Icelandic
Anton Karl Ingason, Johanna Mechler and Lilja Björk Stefánsdóttir 140

Benchmarking Large Language Models on Mandarin Proverb Explanation and Contextual Matching
Xiaojing Zhao, Salima Lamsiyah, Emmanuele Chersoni and Han Xu 150

LUXDIAG-RAG: Diagnostic Evaluation of Retrieval-Augmented Generation for Luxembourgish Reading Comprehension
Keerthana Murugaraj, Hedi Tebourbi, Christophe Friezas Gonçalves and Salima Lamsiyah 161

Assessment of Human-in-the-Loop Multi-LLMs for Low-Resource Educational Data Expansion
Christophe Friezas Gonçalves, Hedi Tebourbi, Christoph Schommer and Salima Lamsiyah 172

Human-Centred Approaches in Low-Resource Languages for Educational Applications and Language Learning: A Systematic Literature Review
Noura El Moussa 181

SimuLe-Lux: A Neuro-Symbolic Teaching System for Diagnosing L2 Reading Comprehension
Christophe Friezas Gonçalves, Christoph Schommer and Salima Lamsiyah 191

Towards Readability Assessment for Under-Resourced Arabic Dialects: A Study of Moroccan Darija
Houdaifa Atou, Nouran Khallaf, Salima Lamsiyah and Ruslan Mitkov 199

Why Is Current XAI Not Enough for Arabic NLP? A Critical Survey of the Explainability Gap
Salima Lamsiyah and Ruslan Mitkov 226

Benchmarking Speech Foundation Models and LLMs for Sentiment Analysis in Low-Resource Najdi Arabic of Saudi Arabia
Nadia Ghezaiel Hammouda, Maram I. Alharbi and Ruslan Mitkov 237

Spatial Entity Extraction Methods from Arabic Texts in the Context of Epidemiological Surveillance: A Comparative Study
Fatima Ezzahra El Houbri, Najlae Idrissi, Mathieu Roche and Sarah Valentin 246

Building an Aligned Speech Corpus for Northern Russian
Kseniia Protonina and Daniil Ignatev 257

Probing Geographic Performance Gaps in Moroccan ASR with ALAMA: Annotated Local Audio of Moroccan Arabic
Avery Cole Kanel, Christian Schuler, Bouazza Laracha, Imrane Lbouhli, Yassine Chaouri, Yusser Al Ghussin and Timo Baumann 264

Generator-Guided Amount Recovery for Voice-Based Financial Record-Keeping in Mooré-French Code-Switched Speech
Maimouna Ouattara, El-Hacen Diallo, Fred Philippy, Abdoul Kader Kaboré, Jacques Klein and Tegawendé F. Bissyandé 272

LëtzCross: A Cross-Lingual Page-Level Benchmark for Multimodal Retrieval over Luxembourgish Documents
Omar El Bachyr, Fred Philippy, Laura Maria Bernardy, Saad Ezzini, Jacques Klein and Tegawendé Bissyandé 284

Improving OCR for a Latvian Pronunciation Dictionary
Viesturs Jūlijs Lasmanis 295

LLM-based conversational systems for Sinhala: Investigating model limitations and exploring a framework for agentic retrieval
Malithi P. Alahapperuma, Andreas Vlachidis and Antonis Bikakis 304

Clinical Entity Recognition from Electronic Health Records and Linking to Biomedical Knowledge Bases in Low-resource Language Settings
Eno-Martin Lotman, David Hübner, Kris Collins, Svetla Boytcheva and Ivelina Nikolova-Koleva 309

Beyond RAG: A Multi-Graph, Multi-Agent, Recursive Retrieval Architecture for Traceable and Source-Grounded Legal Answers
Assia Bouamir, Marie Bonnin, Youssef Al Mouatamid and Jihad Zahir 318

FLICK: Few-Label Incremental Learning for Low-Resource Dialects
Ali Almutairi, Abdullah Alsuhaibani, Shoaib Jameel, Aditya Joshi, Gelareh Mohammadi and Imran Razzak 326

Linguistic Proximity Enables Pivot-Based Machine Translation for an Under-Resourced Tribal Language
Pooja Singh, M Kaab Bin Shahid, Atai Waris Khan, Aryan Kumar Jha and Sandeep Kumar . . . 339

Cross-Lingual Hate Speech Detection in Low-Resource Languages: The Case of Persian and Kurdish
Shahin Yousefi, Ernesto Luis Estevanell-Valladares and Ruslan Mitkov 350

Is Overconfidence Language-Specific? Cross-Lingual Calibration and Recalibration Transfer in Base and Instruction-Tuned LLMs
Divya Gupta and Ruslan Mitkov 361

Knowledge Tracing for Early Childhood Learners with Minimal Telemetry from Low-Connectivity Environments: Evidence from India's Anganwadi Ecosystem
Badmavasan Kirouchenassamy, Rahul Singh, Vishnu Dev, Meenakshi Garg, Sukhna Sawhney and Sudeep Gowrishankar 370

Kuwain 1.5B: An Arabic SLM via Language Injection
Khalil Hennara, Sara Chrouf, Mohamed Motasim Hamed, Zeina Aldallal and Safwan AlModhayan 378

A Neuro-Symbolic RAG System for Marine Environmental Law: from Domain Ontology to Knowledge Graphs and Context-Enriched Legal Question Answering
Assoumana Souley Hadiza, Youssef Al Mouatamid, Marie Bonnin and Jihad Zahir 388

Evaluating Model-Task Fit in Arabic Word Sense Disambiguation
Yousef Younes, Abdelhalim Hafedh Dahou and Brigitte Mathiak 398

Enhancing Urdu ASR with Whisper v3: Fine-Tuning on Latest Datasets and Realistic Multi-Speaker Evaluation with SLM Post-Processing
Zehra Ahmed, Farah Inayat, Zuha Aqib and Sajjad Haider 408

Baseer: A Vision-Language Model for Arabic Document-to-Markdown OCR
Khalil Hennara, Muhammad Hreden, Mohamed Motasim Hamed, Ahmad Bastati, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan 418

Worldview Annotation for Low-Resource Languages
Tetiana Ilman 429

A Multilingual Sentence-Transformer Baseline for Serbian News Topic Classification
Saša Petalinkar, Milica Ikonić Nešić, Ranka Stanković and Jelena Graovac 438

Wasm: A Pipeline for Constructing Structured Arabic Interleaved Multimodal Corpora
Khalil Hennara, Ahmad Bastati, Muhammad Hreden, Mohamed Motasim Hamed, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan 449

Linguistic Specialization of Arabic in Text Embedding Models via Multi-Teacher Knowledge Distillation
Ahmad Abdelfattah, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan 459

Conference Program

Monday September 28, 2026

10:50–11:50 **Session 1: Corpora, Annotation and Resource Development**
Chair: Yoan Gutiérrez

10:50–11:10 *A Kazakh–Russian Corpus of Child-Directed Language for Low-Resource Languages*
Albina Mukusheva, Achille Fusco and Cristiano Chesi

11:10–11:30 *Data Curation, Annotation Quality, and Error Patterns in Old English Automatic Lemmatisation*
Javier Martín Arista

11:30–11:40 *Creating the RozMuz Corpus: Applying Ethics and Technology in Language Research*
Alicja Helena Derych, Bartłomiej Alberski, Hubert Jankowski and Paweł Dembowski

11:40–11:50 *Towards a Universal Dependencies Treebank for Amazigh: A Tarifit Pilot Study*
Azzeddine Afrouni, Fadoua Ataa Allah and Jamal Abarnous

11:50–12:30 **Session 2: Machine Translation and Cross-Lingual Transfer**
Chair: Salima Lamsiyah

11:50–12:10 *Arabic Dialect-to-MSA Translation: A Comparative Evaluation of Large Language Models and Neural Machine Translation*
Maram I. Alharbi, Jihad R'baiti and Ruslan Mitkov

12:10–12:20 *Cross-Lingual Transfer from Portuguese to Nheengatu: Evidence for Contact-Induced Convergence as a Computational Bridge*
Rafael Macario Fernandes

12:20–12:30 *Retrieval-Augmented Machine Translation for Bohairic Coptic: A Pilot Case Study*
So Miyagawa

Monday September 28, 2026 (continued)

14:20–15:30 **Session 3: Morphology, Syntax and Formal Modelling**
Chair: Javier Martín Arista

14:20–14:40 *Automating Mwotlap morphology using formal grammar*
Fabio Meroni, Alexandre François and Max Silberztein

14:40–15:00 *Dependency Parsing Across the Resource Spectrum: Evaluating Architectures on High and Low-Resource Languages*
Kevin Guan, Happy Buzaaba and Christiane Fellbaum

15:00–15:20 *A finite-state model of Innu verbal inflection*
Loïc Daignault-Pichette and François Lareau

15:20–15:30 *Morphological Operations in the Persian Verbal System*
Marzieh Rabiei and Max Silberztein

15:50–17:10 **Session 4: Training Data and Model Adaptation**
Chair: Saad Ezzini

15:50–16:10 *A Massive Open-Source Corpus for Valencian: Over 4.7 Billion Tokens for Low-Resource Language Modelling*
Yoan Gutiérrez, Juan Pablo Consuegra-Ayala, Robiert Sepúlveda-Torres and Rafael Muñoz Guillena

16:10–16:30 *LuxIT: A Luxembourgish Instruction Tuning Dataset from Monolingual Seed Data*
Julian Valline, Cedric Lothritz, Siwen Guo and Jordi Cabot Sagrera

16:30–16:50 *Topic Modeling for Moroccan Darija: A Comparative Study of Classical Machine Learning Approaches*
Salma Mekaoui, Ilham Chaker, Arsalane Zarghili and Nikola S. Nikolov

16:50–17:10 *Distil and Evolve: Compact Adaptive Document Classification with Continual Learning for Low-Resource Settings*
Mehdi Soufiane, Mourad Hammale, Kawtar Zerhouni, Adil Ahidar-Coutrix and Rachid Benouini

Tuesday September 29, 2026

09:00–10:00 **Session 5: Language Model Behaviour and Evaluation**
Chair: Ernesto Luis Estevanell-Valladares

09:00–09:20 *Do Large Language Models Style-Shift? Register-Conditioned Stylistic Fronting in AI-Generated Icelandic*
Anton Karl Ingason, Johanna Mechler and Lilja Björk Stefánsdóttir

09:20–09:40 *Benchmarking Large Language Models on Mandarin Proverb Explanation and Contextual Matching*
Xiaojing Zhao, Salima Lamsiyah, Emmanuele Chersoni and Han Xu

09:40–10:00 *LUXDIAG-RAG: Diagnostic Evaluation of Retrieval-Augmented Generation for Luxembourgish Reading Comprehension*
Keerthana Murugaraj, Hedi Tebourbi, Christophe Friezas Gonçalves and Salima Lamsiyah

10:50–11:50 **Session 6: Human-Centred NLP and Education**
Chair: Salima Lamsiyah

10:50–11:10 *Assessment of Human-in-the-Loop Multi-LLMs for Low-Resource Educational Data Expansion*
Christophe Friezas Gonçalves, Hedi Tebourbi, Christoph Schommer and Salima Lamsiyah

11:10–11:30 *Human-Centred Approaches in Low-Resource Languages for Educational Applications and Language Learning: A Systematic Literature Review*
Noura El Moussa

11:30–11:50 *SimuLe-Lux: A Neuro-Symbolic Teaching System for Diagnosing L2 Reading Comprehension*
Christophe Friezas Gonçalves, Christoph Schommer and Salima Lamsiyah

Tuesday September 29, 2026 (continued)

11:50–12:50 **Session 7: Arabic and Dialect NLP**
Chair: Salmane Chafik

11:50–12:10 *Towards Readability Assessment for Under-Resourced Arabic Dialects: A Study of Moroccan Darija*
Houdaifa Atou, Nouran Khallaf, Salima Lamsiyah and Ruslan Mitkov

12:10–12:30 *Why Is Current XAI Not Enough for Arabic NLP? A Critical Survey of the Explainability Gap*
Salima Lamsiyah and Ruslan Mitkov

12:30–12:40 *Benchmarking Speech Foundation Models and LLMs for Sentiment Analysis in Low-Resource Najdi Arabic of Saudi Arabia*
Nadia Ghezaiel Hammouda, Maram I. Alharbi and Ruslan Mitkov

12:40–12:50 *Spatial Entity Extraction Methods from Arabic Texts in the Context of Epidemiological Surveillance: A Comparative Study*
Fatima Ezzahra El Houbri, Najlae Idrissi, Mathieu Roche and Sarah Valentin

14:20–15:30 **Session 8: Speech and Spoken-Language Technologies**
Chair: Nadia Ghezaiel

14:20–14:40 *Building an Aligned Speech Corpus for Northern Russian*
Kseniia Protonina and Daniil Ignatev

14:40–15:00 *Probing Geographic Performance Gaps in Moroccan ASR with ALAMA: Annotated Local Audio of Moroccan Arabic*
Avery Cole Kanel, Christian Schuler, Bouazza Laracha, Imrane Lbouhli, Yassine Chaouri, Yusser Al Ghussin and Timo Baumann

15:00–15:20 *Generator-Guided Amount Recovery for Voice-Based Financial Record-Keeping in Mooré-French Code-Switched Speech*
Maimouna Ouattara, El-Hacen Diallo, Fred Philippy, Abdoul Kader Kaboré, Jacques Klein and Tegawendé F. Bissyandé

Tuesday September 29, 2026 (continued)

15:50–17:00 **Session 9: Retrieval, RAG and Document Processing**
Chair: Houdaifa Atou

15:50–16:10 *LëtzCross: A Cross-Lingual Page-Level Benchmark for Multimodal Retrieval over Luxembourgish Documents*
Omar El Bachyr, Fred Philippy, Laura Maria Bernardy, Saad Ezzini, Jacques Klein and Tegawendé Bissyandé

16:10–16:30 *Improving OCR for a Latvian Pronunciation Dictionary*
Viesturs Jūlijs Lasmanis

16:30–16:40 *LLM-based conversational systems for Sinhala: Investigating model limitations and exploring a framework for agentic retrieval*
Malithi P. Alahapperuma, Andreas Vlachidis and Antonis Bikakis

16:40–16:50 *Clinical Entity Recognition from Electronic Health Records and Linking to Biomedical Knowledge Bases in Low-resource Language Settings*
Eno-Martin Lotman, David Hübner, Kris Collins, Svetla Boytcheva and Ivelina Nikolova-Koleva

16:50–17:00 *Beyond RAG: A Multi-Graph, Multi-Agent, Recursive Retrieval Architecture for Traceable and Source-Grounded Legal Answers*
Assia Bouamir, Marie Bonnin, Youssef Al Mouatamid and Jihad Zahir

Session VID: Video Presentations (on the conference website)

FLICK: Few-Label Incremental Learning for Low-Resource Dialects
Ali Almutairi, Abdullah Alsuhaibani, Shoaib Jameel, Aditya Joshi, Gelareh Mohammadi and Imran Razzak

Linguistic Proximity Enables Pivot-Based Machine Translation for an Under-Resourced Tribal Language
Pooja Singh, M Kaab Bin Shahid, Atai Waris Khan, Aryan Kumar Jha and Sandeep Kumar

Cross-Lingual Hate Speech Detection in Low-Resource Languages: The Case of Persian and Kurdish
Shahin Yousefi, Ernesto Luis Estevanell-Valladares and Ruslan Mitkov

Is Overconfidence Language-Specific? Cross-Lingual Calibration and Recalibration Transfer in Base and Instruction-Tuned LLMs
Divya Gupta and Ruslan Mitkov

Knowledge Tracing for Early Childhood Learners with Minimal Telemetry from Low-Connectivity Environments: Evidence from India's Anganwadi Ecosystem
Badmavasan Kirouchenassamy, Rahul Singh, Vishnu Dev, Meenakshi Garg, Sukhna Sawhney and Sudeep Gowrishankar

Tuesday September 29, 2026 (continued)

Kuwain 1.5B: An Arabic SLM via Language Injection
Khalil Hennara, Sara Chrouf, Mohamed Motasim Hamed, Zeina Aldallal and Safwan AlModhayan

A Neuro-Symbolic RAG System for Marine Environmental Law: from Domain Ontology to Knowledge Graphs and Context-Enriched Legal Question Answering
Assoumana Souley Hadiza, Youssef Al Mouatamid, Marie Bonnin and Jihad Zahir

Evaluating Model-Task Fit in Arabic Word Sense Disambiguation
Yousef Younes, Abdelhalim Hafedh Dahou and Brigitte Mathiak

Enhancing Urdu ASR with Whisper v3: Fine-Tuning on Latest Datasets and Realistic Multi-Speaker Evaluation with SLM Post-Processing
Zehra Ahmed, Farah Inayat, Zuha Aqib and Sajjad Haider

Baseer: A Vision-Language Model for Arabic Document-to-Markdown OCR
Khalil Hennara, Muhammad Hreden, Mohamed Motasim Hamed, Ahmad Bastati, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan

Worldview Annotation for Low-Resource Languages
Tetiana Ilman

A Multilingual Sentence-Transformer Baseline for Serbian News Topic Classification
Saša Petalinkar, Milica Ikonić Nešić, Ranka Stanković and Jelena Graovac

Wasm: A Pipeline for Constructing Structured Arabic Interleaved Multimodal Corpora
Khalil Hennara, Ahmad Bastati, Muhammad Hreden, Mohamed Motasim Hamed, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan

Linguistic Specialization of Arabic in Text Embedding Models via Multi-Teacher Knowledge Distillation
Ahmad Abdelfattah, Zeina Aldallal, Sara Chrouf and Safwan AlModhayan