CoNLL 2024

The 28th Conference on Computational Natural Language Learning

Proceedings of the Conference

November 15-16, 2024

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL) 317 Sidney Baker St. S Suite 400 - 134 Kerrville, TX 78028 USA Tel: +1-855-225-1962 acl@aclweb.org

ISBN 979-8-89176-178-0

Introduction

CoNLL is a conference organized yearly by SIGNLL (ACL's Special Interest Group on Natural Language Learning), focusing on theoretically, cognitively and scientifically motivated approaches to computational linguistics. This year, CoNLL was held alongside EMNLP 2024.

The program of CoNLL 2024 comprises 40 papers. This was the result of a careful selection process. Reviewing 97 received submissions resulted in a 41% acceptance rate.

Reviewing was organized into 10 tracks, each of them headed by one or two area chairs:

- Computational Psycholinguistics, Cognition and Linguistics (Nathan Schneider)
- Computational Social Science (Kate Atwell)
- Interaction and Grounded Language Learning (Anthony Sicilia)
- Lexical, Compositional and Discourse Semantics (Shira Wein)
- Multilingual Work and Translation (Yuval Marton)
- Natural Language Generation (Tuhin Chakrabarty)
- Resources and Tools for Scientifically Motivated Research (Venkat)
- Speech and Phonology (Huteng Dai)
- Syntax and Morphology (Leshem Choshen)
- Theoretical Analysis and Interpretation of ML Models for NLP (Kevin Small)

We thank our reviewers and area chairs for curating the program. The conference also invited Thamar Solorio and Lorna Quandt to present keynotes, and included a session of additional papers on the BabyLM Challenge, a shared task that challenges community members to train a language model from scratch on the same amount of linguistic data available to a child in addition to multi-modal data.

We would like to acknowledge support from our sponsor, Google DeepMind.

Malihe Alikhani (Northeastern University) Libby Barak (Montclair State University) CoNLL 2024 conference co-chairs

Organizing Committee

Program Chairs

Malihe Alikhani, Northeastern University, MA, USA Libby Barak, Montclair State University, NJ, USA

Publication Chairs

Mert Inan, Northeastern University, MA, USA Julia Watson, University of Toronto, ON, Canada

SIGNLL Officers

President: Omri Abend, Hebrew University of Jerusalem, Israel Secretary: Antske Fokkens, Vrije Universiteit Amsterdam, Netherlands

Area Chairs

Katherine Atwell, Northeastern University and University of Pittsburgh Tuhin Chakrabarty, SalesForce Research Leshem Choshen, Massachusetts Institute of Technology and International Business Machines Huteng Dai, University of Michigan - Ann Arbor Venkata Govindarajan, Ithaca College Yuval Marton, Genentech and University of Washington Nathan Schneider, Georgetown University Anthony Sicilia, Northeastern University Kevin Small, Amazon Shira Wein, Amherst College

Invited Talks

Thamar Solorio, Mohamed bin Zayed University of Artificial Intelligence, MBZUAI Lorna Quandt, Gallaudet University

Program Committee

Reviewers

Adnen Abdessaied, Omri Abend, Kahaerjiang Abiderexiti, Gulinigeer Abudouwaili, Sayantan Adak, Mohammad Akbari, Albina Akhmetgareeva, Syeda Nahida Akter, Bashar Alhafni, Afra Alishahi, Milad Alshomary, Samuel Joseph Amouyal, Avinash Anand, Aryaman Arora, Daiki Asami, Hayastan Avetisyan

Siddhesh Bangar, Anish Bhanushali, Swarnadeep Bhar, Debasmita Bhattacharya, Arianna Bisazza, Gemma Boleda, Angana Borah, Aaron Ari Bornstein, Digbalay Bose, Marwen Bouabid, Jonathan Brennan, David Broneske, Nam Khac-Hoai Bui, Luana Bulla, Davide Buscaldi

Kranti CH, Jie Cao, Georgia Carter, Giovanni Cassani, Franklin Chang, Guanyi Chen, Yangbin Chen, Yiwen Chen, Fei Cheng, Santhosh Cherian, Emmanuele Chersoni, Ta-Chung Chi, Leshem Choshen, Philipp Cimiano, Caio Filippo Corro, Ailis Cournane, Francisco M. Couto

Walter Daelemans, Forrest Davis, Andrea Gregor De Varda, Shumin Deng, Aniket Deroy, Zi-Yi Dou, Rotem Dror, Xiaotang Du, Jonathan Dunn, Esra Dönmez

Oliver Eberle, Yo Ehara, Niama El Khbir, Saman Enayati

Abdellah Fourtassi, Diego Frassinelli, Daniel Freudenthal, Quchen Fu, Yingxue Fu, Yoshinari Fujinuma

Chongyang Gao, Lingyu Gao, Mengshi Ge, Luke Gessler, Mohammad Reza Ghasemi Madani, Shinjini Ghosh, Mario Giulianelli, Aniket Goel, Carlos-Emiliano González-Gallardo, Michael Eric Goodale, Emily Goodwin, Calbert Graham, Ximena Gutierrez

Sherzod Hakimov, Caren Han, Sadid A. Hasan, Shizhu He, Daniel Hershcovich, Xiaoyu Hu, Yebowen Hu, Zhe Hu, Xinting Huang

Katsumi Ibaraki, Mert Inan

Cassandra L Jacobs, Abhik Jana, Ganesh Jawahar, Aditya Joshi, Daksh Joshi

N J Karthika, Marc A. Kastner, Sedrick Keh, Casey Kennington, Tracy Holloway King, Christo Kirov, Thomas H Kober, Tom Kouwenhoven, Alex Krauska, Alexandra Krauska, Tatsuki Kuribayashi

Philippe Langlais, Patrick Lee, Jochen L. Leidner, Ruosen Li, Xia Li, Xiangci Li, Jiangming Liu, Wei Lu, Yunfei Luo, Kai Lv

Zhiyuan Ma, Brielen Madureira, Ayush Maheshwari, Subhankar Maity, Andreas Maletti, Biswadip Mandal, Stella Markantonatou, Edison Marrese-Taylor, Bruno Martins, Jacob A. Matthews, Yevgen Matusevych, Kate McCurdy, Yisong Miao, Timothee Mickus, Manuel Montes, Steven Moran

Nona Naderi, Vinh Van Nguyen, Ratna Nirupama, Ahmad Pir Noman, Tadashi Nomoto

Kazumasa Omura, Jessica Ouyang

Vishakh Padmakumar, Koyena Pal, Saurabh Kumar Pandey, Iñigo Parra, Tiago Pimentel, Priya Pitre, Lidia Pivovarova, Łukasz Pszenny

Siya Qi

Ella Rabinovich, Alexandre Rademaker, Peter A. Rankel, Jesse Roberts, Nathan Roll, Tanya Roosta, Guy Rotman, Sarthak Roy, Alla Rozovskaya

Parisa Safikhani, Rana Salama, Elizabeth Salesky, Tanja Samardzic, Malaikannan Sankarasubbu, Sashank Santhanam, Giorgio Satta, Marten Van Schijndel, Peter Schulam, William Schuler, Aditi Seetha, Mong Yuan Sim, Atul Kumar Singh, Kanishk Singh, Sudipta Singha Roy, Prasanna Srinivasa Murthy, Shane Steinert-Threlkeld, Ruolin Su, Jiuding Sun, Kaiser Sun, Renliang Sun, Wenjun Sun

Wei Tao, R Tharaniya Sairaj, Pascal Tilli, Gaurav Singh Tomar, Rong Tong, Thi Hong Hanh Tran

Marco Valentino, Bram Van Dijk, Tessa Verhoef, Prashanth Vijayaraghavan, Esaú Villatoro-tello, Marta Villegas, Pavlos Vougiouklis, Ivan Vulić

Yuiga Wada, Jiajing Wan, Bin Wang, Jiaan Wang, Jianyu Wang, Xiaomeng Wang, Yiwei Wang, Taro Watanabe, Gijs Wijnholds, Alina Wróblewska, Ting-Wei Wu

Qihui Xu

Hsiu-Yu Yang, Yizhe Yang, Roman Yangarber, Shoubin Yu

Yuan Zang, Sina Zarrieß, Nan Zhang, Ningyu Zhang, Tianlin Zhang, Wei Emma Zhang, Xiang Zhang, Yijia Zhang, Ying Zhang, Yuhan Zhang, Yunxiang Zhang, Yuqing Zhang, Yusen Zhang, Zhisong Zhang, Guangzhen Zhao, Jin Zhao, Yu Zhao, Zhihong Zhu, Heike Zinsmeister, Jinan Zou

Keynote Talk

Towards AI models that can help us to become better global social beings

Thamar Solorio

Mohamed bin Zayed University of Artificial Intelligence, MBZUAI

Abstract: Cultural norms and values fundamentally shape our social interactions. Communication within any society reflects these cultural contexts. For example, while direct eye contact is often seen as a sign of confidence in many Western cultures, it may be viewed as disrespectful in other parts of the world. Moreover, human-human interactions include so much more than just the words we utter; nonverbal communication, including body language and other cues, provides rich signals to those around us.

As vision language models (VLMs) are increasingly integrated into user-facing applications, it is becoming relevant to wonder if and to what extent this technology can robustly process these signals. My research group is interested in developing evaluation frameworks to assess the abilities of VLMs concerning interpreting social cues and in developing new approaches that can assist us and, perhaps, enhance our cross-cultural human-human interactions.

Bio: Thamar Solorio is a professor in the NLP department at MBZUAI. She is also a tenured professor of Computer Science at the University of Houston. She is the director and founder of the RiTUAL Lab. Her research interests include NLP for low-resource settings and multilingual data, including code-switching and information extraction. More recently, she was moved towards language and vision problems, focusing on developing inclusive NLP. She received a National Science Foundation (NSF) CAREER award for her work on authorship attribution and was awarded the 2014 Emerging Leader ABIE Award in Honor of Denice Denton. She served two terms as an elected board member of the North American Chapter of the Association of Computational Linguistics (NAACL) and was PC co-chair for NAACL 2019. She is an Editor in Chief for the ACL Rolling Review (ARR) initiative and was a member of the advisory board for ARR. She serves as general chair for the 2024 Conference on Empirical Methods in Natural Language Processing.

Keynote Talk

Integrating AI-Driven Sign Language Technologies in Education: Recognition, Generation, and Interaction

Lorna Quandt

Gallaudet University

Abstract: This talk explores integrating AI-driven technologies in sign language research, covering the unique challenges of sign language recognition and generation. Dr. Quandt will explore these cuttingedge considerations through the lens of two research projects, ASL Champ! and BRIDGE. Both projects focus on sign language recognition and generation, which is crucial for advancing interaction in virtual and educational environments. ASL Champ! utilizes a dataset of 3D signs to enhance deep-learning-powered sign recognition in virtual reality. At the same time, BRIDGE extends this work by incorporating both recognition and generation of signs to create a more robust, interactive experience. This dual focus underscores the importance of pursuing recognition and generation in tandem rather than treating them as entirely distinct challenges. By leveraging advances in AI and natural language processing (NLP), we can create technologies that recognize and generate signs and facilitate deeper understanding and use of signed languages. These advancements hold great educational potential, particularly in providing more accessible tools for deaf students and enabling broader instruction in sign language. The talk will also address how these innovations can reshape the NLP field by widening the focus beyond spoken/written language and into multimodal, signed, and nonverbal aspects of language, which can inform all linguistic research.

Bio: Dr. Lorna Quandt is the Action & Brain Lab director at Gallaudet University in Washington, D.C. She serves as Co-Director of the VL2 Research Center alongside Melissa Malzkuhn. Dr. Quandt is an Associate Professor in the Ph.D. in Educational Neuroscience (PEN) program and the Science Director of the Motion Light Lab. Dr. Quandt founded the Action & Brain lab in early 2016. Before that, Dr. Quandt obtained her BA in Psychology from Haverford College and a PhD in Psychology, specializing in Brain & Cognitive Sciences, from Temple University. She completed a postdoctoral fellowship at the University of Pennsylvania, working with Dr. Anjan Chatterjee. Her research examines how knowledge of sign language changes perception, particularly visuospatial processing. Dr. Quandt is also pursuing the development of research-based educational technology to create new ways to learn signed languages in virtual reality.

Table of Contents

Words That Stick: Using Keyword Cohesion to Improve Text Segmentation Amit Maraj, Miguel Vargas Martin and Masoud Makrehchi1
Investigating large language models for their competence in extracting grammatically sound sentences from transcribed noisy utterances Alina Wróblewska
Multi-Cultural Norm Base: Frame-based Norm Discovery in Multi-Cultural SettingsViet Thanh Pham, Shilin QU, Farhad Moghimifar, Suraj Sharma, Yuan-Fang Li, Weiqing Wangand Reza Haf24
Lossy Context Surprisal Predicts Task-Dependent Patterns in Relative Clause Processing Kate McCurdy and Michael Hahn
Global-Pruner: A Stable and Efficient Pruner for Retraining-Free Pruning of Encoder-Based Language Models
Guangzhen Yao, Yuehan Wang, Hui Xu, Long Zhang and MiaoQI MiaoQI46
<i>Transformer verbatim in-context retrieval across time and scale</i> Kristijan Armeni, Marko Pranjić and Senja Pollak
<i>EditEval: An Instruction-Based Benchmark for Text Improvements</i> Jane Dwivedi-Yu, Timo Schick, Zhengbao Jiang, Maria Lomeli, Patrick Lewis, Gautier Izacard, Edouard Grave, Sebastian Riedel and Fabio Petroni
An Empirical Comparison of Vocabulary Expansion and Initialization Approaches For Language Mo- dels
Nandini Mundra, Aditya Nanda Kishore Khandavally, Raj Dabre, Ratish Puduppully, Anoop Kun- chukuttan and Mitesh M Khapra
Critical Questions Generation: Motivation and Challenges Blanca Calvo Figueras and Rodrigo Agerri
Information Association for Language Model Updating by Mitigating LM-Logical Discrepancy Pengfei Yu and Heng Ji
Causal ATE Mitigates Unintended Bias in Controlled Text Generation Rahul Madhavan and Kahini Wadhawan
<i>On Functional Competence of LLMs for Linguistic Disambiguation</i> Raihan Kibria, Sheikh Intiser Uddin Dipta and Muhammad Abdullah Adnan
AIStorySimilarity: Quantifying Story Similarity Using Narrative for Search, IP Infringement, and Gui- ded Creativity Jon Chun
SPAWNing Structural Priming Predictions from a Cognitively Motivated Parser Grusha Prasad and Tal Linzen
Global Learning with Triplet Relations in Abstractive Summarization Fengyu Lu, Jiaxin Duan and Junfei Liu
TpT-ADE: Transformer Based Two-Phase ADE Extraction Suryamukhi Kuchibhotla and Manish Singh

<i>The Effect of Surprisal on Reading Times in Information Seeking and Repeated Reading</i> Keren Gruteke Klein, Yoav Meiri, Omer Shubi and Yevgeni Berzak
<i>Revisiting Hierarchical Text Classification: Inference and Metrics</i> Roman Plaud, Matthieu Labeau, Antoine Saillenfest and Thomas Bonald
NeLLCom-X: A Comprehensive Neural-Agent Framework to Simulate Language Learning and Group Communication Yuchen Lian, Tessa Verhoef and Arianna Bisazza
A Novel Instruction Tuning Method for Vietnamese Mathematical Reasoning using Trainable Open- Source Large Language Models Nguyen Quang Vinh, Thanh-Do Nguyen, Vinh Van Nguyen and Nam Khac-Hoai Bui259
<i>Generalizations across filler-gap dependencies in neural language models</i> Katherine Howitt, Sathvik Nair, Allison Dods and Robert Melvin Hopkins
<i>Of Models and Men: Probing Neural Networks for Agreement Attraction with Psycholinguistic Data</i> Maxim Bazhukov, Ekaterina Voloshina, Sergey Pletenev, Arseny Anisimov, Oleg Serikov and Svetlana Toldova
<i>Is Structure Dependence Shaped for Efficient Communication?: A Case Study on Coordination</i> Kohei Kajikawa, Yusuke Kubota and Yohei Oseki
Large Language Model Recall Uncertainty is Modulated by the Fan Effect Jesse Roberts, Kyle Moore, Douglas Fisher, Oseremhen Ewaleifoh and Thao Pham
Continuous Attentive Multimodal Prompt Tuning for Few-Shot Multimodal Sarcasm Detection Soumyadeep Jana, Animesh Dey and Ranbir Singh Sanasam
Aligning Alignments: Do Colexification and Distributional Similarity Align as Measures of cross- lingual Lexical Alignment? Taelin Karidi, Eitan Grossman and Omri Abend
<i>Text2Afford: Probing Object Affordance Prediction abilities of Language Models solely from Text</i> Sayantan Adak, Daivik Agrawal, Animesh Mukherjee and Somak Aditya
 How Are Metaphors Processed by Language Models? The Case of Analogies Joanne Boisson, Asahi Ushio, Hsuvas Borkakoty, Kiamehr Rezaee, Dimosthenis Antypas, Zara Siddique, Nina White and Jose Camacho-Collados
Further Compressing Distilled Language Models via Frequency-aware Partial Sparse Coding of Embeddings Kohki Tamura, Naoki Yoshinaga and Masato Neishi 388
Translating Across Cultures: LLMs for Intralingual Cultural Adaptation Pushpdeep Singh, Mayur Patidar and Lovekesh Vig
<i>Explaining the Hardest Errors of Contextual Embedding Based Classifiers</i> Claudio Moisés Valiense De Andrade, Washington Cunha, Guilherme Fonseca, Ana Clara Sou- za Pagano, Luana De Castro Santos, Adriana Silvina Pagano, Leonardo Chaves Dutra Da Rocha and Marcos André Gonçalves
A Multime del Lance Lance and del "Fernere" Obierte David en Verb Laferna di anti hat Net Cardan

A Multimoo	dal Large	Language I	Model "F	Foresees"	Objects	Based o	n Verb	Information	but Not	Gender
Shuqi	Wang, X	ufeng Duai	n and Zhe	enguang (Cai					435

PRACT: Optimizing Principled Reasoning and Acting of LLM Agent

Zhiwei Liu, Weiran Yao, Jianguo Zhang, Zuxin Liu, Liangwei Yang, Rithesh R N, Tian Lan,
Ming Zhu, Juntao Tan, Shirley Kokane, Thai Quoc Hoang, Juan Carlos Niebles, Shelby Heinecke,
Huan Wang, Silvio Savarese and Caiming Xiong
Image-conditioned human language comprehension and psychometric benchmarking of visual language models
Subha Nawer Pushpita and Roger P. Levy
Self-supervised speech representations display some human-like cross-linguistic perceptual abilities Joselyn Rodriguez, Kamala Sreepada, Ruolan Leslie Famularo, Sharon Goldwater and Naomi Feldman
One-Vs-Rest Neural Network English Grapheme Segmentation: A Linguistic Perspective Samuel Rose, Nina Dethlefs and C. Kambhampati
CrowdCounter: A benchmark type-specific multi-target counterspeech dataset
Punyajoy Saha, Abhilash Datta, Abhik Jana and Animesh Mukherjee
Solving the Challenge Set without Solving the Task: On Winograd Schemas as a Test of Pronominal Coreference Resolution
Ian Porada and Jackie CK Cheung
Advancing Arabic Sentiment Analysis: ArSen Benchmark and the Improved Fuzzy Deep Hybrid Network
Yang Fang, Cheng Xu, Shuhao Guan, Nan Yan and Yuke Mei
Leveraging a Cognitive Model to Measure Subjective Similarity of Human and GPT-4 Written Content
Tyler Malloy, Maria José Ferreira, Fei Fang and Cleotilde Gonzalez

Program

Friday, November 15, 2024

- 09:00 09:10 Opening Remarks
- 09:10 10:30 Keynote 1 Lorna Quandt
- 10:30 11:00 *Coffee Break*
- 11:00 12:30 Oral Session 1: Psycholinguistic Session (chair: Libby Barak)

Leveraging a Cognitive Model to Measure Subjective Similarity of Human and GPT-4 Written Content Tyler Malloy, Maria José Ferreira, Fei Fang and Cleotilde Gonzalez

SPAWNing Structural Priming Predictions from a Cognitively Motivated Parser Grusha Prasad and Tal Linzen

Lossy Context Surprisal Predicts Task-Dependent Patterns in Relative Clause Processing Kate McCurdy and Michael Hahn

A Multimodal Large Language Model "Foresees" Objects Based on Verb Information but Not Gender Shuqi Wang, Xufeng Duan and Zhenguang Cai

- 12:30 13:45 Lunch
- 13:45 15:30 *Poster Session 1*
- 15:30 16:00 *Coffee Break*
- 16:00 17:30 Oral Session 2: Syntax and Structure Session (chair: Omri Abend)

Is Structure Dependence Shaped for Efficient Communication?: A Case Study on Coordination Kohei Kajikawa, Yusuke Kubota and Yohei Oseki

NeLLCom-X: A Comprehensive Neural-Agent Framework to Simulate Language Learning and Group Communication Yuchen Lian, Tessa Verhoef and Arianna Bisazza

Friday, November 15, 2024 (continued)

Solving the Challenge Set without Solving the Task: On Winograd Schemas as a Test of Pronominal Coreference Resolution Ian Porada and Jackie CK Cheung

Global Learning with Triplet Relations in Abstractive Summarization Fengyu Lu, Jiaxin Duan and Junfei Liu

Saturday, November 16, 2024

- 09:00 09:10 Best Paper Awards
- 09:10 10:30 Keynote 2 Thamar Solorio
- 10:30 10:45 *Coffee Break*
- 10:45 12:15 Oral Session 3: LLM Session (chair: Malihe Alikhani)

Global-Pruner: A Stable and Efficient Pruner for Retraining-Free Pruning of Encoder-Based Language Models Guangzhen Yao, Yuehan Wang, Hui Xu, Long Zhang and MiaoQI MiaoQI

The second se

Investigating large language models for their competence in extracting grammatically sound sentences from transcribed noisy utterances Alina Wróblewska

The Effect of Surprisal on Reading Times in Information Seeking and Repeated Reading

Keren Gruteke Klein, Yoav Meiri, Omer Shubi and Yevgeni Berzak

Multi-Cultural Norm Base: Frame-based Norm Discovery in Multi-Cultural Settings

Viet Thanh Pham, Shilin QU, Farhad Moghimifar, Suraj Sharma, Yuan-Fang Li, Weiqing Wang and Reza Haf

- 12:45 13:45 Lunch
- 13:45 15:00 *Poster Session 2*
- 15:00 15:30 BabyLM Challenge (oral session)
- 15:30 16:00 *Coffee Break*
- 16:00 17:20 BabyLM Challenge (poster session)
- 17:20 17:30 Closing Remarks

Saturday, November 16, 2024 (continued)