IWCS 2021

Multimodal Semantic Representations

Proceedings of the First Workshop

June 16, 2021

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 978-1-954085-21-3

Preface

The demand for more sophisticated natural human-computer and human-robot interactions is rapidly increasing as users become more accustomed to conversation-like interactions with AI and NLP systems. Such interactions require not only the robust recognition and generation of expressions through multiple modalities (language, gesture, vision, action, etc.), but also the encoding of situated meaning.

When communications become multimodal, each modality in operation provides an orthogonal angle through which to probe the computational model of the other modalities, including the behaviors and communicative capabilities afforded by each. Multimodal interactions thus require a unified framework and control language through which systems interpret inputs and behaviors and generate informative outputs. This is vital for intelligent and often embodied systems to understand the situation and context that they inhabit, whether in the real world or in a mixed-reality environment shared with humans.

This workshop intends to bring together researchers who aim to capture elements of multimodal interaction such as language, gesture, gaze, and facial expression with formal semantic representations. We provide a space for both theoretical and practical discussion of how linguistic co-modalities support, inform, and align with “meaning” found in the linguistic signal alone. In so doing, the MMSR workshop has several goals:

1. To provide an opportunity for computational semanticists to critically examine existing NLP semantic frameworks for their validity to express multimodal elements;

2. To explore and identify challenges in the semantic representation of co-modalities cross-linguistically and cross-culturally;

3. To gain understanding of domains and tasks where certain semantic frameworks (multimodal or not) are most effective and why.

We would like to thank the authors, reviewers, invited speakers, and IWCS 2021 organizers for making this workshop possible. We look forward to an exciting workshop.

Lucia Donatelli, Nikhil Krishnaswamy, Kenneth Lai, and James Pustejovsky

Organizers:

Lucia Donatelli, Saarland University
Nikhil Krishnaswamy, Colorado State University
Kenneth Lai, Brandeis University
James Pustejovsky, Brandeis University

Program Committee:

Nicholas Asher, Institute de Recherche en Informatique de Toulouse
Claire Bonial, Army Research Lab
Harry Bunt, Tilburg University
Stergios Chatzikyriakidis, University of Gothenburg
Sandy Ciroux, University of Konstanz
Robin Cooper, University of Gothenburg
Simon Dobnik, University of Gothenburg
Maria (Masha) Esipova, University of Oslo
Anette Frank, Heidelberg University
Felix Gervits, Army Research Lab
Jonathan Ginzburg, Université de Paris
Casey Kennington, Boise State University
Stefan Kopp, Bielefeld University
Staffan Larsson, University of Gothenburg
Andy Lücking, Université de Paris, Goethe Universty Frankfurt
Larry Moss, Indiana University
Francisco Ortega, Colorado State University
Gözde Gül Şahin, Technical University of Darmstadt
Philippe Schlenker, Institut Jean-Nicod - Ecole Normale Supérieure, Paris
Nathan Schneider, Georgetown University
Candy Sidner, Worcester Polytechnic Institute
Jurģis Šķilters, University of Latvia
Benjamin Spector, Institut Jean-Nicod - Ecole Normale Supérieure, Paris
David Traum, University of Southern California
Alexis Wellwood, University of Southern California
Bram Willemsen, KTH Royal Institute of Technology

Invited Speakers:

Chiara Bonsignori, Consiglio Nazionale delle Ricerche
Matthias Scheutz, Tufts University
Virginia Volterra, Consiglio Nazionale delle Ricerche

Table of Contents

Workshop Program

Wednesday, June 16, 2021

16:00–16:15 ***Introduction***

16:15–17:00 *Invited Talk: From action to language through gesture*
Virginia Volterra and Chiara Bonsignori

17:05–17:35 **Oral Session 1**

What is Multimodality?
Letitia Parcalabescu, Nils Trost and Anette Frank

Are Gestures Worth a Thousand Words? An Analysis of Interviews in the Political Domain
Daniela Trotta and Sara Tonelli

Requesting clarifications with speech and gestures
Jonathan Ginzburg and Andy Luecking

17:40–18:25 *Invited Talk: Attention, Incrementality, and Meaning: On the Interplay between Language and Vision in Reference Resolution*
Matthias Scheutz

18:30–19:10 **Oral Session 2**

Seeing past words: Testing the cross-modal capabilities of pretrained V&L models on counting tasks
Letitia Parcalabescu, Albert Gatt, Anette Frank and Iacer Calixto

How Vision Affects Language: Comparing Masked Self-Attention in Uni-Modal and Multi-Modal Transformer
Nikolai Ilinykh and Simon Dobnik

EMISSOR: A platform for capturing multimodal interactions as Episodic Memories and Interpretations with Situated Scenario-based Ontological References
Selene Baez Santamaria, Thomas Baier, Taewoon Kim, Lea Krause, Jaap Kruijt and Piek Vossen

Annotating anaphoric phenomena in situated dialogue
Sharid Loáiciga, Simon Dobnik and David Schlangen

Wednesday, June 16, 2021 (continued)

19:15–19:45 Poster Session

Incremental Unit Networks for Multimodal, Fine-grained Information State Representation
Casey Kennington and David Schlangen

Teaching Arm and Head Gestures to a Humanoid Robot through Interactive Demonstration and Spoken Instruction
Michael Brady and Han Du

Building a Video-and-Language Dataset with Human Actions for Multimodal Logical Inference
Riko Suzuki, Hitomi Yanaka, Koji Mineshima and Daisuke Bekki

19:45–20:00 ***Closing***