

CRAC 2025

**The Eighth Workshop on Computational Models of
Reference, Anaphora and Coreference**

Proceedings of the Workshop

November 9, 2025

©2025 The 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 979-8-89176-342-5

Message from the Program Chairs

If you are one of the staunch supporters of CRAC, you should probably know that this is the 8th edition of CRAC (and the 10th edition if you also count the two CORBON workshops). Following the CRAC tradition, we requested that CRAC 2025 be co-located with the EMNLP conference. At the time of proposal submission, the location of EMNLP 2025 had not been finalized. It was only after we were notified of the acceptance of the proposal that we knew that EMNLP would take place in China, which was certainly a pleasant surprise to us. This will be the third time CRAC takes place in Asia, after CRAC 2022 in South Korea and CRAC 2023 in Singapore.

What is special about this year’s workshop is that this is the first time CRAC is held jointly with CODI (Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences) despite the fact that the two workshops organized two shared tasks together in 2021 and 2022. The organizing committee of the joint workshop is composed of the organizers from CRAC and CODI, who worked on the timeline for the workshop, the call for papers, the list of potential invited speakers, and the program schedule. The two workshops, however, had separate program committees, submission sites, and proceedings, and made acceptance decisions independently of each other.

This year, CRAC received 15 submissions, including nine research papers and six shared task papers. Each research paper was rigorously reviewed by three program committee members, and each shared task paper was reviewed by two. Based on their recommendations, we accepted all of the shared task papers, and among the nine research papers, we accepted six, conditionally accepted two, and rejected one. The two conditionally accepted papers were eventually accepted to the workshop after we made sure that the authors adequately addressed the reviewers’ comments in the final camera-ready version.

This year we continued to partner with our colleagues at Charles University, Prague and hosted the shared task on Multilingual Coreference Resolution for the fourth time at CRAC. The shared task allowed researchers who did not participate in the workshop to disseminate their work to a smaller and more focused audience which should promote interesting discussions. Following what we did last year, we similarly merged the shared task proceedings with the CRAC workshop proceedings this year. In other words, you can enjoy both the workshop papers and the shared task papers in this proceedings.

As you can imagine, fitting two invited talks, two shared tasks (the Multilingual Coreference Resolution shared task and the DISRPT shared task), and a large number of presentations of papers accepted to CODI and CRAC to a one-day program is by no means an easy task. In the end, the organizing committee decided to have two poster sessions (one in the morning and one in the afternoon) where the majority of the papers will be presented, selecting only a small number of papers for oral presentations. Even so, it has been logistically challenging for us to arrange for virtual paper presentations during the poster sessions.

We are grateful to the following people, without whom we could not have assembled an interesting program for the joint workshop. First, we are indebted to the CRAC program committee members. This year the average reviewing load was the equivalent of two long papers per reviewer. All of our program committee members did the incredible job of completing their reviews in a short reviewing period. Second, we thank Tanya Goyal and Nancy F. Chen, who are established researchers in Discourse, for accepting our invitation to be this year’s invited speakers. Finally, we would like to thank the workshop participants for joining us.

We hope you will enjoy the workshop and Suzhou as much as we do!

— Maciej Ogrodniczuk, Michal Novák, Massimo Poesio, Sameer Pradhan, and Vincent Ng

Organizers

Organizing Committee:

Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences, Poland
Michal Novák, Charles University in Prague, Czechia
Massimo Poesio, Queen Mary University of London, UK
Sameer Pradhan, University of Pennsylvania and cemantix.org, USA
Vincent Ng, University of Texas at Dallas, USA

Program Committee:

Jackie Chi Kit Cheung, Mila / McGill University, Canada
Loic De Langhe, Ghent University, Belgium
Elisa Ferracane, Abridge AI, Inc., USA
Yulia Grishina, Amazon, USA
Lars Hellan, Norwegian University of Science and Technology, Norway
Veronique Hoste, Ghent University, Belgium
Ekaterina Lapshinova-Koltunski, University of Hildesheim, Germany
Sharid Loáiciga, University of Gothenburg, Sweden.
Costanza Navaretta, University of Copenhagen, Denmark
Michal Novák, Charles University in Prague, Czechia
Constantin Orasan, University of Surrey, UK
Massimo Poesio, Queen Mary University of London, UK
Ian Porada, Mila – Quebec Artificial Intelligence Institute, Canada
Bonnie Webber, University of Edinburgh, UK
Juntao Yu, Queen Mary University of London, UK
Yilun Zhu, Georgetown University, USA
Heike Zinsmeister, University of Hamburg, Germany

Invited Talk 1

From Speech to Sense: The Art of Listening in Artificial Intelligence

Nancy F. Chen

Abstract

Unlike sight, which we can shut off with a blink, sound is inescapable. We are always listening, even when we wish not to. Hearing comes naturally, but understanding what we hear requires learning, knowledge, focus, and interpretation. Yet it is sound — be it the quiet drone of an air conditioner, a lover’s tender whisper, or the distant rush of a waterfall — that anchors us to our physical surroundings, social connections, and the present moment.

In this talk, I will share our experience in modelling the audio signal in multimodal generative AI to drive translational impact across domain applications. In particular, we exploit the audio modality to strengthen contextualization, reasoning, and grounding. Cultural nuances and multilingual peculiarities add another layer of complexity in understanding verbal interactions. Examples include our generative AI efforts in Singapore’s National Multimodal Large Language Model Programme has led to MERaLiON (Multimodal Empathetic Reasoning and Learning In One Network), the first multimodal large language model developed for Southeast Asia context. Such endeavors complement North American centric models to make generative AI more widely deployable for localized needs. Another case in point is SingaKids AI Tutor, which enables young children to learn ethnic languages such as Malay, Mandarin and Tamil. We are currently expanding applications to embodied agentic AI, aviation, and healthcare.

Speaker Bio

Nancy F. Chen is an ISCA Fellow (2025), AAIA Fellow (2025), and A*STAR Fellow (2023), and was recognized with the Asian Women Tech Leaders Award (2025). She is also a 2025 inductee of IEEE Eta Kappa Nu (HKN), the honor society of IEEE recognizing outstanding engineers.

At A*STAR, Dr. Chen leads the Multimodal Generative AI group and the AI for Education Programme. Dr. Chen is a serial best paper award winner across major conferences - including ICASSP, ACL, EMNLP, MICAIA, COLING, APSIPA, SIGDIAL and EACL – her research spans applications in education, healthcare, neuroscience, social media, security and forensics. Dr. Chen’s multimodal, multilingual technologies have led to commercial spin-offs and adoption by Singapore’s Ministry of Education.

Invited Talk 2

Climbing the Right Hill: On Benchmarking Progress in Long-Form Text Processing

Tanya Goyal

Abstract

Large Language Models (LLMs) are now functionally capable of ingesting very long documents as input, but can they truly process and reason over these massive contexts? In this talk, I will discuss our efforts at answering this question through the lens of long narrative summarization, a setting that naturally requires information synthesis and reasoning over long range dependencies. In the first part, I will describe our work highlighting shortcomings of current models along two key summary quality axes - coherence and factuality - and discuss challenges in automating their evaluation. Next, I will present NoCha, our methodology for constructing realistic and uncontaminated benchmarks for long context narrative reasoning. I will discuss results that show that NoCha is challenging for frontier LLMs; GPT-5 reports <30% worse performance compared to humans, and provide a recipe for building the next generation of robust long context benchmarks.

Speaker Bio

Tanya Goyal is an assistant professor in the Computer Science department at Cornell University.

Her research interests include building reliable and sustainable evaluation frameworks for large language models (LLMs) as well as understanding LLM behaviors as a function of training data and/or alignment strategies. Previously, she was a postdoctoral scholar at Princeton Language and Intelligence Center (2023-2024). Tanya completed her Ph.D. in Computer Science at UT Austin in 2023 where her thesis was awarded UTCS's Bert Kay Dissertation award. Her research is supported by NSF and a gift from Google.

Table of Contents

<i>Referential ambiguity and clarification requests: comparing human and LLM behaviour</i> Chris Madge, Matthew Purver and Massimo Poesio	1
<i>Coreference in simplified German: Linguistic features and challenges of automatic annotation</i> Sarah Jablotschkin, Ekaterina Lapshinova-Koltunski and Heike Zinsmeister	12
<i>Revisiting the Givenness Hierarchy. A Corpus-Based Evaluation</i> Christian Chiarcos	24
<i>Mention detection with LLMs in pair-programming dialogue</i> Cecilia Domingo, Paul Piwek, Svetlana Stoyanchev and Michel Wermelinger	42
<i>The Elephant in the Coreference Room: Resolving Coreference in Full-Length French Fiction Works</i> Antoine Bourgois and Thierry Poibeau	55
<i>Towards Adding Arabic to CorefUD</i> Dima Taji and Daniel Zeman	70
<i>Exploring Coreference Resolution in Glosses of German Sign Language</i> Yuzheng Bao and Haixia Chai	77
<i>Impact of ASR Transcriptions on French Spoken Coreference Resolution</i> Kirill Milintsevich	85
<i>Findings of the Fourth Shared Task on Multilingual Coreference Resolution: Can LLMs Dethrone Traditional Approaches?</i> Michal Novák, Miloslav Konopik, Anna Nedoluzhko, Martin Popel, Ondrej Prazak, Jakub Sido, Milan Straka, Zdeněk Žabokrtský and Daniel Zeman	95
<i>GLaRef@CRAC2025: Should we transform coreference resolution into a text generation task?</i> Olga Seminck, Antoine Bourgois, Yoann Dupont, Mathieu Dehouck and Marine Delaborde ...	119
<i>CorPipe at CRAC 2025: Evaluating Multilingual Encoders for Multilingual Coreference Resolution</i> Milan Straka	130
<i>Fine-Tuned Llama for Multilingual Text-to-Text Coreference Resolution</i> Jakub Hejman, Ondrej Prazak and Miloslav Konopík	140
<i>Few-Shot Coreference Resolution with Semantic Difficulty Metrics and In-Context Learning</i> Nguyen Xuan Phuc and Dang Van Thin	149
<i>Few-Shot Multilingual Coreference Resolution Using Long-Context Large Language Models</i> Moiz Sajid, Muhammad Fraz, Seemab Latif and Zuhair Zafar	154

Workshop Program

Sunday, November 9, 2025

Opening Remarks

09:00–09:10 *Opening and Welcome*
Maciej Ogrodniczuk and Michael Strube

Invited Talk 1

09:10–10:00 *From Speech to Sense: The Art of Listening in Artificial Intelligence*
Nancy F. Chen

Shared Tasks Overview

10:00–10:15 *Findings of the Fourth Shared Task on Multilingual Coreference Resolution: Can LLMs Dethrone Traditional Approaches?*
Michal Novák, Miloslav Konopik, Anna Nedoluzhko, Martin Popel, Ondrej Prazak, Jakub Sido, Milan Straka, Zdeněk Žabokrtský and Daniel Zeman

10:15–10:30 *The DISRPT 2025 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification*
Chloé Braud, Amir Zeldes, Chuyuan Li, Yang Janet Liu and Philippe Muller

10:30–11:00 *Coffee Break*

11:00–12:00 Poster Session 1

DisCuT and DiscReT: MELODI at DISRPT 2025 Multilingual discourse segmentation, connective tagging and relation classification
Robin Pujol, Firmin Rousseau, Philippe Muller and Chloé Braud

CLaC at DISRPT 2025: Hierarchical Adapters for Cross-Framework & Multilingual Discourse Relation Classification
Nawar Turk, Daniele Comitogianni and Leila Kosseim

DeDisCo at the DISRPT 2025 Shared Task: A System for Discourse Relation Classification
Zhuoxuan Ju, Jingni Wu, Abhishek Purushothama and Amir Zeldes

HITS at DISRPT 2025: Discourse Segmentation, Connective Detection, and Relation Classification
Souvik Banerjee, YI FAN and Michael Strube

SeCoRel: Multilingual Discourse Analysis in DISRPT 2025
Sobha Lalitha Devi, Pattabhi RK Rao and Vijay Sundar Ram

11:00–12:00 **Poster Session 1** (continued)

Few-Shot Coreference Resolution with Semantic Difficulty Metrics and In-Context Learning

Nguyen Xuan Phuc and Dang Van Thin

Fine-Tuned Llama for Multilingual Text-to-Text Coreference Resolution

Jakub Hejman, Ondrej Prazak and Miloslav Konopík

Few-Shot Multilingual Coreference Resolution Using Long-Context Large Language Models

Moiz Sajid, Muhammad Fraz, Seemab Latif and Zuhair Zafar

CorPipe at CRAC 2025: Evaluating Multilingual Encoders for Multilingual Coreference Resolution

Milan Straka

Code-switching in Context: Investigating the Role of Discourse Topic in Bilingual Speech Production

Debasmita Bhattacharya, Anxin Yi, Siying Ding and Julia Hirschberg

Discourse Relation Recognition with Language Models Under Different Data Availability

Shuhaib Mehri, Chuyuan Li and Giuseppe Carenini

Where Frameworks Disagree: A Study of Discourse Segmentation

Maciej Ogrodniczuk, Anna Latusek, Karolina Saputa, Alina Wróblewska, Daniel Ziembicki, Bartosz Żuk, Martyna Lewandowska, Adam Okrański, Paulina Rosalska, Anna Śliwicka, Aleksandra Tomaszewska and Sebastian Żurowski

Information-Theoretic and Prompt-Based Evaluation of Discourse Connective Edits in Instructional Text Revisions

Berfin Aktas and Michael Roth

Joint Modeling of Entities and Discourse Relations for Coherence Assessment

Wei Liu and Michael Strube

Towards Adding Arabic to CorefUD

Dima Taji and Daniel Zeman

Exploring Coreference Resolution in Glosses of German Sign Language

Yuzheng Bao and Haixia Chai

12:00–13:30 *Lunch Break*

Invited Talk 2

13:30–14:20 *Climbing the Right Hill: On Benchmarking Progress in Long-Form Text Processing*

Tanya Goyal

Long Context Benchmark for the Russian Language

Igor Churin, Murat Apishev, Maria Tikhonova, Denis Shevelev, Aydar S. Bulatov, Yuri Kuratov, Sergei A. Averkiev and Alena Fenogenova

Enhancing the Automatic Classification of Metadiscourse in Low-Proficiency Learners' Spoken and Written English Texts Using XLNet

Wenwen Guan, Marijn Alta and Jelke Bloem

Stance Detection on Nigerian 2023 Election Tweets Using BERT: A Low-Resource Transformer-Based Approach

Mahmoud Said Ahmad and Habeebah A. Kakudi

"Otherwise" in Context: Exploring Discourse Functions with Language Models

Guifu Liu, Bonnie Webber and Hannah Rohde

On the Role of Context for Discourse Relation Classification in Scientific Writing

Stephen Wan, Wei Liu and Michael Strube

Automated Conspiracy Narrative Detection Across Social Media Platforms

Calvin Yixiang Cheng and Mohsen Mosleh

Zero-Shot Belief: A Hard Problem for LLMs

John Murzaku and Owen Rambow

Probing the Limits of Multilingual Language Understanding: Nepali Proverbs as LLM Benchmark for AI Wisdom

Surendrabikram Thapa, Kritesh Rauniyar, Hariram Veeramani, Surabhi Adhikari, Imran Razzak and Usman Naseem

Measuring Sexism in US Elections: A Comparative Analysis of X Discourse from 2020 to 2024

Anna Fuchs, Elisa Noltenius, Caroline Weinzierl, Bolei Ma and Anna-Carolina Haensch

EmbiText: Embracing Ambiguity by Annotation, Recognition and Generation of Pronominal Reference with Event-Entity Ambiguity

Amna Sheikh and Christian Hardmeier

Human and LLM-based Assessment of Teaching Acts in Expert-led Explanatory Dialogues

Aliki Anagnostopoulou, Nils Feldhus, Yi-Sheng Hsu, Milad Alshomary, Henning Wachsmuth and Daniel Sonntag

Multi-token Mask-filling and Implicit Discourse Relations

Meinan Liu, Yunfang Dong, Xixian Liao and Bonnie Webber

Consistent Discourse-level Temporal Relation Extraction Using Large Language Models

Yi Fan and Michael Strube

Coreference in simplified German: Linguistic features and challenges of automatic annotation

Sarah Jablotschkin, Ekaterina Lapshinova-Koltunski and Heike Zinsmeister

Mention detection with LLMs in pair-programming dialogue

Cecilia Domingo, Paul Piwek, Svetlana Stoyanchev and Michel Wermelinger

The Elephant in the Coreference Room: Resolving Coreference in Full-Length French Fiction Works

Antoine Bourgois and Thierry Poibeau

14:20–15:30 **Poster Session 2** (continued)

Referential ambiguity and clarification requests: comparing human and LLM behaviour

Chris Madge, Matthew Purver and Massimo Poesio

Revisiting the Givenness Hierarchy. A Corpus-Based Evaluation

Christian Chiarcos

15:30–16:00 *Coffee break*

Oral Session

16:00–16:15 *Unpacking Ambiguity: The Interaction of Polysemous Discourse Markers and Non-DM Signals*

Jingni Wu and Amir Zeldes

16:15–16:30 *Impact of ASR Transcriptions on French Spoken Coreference Resolution*

Kirill Milintsevich

16:30–16:45 *GLaRef@CRAC2025: Should we transform coreference resolution into a text generation task?*

Olga Seminck, Antoine Bourgois, Yoann Dupont, Mathieu Dehouck and Marine Delaborde

16:45–17:00 *Entity Tracking in Small Language Models: An Attention-Based Study of Parameter-Efficient Fine-Tuning*

Sungho Jeon and Michael Strube

17:00–17:15 *Corpus-Oriented Stance Target Extraction*

Benjamin David Steel and Derek Ruths

17:15–17:30 *Bridging Discourse Treebanks with a Unified Rhetorical Structure Parser*

Elena Chistova

Closing Remarks

17:30–17:45 *Closing the workshop with Best Paper Awards*

Maciej Ogrodniczuk, Michal Novák, Michael Strube and Janet Liu