

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