



ICNLSP 2026

**Proceedings of the 9th
International Conference on
Natural Language and Speech
Processing (ICNLSP-2026)**

25–26 September, 2026



©2026 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-458-3

<https://www.icnlsp.org/>

Introduction

Welcome to the proceedings of the **9th International Conference on Natural Language and Speech Processing (ICNLSP 2026)**.

Every volume of proceedings serves a dual purpose. In its most visible form, it is an archive of scientific progress. More fundamentally, though, it is a record of our community's changing curiosities. The papers selected in these pages reflect not just what our field can solve today, but what we are choosing to care about next.

In response to clear feedback from our participants—with the vast majority of our authors expressing a preference for virtual participation—this year's conference, originally planned as a hybrid event, transitioned to a fully online format, ensuring our research remains widely accessible across geographic boundaries.

This year, we received 98 submissions from 29 countries. Following a rigorous double-blind peer-review process, forty 40 papers were selected for presentation (40.8%). While these metrics reflect the focused scope and high selectivity of our venue, the true scientific value of these proceedings resides in the substance of the papers themselves.

We owe a debt of gratitude to our Program Committee and reviewers, who devoted their time and expertise evaluating each submission with great care. We appreciate their thoughtful evaluations and generous commitment of time.

Our sincere thanks go to all submitting authors. Regardless of the review outcome, every submission contributes meaningfully to the advancement of our field.

Finally, we express our appreciation to DataScientia for its willingness to host the conference and for its vital support throughout our transition from a hybrid to a virtual conference.

We hope these proceedings serve as both a reliable reference for current practice and a catalyst for future ideas across our research community.

Mourad Abbas and Abed Alhakim Freihat

Organizers:**Chairs:**

Prof. Mourad Abbas

Dr. Abed Alhakim Freihat

Program Committee Chair:

Prof. Mourad Abbas

Program Committee:

Mourad Abbas, USTHB, Algeria.

Ahmed Abdelali, QCRI, Qatar.

Ahmad Alkhodre, IUM, KSA.

Yuan An, Drexel University, USA.

Nicklas Sindlev Andersen, SDU, Denmark.

Nicolas Ballier, University Paris Cité, France.

Gianluca Barmina, SDU, Denmark.

Homayoon Beigi, Columbia University, USA.

Fayssal Bouarourou, University of Strasbourg, France.

Daniel Braun, University of Marburg, Germany.

Giuseppe Celano, Leipzig University, Germany.

Mohamed Amine Cheragui, Ahmed Draia University, Algeria.

Gérard Chollet, CNRS, France.

Hadda Cherroun, Amar Telidji University, Algeria.

Mounim El Yacoubi, Telecom SudParis, France.

Ashraf Elnagar, University of Sharjah, UAE.

Pascale Feldkamp, Aarhus University, Denmark.

Abed Alhakim Freihat, MBZUAI, UAE.

Lavanya Gupta, Carnegie Mellon University, USA.

Kristiina Jokinen, University of Helsinki, Finland.

Mohammed Mediani, UAE University, UAE.

Jozsef Mezei, Åbo Akademi University, Finland.

Zoltan Miklos, IRISA, France.

Martin Rehm, SDU, Denmark.

Maxime Holmberg Sainte-Marie, SDU, Denmark.

Hugues Sansen, InstantonWave, France.

Hassan Satori, FSDM, Morocco.

Yücel Saygin, Sabanci University, Turkey.

Nasredine Semmar, CEA, France.

Peter Sullivan, University of British Columbia, Canada.

Irina Temnikova, Big Data for Smart Society Institute, Bulgaria.

María Inés Torres, University of the Basque Country, Spain.

Nikos Tsourakis, University of Geneva, Switzerland.

Christian Vedel, SDU, Denmark.

Christian Wartena, University of Applied Sciences and Arts Hannover, Germany.
Ramin Yahyapour, University of Göttingen, Germany.
Fayçal Ykhlef, CDTA, Algeria.
Tariq Yousef, SDU, Denmark.
Mai Zaki, American University of Sharjah, UAE.
Zaifu Zhan, University of Minnesota, USA.

Table of Contents

Does Financial Speech Need Continual Learning Regularization? A Study of EWC and mini-KFAC on Earnings Call ASR	1
<i>Thai On and Homayoon Beigi</i>	
Budgeting Speech-to-LLM ASR for Underrepresented Dutch Speech	11
<i>Xinnian Zhao, Hugo Vanhamme and Pu Wang</i>	
Textonomy: A TnT-LLM-Based Approach for Interpretable Topic Modeling at Scale	17
<i>Noah Zuchna, Martin Stoll and Jan Blechschmidt</i>	
Thread-Preserving Chunking for LLM-Based Action Item Extraction from Enterprise Email under Constrained Context Budget	33
<i>kiyeop kim and Donghoon Cho</i>	
CURB: Civic Urban Reasoning Benchmark	38
<i>Oliver Fox and Ellis Solaiman</i>	
When Does Uncertainty Measure RAG Hallucination Risk?	53
<i>Rajeel Ansari and Tautik Agrahari</i>	
One Bad Juror: Exploring Adversarial Influence and Deliberation Dynamics in Multi-Agent LLM Panels	59
<i>Vaishakh Menon</i>	
AutoSpecNER: A Fine-Grained Named Entity Recognition Dataset for Vehicle Specification Extraction	67
<i>Jordan Lee, Filippos Ventirozos, Abdirahman Abdullahm, Ioanna Nteka, Peter Appleby and Matthew Shardlow</i>	
Making the Implicit Explicit: A Domain Ontology-Based Approach for Comprehensive Attribute-Level Knowledge Extraction in Technical Texts	82
<i>Vatosoa Razafniary</i>	
Verbal and Nonverbal Coherence as an Index for Second Language Proficiency	96
<i>Katsunori Kotani and Takehiko Yoshimi</i>	
IntentifierBench: An Open-Source Web Tool for Benchmarking Authorship Identification using Illocutionary Force Large Language Models Based Classifiers	107
<i>Godfrey Inyama</i>	
Revisiting Accent Classification with DistilHuBERT: Speaker-Disjoint Evaluation and Cross-Corpus Generalization	117
<i>Xu He and Homayoon Beigi</i>	
Token Merging for Multilingual Speech Recognition: A Systematic Study Across Model Scale and Fine-Tuning	126
<i>Dylan Holyoak</i>	
Improving Persian Sentiment Classification through LLM-Assisted Label Refinement and Small Language Models	138
<i>Maziyar Amini Alashti and Seyed Abbas Bagheri</i>	
Masked Temporal Alignment for Automated Speech Scoring	148

Akshay Badola and Ikkyu Choi

RomMa-FT: Static Word Embeddings for Romanized Marathi-English Code-Mixed Hate Speech Detection 157

Nitin Chavan and Badal Soni

Persona as Infrastructure: Invisible Structural Control in LLM-Mediated Social Networks 164

Amir Reza Asadi, Jess Kropczynski and Hazem Said

KidLit by Numbers: Linguistic Profiles of Original and Translated Russian 0+ Books 179

Daria Gavrilik and Aleksandra Ostraiia

Human-in-the-loop Generation System for Analogical Explanations of Target Concepts 191

Ryo Takahashi and Asuka Terai

GEMCo: A Validated, Ethically Releasable Proxy for Inaccessible Counselling Data 199

Philipp Steigerwald, Eric Rudolph, Mara Stieler, Jennifer Burghardt, Robert Lehmann and Jens Albrecht

Text Segmentation is Subjective: Empirical Evidence of Human Variability and its Implications for Downstream Tasks 216

Sophia Conrad and Florian Frenken

FrIdéo: French News Outlet Ideology Scoring via Multi-Source Evidence Fusion and Distancing-Aware Lexical Scoring 223

Amr Sobhy

Framing by Wording, Framing by Selection: A Large-Scale Two-Dimensional Audit of French News Headlines, 2022–2025 239

Amr Sobhy

Cross-Lingual Transfer Learning for Multilingual Emotion Classification and Generation 259

Adrian Maldonado and Oscar Araque

Combined Learning of Morphological Segmentation, POS Tagging and Contextual Lemmatization 269

Christian Wartena

Evaluating Transformers with NepSum: A Large-Scale Dataset for Nepali Abstractive Summarization 279

Saskar Khadka and Pawan Khatri

When Silver Isn't Enough: Data Quality Effects on Pnar-English Machine Translation 290

Riya Tekcham, Fitha Asma Sulfeekhar and Badal Nyalang

A Multimodal Corpus for Natural Language Understanding in Russian and Tatar 297

Alsu Khabirzyanova, Zaryana Damashova and Ilseyar Alimova

LLM-as-Annotator at 4B Scale: Strategy-Specific Instruction Tuning for Chinese Metaphor Identification 310

Yufeng WU

Probing the phon* representations of the Mimi codebooks 322

Artem Saloev, Nicolas Ballier and Erin Pacquetet

Seeds Before Objectives: Rethinking Evaluation for Low-Resource Garhwali ASR 332

Karamvir Singh Batra, Prathamjyot Singh, Ashima Sood, Jasmeet Singh and Sahil Sharma

Towards a clinical adaptation of the Whisper model for French 350

Haeul Hwang, Deok-Hee Kim-Dufor, Jalal Al-Tamimi, Motasem Alrahabi, Julien Desclés, Valeria Lu-

carini, Louis Pichereau and Christophe Lemey

Adaptive Multi-Teacher Distillation for Accent-Robust Speech Recognition	359
<i>Mehedi Bijoy, Dejan Porjazovski, Tamás Grósz and Mikko Kurimo</i>	
ResPL: Reasoning-Based Pseudo-Labels for Sarcasm Detection in Low-Resource Arabic Dialects	367
<i>Ali Almutairi, Shoaib Jameel, Aditya Joshi, Gelareh Mohammadi and Imran Razzak</i>	
MetaHOPE: Metaphor Translation Evaluation Framework Investigating Open-Source LLMs and State-of-the-Art Neural Translation Models	380
<i>Jiahui Liang and Lifeng Han</i>	
ArImpliSent: A Large-Scale Dataset for Arabic Implicit Sentiment Analysis	402
<i>Ghozlene Chenni, Sadik Bessou, Hanane Boutouta and Nour elhouda Chenni</i>	
Evaluating Latent Semantic Pre-training for Fine-grained Word Sense Disambiguation	411
<i>Soma Simon and Gábor Berend</i>	
Looking under the Wrong Lamppost: On the Limitations of Automated Translation Quality Estimation	423
<i>Serge Gladkoff, Angelika Vaasa, Sue Ellen Wright, Ingemar Strandvik and Lifeng Han</i>	
Listening for Speaker Identity in Multilingual Speech Models: Phonetic Similarity Predicts Transfer in Cross-Lingual Speaker Verification	437
<i>AriaRay Brown, Yusser Al Ghussin and Dietrich Klakow</i>	
From Folklore to Fine-Tuning: An Authentic Angika-Hindi Parallel Corpus and Warm-Initialized Translation Baseline	448
<i>Jyoti Kumari</i>	