

CLARIN-CH Day 2026 Program

Time	Program
9:30-10:00	Welcome Coffee
10:00-11:00	Keynote talk
11:00-11:30	Thematic session 1: 30 minutes (3x9min)
11:30-12:00	Thematic session 2: 30minutes (3x9min)
12:00-13:30	Lunch
13:30-14:00	Showcases & Demos: 30 minutes
14:00-14:30	Thematic session 3: 30 min (3x9min)
14:30-15:00	Thematic session 4: 30 minutes (3x9min)
15:00-15:30	Coffee break
15:30-16:00	Thematic session 5: 30 min (3x9min)
16:00-16:30	Poster session and networking (6 posters)
16:30-17:30	Round table about DMPs for language data and Conclusive remarks
17:30	End of the event

Thematic sessions

Session 1. AI and Language Technologies

This session explores the use of AI and computational methods for language research and processing.

- *Integrating Generative AI into Corpus-Assisted Analysis of Public Discourse* | Dolores Lemmenmeier (ZHAW) (abstract)
- *Whose dialect, whose variety? AI transcription as a hidden technology of language planning* | Francine González Borrell and Yvette Bürki (UNIBE)
- *Towards a Fine-Grained Persian Dataset for Sexualized and Gender-Based Hostile Language Detection* | Hanieh Habibi, Amir H. Payberah and Davide Picca (UNIL)

Session 2. Corpora and Discourse Analysis

This session presents corpus-driven analyses of contemporary public communication and online discourse.

- *Building a Corpus of Russian Military Telegram Blogs: Methodology, Challenges, and Ethical Considerations* | Olga Bikkulova (UNIBE)

- *A multi-level approach to annotating deniability in political and legal discourse* | Bruna Paz Schmid, Steve Oswald and Lou Odermatt (UNIFR)
- *Digital stancetaking towards gender diversity on the platform gutefrage.net* | Aline Siegenthaler (UNIFR)

Session 3. Open Research Data & FAIR

This session addresses Open Research Data, FAIR principles, research infrastructure adoption, and data management practices.

- *Metamorphosis of a DMP: Navigating Consent and De-identification Across Three Stages of an MA Project* | Oscar Jordan (UNIL)
- *Language Data Across Disciplines: Survey Insights on ORD Practices, FAIR Awareness, and Training Needs* | Julia Krasselt (ZHAW), Elsa Liste Lamas (ZHAW), Maike Fischer (ZHAW), Cristina Grisot (UZH) and Joanna Blochowiak (UZH)
- *From Infrastructure to Impact: Understanding Adoption in European Digital Research Infrastructures* | Beliz Gökmen, Andrew A. Clark, Gorka Fraga Gonzalez and Steven Moran (UZH)

Session 4. Language Resources and Infrastructures

This session focuses on the creation, documentation, and sustainability of language resources and research infrastructures.

- *Making the Ephemeral Durable: The Swiss Sticker Corpus (SSC) as Open Research Data* | Yvette Buerki, Kellie Gonçalves and Janis Schneider (UNIBE)
- *VotingBooklets: A Multilingual Parallel Corpus of Swiss Federal Referendum Texts* | Elina Stüssi and Jannis Vamvas (UZH)
- *Database-first lexical infrastructure for Proto-Albanian research: interoperability, FAIR data, and reusable workflows for Albanian–Romanian cognacy and the Proto-Albanian–Thracian substrate segment* | Edmond Cane (Luarasi University)

Session 5. Multilingualism & Linguistic Diversity

This session focuses on multilingual data, language variation, language contact, and language processing across languages.

- *The importance of multilingual and cross-linguistic corpora for the study of language contact* | Olivier Winistörfer (UZH), Maxim Makartsev (University of Oldenburg) and Anastasia Escher (ETH Zurich)
- *Linguistic data structures and standardization are not theory-neutral* | Sandra Auderset (UNIBE) and Tiago Tresoldi
- *At-Issue Sensitivity in L2 Misinformation Detection: The Role of Cognitive Load* | Alan Lombardini (UNIFR), Giulia Giunta (UNINE), Diana Mazzarella (UNINE), Didier Maillat (UNIFR)

Showcases & Demos

- *AViRI – Towards a Shared Infrastructure for Audiovisual Data and Metadata* | Josephine Diecke, Simon Spiegel and Teodora Vukovic (UZH)
- *Audio-video data processing pipelines* | Masoumeth Chapariniya and Teodora Vukovic (UZH)
- *MAVA and DAVA: building a cluster of infrastructure for processing multimodal audio-visual data* | Teodora Vukovic (UZH)

Posters

- *RefCo2 from the fields to the archive: multi-perspective testing of a corpus-quality platform for reusable language data* | Jocelyn Aznar (UNIBE), Miranda Dickerman (UZH) and Jonathan Reich (UNIBE)
- *The “Person Marking in South-Central Trans-Himalayan” (PMST) database* | Linda Konnerth, Sandra Auderset, Jonathan Reich, and Susie Kanshouwa (UNIBE)
- *Infrastructural support for data analysis and reuse in Interactional Linguistics: potential of LCP-Videoscope* | Johanna Miecznikowski (USI), Elena Battaglia (University of Bologna), Jérôme Jacquin (UNIL), Thomas Schmidt (University of Duisburg-Essen / linguisticbits), Teodora Vuković (UZH) and Jeremy Zehr (UZH)
- *Beyond Clinical Pain Descriptors: Detecting Figurative Pain Descriptions in Patient Narratives* | Dolores Lemmenmeier (ZHAW), Nikolai Shurakov (UZH), Yvonne Ilg (UZH), Lucien Baumgartner (UZH) and Sabina Maria Rätz (USZ)
- *FAIRifying a Multilingual European Parliament Corpus for the Study of Gender-Inclusive Lexical Innovation* | Nikos Tsourakis, Aurélie Picton, Julie Humbert-Droz, Charlotte Jeanneret and Anne Kosakevitch (UNIGE)
- *CLARIN – a pan-european distributed digital infrastructure for language data and technology* | Cristina Grisot (UZH)
- *Two communities, one shared practice: presentation of CLARIN-CH RDM Working Group and SRDSN L-RDM Node* | Cristina Grisot and Joanna Blochowiak (UZH)
- *LiRI and the Lia Rumantscha work together to build new infrastructure and collect data to promote Romansh* | Ignacio Pérez Prat (Lia Rumantscha), Jannis Vamvas, Nikolina Rajović and Igor Mustač (UZH)
- *LiRI and Linguist List* | Steven Moran and Valeriia Vyshnevetska (UZH)
- *Swissxox RAG: Retrieval-First AI for Large Multilingual News Archives* | Igor Mustač and Nikolina Rajović (UZH)