

COLING 2025

Proceedings of CoMeDi: Context and Meaning—Navigating  
Disagreements in NLP Annotations

January 19, 2025  
Abu Dhabi, UAE

©2025 International Committee on Computational Linguistics

ISBN 979-8-89176-202-2

## Introduction

Welcome to CoMeDi, the workshop on Context and Meaning—Navigating Disagreements in NLP Annotations!<sup>1</sup>

This workshop is taking place for the first time and explores causes of disagreements in NLP annotations as well as strategies for effective handling or resolution. This iteration of the workshop includes research on representing ambiguity in meaning representations, annotating implicit discourse relations, analyzing disagreement in politeness annotations, and predicting disagreement in tasks about word sense and hateful content, among others.

The workshop also hosted a shared task on ordinal word-in-context judgments. In the first subtask, participants were asked to predict the median of annotator judgments. In the second subtask, they needed to predict the mean of pairwise absolute judgment differences between annotators.

In total, we received 21 paper submissions (among them 7 shared task papers), out of which 17 were accepted. All workshop papers are presented as talks, while the shared task papers are presented in a poster session. In addition, the workshop includes two non-archival paper presentations and one invited talk on human label variation.

The program committee consisted of 20 researchers, who we would like to thank for providing helpful and constructive reviews on the papers. We would also like to thank all authors for their submissions and interest in our workshop.

Michael and Dominik

---

<sup>1</sup><https://comedinlp.github.io/>



**Organizers:**

Michael Roth, University of Technology Nuremberg

Dominik Schlechtweg, University of Stuttgart

**Program Committee:**

David Alfter, University of Gothenburg

Valerio Basile, University of Turin

Felipe Bravo, University of Chile

Jing Chen, Hong Kong Polytechnic University

Diego Frassinelli, University of Konstanz & LMU Munich

Dubossarsky Haim, Queen Mary University

Simon Hengchen, iguanodon.ai & Université de Genève

Snigdha Khanna, Indiana University

Sandra Kübler, Indiana University

Andrei Kutuzov, University of Oslo

Elisa Leonardelli, Fondazione Bruno Kessler

Melissa Lieffers, Indiana University

Marie-Catherine de Marneffe, UCLouvain

Maja Pavlovic, Queen Mary University

Siyao Peng, LMU Munich

Pauline Sander, University of Stuttgart

Pia Sommerauer, Vrije Universiteit Amsterdam

Alexandra Uma

Frank D. Zamora-Reina, University of Chile

Wei Zhao, University of Aberdeen

**Invited Speaker:**

Barbara Plank, LMU Munich & IT University of Copenhagen



## Table of Contents

<i>Is a bunch of words enough to detect disagreement in hateful content?</i> Giulia Rizzi, Paolo Rosso and Elisabetta Fersini .....	1
<i>On Crowdsourcing Task Design for Discourse Relation Annotation</i> Frances Yung and Vera Demberg .....	12
<i>Sources of Disagreement in Data for LLM Instruction Tuning</i> Russel Dsouza and Venelin Kovatchev .....	20
<i>CoMeDi Shared Task: Median Judgment Classification &amp; Mean Disagreement Ranking with Ordinal Word-in-Context Judgments</i> Dominik Schlechtweg, Tejaswi Choppa, Wei Zhao and Michael Roth .....	33
<i>Deep-change at CoMeDi: the Cross-Entropy Loss is not All You Need</i> Mikhail Kuklin and Nikolay Arefyev .....	48
<i>Predicting Median, Disagreement and Noise Label in Ordinal Word-in-Context Data</i> Tejaswi Choppa, Michael Roth and Dominik Schlechtweg .....	65
<i>GRASP at CoMeDi Shared Task: Multi-Strategy Modeling of Annotator Behavior in Multi-Lingual Semantic Judgments</i> David Alfter and Mattias Appelpgren .....	78
<i>Funzac at CoMeDi Shared Task: Modeling Annotator Disagreement from Word-In-Context Perspectives</i> Olufunke O. Sarumi, Charles Welch, Lucie Flek and Jörg Schlötterer .....	90
<i>FuocChuVIP123 at CoMeDi Shared Task: Disagreement Ranking with XLM-Roberta Sentence Embeddings and Deep Neural Regression</i> Phuoc Duong Huy Chu .....	97
<i>JuniperLiu at CoMeDi Shared Task: Models as Annotators in Lexical Semantics Disagreements</i> Zhu Liu, Zhen Hu and Ying Liu .....	103
<i>MMLabUIT at CoMeDiShared Task: Text Embedding Techniques versus Generation-Based NLI for Median Judgment Classification</i> Tai Duc Le and Thin Dang Van .....	113
<i>ABDN-NLP at CoMeDi Shared Task: Predicting the Aggregated Human Judgment via Weighted Few-Shot Prompting</i> Ying Xuan Loke, Dominik Schlechtweg and Wei Zhao .....	122
<i>Automating Annotation Guideline Improvements using LLMs: A Case Study</i> Adrien Bibal, Nathaniel Gerlek, Goran Muric, Elizabeth Boschee, Steven C. Fincke, Mike Ross and Steven N. Minton .....	129
<i>Ambiguity and Disagreement in Abstract Meaning Representation</i> Shira Wein .....	145
<i>Disagreement in Metaphor Annotation of Mexican Spanish Science Tweets</i> Alec M. Sanchez-Montero, Gemma Bel-Enguix, SERGIO LUIS OJEDA TRUEBA and Gerardo Sierra Martínez .....	155





# Workshop Program

**Sunday, January 19, 2025**

**9:20–11:00**    **Morning Session**

**9:20–9:30**    *Introduction*

9:30–10:30    *Invited talk: The Spectrum of Human Label Variation: Reflections from the Last Ten Years in NLP*  
Barbara Plank

**10:30–11:00**    *Coffee break*

**11:00–14:00**    **Pre-lunch Session**

11:00–11:20    *Is a bunch of words enough to detect disagreement in hateful content?*  
Giulia Rizzi, Paolo Rosso and Elisabetta Fersini

11:20–11:40    *On Crowdsourcing Task Design for Discourse Relation Annotation*  
Frances Yung and Vera Demberg

11:40–12:00    *Sources of Disagreement in Data for LLM Instruction Tuning*  
Russel Dsouza and Venelin Kovatchev

12:00–12:20    *CoMeDi Shared Task: Median Judgment Classification & Mean Disagreement Ranking with Ordinal Word-in-Context Judgments*  
Dominik Schlechtweg, Tejaswi Chopra, Wei Zhao and Michael Roth

**12:20–14:00**    *Lunch break*

**Sunday, January 19, 2025 (continued)**

**14:00–16:00 Post-lunch Session**

- 14:00–14:20 *Deep-change at CoMeDi: the Cross-Entropy Loss is not All You Need*  
Mikhail Kuklin and Nikolay Arefyev
- 14:20–14:40 *Predicting Median, Disagreement and Noise Label in Ordinal Word-in-Context Data*  
Tejaswi Choppa, Michael Roth and Dominik Schlechtweg

**14:40–15:30 Poster session**

- 14:40–15:30 *GRASP at CoMeDi Shared Task: Multi-Strategy Modeling of Annotator Behavior in Multi-Lingual Semantic Judgments*  
David Alfter and Mattias Appelgren
- 14:40–15:30 *Funzac at CoMeDi Shared Task: Modeling Annotator Disagreement from Word-In-Context Perspectives*  
Olufunke O. Sarumi, Charles Welch, Lucie Flek and Jörg Schlötterer
- 14:40–15:30 *FuocChuVIP123 at CoMeDi Shared Task: Disagreement Ranking with XLM-Roberta Sentence Embeddings and Deep Neural Regression*  
Phuoc Duong Huy Chu
- 14:40–15:30 *JuniperLiu at CoMeDi Shared Task: Models as Annotators in Lexical Semantics Disagreements*  
Zhu Liu, Zhen Hu and Ying Liu
- 14:40–15:30 *MMLabUIT at CoMeDiShared Task: Text Embedding Techniques versus Generation-Based NLI for Median Judgment Classification*  
Tai Duc Le and Thin Dang Van
- 14:40–15:30 *ABDN-NLP at CoMeDi Shared Task: Predicting the Aggregated Human Judgment via Weighted Few-Shot Prompting*  
Ying Xuan Loke, Dominik Schlechtweg and Wei Zhao
- 14:40–15:30 *Automating Annotation Guideline Improvements using LLMs: A Case Study*  
Adrien Bibal, Nathaniel Gerlek, Goran Muric, Elizabeth Boschee, Steven C. Fincke, Mike Ross and Steven N. Minton
- 14:40–15:30 *Ambiguity and Disagreement in Abstract Meaning Representation*  
Shira Wein
- 14:40–15:30 *The Impact of Annotation Choices on Computational Representations of Semantic and Phonological Distance in Sign Languages*  
Lisa Loy

**Sunday, January 19, 2025 (continued)**

14:40–15:30 *Ambiguity Meets Uncertainty: Investigating Uncertainty Estimation for Word Sense Disambiguation*  
Zhu Liu and Ying Liu

14:40–15:30 *Disagreement in Metaphor Annotation of Mexican Spanish Science Tweets*  
Alec M. Sanchez-Montero, Gemma Bel-Enguix, SERGIO LUIS OJEDA TRUEBA  
and Gerardo Sierra Martínez

15:30–16:00 *Coffee break*

16:00–16:10 **Closing Session**

16:00–16:10 *Closing*

