

CODI 2023

4th Workshop on Computational Approaches to Discourse

Proceedings of the Workshop

July 13-14, 2023

©2023 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 978-1-959429-89-0

Preface

Welcome to the 4th 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 ACL 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 are pleased that CODI 2023 features the third edition of the DISRPT (DIScourse Relation Parsing and Treebanking) shared task on Discourse Segmentation, Connective and Relation Identification across Formalisms. As we hope that the next CODI workshops will also feature shared tasks and other special events, the workshop also includes a discussion on future shared tasks, special sessions on discourse representation and parsing, coreference resolution, and multilingual discourse processing and machine translation.

We thank our invited speakers, **Yufang Hou**, IBM Research Ireland and adjunct senior lecturer and co-supervisor at UKP Lab-TU Darmstadt, who works on referential discourse modeling, argument mining, and scholarly document processing; and **Giuseppe Carenini**, University of British Columbia, known for his work on discourse parsing, summarization, and generation. We would also like to thank our reviewers for their thoughtful and instructive comments. They helped us to prepare an excellent and inclusive workshop program. Finally we would like to thank the ACL 2023 workshop chairs Eduardo Blanco, Yang Feng, and Annie Louis who organized the ACL workshops program.

The CODI Organizers,

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

Program Committee

Chairs

Chloé Braud, IRIT - CNRS - ANITI
Christian Hardmeier, IT University of Copenhagen
Junyi Jessy Li, University of Texas at Austin
Sharid Loáiciga, University of Gothenburg
Michael Strube, Heidelberg Institute for Theoretical Studies
Amir Zeldes, Georgetown University

Program Committee

Katherine Atwell, University of Pittsburgh
Giuseppe Carenini, university of british columbia
Haixia Chai, Heidelberg Institute for Theoretical Studies gGmbH
Jackie Chi Kit Cheung, Mila / McGill University
Vera Demberg, Saarland University
Pascal Denis, INRIA
Elisa Ferracane, Abridge AI, Inc.
Zhengxian Gong, Computer Science and Technology School, Soochow University
Jie He, University of Edinburgh
Ryuichiro Higashinaka, Nagoya University/NTT
Sungho Jeon, Heidelberg Institute for Theoretical Studies
Prathyusha Jwalapuram, Rakuten
Murathan Kurfali, Stockholm University
Ekaterina Lapshinova-Koltunski, Stiftung Universität Hildesheim
Ramesh Manuvinakurike, Intel labs
Philippe Muller, IRIT, University of Toulouse
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Carolyn Rosé, Carnegie Mellon University
Manfred Stede, University of Potsdam
Francielle Vargas, University of São Paulo
Bonnie Webber, University of Edinburgh
Deniz Zeyrek, Middle East Technical University
Heike Zinsmeister, Universität Hamburg

Table of Contents

<i>MuLMS-AZ: An Argumentative Zoning Dataset for the Materials Science Domain</i> Timo Schrader, Teresa Bürkle, Sophie Henning, Sherry Tan, Matteo Finco, Stefan Grünewald, Maira Indrikova, Felix Hildebrand and Annemarie Friedrich	1
<i>A Side-by-side Comparison of Transformers for Implicit Discourse Relation Classification</i> Bruce W. Lee, Bongseok Yang and Jason Lee	16
<i>Ensemble Transfer Learning for Multilingual Coreference Resolution</i> Tuan Lai and Heng Ji	24
<i>Contrastive Hierarchical Discourse Graph for Scientific Document Summarization</i> Haopeng Zhang, Xiao Liu and Jiawei Zhang	37
<i>Leveraging Structural Discourse Information for Event Coreference Resolution in Dutch</i> Loic De Langhe, Orphee De Clercq and Veronique Hoste	48
<i>Entity Coreference and Co-occurrence Aware Argument Mining from Biomedical Literature</i> Boyang Liu, Viktor Schlegel, Riza Batista-navarro and Sophia Ananiadou	54
<i>A Weakly-Supervised Learning Approach to the Identification of Alternative Lexicalizations in Shallow Discourse Parsing</i> René Knaebel	61
<i>Entity-based SpanCopy for Abstractive Summarization to Improve the Factual Consistency</i> Wen Xiao and Giuseppe Carenini	70
<i>Discourse Information for Document-Level Temporal Dependency Parsing</i> Jingcheng Niu, Victoria Ng, Erin Rees, Simon De Montigny and Gerald Penn	82
<i>Encoding Discourse Structure: Comparison of RST and QUD</i> Sara Shahmohammadi, Hannah Seemann, Manfred Stede and Tatjana Scheffler	89
<i>Exploiting Knowledge about Discourse Relations for Implicit Discourse Relation Classification</i> Nobel Varghese, Frances Yung, Kaveri Anuranjana and Vera Demberg	99
<i>SAE-NTM: Sentence-Aware Encoder for Neural Topic Modeling</i> Hao Liu, Jingsheng Gao, Suncheng Xiang, Ting Liu and Yuzhuo Fu	106
<i>Improving Long Context Document-Level Machine Translation</i> Christian Herold and Hermann Ney	112
<i>Unpacking Ambiguous Structure: A Dataset for Ambiguous Implicit Discourse Relations for English and Egyptian Arabic</i> Ahmed Ruby, Sara Stymne and Christian Hardmeier	126
<i>Two-step Text Summarization for Long-form Biographical Narrative Genre</i> Avi Bleiweiss	145
<i>The distribution of discourse relations within and across turns in spontaneous conversation</i> S. Magalí López Cortez and Cassandra L. Jacobs	156
<i>Embedding Mental Health Discourse for Community Recommendation</i> Hy Dang, Bang Nguyen, Noah Ziems and Meng Jiang	163
<i>APA-RST: A Text Simplification Corpus with RST Annotations</i> Freya Hewett	173

Program

Thursday, July 13, 2023

09:15 - 09:30 *Opening Remarks*

09:30 - 10:30 *Invited Talk - Yufang Hou: Bridging Resolution: A Journey Towards Modelling Referential Discourse Entities*

10:30 - 11:00 *Coffee Break*

11:00 - 12:20 *Session - Summarization/Generation*

Contrastive Hierarchical Discourse Graph for Scientific Document Summarization

Hopeng Zhang, Xiao Liu and Jiawei Zhang

Entity-based SpanCopy for Abstractive Summarization to Improve the Factual Consistency

Wen Xiao and Giuseppe Carenini

Two-step Text Summarization for Long-form Biographical Narrative Genre

Avi Bleiweiss

Improving Long Context Document-Level Machine Translation

Christian Herold and Hermann Ney

12:30 - 14:00 *Lunch*

14:00 - 14:50 *Session - Coreference*

Ensemble Transfer Learning for Multilingual Coreference Resolution

Tuan Lai and Heng Ji

Leveraging Structural Discourse Information for Event Coreference Resolution in Dutch

Loic De Langhe, Orphee De Clercq and Veronique Hoste

Entity Coreference and Co-occurrence Aware Argument Mining from Biomedical Literature

Boyang Liu, Viktor Schlegel, Riza Batista-navarro and Sophia Ananiadou

Thursday, July 13, 2023 (continued)

14:50 - 15:30 *Session - Discourse Relations/Other*

Unpacking Ambiguous Structure: A Dataset for Ambiguous Implicit Discourse Relations for English and Egyptian Arabic

Ahmed Ruby, Sara Stymne and Christian Hardmeier

Embedding Mental Health Discourse for Community Recommendation

Hy Dang, Bang Nguyen, Noah Ziemis and Meng Jiang

15:30 - 16:00 *Coffee Break*

16:00 - 16:35 *Session - Rhetorical Structure Theory*

Encoding Discourse Structure: Comparison of RST and QUD

Sara Shahmohammadi, Hannah Seemann, Manfred Stede and Tatjana Scheffler

APA-RST: A Text Simplification Corpus with RST Annotations

Freya Hewett

16:35 - 17:25 *Session - Other*

Discourse Information for Document-Level Temporal Dependency Parsing

Jingcheng Niu, Victoria Ng, Erin Rees, Simon De Montigny and Gerald Penn

SAE-NTM: Sentence-Aware Encoder for Neural Topic Modeling

Hao Liu, Jingsheng Gao, Suncheng Xiang, Ting Liu and Yuzhuo Fu

Friday, July 14, 2023

09:30 - 10:30 *Invited Talk - Giuseppe Carenini: Discourse Processing in the era of Large Language Models*

10:30 - 11:00 *Coffee Break*

11:00 - 12:10 *Discourse Relation Parsing and Treebanking (DISRPT) Shared Task*

12:10 - 12:30 *Discussion of Future Shared Tasks*

12:30 - 14:00 *Lunch*

14:00 - 15:20 *Session - Discourse Relations/Other*

MuLMS-AZ: An Argumentative Zoning Dataset for the Materials Science Domain
Timo Schrader, Teresa Bürkle, Sophie Henning, Sherry Tan, Matteo Finco, Stefan Grünewald, Maira Indrikova, Felix Hildebrand and Annemarie Friedrich

A Side-by-side Comparison of Transformers for Implicit Discourse Relation Classification
Bruce W. Lee, Bongseok Yang and Jason Lee

A Weakly-Supervised Learning Approach to the Identification of Alternative Lexicalizations in Shallow Discourse Parsing
René Knaebel

Exploiting Knowledge about Discourse Relations for Implicit Discourse Relation Classification
Nobel Varghese, Frances Yung, Kaveri Anuranjana and Vera Demberg

The distribution of discourse relations within and across turns in spontaneous conversation
S. Magalí López Cortez and Cassandra L. Jacobs

15:20 - 17:00 *Coffee and Poster Session*

Chinese-DiMLex: A Lexicon of Chinese Discourse Connectives
Shujun Wan, Peter Bourgonje, Hongling Xiao, Clara Wan Ching Ho and Manfred Stede

Friday, July 14, 2023 (continued)

Discourse Structure Extraction from Pre-Trained and Fine-Tuned Language Models in Dialogues

Chuyuan Li, Patrick Huber, Wen Xiao, Maxime Amblard, Chloé Braud and Giuseppe Carenini

Discourse Analysis via Questions and Answers: Parsing Dependency Structures of Questions Under Discussion

Wei-jen Ko, Yating Wu, Cutter Dalton, Dananjay Srinivas, Greg Durrett and Junyi Jessy Li

SWiPE: A Dataset for Document-Level Simplification of Wikipedia Pages

Philippe Laban, Jesse Vig, Wojciech Kryscinski, Shafiq Joty, Caiming Xiong and Chien-sheng Wu

Replicate and Compare with Humans: LLMs Represent Partial Semantic Knowledge in Pronoun Interpretation

Suet-ying Lam, Qingcheng Zeng, Kexun Zhang, Chenyu You and Rob Voigt

Context-Aware Document Simplification

Liam Cripwell, Jol Legrand and Claire Gardent

An Email Dataset for Analyzing Large-Group Decision-Making

Mladen Karan, Prashant Khare, Ravi Shekhar, Stephen Mcquistin, Colin Perkins, Ignacio Castro, Gareth Tyson, Patrick Healey and Matthew Purver

GUMSum: Multi-Genre Data and Evaluation for English Abstractive Summarization

Yang Janet Liu and Amir Zeldes

An Integrated Approach for Political Bias Prediction and Explanation Based on Discursive Structure

Nicolas Devatine, Philippe Muller and Chloé Braud

*$2 * n$ is better than n^2 : Decomposing Event Coreference Resolution into Two Tractable Problems*

Shafiuddin Rehan Ahmed, Abhijnan Nath, James H. Martin and Nikhil Krishnaswamy

Imagination is All You Need! Curved Contrastive Learning for Abstract Sequence Modeling Utilized on Long Short-Term Dialogue Planning

Justus-jonas Erker, Stefan Schaffer and Gerasimos Spanakis

17:00 - 17:30 *Closing Remarks and Awards*

Friday, July 14, 2023 (continued)