LIMO 2023

The 1st Workshop on Linguistic Insights from and for Multimodal Language Processing

Proceedings of the Workshop

September 22, 2023

Order copies of this and other ACL proceedings from:

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

ISBN XXX-X-XXXXXX-XX-X

Foreword

Processing multimodal information (like visual representations of the environment, auditory cues, images, gestures, gaze etc.) and integrating them is a constant and effortless process in human language processing. Recent progress in the area of language & vision, large-scale visually grounded language models, and multimodal learning (e. g. CLIP (Radford et al., 2021), VILBERT (Lu et al., 2019) etc.) have led to breakthroughs in challenging multimodal NLP applications like image-text retrieval, image captioning (Cornia et al., 2020) or visual question answering (Antol et al., 2015). Yet, modeling the semantics and pragmatics of situated language understanding and generation and, generally, language processing beyond the linguistic context, i. e. in combination with multiple other modalities, is still one of the biggest challenges in NLP and Computational Linguistics (Bisk et al., 2020).

Recent efforts in understanding complex multimodal phenomena in language and dialogue have explored a variety of aspects of multimodality and produced a substantial amount of valuable multimodal datasets and models that include various types of text (from short and informal social media comments to more formal news, instructions/manuals and legal documents, they are also usually accompanied by an image, meme, animation or video) and dialogue (from reference games, instruction dialogues to fully situated interaction with agents and robots). The variety in this wide problem space and the downstream tasks also require variety in the approaches to tackle them. As a result, Multimodal Language Processing is approached by many different sub-areas of Computational Linguistics and NLP—in computational semantics and pragmatics, dialogue modeling, language modeling, and grounding, multimodal and crossmodal learning, and beyond, including physical or robotic actions.

While there have been recent venues and workshops targeting multimodal representation learning and large-scale Language and Vision models, there is a lack of discussion in the community that focuses on linguistic multimodal phenomena, domain- and task-specific analyses of multimodality and, generally, contributions of computational linguistics to multimodal learning and vice versa (Parcalabescu et al., 2022). With this workshop, we aim to bring together researchers who work on various linguistic aspects of multimodal language processing to discuss and share the recent advances in this interdisciplinary field.

The main goals of this workshop are to

- Discuss various tasks, phenomena, models, and problems in multimodal language processing
- Discuss how insights from (computational) linguistics can inform multimodal learning and modeling
- Facilitate networking and encourage collaboration between researchers working on different aspects of multimodality in computational linguistics and language processing

The LIMO 2023 workshop organizers:
Piush Aggarwal, Özge Alaçam, Carina Silberer, Sina Zarrieß and Torsten Zesch

Organizing Committee

Piush Aggarwal (FernUniversität in Hagen)

Özge Alaçam (Universität Bielefeld)

Carina Silberer (Universität Stuttgart)

Sina Zarrieß (Universität Bielefeld)

Torsten Zesch (FernUniversität in Hagen)

Program Committee

Albert Gatt (Utrecht University, Netherlands)

Animesh Mukherjee (IIT-Kharagpur, India)

Asif Ekbal (IIT-Patna, India)

Barbara Plank (LMU Munich, Germany)

Corentin Kervadec (Universitat Pompeu Fabra, Spain)

David Schlangen (University of Potsdam, Germany)

Desmond Elliott (University of Copenhagen, Denmark)

Hsiu-Yu Yang (Institute for Computational Linguistics, Stuttgart University, Germany)

Jana Götze (University of Potsdam, Germany)

Letitia Parcalabescu (Heidelberg University, Germany)

Nikolai Ilinykh (University of Gothenburg, Sweden)

Sabine Schulte im Walde (Universität Stuttgart, Germany)

Sandro Pezzelle (ILLC, University of Amsterdam, Netherlands)

Seid Muhie Yimam (University of Hamburg, Germany)

Sherzod Hakimovn (University of Potsdam, Germany)

Simon Dobnik (University of Gothenburg, Sweden)

Timo Baumann (OTH Regensburg, Germany)

Invited Speakers

Letitia Parcalabescu (Heidelberg University, Germany)

Sandro Pezzelle (ILLC, University of Amsterdam, Netherlands)

Table of Contents

Conference Program

Friday, September 22, 2023

9:00–9:10 ***Kick-off for LIMO workshop***

9:10–9:30 **2 Min. Teaser for Accepted Papers**

A Pipeline for the Creation of Multimodal Corpora from YouTube Videos
Nathan Dykes, Anna Wilson and Peter Uhrig

Multi-Modal Learning Application – Support Language Learners with NLP Techniques and Eye-Tracking
Robert Geislinger, Ali Ebrahimi Pourasad, Deniz Gül, Daniel Djahangir, Seid Muhie Yimam, Steffen Remus and Chris Biemann

Context matters: evaluation of target and context features on variation of object naming
Nikolai Ilinykh and Simon Dobnik

The Scenario Refiner: Grounding subjects in images at the morphological level
Claudia C. Tagliaferri, Denis Paperno, Albert Gatt and Sofia Axioti

FlowchartQA: The First Large-Scale Benchmark for Reasoning over Flowcharts
Simon Tannert, Marcelo G. Feighelstein, Jasmina Bogojeska, Joseph Shtok, Assaf Arbelle, Peter W. J. Staar, Anika Schumann, Jonas Kuhn and Leonid Karlinsky

Presenting an Annotation Pipeline for Fine-grained Linguistic Analyses of Multimodal Corpora
Elena Volkanovska, Sherry Tan, Changxu Duan, Debajyoti Chowdhury and Sabine Bartsch

9:30–10:30 *Invited Talk by Dr. Letitia Parcalabescu*

10:30–10:50 ***Coffee Break***

Friday, September 22, 2023 (continued)

10:50–12:00 Poster Session

12:00–13:00 *Invited Talk by Dr. Sandro Pezzelle*