CMCL 2022

Workshop on Cognitive Modeling and Computational Linguistics

Proceedings of the Workshop

May 26, 2022

The CMCL organizers gratefully acknowledge the support from the following sponsors.

In cooperation with Laboratoire Parole et Langage (France) and Japan Society for the Promotion of Science (Japan)

©2022 Association for Computational Linguistics

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 978-1-955917-29-2

Introduction

Welcome to the 12th edition of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL)!!

CMCL is traditionally the workshop of reference for research at the intersection between Computational Linguistics and Cognitive Science. This year, for the first time CMCL will be held in hybrid mode: virtual attendance will still be allowed, given the persistence of the COVID-19 pandemic, while the in-person meeting will take place in the beautiful Dublin.

This year, we received 20 regular workshop submissions and we accepted 10 of them, for a global 50% acceptance rate. We also received two extended abstracts as non-archival submissions, and both of them will be presented during the poster session. As in previous years, submissions have been highly varied across the cognitive sciences, with topics ranging from the relationship between vision and human linguistic-semantic knowledge, the relationship between eye gaze and self-attention in Transformer language models, and an account of the game Codenames. Work ranges from deep neural network approaches to Bayesian cognitive models, learning of phonetic and phonological categories, analyses of neurolinguistic data, and much more. We are thrilled to continue a workshop with the breadth and depth that is emblematic of the fields of cognitive science and natural language processing.

Last year, we held a shared task on eye-tracking prediction in a variety of measures. This year, we led an additional shared task that built on the success of the previous edition. In the second edition of the shared task on eye-tracking data prediction, this time we included multilingual data from English, Russian, German, Hindi, Chinese, Dutch and Danish, enabling research teams to try a variety of methods and language models far beyond prior eye tracking tasks. A total of six teams participated, of which 5 submitted papers describing their systems.

As always, we are extremely grateful to the PC members, without whose efforts we would be unable to ensure high-quality reviews and high-quality work for presentation at the workshop. We are indebted to their generosity and are proud of the community that supports CMCL. We also thank our invited speakers, Andrea E Martin and Vera Demberg for kindly accepting our invitation.

Finally, we thank our sponsors: the Japanese Society for the Promotion of Sciences and the Laboratoire Parole et Langage. Through their generous support, we are able to offer fee waivers to PhD students who were first authors of accepted papers, and to offset the participation costs of the invited speakers.

The CMCL 2022 Organizing Committee

Organizing Committee

Workshop Organizers

Emmanuele Chersoni, The Hong Kong Polytechnic University, China
Nora Hollenstein, University of Copenhagen, Denmark
Cassandra L Jacobs, University of Buffalo, USA
Yohei Oseki, University of Tokyo, Japan
Laurent Prévot, Aix-Marseille University, France
Enrico Santus, Bayer, USA

Program Committee

Program Committee

Laura Aina, Pompeu Fabre University of Barcelona
Raquel Garrido Alhama, Google
Philippe Blache, Aix-Marseille University
Christos Christodoulopoulos, Amazon
Aniello De Santo, University of Utah
Vesna G Djokic, University of Amsterdam
Micha Elsner, Ohio State University
Raquel Fernández, University of Amsterdam
Abdellah Fourtassi, Aix-Marseille University
Michael Frank, Stanford University
Robert Frank, Yale University
Diego Frassinelli, University of Konstanz
John Hale, University of Georgia
Yu-Yin Hsu, The Hong Kong Polytechnic University
Tim Hunter, UCLA
Samar Husain, IIT Delhi
Anna A Ivanova, MIT
Carina Kauf, MIT
Jordan Kodner, Stony Brook University
Gianluca Lebani, University Ca' Foscari of Venice
Fred Mailhot, Dialpad
Karl David Neergaard, University of Macau
Stephen Politzer-Ahles, The Hong Kong Polytechnic University
Giulia Rambelli, University of Pisa
Roi Reichart, Technion – Israel Institute of Technology
Rachel A Ryskin, University of California Merced
Lavinia Salicchi, The Hong Kong Polytechnic University
William Schuler, Ohio State University
Cory Shain, MIT
Ece Takmaz, University of Amsterdam
Lonneke van der Plas, Idiap Research Institute
Yao Yao, The Hong Kong Polytechnic University

Shared Task Program Committee

Sunit Bhattacharya, Charles University of Prague
Stephanie Brandl, University of Copenaghen
Patrick Haller, University of Zurich
Joseph Marvin Imperial, National University of the Philippines
Mitja Nikolaus, Aix-Marseille University
Lavinia Salicchi, The Hong Kong Polytechnic University
Ece Takmaz, University of Amsterdam
Zining Zhu, University of Toronto

Invited Speakers

Vera Demberg, Saarland University, Germany
Andrea E Martin, Max Planck Institute for Psycholinguistics, Netherlands

Table of Contents

Seeing the advantage: visually grounding word embeddings to better capture human semantic knowledge
Danny Merkx, Stefan Frank and Mirjam Ernestus 1

A Neural Model for Compositional Word Embeddings and Sentence Processing
Shalom Lappin and Jean-Philippe Bernardy 12

Visually Grounded Interpretation of Noun-Noun Compounds in English
Inga Lang, Lonneke Van Der Plas, Malvina Nissim and Albert Gatt 23

Less Descriptive yet Discriminative: Quantifying the Properties of Multimodal Referring Utterances via CLIP
Ece Takmaz, Sandro Pezzelle and Raquel Fernández 36

Codenames as a Game of Co-occurrence Counting
Réka Cserháti, Istvan Kollath, András Kicsi and Gábor Berend 43

Estimating word co-occurrence probabilities from pretrained static embeddings using a log-bilinear model
Richard Futrell 54

Modeling the Relationship between Input Distributions and Learning Trajectories with the Tolerance Principle
Jordan Kodner 61

Predicting scalar diversity with context-driven uncertainty over alternatives
Jennifer Hu, Roger P. Levy and Sebastian Schuster 68

Eye Gaze and Self-attention: How Humans and Transformers Attend Words in Sentences
Joshua Bensemann, Alex Yuxuan Peng, Diana Benavides Prado, Yang Chen, Neset Tan, Paul Michael Corballis, Patricia Riddle and Michael Witbrock 75

About Time: Do Transformers Learn Temporal Verbal Aspect?
Eleni Metheniti, Tim Van De Cruys and Nabil Hathout 88

Poirot at CMCL 2022 Shared Task: Zero Shot Crosslingual Eye-Tracking Data Prediction using Multilingual Transformer Models
Harshvardhan Srivastava 102

NU HLT at CMCL 2022 Shared Task: Multilingual and Crosslingual Prediction of Human Reading Behavior in Universal Language Space
Joseph Marvin Imperial 108

HkAmsters at CMCL 2022 Shared Task: Predicting Eye-Tracking Data from a Gradient Boosting Framework with Linguistic Features
Lavinia Salicchi, Rong Xiang and Yu-Yin Hsu 114

CMCL 2022 Shared Task on Multilingual and Crosslingual Prediction of Human Reading Behavior
Nora Hollenstein, Emmanuele Chersoni, Cassandra L Jacobs, Yohei Oseki, Laurent Prévot and Enrico Santus 121

Team ÚFAL at CMCL 2022 Shared Task: Figuring out the correct recipe for predicting Eye-Tracking features using Pretrained Language Models
Sunit Bhattacharya, Rishu Kumar and Ondrej Bojar 130

Team DMG at CMCL 2022 Shared Task: Transformer Adapters for the Multi- and Cross-Lingual Prediction of Human Reading Behavior
Ece Takmaz 136

Program

Thursday, May 26, 2022

09:30 - 09:45 *Opening Remarks*

09:45 - 10:45 *Keynote Talk by Andrea E. Martin*

10:45 - 11:00 *Coffee Break*

11:00 - 12:30 *Session 1 (Oral Presentations)*

Eye Gaze and Self-attention: How Humans and Transformers Attend Words in Sentences
Joshua Bensemann, Alex Yuxuan Peng, Diana Benavides Prado, Yang Chen, Neset Tan, Paul Michael Corballis, Patricia Riddle and Michael Witbrock

Seeing the advantage: visually grounding word embeddings to better capture human semantic knowledge
Danny Merkx, Stefan Frank and Mirjam Ernestus

Visually Grounded Interpretation of Noun-Noun Compounds in English
Inga Lang, Lonneke Van Der Plas, Malvina Nissim and Albert Gatt

12:30 - 13:30 *Lunch Break*

13:30 - 15:00 *Session 2 (Oral Presentations)*

A Neural Model for Compositional Word Embeddings and Sentence Processing
Shalom Lappin and Jean-Philippe Bernardy

Codenames as a Game of Co-occurrence Counting
Réka Cserháti, Istvan Kollath, András Kicsi and Gábor Berend

About Time: Do Transformers Learn Temporal Verbal Aspect?
Eleni Metheniti, Tim Van De Cruys and Nabil Hathout

15:00 - 15:15 *Coffee Break*

15:15 - 15:30 *Shared Task Presentation*

CMCL 2022 Shared Task on Multilingual and Crosslingual Prediction of Human Reading Behavior
Nora Hollenstein, Emmanuele Chersoni, Cassandra L Jacobs, Yohei Oseki, Laurent Prévot and Enrico Santus

Thursday, May 26, 2022 (continued)

15:30 - 17:00 *Poster Session*

Estimating word co-occurrence probabilities from pretrained static embeddings using a log-bilinear model
Richard Futrell

Predicting scalar diversity with context-driven uncertainty over alternatives
Jennifer Hu, Roger P. Levy and Sebastian Schuster

Less Descriptive yet Discriminative: Quantifying the Properties of Multimodal Referring Utterances via CLIP
Ece Takmaz, Sandro Pezzelle and Raquel Fernández

Modeling the Relationship between Input Distributions and Learning Trajectories with the Tolerance Principle
Jordan Kodner

NU HLT at CMCL 2022 Shared Task: Multilingual and Crosslingual Prediction of Human Reading Behavior in Universal Language Space
Joseph Marvin Imperial

Team DMG at CMCL 2022 Shared Task: Transformer Adapters for the Multi- and Cross-Lingual Prediction of Human Reading Behavior
Ece Takmaz

Team ÚFAL at CMCL 2022 Shared Task: Figuring out the correct recipe for predicting Eye-Tracking features using Pretrained Language Models
Sunit Bhattacharya, Rishu Kumar and Ondrej Bojar

HkAmsters at CMCL 2022 Shared Task: Predicting Eye-Tracking Data from a Gradient Boosting Framework with Linguistic Features
Lavinia Salicchi, Rong Xiang and Yu-Yin Hsu

Poirot at CMCL 2022 Shared Task: Zero Shot Crosslingual Eye-Tracking Data Prediction using Multilingual Transformer Models
Harshvardhan Srivastava

17:00 - 18:00 *Keynote Talk by Vera Demberg*

18:00 - 18:15 *Closing Remarks*