

IWCS 2025

16th International Conference on Computational Semantics

Proceedings of the Conference

September 22-23, 2025

©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

ISBN 979-8-89176-316-6

Introduction

These are the proceedings of the 16h International Conference on Computational Semantics, held at Heinrich Heine University, Düsseldorf, Germany, from 22 to 24 September 2025.

The aim of IWCS is to bring together researchers interested in all areas of computational semantics and computational aspects of meaning of natural language within written, spoken, signed, or multimodal communication. As shown in these proceedings, covered topics embrace both symbolic and machine learning approaches to computational semantics, in relation with multimodality, external knowledge, cognitive analysis, and many resources, e.g., annotation and software. We invited three keynote speakers to present their work: Oana-Maria Camburu (Department of Computing, Imperial College London, UK), Alexander Koller (Department of Language Science and Technology, Saarland University, Germany), and Denis Paperno (Utrecht University, Netherlands). The program also includes five oral presentation sessions and two poster sessions. Three satellite workshops will be held on the day after the conference:

- Bridges and Gaps between Formal and Computational Linguistics (BriGap2);
- The Second International Workshop on Construction Grammars and Natural Language Processing (CxGs+NLP 2025);
- A workshop joining the Second Workshop on Multimodal Semantic Representations (MMSR II) and the 21st Joint ACL – ISO Workshop on Interoperable Semantic Annotation (ISA-21).

We received 50 submissions (39 long and 11 short submissions) that each was assigned three reviewers. One long submission was withdrawn before review. Out of the remaining ones, 31 papers were accepted for the conference (25 long and 6 short), resulting in 14 oral presentations and 18 poster presentations, and a final acceptance rate of 63% (66% for long papers and 55% for short papers). Two long papers were withdrawn after acceptance. We are very grateful to the reviewers for their work and discussions that allowed us to produce a high-quality program for the conference.

In addition to this scientific work, this conference was made possible thanks to the local organizing team and the support of Haus der Universität.

We hope that IWCS 2025 will be an exciting edition and a lively forum to the computational semantics community.

Kilian Evang, Laura Kallmeyer, and Sylvain Pogodalla
September 2025

Organizing Committee

Program Chairs

Kilian Evang, Heinrich Heine University Düsseldorf
Laura Kallmeyer, Heinrich Heine University Düsseldorf
Sylvain Pogodalla, LORIA/Inria Nancy

Local Organizers

Long Chen, Heinrich Heine University Düsseldorf
Rafael Ehren, Heinrich Heine University Düsseldorf
Kilian Evang, Heinrich Heine University Düsseldorf
Laura Kallmeyer, Heinrich Heine University Düsseldorf
Rainer Osswald, Heinrich Heine University Düsseldorf
Christian Wurm, Heinrich Heine University Düsseldorf
Deniz Ekin Yavaş, Heinrich Heine University Düsseldorf

Program Committee

Program Committee

Rodrigo Agerri, University of the Basque Country
Maxime Amblard, Université de Lorraine
Daisuke Bekki, Ochanomizu University
Moritz Blum, Universität Bielefeld
Claire Bonial, Georgetown University
Johan Bos, University of Groningen
Susan Windisch Brown, University of Colorado at Boulder
Harry Bunt, Tilburg University
Aljoscha Burchardt, German Research Center for AI
Miriam Butt, Universität Konstanz
Long Chen, Heinrich-Heine Universität Düsseldorf
Emmanuele Chersoni, The Hong Kong Polytechnic University
Philipp Cimiano, Bielefeld University
Robin Cooper, University of Gothenburg
Marie Cousin, Université de Lorraine
Rodolfo Delmonte, Monash University
Markus Egg, Humboldt-Universität zu Berlin
Katrín Erk, University of Texas
Neele Falk, University of Stuttgart
Pascale Feldkamp, School of Communication and Culture
Francis Ferraro, University of Maryland
James Fodor, University of Melbourne
Philippe de Groote, INRIA
Bruno Guillaume, INRIA
Simon Guillot, Université du Maine
Dag Trygve Truslew Haug, University of Oslo
Johannes Heinecke, Orange-labs
Elisabetta Jezek, University of Pavia
Rohit J. Kate, University of Wisconsin - Milwaukee
Gene Louis Kim, University of South Florida
Philippe Langlais, Université de Montréal
Staffan Larsson, Göteborg University
Chuyuan Li, University of British Columbia
Zhaohui Luo, Royal Holloway University of London
Vladislav Maraev, Göteborg University
Gaëtan Margueritte, Tokyo University
Philipp Meier, Heinrich-Heine Universität Düsseldorf
Timothee Mickus, University of Helsinki
Yusuke Miyao, The University of Tokyo
Richard Moot, CNRS
Lawrence S. Moss, Indiana University at Bloomington
Dmitry Nikolaev, University of Manchester
Bill Noble, Göteborg University
Rik van Noord, University of Groningen
Juri Opitz, University of Zurich
Martha Palmer, University of Colorado at Boulder

Violaine Prince, University of Montpellier
Stephen Pulman, Apple
Matthew Purver, Queen Mary University of London
Giulia Rambelli, University of Bologna
Valentin D. Richard, University of Amsterdam
Kyeongmin Rim, Brandeis University
Mehrnoosh Sadrzadeh, University College London
Asad B. Sayeed, University of Gothenburg
Sabine Schulte im Walde, University of Stuttgart
Ravi Shekhar, University of Essex
Mollie Shichman, DEVCOM Army Research Laboratory
Mark Steedman, University of Edinburgh
Bonnie Webber, Edinburgh University
Shira Wein, Amherst College
Hitomi Yanaka, the University of Tokyo
Deniz Ekin Yavas, Heinrich Heine University Duesseldorf
Bingyang Ye, Brandeis University
Annie Zaenen, Stanford University
Alessandra Zarcone, Technische Hochschule Augsburg
Sina Zarri , Bielefeld University

Keynote Talk

**Thoughts You Can Trust? Evaluating the Faithfulness of
Model-Generated Explanations and Their Effects on Human
Performance**

Oana-Maria Camburu
Imperial College London

Abstract: Large Language Models (LLMs) can readily generate natural language explanations—or chain-of-thoughts (CoTs)—to justify their outputs. In this talk, I will first introduce methods for evaluating whether such explanations faithfully reflect the decision-making processes of the models that produce them. Second, I present the results of a user study involving 85 clinicians and medical students diagnosing chest X-rays. The study compares the effectiveness of natural language explanations, saliency maps, and their combination in supporting clinical decision-making.

Bio: Oana-Maria Camburu is an Assistant Professor in the Department of Computing at Imperial College London. Prior to that, she was a Principal Research Fellow in the Department of Computer Science at the University College London, holding an Early Career Leverhulme Fellowship. Oana was also a postdoc at the University of Oxford, from where she obtained her PhD in “Explaining Deep Neural Networks”. Her main research interests lie in explainability for deep learning models and AI safety and alignment.

Keynote Talk

Compositionality, Intensionality and LLMs: The Case of the Personal Relations Task

Denis Paperno
Utrecht University

Abstract: Semanticists have long considered compositionality to be at the heart of natural language interpretation. Modern large language models (LLMs) achieve impressive results at tasks involving semantics, but in most cases it is hard to judge to what extent they rely on compositional mechanisms. Since the training data is enormous and could contain many complex phrases, much of LLM’s performance could potentially be attributed to non-compositional pattern memorization, leaving little space for compositional ability. For example, “mother of Elon Musk” could be processed by a language model as a holistic unit since the phrase occurs in this form in the training corpora. The talk will discuss ongoing work based on the Personal Relations task (Paperno, 2022), designed to assess semantic compositionality properly. The Personal Relations task relies on a small universe with randomly generated relations which can not be present in language model pretraining, therefore offering a testing ground for compositional abilities of models at phrase level. Furthermore, the Personal Relations task allows us to contrast intensional and extensional semantic interpretation. We find that language models (still) exhibit different compositional abilities than humans, with intensionality playing a substantial role.

Denis Paperno. 2022. On Learning Interpreted Languages with Recurrent Models. *Computational Linguistics*, 48(2):471–482.

Bio: Denis Paperno is assistant professor of computational linguistics at Utrecht University. He received a PhD in Linguistics from the University of California Los Angeles, and subsequently worked at the University of Trento (CLIC lab, Rovereto) as a postdoc and at the Loria lab (Nancy) as a CNRS researcher. Denis has published extensively in the fields of semantics, language model evaluation, and vector space representations of meaning. His research contributions include work on compositionality in computational models of semantics, visual grounding, and representation probing.

Keynote Talk

Solving Complex Problems with Large Language Models

Alexander Koller
Saarland University

Abstract: One of the great promises that people connect with LLMs is that they can make complex problem-solving with computers accessible to lay users. Unlike traditional symbolic solvers (e.g. for planning or constraint solving), LLMs accept natural-language input and require no expert training; unlike earlier task-oriented dialogue systems, they can be applied across arbitrary domains. In my talk, I will explore the degree to which LLMs are fulfilling this promise. I will present recent work on whether current LLMs “reason” or “recite”, discuss the role of symbolic representations in LLM-based problem-solving, and offer some thoughts on trustworthy problem-solving with LLMs.

Bio: Alexander Koller is a Professor of Computational Linguistics at Saarland University in Saarbrücken, Germany. His research interests include planning and reasoning with LLMs, syntactic and semantic processing, natural language generation, and dialogue systems. He is particularly interested in neuro-symbolic models that bring together principled linguistic modeling and correctness guarantees with the coverage and robustness of neural approaches. Alexander received his PhD from Saarland University and was previously a postdoc at Columbia University and the University of Edinburgh, faculty at the University of Potsdam, and Visiting Senior Research Scientist at the Allen Institute for AI.

Table of Contents

<i>Advancing the Database of Cross-Linguistic Colexifications with New Workflows and Data</i> Annika Tjuka, Robert Forkel, Christoph Rzymiski and Johann-Mattis List	1
<i>FRIDA to the Rescue! Analyzing Synthetic Data Effectiveness in Object-Based Common Sense Reasoning for Disaster Response</i> Mollie Shichman, Claire Bonial, Austin Blodgett, Taylor Pellegrin, Francis Ferraro and Rachel Rudinger	16
<i>SemToken: Semantic-Aware Tokenization for Efficient Long-Context Language Modeling</i> Dong Liu and Yanxuan Yu	30
<i>ding-01 :ARG0: An AMR Corpus for Spontaneous French Dialogue</i> Jeongwoo Kang, Maria Boritchev and Maximin Coavoux	40
<i>A Graph Autoencoder Approach for Gesture Classification with Gesture AMR</i> Huma Jamil, Ibrahim Khebour, Kenneth Lai, James Pustejovsky and Nikhil Krishnaswamy ...	51
<i>Retrieval-Augmented Semantic Parsing: Improving Generalization with Lexical Knowledge</i> Xiao Zhang, Qianru Meng and Johan Bos	59
<i>Not Just Who or What: Modeling the Interaction of Linguistic and Annotator Variation in Hateful Word Interpretation</i> Sanne Hoeken, Özge Alacam, Dong Nguyen, Massimo Poesio and Sina Zarriß	73
<i>Context Effects on the Interpretation of Complement Coercion: A Comparative Study with Language Models in Norwegian</i> Matteo Radaelli, Emmanuele Chersoni, Alessandro Lenci and Giosuè Baggio	88
<i>LLMs Struggle with NLI for Perfect Aspect: A Cross-Linguistic Study in Chinese and Japanese</i> LU Jie, Du Jin and Hitomi Yanaka	99
<i>Assessing LLMs' Understanding of Structural Contrasts in the Lexicon</i> Shuxu LI, Antoine Venant, Philippe Langlais and François Lareau	108
<i>A German WSC dataset comparing coreference resolution by humans and machines</i> Wiebke Petersen and Katharina Spalek	120
<i>Finding Answers to Questions: Bridging between Type-based and Computational Neuroscience Approaches</i> Staffan Larsson, Jonathan Ginzburg, Robin Cooper and Andy Lücking	128
<i>Can Large Language Models Robustly Perform Natural Language Inference for Japanese Comparatives?</i> Yosuke Mikami, Daiki Matsuoka and Hitomi Yanaka	137
<i>Is neural semantic parsing good at ellipsis resolution, or isn't it?</i> Xiao Zhang and Johan Bos	147
<i>Extracting Behaviors from German Clinical Interviews in Support of Autism Spectrum Diagnosis</i> Margareta A. Kulcsar, Ian Paul Grant and Massimo Poesio	153
<i>The Proper Treatment of Verbal Idioms in German Discourse Representation Structure Parsing</i> Kilian Evang, Rafael Ehren and Laura Kallmeyer	166

<i>Does discourse structure help action prediction? A look at Correction Triangles.</i>	
Kate Thompson, Akshay Chaturvedi and Nicholas Asher	176
<i>FAMWA: A new taxonomy for classifying word associations (which humans improve at but LLMs still struggle with)</i>	
Maria A. Rodriguez, Marie Candito and Richard Huyghe	185
<i>Computational Semantics Tools for Glue Semantics</i>	
Mark-Matthias Zymla, Mary Dalrymple and Agnieszka Patejuk	199
<i>Which Model Mimics Human Mental Lexicon Better? A Comparative Study of Word Embedding and Generative Models</i>	
Huacheng Song, Zhaoxin Feng, Emmanuele Chersoni and Chu-Ren Huang	218
<i>Semantic Analysis Experiments for French Citizens' Contribution : Combinations of Language Models and Community Detection Algorithms</i>	
Sami Guembour, Dominguès Dominguès and Sabine Ploux	241
<i>Neurosymbolic AI for Natural Language Inference in French : combining LLMs and theorem provers for semantic parsing and natural language reasoning</i>	
Maximos Skandalis, Lasha Abzianidze, Richard Moot, Christian Retoré and Simon Robillard	252
<i>ProPara-CRTS: Canonical Referent Tracking for Reliable Evaluation of Entity State Tracking in Process Narratives</i>	
Bingyang Ye, Timothy Obiso, Jingxuan Tu and James Pustejovsky	264
<i>The Difficult Case of Intended and Perceived Sarcasm: a Challenge for Humans and Large Language Models</i>	
Hyewon Jang and Diego Frassinelli	279
<i>A Model of Information State in Situated Multimodal Dialogue</i>	
Kenneth Lai, Lucia Donatelli, Richard Brutti and James Pustejovsky	292
<i>Learning to Refer: How Scene Complexity Affects Emergent Communication in Neural Agents</i>	
Dominik Künkele and Simon Dobnik	299
<i>On the Role of Linguistic Features in LLM Performance on Theory of Mind Tasks</i>	
Ekaterina Kozachenko, Gonçalo Guimar and Karolina Stanczak	308
<i>Mapping Semantic Domains Across India's Social Media: Networks, Geography, and Social Factors</i>	
Gunjan Anand and Jonathan Dunn	317
<i>Disentangling lexical and grammatical information in word embeddings</i>	
Li Liu and François Lareau	331

Program

Haus der Universität, Schadowplatz 14, 40212 Düsseldorf

Monday, September 22, 2025

08:30 - 09:15 *Registration (Level 0)*

09:15 - 09:30 *Opening remarks (Level -1)*

09:30 - 10:30 *Invited speaker: Oana-Maria Camburu (Level -1)*

10:30 - 11:00 *Coffee break (Level 0)*

11:00 - 12:30 *Main session 1 (semantic parsing, linguistic phenomena) (Level -1)*

Neurosymbolic AI for Natural Language Inference in French : combining LLMs and theorem provers for semantic parsing and natural language reasoning

Maximos Skandalis, Lasha Abzianidze, Richard Moot, Christian Retoré and Simon Robillard

Is neural semantic parsing good at ellipsis resolution, or isn't it?

Xiao Zhang and Johan Bos

Retrieval-Augmented Semantic Parsing: Improving Generalization with Lexical Knowledge

Xiao Zhang, Qianru Meng and Johan Bos

12:30 - 14:00 *Lunch break*

14:00 - 15:00 *Invited speaker: Denis Paperno (Level -1)*

15:00 - 16:30 *Coffee break (Level 0)*

15:00 - 16:30 *Poster session 1a (LLMs) (Level 2)*

Can Large Language Models Robustly Perform Natural Language Inference for Japanese Comparatives?

Yosuke Mikami, Daiki Matsuoka and Hitomi Yanaka

LLMs Struggle with NLI for Perfect Aspect: A Cross-Linguistic Study in Chinese and Japanese

LU Jie, Du Jin and Hitomi Yanaka

Monday, September 22, 2025 (continued)

Assessing LLMs' Understanding of Structural Contrasts in the Lexicon

Shuxu LI, Antoine Venant, Philippe Langlais and François Lareau

The Difficult Case of Intended and Perceived Sarcasm: a Challenge for Humans and Large Language Models

Hyewon Jang and Diego Frassinelli

On the Role of Linguistic Features in LLM Performance on Theory of Mind Tasks

Ekaterina Kozachenko, Gonçalo Guimomar and Karolina Stanczak

15:00 - 16:30 *Poster session 1b (sociolinguistics, applied linguistics) (Level 3)*

Extracting Behaviors from German Clinical Interviews in Support of Autism Spectrum Diagnosis

Margareta A. Kulcsar, Ian Paul Grant and Massimo Poesio

Semantic Analysis Experiments for French Citizens' Contribution : Combinations of Language Models and Community Detection Algorithms

Sami Guembour, Dominguès Dominguès and Sabine Ploux

Mapping Semantic Domains Across India's Social Media: Networks, Geography, and Social Factors

Gunjan Anand and Jonathan Dunn

FRIDA to the Rescue! Analyzing Synthetic Data Effectiveness in Object-Based Common Sense Reasoning for Disaster Response

Mollie Shichman, Claire Bonial, Austin Blodgett, Taylor Pellegrin, Francis Ferraro and Rachel Rudinger

16:30 - 18:00 *Main session 2 (LLMs and compositionality, hybrid approaches) (Level -1)*

Context Effects on the Interpretation of Complement Coercion: A Comparative Study with Language Models in Norwegian

Matteo Radaelli, Emmanuele Chersoni, Alessandro Lenci and Giosuè Baggio

19:00 - 21:00 *Conference dinner (ALEX Düsseldorf, Graf-Adolf-Platz 15, 40213 Düsseldorf)*

Tuesday, September 23, 2025

08:30 - 09:30 *Registration (Level 0)*

09:30 - 10:30 *Invited speaker: Alexander Koller (Level -1)*

10:30 - 11:00 *Coffee break (Level 0)*

11:00 - 12:30 *Main session 3 (language models and linguistic knowledge) (Level -1)*

Disentangling lexical and grammatical information in word embeddings

Li Liu and François Lareau

*The Proper Treatment of Verbal Idioms in German Discourse Representation
Structure Parsing*

Kilian Evang, Rafael Ehren and Laura Kallmeyer

*SemToken: Semantic-Aware Tokenization for Efficient Long-Context Language
Modeling*

Dong Liu and Yanxuan Yu

12:30 - 14:00 *Lunch break*

14:00 - 15:00 *Main session 4 (semantics and cognition) (Level -1)*

*Which Model Mimics Human Mental Lexicon Better? A Comparative Study of
Word Embedding and Generative Models*

Huacheng Song, Zhaoxin Feng, Emmanuele Chersoni and Chu-Ren Huang

*Finding Answers to Questions: Bridging between Type-based and Computational
Neuroscience Approaches*

Staffan Larsson, Jonathan Ginzburg, Robin Cooper and Andy Lücking

15:00 - 16:30 *Coffee break (Level 0)*

15:00 - 16:30 *Poster session 2a (resources) (Level 2)*

A German WSC dataset comparing coreference resolution by humans and machines

Wiebke Petersen and Katharina Spalek

Tuesday, September 23, 2025 (continued)

FAMWA: A new taxonomy for classifying word associations (which humans improve at but LLMs still struggle with)

Maria A. Rodriguez, Marie Candito and Richard Huyghe

Advancing the Database of Cross-Linguistic Colexifications with New Workflows and Data

Annika Tjuka, Robert Forkel, Christoph Rzymiski and Johann-Mattis List

ding-01 :ARG0: An AMR Corpus for Spontaneous French Dialogue

Jeongwoo Kang, Maria Boritchev and Maximin Coavoux

Computational Semantics Tools for Glue Semantics

Mark-Matthias Zymla, Mary Dalrymple and Agnieszka Patejuk

15:00 - 16:30 *Poster session 2b (discourse and context) (Level 3)*

ProPara-CRTS: Canonical Referent Tracking for Reliable Evaluation of Entity State Tracking in Process Narratives

Bingyang Ye, Timothy Obiso, Jingxuan Tu and James Pustejovsky

Does discourse structure help action prediction? A look at Correction Triangles.

Kate Thompson, Akshay Chaturvedi and Nicholas Asher

Learning to Refer: How Scene Complexity Affects Emergent Communication in Neural Agents

Dominik Künkele and Simon Dobnik

A Graph Autoencoder Approach for Gesture Classification with Gesture AMR

Huma Jamil, Ibrahim Khebour, Kenneth Lai, James Pustejovsky and Nikhil Krishnaswamy

16:30 - 17:30 *Main session 5 (situated interpretation) (Level -1)*

Not Just Who or What: Modeling the Interaction of Linguistic and Annotator Variation in Hateful Word Interpretation

Sanne Hoeken, Özge Alacam, Dong Nguyen, Massimo Poesio and Sina Zarriß

A Model of Information State in Situated Multimodal Dialogue

Kenneth Lai, Lucia Donatelli, Richard Brutti and James Pustejovsky

Tuesday, September 23, 2025 (continued)

17:30 - 18:00 *Closing remarks (Level -1)*

Wednesday, September 24, 2025

08:30 - 09:00	<i>Registration (Level 0)</i>
09:00 - 18:00	<i>The Second International Workshop on Construction Grammars and Natural Language Processing (CxGs+NLP 2025) (Level -1)</i>
09:00 - 18:00	<i>ISA-21, The 21st Joint ACL - ISO Workshop on Interoperable Semantic Annotation (Level 2)</i>
09:00 - 18:00	<i>Bridges and Gaps between Formal and Computational Linguistics (BriGap2) (Level 3)</i>
10:30 - 11:00	<i>Coffee break (Level 0)</i>
15:30 - 16:00	<i>Coffee break (Level 0)</i>