COLING 2020

# Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN)

## Proceedings of the Second Workshop

December 13, 2020
Barcelona, Spain (Online)

IDeaL SFB 1102 HUGGING FACE

ISBN 978-1-952148-51-4

# Introduction

Welcome to LANTERN 2020, The Second Workshop Beyond Vision and Language: Integrating Real-World Knowledge, co-located with COLING 2020. Building on the success of the first edition co-located with EMNLP-IJCNLP 2019, the second edition aims to bring together and interconnect researchers focusing on language from a multimodal perspective. In particular, the main goal of the workshop is to promote and foster research which uses machine- and deep-learning techniques to interconnect language with vision and other modalities by leveraging some external knowledge. By encouraging contributions exploiting very diverse sources of external knowledge (knowledge graphs, fixed and dynamic environments, cognitive and neuroscience data, etc.), the workshop is open to all research directions which acknowledge the importance of knowledge in acquiring, using, and evaluating language in real-world settings.

In this second edition, we called for both long and short papers. All the accepted contributions are published in these Proceedings. Moreover, we host presentations from papers accepted to appear in the novel "Findings of EMNLP 2020" series. These papers are not published in the LANTERN Proceedings.

LANTERN 2020 received 7 submissions, all of which were double-blindly reviewed by three highly-qualified reviewers. In total, 4 papers (1 long, 3 short) were accepted to appear in the Proceedings of the workshop, with an acceptance rate of around 57% (comparable to that of the first edition, which was around 53%).

Contributions are representative of a broad range of current problems and approaches and include a novel technique to improve text-to-image synthesis by leveraging visual question answering, an approach to learn commonsense from image captions descriptions, an evaluation of the language used by computational models to characterize people in images, a method that leverages eye-tracking information for the task of dependency parsing. Such richness of approaches and perspectives is in line with the purpose of the workshop, and confirms the growing interest for problems going beyond the task-specific integration of language and vision.

The program of the workshop, besides 4 oral presentations and a session where papers published in the "Findings of EMNLP 2020" are presented, includes invited talks by Yonatan Bisk, Gemma Boleda, Angeliki Lazaridou, and Stefan Lee. The workshop received sponsorship by iDeaL SFB 1102. Best paper award is sponsored by HuggingFace.

The LANTERN Workshop Organizers

**Organizers:**

Aditya Mogadala, Saarland University (Germany)
Sandro Pezzelle, University of Amsterdam (The Netherlands)
Dietrich Klakow, Saarland University (Germany)
Marie-Francine Moens, Katholieke Universiteit Leuven (Belgium)
Zeynep Akata, University of Tübingen (Germany)

**Program Committee:**

Afra Alishahi, Tilburg University
Alane Suhr, Cornell University
Albert Gatt, University of Malta
Alessandro Suglia, Heriot-Watt University
Anand Mishra, IIT Jodhpur
Ashutosh Modi, IIT Kanpur
Carina Silberer, University of Stuttgart
Claudio Greco, University of Trento
David Schlangen, University of Potsdam
Douwe Kiela, Facebook AI Research
Florian Metze, Carnegie Mellon University
Iacer Calixto, UvA/NYU
Ionut-Teodor Sorodoc, Universitat Pompeu Fabra
Jacob Goldberger, Bar Ilan
Parisa Kordjamshidi, Michigan State University
Ravi Shekhar, Queen Mary University London
Sina Zarrieß, University of Jena
Somak Aditya, Microsoft Research India
Spandana Gella, Amazon AI

**Invited Speaker:**

Angeliki Lazaridou, DeepMind
Gemma Boleda, Universitat Pompeu Fabra & ICREA
Stefan Lee, Oregon State University
Yonatan Bisk, Carnegie Mellon University

# Table of Contents

# Conference Program

LANTERN is a virtual event with four oral for the accepted long or short papers and also presentations from “Findings of EMNLP 2020”. It also includes four invited presentations. The full schedule is presented below and all times are listed in CET.

**14:00 - 14:15** Welcome and Opening Remarks

**14:15 - 14:55** Invited Talk: Angeliki Lazaridou (DeepMind)

**14:55 - 15:35** Invited Talk: Gemma Boleda (Universitat Pompeu Fabra & ICREA)

**15:35 - 16:15** Accepted Papers (pre-recorded + live Q & A)

1. Eyes on the Parse: Using Gaze Features in Syntactic Parsing

2. Leveraging Visual Question Answering to Improve Text-to-Image Synthesis

3. Seeing the world through text: Evaluating image descriptions for commonsense reasoning in machine reading comprehension

4. How do image description systems describe people? A targeted assessment of system competence in the PEOPLE domain

**16:15 - 16:40** Coffee Break

**16:40 - 17:20** Invited Talk: Yonatan Bisk (Carnegie Mellon University)

**17:20 - 18:00** Invited Talk: Stefan Lee (Oregon State University)

**18:00 - 18:40** Findings of EMNLP 2020 (live talk + Q & A)

**18:40 - 19:00** Best Talk/Poster Award and Concluding Remarks