# Preface

With the 2025 venue, the 11th issue of the series of Language and Technology Conferences celebrated its 30st birthday. Born as Language and Technology Awareness Days in 1995, LTC, organised in Poznań every two years*, witnessed the evolution of language technologies since 2005 from Symbolic Parsing to the current era of Large Language Models. As is tradition, the 2025 edition brought together a series of presentations on the applications, creation and research of language technologies to shed light on progress, but also problems and threats, in what the general public calls the Artificial Intelligence Revolution. Today, many researchers and experts realised that a surge of novelty and progress in AI and which was enthusiastically received by media was premature: if a Large Language Model enables now to produce text and speech in various media with a quasi perfect form and address semantics and pragmatics with an unprecedented level of performance via direct natural language interaction, known in the field as prompting, machines are still not really «intelligent» and are prone to failing unpredictably in situations where a human being, even a child, would succeed. Regardless of the technological aspects, one has to be aware about the side effects of dynamically developing AI technologies: in particular the phenomena of technological exclusion of insufficiently researched and described natural languages (Low-Resourced Languages), which will result in their marginalization and, consequently, cultural and civilizational losses on a global scale. Today, in the rapidly evolving world of science and technology, LTC conferences are a place where progress in language technologies can be appreciated in the context of the cultural richness and diversity of the natural languages that nurtured them. In today's rapidly changing world of science and technology, LTC conferences provide an opportunity to assess progress in language technologies in the context of the cultural richness and diversity of natural languages that have been shaped by them.

*) outside the COVID 19 pandemic period

* * *

In this year's edition of the LTC'25 conference proceedings, we included papers illustrating the range and diversity of topics covered at the conference. We present to our readers a collection of twenty articles, consisting of works by fifty authors from leading scientific institutions in the Czech Republic, France, Germany, Hungary, India, Italy, Japan, Poland, Spain and the United Kingdom. Below, the reader will find brief notes introducing all twenty papers comprising this collection of works. The order of these notes corresponds to the arrangement of the articles in the volume.

The first one is the article (1)*»Who Speaks About Childbirth? A Corpus-Driven Study of Parliamentary Discourse in Poland»* by Jolanta Bachan and Agnieszka Jeran. This work presents a corpus-driven investigation into the discursive constructions of childbirth in Polish parliamentary debates. (2) In the second article, *«The Game of Prompts – Identifying Auxiliary Constructions with LLMs and Human Participants in a Language Game»*, Tímea Bajzát and Balázs Indig examine Hungarian auxiliary constructions in order to investigate how closely Large Language Models (LLM) align with human linguistic intuition. (3) The next paper *«The Mother Tongue Parallel Text Corpus of India: Design, Development, and Applications»* (by Narayan Choudhary, Rejitha K. S., Rajesha N., Anand Jain, Stephen Fernandes and Writtik Bhattacharya) is a collective work whose objective is to document the corpus creation process, framework design, challenges, and mitigation plan while building Mother Tongue Parallel Text Corpus of India (published by the Linguistic Data Consortium for Indian Languages (LDC-IL)). (4) What follows is the paper *«Revisiting the Discrete System-Prompt Optimization in LLM»*, where Vaclav Divis, Marek Hruz and Bastian Spatz present a general overview and detailed analysis of an approach to optimize the system prompt in a large language model (LLM) via backpropagation. (5) Another highly collective work is *«3NMark: A Psychologist-Annotated Twitter Corpus on Migration for 3N Model Analysis in Spanish and Polish»,* which is a study where the authors (Aritz Gorostiza, Małgorzata Kossowska, Reyes Rodríguez, Sandra Cava, Monika Golab, Karolina Bonarska and Gabriele Di Cicco) present the expert-annotated dataset 3NMark, specifically designed for applications of the 3N model of radicalization in Polish

Zygmunt Vetulani, Patrick Paroubek, Irakli Kardava, *Preface.* In: Z. Vetulani, P. Paroubek, I. Kardava (Eds.), Human Language Technologies as a Challenge for Computer Science and Linguistics – 2025, pp. 5–6. Poznań: Adam Mickiewicz University Press, 2025. © Authors, 2025. https://doi.org/10.14746/amup.9788323245414.0. Open Access chapter, distributed under the terms of the CC licence (BY-NC-ND, https://creativecommons.org/licenses/by-nc-nd/4.0/).

and Spanish. (6) In the article *«Toward Automatic Filling of Case Report Forms: A Case Study on Data from an Italian Emergency Department»* the authors, Gabriela Kaczmarek, Pietro Ferrazzi, Lorenzo Porta, Vicky Rubini and Bernardo Magnini, discuss a new dataset of clinical notes from an Italian Emergency Department annotated with respect to a pre-defined CRFs for further investigations. (7) Devendr Kumar, Girish Nath Jha and Khalid Choukri in their paper *«End-to-End Automatic Speech Recognition System for Sanskrit: Whisper-Based Modelling»* present the development of an end-to-end Automatic Speech Recognition (ASR) system for Sanskrit using OpenAI's Whisper architecture. (8) The authors, Tamta Kvaratskhelia and Irakli Kardava, introduce in their article (*«The Precautionary Alignment Framework (PAF): Navigating Dual-Use AI, Geopolitical Risk, and the Enforcement of Procedural Consensus»*) a new concept of Precautionary Alignment Framework (PAF), a novel governance architecture designed to establish a minimum, universal base layer for frontier AI oversight. (9) Mélanie Lancien and Brigitte Bigi present their research on the so far low-investigated issues of French Cued Speech. In the work *«French Cued Speech rhythm: first findings on the relationship between hand position and segments' duration».* (10) In their paper *«Particles in Odia, Polish and French: a Machine Translation Perspective»* Panchanan Mohanty, Zygmunt Vetulani and Grażyna Vetulani presented and discussed selected pragmatically significant particles in three languages Odia /Indo-Aran language spoken in India/, Polish and French, from the *Machine Translation* point of view. This paper refers to a joint publication from 2021 (with the same authors) and is an introduction to further works towards practical applications supported by Large Language Models and Machine-Translation.(11) In the paper *«Building Digital Resources for Dharmaśāstra Texts: A Concept Mining System for Indian Knowledge System»* the authors, Arooshi Nigam and Subhash Chandra, present the development of digital resources and methodologies for ancient Sanskrit written tradition (*Dharmaśāstra Texts* corpus).(12) Ifeanyi Ogbodo presented the study *Preserving the Communicative Dynamism of Nigeria's 250+ Languages in the Age of Artificial Intelligence».* This study explores how the over 250 Nigerian languages both contribute and complicate the development of inclusive AI tools (speech recognition, machine translation, chatbots, digital archives, and language learning apps). (13) In their paper *«Digraphia of Maithili: Transliteration of Speech Corpus Transcripts to Tirhuta Script»* its authors, Rajesha N and Narayan Choudhary, present the brief account of the digraphia in Maithili, and pioneering effort of LDC-IL to develop Maithili Sentence Aligned Speech dataset in Tirhuta using transliteration. (14) *«Beyond Wierzbicka: LLM-Powered Analysis of the Natural Semantic Metalanguage Boundaries»* is the paper, written by Michał Szczyszek and Bartosz Naskręcki, where the authors present a novel computational approach to analyzing the boundaries of Anna Wierzbicka's Natural Semantic Metalanguage (NSM) using Large Language Models (LLMs). (15) Yoichi Uetake presents the paper *«Theme-Rheme structure and Nakano-Amari-Kohonen-Hopfield networks».* The paper connects the typed λ calculus based formalization of theme-rheme structure in functional linguistics with a reinterpretation of LLMs as large scale Nakano-Amari-Kohonen-Hopfield networks, highlighting their potential for future AI breakthroughs. (16) In their interdisciplinary article *«AI in Business, Economy and Intellectual Property Law: Challenges, Opportunities and Risks»* the authors, Zygmunt Vetulani, Peter Odrakiewicz and Jakub Kępiński, intend to identify the role of AI technologies in business, economy and legislation challenges and to explain their omnipresent influence. (17) Aniela Walczak, Piotr Kondratowicz and Rafał Jaworski are co-authors of the paper *«Assessing the Sentiment and Value of Product Reviews in Brazilian Portuguese»* 1where they present works on automatic evaluation of written product reviews in Brazilian Portuguese (and where, as a standard, the reviews were classified with respect to the authors' (clients') sentiment(s)). (18) Said Yasin and Torsten Zesch in their paper *«Show, don't tell – Providing Visual Error Feedback for Handwritten Documents»* empirically compare both modular and end-to-end handwriting recognition systems and conclude that these approaches currently do not achieve overall acceptable quality. They identify the major challenges and outline an agenda for future research. (19) In the paper *«Guiding Large Language Models with Their Own Linguistic Knowledge for Morphological Inflection»* the authors, Yifei Zhou and Yves Lepage, propose a two-stage prompting approach that integrates linguistic knowledge into LLMs to improve morphological inflection for low-resource, morphologically-rich languages. (20) Andrzej Zydroń, Rafał Jaworski and Szymon Kaczmarek, the authors of the article *«Using Large Language Models to build an efficient Adaptive Machine Translation System»,* propose their own approach when using Large Language Models for MT, one which allows the system to 'learn on the fly'.

We wish you a pleasant reading !

Poznań, December 2025

Zygmunt Vetulani, Patrick Paroubek and Irakli Kardava
Editors