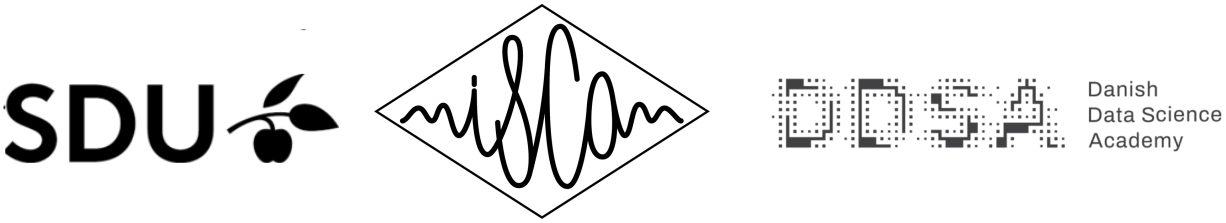
ICNLSP 2025

Proceedings of the 8th International Conference on Natural Language and Speech Processing (ICNLSP-2025)

25–27 August, 2025



©2025 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 979-8-89176-297-8

https://www.icnlsp.org/

Introduction

It is with great pleasure that we welcome you to the 8th International Conference on Natural Language and Speech Processing (ICNLSP 2025), held at Southern Denmark University from August 25-27, 2025. This volume serves as a comprehensive record of the innovative research and groundbreaking insights on different topics discussed during the conference.

This year's conference attracted 130 submissions from around the globe. The acceptance rate was around 34 %. The 45 accepted papers represent the culmination of rigorous inquiry and intellectual dedication, covering a diverse range of topics within NLP field. Indeed, they showcase the current state of knowledge and shed light on new directions for future exploration. We thank the authors for their valuable contributions.

In order to recognize outstanding scientific contributions, we decided this year to present two awards for the best papers (full and short ones). We congratulate the winners and extend our sincere gratitude to the scientific committee who selected the winners based on recommendations from the program committee, and on the originality, significance, and quality of the research, as well as the clarity of presentation.

We are profoundly honored by the participation of our distinguished keynote speakers, Prof. Dr. Barbara Plank, Prof. Dr. Anders Søgaard, and Prof. Peter Schneider-Kamp. whose insights and vision profoundly enriched the conference.

We thank the conference management members for their efforts, and the program committee and reviewers for their diligent work in curating the high-quality content contained within these pages.

Finally, we are deeply grateful to Southern Denmark University, Danish Data Science Academy, and International Speech Communication Association (ISCA) for their support.

Mourad Abbas, Tariq Yousef, and Lukas Galke

Organizers:

Chairs:
Dr. Mourad Abbas
Dr. Tariq Yousef

Program Committee Chair:
Prof. Lukas Galke

Publicity Chair:
Dr. Abed Alhakim Freihat

Program Committee:

Ahmed Abdelali, SDAIA, KSA.
Hend Al-Khalifa, KSU, KSA.
Mehmet Fatih Amasyalı, Yildiz Technical University, Turkey.
Yuan An, Drexel University, USA.
Nicklas Sindlev Andersen, SDU, Denmark.
Nicolas Ballier, University Paris Cité, France.
Fayssal Bouarourou, University of Strasbourg, France.
Pierrette Bouillon, University of Geneva, Switzerland.
William Brach, Slovak University of Technology, Slovakia.
Daniel Braun, University of Twente, Netherlands.
Giuseppe Celano, Leipzig University, Germany.
Gérard Chollet, CNRS, France.
Hadda Cherroun, Amar Telidji University, Algeria.
Christian Møller Dahl, SDU, Denmark.
Najim Dehak, Johns Hopkins University, USA.
Alexandra Diehl, University of Zurich, Switzerland.
Andor Diera, Ulm University, Germany.
Dota Dong, Max Planck Institute for Psycholinguistics, Netherlands.
Ashraf Elnagar, University of Sharjah, UAE.
Pascale Feldkamp, Aarhus University, Denmark.
Abed Alhakim Freihat, University of Trento, Italy.
Mounim El Yacoubi, Telecom SudParis, France.
Lukas Galke, SDU, Denmark.
Christian Heumann, Ludwig Maximilian University of Munich, Germany.
Kevin Hirschi, University of Texas at San Antonio, USA.
Torben Johansen, SDU, Denmark.
Kristiina Jokinen, University of Helsinki, Finland.
Eric Laporte, Gustave Eiffel University, France.
Márton Kardos, Aarhus University, Denmark.
Pierre Lison, University of Oslo, Norway.
Mohammed Mediani, UAE University, UAE.

Hermann Ney, RWTH Aachen University, Germany.
Jacob Nielsen, SDU, Denmark.
Martin Rehm, SDU, Denmark.
Uwe Reichel, University of Munich, Germany.
Michael Richter, Leipzig University, Germany.
Shadi Saleh, Prime Technologies, Czech Republic.
Yücel Saygin, Sabanci University, Turkey.
Thomas Schmidt, University of Regensburg, Germany.
Nasredine Semmar, CEA, France.
Rachele Sprugnoli, University of Parma, Italy.
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.
Jan Trmal, AppTek, USA.
Nikos Tsourakis, University of Geneva, Switzerland.
Iraklis Varlamis, Harokopio University of Athens, Greece.
Christian Vedel, SDU, Denmark.
Christian Wartena, University of Applied Sciences and Arts Hannover, Germany.
Ke Yang, University of Texas at San Antonio, USA.
Fayçal Ykhlef, CDTA, Algeria.
Tariq Yousef, SDU, Denmark.
Mai Zaki, American University of Sharjah, UAE.
Zaifu Zhan, University of Minnesota, USA.
Mike Zhang, Aalborg University, Denmark.

Scientific Committee:

Prof. Christian Møller Dahl, SDU, Denmark.
Dr. Nicklas Sindlev Andersen, SDU, Denmark.

Conference Management:

Lisa Eckford-Soper, SDU, Denmark.
Rula Mreisheh, SDU, Denmark.
Esben Andreas Wrona Bay Sørensen, SDU, Denmark.

Invited Speakers:

Prof. Dr. Barbara Plank, LMU Munich, Germany.
Prof. Dr. Anders Søgaard, University of Copenhagen, Denmark.
Prof. Peter Schneider-Kamp, SDU, Denmark.

Invited Talks

Human-Centered LLMs for Inclusive Language Technology
Prof. Dr. Barbara Plank, LMU Munich, Germany

Prof. Dr. Barbara Plank is full professor and chair for AI and Computational Linguistics at LMU Munich, Head of the Munich AI and NLP (MaiNLP) lab, and co-director of the Center for Information and Language Processing (CIS). She is also a visiting full professor at the IT University of Copenhagen.

What to think of NLP these days?
Prof. Dr. Anders Søgaard, University of Copenhagen.

Prof. Dr. Anders Søgaard is a full professor of natural language processing (NLP) and machine learning at the University of Copenhagen in Denmark. Jointly affiliated with the Dpt. of Computer Science, the Dpt. of Philosophy, the Pioneer Centre for Artificial Intelligence, and the Center for Social Data Science. Previously at University of Potsdam, Amazon Core Machine Learning, and Google Research. Father of three and a published poet.

The Cost of Intelligence: Efficiency Is the Only Path to Democratized AI
Prof. Peter Schneider-Kamp, SDU, Denmark

Prof. Peter Schneider-Kamp is a Professor of Computer Science at the University of Southern Denmark (SDU), where he holds a chair in Artificial Intelligence (AI) within the Center for Machine Learning. He co-leads the Danish Foundation Models (DFM) project on multilingual language models with a focus on Danish.

Table of Contents

Zero-Shot Commonsense Validation and Reasoning with Large Language Models: An Evaluation on SemEval-2020 Task 4 Dataset 1
Rawand Alfugaha, Mohammad AL-Smadi

Bridging the Gap: Design and Evaluation of an Automated System for French Cued Speech . . . 8
Brigitte Bigi

topicwizard - a Modern, Model-agnostic Framework for Topic Model Visualization and Interpretation 19
Márton Kardos, Kenneth Enevoldsen, Kristoffer Nielbo

Scalable Text Vectorization with Hyperdimensional Computing Through Selective Word Encoding 28
Timur Mudarisov, Evgeny Polyachenko, Zsofia Kraussl, Enriqueta Patricia Becerra Sanchez, Tatiana Petrova, Radu State

ASR Models for Traditional Emirati Arabic: Challenges, Adaptations, and Performance Evaluation 42
Maha Alblooki, Kentaro Inui, Shady Shehata

Style-Controlled Response Generation for Dialog Systems with Intimacy Interpretation 50
Takuto Miura, Kiyoaki Shirai, Natthawut Kertkeidkachorn

Next Speaker Prediction for Multi-Speaker Dialogue with Large Language Models 60
Lukas Hilgert, Jan Niehues

Adapting ASR Models for Speech-to-Punctuated-Text Recognition with Utterance Gluing 72
Agata Jakubiak, Piotr Stachyra, Piotr Czubowski, Hubert Borkowski, Sebastian Łątka, Radosław Iżak, Kornel Jankowski, Sonia Janicka, Mateusz Zieliński

Word-level Language Identification using Encoder-Only and Decoder-Only Models 82
Javier Iranzo-Sánchez, Parnia Bahar, Alejandro Pérez-González-de-Martos, Mattia Antonino Di Gangi

Beyond Shallow Heuristics: Leveraging Human Intuition for Curriculum Learning 87
Vanessa Toborek, Sebastian Müller, Tim Selbach, Tamás Horváth, Christian Bauckhage

FActBench: A Benchmark for Fine-grained Automatic Evaluation of LLM-Generated Text in the Medical Domain 93
Anum Afzal, Juraj Vladika, Florian Matthes

LLMs Information Flow Diagnostic: Memory-based Evidence from Random Matrix Theory . . . 103
Sami Diaf

CUPE: Contextless Universal Phoneme Encoder for Language-Agnostic Speech Processing 111
Abdul Rehman, Jian-jun Zhang, Xiasong Yang

The Impact of Annotator Personas on LLM Behavior Across the Perspectivism Spectrum 121
Olufunke Sarumi, Charles Welch, Daniel Braun, Jörg Schlötterer

The Chunking Paradigm: Recursive Semantic for RAG Optimization 137
Seemab Latif, Huma Ameer, Muhammad Hannan Akram, Mehwish Fatima

Towards Robust Urdu Aspect-based Sentiment Analysis through Weakly-Supervised Annotation Framework 146
Zoya Maqsood, Seemab Latif, Rabia Latif

Efficient Continual Learning for Small Language Models with a Discrete Key-Value Bottleneck . 162

Andor Diera, Lukas Galke, Fabian Karl, Ansgar Scherp

Domain adaptation and question-answer pooling for Aphasia modelling 177
Uwe Reichel, Monica Gonzalez Machorro, Lisa Maria Ehlen, Pascal Hecker, Dorothea Peitz, Cornelius Werner, Felix Burkhardt, Christian Kohlschein, Florian Eyben, Björn Schuller

A Retail-Corpus for Aspect-Based Sentiment Analysis with Large Language Models 182
Oleg Şilcenco, Marcos R. Machado, Wallace C. Ugulino, Daniel Braun

The Need for Robust and Inclusive Benchmarks in Evaluating LLMs on Arabic Text 196
Lubana Al Rayes, Ashraf Elnagar

Beyond the Boundaries of Research Fields - Mapping Educational Science in the Broader Academic Discourse about Artifical Intelligence 208
Martin Rehm, Zhiru Sun, Maxime Holmberg Sainte-Marie

From Performance to Process: Temporal Information Dynamics in Language Model Fine-tuning . 214
Frida Hæstrup, Ross Deans Kristensen-McLachlan

GUIDE: A Framework for Improving Functional Software Test Descriptions with Language Models 229
Mathis Ronzon, Thierry Roger, Zoltan Miklos, Annie Foret

Tokenization and Morphology in Multilingual Language Models: A Comparative Analysis of mT5 and ByT5 242
Thao Anh Dang, Limor Raviv, Lukas Galke

Dora explores Clinically Relevant Information in EHRs using NER 258
Martin Laursen, Lina Elkjær Pedersen, Josefine Bak H. Adelhelm, Rasmus Bank Lynggaard, Pernille Just Vinholt

CMC-SC: Cross-Modal Contextualized ASR Spelling Correction via BERT and WavLM using a Soft Fusion Framework 271
Mohammad Reza Peyghan, Sajjad Amini, Shahrokh Ghaemmaghami

On Limitations of LLM as Annotator for Low Resource Languages 277
Suramya Jadhav, Abhay Shanbhag, Amogh Thakurdesai, Ridhima Sinare, Raviraj Joshi

Speech-Based Depressive Mood Detection in the Presence of Multiple Sclerosis: A Cross-Corpus and Cross-Lingual Study 283
Monica Gonzalez-Machorro, Uwe Reichel, Pascal Hecker, Helly Hammer, Hesam Sagha, Florian Eyben, Robert Hoepner, Björn W. Schuller

Signals from Academic Hiring: A Decade of Skill Demands at a Danish University 293
Zhiru Sun, Jakob Mørup Wang

Improving French Synthetic Speech Quality via SSML Prosody Control 302
Nassima Ould ouali, Awais Hussain Sani, Ruben Bueno, Jonah Dauvet, Tim Luka Horstmann, Eric Moulines

Re-Representation in Sentential Relation Extraction with Sequence Routing Algorithm 315
Ramazan Ali Bahrami, Ramin Yahyapour

Building an Ewe Language Dataset: Towards Enhancing Automatic Speech Recognition Technologies for Low Resource Languages 328
Isaac Wiafe, Akon Obu Ekpezu, Raynard Dodzi Helegah, Fiifi Baffoe Payin Winful, Elikem Doe Atsakpo, Charles Nutrokpor, Kafui Kwashie Solaga

CLEAR: Code-Mixed ASR with LLM-Driven Rescoring 339
Shivam Kumar, Md Shad Akhtar

Beyond Labeled Datasets: Advancing TTS with Direct Preference Optimization on Unlabeled Speech Dataset 349
Andrii Zhuravlov, Volodymyr Sydorskyi

L1RA: Dynamic Rank Assignment in LoRA Fine-Tuning 360
Raul Singh, Nicolò Brunello, Vincenzo Scotti, Mark Carman

Evaluating ASR in a Clinical Context : What Whisper Misses 374
Haeeul Hwang, Eric Jordan, Deok-Hee Kim-Dufor, Christophe Lemey, Motasem Alrahabi

Tachelhiyt-Darija: a parallel speech corpus for two underrepresented languages 379
Noureddine Atouf, Elsayed Issa, Said Ouzbayr

From Outliers to Topics in Language Models: Anticipating Trends in News Corpora 385
Evangelia ZVE, Benjamin Icard, Alice Breton, Lila Sainero, Gauvain Bourgne, Jean-Gabriel Ganascia

From Context to Emotion: Leveraging LLMs for Recognizing Implicit Emotions 399
Hanane Boutouta, Abdelaziz Lakhfif, Ferial Senator, Chahrazed Mediani

Cross-Lingual Sentence-Level Skill Identification in English and Danish Job Advertisements . . . 410
Nurlan Musazade, Mike Zhang, József Mezei

Demographics and Democracy: Benchmarking LLMs' Gender Bias and Political Leaning in European Parliament 416
Jinrui Yang, Xudong Han, Timothy Baldwin

Assessing ASR Robustness for Burmese: Impacts of Missing Speech Segments and Interruptions . 440
Ankit Maurya, Manikandan Ravikiran, Rohit Saluja

Following Route Instructions using Large Vision-Language Models: A Comparison between Low-level and Panoramic Action Spaces 449
Vebjørn Kåsene, Pierre Lison

HiSlang-4.9k: A Benchmark Dataset for Hindi Slang Detection and Identification 464
Tanmay Tiwari, Vibhu Gupta, Manikandan Ravikiran, Rohit Saluja

Enhancing Arabic Retrieval Augmented Generation through Language Processing 473
Shadi Saleh, Belkacem EL Jattari, Layth Oud, Maryam Alblooshi, Bouthaina Lakhdari, Ali Alnaqbi