

CODI 2024

**5th Workshop on Computational Approaches to Discourse**

**Proceedings of the Workshop**

March 21, 2024

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

## **Sponsor**



©2024 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](mailto:acl@aclweb.org)

ISBN 979-8-89176-079-0

## Preface

Welcome to the 5th Workshop on Computational Approaches to Discourse, CODI!

CODI provides a venue to bring together researchers working on all aspects of discourse in Computational Linguistics and NLP. Our aim is to provide a venue for the entire discourse processing community where we can present and exchange our theories, algorithms, software, datasets, and tools.

The workshop consists of invited talks, contributed papers, extended abstracts, and EACL Findings presentations. We received paper submissions that span a wide range of topics, addressing issues related to discourse representation and parsing, reference and coreference resolution, summarization, dialogue, pragmatics, applications, and more. As the workshop is hybrid this year, papers are presented live either in person or remotely and discussed during live Q&A sessions. We received 28 submissions, including 14 regular long papers, 6 regular short papers and 8 non-archival communications (Findings, extended abstracts and direct submissions). We accepted 16 articles among the 20 regular submissions and 6 are presented orally. We also organize two poster sessions this year, in order to encourage discussions.

We thank our invited speakers, **Hannah Rohde**, Professor in Linguistics & English Language at the University of Edinburgh, who works in experimental pragmatics, focusing on aspects of communication such as ambiguity, redundancy, deception, and the establishment of discourse coherence. Her presentation is entitled: *Inferences of additional coherence-driven meaning within and across clauses*. Our second invited speaker is **Manfred Stede**, Professor of Applied Computational Linguistics at the University of Potsdam and head of the Applied CompLing Discourse Research Lab, who works on text structure and automatic text analysis, currently mainly for social science issues. His presentation is entitled: *Connectives and Arguments*. They helped us to prepare an excellent and well-rounded workshop program. We would also like to thank the EACL 2024 workshop chairs Zeerak Talat and Nafise Moosavi who organized the ACL workshops program.

Finally, we thank our sponsor HITS, Heidelberg Institute for Theoretical Studies <https://www.h-its.org>.

The CODI Organizers,

Chloé Braud, Christian Hardmeier, Chuyuan Li, Junyi Jessy Li, Sharid Loáiciga, Michael Strube, and Amir Zeldes

# Program Committee

## Program Committee

Giuseppe Carenini, University of British Columbia  
Jackie Chi Kit Cheung, Mila / McGill University  
Vera Demberg, Saarland University  
Elisa Ferracane, Abridge AI, Inc.  
Mark Finlayson, FIU  
Annemarie Friedrich, University of Augsburg  
Jie He, University of Edinburgh  
Freya Hewett, University of Potsdam  
Ryuichiro Higashinaka, Nagoya University/NTT  
Cassandra L. Jacobs, University at Buffalo  
Sungho Jeon, Heidelberg Institute for Theoretical Studies  
Yangfeng Ji, University of Virginia  
René Knaebel, University of Potsdam  
Ekaterina Lapshinova-Koltunski, University of Hildesheim  
Yang Janet Liu, Georgetown University  
Wanqiu Long, The University of Edinburgh  
S. Magalí López Cortez, University at Buffalo  
Philippe Muller, IRIT, University of Toulouse  
Jingcheng Niu, University of Toronto  
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences  
Massimo Poesio, Queen Mary University of London and University of Utrecht  
Hannah Rohde, University of Edinburgh  
Ahmed Ruby, Uppsala University  
Tatjana Scheffler, Ruhr University Bochum  
Manfred Stede, University of Potsdam  
Francielle Vargas, University of São Paulo  
Suncheng Xiang, Shanghai Jiao Tong University  
Wen Xiao, University of British Columbia  
Frances Yung, Saarland University  
Deniz Zeyrek, Middle East Technical University  
Haopeng Zhang, University of California, Davis  
Zheng Zhao, University of Edinburgh

## Table of Contents

<i>An Algorithmic Approach to Analyzing Rhetorical Structures</i> Andrew Potter .....	1
<i>SciPara: A New Dataset for Investigating Paragraph Discourse Structure in Scientific Papers</i> Anna Kiepora, Yingqiang Gao, Jessica Lam, Nianlong Gu and Richard H.r. Hahnloser .....	12
<i>Using Discourse Connectives to Test Genre Bias in Masked Language Models</i> Heidrun Dorgeloh, Lea Kawaletz, Simon Stein, Regina Stodden and Stefan Conrad .....	27
<i>Projecting Annotations for Discourse Relations: Connective Identification for Low-Resource Languages</i> Peter Bourgonje and Pin-Jie Lin .....	39
<i>Experimenting with Discourse Segmentation of Taiwan Southern Min Spontaneous Speech</i> Laurent Prévot and Sheng-Fu Wang .....	50
<i>Actor Identification in Discourse: A Challenge for LLMs?</i> Ana Barić, Sebastian Padó and Sean Papay .....	64
<i>Quantitative metrics to the CARS model in academic discourse in biology introductions</i> Charles Lam and Nonso Nnamoko .....	71
<i>Probing of pretrained multilingual models on the knowledge of discourse</i> Mary Godunova and Ekaterina Voloshina .....	78
<i>Feature-augmented model for multilingual discourse relation classification</i> Eleni Metheniti, Chloé Braud and Philippe Muller .....	91
<i>Complex question generation using discourse-based data augmentation</i> Khushnur Jahangir, Philippe Muller and Chloé Braud .....	105
<i>Exploring Soft-Label Training for Implicit Discourse Relation Recognition</i> Nelson Filipe Costa and Leila Kosseim .....	120
<i>The ARRAU 3.0 Corpus</i> Massimo Poesio, Maris Camilleri, Paloma Carretero Garcia, Juntao Yu and Mark-Christoph Müller .....	127
<i>Signals as Features: Predicting Error/Success in Rhetorical Structure Parsing</i> Martial Pastor and Nelleke Oostdijk .....	139
<i>GroundHog: Dialogue Generation using Multi-Grained Linguistic Input</i> Alexander Chernyavskiy, Lidiia Ostyakova and Dmitry Ilvovsky .....	149
<i>Discourse Relation Prediction and Discourse Parsing in Dialogues with Minimal Supervision</i> Chuyuan Li, Chloé Braud, Maxime Amblard and Giuseppe Carenini .....	161
<i>With a Little Help from my (Linguistic) Friends: Topic segmentation of multi-party casual conversations</i> Amandine Decker and Maxime Amblard .....	177

# Program

**Thursday, March 21, 2024**

09:15 - 09:30     *Opening Remarks*

09:30 - 10:30     *Invited Talk - Hannah Rohde: Inferences of additional coherence-driven meaning within and across clauses*

10:30 - 11:00     *Coffee Break*

11:00 - 12:00     *Poster Session I*

*Discourse Structure Annotations for the Minecraft Dialogue Corpus*  
Kate Thompson, Julie Hunter and Nicholas Asher

*Syntactic Preposing and Discourse Relations*  
Yunfang Dong, Xixian Liao and Bonnie Webber

*Learning Label Embeddings for Mapping Discourse Relations of Different Annotation Frameworks (Extended Abstract)*  
Yingxue Fu

*Actor Identification in Discourse: A Challenge for LLMs?*  
Ana Barić, Sebastian Padó and Sean Papay

*Exploring Soft-Label Training for Implicit Discourse Relation Recognition*  
Nelson Filipe Costa and Leila Kosseim

*Projecting Annotations for Discourse Relations: Connective Identification for Low-Resource Languages*  
Peter Bourgonje and Pin-Jie Lin

*Probing of pretrained multilingual models on the knowledge of discourse*  
Mary Godunova and Ekaterina Voloshina

*Complex question generation using discourse-based data augmentation*  
Khushnur Jahangir, Philippe Muller and Chloé Braud

**Thursday, March 21, 2024 (continued)**

12:00 - 13:30     *Lunch*

13:30 - 14:30     *Invited Talk - Manfred Stede: Connectives and Arguments*

14:30 - 15:30     *Poster Session 2*

*Zero-shot learning for multilingual discourse relation classification*

Eleni Metheniti, Philippe Muller, Chloé Braud and Margarita Hernández Casas

*Exploiting Questions Under Discussion for Discourse Relation Recognition in Dialog*

Shuhaib Mehri, Chuyuan Li and Giuseppe Carenini

*Quantitative metrics to the CARS model in academic discourse in biology introductions*

Charles Lam and Nonso Nnamoko

*On Measuring Context Utilization in Document-Level MT Systems*

Wafaa Mohammed and Vlad Niculae

*An Algorithmic Approach to Analyzing Rhetorical Structures*

Andrew Potter

*SciPara: A New Dataset for Investigating Paragraph Discourse Structure in Scientific Papers*

Anna Kiepura, Yingqiang Gao, Jessica Lam, Nianlong Gu and Richard H.r. Hahnloser

*The ARRAU 3.0 Corpus*

Massimo Poesio, Maris Camilleri, Paloma Carretero Garcia, Juntao Yu and Mark-Christoph Müller

*Discourse Relation Prediction and Discourse Parsing in Dialogues with Minimal Supervision*

Chuyuan Li, Chloé Braud, Maxime Amblard and Giuseppe Carenini

15:30 - 16:00     *Coffee Break*



**Thursday, March 21, 2024 (continued)**

16:00 - 17:30     *Session - Oral presentations*

*Using Discourse Connectives to Test Genre Bias in Masked Language Models*

Heidrun Dorgeloh, Lea Kawaletz, Simon Stein, Regina Stodden and Stefan Conrad

*Experimenting with Discourse Segmentation of Taiwan Southern Min Spontaneous Speech*

Laurent Prévot and Sheng-Fu Wang

*Feature-augmented model for multilingual discourse relation classification*

Eleni Metheniti, Chloé Braud and Philippe Muller

*Signals as Features: Predicting Error/Success in Rhetorical Structure Parsing*

Martial Pastor and Nelleke Oostdijk

*GroundHog: Dialogue Generation using Multi-Grained Linguistic Input*

Alexander Chernyavskiy, Lidiia Ostyakova and Dmitry Ilvovsky

*With a Little Help from my (Linguistic) Friends: Topic segmentation of multi-party casual conversations*

Amandine Decker and Maxime Amblard

17:30 - 17:45     *Closing Remarks*