CMCL 2021

# The Workshop on Cognitive Modeling and Computational Linguistics

## Proceedings of the Workshop

June 10, 2021
Online Event





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 978-1-954085-35-0

# Introduction

Welcome to the Workshop on Cognitive Modeling and Computational Linguistics (CMCL)!!

We reached the 11th edition of CMCL, the workshop of reference for the research at the intersection between Computational Linguistics and Cognitive Science. This is the 2nd edition in a row that will be held entirely online because of the COVID-19 pandemic. Although we won't have the possibility of meeting in person in charming Mexico City, the program of CMCL 2021 is one of the richest and most interesting in the recent history of the workshop. We received 26 regular paper submissions and 17 were accepted for publication, for a total acceptance rate of 65.3%. We also received 4 non-archival submissions (extended abstracts or cross-submissions), 2 of which were accepted for presentation.

This year's accepted papers spanned a highly diverse range of questions centering on language, cognition, and computation. Several papers unified computational methods with neurobehavioral data, including EEG, MEG, and fMRI. Many of the papers leveraged state-of-the-art, transformer-based language models to distinguish between two competing theories of sentence processing. Still others probed the differences between language comprehension and language production, and whether it is feasible to treat them similarly for the purposes of explaining language use. Outside of sentence processing, accepted papers also probed the relationship between language and emotion; the graph structure of phonology; and lexical comprehension. Accepted papers spanned several grammatical formalisms, including Combinatory Categorial Grammar, Construction Grammar, and dependency grammars, in addition to statistical approaches. These diverse perspectives on cognition modeling and computational linguistics promote our scientific community's continued growth.

Additionally, as a novelty of this year's edition, we have organized a shared task on eye-tracking data prediction for English, and we accepted 10 system description papers. The ability to accurately model gaze features is vital to advance our understanding of language processing. Therefore, we posed the challenge of predicting token-level eye-tracking metrics recorded during natural reading. The participating teams submitted predictions generated mainly with two approaches: (1) Tree-based boosting algorithms with extensive feature engineering and (2) neural networks trained for regression such as fine-tuning transformer-based language models. The features for training the systems included surface features, lexical and syntactic features, token probability features, and text complexity metrics, as well as representations from state-of-the-art language models, such as BERT, RoBERTa, and XLNet. The winning team presented a linguistic feature-based approach.

Also for this year, the contribution of our PC members in thoroughly reviewing and selecting the best papers has been invaluable. Here we wish to deeply thank all of them for their time and effort.

We also thank Afra Alishahi and Zoya Bylinskii, our keynote speakers, for having accepted our invitation.

Finally, thanks again to our sponsors: the Japanese Society for the Promotion of Sciences and the Laboratoire Parole et Langage. Through their generous support, we have been 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 2021 Organizing Committee

## Organizing Committee

Emmanuele Chersoni, The Hong Kong Polytechnic University
Nora Hollenstein, University of Copenaghen
Cassandra Jacobs, University of Wisconsin
Yohei Oseki, University of Tokyo
Laurent Prévot, Aix-Marseille University
Enrico Santus, Bayer

## Program Committee

Laura Aina, Pompeu Fabre University of Barcelona
Raquel Garrido Alhama, Tilburg University
Louise Gillian Bautista, University of the Philippines
Klinton Bicknell, Duolingo
Philippe Blache, Aix-Marseille University
Lucia Busso, Aston University
Christos Christodoulopoulos, Amazon
Aniello De Santo. University of Utah
Vesna Djokic, University of Amsterdam
Micha Elsner, Ohio State University
Raquel Fernández, University of Amsterdam
Thomas François, Catholic University of Louvain
Robert Frank, Yale University
Stefan Frank, Radboud University of Nijmegen
Stella Frank, University of Trento
Diego Frassinelli, University of Kostanz
Abdellah Fourtassi, Aix-Marseille University
John Hale, University of Georgia
Yu-Yin Hsu, The Hong Kong Polytechnic University
Tim Hunter, UCLA
Samar Husain, IIT Delhi
Jordan Kodner, Stony Brook University
Gianluca Lebani, University Ca' Foscari Venezia
Alessandro Lenci, University of Pisa
Ping Li, The Hong Kong Polytechnic University
Fred Mailhot, DialPad
Mohammad Momenian, The Hong Kong Polytechnic University
Karl Neergaard, University of Macau
Ludovica Pannitto, University of Trento
Bo Peng, Yunnan University
Sandro Pezzelle, University of Amsterdam
Stephen Politzer-Ahles, The Hong Kong Polytechnic University
Vito Pirrelli, ILC-CNR Pisa
Jakob Prange, Georgetown University
Carlos Ramisch, Aix-Marseille 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
Marco Senaldi, McGill University
Friederike Seyfried, The Hong Kong Polytechnic University
William Schuler, Ohio State University
Cory Shain, Ohio State University
Lonneke Van Der Plas, University of Malta
Yao Yao, The Hong Kong Polytechnic University

# Table of Contents

# Conference Program

**June 10, 2021, Mexico City (GMT-5)**

**9:00–9:15 Introduction**

**9:15–10:15 Keynote Talk 1**

9:15–10:15 *Grounded Language Learning, from Sounds and Images to Meaning*
Afra Alishahi

**10:15–10:30 Break**

**10:30–12:00 Oral Presentations 1**

*Non-Complementarity of Information in Word-Embedding and Brain Representations in Distinguishing between Concrete and Abstract Words*
Kalyan Ramakrishnan and Fatma Deniz

*Human Sentence Processing: Recurrence or Attention?*
Danny Merkx and Stefan L. Frank

*Modeling Incremental Language Comprehension in the Brain with Combinatory Categorial Grammar*
Miloš Stanojević, Shohini Bhattasali, Donald Dunagan, Luca Campanelli, Mark Steedman, Jonathan Brennan and John Hale

**June 10, 2021, Mexico City (GMT-5) (continued)**

**12:00–13:00 Lunch break**

**13:00–14:30 Oral Presentations 2**

*A Multinomial Processing Tree Model of RC Attachment*
Pavel Logacev and Noyan Dokudan

*That Looks Hard: Characterizing Linguistic Complexity in Humans and Language Models*
Gabriele Sarti, Dominique Brunato and Felice Dell'Orletta

*Accounting for Agreement Phenomena in Sentence Comprehension with Transformer Language Models: Effects of Similarity-based Interference on Surprisal and Attention*
Soo Hyun Ryu and Richard Lewis

**14:30–14:45 Break**

**14:45–15:00 Shared Task Presentation**

*CMCL 2021 Shared Task on Eye-Tracking Prediction*
Nora Hollenstein, Emmanuele Chersoni, Cassandra L. Jacobs, Yohei Oseki, Laurent Prévot and Enrico Santus

**15:00–16:30 Poster Session**

*LangResearchLab_NC at CMCL2021 Shared Task: Predicting Gaze Behaviour Using Linguistic Features and Tree Regressors*
Raksha Agarwal and Niladri Chatterjee

*TorontoCL at CMCL 2021 Shared Task: RoBERTa with Multi-Stage Fine-Tuning for Eye-Tracking Prediction*
Bai Li and Frank Rudzicz

*LAST at CMCL 2021 Shared Task: Predicting Gaze Data During Reading with a Gradient Boosting Decision Tree Approach*
Yves Bestgen

*Team Ohio State at CMCL 2021 Shared Task: Fine-Tuned RoBERTa for Eye-Tracking Data Prediction*
Byung-Doh Oh

**June 10, 2021, Mexico City (GMT-5) (continued)**

*PIHKers at CMCL 2021 Shared Task: Cosine Similarity and Surprisal to Predict Human Reading Patterns.*
Lavinia Salicchi and Alessandro Lenci

*TALEP at CMCL 2021 Shared Task: Non Linear Combination of Low and High-Level Features for Predicting Eye-Tracking Data*
Franck Dary, Alexis Nasr and Abdellah Fourtassi

*MTL782_IITD at CMCL 2021 Shared Task: Prediction of Eye-Tracking Features Using BERT Embeddings and Linguistic Features*
Shivani Choudhary, Kushagri Tandon, Raksha Agarwal and Niladri Chatterjee

*KonTra at CMCL 2021 Shared Task: Predicting Eye Movements by Combining BERT with Surface, Linguistic and Behavioral Information*
Qi Yu, Aikaterini-Lida Kalouli and Diego Frassinelli

*CogNLP-Sheffield at CMCL 2021 Shared Task: Blending Cognitively Inspired Features with Transformer-based Language Models for Predicting Eye Tracking Patterns*
Peter Vickers, Rosa Wainwright, Harish Tayyar Madabushi and Aline Villavicencio

*Team ReadMe at CMCL 2021 Shared Task: Predicting Human Reading Patterns by Traditional Oculomotor Control Models and Machine Learning*
Alisan Balkoca, Abdullah Algan, Cengiz Acarturk and Çağrı Çöltekin

*Enhancing Cognitive Models of Emotions with Representation Learning*
Yuting Guo and Jinho D. Choi

*Production vs Perception: The Role of Individuality in Usage-Based Grammar Induction*
Jonathan Dunn and Andrea Nini

*Clause Final Verb Prediction in Hindi: Evidence for Noisy Channel Model of Communication*
Kartik Sharma, Niyati Bafna and Samar Husain

*Dependency Locality and Neural Surprisal as Predictors of Processing Difficulty: Evidence from Reading Times*
Neil Rathi

*Modeling Sentence Comprehension Deficits in Aphasia: A Computational Evaluation of the Direct-access Model of Retrieval*
Paula Lissón, Dorothea Pregla, Dario Paape, Frank Burchert, Nicole Stadie and Shravan Vasishth

*Sentence Complexity in Context*
Benedetta Iavarone, Dominique Brunato and Felice Dell'Orletta

*Evaluating the Acquisition of Semantic Knowledge from Cross-situational Learning in Artificial Neural Networks*
Mitja Nikolaus and Abdellah Fourtassi

*Representation and Pre-Activation of Lexical-Semantic Knowledge in Neural Language Models*
Steven Derby, Barry Devereux and Paul miller

*Relation Classification with Cognitive Attention Supervision*
Erik McGuire and Noriko Tomuro

*Graph-theoretic Properties of the Class of Phonological Neighbourhood Networks*
Rory Turnbull

*Contributions of Propositional Content and Syntactic Category Information in Sentence Processing*
Byung-Doh Oh and William Schuler

*The Effect of Efficient Messaging and Input Variability on Neural-Agent Iterated Language Learning*
Yuchen Lian, Arianna Bisazza and Tessa Verhoef

*Capturing Phonotactic Learning Biases with a Simple RNN*
Max Nelson, Brandon Prickett and Joe Pater

**16:30–17:30 Keynote Talk 2**

16:30–17:30 *The Importance of Individualized Text Formats for Readability*
Zoya Bylinskii

**June 10, 2021, Mexico City (GMT-5) (continued)**

**17:30–17:45 Closing Remarks**