

LREC 20 *Palma* 26

Conference Programme

Language Resources and Evaluation Conference



Workshops & Tutorials
11–12 & 16 May 2026

Main Conference
13–15 May 2026



15TH EDITION OF THE INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION

PALAU DE CONGRESSOS

PALMA DE MALLORCA, SPAIN

MAY 11 – 16, 2026

CONFERENCE PROGRAMME

For more information about ELRA:

ELRA-ELDA

9, rue des Cordelières

75013 Paris (France)

info@elda.org

LREC 2026 Chairs

General Chair

Stelios Piperidis Athena Research Center/ILSP, Athens, Greece

Program Committee Chairs

Núria Bel Pompeu Fabra University, Barcelona, Spain
Henk van den Heuvel Radboud University, Nijmegen, Netherlands
Nancy Ide Vassar College / Brandeis University, USA
Simon Krek Jožef Stefan Institute / University of Ljubljana, Slovenia
Antonio Toral Universitat d'Alacant, Spain

Management Committee Chairs

Khalid Choukri ELRA, Paris, France
Sara Goggi ILC/CNR, Pisa, Italy
Hélène Mazo ELDA/ELRA, Paris, France

Workshop Chairs

Luis Chiruzzo Universidad de la República, Uruguay
Darja Fiser CLARIN ERIC, The Netherlands
Menno van Zaenen North-West University, South Africa

Tutorial Chairs

Francesca Frontini ILC-CNR, Italy
Atul Kumar Ojha University of Galway, Ireland

Industry Day Chairs

Teresa Lynn Mohamed bin Zayed University of Artificial Intelligence, UAE
Natalie Schluter Apple

Local Organisation Chairs

German Rigau HiTZ, University of the Basque Country, Spain
Maria del Mar Vanrell Universitat de les Illes Balears, Spain
Melania Sánchez Masià Universitat de les Illes Balears, Spain

Rafael Muñoz Guillena

Universidad de Alicante, Spain

Publication Chairs

John McCrae

University of Galway, Ireland

Gilles Sérasset

Université Grenoble Alpes, France

Sponsorship Chairs

Patrick Paroubek

Université Paris-Saclay, France

Alexandros Poulis

Transperfect/Dataforce, Luxembourg

Jimmy Kunzmann

Retired Amazon Alexa Hybrid Science

Raquel Lázaro Gutiérrez

Universidad de Alcalá, Spain

Exhibition Chair

Rafael Muñoz Guillena

Universidad de Alicante, Spain

Publicity Chairs

Elena Volodina

University of Gothenburg, Sweden

Rafael Valencia García

Universidad de Murcia, Spain

A. Seza Doğruöz

Ghent University, Belgium

Ethics Chairs

Denise DiPersio

Linguistic Data Consortium, USA

Thomas Margoni

KU Leuven, Belgium

Claudia Cevenini

Università di Bologna, Italy

Diversity and Inclusion Chair

Sakriani Sakti

Nara Institute of Science and Technology, Japan

Advisor

Nicoletta Calzolari

Scientific Committee, Senior Area Chairs and Area Chairs

The Programme Chairs and the LREC 2026 organizers are very grateful to the Senior Area Chairs, the Area Chairs and the members of the Scientific Committee who reviewed the submissions and contributed to designing the conference programme.

The list of the Senior Area Chairs, the Area Chairs and the members of the Scientific Committee is published on the LREC 2026 web site and the LREC 2026 Proceedings.

Organizing Team at ELDA and ILC-CNR Pisa

Caroline Merlieux-Petit	ELDA, Paris, France
Alexandre Sicard	ELDA/ELRA, Paris, France
Fernanda González Campo	ELDA, Paris, France
Valérie Mapelli	ELDA, Paris, France
Inae Song	ELDA, Paris, France
Florian Philippe	ELDA, Paris, France
Vincenzo Parinelli	ILC-CNR, Pisa, Italy
Gabriele Chignoli	ELDA, Paris, France
Shi Yu	ELDA, Paris, France
Kossay Talmoudi	ELDA, Paris, France

Acknowledgements

The European Language Resources Association, ELRA, and all the LREC Committees acknowledge with gratitude the support and sponsoring of the following institutions.

Sponsors and supporters

Gold Sponsors

- Transperfect DataForce
- Mozilla Data Collective

Silver Sponsors

- HITZ
- OpenEuroLLM

Bronze Sponsors

- ALT-EDIC
- Language Data Space (LDS)
- Multiverse Computing
- Reverso
- Translated

Supporters

- LinguisList

Publisher

- The CJK Dictionary Institute, Inc.

Supporting Institutions

ELDA; Paris, France

ILC-CNR, Pisa, Italy

UIB, Palma (Balearic Islands)

Workshops and Tutorials Schedule

		CABRERA 2	CABRERA 1	CALVIA	EIVISSA 1	EIVISSA 2	ELYXIR
May 11	AM	W40 SPEAKABLE Speech Language Models in Low- Resource Settings	W24 CMLC-12 12th Workshop on Challenges in Management of Large Corpora	W32 BUCC 19th Workshop on Building and Using Comparable Corpora	W43 LLMs4SSH Shaping multilingual, multimodal AI for social sciences & humanities	W36 SIGUL 2026 Towards Inclusivity: LR&T for Under- Resourced & Endangered Languages	W05 DMR 2026 7th Int'l Workshop on Designing Meaning Representations
	PM		W53 HTRes-2026 Holocaust Testimonies as Language Resources				

May 12	AM	W29 LDL 2026 10th Workshop on Linked Data in Linguistics	W51 CAWL 2026 3rd Workshop on Computation and Written Language	W18 ISA 2026 22nd Joint ACL- ISO Workshop on Interoperable Semantic Annotation	W20 Social Context The Social Context Workshop	W36 SIGUL 2026 Towards Inclusivity: LR&T for Under- Resourced & Endangered Languages	T15 Tutorial Hallucination in Large Foundation Models
	PM	W13 Resources & Processing of Data from People with Cognitive/Psychiatric Impairments	W04 NLP4Ecology 2026 2nd Workshop on Ecology, Environment, and NLP	W34 DTF Leveraging Derived Text Formats to Unlock Copyrighted Collections for Open Science		T18 Tutorial Multilingual & Multimodal LLMs in the Wild: Building for Low- Resource Languages	W45 InDor The Information Disorder Workshop

May 16	AM	W01 FNP 2026 7th Financial Narrative Processing Workshop	W14 DialRes Dialects in NLP : A Resource Perspective	W33 PoliticalNLP 2026 3rd Workshop on NLP for Political Sciences	W35 KG&LLM Knowledge Graphs and Large Language Models	W16 UDW 2026 9th Workshop on Universal Dependencies	W06 CMCL 2026 15th Workshop on Cognitive Modeling and Computational Linguistics
	PM		W21 ChiPSAL-2026 2nd Workshop on Challenges in Processing South Asian Languages	W27 Neology & LLM Neology and Large Language Models			

Workshops and Tutorials Schedule

FORMENTERA	MENORCA 1	MENORCA 2	PORTIXOL 1	PORTIXOL 2	PORTOPI	SANTA CATALINA
W37 Non-Literal Learning Non-Literal Expressions with Small Data	W30 SLIDE Workshop on Structured Linguistic Data	W12 LT4HALA 2026 4th Workshop on LT for Historical and Ancient Languages	T10 FAIRTAG4ALL FAIR Terminology-Augmented Generation for All	W55 RESOURCEFUL 2026 Role of Resources in the Age of Large Language Models	W03 OSACT7 7th Workshop on Open-Source Arabic Corpora & Processing Tools	W49 READixTSAR READixTSAR
W17 DeTermIt! 2026 2nd Workshop on Evaluating Text Difficulty in a Multilingual Context			T14 Tutorial Metaphor Processing: Methods, Theories and Resources		W08 NakbaNLP 2026 2nd Int'l Workshop on Nakba Narratives as Language Resources	T12 Tutorial Biases we Research by

W15 NSLP 2026 3rd Int'l Workshop on Natural Scientific Language Processing	W44 LEGAL2026 & CALD-pseudo Joint Workshop on Legal & Ethical Issues in HLT + Data Pseudonymization	W26 NLPerspectives 5th Workshop on Perspectivist Approaches to NLP	W42 Gaze4NLP Eye-Tracking Resources & Evaluation for Human-Aligned NLP	T13 HealthcareNLP HealthcareNLP: where we are and what's next?	W47 WILDRE-8 8th Workshop on Indian Language Data: Resources & Evaluation	W02 CL4Health 3rd Workshop on Patient-oriented Language Processing
			T6 Tutorial Linguistic Creativity in the Age of LLMs: From Diversity to Innovation	W31 RAIL 7th Workshop on Resources for African Indigenous Languages	T7 Tutorial Distributed Compositionality: Formal Foundations for Distributional Semantics	

W28 PressMint 1st Workshop on Creating Interoperable Corpora of Historical Newspapers	W39 SignLang 12th Workshop on Representation & Processing of Sign Languages: Language in Motion	W07 Clinical NLP 8th Clinical NLP Workshop	T2 Tutorial Low-Resource, High-Impact: Building Corpora for Inclusive Language Technologies	W46 Identity-Aware NLP Ethical & Technical Challenges for Identity-Aware NLP	T11 Tutorial Meaning Representation in NLP: Foundations, Frameworks, and Evaluation	LANNLP LANNLP Bridging Ibero and Latin American NLP Communities
W38 ParlaCLARIN V Workshop on Interoperability, Multilinguality & Multimodality in Parliamentary Corpora		W25 CAS 2026 1st Workshop on Computational Affective Science	T3 Tutorial Under-Resourced Studies of Under-Resourced Languages: LLM-as-Annotator Pipelines	W48 DELITE 2026 2nd Workshop on Language-driven Deliberation Technology	T4 Tutorial From Shared Tasks to Benchmark Datasets: Lessons from NTCIR's 26-Year Journey	

Conference Programme at a glance, May 13

		AUDITORIUM ILLES BALEARS	AUDITORIUM MALLORCA	MENORCA 1	EIVISSA 1	POSTER AREA
9:30	11:00	Opening Session (AUDITORIUM ILLES BALEARS)				
11:00	11:20	Coffee break				
11:20	11:40	O1 Dialogue, Conversational Systems, Chatbots, Human-Robot Interaction	O2 Interpretability, Explainability I	O3 Document Classification, Information Retrieval and Cross-lingual Retrieval	O4 Evaluation, Validation, Quality Assurance and Benchmarking Methodologies	Poster Area 1 P1.1 Applications: Datasets and Benchmarks P1.2 Applications: LLMs P1.3 Applications P1.4.1 Digital Humanities I P1.4.2 Digital Humanities II P1.5 Simplification, Accessibility P1.6 Infrastructures, Policy and Legal Issues I
11:40	12:00					
12:00	12:20					
12:20	12:40					
12:40	13:00					
13:00	14:30	Lunch				
14:30	15:15	Keynote Speaker 1: Nancy Chen (AUDITORIUM ILLES BALEARS) - Chair: Simon Krek				
15:15	15:20	Break to change rooms				
15:20	15:40	O5 Inference, Reasoning, Question Answering I	O6 Information Extraction and Text Mining I	O7 Language Modeling and LRs I	O8 Less-Resourced/Endangered/Less-studied Languages	Poster Area 2 P2.1.1 Corpora and Treebanks I P2.1.2 Corpora and Treebanks II P2.1.3 Corpora and Treebanks III P2.2 Discourse and Pragmatics I P2.3.1 Interpretability, Explainability II P2.3.2 Interpretability, Explainability III
15:40	16:00					
16:00	16:20					
16:20	16:40					
16:40	17:00					
17:00	17:20	Coffee break				
17:20	17:40	O9 Corpora, Treebanks and Annotation; Tools, Systems and Platforms	O10 Information Extraction and Text Mining II	O11 Language Modeling and LRs II	O12 Applications Involving LRs and Evaluation I	Poster Area 1 P3.1.1 Dialogue, Conversational Systems I P3.1.2 Dialogue, Conversational Systems II P3.2.1 Less-Resourced/Studied Languages I P3.2.2 Less-Resourced/Studied Languages II P3.2.3 Less-Resourced/Studied Languages III
17:40	18:00					
18:00	18:20					
18:20	18:40					
18:40	19:00					
19:00	20:00	ELRA Members Meeting				
20:00		Welcome Reception				

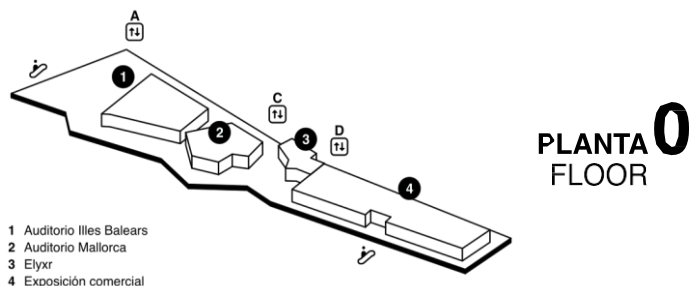
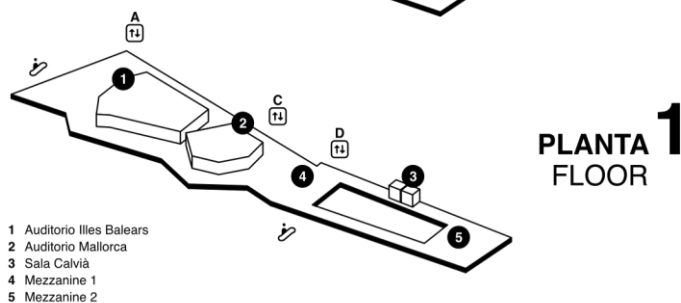
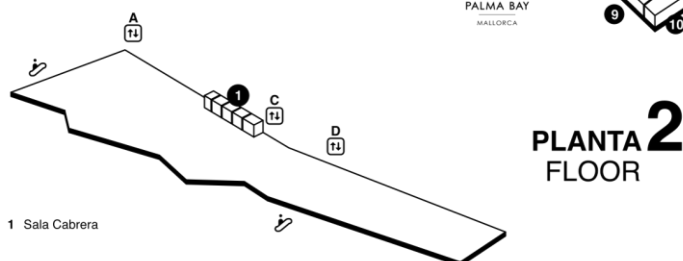
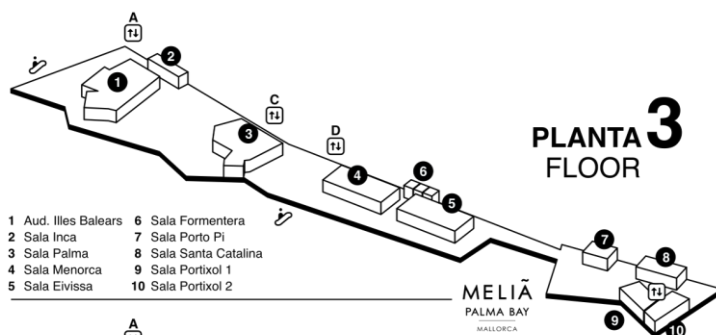
Conference Programme at a glance, May 14

		AUDITORIUM ILLES BALEARS	AUDITORIUM MALLORCA	MENORCA 1	EIVISSA 1	EIVISSA 2	POSTER AREA
9:00	09:20	O13 Digital Humanities and Related Corpora	O14 Lexicon	O15 Multilinguality, Machine Translation	O16 Natural Language Generation and Summarization	Industry Day	Poster Area 2 P4.1.1 Bias, Offensive Content, Guardrails I P4.1.2 Bias, Offensive Content, Guardrails II P4.2.1 Evaluation, Validation I P4.2.2 Evaluation, Validation II P4.2.3 Evaluation, Validation III
9:20	09:40						
9:40	10:00						
10:00	10:20						
10:20	10:40						
10:40	11:00	Coffee Break					
11:00	11:20	O17 Evaluation, Validation IV	O18 Lexicon and Semantics I	O19 Multilinguality, Machine Translation Evaluation	O20 Discourse and Pragmatics II	Industry Day	Poster Area 1 P5.1.1 Inference, Reasoning, Question Answering II P5.1.2 Inference, Reasoning, Question Answering III P5.2.1 Speech Resources and Processing I P5.2.2 Speech Resources and Processing II P5.2.3 Speech Resources and Processing III
11:20	11:40						
11:40	12:00						
12:00	12:20						
12:20	12:40						
12:40	13:20	Antonio Zampolli Prize Talk (AUDITORIUM ILLES BALEARS) - Chair: German Rigau					
13:20	14:45	Lunch					
14:45	15:15	Invited Local Talk: Nicolau Dols (AUDITORIUM ILLES BALEARS) - Chair: Antonio Toral					
15:15	15:20	Break to change rooms					
15:20	15:40	O21 Evaluation, Validation V	O22 Information Extraction and Text Mining III	O23 Simplification, Plain Language	O24 Machine Learning I	Industry Day	Poster Area 2 P6.1.1 Corpora and Treebanks IV P6.1.2 Corpora & Treebanks V P6.1.3 Corpora & Treebanks VI P6.1.4 Corpora & Treebanks VII
15:40	16:00						
16:00	16:20						
16:20	16:40						
16:40	17:00						
17:00	17:20	Coffee Break					
17:20	17:40	O25 Corpora, Treebanks and Annotation	O26 Named Entity Recognition, Speech Resources	O27 Simplification, Plain Language and Assistive Technologies	O28 Applications Involving LRs and Evaluation II	Industry Day	Poster Area 1 P7.1 Document Classification P7.2.1 Information Extraction and Text Mining IV P7.2.2 Information Extraction and Text Mining IV P7.3 Knowledge Representation and Graphs P7.4 Opinion, Sentiment, Emotion Analysis P7.5 Argument Mining and Emotion Classification
17:40	18:00						
18:00	18:20						
18:20	18:40						
18:40	19:00						

Conference Programme at a glance, May 15

		AUDITORIUM ILLES BALEARS	AUDITORIUM MALLORCA	MENORCA 1	EIVISSA 1	POSTER AREA
9:00	9:20	O29 Infrastructures, Policy and Legal Issues II	O30 Opinion and Argument Mining, Sentiment Analysis	O31 Bias, Offensive and Non-inclusive Language	O32 Speech Resources, Processing, Applications	Poster Area 2 P8.1.1 Machine Translation I P8.1.2 Machine Translation II P8.2 Multilinguality and Translation Aids P8.3 Multimodality P8.4 Cross-modality P8.5 Sign Languages
9:20	9:40					
9:40	10:00					
10:00	10:20					
10:20	10:40					
10:40	11:00	Coffee break				
11:00	11:20	O33 Psycholinguistics, Cognitive Linguistics and Linguistic Theories	O34 Opinion and Argument Mining	O35 Parsing	O36 Multimodality and Speech	Poster Area 1 P9.1 Natural Language Generation P9.2.1 Machine Learning II P9.2.2 Machine Learning III P9.3.1 Language Modeling and LRs III P9.3.2 Language Modeling and LRs IV P9.3.3 Language Modeling and LRs V
11:20	11:40					
11:40	12:00					
12:00	12:20					
12:20	12:40					
12:40	14:10	Lunch				
14:10	14:55	Keynote Speaker 2: Dan Jurafsky (AUDITORIUM ILLES BALEARS) - Chair: Nancy Ide				
14:55	15:00	Break to change rooms				
15:00	15:20	O37 Evaluation, Validation, Quality Assurance	O38 Knowledge Discovery and Representation	O39 Applications Involving LRs and Evaluation III	O40 Multimodality, Cross- modality	Poster Area 2 P10.1 Social Media P10.2.1 Linguistics and Psycholinguistics I P10.2.2 Linguistics and Psycholinguistics II P10.3.1 Parsing and Tagging I P10.3.2 Parsing and Tagging II P10.4.1 Lexicon and Semantics II P10.4.2 Lexicon and Semantics III
15:20	15:40					
15:40	16:00					
16:00	16:20					
16:20	16:40					
16:40	17:00	Coffee break				
17:00	18:20	Closing Session				
20:00		Gala Dinner				

Venue Map



LREC 2026 Speakers



Nancy F. Chen

A*STAR

Keynote Talk

Wednesday, 13 May, 2026

14:30 - 15:15

Auditorium Illes Balears

The Social Failures of Language Models as Conversational Partners

Abstract: Nancy Chen traces how language resources have evolved across AI's major eras, from hand-crafted linguistic annotations, through large-scale statistical datasets, to multimodal pretraining, and now toward *alignment*, where data shapes not just model capabilities but also societal behaviour. She argues that these resources have always reflected the scientific and institutional priorities of their time (DARPA, LDC, ELRA, etc.), and that today's convergence of speech, text, and multimodality raises urgent questions around cultural representation and responsible deployment. She illustrates this with three real-world systems from her lab: MERaLION (a multimodal LLM for Southeast Asia), SingaKids AI Tutor (a multilingual agent for children), and Siu Dai (a telebot for chronic disease patients).

Short bio: ISCA Fellow and inaugural A*STAR Fellow, Nancy Chen leads Multimodal Generative AI research at I2R (A*STAR, Singapore). Her lab advances multilingual LLMs for agentic reasoning, inclusive AI, and AI-augmented learning, with spin-offs deployed at Singapore's Ministry of Education. She has won 10 Best Paper Awards, delivered 40+ keynotes, and served as Program Chair for NeurIPS 2025 and ICLR 2023.



Nicolau Dols

Universitat de les Illes
Balears

Invited Local Talk

Thursday, 14 May, 2026

14:45 - 15:15

Auditorium Illes Balears

Catalan in Majorca: main features, challenges, and research tools

Abstract: Catalan has been the language of the Balearics since the 13th century. After 1715, Catalan was gradually replaced by Spanish in official usage, while remaining the main everyday language in a mostly monolingual society. This situation was reversed at the end of the 20th century, when Catalan regained official status (together with Spanish), although monolinguals in Catalan no longer existed. Nicolau Dols plans to produce a general overview of Catalan, focusing on the challenges it faces to maintain its vitality and relevance. In addition, issues related to digital resources for research on Catalan, both written and oral, will be presented, with special attention to corpora, dictionaries, and grammars.

Short bio: Full professor of Catalan at the Universitat de les Illes Balears and Chair of the Philological Section of the Institut d'Estudis Catalans (the official academy for Catalan). His research spans phonology, grammar, and sociolinguistics, and he leads the team building the oral Catalan language corpus at the IEC.



Dan Jurafsky
Stanford University

Keynote Talk

Friday, 15 May, 2026

14:10 - 14:55

Auditorium Illes Balears

The Social Failures of Language Models as Conversational Partners

Abstract: Language models are increasingly used in conversation for information, advice, and emotional support. In this talk I'll summarize studies in our lab showing that models fail in systematic ways as social interlocutors. We find that language models are socially sycophantic, linguistically overconfident, overly anthropomorphic, and epistemically self-centered. We then show that these flaws have real consequences for users: people interacting with models suffer consequences including overreliance, distorted judgment, and reduced personal responsibility. I'll discuss datasets and metrics, explore mitigations, and call for design, evaluation, and accountability mechanisms to protect user well-being.

Short bio: Dan Jurafsky is a Stanford professor of Linguistics and Computer Science, with affiliations in the humanities. A MacArthur Fellow and member of the American Academy of Arts and Sciences, he researches natural language processing and its applications across cognitive, linguistic, and social sciences. He co-authored the widely used textbook *Speech and Language Processing* and wrote *The Language of Food* (2014), a 2014 international bestseller and James Beard Award nominee.

Wednesday, 13 May 2026

11:00

Chair: Stelios Piperidis

11:20

11:00	Coffee Break
11:20	

O1 • Dialogue, Conversational Systems, Chatbots, Human-Robot Interaction

Chair: Frédéric Béchet

11:40 Yen-Yu Chang, Daryna Dementieva, Alexander Fraser

Kazuya Matsuo, Atsushi Otsuka, Narichika Nomoto, Makoto Nakatsuji

Iva Bojic, Mathieu Ravaut, Stephanie Hilary Xinyi Ma, Doreen Tan, Andy Hau Yan Ho, Andy Khong

Priyanshu Priya, Hardik Goyal, Asif Ekbal

13:00 *Sofia Brenna, Elisabetta Jezek, Matthias Kraus, Bernardo Magnini*

O2 • Interpretability, Explainability I

Chair: Andrey Kutuzov

12:40 Prompts with What They Represent in Their Embeddings

11:40 Why So Separate: Analyzing In-Context Learning from a Vector Space Perspective

11:20	Investigating How LLMs Propagate Female Stereotypes: Comparing What Models Say via
12:40	Prompts with What They Represent in Their Embeddings
	<i>Andrea Valderrey Nuñez and Jelke Bloem</i>
11:40	Why So Separate: Analyzing In-Context Learning from a Vector Space Perspective
12:00	<i>Tobias Kalmbach and Sandipan Sikdar</i>

12:00	Explaining Explanations: Interpretability Methods for Discourse Analysis of Transformer
12:20	Attention Maps
	<i>Louis Escoufflaire, Jérémie Bogaert, Antonin Descampe, Cédric Fairon, Francois-Xavier Standaert</i>
12:20	TempPerturb-Eval: On the Joint Effects of Internal Temperature and External Perturbations
12:40	in RAG Robustness
	<i>Yongxin Zhou, Philippe Mulhem, Didier Schwab</i>
12:40	Refusal Steering: Fine-grained Control over LLM Refusal Behaviour for Sensitive Topics
13:00	<i>Iker García-Ferrero, David Montero, Roman Orus</i>

03 • Document Classification, Information Retrieval and Cross-lingual Retrieval

Menorca 1

Chair: William Hersh

11:20	To Predict or Not to Predict? Towards Reliable Uncertainty Estimation in the Presence of
11:40	Noise
	<i>Nouran Khallaf and Serge Sharoff</i>
11:40	An Extreme Multi-label Text Classification (XMTC) Library Dataset: What If We Took "Use of
12:00	Practical AI in Digital Libraries" Seriously?
	<i>Jennifer D'Souza, Sameer Sadruddin, Maximilian Kaehler, Andrea Salfinger, Luca Zaccagna, Francesca Incitti, Lauro Snidaro, Osmo Suominen</i>
12:00	A Historical Database for the Study of Obstruent-Lateral Palatalization in Ibero-Romance
12:20	<i>Andrea García Covelo</i>
12:20	Is Clinical Text Enough? A Multimodal Study on Mortality Prediction in Heart Failure Patients
12:40	<i>Oumaima El Khettari, Virgile Barthet, Guillaume Hocquet, Joconde Weller, Emmanuel Morin, Pierre Zweigenbaum</i>
12:40	HistoriQA-ThirdRepublic: Multi-Hop Question Answering Corpus for Historical Research,
13:00	Parliamentary Debates from the French Third Republic (1870-1940)
	<i>Aurelien Pellet, Marie Puren, Julien PEREZ</i>

04 • Evaluation, Validation, Quality Assurance and Benchmarking Methodologies

Eivissa 1

Chair: Roman Klinger

11:20	Assessing the Political Fairness of Multilingual LLMs: A Case Study Based on a 21-Way
11:40	Multiparallel EuroParl Dataset
	<i>Paul Lerner and François Yvon</i>
11:40	AfriStereo: A Culturally Grounded Dataset for Evaluating Stereotypical Bias in Large
12:00	Language Models
	<i>Yann Le Beux, Oluchi Audu, Oche Ankele, Dhananjay Balakrishnan, Melissah Weya, Marie Ralaiarinosy, Ignatius Ezeani</i>
12:00	Judging Instruction Responses in a Low-Resource Language: A Case Study on Basque
12:20	<i>David Ponce, Harritxu Gete, Thierry Etchegoyhen, Irune Zubiaga, Aitor Soroa</i>

12:20	Appeal, Align, Divide? Stance Detection for Group-Directed Messages in German
12:40	Parliamentary Debates <i>Ines Rehbein, Maris Buttmann, Julian Schlenker, Simone Paolo Ponzetto</i>
12:40	BURMESE-SAN: Burmese NLP Benchmark for Evaluating Large Language Models
13:00	<i>Thura Aung, Jann Montalan, Jian Ngui, Peerat Limkonchotiwat</i>

11:20 **Poster Sessions - Poster Area 1**
13:00

P1.1 • Applications: Datasets and Benchmarks

Poster Area 1

Chair: Andreas Witt

Report-based Recommendations for Policy Making and Agency Operations: Dataset and LLM Evaluation

Aleksandra Edwards, Thomas Edwards, Jose Camacho-Collados, Alun Preece

ConceptKT: A Benchmark for Concept-Level Deficiency Prediction in Knowledge Tracing
Yu-Chen Kang, Yu-Chien Tang, An-Zi Yen

Open-access Dataset on Acceptability Ratings of Korean Clausal Constructions by Humans and GPT Models

Gyu-Ho Shin, Soo-Hwan Lee, Chanyoung Lee

Talk2Ref: A Dataset for Reference Prediction from Scientific Talks

Frederik Broy, Maïke Züfle, Jan Niehues

MuSaG: A Multimodal German Sarcasm Dataset with Full-Modal Annotations

Aaron Scott, Maïke Züfle, Jan Niehues

Icelandic Math Eval: A Competitive Mathematics Benchmark for Large Language Models

Hafsteinn Einarsson, Jökull Haraldsson, Ívar Derayat, Sigrún Lund, Benedikt Magnússon

MazeEval: A Benchmark for Testing Sequential Decision-Making in Language Models

Hafsteinn Einarsson

J-ClinicalBench: A Benchmark for Evaluating Large Language Models on Practical Clinical Tasks in Japanese

Seiji Shimizu, Tomohiro Nishiyama, HISADA Shohei, Yamato Himi, Shoko Wakamiya, Yuki Yanagisawa, Masami Tsuchiya, Satoko Hori, Eiji ARAMAKI

Is One Dataset Enough for Evaluation? Studying Generalizability of Automated Essay Scoring Models

Sohaila Eltanbouly, Marwan Sayed, Tamer Elsayed

HiFi-KPI: A Dataset for Hierarchical KPI Extraction from Earnings Filings

Rasmus Jensen, Giovanni Rizzi, Rasmus Tjalk-Bøggild, Alexandre Iolov, Mike Zhang, Johannes Bjerva

UniSkill: A Dataset for Matching University Curricula to Professional Competencies

Nurlan Musazade, József Mezei, Mike Zhang

A Dataset for Evaluating ASR on Specialized Vocabulary

Emily Haubert Klering, Eduardo Cortes, Tatjana Chernenko, Mariana Vargas Trarbach, Gabriel de Oliveira Ramos, Sandro José Rigo, Maitê Dupont, Ana Treichel Vianna, Gabriela Krause dos Santos, Vinicius Meirelles Pereira, Denis de Araujo, Rafael Kunst

SommBench: Assessing Sommelier Expertise of Language Models

William Brach, Tomas Bedej, Jacob Nielsen, Jacob Pichna, Juraj Bedej, Eemeli Saarensilta, Julie Dupouy, Gianluca Barmina, Andrea Blasi Núñez, Peter Schneider-Kamp, Kristian Košťál, Michal Ries, Lukas Galke Poech

CzechDocs: A Multiway Parallel Dataset of Formatted Documents for Minority Languages in Czechia

Josef Jon and Ondřej Bojar

P1.2 • Applications: LLMs

Poster Area 1

Chair: Parameswari Krishnamurthy

An LLM-Based Assistant for Debt Waiver Court Procedures

Lluís Padro, Daniel Ferrés, Roser Saurí, Mireia Artigot

Enhancing Clinical Trial Analysis through Large Language Models for Multi-Evidence Natural Language Inference

Shobanapriyan Chandrasegaran and Amal Htait

A Systematic Comparison of Large Language Models for Data Annotation in NER Tasks

Muhammad Uzair Ul Haq, Davide Rigoni, Alessandro Sperduti

Toward Generalized Cross-Lingual Hateful Language Detection with Web-Scale Data and Ensemble LLM Annotations

Dang Dang, Jelena Mitrović, Michael Granitzer

Can LLMs Faithfully Explain Themselves in Low-Resource Languages? A Case Study on Emotion Detection in Persian

Mobina Mehrazar, Mohammad Amin Yousefi, Parisa Beygi, Behnam Bahrak

Are LLMs Good Text Diacritizers? An Arabic and Yoruba Case Study

Hawau Olamide Toyin, Samar Mohamed Magdy, Hanan Aldarmaki

Automatic Suggestions of Supplements in the Herculaneum Papyri: Language Models and RESTful API

Angelo Mario Del Grosso, Gabriele Giannesi, Simone Zenzaro, Federico Boschetti

Designing LLM Agents for User-Centered Language Service Selection

Ryoichiro Ogawa, Donghui Lin, Fumito Uwano

User Profiling for Specification-Sensitive Recommendations with Large Language Model Prompting

Chih-Yu Chien, An-Zi Yen, Hen-Hsen Huang, Hsin-Hsi Chen

Comparing Traditional and LLM-based Approaches for Automated Scoring of Dutch Writing Products

Joni Kruijsbergen and Orphee De Clercq

“Decode the Law”: Towards Legal Text Simplification with Large Language Models

Mohammed Rabbani, Subhadeep Roy, Sayantan Mitra, Tulika Saha

P1.3 • Applications

Poster Area 1

Chair: Mark Finlayson

CLASE: A Hybrid Method for Chinese Legalese Stylistic Evaluation

Yiran Ma, Yuxiao Ye, Huiyuan Xie

Neural Network-assisted Analysis of Tube Vocal Tract Models

Runhui Song, Johan Sjons, Axel Ekstrom

Central Kurdish Text-to-Speech and Its Application in Speech-to-Text Translation

Mohammad Mohammadamini, Meysam Shamsi, Marie Tahon

QuALA-NL: Question & Answer with Legal Attribution in Dutch

Romy van Drie, Roos Bakker, Daan Di Scala, Maaïke de Boer

SouDeC: Source Detection and Classification in Czech

Jiří Mirovský and Barbora Hladka

Frame Semantic Patterns for Identifying Underreporting of Notifiable Events in Healthcare:
The Case of Gender-Based Violence

Livia Dutra, Arthur Lorenzi, Luis Berno, Franciany Campos, Karoline Biscardi, Kenneth Brown, Marcelo Viridiano, Frederico Belcavello, Ely Matos, Olivia Guaranha, Erik Santos, Sofia Reinach, Tiago Timponi Torrent

PrePPER: A Preference Pattern-based Profiling Framework for Explainable Recommendation

Taisuke Usumi, Akiko Masaki, Sanae Muramatsu, Akira Sakamoto, Takeharu Eda

Evaluating the Impact of Source Diversity for RAG in Historical Research

Ruhi Mahadeshwar, Andreas van Cranenburgh, Tommaso Caselli, Malvina Nissim

Automatic Essay Scoring and Feedback Generation in Basque Language Learning

Ekhi Azurmendi, Xabier Arregi, Oier Lopez de Lacalle

Paragraph Segmentation Revisited: Towards a Standard Task for Structuring Speech

Fabian Retkowski and Alexander Waibel

High-Order Question Generation in a Multilingual Educational Context

Suna Uçar, Itziar Aldabe, Nora Aranberri, Orphee De Clercq

From Print to Digital and beyond: The Retrodigitization of a Historical Dictionary of Italian as a Hybrid Lexical Resource

Marco Biffi, Sebastiana Cucurullo, Manuel Favaro, Elisa Guadagnini, Simonetta Montemagni, Eva Sassolini

Learning through News: Bridging the Gap between Algorithmic Recommendation and Human Curation

Florian Debaene, Loïc De Langhe, Orphee De Clercq, Veronique Hoste

MaskedVerbalizer: Automatic Verbalizer Construction for Few-Shot Text Classification in Low-Resource Right-to-Left Languages

Faizad Ullah, Furqan Sikandar, Areeba Waqar, Faizan Ali, Muhammad Sohaib Ayub, Mubashar Mushtaq, Asim Karim

RBR: RAG-Based Open-Domain Question Answering Using a Ranking Approach to Document Retrieval

Priyatam Naravajhula and Vincent Ng

Sentence-Level Back-Transliteration of Romanized Indian Languages: Performance Analysis and Challenges

Saurabh Kumar, Dhruvkumar Kakadiya, Sanasam Ranbir Singh, Sukumar Nandi

Cross-Corpus CEFR Classification through Artificial Learners Perplexities

bernardo stearns, John Mccrae, Thomas Gaillat

P1.4.1 • Digital Humanities I

Poster Area 1

Chair: Natalie Parde

CorpusClues: Scalable Unsupervised Similarity Search for Historical Texts Using MinHash-LSH

Paulien Lemay, Klaas Bentein, Els Lefever

BenCSSmark: Making the Social Sciences Count in LLM Research

Arnault Chatelain, Etienne Ollion, Qianwen Guan, Diandra Fabre, Lorraine Goeuriot, emile chapuis, Abdelkrim Beloued, Marie Candito, Nicolas Hervé, Didier Schwab

Predicting Topic (Co-)Occurrence Using Topic Networks Built from the Project Gutenberg Corpus

Bhuvanesh Verma and Alexander Mehler

AraHopeCorpus: Annotation Guidelines and Dataset for Hope Speech in Arabic Social Media Crisis Discourse

Esra'a Sharqawi and Wajdi Zaghouni

Cohesion-6K: An Arabic Dataset for Analyzing Social Cohesion and Conflict in Online Discourse

Aisha Al-Athba and Wajdi Zaghouni

Reference-free Evaluation at Inference for NER/NEL over OCRed Historical Texts

Tien-Nam Nguyen, Adam Jatowt, Ahmed Hamdi, Mickael Coustaty, Thi Hong Hanh Tran, Antoine Doucet

Echoes of the Troubadours: A Corpus of Troubadour Poetry for Stylometric Analysis and Authorship Attribution

Loic De Langhe, Orphee De Clercq, Veronique Hoste

Gretino: A Greek and Latin Dataset to Benchmark Retrieval Systems in Classical Languages

Hawau Olamide Toyin, Federico Iezzi, Elia Scapini, Giulio Federico, Giovanni Puccetti

A Recipe for Adapting Multilingual Embedders to OCR-Error Robustness and Historical Texts

Andrianos Michail, Stylianos Psychias, Juri Opitz, Simon Clematide

Phrase-Level Segmentation on Medieval Corpora for Aligning Multilingual Texts

Lucence Ing, Matthias Gille Levenson, Carolina Macedo

P1.4.2 • Digital Humanities II

Poster Area 1

Chair: Paul Rayson

RAGE: Roman and Greek Emotions

Frederick Riemenschneider, Jonathan Geiger, Thomas Kuhn-Treichel, Anette Frank

From Variance to Invariance: Qualitative Content Analysis for Narrative Graph Annotation
Junbo Huang, Max Weinig, Ulrich Fritsche, Ricardo Usbeck

A Dataset of Historical Medical Periodicals Annotated with Textual Genre
Vera Danilova and Sara Stymne

Preserving Endangered Linguistic Heritage: Developing a Corpus for the Study of Contact-induced Changes in Corfioto
Giorgio Maria Di Nunzio and Georgios Vardakis

To Eat and beyond: A FrameNet-Inspired Annotation of Food and Its Uses over Time
Teresa Paccosi, Gauri Bhagwat, Marieke van Erp

To Overfit or Not to Overfit? An Evaluation of HTR Workflow on 17Th-18Th Century French Corpus
Marine Tiger

Automatic Segmentation of Classical Tibetan Texts into Autochthonous and Allochthonous Regions
Guy Bilitski, Lev Shechter, Sonam Jamtsho, Nir Marciano, Nicola Bajetta, Rebecca Sunden, Omri Drori, Kai Golan Hashiloni, Orr Zwebner, Asaf Shina, Orna Almogi, Dorji Wangchuk, Kfir Bar

RespondeoQA: A Benchmark for Bilingual Latin-English Question Answering
Marisa Hudspeth, Patrick Burns, Brendan O'Connor

Transformer-Enabled Diachronic Analysis of Vedic Sanskrit: Neural Methods for Quantifying Types of Language Change
Ananth Hariharan and David R. Mortensen

Ithaca Revisited: Benchmarking a Domain-Specific Model for Epigraphy in the Age of LLMs
Alessandro Locaputo, Andrea Brunello, Nicola Saccomanno, Paraskevi Platanou, Giuseppe Serra

P1.5 • Simplification, Accessibility

Poster Area 1

Chair: Seid Yimam

CEFR Level Prediction for Short Russian L2 Texts: Evaluating Classifiers and Instruction-Based LLMs
Anna Glazkova, Antonina Laposhina, Dmitry Morozov

Evaluation of Document-Level Text Simplification in Japanese
Iori Yamashita, Hikari Tanaka, Hajime Kiyama, Kexin Bian, Zhousi Chen, Mamoru Komachi

Parallel Corpus Filtering Based on Semantic Similarity and Surface Dissimilarity for Japanese Text Simplification with LLMs

Daisuke Maekawa, Tomoyuki Kajiwara, Takashi Ninomiya

A Multilingual Human Annotated Corpus of Original and Easy-to-Read Texts to Support Access to Democratic Participatory Processes

Verena Riegler, Stefan Bott, Horacio Saggion, Almudena Rascón Alcaina, Nouran Khallaf

Proffiliadur: Welsh Language Text Profiling Toolkit

Nicolás Gutiérrez-Rolón, Jonathan Davies, Tomos Williams, Dawn Knight, Fernando Alva-Manchego

Recovering Registers from Leveled Wordlists

Yo Ehara

P1.6 • Infrastructures, Policy and Legal Issues I

Poster Area 1

Chair: Davide Picca

Fill-in-the-Blanks: Automatic Generation and Evaluation of Language Models' Pseudonyms for English and Swedish Texts

Maria Irena Szawerna and Jacob Suchardt

Integrating Services, Platforms and Resources into a National Infrastructure Cluster for FAIR Language and Cultural Data

Giulia Pedonese, Daniele Melaccio, Michele Mallia, Monica Monachini, Francesca Frontini, Valeria Quochi, Fahad Khan, Angelo Mario Del Grosso, Federico Boschetti, Riccardo Del Gratta

Common European Language Data Space: Development, Current Status, and Future Perspectives

Stelios Piperidis, Penny Labropoulou, Dimitrios Galanis, Khalid Choukri, Andrejs Vasiljevs, Mitos Deligiannis, Katerina Gkirtzou, Dimitris Gkoumas, Athanasia Kolovou, Leon Voukoutis, Kanella Pouli, Maria Giagkou, Maria Gavrilidou, Katrin Marheinecke, Elena Leitner, Simon Ostermann, Stefania Raccioppa, Kossay Talmoudi, Victoria Arranz, Valérie Mapelli, Hélène Mazo, Fernanda González Campo, Shi Yu, Aivars Bērziņš, Andis Lagzdīņš, Georg Rehm

Euskorpora: A Strategic Framework for Digital Sovereignty and Linguistic Inclusion of Basque in the Era of AI

Victoria Arranz, Sara Arregi, Leire Barañano, Aitor García-Pablos

Automating FAIRness: A FAIRification Tool within the Language Resources Infrastructure

Daniele Melaccio and Monica Monachini

FIBER: A Multilingual Evaluation Resource for Factual Inference Bias

Evren Ayberk Munis, Deniz Yilmaz, Arianna Muti, Cagri Toraman

EthiQuest: LLM-Powered Ethical Questionnaire Generation for Research Review

ishank kapania, Radhika Mamidi, Rahul Mishra

13:00

14:30

Lunch Break

Keynote Speaker: Nancy Chen	
14:30	The Long Arc of Language Resources: From Annotation to Alignment
15:15	Auditorium Illes Balears
	Chair: Simon Krek
15:15	Short Break (5mn)
15:20	
O5 • Inference, Reasoning, Question Answering I	
Auditorium Illes Balears	
Chair: Piek Vossen	
15:20	NegNLI-BR: A Brazilian Portuguese Benchmark for Negation in Natural Language Inference
15:40	<i>Matheus Westhelle and Viviane Moreira</i>
15:40	SWE-QA: A Dataset and Benchmark for Complex Code Understanding
16:00	<i>Laila ELKOUSSY and Julien PEREZ</i>
16:00	Augmenting LLM Reasoning with Dynamic Notes Writing for Complex MultiHop QA
16:20	<i>Rishabh Maheshwary, Masoud Hashemi, Khyati Mahajan, Shiva Krishna Reddy Malay, sai rajeswar mudumba, Sathwik Tejaswi Madhusudhan, Spandana Gella, Vikas Yadav</i>
16:20	Information Asymmetry across Language Varieties: A Case Study on Cantonese-Mandarin
16:40	and Bavarian-German QA
	<i>Renhao Pei, Siyao Peng, Verena Blaschke, Robert Litschko, Barbara Plank</i>
16:40	FRASE: Frame-based Structured Representations for Generalizable SPARQL Query
17:00	Generation
	<i>Papa Abdou Karim Karou Diallo and Amal Zouaq</i>
O6 • Information Extraction and Text Mining I	
Auditorium Mallorca	
Chair: Giorgio Maria Di Nunzio	
15:20	Representing Multimodality in Terminology Resources
15:40	<i>Federica Vezzani</i>
15:40	EPOP: A Benchmark Corpus for Assessing NLP Models on Structured Information Extraction
16:00	in Plant Health
	<i>Claire Nedellec, Marine Courtin, Xinzhi Yao, Marie Grosdidier, Isabelle Pieretti, Sandy Duperier, Robert Bossy</i>
16:00	ReTaT: A Unified Benchmark for Relation Extraction across Text and Table
16:20	<i>Mohamed Ettaleb, Thibault Ehrhart, Nathalie Aussenac-Gilles, Yoan Chabot, Mouna Kamel, Véronique MORICEAU, Raphael Troncy, Fanfu Wei</i>
16:20	LitTx: A New Treatment Relation Extraction Dataset
16:40	<i>Yuhang Jiang, Md Sultan Al Nahian, Li Hao Richie Xu, Rani Chikkanna, Ramakanth Kavuluru</i>

16:40 LegitimNarrate: A Dataset for Analyzing Legitimation Mechanisms in Crowdfunding
 17:00 Narratives
Asmaa Lagrid, Sebastien Fournier, Benedicte ALDEBERT, Ali Ghods, Daisy Bertrand, Gael Leboeuf

O7 • Language Modeling and LRs I

Menorca 1

Chair: Nigel Ward

15:20 A Fine-tuned ASR Model for Historical American Dialect Recordings
 15:40 *Steven Coats*

15:40 A Comprehensive Full-Form Lexicon for Arabic NLP and Speech Technology
 16:00 *Yannis Haralambous and Jack Halpern*

16:00 MzansiText and MzansiLM: An Open Corpus and Decoder-Only Language Model for South
 16:20 African Languages
Anri Lombard, Temi Aina, Ethan Wolff, Elan Norvick, Sbonelo Gumede, Simbarashe Mawere, Francois Meyer, Jan Buys

16:20 Very Large-Scale Multilingual Resources for LLMs and MT. Mono- and Bi-lingual Data,
 16:40 Multilingual Evaluation, and Pre-Trained Models
Stephan Oepen, Nikolay Arefyev, Mikko Aulamo, Marta Bañón, Maja Buljan, Laurie Burchell, Lucas Charpentier, Pinzhen Chen, Mariia Fedorova, Ona de Gibert, Barry Haddow, Jan Hajič, Jindrich Helcl, Andrey Kutuzov, Veronika Laippala, Zihao Li, Bhavitvya Malik, Vladislav Mikhailov, Amanda Myntti, Dayyán O'Brien, Lucie Polakova, Gema Ramirez-Sánchez, Janine Siewert, Pavel Stepachev, Joerg Tiedemann, Teemu Vahtola, Dusan Varis, Fedor Vitiugin, Jaume Zaragoza

16:40 Generation of Instruction and Preference Dataset for Improving Japanese Instruction
 17:00 Following in LLMs
Kei Moriyama, Takashi Kodama, Kouta Nakayama

O8 • Less-Resourced/Endangered/Less-studied Languages

Eivissa 1

Chair: Martin Wynne

15:20 Adapting Pretrained Models to Endangered Languages in Japan: A Comparative Study on
 15:40 Ryukyuan and Ainu Speech Recognition
Kohei Matsuura, Takanori Ashihara, Tatsuya Kawahara

15:40 Prerequisites for Advancing Automatic Speech Recognition in Breton
 16:00 *Morgan Grobol, Alice Millour, Wassim Zemouri, Yuna Drapier, Mélanie Jouitteau*

16:00 Integrating TEI, NER/NEL, Textometry, and Linked Data for a Semantically Enriched
 16:20 Interview Corpus
Ranka Stankovic, Tamara Vučenović, Biljana Rujević, Milica Ikončić Nešić, Mihailo Škorić

16:20	Uhura: A Benchmark for Evaluating Scientific Question Answering and Truthfulness in Low-Resource African Languages
16:40	<i>Edward Bayes, Israel Abebe Azime, Jesujoba Alabi, Jonas Kgomo, Tyna Eloundou, Elizabeth Proehl, Kai Chen, Imaan Khadir, Naome Etori, Shamsuddeen Hassan Muhammad, Choice Mpanza, Igneciah Pocia Thete, Dietrich Klakow, David Ifeoluwa Adelani</i>
16:40	Dialectal Filtering: Synthesizing Kurdish Corpora for Low-Resource Varieties by Utilizing "Noise" in Large Textual Data
17:00	<i>Christian Schuler, Raman Ahmad, Ānrán Wáng, Daniil Gurgurov, Timo Baumann, Simon Ostermann, Josef van Genabith</i>
15:20	Poster Sessions - Poster Area 2
17:00	

P2.1.1 • Corpora and Treebanks I

Poster Area 2

Chair: Luis Chiruzzo

HybridCodeAuthorship: A Benchmark Dataset for Line-Level Code Authorship Detection
Luke Patterson, Li Wang, Adam Faulkner

CorEGe-PT: Compiling a Large Corpus of Academic Texts in Portuguese
Tanara Zingano Kuhn, José Matos, Bruno Neves, Daniela Pereira, Elisabete Cação, Ivo Simões, Jacinto Estima, Delfim Leão, Hugo Goncalo Oliveira

SLURP-TN : Resource for Tunisian Dialect Spoken Language Understanding
Haroun Elleuch, Salima Mdhaffar, Yannick Estève, Fethi Bougares

From Semi-Digital Edition to Historical NLP Resource: Constructing and Annotating Historical Multilingual Parallel Text Collections on the TEITOK Platform
Maarten Janssen, Anna Jouravel, Piroska Lendvai

Toward Conversational Hungarian Speech Recognition: Introducing the BEA-Large and BEA-Dialogue Datasets

Máté Gedeon, Piroska Barta, Peter Mihajlik, Tekla Etelka Graczi, Anna Kohári, Katalin Mády

Developing the German Medical Text Corpus (GeMTeX): Legal Compliance and Semantic Enrichment

Justin Hofenbitzer, Christina Lohr, Andrea Riedel, Rebekka Kiser, Aliaksandra Shutsko, Abanoub Abdelmalak, Peter Klügl, Jutta Romberg, Sarah Riepenhausen, Miriam Schechner, Jakob Faller, Frank Meineke, Luise Modersohn, Markus Löffler, Juliane Fluck, Udo Hahn, Stefan Schulz, Martin Boeker

MaiChat: A Text-based Dialogue Corpus Rich in Conversational Features
Mai Hoang Dao, Catherine Lai, Peter Bell

Saudi ASWAT: A Large-Scale Corpus of Spontaneous Saudi Arabic Speech
Abdullah I. Alharbi, Afrah Altamimi, Muneera Alhoshan, Amal Almazrui, Halah Alharbi, Bayan Almuqhim, Hawra Aljasim, Abdulrahman Alosaimy, Yahya Asiri, Abdullah Alfaifi

SciCiteVal: A Multi-Domain Dataset for Scientific Citation Verification
Qinyue Liu, Yongxin Zhou, Cyril Labbe

RuznamceNER: A Named Entity Recognition Dataset for Ottoman Turkish
Esma Bilgin Tasdemir, Dilara Güler, Saziye Ozates

Scripting History: A Diachronic Urdu Text and Image Corpus from the 18Th to 19Th Centuries

Sana Shams, Sahar Rauf, Asad Mustafa, Muhammad Javed, Qurat-ul-Ain Akram, Sarmad Hussain, Miriam Butt

IREKIER: An Easy Read Corpus for Basque and Spanish

Jesús Calleja and Thierry Etchegoyhen

P2.1.2 • Corpora and Treebanks II

Poster Area 2

Chair: Milos Jakubicek

MekongPhon: A Large-Scale Parallel IPA Corpus for Lao and Khmer

Ammon Shurtz, Christian Richardson, Stephen Richardson

CorSpell: Introducing a Semiautomatic Tool for Spelling Normalization in Brazilian Portuguese

Juliana Schoffen, Dennis Giovanni Balreira, Elisa Marchioro Stumpf, Larissa Goulart, Tanara Zingano Kuhn, Rafael Oleques Nunes, Gabriel Ricci Pazzinato, Isadora Dahmer Hanauer, José Henrique de Souza Silva, Luiza Sarmento Divino, Marine Matte

Meta4XNLI-ptBR: Brazilian Portuguese Extension of Meta4XNLI Corpus

Karina Johansson, Fernanda Assi, Isabella da Silva, Rafael Passador, Isabela Rodrigues, Aline Paes, Helena Caseli

More than "Oh": Grounding Observable Events with Grunts in Multimodal Dialogue

Richard Brutti and James Pustejovsky

COME-ALPs: Coreference Annotation with Merging Heuristics Using Alignment-based Projection in Parallel Corpora

gabriela gonzalez saez, Mariam Nakhle, Illia Kholosha, Rachel Atherly, Marco Dinarelli

MEUR: A Benchmark for Evaluating Vision-Language Models on Multimodal Event Understanding and Reasoning

Zimu Wang, Yuqi Wang, Tong Chen, Changyu Zeng, Hongbin Na, Nijia Han, Fuyu Xing, Qi Chen, Qiufeng Wang, Anh Nguyen, Shuihua Wang, Ling Chen, Jionglong Su, Haiyang Zhang, Wei Wang

Building Collaborative Speech Corpora for Low-Resource Languages: The Galician Dataset in Mozilla Common Voice

Adina Vladu, Elisa Fernández Rei, María Pérez Lago

Frame-Guided Synthetic Claim Generation for Automatic Fact-Checking Using High-Volume Tabular Data

Jacob Devasier, Akshith Putta, Qing Wang, Alankrit Moses, Chengkai Li

A Bilingual Bimodal Benchmark for Arabic-English NLP across Grammatical Correction, Essay Scoring, Morphological Tagging, and Speech Recognition

Bashar Alhafni, Injy Hamed, Fadhl Eryani, David Palfreyman, Nizar Habash

Developing a Guideline for the Labovian-Structural Analysis of Oral Narratives in Japanese

Amane Watahiki, Tomoki Doi, Akari Kikuchi, Hiroshi Ohata, Yuki Nakata, Takuya Niikawa, Taiga Shinozaki, Hitomi Yanaka

German General Social Survey Personas: A Survey-Derived Persona Prompt Collection for Population-Aligned LLM Studies

Jens Rupprecht, Leon Froehling, Claudia Wagner, Markus Strohmaier

Slovene Morphological and Word Formation Segmentation: A Novel Dataset and Evaluation

Marko Pranič, Boris Kern, Ines Voršič, Senja Pollak

GePaDeU - a Multi-layer Corpus of German Parliamentary Debates with Rich Semantic and Pragmatic Annotations

Ines Rehbein, Julian Schlenker, Lars Ostertag, Simone Paolo Ponzetto

P2.1.3 • Corpora and Treebanks III

Poster Area 2

Chair: Mustafa Jarrar

What Are LLMs Doing to Scientific Communication? Measuring Changes in Writing Practices and Reading Experience

Filip Miletic and Neele Falk

GeneFRDebate: Generated French Debates from News Articles with Industrial-Expert Summaries

Rim Abrougui, Guillaume Lechien, Elisabeth Savatier, Benoît Laurent

AmbiCoRefVis: A Tool for Visualizing Coreferential Ambiguity

Patrick Paetzold, Lukas Beiske, Mark-Matthias Zymla, Massimo Poesio, Miriam Butt, Daniel Weiskopf, Oliver Deussen

Fables-DTR: A Corpus of Fables Annotated for Discourse and Temporal Relations

Purificação Silvano, António Leal, Maciej Ogrodniczuk, Aleksandra Tomaszewska, Joana Gomes, Luis Cunha, Evelin Amorim, Martyna Lewandowska, Anna Śliwicka, Alípio Jorge

A Benchmark Corpus for the Diagnostic Assessment of Content in L2 English Speech

Kosuke Doi, Justin Vasselli, Taro Watanabe

Insights from Romanized Manipuri Social Media Text: A Transliteration Corpus and Variation Analysis

Maisang Salice, Sanasam Ranbir Singh, Priyankoo Sarmah

MELD: Melding Diverse Multilingual and Multi-Domain Datasets for Named Entity Recognition Evaluation

Kevin Glocker and Marco Kuhlmann

FinER-ABSA: A Benchmark for Implicit and Explicit Entity Recognition and Aspect-Based Sentiment Analysis in Financial News

Pachara Akkanwanich, Pavorn Thongyoo, Mahannop Thabua, Konlakorn Wongpatikaseree, Natthawut Kertkeidkachorn

MUSIA: Multilingual Story Illustration Corpus for Cross-Cultural Alignment and Generation

Krishna Tewari, Supriya Chanda, Nirmal Patil, Sukomal Pal

MUDiC: A Dataset for Multi-User Dialogue and Collaboration in Chatbot Interaction

Nicolas Wagner, Cristina Luna Jimenez, Elisabeth Andre, Wolfgang Minker, Stefan Ultes

StoryCCDial: Collecting and Analyzing Human-Human Co-Creation Dialogues for Personalized Creative Support

Natsumi Ezure and Michimasa Inaba

P2.2 • Discourse and Pragmatics I

Poster Area 2

Chair: Nikhil Krishnaswamy

Evaluating Social Intelligence in LLMs via Japanese Honorifics in Email Generation: A Social Semiotic System Perspective

Muxuan Liu, Tatsuya Ishigaki, Yusuke Miyao, Hiroya Takamura, Ichiro Kobayashi

Do Language Models Know Theo Has a Wife? Investigating the Proviso Problem

Tara Azin, Daniel Dumitrescu, Diana Inkpen, Raj Singh

Cross-Lingual and Cross-Cultural Transfer of Talk Move Classification to German Science Classrooms

Christian Wartena, Christian Schumburg, Andreas Nehring, Marcel Ebert, Friederike Korneck, David Schmitt, Marie Irmer, Birgit Neuhaus

IHPP: A Paragraph-Level Dataset for Investigating the Pragmatics of Hyperpartisan Italian News

Michele Maggini, Davide Bassi, Angelo Valente, Gaël Dias, Pablo Gamallo

Detecting Potentially Under-annotated Explicit Discourse Connectives in the Penn Discourse Treebank (PDTB-3) with LLMs

Yueh-Ting Chuang, Xixian Liao, Bonnie Webber

Can LLMs Understand Punchlines? LLMs' Narrative Understanding Evaluation with Short-shorts

Jiashi Cheng and Takehito Utsuro

Building the AURIS Corpus of Reference and Information Structure

Christian Chiarcos, Christian Fäth, Tabea Gröger, Quentin Frey

There Is No Spoon: Existential Presupposition in Large Language Models

Marie-Léontine Wörgötter, Shikai Lai, Sebastian Schuster

DiscoRAG: A Discourse-Aware Agent for Query-Based Summarization of Long Documents

Alexander Chernyavskiy, Lidiia Ostyakova, Dmitry Ilvovsky

P2.3.1 • Interpretability, Explainability II

Poster Area 2

Chair: Luis Fernando D'Haro

In-Distribution Steering: Balancing Control and Coherence in Language Model Generation

Arthur Vogels, Benjamin Wong, Yann Choho, Annabelle Blangero, Milan Bhan

Improving Multilingual Language Models by Aligning Representations through Steering

Omar Mahmoud, Buddhika Semage, Thommen Karimpanal, Santu Rana

Explainable AI for Ethical Counter Speech Generation in Hate Speech Mitigation

Ashiful Islam Ridoy, Mohammed Faisal, Yogesh Kumar, Md Mamun-Ur Rashid, Marina Ernst, Frank Hopfgartner

Do Language Models Encode Semantic Relations? Probing and Sparse Feature Analysis
Andor Diera and Ansgar Scherp

The Sufficiency-Conciseness Trade-off in LLM Self-Explanation from an Information Bottleneck Perspective
Ali Zahedzadeh and Behnam Bahrak

Node-Level Uncertainty Estimation in LLM-Generated SQL
Hilaf Hasson and Ruocheng Guo

A Typologically Grounded Evaluation Framework for Word Order and Morphology Sensitivity in Multilingual Masked LMs
Anna Feldman, Libby Barak, Jing Peng

From Generation to Evaluation: A Resource for Error-Categorized Question Generation from Video Transcripts
Joshua Berger, Markos Stamatakis, Anett Hoppe, Ralph Ewerth, Christian Wartena

From Behavior to Geometry: A Causal and Geometric Analysis of LoRA-Based Domain Adaptation
Yizhe WANG, Liu He, Zhenhua Ling

Explainable Semantic Textual Similarity via Dissimilar Span Detection
Diego Miguel Lozano, Daryna Dementieva, Alexander Fraser

BIS Reasoning 1.0: The First Large-Scale Japanese Benchmark for Belief-Inconsistent Syllogistic Reasoning
Ha Thanh Nguyen, Hideyuki Tachibana, Chaoran Liu, Qianying Liu, Su Myat Noe, Koichi Takeda, Sadao Kurohashi

A Discourse-based Tool Series for Logical Validation of LLMs
Boris Galitsky and Dmitry Ilvovsky

P2.3.2 • Interpretability, Explainability III

Poster Area 2

Chair: Francesca Frontini

Voice, Bias, and Coreference: An Interpretability Study of Gender in Speech Translation
Lina Conti, Dennis Fucci, Marco Gaido, Matteo Negri, Guillaume Wisniewski, Luisa Bentivogli

MUCH: A Multilingual Claim Hallucination Benchmark
J  r  mie Dentan, Alexi Canesse, Davide Buscaldi, Aymen Shabou, Sonia Vanier

AgriChain: Visually-Grounded Expert-Verified Reasoning for Interpretable Agricultural Vision-Language Models
Hazza Mahmood, Yongqiang Yu, Rao Anwer

SyntaxGym for French: Resource, Annotation, and Evaluation of French and Multilingual LLMs
Tatiana Bladier, Henri-Jos   Deulofeu, Alexis Nasr

Modeling the Human Lexicon under Temperature Variations: Linguistic Factors, Diversity and Typicality in LLM Word Associations
Maria A. Rodr  guez, Marie Candito, Richard Huyghe

Object Realisation in Spoken Guadeloupan French: Evaluating NLP Models for an Under-Resourced Variety
Amalia Canes N  poles and Sophie Repp

Reason2Decide: Rationale-Driven Multi-Task Learning

H M QUAMRAN HASAN, Housam Khalifa Bashier, Jiayi Dai, Mi-Young Kim, Randy Goebel

Ragability Benchmark: A Dataset and Library to Test LLMs on Inter-context Conflicts

Stephanie Gross, Johann Petrak, Brigitte Krenn

Evaluating the Adaptability of Large Language Models to Linguistic Variation

Ziyan Xu, Marina Seghier, Alice Millour, Carlos-Emiliano Gonzalez-Gallardo, Jean-Yves Antoine

Probing Discrete Speech Tokens of Spoken Language Models

Sven Naber, Julia Koch, Pranav Singh, Alberto Saponaro, Ioanna Karagianni, Ngoc Thang Vu

When Consistency Becomes Bias: Interviewer Effects in Semi-Structured Clinical Interviews

Hasindri Watawana, Sergio Burdisso, Diego Moreno-Galvan, Fernando Sanchez-Vega, Adrian Pastor Lopez Monroy, Petr Motlicek, Esau Villatoro-Tello

17:00

17:20

Coffee Break

O9 • Corpora, Treebanks and Annotation; Tools, Systems and Platforms

Auditorium Illes Balears

Chair: Nizar Habash

17:20 Constructing a Japanese Claim Decomposition Dataset for Fact-Checking of LLM-Generated
17:40 Texts

Miwa Masano, Ribeka Keyaki, Atsushi Keyaki, Rei Minamoto, Kaito Horio, Hirokazu Kiyomaru, Kouta Nakayama, Hideyuki Tachibana, Daisuke Kawahara

17:40 Using LLMs for Automatic Discipline Annotation in a Diachronic Corpus of English Scientific
18:00 Papers

Sergei Bagdasarov, Diego Alves, Stefan Fischer, Elke Teich

18:00 COCOA: Creation and Exploratory Investigation of a COrpus of Claims frOm NLP Articles

18:20 *Clémentine Bleuze, Fanny Ducel, Maxime Amblard, Karen Fort*

18:20 SPOT: An Annotated French Corpus and Benchmark for Detecting Critical Interventions in
18:40 Online Conversations

Manon Berriche, Célia Nouri, Chloé Clavel, jean-philippe cointet

18:40 MedPT: A Massive Medical Question Answering Dataset for Brazilian-Portuguese Speakers

19:00 *Fernanda Farber, Iago Brito, Julia Dollis, Pedro Schindler Freire Brasil Ribeiro, Rafael Sousa, Arlindo Galvão Filho*

O10 • Information Extraction and Text Mining II

Auditorium Mallorca

Chair: Claire Nedellec

17:20 Large Language Models for Citation Function Classification

17:40 *Daniel Vodička, Pavel Kral, Christophe Cerisara, Jakub Šmíd*

17:40 Small LLMs for Medical NLP: A Systematic Analysis of Few-Shot, Constraint Decoding, Fine-
18:00 Tuning and Continual Pre-Training in Italian

Pietro Ferrazzi, Mattia Franzin, Alberto Lavelli, Bernardo Magnini

18:00	Analysing Lightweight Large Language Models for Biomedical Named Entity Recognition on
18:20	Diverse Ouput Formats <i>Pierre Epron, Adrien Coulet, Mehwish Alam</i>
18:20	WISTERIA: Weak Implicit Signal-based Temporal Relation Extraction with Attention
18:40	<i>Duy Dao DO, Anaïs Halftermeyer, Thi Bich Hanh DAO</i>
18:40	Dynamic Model Switching to Mitigate Outdated Knowledge in Large Language Models
19:00	<i>Ramakrishna Pinninti, Sabyasachi Kamila, Ayan Mazumder, Mohammed Hasanuzzaman</i>

O11 • Language Modeling and LRs II

Menorca 1

Chair: Stephan Oepen

17:20	Multi-Scale Model Compression via Nested Matrix Learning
17:40	<i>Xiangjue Dong, Aditya Anantharaman, Hemant Pugaliya, Kai Zhong</i>
17:40	Confabulations from ACL Publications (CAP): A Dataset for Scientific Hallucination Detection
18:00	<i>Federica Gamba, Aman Sinha, Timothee Mickus, Raul Vazquez, Patanjali Bhamidipati, Claudio Savelli, Ahana Chattopadhyay, Laura Zanella, Yash Kankanampati, Binesh Remesh, Aryan Chandramania, Rohit Agarwal, Chuyuan Li, Ioana Buhnila, Radhika Mamidi</i>
18:00	MedInjection-FR: Exploring the Role of Native, Synthetic, and Translated Data in Biomedical
18:20	Instruction Tuning <i>Ikram Belmadani, Oumaima El Khettari, pacome constant dit beaufils, Benoit Favre, Richard Dufour</i>
18:20	The Impact of Tokenization Algorithms on Hungarian Language Model Performance
18:40	<i>Mátyás Osváth, Máté Norbert Molnár, Roland Gunics, Noémi Ligeti-Nagy</i>
18:40	FAME: Fictional Actors for Multilingual Erasure
19:00	<i>Claudio Savelli, Moreno La Quatra, Alkis Koudounas, Flavio Giobergia</i>

O12 • Applications Involving LRs and Evaluation I

Eivissa 1

Chair: Georg Rehm

17:20	Detecting Risky Behavior Related to Alcohol and Drug Use within Adolescents' Private
17:40	Messenger Conversations <i>Jaromír Plhák, Michaela Lebedíková, Ondrej Sotolar, David Smahel</i>
17:40	Voices and Echoes in Fictional Dialogue: A Study of Linguistic Coordination in Literary Texts
18:00	<i>Ioana-Roxana Boriceanu, Alina Iacob, Liviu Dinu</i>
18:00	Bridging the Domain Divide: Supervised vs. Zero-Shot Clinical Section Segmentation from
18:20	MIMIC-III to Obstetrics <i>Baris Karacan, Barbara Di Eugenio, Patrick Thornton</i>
18:20	Reading Dynamics and Comprehension in Cognitive Aging: A Multimodal Language
18:40	Resource <i>Claudia Marzi, Noemi Boni, Alice Todesco, Andrea Nadalini, Giorgia Albertin, Cristina Dolciotti, Paolo Bongioanni, Marcello Ferro, Fabio Tamburini, Gloria Gagliardi, Vito Pirrelli</i>

18:40 Evaluating Style Embeddings for Machine-Generated Text Detection
19:00 *Noé Durandard, Saurabh Dhawan, Thierry Poibeau*

17:20
19:00 **Poster Sessions - Poster Area 1**

P3.1.1 • Dialogue, Conversational Systems I

Poster Area 1

Chair: Pierre Nugues

The Speech-LLM Takes It All: A Truly Fully End-to-End Spoken Dialog State Tracking Approach
Nizar El Ghazal, Antoine Caubrière, Valentin Vielzeuf

Off the Hamster Wheel: Rethinking Dialogue Research through a Meta-Analysis of the ACL Anthology 2024
Amandine Decker, Maxime Amblard, Ellen Breitholtz

VDAct 2.0: Scaling Video-Grounded Dialogue for Event-driven Activity Understanding with LLM-Assisted Filtering

Wiradee Imrattanatnai, Masaki Asada, Kimihiro Hasegawa, Ken Fukuda, Teruko Mitamura

Multi-dimensional Evaluation of Character-Authentic Dialogue Models Learned from Question-Answer Data

Atsushi Otsuka, Kazuya Matsuo, Kenta Hama, Masahiro Mizukami, Tsunehiro Arimoto, Hiroaki Sugiyama, Makoto Nakatsuji, Narichika Nomoto

Empathy in Greek Exam-Related Support Conversations: A Comparative Evaluation of LLM Responses

Panagiota Kyriazi and Prokopis Prokopidis

Evaluation of Two Leading Polish Language Models in a Real-world RAG Scenario
Szymon Bartanowicz and Krzysztof Jassem

A Mental State Extraction Dataset for Theory-of-Mind-based Reasoning in Emotional Support Conversations

Seulgi Kim and Harksoo Kim

Construction and Analysis of Japanese Parent-Child Dialogic Reading Corpus for Conversational Agents

Yuko Nakagi, Yuya Chiba, Sanae Fujita, Shoko Araki

ACLBOT: A Knowledge Graph-Driven Assistant for ACL Anthology Research

Jan Buchmann, Steven Lynden, Kristiina Jokinen

This House Debates AI: Evaluating a Language Model in Oxford-Style Debates against Human Experts

Umberto Belluzzo, Kobi Hackenburg, Hannah Kirk, Scott Hale, Paul Röttger

PAIR: A Pilot Dataset for Dual Perspective-based Video-Grounded Dialogue and Reconciliation

Lewis Watson, Carl Strathearn, Kenny Mitchell, Yanchao Yu

I Am Not Them: Persistent Outgroup Bias in Large Language Models Arising from Social Identity Persona Setting

Wenchao Dong, Assem Zhunis, Dongyoung Jeong, Hyojin Chin, Jiyoung Han, Meeyoung Cha

CONVERSE: Annotation Scheme and Dataset for Multimodal Conversational Engagement Analysis in Human-Human and Human-Robot Interaction

Ekaterina Torubarova, Oskar Ljung, Julia Uddén, André Pereira

FineDialFact: A Benchmark for Fine-Grained Dialogue Fact Verification

Xiangyan Chen, Yufeng Li, Yujian Gan, Arkaitz Zubiaga, Matthew Purver

P3.1.2 • Dialogue, Conversational Systems II

Poster Area 1

Chair: Ioannis Konstas

Meta-Prompting Follow-Ups for Unsupervised Dialogue Evaluation Using Open-Source Large Language Models

Gaetano Cimino, Chuyuan Li, Giuseppe Carenini, Vincenzo Deufemia

HumanICA: A Benchmark Resource for the Detection of Users' Ascription of Humanness to Conversational Agents

Sabrina Villata, Amon Rapp, Luigi Di Caro, Federica Cena

Towards Reliable Evaluation of Emotional Text Generation in LLMs: Human vs. Automatic Metrics

sadegh jafari, Els Lefever, Veronique Hoste

Question and Response Dynamics in Public Service Encounters

Wassiliki Siskou, Ingrid Espinoza, Laurin Friedrich, Steffen Eckhard, Annette Hautli-Janisz

Reasoning over Object Descriptions Improves Coreference Resolution in Task-Based Dialogue Systems

Oier Ijurco and Oier Lopez de Lacalle

Evaluating the Effect of Question Wording Variations on Answer Consistency in Large Language Models

Junya Takayama, Masaya Ohagi, Tomoya Mizumoto, Katsumasa Yoshikawa

Knowledge-Infused Hierarchy-Aware Emotion Recognition in Code-mixed Mental Health Counseling Conversations

Aseem Srivastava, Kushagra Mittal, Anusha Tiwari, Md. Shad Akhtar

A Corpus for Personalized Dialogue Breakdown Repair in Japanese Open-Domain Conversations

Kazuya Tsubokura, Yurie Iribe, Norihide Kitaoka

Conversational Assistants to Support Patients with Heart Failure: \\\ Comparing a Neurosymbolic Architecture with GPT

Anuja Tayal, Devika Salunke, Barbara Di Eugenio, Paula Allen-Meares, Eulalia Abril, Olga Garcia-Bedoya, Carolyn Dickens, Andrew Boyd

Disentangling Approaches to Conversation Disentanglement: Fine-Tune or Learn from Scratch?

Debaditya Pal, Anton Leuski, Ron Artstein, David Traum, Kallirroi Georgila

Evaluation of Failure Communication Strategies for Trust Repair in Human-AI Collaboration

Stina Klein, Alexandru Wurm, Elisabeth Andre, Matthias Kraus

Multi-Session Client-Centered Treatment Outcome Evaluation in Psychotherapy

Hongbin Na, Tao Shen, Shumao Yu, Ling Chen

Towards Reward Modeling for AI Tutors in Math Mistake Remediation

Kseniia Petukhova and Ekaterina Kochmar

HOTATE: A Japanese Dialogue Corpus Annotated with Responses of Private Thoughts and Public Statements

Yuko Toda, Daisuke Maekawa, Kota Manabe, Eito Yoneyama, Kanade Nonomura, Yuki Fujiwara, Tomoyuki Kajiware

P3.2.1 • Less-Resourced/Studied Languages I

Poster Area 1

Chair: Serge Sharoff

Mining Naturally Romanized Seed Corpora without Romanizations

Adrian Benton, Alexander Gutkin, Christo Kirov, Brian Roark

From Press to Pixels: Evolving Urdu Text Recognition

Samee Arif and Sualeha Farid

HalleluBERT: Let Every Token That Has Meaning Bear Its Weight

Raphael Scheible-Schmitt

Sentiment Analysis and Language Models for Kwanyama

Ndapa Nakashole

TigerCoder: A Novel Suite of LLMs for Code Generation in Bangla

Nishat Raihan, Antonios Anastasopoulos, Marcos Zampieri

ViX-Ray: A Vietnamese Chest X-Ray Dataset for Vision-Language Models

Duy Nguyen, Chinh Truong, Trần Phúc, Hung Le, Nguyen Dat, Trung Hieu Pham, Kiet Nguyen

Creating Task-Specific Speech Recognition Datasets from Scratch for Low-Resource Languages: Assessing the Impact of Token Sequence Overlap

Adwoa Bremang, Dennis Asamoah Owusu, Victor Quagrainne, Leanne Annor-Adjaye

Radio Haiti-Inter: A Large-Scale Annotated Corpus of Spoken Haitian Creole

William Havard, Rayan Ziane, Mélissa Menclé, Maximin Coavoux, Benjamin Lecouteux, Emmanuel Schang

Synthetic Function Demonstrations Improve Generation in Low-Resource Programming Languages

Nick McKenna, Xinnuo Xu, Jack Williams, Nicholas Wilson, Benjamin Van Durme, Christian Poelitz

PerHalluEval: Persian Hallucination Evaluation Benchmark for Large Language Models

Mohammad Hosseini, Kimia Hosseini, Shayan Bali, Zahra Zanjani, Saeedeh Momtazi

ADAB: Arabic Dataset for Automated Politeness Benchmarking - a Large-Scale Resource for Computational Sociopragmatics

Hend Al-Khalifa, Nadia Ghezael, Maria Bounnit, Hend Alhazmi, Noof Alfear, Reem Alqifari, Ameer Almasoud, Sharefah Al-Ghamdi

GRDD+: An Extended Greek Dialectal Dataset with Cross-Architecture Fine-tuning Evaluation

Stergios Chatzikyriakidis, Dimitrios Papadakis, Sevasti Papaioannou, Erofilis Psaltaki

Same-Language Subtitles for Low-resource Languages: A Case of Bundelkhandi

Anirudh Pradhan, Ayushi Pandey, Divyansh Kushwaha, Akshita Tiwary, Vivek Seshadri

The Chulalongkorn Corpus of Spoken Thai (CCOST)

Pittayawat Pittayaporn, Cathryn Yang, Sujinat Jitwiriyonant, James Kirby

Nepal Script Text Recognition from Ancient Artifacts: Challenges and Opportunities

Swornim Nakarmi, Sarin Sthapit, Sahil Tuladhar, Arya Shakya, Bal Krishna Bal, Rajani Chulyadyo

LuxBorrow: From Pompier to Pompjee, Tracing Borrowing in Luxembourgish

Nina Hosseini-Kivanani and Fred Philippy

Ramsa: A Large Sociolinguistically Rich Emirati Arabic Speech Corpus for ASR and TTS

Rania Al-Sabbagh

DialectalArabicMMLU: Benchmarking Dialectal Capabilities in Arabic and Multilingual Language Models

Malik Altakrori, Nizar Habash, Teresa Lynn, Younes Samih, Abed Alhakim Freihat, Kirill Chirkunov, Muhammed AbuOdeh, Radu Florian, Preslav Nakov, Alham Fikri Aji

P3.2.2 • Less-Resourced/Studied Languages II

Poster Area 1

Chair: Tanel Alumäe

ForumOccitania: A Corpus of User-Generated Content for Multiple Occitan Varieties

Oriane Nédey, Juliette Janès, Rachel Bawden, Thibault Clérie, Benoît Sagot

A Dataset of Wolof Ajami Manuscripts for HTR and OCR

Oreen Yousuf, Elhadji Djibril Diagne, Christian Høgel, Beata Megyesi, Joakim Nivre

TDMulti: A Tunisian Dialect-Modern Standard Arabic Multitask Corpus with a Context-Aware Cross-Attention BERT Model

Roua Torjmen and Kais HADDAR

The Megrelian Language Corpus (MLC): Creation, Annotation, and Initial Steps toward a UD Treebank

Irina Lobzhanidze, Rusudan Gersamia, Tamar Gogia

Steering LLMs toward Korean Local Speech: Iterative Refinement Framework for Faithful Dialect Translation

keunhyeung park, Seunguk Yu, Youngbin Kim

LombardoGraphia: Automatic Classification of Lombard Orthography Variants

Edoardo Signoroni and Pavel Rychly

Meenz bleibt Meenz, but Large Language Models Do Not Speak Its Dialect

Minh Duc Bui, Manuel Mager, Peter Kann, Katharina von der Wense

Bootstrapping NLP for Sakha: Named Entity Recognition and Sentiment Analysis in an Extremely Low-Resource Setting

Mariia Everstova, Nikolai Efimov, Valerio Basile

Lightweight Cross-Lingual Federated Prompt Tuning for Low-Resource Languages

Ubaid Azam, Imran Razzak, Shoaib Jameel

A Parallel Corpus of the Parable of the Prodigal Son: Building a Resource for Documenting Language Varieties in Mainland France

Lucence Ing, Juliette Janès, Sven Ködel, Benoît Sagot

Developing Zila: A Spoken Language Resource for the Endangered Slovenian Gail Valley Dialect
Andrej Zgank, Gregor Donaj, Urh Kolaric, Usi Sereinig, Tatjana Koren-Zwitter, Sanja Boto, Sabina Zwitter-Grilc, Jasna Vidinic, Darinka Verdonik

Nawatl Context-Free Grammars for Natural Language Processing
Juan Jose Guzman Landa, Juan-Manuel Torres-Moreno, Graham Ranger, Miguel Figueroa-Saavedra, Ligia Quintana Torres, Carlos-Emiliano Gonzalez-Gallardo, Luis Moreno Jimenez, Martha Lorena Avendaño Garrido

Physical Commonsense Reasoning for Lower-Resourced Languages and Dialects: A Study on Basque
Jaione Bengoetxea, Itziar Gonzalez-Dios, Rodrigo Agerri

Common Voice for Pakistan: Developing an Open Speech Corpus for Low-Resource Pakistani Languages
Meesum Alam and Francis Tyers

Amulwe Kimün: A Community-Grounded Demo, Resource, and ASR Baseline for Mapuzugun
Cristian Ahumada Oliva and Fatiha Sadat

Development of Serbian QA Datasets through Prompt-Based Generation and Human Validation
Jovana Rađenović, Olivera Kitanović, Ranka Stankovic, Mihailo Škorić

An Enhanced Pipeline for the Manzini-Savoia Dialect Corpus
Achille Fusco, Greta Mazzaggio, Carlo Zoli

P3.2.3 • Less-Resourced/Studied Languages III

Poster Area 1

Chair: Aitor Soroa

Are Language Models Borrowing-Blind? A Multilingual Evaluation of Loanword Identification across 10 Languages
Merilin Sousa Silva and Sina Ahmadi

Comparing Approaches to Automatic Summarization in Less-Resourced Languages
Chester Palen-Michel and Constantine Lignos

PsihoRo: Depression and Anxiety Romanian Text Corpus
Alexandra Ciobotaru, Ana-Maria Bucur, Liviu Dinu

Aligned Parallel Corpus of the Vedic Samhitās for Machine Translation
Yuzuki Tsukagoshi and Ikki Ohmukai

FormosanMT: A Multilingual Parallel Corpus of the Formosan Language Family
Hunter Scheppat, Joshua K. Hartshorne, Sema Koc, Éric Le Ferrand, Emily Prud'hommeaux

The Construction of a Mixe Variant Parallel Corpus
Ivan Vladimir Meza Ruiz, Delfino Zacarias Marquez, Martha Elba Ramírez Andrés, Victoriano Santiago Cayetano, Jonathan Santiago Antonio, Carlos Daniel Hernández Mena

Nepali Lemmatization with Multilingual Transformers: Intrinsic and Extrinsic Evaluation in a Low-Resource Setting
Sunil Regmi, Sundeep Dawadi, Bal Krishna Bal

Diacritic Restoration for Low-Resource Indigenous Languages: Case Study with Bribri and Cook Islands Māori

Rolando Coto-Solano, Daisy Li, Manoela Teleginski Ferraz, Olivia Sasse, Cha Krupka, Sharid Loaigiga, Sally Akevai Nicholas

A Modern Online Learning Platform for ‘Ōlelo Hawai‘i Classrooms

Christian Castro, Keneth Martin, Winston Wu, William Wilson

Glossed Data in Northern Interior Salish

Anna Stacey

CEFR-Cymraeg: A Dataset and Baseline Models for Language Proficiency Assessment in Welsh

Eeshan Waqar, Jonathan Davies, Dawn Knight, Fernando Alva-Manchego

Singlish to English Translation with Precision: A Dataset and Language Detection-Driven Masked Modeling for Singlish to English Translation

Sujit Kumar, Gerome Ang, Stephanie Hilary Xinyi Ma, Andy Hau Yan Ho, Andy Khong

LLMs in Ottoman Turkish: From MLM to NER

Enes Yilandiloğlu

SloPal: A 60-Million-Word Slovak Parliamentary Corpus with Aligned Speech and Fine-Tuned ASR Models

Erik Božík and Marek Suppa

SlovKE: A Large-Scale Dataset and LLM Evaluation for Slovak Keyphrase Extraction

Dávid Števaňák and Marek Suppa

Automatic Speech Recognition for Documenting Endangered Languages: Case Study of Ikema Miyakoan

Chihiro Taguchi, Yukinori Takubo, David Chiang

Adaptive Method for Self-Supervised Learning Models on Automatic Dialect Speech Recognition Based on Shared Knowledge of Japanese Dialects and Standard Japanese

Naoru Asakawa, Naoki Takahashi, Atsuhiko Kai, Seiichi Nakagawa

ELRA General Meeting

Auditorium Illes Balears

Chairs: German Rigau, Khalid Choukri

19:00

20:00

20:00

22:00

Welcome Reception

Thursday, 14 May, 2026

O13 • Digital Humanities and Related Corpora

Auditorium Illes Balears

Chair: Debora Nozza

- 09:00 ATLAS: Article Tracking, Linking, and Analysis of Swedish Encyclopedias
09:20 *Albin Andersson, Salam Jonasson, Fredrik Wastring, Pierre Nugues*
- 09:20 Evaluating Embedding Models on Danish Historical Newspapers: A Corpus and Benchmark
09:40 Resource
Alie Lassche, Pascale Feldkamp, Yuri Bizzoni, Katrine Baunvig, Kristoffer Nielbo, Johan Heinsen
- 09:40 Leveraging Linguistic Similarity for Low-Resource Speech Transcription
10:00 *Valentina Fedchenko and Eric Jordan*
- 10:00 A Corpus of Persuasion Techniques in Slavic Languages
10:20 *Jakub Piskorski, Dimitar Dimitrov, Marina Ernst, Jacek Haneczok, Michal Marcinczuk, Arkadiusz Modzelewski, Roman Yangarber*
- 10:20 GePaDeSE: A New Resource for Clause-Level Aspect in German Parliamentary Debates
10:40 *Julian Schlenker, Ines Rehbein, Lilly Brauner, Florian Ertz, Ines Reinig, Simone Paolo Ponzetto*
-

O14 • Lexicon

Auditorium Mallorca

Chair: Jose Camacho-Collados

- 09:00 FrameNet Semantic Role Classification by Analogy
09:20 *Van Duy Ngo, Stergos Afantenos, Emiliano Lorini, Miguel Couceiro*
- 09:20 CEFR-Annotated WordNet: LLM-Based Proficiency-Guided Semantic Database for Language
09:40 Learning
Masato Kikuchi, Masatsugu Ono, Toshioki Soga, Tetsu Tanabe, Tadachika Ozono
- 09:40 Towards a Gold Standard for Adjectival Hypernymy: Enriching the Open English WordNet
10:00 with a Hybrid Approach
Lorenzo Augello, John P. McCrae, Marco Passarotti
- 10:00 REMOVE in LiLa: Integrating Latin Preverbed Motion Verbs with WordNet and VerbNet
10:20 *Andrea Farina, Marco Passarotti, Francesco Mambrini, Matteo Pellegrini, Eleonora Litta, Giovanni Moretti*
- 10:20 From Incidents to Framing: A Dutch and English Frame-semantic Corpus and Lexicon
10:40 *Piek Vossen, Pia Sommerauer, Levi Remijnse*
-

O15 • Multilinguality, Machine Translation

Menorca 1

Chair: Bonnie Dorr

- 09:00 AI Safety Lost in Translation: Evaluating the Effectiveness of English-Italian Cross-Lingual
09:20 LLM Safety Alignment
Alessio Wu and Martim Brandao
- 09:20 Semantic Label Drift in Cross-Cultural Translation
09:40 *Mohsinul Kabir, Tasnim Ahmed, Md Mezbaur Rahman, Polydoros Giannouris, Sophia Ananiadou*
- 09:40 Chain-of-Thought Reasoning Improves Context-Aware Translation with Large Language
10:00 Models
Shabnam Ataei, Hugo Huart, Andrei Popescu-Belis
- 10:00 Adja-French Parallel Corpus: A New Resource for Machine Translation of a West African
10:20 Under-Resourced Language
Josue Godeme and Rolando Coto-Solano
- 10:20 Goldfish: Monolingual Language Models for 350 Languages
10:40 *Tyler Chang, Catherine Arnett, Zhuowen Tu, Benjamin Bergen*
-

O16 • Natural Language Generation and Summarization

Eivissa 1

Chair: Horacio Saggion

- 09:00 Dynaword: From One-shot to Continuously Developed Datasets
09:20 *Kenneth Enevoldsen, Kristian Jensen, Jan Kostkan, Balázs Szabó, Márton Kardos, Kirsten Vad, Johan Heinsen, Andrea Núñez, Gianluca Barmina, Jacob Nielsen, Rasmus Larsen, Rob van der Goot, Peter Vahlstrup, Per Dalum, Desmond Elliott, Lukas Poeh, Peter Schneider-Kamp, Kristoffer Nielbo*
- 09:20 From Bones to Rocks: A Systematic Evaluation of Specialized Definition Generation for
09:40 Portuguese
Rafael Oleques Nunes, Dennis Giovani Balreira, Joel Luís Carbonera
- 09:40 Beyond Lemmas and Syntax: Comparing Human and LLM-Generated Scientific Abstracts
10:00 *Sergei Bagdasarov and Diego Alves*
- 10:00 Systematic Multi-Aspect Evaluation of Time Series-Based Report Generation: The Case of
10:20 Financial Analysis from Stock Data
Elizabeth Fons, Elena Kochkina, Rachneet Kaur, Zhen Zeng, Berowne Hlavaty, Charese Smiley, Svitlana Vyetenko, Manuela Veloso
- 10:20 Bangla Key2Text: Text Generation from Keywords for a Low Resource Language
10:40 *Tonmoy Talukder and G M Shahariar*

Industry Day

Eivissa 2

Chairs: Teresa Lynn, Natalie Schluter

08:50	Opening
09:00	<i>Teresa Lynn, Natalie Schluter</i>
09:00	Tiny Aya: Bridging Scale and Multilingual Depth
09:30	<i>Beyza Ermis (Cohere)</i>
09:30	PronunCise Phoneme-level Speech Analysis
09:40	<i>Youssef Ouarrak (PronunCise)</i>
09:40	LoCar: Localization-Aware Evaluation of In-Vehicle Assistants through Fine-Grained Sociolinguistic control
09:50	<i>Kiwoong Park (KAIST), Seogyong Jeong (KAIST), Ken Freidl (BMW)</i>
09:50	ELDA at the crossroads: bridging language resources and responsible AI,
10:00	<i>Gabriele Chignoli (ELDA)</i>
10:00	A Comprehensive Full-Form Lexicon for Computational Processing of Arabic Dialects
10:10	<i>Jack Halpern (The CJK Dictionary Institute)</i>
10:10	Deep Adaptive AI Translation (DAAIT):
10:20	<i>Manuel Herranz (Pangeanic), Marina Albert (Pangeanic)</i>
10:20	VisualGPT: A Multimodal AI Tool for DeafDidactic Materials in Vocational Training
10:30	<i>Jörg Rutenberg (OpenMind Software)</i>
10:30	Mozilla Data Collective: The data platform for human agency and fair value exchange
10:40	<i>Robert Pugh (Mozilla)</i>

09:00	Poster Sessions - Poster Area 2
10:40	

P4.1.1 • Bias, Offensive Content, Guardrails I

Poster Area 2

Chair: Veronique Hoste

Towards Reliable AI Fairness: Challenges in Steering Features within Bias-Implicated Neurons
Ismael Garrido-Munoz, Arturo Montejo-Raez, Fernando Martí-nez-Santiago

From Body to Mind: Analyzing Gender Representation in Spanish Generative Language Models
Ismael Garrido-Munoz, Fernando Martí-nez-Santiago, Arturo Montejo-Raez

Incivility and Rigidity: Evaluating the Risks of Fine-Tuning LLMs for Political Argumentation
Svetlana Churina and Kokil Jaidka

EsBBQ and CaBBQ: The Spanish and Catalan Bias Benchmarks for Question Answering
Valle Ruiz-Fernández, Mario Mina, Júlia Falcão, Luis Antonio Vasquez Reina, Anna Salles, Aitor Gonzalez-Agirre, Olatz Perez-de-Viñaspre

ToxSyn-PT: A Synthetic Fine-Grained Dataset of Minority-Targeted Toxic Language in Portuguese
Iago Brito, Julia Dollis, Fernanda Farber, diogo fernandes, Arlindo Galvão Filho

AnswerCarefully: Creating a Dataset for LLM Safety in Japanese

Hisami Suzuki, Satoru Katsumata, Takashi Kodama, Tetsuro Takahashi, Kouta Nakayama, Satoshi Sekine

A Dutch Benchmark to Assess Social Bias in LLMs within a Hiring Decision Setting

Renate Burema, Anne Schuth, Christopher Spelt, Dong Nguyen

PBBQ: A Persian Bias Benchmark Dataset Curated with Human-AI Collaboration for Large Language Models

Farhan Farsi, Shayan Bali, Fatemeh Valeh, Parsa Ghofrani, Alireza Pakniat, Seyedkian Kashfipour, Amir H. Payberah

Contextualizing Toxicity: An Annotation Framework for Unveiling Pragmatics in Conversations of Online Discussion Forums

Yingxue Fu and Anaïs Ollagnier

How Far Can Bias Go? Tracing Bias from Pre-Training Data to Alignment

Marion Thaler, Abdullatif Köksal, Alina Leiding, Anna Anna Korhonen, Hinrich Schütze

P4.1.2 • Bias, Offensive Content, Guardrails II

Poster Area 2

Chair: Sandra Kübler

Robust Bias Evaluation with FilBBQ: A Filipino Bias Benchmark for Question-Answering Language Models

Lance Calvin Gamboa, Yue Feng, Mark Lee

Uncovering Hidden Violent Tendencies in LLMs: A Demographic Analysis via Behavioral Vignettes

Quintin Myers and Yanjun Gao

Exploring Social Bias in Slovenia: The EEC-SL Dataset

Jaya Caporusso, Damar Hoogland, Boshko Koloski, Matthew Purver, Senja Pollak, Spela Vintar

The MISOMEM-Val Dataset for Identifying Human Values in Misogynistic Memes

Rakshitha Rao Ailneni and Sanda Harabagiu

ConGA: Guidelines for Contextual Gender Annotation. a Framework for Annotating Gender in Machine Translation

Argentina Rescigno, Eva Vanmassenhove, Johanna Monti

University Speaking for Everyone: Assessing Changes in Italian Higher Education Statutes toward Gender-Inclusive Language

Sebastiano Vecellio Salto, Camilla Casula, Alessio Palmero Aprosio, Sara Tonelli

Breaking the Benchmark: Revealing LLM Bias via Minimal Contextual Augmentation

Kaveh Eskandari Miandoab, Mahammed Kamruzzaman, Arshia Gharooni, Gene Kim, Vasanth Sarathy, Ninareh Mehrabi

TryggLLM: A Benchmark for Evaluating LLM Safety in Norwegian

Samia Touileb, Truls Pedersen, Isabell Haugen

KOCOH: Korean Context-Dependent Hate Speech Dataset

Eunah Park and Sanghoun Song

P4.2.1 • Evaluation, Validation I

Poster Area 2

Chair: Dieter Van Uytvanck

RuBIN: A Russian Benchmark for Evaluating LLMs with Cultural Insights

Polina Lazukova and Irina Piontkovskaya

Evaluating Phonetically Weighted and Unweighted Distance Measures in Dialectometry

Alfred Lameli

Piecing Together Cross-Document Coreference Resolution Datasets: Systematic Dataset Analysis and Unification

Anastasia Zhukova, Terry Lima Ruas, Jan Philip Wahle, Bela Gipp

Spotlights and Blindspots: Evaluating Machine-Generated Text Detection

Kevin Stowe and Kailash Patil

JAPAS: A Benchmark and Neural Approach for Japanese Patent Support Relation Extraction

Katsuki Chousa and Ryosuke Sugiura

A Teacher-Student Approach to Creating Verified Synthetic Clarification and Correction Dialogues for TableQA Tasks

Christian Poelitz and Nick McKenna

Persona-Aware Evaluation of Cognitive Bias in LLMs: From Benchmark to Applied Decision-Making

Katsumasa Yoshikawa, Junya Takayama, Takato Yamazaki

ArtistMus: A Globally Diverse, Artist-Centric Benchmark for Retrieval-Augmented Music Question Answering

Daeyong Kwon, SeungHeon Doh, Juhan Nam

MATA (మాతా): Mindful Assessment of the Telugu Abilities of Large Language Models

Chalamalasetti Kranti and Sowmya Vajjala

Estonian Native Large Language Model Benchmark

Helena Grete Lillepalu and Tanel Alumäe

Indirect Question Answering in English, German and Bavarian: A Challenging Task for High- and Low-Resource Languages Alike

Miriam Winkler, Verena Blaschke, Barbara Plank

Benchmarking Large Language Models for Chinese and Japanese IMEs: Phonetic-to-Character Generation and Textual Error Correction

Yuchun Zou, Tedd Lee, Xiaodi Fan, Jun Li

DaLA: Danish Linguistic Acceptability Evaluation Guided by Real World Errors

Gianluca Barmina, Nathalie Norman, Peter Schneider-Kamp, Lukas Poech

KCIF: Knowledge-Conditioned Instruction Following

Rudra Murthy, Praveen Venkateswaran, Prince Kumar, Danish Contractor

GAIN: A Benchmark for Goal-Aligned Decision-Making of Large Language Models under Imperfect Norms

Masayuki Kwarada, Kodai Watanabe, Soichiro Murakami

Can LLMs Evaluate What They Cannot Annotate? Revisiting LLM Reliability in Hate Speech Detection

Paloma Piot, David Otero, Patricia Martin-Rodilla, Javier Parapar

P4.2.2 • Evaluation, Validation II

Poster Area 2

Chair: Emmanuele Chersoni

PersianMedQA: Evaluating Large Language Models on a Persian-English Bilingual Medical Question Answering Benchmark

Mohammad Javad Ranjbar Kalahroodi, Amirhossein Sheikholselami, Sepehr Karimi Arpanahi, Sepideh Ranjbar Kalahroodi, Hshaam Faili, Azadeh Shakeri

HatePrototypes: Interpretable and Transferable Representations for Implicit and Explicit Hate Speech Detection

Irina Proskurina, Marc-Antoine Carpentier, Julien Velcin

Investigating Memorization in Language Models Trained via Knowledge Distillation

Maarten Mäcking and Michaela Regneri

Redefining Evaluation Standards: A Unified Framework for Evaluating the Korean Capabilities of Language Models

Hanwool Lee, Dasol Choi, Sooyong Kim, Ilgyun Jung, Sangwon Baek, Guijin Son, Inseong Hwang, Naeun Lee, Seunghyeok Hong

Cross-Lingual Stability and Bias in Instruction-Tuned Language Models for Humanitarian NLP

Poli Nemkova, Amrit Adhikari, Matthew Pearson, Vamsi Krishna Sadu, Albert Mark

Counting on Consensus: Selecting the Right Inter-Annotator Agreement Metric for NLP Annotation and Evaluation

Joseph James

Quadratic Weighted Kappa Is Not Enough for Evaluating Automated Essay Scoring Models

Salam Albatarni and Tamer Elsayed

Evaluating the Homogeneity of Keyphrase Prediction Models

Mael Houbre, Florian Boudin, Beatrice Daille

A Taxonomy of Safety: Harmonizing LLM Benchmarks in a Fragmented Landscape

Shadi Rastegar, Viktor Hangya, Fabian Kuech, Darina Gold

Consistency of LLMs to Comparative Statements in Mathematical Reasoning Tasks

Aidan San, Daniel Son, Xiaodong Liu, Yangfeng Ji

How Many Samples Do We Need? A Toolkit for Power-Aware Evaluation Design

Angelo Basile, Areg Mikael Sarvazyan, José González

Of Words and Meaning: A Grammatical and Semantic Benchmark for Faroese LLM Understanding

Iben Debess, Barbara Scalvini, Bolette Pedersen

TURING: Evaluating Human Abilities to Identify AI-Generated Texts
Natalia Kalashnikova, Nicolas De Bufala, Sophie Fayad, Laurent Cervoni

JamC-QA: A Multiple-Choice Question Answering Benchmark for Japan-Specific Knowledge
Teruaki Oka, Tomohide Shibata, Nao Yoshida

P4.2.3 • Evaluation, Validation III

Poster Area 2

Chair: Wajdi Zaghouani

Evaluating Text Style Transfer: A Nine-language Benchmark for Text Detoxification
Vitaly Protasov, Nikolay Babakov, Daryna Dementieva, Alexander Panchenko

Irish-BLIMP: A Linguistic Benchmark for Evaluating Human and Language Model Performance in a Low-Resource Setting

Josh McGiff, Tung Tran, William Mulcahy, Dáibhidh Ó Luínín, Jake Dalzell, Róisín Ní Bhroin, Adam Burke, Barry O'Sullivan, Hoang Nguyen, Nikola Nikolov

EduBench: A Portuguese Benchmark for Open-Ended Discursive Question Answering
Pedro Paiola, Luís Gabriel Mendes, Bruno Monchelato, André Schuck, Gabriel Garcia, Douglas Rodrigues, Helena Caseli, João Papa

SemBench: A Universal Semantic Framework for LLM Evaluation
Mikel Zubillaga, Naiara Perez, Oscar Sainz, German Rigau

EL-MIA: Quantifying Membership Inference Risks of Sensitive Entities in LLMs
Ali Satvaty, Suzan Verberne, Fatih Turkmen

Same Meaning, Different Scores: Lexical and Syntactic Sensitivity in LLM Evaluation
Bogdan Kostić, Conor Fallon, Julian Risch, Alexander Loeser

The Potential for Misleading Results in Text Sanitisation with Standard Evaluation Metrics
Dan Zhang and Mark Anderson

Mind the Language Gap: Assessing LLM Safety in Italian
Elena Marafatto and Roberto Navigli

Bulgarian Massive Multitask Language Understanding Benchmark
Svetla Koeva, Ivelina Stoyanova, Dimitar Georgiev, Svetlozara Leseva, Valentina Stefanova, Maria Todorova, Tsvetana Dimitrova, Hristina Kukova, Mihaela Moskova, Tinko Tinchev

PHEB: An European Portuguese High School-Level LLM Benchmark
Diogo Tavares, Rafael Ferreira, Afonso Simplicio, Gonçalo Vinagre, Ana Condez, Inês Calvo, Inês Vieira, David Semedo, Joao Magalhaes

S-GRADES - Studying Generalization of Student Response Assessments in Diverse Evaluative Settings
Tasfia Seuti and Sagnik Ray Choudhury

Who Benchmarks the Benchmarks? A Case Study of LLM Evaluation in Icelandic
Finnur Ingimundarson, Steinunn Rut Friðriksdóttir, Bjarki Ármannsson, Iris Nowenstein, Steinþór Steingrímsson

Is This Idea Novel? An Automated Benchmark for Judgment of Research Ideas
Tim Schopf and Michael Färber

Questionnaire Meets LLM: A Benchmark and Empirical Study of Structural Skills for Understanding Questions and Responses

Duc-Hai Nguyen, Vijayakumar Nanjappan, Barry O'Sullivan, Hoang Nguyen

Assessing the Effectiveness of LLMs in Delivering Cognitive Behavioral Therapy

Navdeep Singh Bedi, Ana-Maria Bucur, Noriko Kando, Fabio Crestani

10:40

Coffee Break

11:00

O17 • Evaluation, Validation IV

Auditorium Illes Balears

Chair: Patrizia Paggio

11:00 Transcription Accuracy in the Icelandic Gigaword Corpus: Evaluating Automatic and Manual
11:20 Annotation

Johanna Mechler, Lilja Stefánsdóttir, Anton Ingason

11:20 Benchmark Data Contamination in Underrepresented Languages: A Comprehensive Analysis
11:40 Using Brazilian Data

Iriedson Vilar, David Maia, João Brunet, Fabio Morais, Leandro Marinho

11:40 TTSVowelViz: A Tool for Visualising Text-to-Speech Model Training via Vowel Spaces

12:00 *Pasindu Udawatta, Jesin James, Balamurali B T, Catherine Watson, Ake Nicholas, Binu Abeyasinghe*

12:00 A Sociophonetic Analysis of Racial Bias in Commercial ASR Systems Using the Pacific
12:20 Northwest English Corpus

Michael Scott, Siyu Liang, Alicia Wassink, Gina-Anne Levow

12:20 ParliaBench: An Evaluation and Benchmarking Framework for LLM-Generated Parliamentary
12:40 Speech

Marios Koniaris, Argyro Tsiipi, Panayiotis Tsanakas

O18 • Lexicon and Semantics I

Auditorium Mallorca

Chair: Bolette Pedersen

11:00 PARSEME 2.0 Multilingual Corpus of Multiword Expressions

11:20 *Agata Savary, Manon Scholivet, Carlos Ramisch, Takuya Nakamura, Eric Bilinski, Sara Stymne, Voula Giouli, Stella Markantonatou, Vasile Pais, Maria Mitrofan, Louis Estève, Bruno Guillaume, Verginica Barbu Mititelu, Jaka Čibej, Roberto Díaz Hernández, Victoria Fendel, Polona Gantar, Olha Kanishcheva, Cvetana Krstev, Chaya Liebeskind, Irina Lobzhanidze, Aleksandra Marković, Gunta Nešpore-Bērzkalne, Adriana Pagano, Mehrnough Shamsfard, Ranka Stankovic, Vahide Tajalli, Carole Tiberius, Aakanksha Padhye*

11:20 Multi-SimLex for Dutch: Benchmarking Embedding- and Prompt-Based Model Performance
11:40 on Semantic Similarity

Lizzy Brans and Jelke Bloem

11:40 MultiCoS: A Multilingual Dataset of Connective Semantics with Context–Sentence
12:00 Compatibility

Anne Mucha, Ciyang Qing, Wataru Uegaki

- 12:00 Adverbs Revisited: Enhancing WordNet Coverage of Adverbs with a Supersense Taxonomy
 12:20 Jooyoung Lee, Jader Camboim de Sá, Cedric Pruski
- 12:20 Introducing PerMet 1.0: A Metaphor-Annotated Corpus for Persian
 12:40 Mohammad Saeid Miri
-

O19 • Multilinguality, Machine Translation Evaluation

Menorca 1

Chair: Andrei Popescu-Belis

- 11:00 KinyCOMET: Automatic Evaluation of Machine Translation Systems for Kinyarwanda--English
 11:20 Prince Mazimpaka, Jan Nehring, Samuel Rutunda, Cristina España-Bonet
- 11:20 Multiway Parallel Corpus in Forced Migration Domain for Multilingual Machine Translation
 11:40 Fatemeh Azadi, Samuel Larkin, Chi-kiu Lo
- 11:40 Context-8: A Data Set for Evaluating Context Sensitivity in Machine Translation
 12:00 Dongyue Wang and Kyo Kageura
- 12:00 AssamLegalTrans: A Parallel Corpus, Benchmark and Analysis for English-Assamese Machine
 12:20 Translation of Legal Judgments
 Telem Joyson Singh, Hemanta Baruah, Sanasam Ranbir Singh, Anindita Talukdar, Nasrin Shahnaz, Okram Jimmy Singh, Priyankoo Sarmah, Pallav Dutta, Sukumar Nandi, Pranab Duara
- 12:20 Coordinate Structure Extraction for Patent Claims Using Multilingual LLMs
 12:40 Tsukasa Ishimaru, Takehito Utsuro, Masaaki Nagata
-

O20 • Discourse and Pragmatics II

Eivissa 1

Chair: Stergos Afantenos

- 11:00 Human Label Variation in Implicit Discourse Relation Recognition
 11:20 Frances Yung, Daniil Ignatev, Merel Scholman, Vera Demberg, Massimo Poesio
- 11:20 Conversational Implicatures through the Lens of LLMs
 11:40 Agnese Lombardi and Alessandro Lenci
- 11:40 The Emergence of the Pragmatic Dimension in Instructed-LMs
 12:00 Davide Mazzaccara and Raffaella Bernardi
- 12:00 Distributed Partial Information Puzzles: Examining Common Ground Construction under
 12:20 Epistemic Asymmetry
 Yifan Zhu, Mariah Bradford, Kenneth Lai, Timothy Obiso, Videep Venkatesha, James Pustejovsky, Nikhil Krishnaswamy
- 12:20 Grounded Misunderstandings in Asymmetric Dialogue: A Perspectivist Annotation Scheme
 12:40 for MapTask
 Nan Li, Albert Gatt, Massimo Poesio

Industry Day

Eivissa 2

Chairs: Teresa Lynn, Natalie Schluter

11:00	Equitable Voice Technologies For Real World Classrooms
11:30	<i>Emma O'Neill (AI Labs Curriculum Associates)</i>
11:30	From Hype to Metrics: Assessing Agentic LLMs for Complex Analysis
12:00	<i>Stacey Bailey (MITRE), Christy Doran (MITRE), Keith Miller (MITRE)</i>
12:00	FUSE: Record-Agnostic Multimodal Semantic Embeddings for Industrial Data
12:30	<i>William Dyer (Terumo Blood & Cell Technologies)</i>

11:00	Poster Sessions - Poster Area 1
12:40	

P5.1.1 • Inference, Reasoning, Question Answering II

Poster Area 1

Chair: Orphee De Clercq

Assessing LLM Reasoning through Implicit Causal Chain Discovery in Climate Discourse
Liesbeth Allein, Nataly Pineda-Castañeda, Andrea Rocci, Marie-Francine Moens

AccurateRAG: A Framework for Building Accurate Retrieval-Augmented Question-Answering Applications

Linh The Nguyen, Chi Tran, Dung Ngoc Nguyen, Van-Cuong Pham, Hoang Ngo, Dat Quoc Nguyen

VideoEvent: Leveraging Relevance and LLMs for Video Question Answering
Chen-Chen Lin, Ming-Han Lee, KunRu Wu, Yu-Chee Tseng

MORQA: Benchmarking Evaluation Metrics for Medical Open-Ended Question Answering
Wen-wai Yim, Asma Ben Abacha, Zixuan Yu, Robert Doerning, Fei Xia, Meliha Yetisgen

LegalRikai: Open Benchmark - a Benchmark for Complex Japanese Corporate Legal Tasks
Shogo Fujita, Yuji Naraki, Yiqing Zhu, Shinsuke Mori

Integrating Arithmetic Learning Improves Mathematical Reasoning in Smaller Models
Neeraj Gangwar, Suma Bhat, Nickvash Kani

mSCoRe: A Multilingual and Scalable Benchmark for Skill-based Commonsense Reasoning
Nghia Ngo, Franck Dernoncourt, Thien Nguyen

A Binary Problem in Binary QA: Diverse LLMs or Diverse Question Interpretations? That Is the Ensembling Question
Rafael Rosales and Santiago Miret

ObfusQAt: A Proposed Framework to Evaluate LLM Robustness on Obfuscated Factual Question Answering
Shubhra Ghosh, Abhilekh Borah, Aditya Guru, Kripabandhu Ghosh

POLAR: A Corpus of Questions, Responses and Argumentation in Polish Political Radio Discourse
Daniel Ziembicki, Aleksandra Zwierzchowska, Ewelina Sobol, Katarzyna Przerada

MHTS: Multi-Hop Tree Structure Framework for Generating Difficulty-Controllable QA Datasets for RAG Evaluation
Jeongsoo Lee, Daeyong Kwon, Kyohoon Jin, JunNyeong Jeong, Minwoo Sim, Minwoo Kim

CareMedEval Dataset: Evaluating Critical Appraisal and Reasoning in the Biomedical Field
Doria Bonzi, Alexandre Guiggi, Frederic Bechet, Carlos Ramisch, Benoit Favre

LongTailQA: Benchmarking LLMs and RAG Models on Disambiguated Long-Tail Entities
William Xion, Uwe Hadler, Tim Cofala, Maximilian Idahl, Soumyadeep Roy, Wolfgang Nejdl

CRAFT: An Explanation-Based Framework for Evaluating Cultural Reasoning in Multilingual Language Models

Shehenaz Hossain and Haithem Afli

HEAD-QA v2: Expanding a Healthcare Benchmark for Reasoning

Alexis Correa, Carlos Gómez-Rodríguez, David Vilares

Beyond MCQ: An Open-Ended Arabic Cultural QA Benchmark with Dialect Variants

Hunzalah Hassan Bhatti and Firoj Alam

Automatic Inter-document Multi-hop Scientific QA Generation

Seungmin Lee, Dongha Kim, Yuni Jeon, Junyoung Koh, Min Song

CRIT-QA: Evaluating Multi-hop Reasoning with Counterfactual Chains and Distractor Traps

Jungmin Yun, June Hyoung Kwon, Youngbin Kim

TARAZ: Persian Short-Answer Question Benchmark for Cultural Evaluation of Language Models

Reihaneh Iranmanesh, Saeedeh Davoudi, Pasha Abrishamchian, Ophir Frieder, Nazli Goharian

P5.1.2 • Inference, Reasoning, Question Answering III

Poster Area 1

Chair: Christian Chiacros

Benchmarking Mathematical Reasoning in a Low-Resource Language: Structured Prompting and Evaluation in Basque

Inigo Martinez-Criado, Aitor Soroa, Jeremy Barnes

Assessing the Difficulty of Inference Types in Natural Language Inference for Clinical Trials

Mathilde Aguiar, Pierre Zweigenbaum, Nona Naderi

Reasoning Graph-Structured Question Answering: Datasets and Insights from LLM Benchmarking

Khin Yone, Devasha Trivedi, Anish Pahilajani, Jincen Shuai, Samyak Rajesh Jain, Ryan Rossi, Nesreen Ahmed, Franck Dernoncourt, Yu Wang, Namyong Park

JBE-QA: Japanese Bar Exam QA Dataset for Assessing Legal Domain Knowledge

Zhihan Cao, Fumihito Nishino, Hiroaki Yamada, Ha Thanh Nguyen, Yusuke Miyao, Ken Satoh

A Diagnostic Benchmark for Sweden-Related Factual Knowledge

Jenny Kunz

GeoBenchmark: Probing Large Language Models for Geo-Spatial Knowledge

Ayomide Abayomi, Jose G. Moreno, Karim Radouane, Lynda Tamine

FactORes: Fact-checking with an Evidence-based Open Resource in Spanish

Nagore Bravo, Jaione Bengoetxea, Iker García-Ferrero, Alba Bonet Jover, Estela Saquete, Rodrigo Aggeri

Stands to Reason: Investigating the Effect of Reasoning on Idiomaticity Detection

Dylan Phelps, Rodrigo Wilkens, Edward Gow-Smith, Thomas Pickard, Maggie Mi, Marco Idiart, Aline Villavicencio

ESG-QA: Building a Dataset for Question Answering on Environmental, Social, and Governance Pillars

Gabriel Assis, Ayrton Surica, Pedro Kroll, Gabriela Mendes, Darian Rabbani, Edson Bollis, Lucas Francisco Pellicer, Aline Paes

Enhancing and Evaluating Tabular Models on the Fly via Synthetic Question–Answer Generation

Jorge Osés Grijalba, Eugenio Martí-nez Cámara, L. Alfonso Ureñ-López, Jose Camacho-Collados

VIVID: A Culturally Grounded Benchmark Exposing the Figurative Language Gap in Vietnamese NLP

Tu Do, Nhat Nguyen, Tung Tran, Hoang Nguyen, Tu Phuong, Long Dang

Assessing Logical Coherence of LLMs via Fine-Grained NLI

Jon Apaolaza Larraya, Begoña Altuna, Aitor Soroa, Inigo Lopez-Gazpio

Counter-Hypothesis Generation: Towards Evaluating How LLMs Reason about Alternatives

Marzieh Abdolmaleki, Aaron Maladry, Veronique Hoste, Els Lefever

LFQA-HP-1M: A Large-Scale Human Preference Dataset for Long-Form Question Answering

Rafid Ishrak Jahan, FAHMID SHAHRIAR IQBAL, Sagnik Ray Choudhury

Orthographic Constraint Satisfaction and Human Difficulty Alignment in Large Language Models

Bryan Tuck and Rakesh Verma

LIT-RAGBench: Benchmarking Generator Capabilities of Large Language Models in Retrieval-Augmented Generation

Koki Itai, Shunichi Hasegawa, Yuta Yamamoto, Gouki Minegishi, Masaki Otsuki

Investigating Reasoning with Hypotheses: The RIP2 Corpus

Ella Schad, Clara Seyfried, Chris Reed

Can Multimodal LLMs Generate Pedagogical Questions?

Thomas Gerald, Sahar Ghannay, Julie Lascar, Paul Lerner, Anne Vilnat

The Riddle of Reflection: Evaluating Reasoning and Self-Awareness in Multilingual LLMs Using Indian Riddles

Abhinav P M, Ojasva Saxena, Oswald C, Parameswari Krishnamurthy

P5.2.1 • Speech Resources and Processing I

Poster Area 1

Chair: Nianwen Xue

Using Songs to Improve Kazakh Automatic Speech Recognition

Rustem Yeshpanov

Southern Kurdish Speech Recognition Resources and Benchmarking

Mohammad Mohammadamini and Marie Tahn

MASA: A Novel Multimodal Foundation Model for L2 Speaking Assessment in Picture-description Scenarios

Bi-Cheng Yan, Fu-An Chao, Hong-Yun Lin, Berlin Chen

Tools for Estimating the Perceived Level of Phonetic Reduction

Nigel Ward, Javier Vazquez, Emma (Danny) Boushka, Oliver Niebuhr

FaLAR: A Large-scale Speaker-Annotated European Portuguese Speech Corpus of Parliamentary Sessions

Francisco Teixeira, Carlos Carvalho, Mariana Julião, Catarina Botelho, Rubén Solera-Ureña, Sérgio Paulo, Thomas Rolland, Ben Peters, Isabel Trancoso, Alberto Abad

English to Central Kurdish Speech Translation: Corpus Creation, Evaluation, and Orthographic Standardization

Mohammad Mohammadini, Daban Jaff, Josep Crego, Marie Tahon, Antoine LAURENT

Automatic Prediction of Prominence and Boundary Strength from Text

Pauline Mas, Kévin Vythelingum, Jonathan Chevelu, Marion Ouédraogo, Damien Lolive, Olivier Rosec

SOMVOICE: A First Dataset to Study the Effects of Sleep Deprivation on Voice Characteristics of Healthy French Speakers

Vincent P. Martin, Jean-Luc Rouas, Colleen Beaumard, Pierre Philip

Automatic Prediction of Child Speech Fluency with Game-Based Data from German Preschoolers

Valentin Kany, Bernd Möbius, Jürgen Trouvain

Selective Augmentation: Improving Universal Automatic Phonetic Transcription via G2P Bootstrapping

Tobias Bystrich, Julia Pritzen, Christoph Schmidt, Claudia Wich-Reif

AURORA Model of Formant-to-tongue Inversion for Didactic and Clinical Applications

Patrycja Strycharczuk and Sam Kirkham

Investigating the Role of Synthetic Data Augmentation and Training Strategies on Improving Low-Resource Language ASR

Yun Hao, Reihaneh Amooie, Wietse de Vries, Rik van Noord, Martijn Wieling

AutoRPT: A Tool for Bootstrapping Prosodic Annotation

Seth Heiney, Thomas Hicks, Sally Little, Fernanda Lourenco, Kai Retana, Eliana Stevens, Jonathan Howell

J-CHAT: Japanese Large-scale Spoken Dialogue Corpus for Spoken Dialogue Language Modeling

Wataru Nakata, Kentaro Seki, Hitomi Yanaka, Yuki Saito, Shinnosuke Takamichi, Hiroshi Saruwatari

ViMedCSS: A Vietnamese Medical Code-Switching Speech Dataset & Benchmark

Tung Nguyen, Nhu Vo, Giang Son Nguyen, Duy Hoang, Chien Huynh, Inigo Jauregi Unanue, Massimo Piccardi, Wray Buntine, Dung Le

P5.2.2 • Speech Resources and Processing II

Poster Area 1

Chair: Maria Gavriilidou

Towards Privacy-Preserving Fine-Tuning: Anonymization of Aphasic Speech for Effective ASR

Sebastian Hofstetter and Timo Baumann

ParlaSpeech 3.0: Richly Annotated Spoken Parliamentary Corpora of Croatian, Czech, Polish, and Serbian

Nikola Ljubešić, Peter Rupnik, Ivan Porupski, Taja Kuzman Pungeršek

LexiPhon: A Collection of Phonetically Transcribed Lexicons from Wikipedia
Amanda Doucette, Timothy J. O'Donnell, Morgan Sonderegger

ROG: A Multi-Layer Manually Annotated Corpus of Spoken Slovenian
Kaja Dobrovoljc, Darinka Verdonik, Jaka Čibej, Peter Rupnik, Nikola Ljubešić

Building a Dataset for French Accent Classification Evaluation: Are We There Yet?
Diandra Fabre, Mathieu Avanzi, François Portet

M3-SLU: Evaluating Speaker-Attributed Reasoning in Multimodal Large Language Models
Yejin Kwon, TAEWOO KANG, Hyunsoo Yoon, Chang Ouk Kim

Medispeech: A French Reading and Spontaneous Speech Corpus for Sleepiness Estimation
Colleen Beaumard, Vincent P. Martin, Charles Brazier, Julien Coelho, Jean-Luc Rouas, Pierre Philip

StarDrinks: An English and Korean Test Set for SLU Evaluation in a Drink Ordering Scenario
Marcely Zanon Boito, Caroline Brun, Inyoung Kim, Denys PROUX, Salah Ait-Mokhtar, Nikolaos Lagos, Jean-Luc Meunier, Ioana Calapodescu

Audio-Lyrics Alignment Dataset for Italian Arias
Pushkar Jajoria, Arianna Graciotti, Giovanna Casali, Jesujoba Alabi, Rodolfo Delmonte, Angelo Pompilio, Rocco Tripodi, James McDermott, Dietrich Klakow

The Added Value of Metadata and Annotations: Evidence from Two Large-Scale, Naturalistic Corpus Studies
Anisia Popescu, Johanna Cronenberg, Ioana Vasilescu, Ioana Chitoran, Lori Lamel, Martine Adda-Decker

CS-YODAS: A Mined Dataset of In-the-Wild Code-Switched Speech
Brian Yan, Qingzheng Wang, Matthew Wiesner, Anuj Diwan, Olga Iakovenko, Alex Polok, Injy Hamed, Shuichiro Shimizu, Iris Emerman, Thomas Hain, David R. Mortensen, Peter Viechnicki, Shinji Watanabe

The Limits of Data Scaling: Sub-token Utilization and Acoustic Saturation in Multilingual ASR
Siyu Liang, Nicolas Ballier, Gina-Anne Levow, Richard Wright

AusKidTalk: Developing Transcription Guidelines for Continuous Australian English Child Speech
Tuende Szalay, Zheng Nan, Renata Huang, Mostafa Shahin, Sirojan Tharmakulasingam, Kirrie Ballard, Beena Ahmed

P5.2.3 • Speech Resources and Processing III

Poster Area 1

Chair: Marc Kupietz

spINAch: A Diachronic Corpus of French Broadcast Speech Controlled for Speakers' Age and Gender

Simon Devauchelle, David Doukhan, Remi Uro, Lucas Ondel, Valentin Pelloin, Olympia Imbert-Brégégère, Véronique Lefort, Kévin Picard, Emeline Seignobos, Albert Rilliard

SALAN: A Massive ASR Dataset for the Languages of Niger

Mamadou K KEITA, Christopher Homan, Emily Prud'hommeaux, Abdoulaye SAKO, Seydou Diallo

Listening for Ideology: Automatic Analysis of Character Speech in Historical Nazi Propaganda Films

Nicolas Ruth, Manuel Burghardt, Andreas Niekler

Supplementary Resources and Analysis for Automatic Speech Recognition Systems Trained on the Loquacious Dataset

Nick Rossenbach, Robin Schmitt, Tina Raissi, Simon Berger, Larissa Kleppel, Ralf Schlüter

WhiteHouse: Translation of the Casablanca Corpus for Multi-dialectal Arabic Speech Translation

Fethi Bougares, Salima Mdhaffar, Yannick Estève

ToneSwiper: Facilitating Manual ToDI-annotation of Dutch Prosody

Matthijs Westera and Ariëlle Reitsema

IMaSC: A Malayalam Speech Corpus for High-Quality Text-to-Speech Synthesis

Deepa Gopinath, Thennal D K, Vrinda Nair, Swaraj S, Sachin G

Speak in Context: Multilingual ASR with Speech–Context Alignment via Contrastive Learning

Yuchen Zhang, Haralambos Mouratidis, Ravi Shekhar

Task-Lens: Cross-Task Utility Based Speech Dataset Profiling for Low-Resource Indian Languages

Swati Sharma, Divya Sharma, Anubha Gupta

Introducing MELI: The Mandarin-English Language Interview Corpus

Suyuan Liu and Molly Babel

PhonemeDF: A Synthetic Speech Dataset for Audio Deepfake Detection and Naturalness Evaluation

Vamshi Nallaguntla, Aishwarya Fursule, Shruti Kshirsagar, Anderson Avila

How Much Data for Stable Formant Values? Pipeline for Convergence Detection Based on Read Speech

Kayla Sward, Johan Sjons, Axel Ekstrom

MUSCAT: MULTilingual, SCientific ConversATion Benchmark

Supriti Sinhamahapatra, Thai-Binh Nguyen, Yiğit Oğuz, Enes Ugan, Jan Niehues, Alexander Waibel

12 :40

Antonio Zampolli Prize Winner Talk
Lori Levin, Carnegie Mellon University

13:20

Auditorium Illes Balears

Chair: German Rigau, ELRA President and Teresa Lynn, ELRA Vice-President

13:20

Lunch Break

14:45

14:45

15:15

Invited Local Speaker: Nicolau Dols

Catalan in Majorca: main features, challenges, and research tools

Auditorium Illes Balears

Chair: Antonio Toral

15:15

Short Break (5mn)

15:20

O21 • Evaluation, Validation V

Auditorium Illes Balears

Chair: Sarah Ebling

- 15:20 Towards a Diagnostic and Predictive Evaluation Methodology for Sequence Labeling Tasks
15:40 *Elena Alvarez-Mellado and Julio Gonzalo*
- 15:40 Memorization or Lucky Guesses: Detecting Short Sequences from Copyrighted Dutch News
16:00 in LLM Output
Joris Veerbeek, Kas Berendsen, Alessandra Polimeno, Antal van den Bosch
- 16:00 When Numbers Tell Half the Story: Human-Metric Alignment in Topic Model Evaluation
16:20 *Thibault Prouteau, Francis Lareau, Nicolas Dugue, Jean-Charles Lamirel, Christophe Malaterre*
- 16:20 Detecting Hallucinations in Authentic LLM–Human Interactions
16:40 *Yujie Ren, Niklas Gruhlke, Anne Lauscher*
- 16:40 Issue Detection and Category Classification in Domain-Specific Technical Logbooks
17:00 *Afshin Karimi, Ingmar Hartl, Henrik Tuennermann, Anne Lauscher*
-

O22 • Information Extraction and Text Mining III

Auditorium Mallorca

Chair: Jennifer D'Souza

- 15:20 Once upon a Kernel: Extracting Important Events from Narratives
15:40 *Anshu Sharma, Miguel Castiblanco-Melendez, Alejandro Morales, Mark Finlayson*
- 15:40 Temporal Expression Recognition in Legal Transcripts
16:00 *Elizabeth Goldstein and Maria Berger*
- 16:00 Multilingual, Multimodal Pipeline for Creating Authentic and Structured Fact-Checked
16:20 Claim Dataset
Z. Melce Hüsünbeyi, Virginie Moulleron, Leonie Uhling, Daniel Foppe, Tatjana Scheffler, Djamé Seddah
- 16:20 A Study on Building Efficient Zero-Shot Relation Extraction Models
16:40 *Hugo THOMAS, Caio Corro, Guillaume Gravier, Pascale Sébillot*
- 16:40 Beyond Catalogue Counts: The Dataset Visibility Asymmetry in Low-Resource Multilingual NLP
17:00 *Zhiyin Tan and Changxu Duan*
-

O23 • Simplification, Plain Language

Menorca 1

Chair: Els Lefever

- 15:20 BLooP: Zero-Shot Abstractive Summarization Using Large Language Models with Bigram
15:40 Lookahead Promotion
Varun Iyer and Cornelia Caragea

15:40	OasisSimp: An Open-source Asian-English Sentence Simplification Dataset
16:00	<i>Hannah Liu, Murphy Tian, Iqra Ali, Haonan Gao, Qiaoyiwen Wu, Blair Yang, Uthayasanker Thayasivam, Annie Lee, Pakawat Nakwijit, Surangika Ranathunga, Ravi Shekhar</i>
16:00	Fully Automated Identification of Lexical Alignment and Preference-Stage Shifts in Large Language Models
16:20	<i>Thomas Stephan Juzek, Xiaoyang Ming, Jose Hernandez</i>
16:20	How Much Noise Can BERT Handle? Insights from Multilingual Sentence Difficulty Detection
16:40	<i>Nouran Khallaf and Serge Sharoff</i>
16:40	Comparing Reading Behavior across Reader Expertise and Text Complexity: Insights from the French Eye-Tracking Corpus (FETA)
17:00	<i>Oksana Ivchenko and Natalia Grabar</i>

O24 • Machine Learning I

Eivissa 1

Chair: Edwin Simpson

15:20	Scaling LLM Reasoning from Minimal Labels: A Semi-Supervised Framework with a Lightweight Verifier
15:40	<i>Keizo Kato, Chenhui Chu, Yugo Murawaki, Sadao Kurohashi</i>
15:40	PARL: Prompt-based Agents for Reinforcement Learning
16:00	<i>Yarik Menchaca Resendiz and Roman Klingner</i>
16:00	SPQ: An Ensemble Technique for Large Language Model Compression
16:20	<i>Jiamin Yao and Eren Gultepe</i>
16:20	FPSC: A Sustainable Pipeline for Building a Faroese Parliamentary Speech Corpus
16:40	<i>Dávid í Lág, Barbara Scalvini, Carlos Hernandez Mena, Jon Gudnason</i>
16:40	Efficient Dialect-Aware Modeling and Conditioning for Low-Resource Taiwanese Hakka Speech Processing
17:00	<i>Peng An-Ci, Kuan-Tang Huang, Tien-Hong Lo, Hung-Shin Lee, Hsin-Min Wang, Berlin Chen</i>

Industry Day

Eivissa 2

Chairs: Teresa Lynn, Natalie Schluter

15:30	Is Agentic RAG worth it?
15:40	<i>Pietro Ferrazzi (Fondazione Bruno Kessler)</i>
15:40	AskRegula: Evidence-Grounded Regulatory QA with Integrated Benchmarking and Compliance-Oriented Evaluation
15:50	<i>Tuba Gokhan (Mohamed bin Zayed University of Artificial Intelligence)</i>
15:50	From Zero-Shot to Production: Building Robust NER for Automotive Service Analytics
16:00	<i>Adeel Zafar (Halmstad University)</i>
16:00	SNOFlow: An Interface for Human-Led Medical Entity Linking to SNOMED CT with Agent Assistance
16:10	<i>Demian Gholipour (Quantexa)</i>

16:10	Building Trust in LLM Products: A Production-Ready Evaluation Framework
16:20	<i>Erin Pacquetet (SCIAM), Christophe Séjourné (SCIAM)</i>
16:20	Efficient Knowledge Base Population with Large Language Models for Industrial Applications: A Space Mission Case Study.
16:30	<i>José Manuel Gómez-Pérez (Expert AI)</i>
16:30	FREC: A Model Context Protocol Infrastructure at the service of SMEs in the Catalan Digital Ecosystem
16:40	<i>Lluís Padró (Process Talks S.L./ Universitat Politècnica de Catalunya)</i>
16:40	The need for problem-first research for European industry and the public sector.
16:50	<i>Toms Bergamis (Tilde)</i>

15:20
17:00

Poster Sessions - Poster Area 2

P6.1.1 • Corpora and Treebanks IV

Poster Area 2

Chair: Carlos Ramisch

Construction of Japanese Prefectural Assembly Minutes Datasets across Three Electoral Terms: Comparative Analysis of 2011, 2015, and 2019 Four-Year Periods

Keiichi Takamaru, Hokuto Ototake, Yuzu Uchida, Yasutomo Kimura

EDDA-Coordinata: An Annotated Dataset of Historical Geographic Coordinates

Ludovic Moncla, Pierre Nugues, Thierry Joliveau, Katherine McDonough

Mental Health Disorder Detection beyond Social Media: A Systematic Review of Available Datasets

Sadiya Sayara Chowdhury Puspo, Ana-Maria Bucur, Stevie Chancellor, Özlem Uzuner, Marcos Zampieri

German Counseling Grounding-Act Corpus (GRACO)

Milena Belosevic

Presenting the Prague Discourse Treebank 4.0

Jiří Mírovský and Pavlína Synková

Evaluation of Co-Speech Gesture Tracking Techniques in Naturalistic Interactions

Victoria Ivanova and Naomi Harte

Voices across Decades: A Multimodal Diachronic Corpus of German Bundestag Debates (GerParlDia-MM)

Ingo Siegert

MultiWikiQA: A Reading Comprehension Benchmark in 300+ Languages

Dan Smart

SALOMO: An Annotation Tool for Complex Annotation Tasks with a Large Number of Labels

Tim Menzner

VietJobs: A Vietnamese Job Advertisement Dataset

Hieu Pham Dinh, Hung Nguyen Huy, Mo El-Haj

A Resource on Dialogical Moves in Native and Non-Native Academic Writers of English

Giulia D'Agostino, Narjes Sheikh Asadi, Elena Musi

A Corpus-Based Profiling of Regional English Variants in Global Media: Insights from Olympic Journalism

Felix Mao

JFC-Recipe: A Dataset for Nutrient Estimation from Japanese User-Generated Cooking Recipes

Keisuke Shirai, Yoko Yamakata, Hirokata Kameko, Akiko Sunto, Jun Harashima, Shinsuke Mori

Annotating Conversational Phases and Communication Techniques: A Corpus of German Teacher-Parent Counseling Conversations

Tobias Hallmen, Kathrin Gietl, Karoline Hillesheim, Annemarie Friedrich, Elisabeth André

RO-ABSA: A Romanian Dataset and Baselines for Aspect-Based Sentiment Analysis

Gheorghe Alina, Andrei Claudia, Ionescu Elena, Ruseti Stefan, Dascalu Mihai

The Moral Foundations Reddit Corpus

Jackson Trager, Alireza S. Ziabari, Elnaz Rahmati, Aida Mostafazadeh Davani, Prenti Golazizian, Farzan Karimi-Malekabadi, Ali Omrani, Zhihe Li, Brendan Kennedy, Georgios Chochlakis, Nils Karl Reimer, Melissa Reyes, Kesley Cheng, Mellow Wei, Christina Merrifield, Arta Khosravi, Evans Alvarez, Morteza Dehghani

A Semi-Automatic Workflow for Transcribing and Annotating Broadcast News

Christoph Draxler, Sven Grawunder, Jürgen Trouvain, Felicitas Kleber

P6.1.2 • Corpora and Treebanks V

Poster Area 2

Chair: Aline Villavicencio

From Rosetta to Match-Up: A Paired Corpus of Linguistic Puzzles with Human and LLM Benchmarks

Neh Majmudar, Anne Huang, Jinfan Frank Hu, Elena Filatova

Tracing How Annotators Think: Augmenting Preference Judgments with Reading Processes

Karin de Langis, William Walker, Khanh Le, Dongyeop Kang

CodeClarity: A Framework and Benchmark for Evaluating Multilingual Code Summarization

Madhurima Chakraborty, Ba Drishti Sharma, Maryam Sikander, Eman Nisar

A Longitudinal, Multinational, and Multilingual Corpus of News Coverage of the Russo-Ukrainian War

Dikshya Mohanty, Taisiia Sabadyn, Jelwin Rodrigues, Chenlu Wang, Abhishek Kalugade, Ritwik Banerjee

SKILL-IR-Discourse: A Large, Annotated Corpus of Argumentation and Domain Discourse on International Relations

Magdalena Wolska, Matti Wiegmann, Sassan Gholiagh, Mitja Sienknecht, Dora Kiesel, Irene Lopez Garcia, Patrick Riehm, Bernd Fröhlich, Katrin Girsensohn, Jürgen Neyer, Benno Stein

Building Multimodal Corpora Using Microtask Pipelines and Local Annotators

Helmiina Hotti, Raul Vazquez, Anna-Kaisa Jokipohja, Timo Kallioikoski, Henna Paakki, Rosa Suviranta, Tuomo Hiipala

Beyond Fake News Detection: A Community-based Study of the Multicultural Nature of Information Disorder

Sara Gemelli, Giulia Di Cristina, Yiran Zhang, Md Azizul Hoque, Alberto De La Torre Solís, Mohamad Behboudi Eshkiki, Nikolai Efimov, Mariia Everstova, Caterina Cappello, Maziar Kianimoghadam Jouneghani, Payam Latifi, Yashar Mahboudi, Farzaneh Mohseni, Dario Placenti, Tommaso Caselli, Manuela Sanguinetti, Aurora Scarpellini, Chiara Zanchi, Usman Naseem, Marco Antonio Stranisci, Simona Frenda

FreeTxt-Vi: A Benchmarked Vietnamese-English Toolkit for Segmentation, Sentiment, and Summarisation

Hung Nguyen, Mo El-Haj, Paul Rayson, Dawn Knight

The Patrologia Graeca Corpus: OCR, Annotation, and Open Release of Noisy Nineteenth-Century Polytonic Greek Editions

Chahan Vidal-Gorène and Bastien Kindt

National Library as Corpus: DeLiKo-2025@DNB - a Very Large Corpus of German-language Contemporary Literature

Marc Kupietz, Nils Diewald, Philippe Genêt, Andreas Witt

Multi-party Conversational Corpus of L1 and L2 for Speech Alignment Research (Teams-SK): Methodological Approach

Stefan Benus, Viktor Gatjal, Erik György, Mária Hricková, Martin Kažimír, Zuzana Kozáčiková, Lucia Mareková, Róbert Sabo, Marian Trnka, Erik Vráb

Is Semi-Automatic Transcription Useful in Corpus Creation? Preliminary Considerations on the KIParla Corpus

Martina Simonotti, Ludovica Pannitto, Eleonora Zucchini, Silvia Ballarè, Caterina Mauri

Open Korean Historical Corpus: A Millennia-Scale Diachronic Collection of Public Domain Texts

Seyoung Song, Nawon Kim, Songeun Chae, Kiwoong Park, Jiho Jin, Haneul Yoo, Kyunghyun Cho, Alice Oh

NAIST LIFE STORY: A Seven-Year Crowdsourced Dataset of Japanese Emotion-related Episodes

Kazuhiro Ito, Junko Hayashi, Hiroyuki Nagai, Shoko Wakamiya, Eiji ARAMAKI

Audience Engagement with Arabic Women's Social Empowerment and Wellbeing: A Decadal Corpus

Wajdi Zaghouni, Mabrouka Bessghaier, Md. Rafiul Biswas, Shima Ibrahim

ArPoMeme: An Annotated Arabic Multimodal Dataset for Political Ideology and Polarization

Wajdi Zaghouni, Kais Attia, Md. Rafiul Biswas, Fadhl Eryani

P6.1.3 • Corpora and Treebanks VI

Poster Area 2

Chair: Alessio Palmero Aprosio

JobArabi: An Arabic Corpus and Analysis of Job Announcements from Social Media

Wajdi Zaghouni, Shima Ibrahim, Mabrouka Bessghaier, Houda Bouamor

ParaCLEAN: Improving Translation Quality through Systematic Parallel Data Cleaning

Audrey Mash, Ella Bohman, Maite Melero

DReUD: Discourse Relations in Universal Dependencies

Jiří Mírovský and Pavlína Synková

MultiGraSCo: A Multilingual Anonymization Benchmark with Annotations of Personal Identifiers

Ibrahim Baroud, Christoph Otto, Vera Czehmann, Christine Hovhannisyan, Lisa Raithe, Sebastian Möller, Roland Roller

Structured Legal Document Generation in India: A Model-Agnostic Wrapper Approach with VidhikDastaavej

Shubham Nigam, Deepak Patnaik Balaramamahanthi, Noel Shallum, Kripabandhu Ghosh, Arnab Bhattacharya

PolyglotQL: A Pipeline for Multilingual Text-to-SPARQL Dataset Generation

Julio Perez, Fabio Barth, Georg Rehm

Building and Annotating a Large Comparable Corpus for Studying Semantic Quantification - Chinese, French, Japanese, Korean

raoul blin, Jinnam Choi, WU qishen, Yuxin Zhang, Soonhee Hwang, Takahiro Morita, Alexander Delaporte, Ilaine Wang, Chang Liu

Towards the Generation and Application of Dynamic Web-Based Visualization of UIMA-based Annotations for Big-Data Corpora with the Help of Unified Dynamic Annotation Visualizer

Thiemo Dahmann, Julian Schneider, Philipp Stephan, Giuseppe Abrami, Alexander Mehler

The MultipleYE Text Corpus: Towards a Diverse and Ever-Expanding Multilingual Text Corpus

Ramuné Kasperė, Anna Bondar, Sergiu Nisioi, Maja Stegenwallner-Schütz, Hanne B. Søndergaard Knudsen, Ana Matić, Eva Pavlinušić Vilus, Dorota Klimek-Jankowska, Chiara Tschirner, Not Battesta Soliva, Deborah Jakobi, Cui Ding, Dima Abu Romi, Cengiz Acarturk, Matilda Agdler, Anton Alexandru, Mohd Faizan Ansari, Annalisa Arcidiacono, Elizabete Barisa, Ana Bautista, Lisa Beinborn, Yevgeni Berzak, Nedeljka Bjelanović, Anna Bothmann, Jan Brasser, Caterina Cacioli, Anila Çepani, Ilze Ceple, Adelina Cerpja, Dalí Chirino, Jan Chromý, Alessandro Corona Mendoza, Iria de-Dios-Flores, Nazik Dinçtopal Deniz, Ana Došen, Kristian Elersič, Inmaculada Fajardo, Zigmunds Freibergs, Angelina Ganebnaya, Shan Gao, Jéssica Gomes, Annjo Greenall, Alba Haveriku, Miao He, Anamaria Hodivoianu, Yu-Yin Hsu, Amanda Isaksen, Andreia Janeiro, Kristine Jensen de López, Aleksandar Jevremovic, Vojislav Jovanovic, Hanna Kędzierska, Nik Kharlamov, Sara Kosutar, Nelda Kote, Vanja Kovic, Izabela Krejtz, Thyra Krosness, Oleksandra Kuvshynova, Eilam Lavy, Ella Lion, Marta Łockiewicz, Kaidi Lõo, Paula Luegi, Mircea Mihai Marin, Clara Martin, Svitlana Matvieieva, Diane Mézière, Xavier Mínguez-López, Valeriia Modina, Jurgita Motiejūnienė, Marie-Luise Müller, Tolgonai Nasipbek kyzy, Jamal Abdul Nasir, Johanne Nedergård, Ayşegül Özkan, Patrizia Paggio, Marijan Palmović, Maria Christina Panagiotopoulou, Alberto Parola, Helena Pérez, Klaudia Petersen, Anja Podlesek, Eva Pospíšilová, Marta Praulina, Mikuláš Preininger, Loredana Pungă, Diego Rossini, Špela Rot, Habib Sani Yahaya, Irina A. Sekerina, Anne Skadina, Jordi Solé-Casals, Lonneke van der Plas, Saara M. Varjopuro, Spyridoula Varlokosta, João Veríssimo, Oskari Juhapekka Virtanen, Nemanja Vračar, Mila Vulchanova, Ahmad Wali, Peizheng Wu, Nilgün Yücel, Stefan Frank, Nora Hollenstein, Lena Jäger

Sanskrit Travelogue: A Large-Scale Unified and Annotated Corpus of Sanskrit Texts

Giacomo De Luca, Danilo Croce, Roberto Basili

The Foggia Occupator Corpus: Digitisation, Annotation, and Computational Analysis of an Occupation-Era Newspaper (1945-1946)

Michele Ciletti

SiDiaC-v.2.0: Sinhala Diachronic Corpus Version 2.0

Nevidu Jayatileke, Nisansa de Silva, Uthpala Sooriya-Arachchi, Gagani Kulathilaka, Azra Safrullah, Johan Sofalas

ShAnEL-2: A Multilingual Benchmarking Dataset for Short-Answer Language Learning Exercises

Jasper Degraeuwe and Thomas Moerman

The Swedish Parliamentary Motions Corpus 1867-2024

Robert Borges, Fredrik Mohammadi Norén, Lotta Åberg Brorsson, Väinö Yrjänäinen, Hanna Bäck, Robert Klemmensen, Måns Magnusson

The Swedish Benchmark of Linguistic Minimal Pairs

Johan Sjons, Fredrik Heinat, Murathan Kurfali

Exploring the Transfer of Irony Explanation Generation from English to Dutch

Aaron Maladry, Els Lefever, Cynthia Van Hee, Veronique Hoste

DIDECO: An Annotated Dataset for Intent Detection in Digital Communications

Senaid Popovic, Damien Riquet, Maxime Meyer, Fabien Lauer, Yannick Parmentier

P6.1.4 • Corpora and Treebanks VII

Poster Area 2

Chair: Darja Fiser

GUMBridge: A Corpus for Varieties of Bridging Anaphora

Lauren Levine and Amir Zeldes

Beyond Transcripts: Iterative Peer-Editing with Audio Unlocks High-Quality Human Summaries of Conversational Speech

Kaavya Chaparala, Thomas Thebaud, Jesus Villalba Lopez, Laureano Moro-Velazquez, Peter Viechnicki, Najim Dehak

SEEM-CZ: Annotation and Classification of Epistemic Markers in Czech

Barbora Štěpánková, Michal Novák, Tomáš Musil, Lucie Polakova

When Words Don't Mean What They Say: Figurative Understanding in Bengali Idioms

Adib Sakhawat, Shamim Parveen, Md Ruhul Amin, Tahera Khatun, Shamim Mahmud, Md Saiful Islam

Human vs LLM in Conversational Repair Annotation: A New Resource and Comparative Study

Anh Ngo, Nicolas Rollet, Catherine Pelachaud, Chloé Clavel

GPT-NL Public Corpus: A Permissively Licensed, Dutch-First Dataset for LLM Pre-training

Jesse Van Oort, Frank Brinkkemper, Erik de Graaf, Bram Vanroy, Saskia Lensink

Estonian WinoGrande Dataset: Comparative Analysis of LLM Performance on Human and Machine Translation

Marij Ojastu, Hele-Andra Kuulmets, Aleksei Dorkin, Marika Borovikova, Dage Särg, Kairit Sirts

GENIUS Keylog Corpus - a German High School Student Corpus with Keystroke Logging Data

Nils-Jonathan Schaller, Thorben Jansen, Lars Höft, Hannah Pünjer, Andrea Horbach

OTA-BOUN: A Historical Turkish Dependency Treebank
Tarik Tıraş, Nureddin Ünal, Ada Cengiz, Ece Yurtseven, Esma Taşdemir, Saziye Ozates

TCMPHal: A Large-scale Dataset for Hallucination Detection in Traditional Chinese Medicine Pharmacy
Nijia Han, Zimu Wang, Ziwen Xie, Wei Wang, Jia Meng, John Moraros, Shuihua Wang

AraREQ: A Dataset and End-to-End System for Conflict Detection and Resolution in Software Requirements
Tymaa Hammouda, Alaa Aljabari, Nagham Hamad, Mustafa Jarrar

MAD: A Corpus of Multilingual Argumentative Deliberation
Eimear Maguire, Ella Schad, Jacky Visser, Chris Reed, John Lawrence

Infox-QC: A Quebec-Focused French Corpus for Misinformation Detection and AI Robustness Assessment
Moetaz Doghmane, Hazem Amamou, Thiziri Sefsaf, Alan Davoust, Anderson Avila

unarXiv 2024: A Large-Scale Scientific Corpus for Citation-Aware Retrieval and Generation
Ines Besrour and Michael Färber

EPIC-EuroParl-UdS: Information-Theoretic Perspectives on Translation and Interpreting
Maria Kunilovskaya and Christina Pollklaesener

FeedFetcher: A Resilient Web Feed Downloader for Corpus Construction
Ondřej Herman, Jan Kraus, Vit Suchomel

Human-in-the-Loop Mass Transcription and Ground Truth Annotation for Challenging Historical Documents
Norbert Fischer and Frank Puppe

17:00
17:20

Coffee Break

O25 • Corpora, Treebanks and Annotation

Auditorium Illes Balears

Chair: Claudia Borg

17:20 CoMMA, a Large-scale Corpus of Multilingual Medieval Archives
 17:40 *Thibault Clérice, Simon Gabay, Malamatenia Vlachou-Efstathiou, Ariane Pinche, Benoît Sagot*

17:40 Conversion of the Clark Hall Dictionary of Old English to TEI with RDF: An End-to-end
 18:00 Pipeline for Lexicographic Resource Retrodigitization
Sergei Stoliarov, Maxim Ionov, Fahad Khan, Marina Buzzoni, Francesca Frontini

18:00 AMORES: A Spanish Language Resource for an Extended Set of Moral Foundations
 18:20 *Oscar Araque, Daniel Molina, Anny Alvarez Nogales, Carlos A. Iglesias*

18:20 The Moralization Corpus: Frame-Based Annotation and Analysis of Moralizing Speech Acts
 18:40 across Diverse Text Genres
Maria Becker, Mirko Sommer, Lars Tapken, Yi Wan Teh, Bruno Brocai

18:40 Targum - a Multilingual New Testament Translation Corpus
 19:00 *Maciej Rapacz and Aleksander Smywiński-Pohl*

O26 • Named Entity Recognition, Speech Resources

Auditorium Mallorca

Chair: Constantine Lignos

- 17:20 Trigger Warnings Are Grounded in a Shared Vocabulary: A Corpus Analysis with User-
17:40 Generated Labels
Sebastian Heineking, Matti Wiegmann, Magdalena Wolska, Benno Stein, Martin Potthast
- 17:40 ENEIDE: A High Quality Silver Standard Dataset for Named Entity Recognition and Linking in
18:00 Historical Italian
Cristian Santini, Sebastian Barzaghi, Paolo Sernani, Emanuele Frontoni, Laura Melosi, Mehwish Alam
- 18:00 YoNER: A New Yorùbá Multi-domain Named Entity Recognition Dataset
- 18:20 *Peace Falola, Jesujoba Alabi, Solomon Akinola, Folashade Ogunajo, Emmanuel Alabi, David Ifeoluwa Adelani*
- 18:20 Linking Rationale to Decision on Internet Standards: A Retrieval-Based Approach Using
18:40 Synthetic Data
Jie Bian and Michael Welzl
- 18:40 The GELATO Dataset for Legislative NER
- 19:00 *Matthew Flynn, Timothy Obiso, Sam Newman*
-

O27 • Simplification, Plain Language and Assistive Technologies

Menorca 1

Chair: Vincent Vandeghinste

- 17:20 Controllable Sentence Simplification in Italian: Fine-Tuning Large Language Models on
17:40 Automatically Generated Resources
Michele Papucci, Giulia Venturi, Felice Dell'Orletta
- 17:40 Evaluating LLM-based Text Simplification for German: Effects on Post-Editing Effort, Quality
18:00 Ratings, and User Comprehension
Luisa Carrer, Andreas Säuberli, Martin Kappus, Lukas Fischer, Sarah Ebling
- 18:00 Reading Time in the Wild: An Assessment of Readability Predictors Based on Naturally-
18:20 Observed Reading Times
Sijbren van Vaals, Rik van Noord, Malvina Nissim
- 18:20 Document-Level Text Simplification in Estonian Using Large Language Models
- 18:40 *Meeri-Ly Muru and Eduard Barbu*
- 18:40 A Human-in/on-the-Loop Framework for Accessible Text Generation
- 19:00 *Lourdes Moreno and Paloma Martínez*

O28 • Applications Involving LRs and Evaluation II

Eivissa 1

Chair: Piotr Przybyła

- 17:20 Automatic Analysis of Collaboration through Human Conversational Data Resources: A Review
17:40 *Yi Yu, Maria Boritchev, Chloé Clavel*
- 17:40 Benchmarking Arabic Authorship Attribution and Style Transfer with Large Language Models
18:00 *Injy Hamed, Bashar Alhafni, Nizar Habash, Thamar Solorio*
- 18:00 ADHD-Lang: A Large-Scale Social Media Dataset for Verbal Behavior and Digital Phenotyping
18:20 in Adult ADHD
Daniel Wiechmann, Elma Kerz, Edward Kempa, Yu Qiao
- 18:20 SynBullying: A Multi-LLM Synthetic Conversational Dataset for Cyberbullying Detection
18:40 *arefeh kazemi, Hamza Qadeer, Joachim Wagner, hossein hosseini, Sri Balaaji Natarajan Kalaivendan, Brian Davis*
- 18:40 The Multilingual Euphemism Benchmark: Datasets and Baselines for Pragmatic Language
19:00 Understanding
Whitney Poh, Julia Sammartino, Jasper Andrew, Witold Kieraś, Natalia Zawadzka-Palucktau, Iryna Dilai, Libby Barak, Jing Peng, Anna Feldman
-

Industry Day

Eivissa 2

Chairs: Teresa Lynn, Natalie Schluter

- 17:20 Word Is Not Only Text: The Word Benchmark and the Limits of Agentic Document
17:40 Manipulation
Adrian de Wynter (Microsoft/ The University of York)
- 17:40 Panel session
- 18:30 *Moderators: Teresa Lynn, Natalie Schluter*
- 18:30 OpenEuroLLM: safe, diverse and open LLMs à la EU
18:40
- 18:40 Transperfect/Dataforce
18:50 *Alexandros Poulis (Transperfect/Dataforce)*
- 18:50 The HiTZ Chair of Artificial Intelligence and Language Technology
19:00 *Aitor Soroa (Hitz)*
-

17:20

19:00

Poster Sessions - Poster Area 1

P7.1 • Document Classification

Poster Area 1

Chair: Maite Melero

Corpus and Baselines for Distinguishing Authentic, AI-Generated, and AI-Enhanced Resumes
Andrea Loizidou, Anshu Sharma, Adrian Esquivel, Mark Finlayson, Mustafa Ocal

Mute Cods: A Multilingual Telegram Dataset with Benchmark Models for Conspiracy Theory Detection

Katarina Laken, Erik Marino, Paloma Piot, Davide Bassi, Søren Fomsgaard, Michele Maggini, Renata Vieira, Marcos Garcia, Sara Tonelli

Push and Pull: Training Sentence Encoders with Contrastive Losses for Distance-Based Multi-Label Text Classification

Jens Van Nooten and Andriy Kosar

PRIVaThe: An Annotated Dataset of Multi-Objectives Web Search Sessions

Claire Ibarboure, Ludovic Tanguy, Franck Amadieu, Josiane Mothe

Towards Safer Calls for Everyone: Designing a Benchmark Dataset for Evaluating Voice Phishing Detection Models

joeun kang, Gyuri Choi, Chanhuk Yoon, Yongbin Jeong, Younggyun Hahm, Shea Husband, Hansaem Kim

Learning Long-Document Embeddings via Chunk–Context Entailment

Waheed Ahmed Abro, Naim Es-Sebbani, Zied Bouraoui

Scientific Article Section Classification (SASC) Dataset

Nicolau Duran-Silva, Julian Moreno-Schneider, César Parra-Rojas, Georg Rehm

JMTEB and JMTEB-lite: Japanese Massive Text Embedding Benchmark and Its Lightweight Version

Shengzhe Li, Masaya Ohagi, Ryokan Ri, Akihiko Fukuchi, Tomohide Shibata, Daisuke Kawahara

Construction of a Japanese RAG Benchmark Using Synthetic Documents on Non-existent Entities and Events

Shengzhe Li, Masaya Ohagi, Hayato Tsukagoshi, Akihiko Fukuchi, Tomohide Shibata, Daisuke Kawahara

C4: A Multilingual Benchmark for Retrieval-Augmented Generation Based on the Catechism of the Catholic Church and Its Compendium

Pius von Däniken, Mark Cieliebak, Jan Deriu

P7.2.1 • Information Extraction and Text Mining IV

Poster Area 1

Chair: Bernardo Magnini

Contrastively Pre-trained Event Embeddings with Schema-free LLM Annotations

Frank Mtumbuka and Steven Schockaert

A Dataset of Psychiatric Hospital Notes with Temporal Information Annotations

Timothy Miller, Gaby Dinh, David Harris, WonJin Yoon, Spencer Thomas, Boyu Ren, MEIHUA HALL, Guergana Savova

Format Matters: A Critical Evaluation of Output Formats for Prompting LLMs in SLU and NER

Pierre Lepagnol, Sahar Ghannay, Thomas Gerald, Christophe Servan, Sophie Rosset

Identifying Imaging Follow-Up in Radiology Reports: A Comparative Analysis of Traditional ML and LLM Approaches

Namu Park, Giridhar Kaushik Ramachandran, Kevin Lybarger, Fei Xia, Özlem Uzuner, Martin Gunn, Meliha Yetisgen

Efficient Topic Extraction via Graph-Based Labeling: A Lightweight Alternative to Deep Models
SALMA MEKAoui, Hiba Sofyan, Imane Benchrif, Imane Amaaz, Ilham Chaker, Arsalane Zarghili, Nikola Nikolov

From Noise to Signal: When Outliers Seed New Topics
Evangelia Zve, Gauvain Bourgne, Benjamin Icard, Jean-Gabriel Ganascia

Explore Political Discourse with Transformers. Emergent Paradigmatic and Syntagmatic Representations.
Laurent Vanni and Damon Mayaffre

The Growing Gains and Pains of Iterative Web Corpora Crawling: Insights from South Slavic CLASSLA-web 2.0 Corpora
Taja Kuzman Pungeršek, Peter Rupnik, Vit Suhomel, Nikola Ljubešić

MaritimEmails: A Synthetic Dataset for Maritime Chartering Correspondence
Kevin Bruendler and Simon Clematide

eSciBench: An Extensible Scientific PDF Extraction Benchmark
Noah Tremblay Taillon and Phillippe Langlais

Vrittanta-AS: Dataset Development and Benchmarking for Event Trigger Detection and Classification in Assamese
Chaitanya Kirti, Dhruvajyoti Pathak, Ashish Anand, Prithwijit Guha

From Facts to Hypotheses: Joint Detection of Biomedical Relations and Epistemic Commitment Using LLMs
Aleksandra Gabryszak, Phuc Tran Truong, Arne Binder, Nikola Milosevic, Felix-Sebastian Keese, Astrid Rheinländer, Philippe Thomas

SciLaD: A Large-Scale, Transparent, Reproducible Dataset for Natural Scientific Language Processing
Luca Foppiano, Sotaro Takeshita, Pedro Ortiz Suarez, Ekaterina Borisova, Raia Abu Ahmad, Malte Ostendorff, Fabio Barth, Julian Moreno-Schneider, Georg Rehm

CausalSense: Leveraging Common Sense Knowledge and LLMs for Joint Event Extraction and Relation Classification
Youssra REBBOUD, Pasquale Lisena, Raphael Troncy

P7.2.2 • Information Extraction and Text Mining V

Poster Area 1

Chair: Jörg Tiedemann

Large Language Models Are Good Term Extractors: A Systematic Evaluation
Ayla Rigouts Terry

A Large-Scale Dataset for Linking-Based Geocoding
Hibiki Nakatani, Yuichiro Yasui, Ryosuke Wakamoto, Masayuki Ishii, Tetsuhisa Suizu, Hiroki Ouchi, Taro Watanabe

FiNERVINER: Fine-grained Named Entity Recognition for Vulnerable Languages of India's North Eastern Region
Prachurya Kaushik and Ashish Anand

APTFiNER: Annotation Preserving Translation for Fine-grained Named Entity Recognition
Prachuryya Kaushik, Aditya Gupta, Ajanta Maurya, Gautam Sharma, V. Saradhi, Ashish Anand

RelEx-PT: A Portuguese Sentence-Level Relation Extraction Dataset
Tomás Pinto, Catarina Silva, Hugo Goncalo Oliveira

Benchmarking Portuguese Open Information Extraction
Gabriel Silva, Mário Rodrigues, António Teixeira, Marlene Amorim

A Scalable Pipeline for Novelty Detection in Skill Extraction Using Large Language Models
Gian Seifert and Simon Clematide

Do LLMs Judge Distantly Supervised Named Entity Labels Well? Constructing the JudgeWEL Dataset
Alistair Plum, Laura Bernardy, Tharindu Ranasinghe

From Articles to Premises: Building PrimeFacts, an Extraction Methodology and Resource for Fact-Checking Evidence
Premtim Sahitaj, Jawan Kolanowski, Ariana Sahitaj, Veronika Solopova, Max Upravitelev, Daniel Röder, Iffat Maab, Junichi Yamagishi, Sebastian Möller, Vera Schmitt

EpiGator: An Event-based Surveillance System for Infectious Disease Outbreaks
Yiheng Wu, Jue Hou, Trangcasanchai Sathianpong, Lidia Pivovarova, Roman Yangarber

Relation Extraction across Entire Books to Reconstruct Community Networks: The AffilKG Datasets
Erica Cai, Sean McQuade, Kevin Young, Brendan O'Connor

Vrittanta-EN: A Benchmark Dataset for Event Trigger Detection and Classification Advancing Event Understanding in English Narrative Discourse
Chaitanya Kirti, Ashish Anand, Prithwiji Guha

MUC-4 Revisited: Document-level Event Analysis beyond Span-based Arguments
Helene Olsen, Erik Velldal, Lilja Øvrelid

P7.3 • Knowledge Representation and Graphs

Poster Area 1

Chair: John McCrae

Historical Medical Knowledge Graphs and Ontologies from the Medical History of British India Corpus (1850-1950)
Mehrdad Almasi and Tugce Karatas

Graph-TempCZ: A Graph Representation of Software Mentions for Predicting Software Usage in Scientific Publications
Congfeng Cao, Pengyu Zhang, Jelke Bloem

Automatic Suggestions Help Extending Eventive Ontology: A Case Study on SynSemClass
Jana Straková, Eva Fučíková, Zdenka Uresova, Jan Hajič

JPPB: Automatic Construction of a Soft-Labeled Japanese Patient Phrase Bank for Symptom Normalization
Tomohiro Nishiyama, Mana Kuramoto, Shoko Wakamiya, Eiji ARAMAKI

How I Met Your Snowclone: Unsupervised Discovery of Snowclone Patterns in Large Datasets
Julien Bezançon, Gaël Lejeune, Marceau Hernandez

HOME-KGQA: A Benchmark Dataset for Multimodal Knowledge Graph Question Answering on Household Daily Activities

Shusaku Egami, Aoi Ohta, Tomoki Tsujimura, Masaki Asada, Tatsuya Ishigaki, Ken Fukuda, Masahiro Hamasaki, Hiroya Takamura

Extending the Semantic Layer of the Compl-it Italian Lexicon: Traits, Semantic Types, and Definitions

Emiliano Giovannetti, Andrea Bellandi, Simone Marchi, Mafalda Papini

Integrating Knowledge Graph with Large Language Models for Multi-hop Question Generation

Yllias Chali and Al Hasib Mahamud

LocalGovPL: A Corpus of Speaker-Attributed Polish Local Government Transcripts

Dariusz Czernski and Maciej Ogrodniczuk

Amharic DBpedia Chapter: A Knowledge Graph for a Low-Resource Language

Hizkiel Alemayehu, Tilahun Abedissa Taffa, Meti Bayissa, Andargachew Zewge, Hamada Zahera, Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo

Cygnnet: Refactoring the Open Multilingual Wordnet

Rowan Maudslay and Francis Bond

Masrad: Arabic Terminology Management Corpora with Semi-Automatic Construction

Mahdi Nasser, Laura Sayah, Fadi Zaraket

P7.4 • Opinion, Sentiment, Emotion Analysis

Poster Area 1

Chair: Dimitrios Galanis

SentiMalti: A Maltese Sentiment Analysis Dataset and Models

Ian Caruana, Matthew Vella, Fabio Zammit, Kurt Micallef, Claudia Borg

Multilingual Structured Sentiment Analysis for Environmental Sustainability

Muhammad Okky Ibrohim, Tommaso Caselli, Cristina Bosco, Valerio Basile

LLM-as-an-Annotator: Training Lightweight Models with LLM-Annotated Examples for Aspect Sentiment Tuple Prediction

Nils Constantin Hellwig, Jakob Fehle, Udo Kruschwitz, Christian Wolff

Extending Czech Aspect-Based Sentiment Analysis with Opinion Terms: Dataset and LLM Benchmarks

Jakub Šmíd, Pavel Priban, Pavel Kral

AnnoABSA: A Web-Based Annotation Tool for Aspect-Based Sentiment Analysis with Retrieval-Augmented Suggestions

Nils Constantin Hellwig, Jakob Fehle, Udo Kruschwitz, Christian Wolff

Zero-Shot to Full-Resource: Cross-lingual Transfer Strategies for Aspect-Based Sentiment Analysis

Jakob Fehle, Nils Constantin Hellwig, Udo Kruschwitz, Christian Wolff

LoveHate: Stance Detection and Generation for Multiple Topics in User-generated Comments in Russian and English

Natalia Evgrafova, Veronique Hoste, Els Lefever

From Trial by Fire to Sleep like a Baby: A Lexicon of Anxiety Associations for 20K English Multi-Word Expressions

Saif Mohammad

Entity-Level Sentiment Analysis with Sentence Relevance Detection

Egil Rønningstad, Roman Klinger, Lilja Øvrelid, Erik Veldal

Enhancing Multi-Label Emotion Analysis and Corresponding Intensities for Ethiopian Languages

Tadesse Destaw Belay, Dawit Gete, Abinew Ali Ayele, Olga Kolesnikova, Iqra Ameer, Grigori Sidorov, Seid Muhie Yimam

A Japanese Dataset for Aspect-based Sentiment Polarity Classification and Emotion Intensity Estimation

Kentaro Hanafusa, Kota Manabe, Yuki Maeda, Daisuke Maekawa, Tomoyuki Kajiware, Hideaki Hayashi, Yuta Nakashima, Hajime Nagahara

P7.5 • Argument Mining and Emotion Classification

Poster Area 1

Chair: Atul Kumar Ojha

Assessing the Persuasive Effect of AI-Generated Image Support of Arguments

Mackwyn Quadras, Manfred Stede, Henning Wachsmuth

CIARAM: Class Imbalance Aware Generative Framework for Relational Argument Mining

Nilmadhab Das, Sayan Pal, V. Saradhi, Ashish Anand

Surfacing Subtle Stereotypes: A Multilingual, Debate-Oriented Evaluation of Modern LLMs

Muhammed Saeed, Muhammad Abdul-Mageed, Shady Shehata

Prompt-Based Stance Control in German: An Evaluation of LLMs for Experimental Research on Attitude Change

Florian Omieciński, Cornelia Sindermann, Agnieszka Falenska

CoSt-BR: A Language Resource for Conversational Stance Detection

Felipe da Fonseca, Ivandre Paraboni, Luciano Digiampietri

Less Is More? The Role of Demographic Author Information in Emotion Classification of Ambiguous Text

Sabine Weber, Lynn Greschner, Roman Klinger

Big Five Personality Prediction through Emotion-Conditioned Representations and Learnable Psycholinguistic Mapping

Lorenzo Zangari, Antonin Schnyder, Davide Picca

SENSEI-ASG: A Challenging Dataset for Argument Summary Graph Parsing

Jonathan Clayton, Marco Damonte, Robert Gaizauskas

Categorical Emotions or Appraisals - Which Emotion Model Explains Argument Convincingness Better?

Lynn Greschner, Meike Bauer, Sabine Weber, Roman Klinger

Creation of the Estonian Subjectivity Dataset: Assessing the Degree of Subjectivity on a Scale

Karl Gustav Gailit, Kadri Muischnek, Kairit Sirts

Friday, 15 May, 2026

Session O29 • Infrastructures, Policy and Legal Issues II

Auditorium Illes Balears

Chair: Karen Fort

- 09:00 Mitigating Misinterpretation in Policy Documents through Automated Language
09:20 Understanding
Momojit Biswas, Anka Chandrachud Tummepalli, Preethu Rose Anish
- 09:20 Sovereign AI-based Public Services Are Viable and Affordable
09:40 *António Branco, Luis Gomes, Rodrigo Santos, Eduardo Santos, João Ricardo Silva, Nuno Marques, Madalena Rodrigues*
- 09:40 A Typology of Synthetic Datasets for Dialogue Processing in Clinical Contexts
10:00 *Steven Bedrick, A. Seza Dogruoz, Sergiu Nisioi*
- 10:00 Text+: A National Hub Including Legacy Language Data
10:20 *Florian Barth, Christoph Draxler, Jennifer Ecker, Stefan Fischer, Philippe Genêt, Alina Hemmer, Timm Lehmberg, Thorsten Trippel, Andreas Witt, Arden Zimmermann, Claus Zinn*
- 10:20 Can NLP Tackle Hate Speech in the Real World? Stakeholder-Informed Feedback and Survey
10:40 on Counterspeech
Tanvi Dinkar, Aiqi Jiang, Simona Frenda, Poppy Gerrard-Abbott, Nancie Gunson, Gavin Abercrombie, Ioannis Konstas
-

Session O30 • Opinion and Argument Mining, Sentiment Analysis

Auditorium Mallorca

Chair: Diana Inkpen

- 09:00 Towards Complex Debate Understanding: Predicting Claim Impact Scores through the
09:20 Modelling of Claim Interactions
Maxime Brouat, Mihai Surdeanu, Srdjan Vesic, Eduardo Blanco
- 09:20 Is There Anything More Deceptive than an Obvious Fact? Investigating Implicitness in User-
09:40 Generated Argumentative Text
Ekaterina Sviridova, Elena Cabrio, Serena Villata
- 09:40 Best-Worst Scaling of Hype in Biomedical Research: Building an Intensity Lexicon of
10:00 Promotional Adjectives
Neil Millar, Dipesh Satav, Bojan Batalo, Erica K. Shimomoto, Ryosuke Ohniwa
- 10:00 Trust Me, I Can Convince You: The Contextualized Argument Appraisal Framework and the
10:20 ContArgA Corpus
Lynn Greschner, Sabine Weber, Roman Klinger
- 10:20 Towards Clinical Applications of NLP: Detecting Emotion Regulation via Emotional
10:40 Categories and Expression Modes in French Transcriptions
Salome Klein, Amalia Todirascu, Hélène Vassiliadou

Session O31 • Bias, Offensive and Non-inclusive Language

Menorca 1

Chair: Penny Labropoulou

- 09:00 R.U.Psycho? A Framework for Robust Unified Psychometric Testing of Language Models
09:20 *Julian Schelb, Orr Borin, David Garcia, Andreas Spitz*
- 09:20 Code-switching as a Bias Indicator in LLMs: "the Consequences Are Not the Same Para
09:40 Nosotros"
Fanny Ducel, Aurélie Névéol, Vidit Khazanchi, Loïc Leclere, Arthur Pedrini, Léa Bouchet, Benjamin Caissial, Karen Fort
- 09:40 Exploration of How Hate Is Framed on Social Media
10:00 *Rakshitha Rao Ailneni and Sanda Harabagiu*
- 10:00 Are Social Biases in LLMs Consistent across Generative Tasks? A Case Study for Basque
10:20 *Muitze Zulaika, Xabier Saralegi, Julia Shershneva, Lia Gonzalez, Arkaitz Fullaondo*
- 10:20 Fine-grained Narrative Classification in Biased News Articles
10:40 *Zeba Afroz, Harsh Vardhan, pawan bhakuni, Aanchal Punia, Rajdeep Kumar, Md. Shad Akhtar*
-

Session O32 • Speech Resources, Processing, Applications

Eivissa 1

Chair: Nikola Ljubešić

- 09:00 A Shoal of Voices: Parallel Read Speech from Professional Swedish Narrators
09:20 *Christina Tännander, Jim O'Regan, Jens Edlund*
- 09:20 Deep Learning-Based Multi-Aspect Pronunciation Assessment for Individuals with Down
09:40 Syndrome
David Fernández-García, César González-Ferreras, Valentín Cardeñoso-Payo, Mario Corrales-Astorgano
- 09:40 WikiIPA: Integrating WikiPron and Lingua Libre for Multilingual IPA Transcription
10:00 *Pierluigi Cassotti, Jacob Suchardt, Domenico De Cristofaro*
- 10:00 How Pragmatics Shape Articulation: A Computational Case Study in STEM ASL Discourse
10:20 *Saki Imai, Lee Kezar, Laurel Aichler, Mert Inan, Erin Walker, Alicia Wooten, Lorna Quandt, Malihe Alikhani*
- 10:20 Setting the Stage for Disfluency: Implications of Contextual Task Framing Effects for the
10:40 Design of Listening Tasks
Ambika Kirkland and Jens Edlund
-

09:00

10:40

Poster Sessions - Poster Area 2

P8.1.1 • Machine Translation I

Poster Area 2

Chair: Oscar Araque

ACADData: Parallel Dataset of Academic Data for Machine Translation

Iñaki Lacunza, Javier Garcia Gilabert, Francesca De Luca Fornaciari, Javier Aula-Blasco, Aitor Gonzalez-Agirre, Maite Melero, Marta Villegas

A Single Model Ensemble Framework for Neural Machine Translation Using Pivot Translation

Seokjin Oh, Keonwoong Noh, Woohwan Jung

Gender Disambiguation in Machine Translation: Diagnostic Evaluation in Decoder-Only Architectures

Chiara Manna, Hosein Mohebbi, Afra Alishahi, Frederic Blain, Eva Vanmassenhove

Building a One-Million-Pair Bokmål–Nynorsk Translation Corpus: A Quality-First Harvesting and Cleaning Pipeline

Per Kummervold, Thea Tollersrud, Angelina Zanardi

New Trends for Modern Machine Translation with Large Reasoning Models

Sinuo Liu, Chenyang Lyu, Minghao Wu, Zifu Shang, Longyue Wang, Weihua Luo, Kaifu Zhang

MaitH 1.0: A Parallel Corpus and Baseline for Low-Resource Maithili-Hindi Translation

Kamanksha Dubey, Chandresh Maurya, Kumar Padmanabh

NRD: A Hybrid Disentanglement Framework for Mitigating Interference in Multilingual Machine Translation

Jiarui Zhang and Yifan Deng

Linguistic and Demographic Factors in an Online Free Translation Task

Tyler Lee, Irina Stenger, Tania Avgustinova

Biases in Translation: Assessing Opinion Distortion in Machine Translated Texts

Nazanin Shafiabadi and François Yvon

When Translations Surprise: Human Awareness of Predictability in Translations

Cristian García-Romero, Miquel Esplà-Gomis, Felipe Sanchez-Marti-nez

Bidirectional Chinese and English Passive Sentences Dataset for Machine Translation

Xinyue Ma, Pol Pastells, Mireia Farrus, Mariona Taule

CoTERM: A Consistency-Oriented Term Metric for MT System Evaluation

Amir Hazem and Kyo Kageura

SiniticMTError: A Machine Translation Dataset with Error Annotations for Sinitic Languages

Hannah Liu, Junghyun Min, Annie Lee, Ethan Yue Heng Cheung, Shou-Yi Hung, Elsie Chan, Shiyao Qian, RUNTONG LIANG, Kimlan Huynh, Wing Yu Yip, York Hay Ng, Tsz Fung Yau, Ka Ieng Charlotte Lo, You-Wei Wu, Richard Tzong-Han Tsai

Ancient Greek to Modern Greek Machine Translation: A Novel Benchmark and Fine-Tuning Experiments on LLMs and NMT Models

Spyridon Mavroumatidis, Sokratis Sofianopoulos, Prokopis Prokopidis, Maria Giagkou

P8.1.2 • Machine Translation II

Poster Area 2

Chair: Sahar Ghannay

Linguistic Knowledge-Infused Fine-Tuning for Mitigating Gender Bias in Machine Translation

Luis Ernesto Garcia Estrada, Audrey Mash, Carlos Escolano, Maite Melero, Christine Basta

What Triggers My Model? Contrastive Explanations Inform Gender Choices by Translation Models

Janiça Hackenbuchner

ViKhoMT: A Vietnamese–K'Ho Neural Machine Translation Dataset and Evaluation for Community Health Communication

Tram Truong, Vinh Nguyen, Dang Thin, Ngan Nguyen

Hindsight Quality Prediction Experiments in Multi-Candidate Human-Post-Edited Machine Translation

Malik Marmonier, Benoît Sagot, Rachel Bawden

PETra: A Multilingual Corpus of Pragmatic Explication in Translation

Doreen Osmelak, Koel Dutta Chowdhury, Uliana Sentsova, Cristina España-Bonet, Josef van Genabith

A Dataset for Probing Translationese Preferences in English-to-Swedish Translation

Jenny Kunz, Anja Jarochenko, Marcel Bollmann

STAR-IL: A Dataset for Style-Aware Machine Translation of Product Reviews in Indian Languages

Ketaki Shetye, Dipti Sharma, Parameswari Krishnamurthy

Cultural and Knowledge Biases in LLMs through the Lens of Entity-Aware Machine Translation

Lu Xu, Luca Moroni, Roberto Navigli

Referenceless Evaluation of Machine Translation Models by Ranking Performance in Romanian to English Translate-train Settings

Mihail Feraru, Alexandra Diaconu, Bogdan Alexe

Every Word Presented in Context: Syntactic Coverage as Objective for Low-Resource Machine Translation with Large Language Models

Samuel Frontull and Thomas Ströhle

Multilingual KokoroChat: A Multi-LLM Ensemble Translation Method for Creating a Multilingual Counseling Dialogue Dataset

Ryoma Suzuki, Zhiyang Qi, Michimasa Inaba

NepTam: A Nepali-Tamang Parallel Corpus and Baseline Machine Translation Experiments

Rupak Raj Ghimire, Bipesh Subedi, Balaram Prasain, Prakash Poudyal, Praveen Acharya, Nischal Karki, Rupak Tiwari, Rishikesh Sharma, Jenny Poudel, Bal Krishna Bal

Scoring the Translation: On Target Automatic Keyword-Based Evaluation of Machine Translation in the Sports Domain

Steinþór Steingrímsson and Einar Sigurdsson

Towards Improving Multimodal Machine Translation with LLMs: A Focus on Indic Languages

Amulya Ratna Dash, Chirag Wadhwa, Yashvardhan Sharma

P8.2 • Multilinguality and Translation Aids

Poster Area 2

Chair: Jan Niehues

Parallel Sentence Filtering for Low-Resource Language Pairs: A Case Study for Upper Sorbian, German, and Czech

Ruiyang Jiang, Shu Okabe, Alexander Fraser

OpenSubtitles2024: A Massively Parallel Dataset of Movie Subtitles for MT Development and Evaluation

Joerg Tiedemann and Hengyu Luo

CREST: Universal Safety Guardrails through Cluster-Guided Cross-Lingual Transfer

Lavish Bansal and Naman Mishra

Semantic Alignment across Ancient Egyptian Language Stages via Normalization-Aware Multitask Learning

He Huang

Conditioning LLMs to Generate Code-Switched Text

Maite Heredia, Gorka Labaka, Jeremy Barnes, Aitor Soroa

Are the LLMs Capable of Maintaining at Least the Language Genus?

Sandra Mitrović, David Kletz, Ljiljana Dolamic, Fabio Rinaldi

Gender Bias in MT for a Genderless Language: New Benchmarks for Basque

Amaia Murillo, Olatz Perez-de-Viñaspre, Naiara Perez

Optimizing Multilingual LLMs via Federated Learning: A Study of Client Language Composition

Aleix Sant, Jordi Luque, Carlos Escolano

Multilingual Target-Stance Extraction

Ethan Mines and Bonnie Dorr

MUNICHus: Multilingual News Image Captioning Benchmark

Yuji Chen, Alistair Plum, Hansi Hettiarachchi, Diptesh Kanojia, Saroj Basnet, Marcos Zampieri, Tharindu Ranasinghe

GlossMATE: Multi-Agent Translator Explanations for Glosses

Changbing Yang, Patrick Littell, Gabriel Bernier-Colborne, Yanfei Lu, Mengzhe Geng

Diagnosing Translated Benchmarks: An Automated Quality Assurance Study of the EU20 Benchmark Suite

Klaudia Thellmann, Bernhard Stadler, Michael Färber

Resource-Learn Lexicon Induction for German Dialects

Robert Litschko, Barbara Plank, Diego Frassinelli

P8.3 • Multimodality

Poster Area 2

Chair: Tomek Strzalkowski

FENCE: A Financial and Multimodal Jailbreak Detection Dataset

Mirae Kim, Seonghun Jeong, Youngjun Kwak

Evaluating Multimodal Large Language Models on Vertically Written Japanese Text

Keito Sasagawa, Shuhei Kurita, Daisuke Kawahara

ProMQA-Assembly: Multimodal Procedural QA Dataset on Assembly

Kimihiro Hasegawa, Wiradee Imrattanaatrai, Masaki Asada, Susan Holm, Yuran Wang, Xuanang Zhou, Ken Fukuda, Teruko Mitamura

K-MIND: Korean Multimodal Interaction Data for Dyadic Conversation Analysis

Jae Hee Yang, Yuha Shin, Saim Shin, Je Woo Kim, Jin Yea Jang

Do Multimodal LLMs Understand Order? Measuring the Fragility of Multimodal Reasoning under Input Order Perturbations

Sheng-Lun Wei, Yu-Ling Liao, Hen-Hsen Huang, Hsin-Hsi Chen

Early Fusion with Contrastive Learning: A Lightweight Alternative for Multi-modal Classification

Felix Wernlein, Abhik Jana, Sandipan Sikdar

Multimodal Entrainment and Feedback in Online Group Meetings

Patrizia Paggio, Manex Agirrezabal, Giulia Di Cristina, Bart Jongejan, Costanza Navarretta

MMCIg: Multimodal Cover Image Generation for Text-only Documents and Its Dataset Construction via Pseudo-labeling

HYEYON KIM, Sungwoo Han, Jingun Kwon, Hidetaka Kamigaito, Manabu Okumura

Multimodal Reference by Means of the Pronoun We and Hand Gestures in a Novel Corpus of Parliamentary Opening Debates

Costanza Navarretta

Multimodal Large Language Models for Low-Resource Languages: A Case Study for Basque

Lukas Arana, Julen Etxaniz, Ander Salaberria, Gorka Azkune

Real-Time Generation of Game Video Commentary with Multimodal LLMs: Pause-Aware Decoding Approaches

Anum Afzal, Yuki Saito, Hiroya Takamura, Katsuhito Sudoh, Shinnosuke Takamichi, Graham Neubig, Florian Matthes, Tatsuya Ishigaki

ARB: A Comprehensive Arabic Multimodal Reasoning Benchmark

Sara Ghaboura, Shubham Patle, Ketan More, Wafa Alghallabi, Omkar Thawakar, Jorma Laaksonen, Hisham Cholakkal, Salman Khan, Rao Anwer

Event Chronography in Multi-modal Data: The BME Method for Quantitative Analyses

Anaïs Murat, Maria Koutsombogera, Carl Vogel

CANVAS: A Multimodal Dataset of Chinese Textbook Images for Bias and Representation Analysis

Haotian Zhu, Kefan Yu, Min Li

MM-Conv: A Multimodal Dataset and Benchmark for Context-Aware Grounding in 3D Dialogue
Anna Deichler, Jim O'Regan, Fethiye Irmak Dogan, Anna Klezovich, Lubos Marcinek, Iolanda Leite, Jonas Beskow

Erase Persona, Forget Lore: Benchmarking Multimodal Copyright Unlearning in Large Vision Language Models

June Hyoungh Kwon, Jungmin Yun, Youngbin Kim

DREAM: A Multicultural Multimodal Dataset Linking Dialogues and Realistic Image Sequences

Juan Mallo, Marcos Esteche-Garitagoitia, Ricardo Cordoba, Luis Fernando D'Haro

Multimodal Task Interference: A Benchmark and Analysis of History-Target Mismatch in Multimodal LLMs

Masayuki Kawarada, Tatsuya Ishigaki, Hiroya Takamura

P8.4 • Cross-modality

Poster Area 2

Chair: Anton Ingason

Can Video LLMs See Through Illusions? Video-Illusion QA Benchmark Dataset

Souto Ohira, Toshio Hirasawa, Mamoru Komachi

To Skip, to Swap or to Not Swap? Identifying Step Transition Types in Instructional Manuals

Hsiu-Yu Yang, Michael Roth, Andreas Bulling, Carina Silberer

Fruitcakes and Cupcakes Emerging from Noise: The ComposiGen Dataset of Compounds and Their Compositionality

Jule Godbersen, Sinan Kurtayigit, Emma Raimundo Schulz, Tonmoy Rakshit, Diego Frassinelli, Sabine Schulte im Walde, Carina Silberer

Large Language Models' Internal Perception of Symbolic Music

Andrew Shin and Kunitake Kaneko

Entity Image and Mixed-Modal Image Retrieval Datasets

Cristian-Ioan Blaga, Paul Suganthan G C, Sahil Dua, Krishna Srinivasan, Enrique Alfonseca, Peter Dornbach, Tom Duerig, Imed Zitouni, Zhe Dong

Generating Sign Language Poses from HamNoSys and Natural Language Descriptions

Santiago Máximo and Luis Chiruzzo

Evaluating Discriminability of Vision-Language Models

Masayasu Muraoka and Naoaki Okazaki

Seeing the Other Side: Diagnostic Tasks for Viewpoint Reasoning in Vision-Language Models

Makoto Takenaka and Hitomi Yanaka

Multi-modal, Multi-task, Multi-criteria Automatic Evaluation with Vision Language Models

Masanari Oi, Masahiro Kaneko, Naoaki Okazaki, Nakamasa Inoue

Challenges in Image-Caption Association in Portuguese: Evaluating the CLIP Model on the FM30K Dataset

Vitória Colonetti Benedet, Gustavo Lopes Tamiosso, Rafael Oleques Nunes, Dennis Giovanni Balreira

A Large-Scale Instruction-Tuning Dataset and Models for Slovenian Vision-Language Tasks
Matej Martinc and Domen Vrešč

Which Way Does Time Flow? A Psychophysics-Grounded Evaluation for Vision–Language Models
Shiho Matta, Lis Kanashiro Pereira, Peitao Han, Shigeru Kitazawa, Fei Cheng
I Came, I Saw, I Explained: Benchmarking Multimodal LLMs on Figurative Meaning in Memes
Shijia Zhou, Saif Mohammad, Barbara Plank, Diego Frassinelli

DEJIMA: A Novel Large-scale Japanese Dataset for Image Captioning and Visual Question Answering

Toshiki Katsube, Fukuhara Taiga, Kenichiro Ando, Yusuke Mukuta, Kohei Uehara, Tatsuya Harada

CLEVR-3D-DeRef

Mary Martin, Martha Palmer, Maria Pacheco

A Parallel Cross-Lingual Benchmark for Multimodal Idiomaticity Understanding

Dilara Torunoğlu-Selamet, Doğukan Arslan, Rodrigo Wilkens, Wei He, Doruk Eryiğit, Thomas Pickard, Adriana Pagano, Aline Villavicencio, Gülşen Eryiğit, Ágnes Abuczki, Aida Cardoso, Alesia Lazarenka, Dina Almassova, Amália Mendes, Anna Kanellopoulou, Antoni Brosa-Rodriguez, Baiba Valkovska, Beata Wojtowicz, Bolette Pedersen, Carlos Manuel Hidalgo-Ternero, Chaya Liebeskind, Danka Jokić, Diego Alves, Eleni Triantafyllidi, Erik Velldal, Fred Philipp, Giedre Valunaite Oleskeviciene, Ieva Rizgeliene, Inguna Skadina, Irina Lobzhanidze, Isabell Haugen, Jauza Akbar Krito, Jelena Marković, Johanna Monti, Josue Sauca, Kaja Dobrovoljc, Kingsley Ugwuanyi, Laura Rituma, Lilja Øvrelid, Maha Tufail Agro, Manzura Abjalova, Maria Chatzigrigoriou, María del Mar Sánchez Ramos, Marija Pendevska, Masoumeh Seyyedrezaei, Mehrnough Shamsfard, Momina Ahsan, Muhammad Ahsan Khan, Nathalie Norman, Nilay Erdem Ayyıldız, Nina Hosseini-Kivanani, Noémi Ligeti-Nagy, Numaan Naeem, Olha Kanishcheva, Olha Yatsyshyna, Daniil Orel, Petra Giommarelli, Petya Osenova, Radovan Garabik, Regina Semou, Rozane Rebechi, Salsabila Zahirah Pranida, Samia Touileb, Sanni Nimb, Sarfraz Ahmad, Sarvinoz Sharipova, Shahar Golan, Shaoxiong Ji, Sopuruchi Aboh, Srdjan Sucur, Stella Markantonatou, Sussi Olsen, Vahide Tajalli, Veronika Lipp, Voula Giouli, Yelda Yeşildal Eraydın, Zahra Saaberi, Zhuohan Xie

P8.5 • Sign Languages

Poster Area 2

Chair: Dominik Schlechtweg

Bridging Text-to-Sign Translation via Codebook-Oriented Pretraining

Ninlawat Phuangchoke and Chantri Polprasert

A Resource and Evaluation Method for Phonological Continuity in Japanese Sign Language

Jundai Inoue, Daisuke Hara, Makoto Miwa

Sentiment Analysis of German Sign Language Fairy Tales

Fabrizio Nunnari, Siddhant Jain, Patrick Gebhard

A Critical Study of Automatic Evaluation in Sign Language Translation

Shakib Yazdani, Yasser HAMIDULLAH, Cristina España-Bonet, Eleftherios Avramidis, Josef van Genabith

How Much Data Is Enough Data? A New Motion Capture Corpus for Probabilistic Sign Language Generation

Anna Klezovich, Johanna Mesch, Gustav Eje Henter, Jonas Beskow

Decomposing Sign Language Movements: A Multi-Band Visualization Method for Articulatory Analysis

Antonio F. G. Sevilla and José María Lahoz-Bengoechea

10:40

Coffee Break

11:00

Session O33: Psycholinguistics, Cognitive Linguistics and Linguistic Theories

Auditorium Illes Balears

Chair: Jelke Bloem

11:00 Implicit Bias in Peer Review: Through the Lens of Language Abstraction

11:20 *Xulang Zhang, Rui Mao, Erik Cambria*

11:20 The PARLO Dementia Corpus: A German Multi-Center Resource for Alzheimer's Disease

11:40 *Franziska Braun, Christopher Witzl, Florian Hönig, Elmar Nöth, Tobias Bocklet, Korbinian Riedhammer*

11:40 Lexical and Discourse Semantics in a Reading-time Corpus of English

12:00 *Jakub Dotlacil, Laia Fortuny, Li Kloostra, Johan Bos*

12:00 Semantic Capacity in Language Learners and LLMs: A Case Study of Quantifier Scopepe

12:20 *Shaohua Fang, Yue Li, Yan Cong*

12:20 How Long Does a Quick Kiss Take? Studying Event Duration of Light Verb Constructions Using Explicit Word Embeddings

Lin de Huybrecht and Geraint Wiggins

Session O34 • Opinion and Argument Mining

Auditorium Mallorca

Chair: Valerio Basile

11:00 Disambiguation of Emotion Annotations by Contextualizing Events in Plausible Narratives

11:20 *Johannes Schaefer and Roman Klinger*

11:20 Identifying Contexts of Distress in College Students' Reddit Posts: A Comparative Study of

11:40 Classical NLP and Large Language Models

Carine Graff and Nikhil Krishnaswamy

11:40 TiC-MuFormer: Time-Aware Caption-Integrated Multimodal Transformers for User-Level

12:00 Mental Health Modeling

Georgios Tsoumplekas, Yannis Spyridis, Vasileios Argyriou

12:00 Improving Neural Argumentative Stance Classification in Controversial Topics with Emotion-Lexicon Features

12:20 *Mohammad Yeghaneh Abkenar, Weixing Wang, Manfred Stede, Mark Finlayson, Davide Picca, Panagiotis Ioannidis*

- 12:20 Emotion Transcription in Conversation: A Benchmark for Capturing Subtle and Complex
 12:40 Emotional States through Natural Language
Yoshiki Tanaka, Ryuichi Uehara, Koji Inoue, Michimasa Inaba
-

Session O35 • Parsing

Menorca 1

Chair: Menno van Zaanen

- 11:00 SETUP: Sentence-level English-To-Uniform Meaning Representation Parser
 11:20 *Emma Markle, Javier Gutierrez Bach, Shira Wein*
- 11:20 This One or That One? A Study on Accessibility via Demonstratives with Multimodal Large
 11:40 Language Models
Yu Wang, Emmanuele Chersoni, Chu-Ren Huang
- 11:40 AMR Parsing beyond English: An Experiment on Bulgarian, French, Hungarian and Ukrainian
 12:00 *Ivaylo Mitov, Tadzhat Marharian, Zsofia Hauk, Samba FALL, Maxime Amblard, Bruno Guillaume*
- 12:00 Semantic Parsing for Evaluating Large Language Models: Separating Linguistic Abilities
 12:20 with YARN
Rémi DE VERGNETTE and Maxime Amblard
- 12:20 Two Ojibwe Constraint Grammars: Morphological Disambiguation and Dependency Parsing
 12:40 *Matthias Diederichsen and Christopher Hammerly*
-

Session O36 • Multimodality and Speech

Eivissa 1

Chair: Jonas Beskow

- 11:00 Multimodal LLMs Do Not Compose Skills Optimally across Modalities
 11:20 *Paula Ontalvilla, Aitor Ormazabal, Gorka Azkune*
- 11:20 Code-Switching in End-to-End Automatic Speech Recognition: A Systematic Literature Review
 11:40 *Maha Tufail Agro, Atharva Kulkarni, Karima Kadaoui, Zeerak Talat, Hanan Aldarmaki*
- 11:40 MUSTReason: A Benchmark for Diagnosing Pragmatic Reasoning in VideoLMs for
 12:00 Multimodal Sarcasm Detection.
Anisha Saha, Varsha Suresh, Timothy Hospedales, Vera Demberg
- 12:00 Human-Centered Multimodal Fusion for Sexism Detection in Memes with Eye-Tracking,
 12:20 Heart Rate, and EEG Signals
Iván Arcos Gabaldón, Paolo Rosso, Elena Gomis Vicent
- 12:20 Nos_Brais-GL: A FAIR Galician TTS Corpus for Neural Speech Synthesis
 12:40 *Adina Vladu, Antonio Moscoso Sánchez, Carmen Magariños, María Perez Lago, Elisa Fernández Rei*

P9.1 • Natural Language Generation

Poster Area 1

Chair: Victoria Arranz

DR-CUP: A Dataset on Real-time Commentary in U.S. Presidential Debates

Yu-Yu Chang, Huan-Wen Ho, Chung-Chi Chen, Ming-Hung Wang

Russian Generative Spelling, Punctuation and Capitalization Correction

Nikita Martynov, Danil Astafurov, Ulyana Isaeva, Ivan Maksimov, Joqsan Azocar, Dmitrii Kosenko, Alena Fenogenova

Using Multimodal and Language-Agnostic Sentence Embeddings for Abstractive Summarization

Chaimae Chellaf El Hammoud, Salima Mdhaffar, Yannick Estève, Stéphane Huet

Gradient-Controlled Decoding: A Safety Guardrail for LLMs with Dual-Anchor Steering

Purva Chiniya, Kevin Scaria, Sagar Chaturvedi

The Chronicles of RiDiC: Generating Datasets with Controlled Popularity Distribution for Long-form Factuality Evaluation

Pavel Braslavski, Dmitrii Iarosh, Nikita Sushko, Andrey Sakhovskiy, Vasily Konovalov, Elena Tutubalina, Alexander Panchenko

MeteoGalEus: An Iberian Multilingual Weather Dataset in Galician, Euskera, and Spanish

Ainhoa Vivel-Couso, Nella Zabrina Pramata, David Robredo, Aitor Soroa, Jose Maria Alonso-Moral

RadTimeline: Timeline Summarization for Longitudinal Radiological Lung Findings

Sitong Zhou, Meliha Yetisgen, Mari Ostendorf

InstructSum: A Benchmark to Evaluate Instruction-Following Capability of Large Language Models in Summarization

Kosuke Nishida, Kyosuke Nishida, Itsumi Saito

NOVELSUM: Evaluating Long-Form Summary Generation for Historical Scandinavian Novels

Ali Al-Laith, Alexander Conroy, Kirstine Degn, Jens Bjerring-Hansen, Daniel Hershcovich

Evaluating Large Language Models for Text-to-Gloss Translation in Kazakh-Russian Sign Language: A Pilot Study

Zhanibek Kozhimbayev and Alfarabi Imashev

HotelCheckSpan: A Benchmark Dataset for LLM Faithfulness

Patricia Schmidtova, Ondrej Dusek, Saad Mahamood

P9.2.1 • Machine Learning II

Poster Area 1

Chair: Udo Kruschwitz

Procrustes Analysis for Improving Language Model Merging

Olivier Ferret

MetaCORA: A Meta-Learned Curriculum for Adversarial and Contrastive Robustness in Speech Recognition

Yuqian Dai, Chun Fai Chan, Ying Ki Wong, Tsz Ho Pun

Insights from Transfer Learning Experiments with Word-in-Context and Word Sense Disambiguation Models

Alp Mujko and Dominik Schlechtweg

Joint Identification and Induction of Semantic Frames with Scalable Semi-Supervised Graph Clustering

Fabian Barteld, Steffen Remus, Saba Anwar, Julian Stawecki, Alexander Ziem, Chris Biemann

Low-Rank Compression of Language Models via Differentiable Rank Selection

Sidhant Sundrani, Francesco Tudisco, Pasquale Minervini

Self-supervised Data Augmentation for Text Classification in Low-Data Settings

Deyu Ding, Mengying Wang, Andreas Spitz

Distribution-aware Low-bitwidth Quantization for Large Language Models

Bao Huynh, Takashi Tsunakawa, Masafumi Nishida

TG-ASR: Translation-Guided Learning with Parallel Gated Cross Attention for Low-Resource Automatic Speech Recognition

ChengYeh Yang, Chien-Chun Wang, Li-Wei Chen, Hung-Shin Lee, Hsin-Min Wang, Berlin Chen

Harnessing Synergy in Context and Emoji for Joint Detection of Harmful Online Content in Multi-turn Conversations

Feiyan Hu, Ciara Byrne, Jiang Zhou, Rena Maycock, Mark Langan

Dynamic Layer Selection for Efficient Tone Recognition in Self-Supervised Speech Models

Saint Germe BENGONO OBIANG, Norbert TSOPZE, Paulin MELATAGIA YONTA

Intent Recognition in Speech-to-Text Processing in the Context of Natural Interaction with Cognitive Assistive Systems

Behnam Ensan, Magnus Jung, Matthias Busch, Adreas Wendemuth

Merging Continual Pretraining Models for Domain-Specialized LLMs: A Case Study in Finance

Kentaro Ueda, François Portet, Hirohiko Suwa, Keiichi Yasumoto

Phonetic-based Ranking for Improved Pseudo-Labeling in Low-Resource ASR

Marco Matassoni, Roberto Gretter, Falavigna Daniele, Mohamed Nabih Ali Mohamed Nawar, Alessio Brutti, Matteo Negri, Mauro Cettolo, Marco Gaido, Sara Papi, Luisa Bentivogli

Privacy-Preserving Information Extraction with Local LLMs: A Comparative Study on Dutch Debt Collection Letters

Beyza Celep, Natalia Amat-Lefort, Joost Visser

P9.2.2 • Machine Learning III

Poster Area 1

Chair: Marko Tadić

Forewarned Is Forearmed: When Non-Sequential Embedding Turns into an Anomaly Detector

Elys Allesiaro, Antoine Caubrière, Valentin Vielzeuf

A Joint Detection Framework for Latvian Loanwords and Calques Using Monolingual Data

Yelinyun Zhang, Atis Kapenieks, Marina Platonova

Pantagruel: Unified Self-Supervised Encoders for French Text and Speech

Phuong-Hang Le, Valentin Pelloin, Arnault Chatelain, Maryem Bouziane, Mohammed Ghennai, Qianwen Guan, Kirill Milintsevich, Salima Mdahaffar, Aidan Mannion, Nils Defauw, Shuyue Gu, Alexandre Audibert, Marco Dinarelli, Yannick Estève, Lorraine Goeuriot, Steffen Lalande, Nicolas Hervé, Maximin Coavoux, François Portet, Étienne Ollion, Marie Candito, Maxime Peyrard, Solange Rossato, Benjamin Lecouteux, Aurélie Nardy, Gilles Sérasset, Vincent Segonne, Solène Evain, Diandra Fabre, Didier Schwab

Merge and Conquer: Instructing Multilingual Models by Adding Target Language Weights

Eneko Valero, Maria Ribalta i Albado, Oscar Sainz, Naiara Perez, German Rigau

SemiAdapt: Semi-Supervised and Efficient LoRA-Based Domain Adaptation for Low-Resource Irish Machine Translation with Transformers

Josh McGiff and Nikola Nikolov

Data Selection Effects on Self-Supervised Learning of Audio Representations for French Audiovisual Broadcasts

Valentin Pelloin, Lina Bakkali, Reda Dehak, David Doukhan

SENS-ASR: Semantic Embedding Injection in Neural-transducer for Streaming Automatic Speech Recognition

Youness Dkhissi, Valentin Vielzeuf, Elys Allesiaro, Anthony Larcher

Efficient Financial Language Understanding via Distillation with Synthetic Data

Wen-Fong (Xavier) Huang and Edwin Simpson

Rubric-Guided Fine-tuning of SpeechLLMs for Multi-Aspect, Multi-Rater L2 Reading-Speech Assessment

Aditya Kamlesh Parikh, Cristian Tejedor-García, Catia Cucchiari, Helmer Strik

Leveraging Semi-Supervised Learning for Multimodal Hate Speech Data Annotation and Detection

Rathi Adarshi Rammohan, Zhao Ren, Dominik Puchała, Aleksandra Świdarska, Dennis Küster, Tanja Schultz

Lexicalized Constituency Parsing for Middle Dutch: Low-resource Training and Cross-Domain Generalization

Yiming Liang and Fang Zhao

Quantifying the Accuracy and Cost Impact of Design Decisions in Budget-Constrained Agentic LLM Search

Kyle McCleary and James Ghawaly

Reason-to-Learn (R2L): Multi-Agent Knowledge Distillation for Lightweight LLMs in Sentiment Analysis

Le-Huy Tu, Quan Nguyen, Vincent NGUYEN, Johanna Bjorklund, Xuan-Son Vu

PRISM: Partial Ranking via Inter-layer Semantic Measurement for Efficient Fine-tuning of Language Models

Aldrin Biswas, Md Fahim, Md. Amin, Amin Ali, AKM Rahman

P9.3.1 • Language Modeling and LRs III

Poster Area 1

Chair: Monica Monachini

Beyond Literal Meaning: How LLMs Interpret Yemeni Proverbs

Nasser Thmer, Ali Al-Laith, Muhammad Shoaib

SEFL: A Framework for Generating Synthetic Educational Assignment Feedback with LLM Agents

Mike Zhang, Amalie Dilling, Léon Gondelman, Niels Lyngdorf, Euan Lindsay, Johannes Bjerva

LGSE: Lexically Grounded Subword Embedding Initialization for Low-Resource Language Adaptation

Hailay Kidu Teklehaymanot, Dren Fazlija, Wolfgang Nejdl

A Cheap Lunch: Synthetic Annotation With Reduced Human Effort for Medical Text Mining

Shutao Chen and Piek Vossen

Supervised Contrastive Fine-Tuning for Active Few-Shot Learning

Zirui Zhang, Lei Ge, Shengyu Qiao

Simulating Student Interactions for Virtual Pretesting with In-Context Learning

Arthur Thuy, Luca Benedetto, Ekaterina Loginova, Dries Benoit

An Exploration-Analysis-Disambiguation Reasoning Framework for Word Sense Disambiguation with Low-Parameter LLMs

Deshan Sumanathilaka, Nicholas Micallef, Julian Hough

Building Effective Japanese Medical LLMs with an Open Recipe for Domain Adaptation through Continued Pre-training

Akiko Aizawa, Yuki Arase, Fei Cheng, Jiahao Huang, Zhiyi Huang, Junfeng Jiang, Teruhito Kanazawa, Daisuke Kawahara, Kazuma Kobayashi, Takashi Kodama, Sadao Kurohashi, Yusuke Oda, Yuma Tsuta, Zhen Wan, Zhishen Yang, Rio Yokota

New Encoders for German Trained from Scratch: Comparing ModernGBERT with Converted LLM2Vec Models

Julia Wunderle, Anton Ehrmanntraut, Jan Pfister, Fotis Jannidis, Andreas Hotho

Arabic ChartSumm: An English-to-Arabic Benchmark for Metadata-to-Text Summarization

Passant Elchafei and Amany Fashwan

Introducing a Bangla Sentence – Gloss Pair Dataset for Bangla Sign Language Translation and Research

Neelavro Saha, Rafi Shahriyar, Nafis Roudra, Saadman Sakib, Annajiat Rasel

Language Models as Semantic Augmenters for Sequential Recommenders

Mahsa Valizadeh, Xiangjue Dong, Rui Tuo, James Caverlee

Efficient Adaptation of English Language Models for Morphologically Rich and Underrepresented Languages: The Case of Arabic

Ahmed Eldamaty, Mohamed Abdelrahman, Mohamed Elbehery, Mariam Ashraf, Radwa Elshawi

P9.3.2 • Language Modeling and LRs IV

Poster Area 1

Chair: Simonetta Montemagni

GhostWriter: Hidden AI-Generated Texts over Multiple Languages, Domains and Generators

Manuel Schaaf, Kevin Bönisch, Alexander Mehler

Using LLMs to Extract Instances of Schematic Constructions from Unannotated L2 Learner Corpora

Jelena Kallas, Ahto Kiil, Heete Sahkai, Geda Paulsen, Kertu Saul

Corruption-Based Data Augmentation for Arabic Essay Scoring: A Preliminary Study on the Organization Trait

May Bashendy and Tamer Elsayed

Structured Prompting for Arabic Essay Proficiency: A Trait-Centric Evaluation Approach

Salim Al Mandhari, Hieu Pham Dinh, Mo El-Haj, Paul Rayson

ManufactuBERT: Efficient Continual Pretraining for Manufacturing

Robin Armingaud and Romaric Besancon

Śmigiel Dataset: Laying Foundations for Investigating Machine-Generated Text Detection in Polish

Jakub Strebeyko, Alina Wróblewska, Piotr Przybyła

Extracting Medical Image-Related Entities from Spanish Electronic Health Records Using NER Methods

Alexander Platas, Marcos Merino, Elena Zotova, Montse Cuadros, Karen López-Linares, Mikel Pérez de Mendiola, María Gálvez, Cristina Barba, Antón Asla

A Novel Synthetic Dataset for Few-Shot Legal Relation Extraction in German

Shiva Banasaz Nouri, Elena Leitner, Julian Moreno-Schneider, Georg Rehm

LLM-Based Data Generation and Clinical Skills Evaluation for Low-Resource French OSCEs

Tian Huang, Tom Bourgeade, Irina Illina

Instruction-Tuned Urdu LLMs: Efficient Adaptation of Llama Models and Evaluation Resources for Urdu

Munief Tahir, Sana Shams, Sarmad Hussain, Miriam Butt

Is Biomedical Specialization Still Worth It? Insights from Domain-Adaptive Language Modelling with a New French Health Corpus

Aidan Mannion, Cécile Macaire, Armand Violle, Stéphane Ohayon, Xavier Tannier, Didier Schwab, Lorraine Goeuriot, François Portet

TildeOpen LLM: Leveraging Curriculum Learning to Achieve Equitable Language Representation

Toms Bergmanis, Ingus Pretkalniņš, Martins Kronis, Davis Nicmanis, Jeļizaveta Jelinska, Roberts Rozis, Rinalds Viksna, Marcis Pinnis

Common Sense vs. Morality: The Curious Case of Narrative Focus Bias in LLMs

Saugata Purkayastha, Pranav Kushare, Pragya Pal, Sukannya Purkayastha

P9.3.3 • Language Modeling and LRs V

Poster Area 1

Chair: Lluís Padro

“Emphasizing the Commendable”: A Study of Homogenized Transitive Verb Constructions in Machine Generated Peer Reviews

Hing-Yuet Fung, Chi-kiu Lo, Samuel Larkin

CoDAE: Adapting Large Language Models for Education via Chain-of-Thought Data Augmentation

Shuzhou Yuan, William LaCroix, Hardik Ghoshal, Ercong Nie, Michael Färber

Synthetic Instruction Generation for Low-Resource Nordic Languages: Viability and Limitations in LLM Instruction-Tuning

Mathias Stenlund, Annika Simonsen, Lars Bungum, Jan Ebert, Jiangtao Wang, Oleg Filatov, Hemanadhan Myneni, Morris Riedel, Hafsteinn Einarsson

AYN: A Tiny Yet Competitive Indian Legal Language Model Pretrained from Scratch

Mitodru Niyogi, Eric Gaussier, Arnab Bhattacharya

Low-Resource Dialect Adaptation of Large Language Models: A French Dialect Case-Study

Eeham Khan, Firas Saidani, Owen Van Esbroeck, Richard Khoury, Leila Kosseim

Reformulate and Create, Don't Translate: Creating Natural Prompts for Underserved Languages

Annika Simonsen, Mathias Stenlund, Lars Bungum, Marc Volhardt, Hafsteinn Einarsson

Generating High Quality Synthetic Data for Dutch Medical Conversations

Cecilia Kuan, Aditya Kamlesh Parikh, Henk van den Heuvel

DeepICD-R1: Medical Reasoning through Hierarchical Rewards and Unsupervised Distillation

Tom Röhr, Thomas Steffek, Roman Teucher, Keno Bressem, Alexei Figueroa, Paul Grundmann, Peter Troeger, Felix Gers, Alexander Löser

SynthLLM: An LLM-based Scalable Synthetic Data Generation Pipeline for Low-Resource Languages

Solmaz Panahi, Vasudevan Nedumpozhimana, John Kelleher

Persona-Conditioned Generation of Patient Self-Reports from EHRs

Yuxin Wu, jianming wei, Vasile Rus

SocialStep: Fast Prediction of Social Determinants of Health

Paul Landes, Adam Cross, Jimeng Sun

Dynamically Acquiring Text Content to Enable the Classification of Lesser-known Entities for Real-world Tasks

Fahmida Alam and Ellen Riloff

RILEC: Detection and Generation of L1 Russian Interference Errors in English Learner Texts

Darya Kharlamova and Irina Proskurina

12:40

Lunch Break

14:10

14 :10

15:00

Keynote Speaker: Dan Jurafsky

The Social Failures of Language Models as Conversational Partners

Auditorium Illes Balears

Chair: Nancy Ide

14:55

15:00

Short Break (5mn)

Session O37 • Evaluation, Validation, Quality Assurance

Auditorium Illes Balears

Chair: Joakim Nivre

15:00

Critical Foreign Policy Decision (CFPD) Benchmark: Measuring Diplomatic Preferences of Large Language Models

15:20

Benjamin Jensen, Ian Reynolds, Yasir Atalan, Michael Garcia, Austin Woo, Anthony Chen, Trevor Howarth

15:20

CrisisCL: A Domain Incremental Learning Benchmark for Crisis Management

15:40

Paul Le Van Kiem, Romain Meunier, Farah Benamara, Véronique MORICEAU

15:40

Towards Consistent Detection of Cognitive Distortions: LLM-Based Annotation and Dataset-Agnostic Evaluation

16:00

Neha Sharma, Navneet Agarwal, Kairit Sirts

16:00

LLMs as Annotators: Evaluating Model–Human Alignment in Detecting Contentious Language in Historical Corpora

16:20

Yahui Zhao, Clemencia Siro, Laura Hollink

16:20

Widespread Gender and Pronoun Bias in Moral Judgments across LLMs

16:40

Gustavo Fernandes, Jeiverson Santos, Pedro O.S Vaz-de-Melo

Session O38 • Knowledge Discovery and Representation

Auditorium Mallorca

Chair: Gilles Sérasset

15:00

Frame2KG: A Benchmark and Evaluation Toolkit for Interpretable Frame-to-Graph Generation

15:20

Lewis Watson, Carl Strathearn, Kenny Mitchell, Yanchao Yu

- 15:20 Injecting Structured Biomedical Knowledge into Language Models: Continual Pretraining vs.
15:40 GraphRAG
Jaafar Klila, Sondes Bannour Souihi, rahma boujelbane, Nasredine Semmar, Lamia Hadrich-Belguith
- 15:40 Linguistic Knowledge Graphs for Sense Prediction: A Case-study on Latin
16:00 *Eleonora Ghizzota, Paola Marongiu, Pierpaolo Basile, Stefano Ferilli, Barbara McGillivray*
- 16:00 ACID: On the Perception of Online Classism
16:20 *Arianna Muti, Elisa Bassignana, Amanda Cercas Curry, Federica Durante, Dirk Hovy, Debora Nozza*
- 16:20 The Spectrum of Sentiment: Optimistic, Pessimistic, and Neutral Voices in Online
16:40 Depression Discourse
Stefana Tabusca, Ana-Maria Bucur, Liviu Dinu
-

Session O39 • Applications Involving LRs and Evaluation III

Menorca 1

Chair: Andreas Hotho

- 15:00 A Benchmark Dataset and Comparative Evaluation of Phonemized and Romanized Urdu for
15:20 Text-to-Speech
M Kaab Bin Shahid and Muhammed Izharuddin
- 15:20 S-VoCAL: A Dataset and Evaluation Framework for Inferring Speaking Voice Character
15:40 Attributes in Literature
Abigail Berthe-Pardo, Gaspard Michel, Elena Epure, Christophe Cerisara
- 15:40 BankMathBench: A Benchmark for Numerical Reasoning in Banking Scenarios
16:00 *Yunseung Lee, Subin Kim, Youngjun Kwak, Jaegul Choo*
- 16:00 TR-TEB: Turkish Text Embedding Benchmark
16:20 *Omer Arslan, Atalay Celik, Yusuf Aslan, Hasan Durkaya, Mustafa Zenginoglu, Musa Yilmaz, Merve Kantarci, Mehmet Haklidir*
- 16:20 Simple Additions, Substantial Gains: Expanding Scripts, Languages, and Lineage Coverage in
16:40 URIEL+
Mason Shipton, York Hay Ng, Aditya Khan, Phuong Hoang, Xiang Lu, A. Seza Dogruoz, Annie Lee
-

Session O40 • Multimodality, Cross-modality

Eivissa 1

Chair: Costanza Navarretta

- 15:00 SciClaimEval: Cross-modal Claim Verification in Scientific Papers
15:20 *Xanh Ho, Yun-Ang Wu, Sunisth Kumar, Tian Cheng Xia, Florian Boudin, Andre Greiner-Petter, Akiko Aizawa*
- 15:20 Localizing Events in Space: Comparing Humans and AI Models
15:40 *Derrick Eui Gyu Kim, Kenneth Lai, James Pustejovsky*

15:40	STRUDEL: Unrolling a Benchmark for Evaluating Vision-Language Models on Structured
16:00	Diagram Understanding across Domains <i>Daniel Steinigen, Lucie Flek, Sebastian Houben</i>
16:00	VG-CoT: Towards Trustworthy Visual Reasoning via Grounded Chain-of-Thought
16:20	<i>Byeonggeuk Lim, Kyeonghyun Kim, Jungmin Yun, Youngbin Kim</i>
16:20	VectorEdits: A Dataset and Benchmark for Instruction-Based Editing of Vector Graphics
16:40	<i>Josef Kuchar, Marek Kadlcik, Michal Spiegel, Michal Stefanik</i>

15:00

16:40 **Poster Sessions - Poster Area 2**

P10.1 • Social Media

Poster Area 2

Chair: Thorsten Trippel

ViWikiFC: Fact-Checking for Vietnamese Wikipedia-Based Textual Knowledge Source
Hung Le, Long To, Manh Nguyen, Kiet Nguyen

Automated Extraction of Answer Candidates for Question Generation
Claudia Preda, Mihai Dascalu, Stefan Ruseti, Danielle McNamara

Green Bots versus Red Bots: Evaluating Large Language Models for Simulating Persuasion Dynamics in Online Influence Campaigns
Majd Al Ali, Filip Muntean, Lucia Donatelli, Jurriaan van Diggelen

Towards Expectation Detection in Language: A Case Study on Treatment Expectations in Reddit
Aswathy Velutharambath and Amelie Wühl

Empathy Speaks in Metaphors: The Empathy-Metaphor Corpus of Figurative Language in Empathetic Text
Gyeongeun Lee and Natalie Parde

A Computational Diachronic Analysis of Gen Z Mental Health Discourse: A Large-scale Reddit Corpus Study from Pre- to Post-COVID
Felix Mao

"Oat Milk Vegan Chocolate Taste Great!": Monitoring the Food Transition Debate in Reddit
Greta Zella, Jan Willem Bolderdijk, Saskia Peels, Gerry Wakker, Tommaso Caselli

ClimateChat-300K: A Multi-Modal Facebook Dataset for Understanding Diverse Perspectives in Climate Communication
Wajdi Zaghouani, Md. Rafiul Biswas, Mabrouka Bessghaier, Shima Ibrahim, George Mikros

HateMirage: An Explainable Multi-Dimensional Dataset for Decoding Faux Hate and Subtle Online Abuse
Sai Kartheek Reddy Kasu, Shankar Biradar, SUNIL SAUMYA, Md. Shad Akhtar

MindSET: Advancing Mental Health Benchmarking through Large-Scale Social Media Data
Saad Mankarious, Edward Kempa, Daniel Wiechmann, Elma Kerz, Yu Qiao, Ayah Zirikly

A Corpus of Misunderstood Irony on Turkish Social Media
Çağrı Çöltekin and Güliz Güneş

P10.2.1 • Linguistics and Psycholinguistics I

Poster Area 2

Chair: Aditya Kamlesh Parikh

A Corpus of Joint EEG and Self-Paced Reading of Natural Dutch Texts

Sara Østergaard, Lenneke Lichtenberg, Laura Boon, Bruno Nicenboim

Evaluation Drift in LLM Personality Induction: Are We Moving the Goalpost?

Prateek Rajput, Yewei Song, Iyiola Olatunji, Jacques Klein, Tegawendé Bissyandé

A Multi-Dialectal, Longitudinal Corpus of Human-AI Hybrid Language Production

Qiao Gan, Jonathan Dunn, Andrea Nini, Benjamin Adams

Semantic Information: A Difference That Makes a Difference

J. Nathanael Philipp, Max Kölbl, Michael Richter

Modeling the Memory-Surprisal Trade-Off over Time: Communicative Efficiency Decreases with Lexico-Grammatical Change in Scientific English

Julius Steuer, Marie-Pauline Krielke, Stefania Degaetano-Ortlieb, Elke Teich, Dietrich Klakow

Mechanistic Interpretability Meets Cognitive Linguistics: Modelling Locative Image Schemas in the Circuit Framework

Mattia Proietti, Afra Alishahi, Grzegorz Chrupała, Alessandro Lenci

Variation Is the Norm: Embracing Sociolinguistics in NLP

Anne-Marie Lutgen, Alistair Plum, Verena Blaschke, Barbara Plank, Christoph Purschke

Appraisal Theory-Informed Emotion Prediction

Xiaowei Wang, Jayant Teotia, Rui Mao, Wandeep Ratan Singh, Sabrina Tiun, Erik Cambria

The Evolution of Philosophy: A Metaphorical Cognition Perspective

Rui Mao, Dapeng Chen, Zihao Huang, Xulang Zhang, Erik Cambria

P10.2.2 • Linguistics and Psycholinguistics II

Poster Area 2

Chair: Cecilia Kuan

Predicting States of Understanding in Explanatory Interactions Using Cognitive Load-Related Linguistic Cues

Yu Wang, Olcay Türk, Angela Grimminger, Hendrik Buschmeier

Figurative Language in Alzheimer's Discourse: Linguistic and Neural Alignment in Clinical Narratives

Diana Klymnyk, Vitória Tomasel, Helena Caseli, Edward Watkins, Aline Villavicencio, Rodrigo Wilkens

Prompting Instruction-tuned LLMs for Semantic Similarity Values

Xander Snelder, Yunchong Huang, Jelke Bloem

Towards Dynamic Metaphor Identification: Evaluating GPT O-Series Models on Five Metaphoricity Cues in U.S. Trade Corpora

Berkay Bas, Jelke Bloem, Xiaojuan Tan

Rethinking Evaluation in Retrieval-Augmented Personalized Dialogue: A Cognitive and Linguistic Perspective

Tianyi Zhang and David Traum

Evaluating Multimodal Large Language Model Narrative Interpretation through the Lens of Appraisal Theory

Jayant Teotia, Xiaowei Wang, Xulang Zhang, Rui Mao, Erik Cambria

Mapping Liberty Metaphors across Cultures and Time

Sidney Suen, Rui Mao, Kenneth Kwok, Erik Cambria

The Sensorimotor Norms for the Chinese Classifiers

Yimei Shao, Yu-Yin Hsu, Chu-Ren Huang

DeepQuestion: Systematic Generation of Real-World Challenges for Evaluating LLMs Performance

Ali Khoramfar, Ali Ramezani, Mohammad Mahdi Mohajeri, Mohammad Javad Dousti, Majid Nili Ahmadabadi, Hesham Faili

Pragmatic Modelling in Language Learning: Caregiver Question-Answer Feedback in Child-Directed Dialogue

Maryam Bala, Johannes Heim, Elspeth Edelstein, Arabella Sinclair

P10.3.1 • Parsing and Tagging I

Poster Area 2

Chair: Patrick Paroubek

Modular Approach to Automating Morphological Components in Grammar Engineering

Ekaterina Voloshina and Krasimir Angelov

MorfFlex: Handling Rich Morphology

Jaroslava Hlaváčková, Marie Mikulová, Barbora Štěpánková, Milan Straka, Jan Hajič

Using Valency Inheritance in Building a Valency Lexicon

Václava Kettnerová, Veronika Kolářová, Jiří Mirovský, Michal Olbrich

From CHAT to Coded CoNLL-U: A Reproducible Pipeline for the Syntactic Annotation and Querying of Child Language Data

Achim Stein

TækTæk: Syntactic Analysis of Language Use on Danish TikTok

Thea Kristensen and Rob van der Goot

Adaptive Chunking: Optimizing Chunking-Method Selection for RAG

Paulo de Moura Júnior, Jean Lelong, Annabelle Blangero

Do Large Language Models Grasp the Grammar? Evidence from Grammar-Book-Guided Probing in Luxembourgish

Lujun Li, Yewei Song, Lama Sleem, Yiqun Wang, Yangjie Xu, Cedric LOTHRITZ, Niccolo' Gentile, Radu State, Tegawendé Bissyandé, Jacques Klein

Survey of Tools for Manual Linguistic Annotation: Supporting Diversity through Interactive Exploration

Ludovica Pannitto, Kaja Dobrovolic, Bruno Guillaume

TextLens & LeTTuce: Automated Corpus Annotation and Multilingual Tagging as a Service
Cynthia Van Hee, Jonas Doumen, Vincent Prins, Pranaydeep Singh, Vincent Vandeghinste, Els Lefever

The Corpus of Contemporary Polish - a New Reference Corpus with Rich Syntactic Annotations

Witold Kieraś, Małgorzata Marciniak, Marcin Woliński, Katarzyna Krasnowska-Kieraś, Marek Łaziński

Prague Dependency Treebank - Consolidated 2.0: Enriching a Complex Annotation Scheme
Marie Mikulová, Jiří Mírovský, Milan Straka, Pavlína Synková, Jan Štěpánek, Barbora Štěpánková, Jan Hajič

Meet UD_Czech-PDTC: A Large and Genre-Rich Treebank in Universal Dependencies
Marie Mikulová, Barbora Štěpánková, Daniel Zeman, Jan Štěpánek, Milan Straka, Jan Hajič

Encoding Logical Relations of Chinese Complex Sentences within the Universal Dependencies Framework
Hongpu Zhu and Hongzhi Xu

Unsupervised Labelling of Mutation Triggers in Welsh
Nicolás Gutiérrez-Rolón and Fernando Alva-Manchego

P10.3.2 • Parsing and Tagging II

Poster Area 2

Chair: Amy Isard

UzUDT: Uzbek Universal Dependencies Treebank
Sanatbek Matlatipov and Mersaid Aripov

BRAGD: Constrained Multi-Label POS Tagging for Faroese
Annika Simonsen, Barbara Scalvini, Uni Johannesen, Iben Debess, Hafsteinn Einarsson, Vésteinn Snæbjarnarson

Syntactic Sugar for Syntactic Queries: Sequential Representations for Dependency Queries
Niklas Deworetzki and Arianna Masciolini

Context Is (Almost) Everything: Llama-3 on Structured Output and AMR Parsing
Maja Buljan, Stephan Oepen, Lilja Øvrelid

Towards the Morphological Annotation of North Markian (Low German)
Christian Chiacos

Cross-Dataset Inconsistencies in Morphological Annotation: Evidence from Universal Dependencies
Vlasta Ohlidalová

Improving Latvian Morphosyntactic Parsing with Pretrained Encoders and Analyzer-Constrained Decoding
Arturs Znotins

CommonMorph: Participatory Morphological Documentation Platform
Aso Mahmudi, Sina Ahmadi, Kemal Kurniawan, Rico Sennrich, Eduard Hovy, Ekaterina Vylomova

Datasets for Verb Alternations across Languages: BLM Templates and Data Augmentation Strategies

Giuseppe Samo and Paola Merlo

A Large and Balanced Multi-Domain Arabic Corpus Annotated for Morphology, Syntax, and Readability

Khalid Elmadani, Adel Mahmoud Wizani, Hanada Taha Thomure, Nizar Habash

The DELPH-IN Grammar: A Curated Repository of Grammars and Treebanks

Francis Bond and Dan Flickinger

Morphemes without Borders: Evaluating Root–Pattern Morphology in Arabic Tokenizers and LLMs

Yara Alakeel, Chatrine Qwaider, Hanan Aldarmaki, Sawsan Alqahtani

Universal NER v2: Towards a Massively Multilingual Named Entity Recognition Benchmark

Terra Blevins, Stephen Mayhew, Marek Suppa, Hila Gonen, Shachar Mirkin, Vasile Pais, Kaja Dobrovoljc, Voula Giouli, Jun Kevin, Eugene Jang, Eungseo Kim, Jeongyeon Seo, Xenophon Gialis, Yuval Pinter

P10.4.1 • Lexicon and Semantics II

Poster Area 2

Chair: António Branco

APODICTUS: Automatic Processing of DICTIONary Update candidates

Felix Blessing, Johannes Sax, Julian Kaufmann, Wei Zhao, Nikolay Arefyev, Dominik Schlechtweg

A Test Collection for Part-of-Speech Tagging and Word Sense Disambiguation

Robert Krovetz

Creating a Hybrid Rule and Neural Network Based Semantic Tagger Using Silver Standard Data: The PyMUSAS Framework for Multilingual Semantic Annotation

Andrew Moore, Paul Rayson, Dawn Archer, Tim Czerniak, Dawn Knight, Daisy Lal, Gearóid Ó Donnchadha, Mícheál Ó Meachair, Scott Piao, Elaine Uí Dhonnchadha, Johanna Vuorinen, Yan Yabo, Xiaobin Yang

Scare Quotes as Markers of "Questionable" Word Usages and Misalignment in Conversation: An Annotation Study

Aina Garí Soler, Juan Carlos Zevallos Huaco, Matthieu Labeau, Chloé Clavel

Modeling Clinical Uncertainty in Radiology Reports: From Explicit Uncertainty Markers to Implicit Reasoning Pathways

Paloma Rabaey, Jong Hak Moon, Jung-Oh Lee, Min Gwan Kim, Hangyul Yoon, Thomas Demeester, Edward Choi

ArabDiscrim: A Decade-Long Arabic Facebook Corpus on Racism and Discrimination

Wajdi Zaghoulani, Shima Ibrahim, Mabrouka Bessghaier, Houda Bouamor

DAMETA: An LLM Benchmark for Danish Metaphor Interpretation with Systematically Varied Distractors

Nina Schneidermann, Sanni Nimb, Nathalie Norman, Sussi Olsen, Bolette Pedersen

A New Semantic Artifact Based Framework for Studying and Documenting Algospeak and Related Phenomena

Fahad Khan, Elisa Gugliotta, Elisa Squadrito, Maura Tarquini, Francesca Frontini

Creating a High Quality Abstract Meaning Representation Dataset Automatically

Johannes Heinecke, Asadullah Munshi, Frédéric Herledan, Geraldine Damnati

Towards a Comprehensive English Wordnet-Wikidata Mapping

John P. McCrae, Johann Bergh, Krasimir Angelov

AmDi - Ambiguous Words Diachronic Dataset

Felix Thielen and Kai Kugler

P10.4.2 • Lexicon and Semantics III

Poster Area 2

Chair: Gerard De Melo

GerVLP: A CEFR-Graded Vocabulary List of L2 Learners' Productive Vocabulary in German

Noah-Manuel Michael, Anna Huelsing, Andrea Horbach

Building Bridges between Student and Curricular Language: Creating a Corpus of Abstract Meaning Representations for the Classroom

Kristin Wright-Bettner, Zheng Cai, zekun zhao, James H. Martin, Jeffrey Flanigan, Martha Palmer

Mu'jam Arriyadh: A Comprehensive Lexicon for Contemporary Arabic Language

Afrak Altamimi, Abdulrahman Alosaimy, Halah Alharbi, Hawra Aljasim, Muneera Alhoshan, Amal Almazrui, Hanan Alharbi, Abdulrahman Alshehri, Bayan Almuqhim, Maryam Algarny, Yahya Asiri, Abdullah I. Alharbi, SALEH ALBALAWI, Fawziah Asiri, Sara Alhifthi, Abdullah Alfaifi

The Romanian Corpus Annotated with Multiword Expressions. PARSEME-Ro Version 2.0

Verginica Barbu Mititelu, Mihaela Cristescu, Elena Irimia, Carmen Vasile

Missing Links: LLM-Augmentation of Event Triggers of State Changes in the OpenPI Dataset

Kyeongmin Rim and James Pustejovsky

VUPMC: A New Political Metaphor Corpus in Mandarin Chinese

Xiaojuan Tan

Not All Disneys Are the Same: Making Coreference Metonymy-Aware

Bingyang Ye, Jingxuan Tu, James Pustejovsky

JSTS-Neg: Japanese Semantic Textual Similarity Dataset for Evaluating Negation Understanding Ability

Reiko Yuasa, Yoshihide Kato, Shigeki Matsubara

Few-shot Prompting or Supervised Tuning? A Comparative Study of LLMs for Linguistically Distant Language Pairs in BDI

Deepen Naorem, Sanasam Ranbir Singh, Telem Joyson Singh, Priyankoo Sarmah

When Structure Matters: Cross-Lingual Hyperbolic Embeddings for Chinese and English Wordnets

Mao-Chang Ku, Da-Chen Lian, Pin-Er Chen, Po-Ya Angela Wang, Wei-Ling Chen, Shu-Kai HSIEH

16:40 ***Coffee Break***
17:00

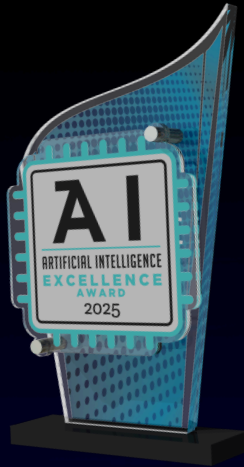
17:00 **LREC 2026 Closing Ceremony**
18:20 ***Auditorium Illes Balears***
Chair: Stelios Piperidis

LREC 2026 Gala Dinner

DATAFORCE

BY TRANSPERFECT

[HUMAN INSIGHTS FOR AI THAT IS RELIABLE. REFINED. RESPECTED]



Excellence in **Natural Language Processing**

[LLM OUTPUT EVALUATION] [SPEECH TRANSCRIPTION] [ANNOTATION WITH META DATA]
[ONSITE DATA COLLECTION THROUGH USER STUDIES] [ON DEMAND LINGUISTIC SUPPORT]

WEBSITE



LINKEDIN



VISIT US AT **BOOTH 3**

A platform for data agency and fair value exchange

We're building the data platform that starts with people – not extraction.

Upload your dataset

Share your datasets whilst retaining total ownership.

Browse datasets

Discover datasets shared directly by organisations and communities.





INTERESTED IN #transparentAIforEU?

COME TALK LLMs à la EU TO BOOTH #5



Co-funded by
the European Union



HiTZ

Hizkuntza Teknologiako
Zentroa
Basque Center
for Language Technology



EUROPEAN LANGUAGE DATA SPACE

The European Marketplace
for Language Data.

Share. Connect. Benefit.



<https://language-data-space.ec.europa.eu>

alt-educ

Alliance for Language Technologies
European Digital Infrastructure Consortium

Promoting excellence in language
technologies and preserving Europe's
linguistic diversity and cultural richness

Main activities



Federate and strengthen
the linguistic data
production



Contribute to the development
of evaluation, certification and
normalisation



Support the
development of
large language
models



Facilitate the
adaptation of
existing models to
user needs



Provide advice to the LT
ecosystem, including
the administration and
citizens

Get in touch



Website
www.alt-educ.eu



Email
contact@alt-educ.eu



Location
1 place Aristide Briand
02600 Villers-Cotterets
France

 translated.

\$100,000

to fund language technology
innovators who share the goal
of making it easier for everyone
to understand and be understood
by all others.

Find out more.



MULTIVERSE

COMPUTING

LREC 20

Palma 26

Organized by **ELRA Language Resources Association**

9, rue des Cordelières

75013 Paris France

www.elra.info

Mozilla
Data
Collective

Create. Curate. Control.

DATAFORCE
BY TRANSPERFECT



HiTZ
Hizkuntza Teknologiako
Zentroa
Basque Center
for Language Technology



alt-edic



Reverso



LINGUIST
List