

CoNLL 2025

**The 29th Conference on Computational Natural Language  
Learning (CoNLL 2025)**

**Proceedings of the Conference**

July 31 - August 1, 2025

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

## **Platinum**



©2025 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](mailto:acl@aclweb.org)

ISBN 979-8-89176-271-8

## Introduction

Welcome to CoNLL, the annual conference organized by the ACL Special Interest Group on Natural Language Learning (SIGNLL). As in previous years, CoNLL is a single-track conference with a focus on theoretically, cognitively and scientifically motivated approaches to computational linguistics, particularly exploring interactions between theoretical issues and computational modeling. This year, CoNLL is held alongside ACL 2025.

For our conference program, we received a total of 217 archival and 20 non-archival paper submissions, out of which we accepted 40 archival submissions (18%) and 6 non-archival submissions (30%). In addition to direct submissions, our program includes one presentation of a Findings paper accepted to the ACL main conference. Accepted papers are presented in two poster sessions and four oral sessions. In addition, the program includes a discussion session as well as two invited talks, by Raquel Fernández and Jean-Rémi King.

Our program committee consisted of 22 area chairs and 189 reviewers, whom we'd like to thank very much for providing constructive reviews and thoughtful recommendations. Special thanks go to our publication chairs, Emily Cheng and Selina Meyer, and publicity chairs, Snigdha Chaturvedi and Anvesh Rao Vijjini, for their crucial support. The SIGNLL board and last year's organizers provided numerous useful tips and important information. We are grateful for Google DeepMind's sponsoring of the Best Paper Award.

And, last but very much not least, a big thank you to the authors and the community, without whom this conference would not exist.

Gemma and Michael  
CoNLL 2025 conference co-chairs

# Organizing Committee

## **Program Chairs**

Gemma Boleda, Universitat Pompeu Fabra / ICREA, Spain  
Michael Roth, University of Technology Nuremberg, Germany

## **Publicity Chairs**

Snigdha Chaturvedi, University of North Carolina at Chapel Hill, USA  
Anvesh Rao Vijjini, University of North Carolina at Chapel Hill, USA

## **Publication Chairs**

Emily Cheng, Universitat Pompeu Fabra, Spain  
Selina Meyer, University of Technology Nuremberg, Germany

# Program Committee

## Program Chairs

Gemma Boleda, ICREA and Universitat Pompeu Fabra

Michael Roth, University of Technology Nuremberg

## Area Chairs

Christian Bentz, Universität Passau

Zhenguang Cai, Chinese University of Hong Kong

Tanise Ceron, Bocconi University

Jackie CK Cheung, McGill University

Kilian Evang, Heinrich Heine University Düsseldorf

Agnieszka Falenska, University of Stuttgart

Aina Garí Soler, Télécom-Paris

Ximena Gutierrez, Universidad Nacional Autónoma de México

Dieuwke Hupkes, Meta

Xixian Liao, Barcelona Supercomputing Center

Brielen Madureira, University of Potsdam

Yohei Oseki, University of Tokyo

Sandro Pezzelle, University of Amsterdam

Emily Prud'hommeaux, Boston College

Carina Silberer, University of Stuttgart

Ece Takmaz, Utrecht University

Max Johannes Van Duijn, Leiden University

Tessa Verhoef, Leiden University

Yang Xu, University of Toronto

Wei Zhao, University of Aberdeen

Iria de-Dios-Flores, Universitat Pompeu Fabra

## Reviewers

The list of all reviewers for CoNLL 2025 (in alphabetical order). Area chairs indicated outstanding reviewers in their track. These are marked \*.

Omri Abend, Gulinigeer Abudouwaili, Rodrigo Agerri, Albina Akhmetgareeva, Afra Alishahi, Samuel Joseph Amouyal, Avinash Anand, Jacob Andreas\*, Tatiana Anikina, David Arps\*, Daiki Asami, Eleftherios Avramidis

Roberto Basili, Anna Bavaresco\*, Christian Bentz\*, Gábor Berend, Swarnadeep Bhar\*, Debasmita Bhattacharya, Shohini Bhattasali, Arianna Bisazza\*, Tatiana Bladier, Gosse Bouma, Tomas J.f. Breedveld\*, David Broneske, Khac-Hoai Nam Bui

Jose Camacho-Collados\*, Jie Cao, Spencer Caplan, Giovanni Cassani\*, Tanise Ceron, Guanyi Chen, Long Chen, Xinchu Chen, Yiwen Chen, Emmanuele Chersoni, Pelin Çelikkol

Verna Dankers\*, Forrest Davis, Marianne De Heer Kloots\*, Mathieu Dehouck\*, Shumin Deng, Aniket Deroy, Antoine Doucet, Rotem Dror, Xiaotang Du, Jonathan Dunn, Ritam Dutt\*, Iria de-Dios-Flores

Yo Ehara, Saman Enayati

Abdellah Fourtassi, Dayuan Fu, Yingxue Fu\*, Yoshinari Fujinuma, Richard Futrell

Lingyu Gao, Aina Garí Soler\*, Mengshi Ge, Mohammad Reza Ghasemi Madani, Shinjini Ghosh, Maria Gini, Mario Giulianelli\*, Goran Glavaš, Carlos-Emiliano González-Gallardo, Michael Eric Goodale, Karan Gupta

Sherzod Hakimov, Lifeng Han, Caren Han\*, Daniel Hershcovich\*, Mark Hopkins, Xiaoyu Hu

Cassandra L Jacobs, Abhik Jana, Ganesh Jawahar, Yichen Jiang

Kohei Kajikawa\*, Ratna Kandala, N J Karthika, Marc A. Kastner, Sedrick Keh, Casey Kennington, Tracy Holloway King, Tom Kouwenhoven\*, Chalamalasetti Kranti, Tatsuki Kuribayashi

Leo Laugier, Jochen L. Leidner\*, Martha Lewis\*, Junlin Li, Meng Li, Xia Li, Xixian Liao, Dongqi Liu, Jiangming Liu, Yunfei Luo, Kai Lv

Subhankar Maity\*, Biswadip Mandal, Stella Markantonatou, Yuichiroh Matsubayashi, Yevgen Matuskevych\*, Maximilian Maurer, R. Thomas McCoy\*, Kate McCurdy\*, Robert Mercer\*, Selina Meyer\*, Yisong Miao, Timothee Mickus, Yusuke Miyao, Misael Mongiovi\*, Manuel Montes, Steven Moran, Samuel Musker, Arianna Muti

Yee Man Ng, Vinh Van Nguyen, Dmitry Nikolaev, Tadashi Nomoto

Kazumasa Omura, Yohei Oseki, Jessica Ouyang\*

Vishakh Padmakumar\*, Minkyung Park, Olga Pelloni, Simon Petitjean, Tiago Pimentel\*, Priya Pitre, Lidia Pivovarova, Flor Miriam Plaza-del-Arco\*, Maja Popovic, Emily Prud'hommeaux\*, Peter Van Der Putten

Siya Qi\*

Ella Rabinovich, Alexandre Rademaker, Aida Ramezani, Anthony Rios, Jesse Roberts, Nathan Roll\*, Carolyn Rose, Sumegh Roychowdhury, Adrián Rubio, Maria Ryskina

Philipp Sadler, Parisa Safikhani, Tanja Samardzic, Malaikannan Sankarasubbu, David Schlangen, William Schuler, Sabine Schulte Im Walde\*, Carina Silberer\*, Mong Yuan Sim, Kanishk Singh, Kevin Small\*, Yingjin Song\*, Shane Steinert-Threlkeld, Egon Stemle, Ruolin Su, Wenjun Sun

Thi Hong Hanh Tran

Michiel Van Der Meer, Bram Van Dijk, Jacob Hoover Vigly\*, Prashanth Vijayaraghavan, Esaú Villatoro-tello, Pavlos Vougiouklis

Yuiga Wada, Jianyu Wang, Xiaomeng Wang, Franziska Weeber\*, Alina Wróblewska\*, Ting-Wei Wu

Qihui Xu

Zukang Yang, Roman Yangarber

Sina Zarriß, Ningyu Zhang, Wei Emma Zhang, Xiang Zhang, Yuhan Zhang, Yuqing Zhang, Yusen Zhang, Guangzhen Zhao, Jiachen Zhao, Jin Zhao, Kai Zhao, Wei Zhao, Heike Zinsmeister\*, Jinan Zou



## Table of Contents

<i>HKCanto-Eval: A Benchmark for Evaluating Cantonese Language Understanding and Cultural Comprehension in LLMs</i>	
Tsz Chung Cheng, Chung Shing Cheng, Chaak-ming Lau, Eugene Lam, Wong Chun Yat, Hoi On Yu and Cheuk Hei Chong .....	1
<i>Quasi-symbolic Semantic Geometry over Transformer-based Variational AutoEncoder</i>	
Yingji Zhang, Danilo Carvalho and Andre Freitas .....	12
<i>LawToken: a single token worth more than its constituents</i>	
Yu-Hsiang Tseng, Hsin-Yu Chou and Shu-Kai Hsieh .....	30
<i>Interpersonal Memory Matters: A New Task for Proactive Dialogue Utilizing Conversational History</i>	
Bowen Wu, Wenqing Wang, Lihaoran Lihaoran, Yunhan Deng, Ying Li, Jingsong Yu and Baoxun Wang .....	47
<i>WinoWhat: A Parallel Corpus of Paraphrased WinoGrande Sentences with Common Sense Categorization</i>	
Ine Gevers, Victor De Marez, Luna De Bruyne and Walter Daelemans .....	68
<i>Planning for Success: Exploring LLM Long-term Planning Capabilities in Table Understanding</i>	
Thi-Nhung Nguyen, Hoang Ngo, Dinh Phung, Thuy-Trang Vu and Dat Quoc Nguyen .....	81
<i>Derivational Probing: Unveiling the Layer-wise Derivation of Syntactic Structures in Neural Language Models</i>	
Taiga Someya, Ryo Yoshida, Hitomi Yanaka and Yohei Oseki .....	93
<i>Short-circuiting Shortcuts: Mechanistic Investigation of Shortcuts in Text Classification</i>	
Leon Eshuijs, Shihan Wang and Antske Fokkens .....	105
<i>A Linguistically Motivated Analysis of Intonational Phrasing in Text-to-Speech Systems: Revealing Gaps in Syntactic Sensitivity</i>	
Charlotte Pouw, Afra Alishahi and Willem Zuidema .....	126
<i>Experiential Semantic Information and Brain Alignment: Are Multimodal Models Better than Language Models?</i>	
Anna Bavaresco and Raquel Fernández .....	141
<i>What is an Abstract Reasoner"? Revisiting Experiments and Arguments about Large Language Models</i>	
Tian Yun, Chen Sun and Ellie Pavlick .....	156
<i>Do Construction Distributions Shape Formal Language Learning In German BabyLMs?</i>	
Bastian Bunzeck, Daniel Duran and Sina Zarri�� .....	169
<i>Adapting Large Language Models for Movie Domain with Narrative Understanding Tasks</i>	
Siqi Shen and Amanmeet Garg .....	187
<i>From Stories to Statistics: Methodological Biases in LLM-Based Narrative Flow Quantification</i>	
Amal Sunny, Advay Gupta, Yashashree Chandak and Vishnu Sreekumar .....	201
<i>Components of Creativity: Language Model-based Predictors for Clustering and Switching in Verbal Fluency</i>	
Sina Zarri��, Simeon Junker, Judith Sieker and ��zge Alacam .....	216

<i>An Appraisal Theoretic Approach to Modelling Affect Flow in Conversation Corpora</i> Alok Debnath, Yvette Graham and Owen Conlan .....	233
<i>Principal Parts Detection for Computational Morphology: Task, Models and Benchmark</i> Dorin Keshales, Omer Goldman and Reut Tsarfaty .....	251
<i>Accelerating Large Language Model Pretraining via LFR Pedagogy: Learn, Focus, and Review</i> Neha Prakriya, Jui-Nan Yen, Cho-Jui Hsieh and Jason Cong .....	268
<i>What does memory retrieval leave on the table? Modelling the Cost of Semi-Compositionality with MINERVA2 and sBERT</i> Sydelle De Souza, Ivan Vegner, Francis Mollica and Leonidas A. A. Doumas .....	291
<i>Polarity inversion operators in PLM</i> David Kletz, Pascal Amsili and Marie Candito .....	312
<i>Dynamic Epistemic Friction in Dialogue</i> Timothy Obiso, Kenneth Lai, Abhijnan Nath, Nikhil Krishnaswamy and James Pustejovsky .	323
<i>A Three-Tier LLM Framework for Forecasting Student Engagement from Qualitative Longitudinal Data</i> Ahatsham Hayat, Helen Martinez, Bilal Khan and Mohammad Rashedul Hasan .....	334
<i>Bridging the Socioeconomic Gap in Education: A Hybrid AI and Human Annotation Approach</i> Nahed Abdelgaber, Labiba Jahan, Arham Vinit Doshi, Rishi Suri, Hamza Reza Pavel and Jia Zhang .....	348
<i>Construction Identification and Disambiguation Using BERT: A Case Study of NPN</i> Wesley Scivetti and Nathan Schneider .....	365
<i>Evidence of Generative Syntax in LLMs</i> Mary Kennedy .....	377
<i>Timestep Embeddings Trigger Collapse in Diffusion Text Generation</i> Ryota Nosaka and Takuya Matsuzaki .....	397
<i>Investigating Psychometric Predictive Power of Syntactic Attention</i> Ryo Yoshida, Yushi Sugimoto and Yohei Oseki .....	407
<i>A Continuous Approach to Metaphorically Motivated Regular Polysemy in Language Models</i> Anna Temerko, Marcos Garcia and Pablo Gamallo .....	419
<i>Is Incremental Structure Prediction Process Universal across Languages?: Revisiting Parsing Strategy through Speculation</i> Taiga Ishii and Yusuke Miyao .....	437
<i>Lost in Variation? Evaluating NLI Performance in Basque and Spanish Geographical Variants</i> Jaione Bengoetxea, Itziar Gonzalez-Dios and Rodrigo Agerri .....	452
<i>Compositionality and Event Retrieval in Complement Coercion: A Study of Language Models in a Low-resource Setting</i> Matteo Radaelli, Emmanuele Chersoni, Alessandro Lenci and Giosuè Baggio .....	469
<i>DLU: Dictionary Look-Up Data and Prediction</i> David Strohmaier, Gladys Tyen, Hongyi gu, Diane Nicholls, Zheng Yuan and Paula Buttery .	481
<i>IPA CHILDES &amp; G2P+: Feature-Rich Resources for Cross-Lingual Phonology and Phonemic Language Modeling</i> Zebulon Goriely and Paula Buttery .....	502

<i>BabyLM’s First Words: Word Segmentation as a Phonological Probing Task</i>	
Zebulon Goriely and Paula Buttery .....	522
<i>GCG-Based Artificial Languages for Evaluating Inductive Biases of Neural Language Models</i>	
Nadine El-Naggar, Tatsuki Kuribayashi and Ted Briscoe .....	540
<i>Beyond Accuracy: Revisiting Out-of-Distribution Generalization in NLI Models</i>	
Zahra Delbari and Mohammad Taher Pilehvar .....	557
<i>Spatial relation marking across languages: extraction, evaluation, analysis</i>	
Barend Beekhuizen .....	571
<i>Human-likeness of LLMs in the Mental Lexicon</i>	
Bei Xiao, Xufeng Duan, David A. Haslett and Zhenguang Cai .....	586
<i>Vorm: Translations and a constrained hypothesis space support unsupervised morphological segmentation across languages</i>	
Barend Beekhuizen .....	602
<i>Do large language models solve verbal analogies like children do?</i>	
Tamar Johnson, Mathilde ter Veen, Rochelle Choenni, Han van der Maas, Ekaterina Shutova and Claire E Stevenson .....	627

# Program

## Thursday, July 31, 2025

09:20 - 09:30 *Opening Remarks*

09:30 - 10:30 *Invited talk: Raquel Fernández*

10:30 - 11:00 *Coffee break*

11:00 - 12:20 *Oral session 1: Phonology, Morphology, and Syntax*

*A Linguistically Motivated Analysis of Intonational Phrasing in Text-to-Speech Systems: Revealing Gaps in Syntactic Sensitivity*

Charlotte Pouw, Afra Alishahi and Willem Zuidema

*Principal Parts Detection for Computational Morphology: Task, Models and Benchmark*

Dorin Keshales, Omer Goldman and Reut Tsarfaty

*Is Incremental Structure Prediction Process Universal across Languages?: Revisiting Parsing Strategy through Speculation*

Taiga Ishii and Yusuke Miyao

*GCG-Based Artificial Languages for Evaluating Inductive Biases of Neural Language Models*

Nadine El-Naggar, Tatsuki Kuribayashi and Ted Briscoe

12:20 - 14:00 *Lunch break*

14:00 - 15:30 *Poster session 1*

*HKCanto-Eval: A Benchmark for Evaluating Cantonese Language Understanding and Cultural Comprehension in LLMs*

Tsz Chung Cheng, Chung Shing Cheng, Chaak-ming Lau, Eugene Lam, Wong Chun Yat, Hoi On Yu and Cheuk Hei Chong

*LawToken: a single token worth more than its constituents*

Yu-Hsiang Tseng, Hsin-Yu Chou and Shu-Kai Hsieh

*Interpersonal Memory Matters: A New Task for Proactive Dialogue Utilizing Conversational History*

Bowen Wu, Wenqing Wang, Lihaoran Lihaoran, Yunhan Deng, Ying Li, Jingsong Yu and Baoxun Wang

**Thursday, July 31, 2025 (continued)**

*WinoWhat: A Parallel Corpus of Paraphrased WinoGrande Sentences with Common Sense Categorization*

Ine Gevers, Victor De Marez, Luna De Bruyne and Walter Daelemans

*Planning for Success: Exploring LLM Long-term Planning Capabilities in Table Understanding*

Thi-Nhung Nguyen, Hoang Ngo, Dinh Phung, Thuy-Trang Vu and Dat Quoc Nguyen

*Short-circuiting Shortcuts: Mechanistic Investigation of Shortcuts in Text Classification*

Leon Eshuijs, Shihan Wang and Antske Fokkens

*Do Construction Distributions Shape Formal Language Learning In German BabyLMs?*

Bastian Bunzeck, Daniel Duran and Sina Zarriß

*Adapting Large Language Models for Movie Domain with Narrative Understanding Tasks*

Siqi Shen and Amanmeet Garg

*What does memory retrieval leave on the table? Modelling the Cost of Semi-Compositionality with MINERVA2 and sBERT*

Sydelle De Souza, Ivan Vegner, Francis Mollica and Leonidas A. A. Douras

*Polarity inversion operators in PLM*

David Kletz, Pascal Amsili and Marie Candito

*Dynamic Epistemic Friction in Dialogue*

Timothy Obiso, Kenneth Lai, Abhijnan Nath, Nikhil Krishnaswamy and James Pustejovsky

*A Three-Tier LLM Framework for Forecasting Student Engagement from Qualitative Longitudinal Data*

Ahatsham Hayat, Helen Martinez, Bilal Khan and Mohammad Rashedul Hasan

*Bridging the Socioeconomic Gap in Education: A Hybrid AI and Human Annotation Approach*

Nahed Abdelgaber, Labiba Jahan, Arham Vinit Doshi, Rishi Suri, Hamza Reza Pavel and Jia Zhang

*Construction Identification and Disambiguation Using BERT: A Case Study of NPN*

Wesley Scivetti and Nathan Schneider

Thursday, July 31, 2025 (continued)

*Timestep Embeddings Trigger Collapse in Diffusion Text Generation*

Ryota Nosaka and Takuya Matsuzaki

*Lost in Variation? Evaluating NLI Performance in Basque and Spanish Geographical Variants*

Jaione Bengoetxea, Itziar Gonzalez-Dios and Rodrigo Agerri

*Compositionality and Event Retrieval in Complement Coercion: A Study of Language Models in a Low-resource Setting*

Matteo Radaelli, Emmanuele Chersoni, Alessandro Lenci and Giosuè Baggio

*DLU: Dictionary Look-Up Data and Prediction*

David Strohmaier, Gladys Tyen, Hongyi gu, Diane Nicholls, Zheng Yuan and Paula Buttery

*IPA CHILDES & G2P+: Feature-Rich Resources for Cross-Lingual Phonology and Phonemic Language Modeling*

Zebulon Goriely and Paula Buttery

*BabyLM's First Words: Word Segmentation as a Phonological Probing Task*

Zebulon Goriely and Paula Buttery

*Beyond Accuracy: Revisiting Out-of-Distribution Generalization in NLI Models*

Zahra Delbari and Mohammad Taher Pilehvar

*Spatial relation marking across languages: extraction, evaluation, analysis*

Barend Beekhuizen

*Human-likeness of LLMs in the Mental Lexicon*

Bei Xiao, Xufeng Duan, David A. Haslett and Zhenguang Cai

*Vorm: Translations and a constrained hypothesis space support unsupervised morphological segmentation across languages*

Barend Beekhuizen

*Do large language models solve verbal analogies like children do?*

Tamar Johnson, Mathilde ter Veen, Rochelle Choenni, Han van der Maas, Ekaterina Shutova and Claire E Stevenson

**Thursday, July 31, 2025 (continued)**

15:30 - 16:00     *Coffee break*

16:00 - 17:20     *Oral session 2: Semantics, Affect, and Multimodality*

*Experiential Semantic Information and Brain Alignment: Are Multimodal Models Better than Language Models?*

Anna Bavaresco and Raquel Fernández

*A Continuous Approach to Metaphorically Motivated Regular Polysemy in Language Models*

Anna Temerko, Marcos Garcia and Pablo Gamallo

*What is an Abstract Reasoner"? Revisiting Experiments and Arguments about Large Language Models*

Tian Yun, Chen Sun and Ellie Pavlick

*An Appraisal Theoretic Approach to Modelling Affect Flow in Conversation Corpora*

Alok Debnath, Yvette Graham and Owen Conlan

17:20 - 18:00     *Discussion session*

**Friday, August 1, 2025**

09:30 - 10:30 *Invited talk: Jean-Rémi King*

10:30 - 11:00 *Coffee break*

11:00 - 12:20 *Oral session 3: LLMs, Linguistic Structure, and Psycholinguistics*

*Evidence of Generative Syntax in LLMs*

Mary Kennedy

*Derivational Probing: Unveiling the Layer-wise Derivation of Syntactic Structures in Neural Language Models*

Taiga Someya, Ryo Yoshida, Hitomi Yanaka and Yohei Oseki

*Investigating Psychometric Predictive Power of Syntactic Attention*

Ryo Yoshida, Yushi Sugimoto and Yohei Oseki

*Components of Creativity: Language Model-based Predictors for Clustering and Switching in Verbal Fluency*

Sina Zarriß, Simeon Junker, Judith Sieker and Özge Alacam

12:20 - 14:00 *Lunch break*

14:00 - 15:30 *Poster session 2 (Poster Session 2 will feature the same poster presentations as Poster Session 1)*

15:30 - 16:00 *Coffee break*

16:00 - 17:00 *Oral session 4: Best Paper Session*

*Quasi-symbolic Semantic Geometry over Transformer-based Variational AutoEncoder*

Yingji Zhang, Danilo Carvalho and Andre Freitas

*Accelerating Large Language Model Pretraining via LFR Pedagogy: Learn, Focus, and Review*

Neha Prakriya, Jui-Nan Yen, Cho-Jui Hsieh and Jason Cong



**Friday, August 1, 2025 (continued)**

*From Stories to Statistics: Methodological Biases in LLM-Based Narrative Flow  
Quantification*

Amal Sunny, Advay Gupta, Yashashree Chandak and Vishnu Sreekumar

17:00 - 17:15     *Closing*