

EMNLP 2021

**The Second Workshop on Computational Approaches to
Discourse**

Proceedings of the Workshop

November 10 - 11, 2021

©2021 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 978-1-954085-99-2

Acknowledgements

Christian Hardmeier was supported by the Swedish Research Council under grant 2017-930.

Preface

Welcome to the 2nd Workshop on Computational Approaches to Discourse, CODI 2021! CODI is intended to provide a venue for researchers working on all aspects of discourse. 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 paper and demo presentations, and discussion sessions. We received paper submissions that span a wide range of topics, addressing issues related to discourse representation and parsing, anaphora and coreference resolution, dialogue, 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. There will also be prerecorded videos available.

We are happy that CODI 2021 features two shared tasks. The CODI-CRAC 2021 shared task on anaphora, bridging, and discourse deixis resolution in dialogue is a joint shared task between the CODI and CRAC workshops at EMNLP. This shared task goes beyond previous ones as its input is transcribed spoken dialogue, and in that it includes anaphoric relations beyond coreference. This effort is spearheaded by Carolyn Rosé. The DISRPT 2021 shared task continues a series of successful DISRPT shared tasks on discourse segmentation, relation classification and connective detection. The organization team is led by Amir Zeldes. 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.

We thank our invited speakers, **Jackie Chi Kit Cheung**, McGill University, who works on language understanding and question answering in context, and **Vera Demberg**, Saarland University, who works on computational and cognitive models of text generation and understanding. We would also like to thank our reviewers for their thoughtful and instructive comments. They helped us to prepare an inclusive workshop program. Finally we would like to thank the EMNLP 2021 workshop chairs Parisa Kordjamshidi and Minlie Huang who organized the EMNLP workshop program under very challenging circumstances.

The CODI Organizers,

Chloé Braud, Christian Hardmeier, Junyi Jessy Li, Annie Louis, Michael Strube, and Amir Zeldes

Organizers:

Chloé Braud, CNRS - IRIT
Christian Hardmeier, University of Edinburgh and Uppsala University
Junyi Jessy Li, The University of Texas at Austin
Annie Louis, Google
Michael Strube, Heidelberg Institute for Theoretical Studies
Amir Zeldes, Georgetown University

Program Committee:

Carenini, Guiseppa, University of British Columbia, Canada
Cheung, Jackie Chi Kit, McGill University, Canada
Demberg, Vera, Saarland University, Germany
Denis, Pascal, INRIA, France
Dipper, Stefanie, University of Bochum, Germany
Gong, Zhengxian, Soochow University, China
Ferracane, Elisa, University of Texas at Austin, US
Grishina, Yulia, Amazon (Berlin), Germany
Inui, Kentaro, Tohoku University, Japan
Ji, Yangfeng, University of Virginia, US
Kurfali, Murathan, Stockholm University, Sweden
Kurohashi, Sadao, Kyoto University, Japan
Lapshinova-Koltunski, Ekaterina, Saarland University, Germany
Loáiciga, Sharid, University of Potsdam, Germany
Liu, Qun, Huawei, China
Muller, Philippe, Université Paul Sabatier, France
Müller, Mark-Christoph, HITS gGmbH, Germany
Ng, Vincent, University of Texas at Dallas, US
Ogrodniczuk, Maciej, Warsaw University, Poland
Poesio, Massimo, Queen Mary University of London, UK
Rutherford, Attapol Te, Chulalongkorn University, Thailand
Webber, Bonnie, University of Edinburgh, UK
Xiong, Deyi, Tianjin University, China
Zinsmeister, Heike, Hamburg University, Germany

Invited Speakers:

Jackie Chi Kit Cheung, McGill University, Canada
Vera Demberg, Saarland University, Germany

Table of Contents

<i>"I'll be there for you": The One with Understanding Indirect Answers</i> Cathrine Damgaard, Paulina Toborek, Trine Eriksen and Barbara Plank	1
<i>Developing Conversational Data and Detection of Conversational Humor in Telugu</i> Vaishnavi Pamulapati and Radhika Mamidi	12
<i>Investigating non lexical markers of the language of schizophrenia in spontaneous conversations</i> Chuyuan Li, Maxime Amblard, Chloé Braud, caroline demily, Nicolas Franck and Michel Musiol 20	
<i>Discourse-Driven Integrated Dialogue Development Environment for Open-Domain Dialogue Systems</i> Denis Kuznetsov, Dmitry Evseev, Lidia Ostyakova, Oleg Serikov, Daniel Kornev and Mikhail Burtsev	29
<i>Coreference Chains Categorization by Sequence Clustering</i> Silvia Federzoni, Lydia-Mai Ho-Dac and Cécile Fabre	52
<i>Resolving Implicit References in Instructional Texts</i> Talita Anthonio and Michael Roth	58
<i>A practical perspective on connective generation</i> Frances Yung, Merel Scholman and Vera Demberg	72
<i>Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin</i> Marian Marchal, Merel Scholman and Vera Demberg	84
<i>Comparison of methods for explicit discourse connective identification across various domains</i> Merel Scholman, Tianai Dong, Frances Yung and Vera Demberg	95
<i>Revisiting Shallow Discourse Parsing in the PDTB-3: Handling Intra-sentential Implicits</i> Zheng Zhao and Bonnie Webber	107
<i>Improving Multi-Party Dialogue Discourse Parsing via Domain Integration</i> Zhengyuan Liu and Nancy Chen	122
<i>discopy: A Neural System for Shallow Discourse Parsing</i> René Knaebel	128
<i>Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces</i> Ekaterina Lapshinova-Koltunski, Heike Przybyl and Yuri Bizzoni	134
<i>Capturing document context inside sentence-level neural machine translation models with self-training</i> Elman Mansimov, Gábor Melis and Lei Yu	143
<i>DMRST: A Joint Framework for Document-Level Multilingual RST Discourse Segmentation and Parsing</i> Zhengyuan Liu, Ke Shi and Nancy Chen	154
<i>Visualizing Cross-Lingual Discourse Relations in Multilingual TED Corpora</i> Zae Myung Kim, Vassilina Nikoulina, Dongyeop Kang, Didier Schwab and Laurent Besacier .	165

Conference Program

Nov. 10, 2021

09:00–12:00 **CODI-CRAC Shared Task**

09:05–09:30 *Welcome*

09:05–09:30 *The CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue*

Sopan Khosla, Juntao Yu, Ramesh Manuvinakurike, Vincent Ng, Massimo Poesio, Michael Strube and Carolyn Rosé

09:30–09:45 *Neural Anaphora Resolution in Dialogues*

Hideo Kobayashi, Shengjie Li and Vincent Ng

09:45–10:00 *Anaphora Resolution in Dialogue: Description of the DFKI-TalkingRobots System for the CODI-CRAC 2021 Shared-Task*

Tatiana Anikina, Cennet Oguz, Natalia Skachkova, Siyu Tao, Sharmila Upadhyaya and Ivana Kruijff-Korbyova

10:00–10:15 *The Pipeline Model for Resolution of Anaphoric Reference and Resolution of Entity Reference*

Hongjin Kim, Damrin Kim and Harksoo Kim

10:15–10:30 *An End-to-End Approach for Full Bridging Resolution*

Joseph Renner, Priyansh Trivedi, Gaurav Maheshwari, Rémi Gilleron and Pascal Denis

10:30–11:00 *Coffee Break*

11:00–11:15 *Adapted End-to-End Coreference Resolution System for Anaphoric Identities in Dialogues*

Liyan Xu and Jinho D. D. Choi

11:15–11:30 *Anaphora Resolution in Dialogue: Cross-Team Analysis of the DFKI-TalkingRobots Team Submissions for the CODI-CRAC 2021 Shared-Task*

Natalia Skachkova, Cennet Oguz, Tatiana Anikina, Siyu Tao, Sharmila Upadhyaya and Ivana Kruijff-Korbyova

11:30–11:45 *The CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis Resolution in Dialogue: A Cross-Team Analysis*

Shengjie Li, Hideo Kobayashi, and Vincent Ng

Nov. 10, 2021 (continued)

11:45–12:00 *Visioning Discussion and Next Steps*

12:00–13:30 *Lunch Break*

13:30–14:30 **Invited Talk**

Keeping Track of Entities Over Time, Minds, and Knowledge Sources

Jackie Chi Kit Cheung

14:30–14:45 *Mini Break*

14:45–15:45 **Pragmatics and Applications**

14:45–15:00 *"I'll be there for you": The One with Understanding Indirect Answers*

Cathrine Damgaard, Paulina Toborek, Trine Eriksen and Barbara Plank

15:00–15:10 *Improving Text Generation via Neural Discourse Planning*

Alexander Chernyavskiy and Dmitry Ilvovsky

15:10–15:25 *Developing Conversational Data and Detection of Conversational Humor in Telugu*

Vaishnavi Pamulapati and Radhika Mamidi

15:25–15:35 *Investigating non lexical markers of the language of schizophrenia in spontaneous conversations*

Chuyuan Li, Maxime Amblard, Chloé Braud, Caroline Demily, Nicolas Franck and Michel Musiol

15:35–15:45 *Discourse-Driven Integrated Dialogue Development Environment for Open-Domain Dialogue Systems*

Denis Kuznetsov, Dmitry Evseev, Lidia Ostyakova, Oleg Serikov, Daniel Kornev and Mikhail Burtsev

Nov. 10, 2021 (continued)

15:50–16:15 Anaphora and Coreference

15:50–16:00 *Coreference Chains Categorization by Sequence Clustering*

Silvia Federzoni, Lydia-Mai Ho-Dac and Cécile Fabre

16:00–16:15 *Resolving Implicit References in Instructional Texts*

Talita Anthonio and Michael Roth

16:15–16:45 Coffee Break

16:45–18:15 Discourse Relations

16:45–17:00 *A practical perspective on connective generation*

Frances Yung, Merel Scholman and Vera Demberg

17:00–17:15 *Semi-automatic discourse annotation in a low-resource language: Developing a connective lexicon for Nigerian Pidgin*

Marian Marchal, Merel Scholman and Vera Demberg

17:15–17:30 *Comparison of methods for explicit discourse connective identification across various domains*

Merel Scholman, Tianai Dong, Frances Yung and Vera Demberg

17:30–17:40 *A Novel Corpus of Discourse Structure in Humans and Computers*

Babak Hemmatian, Sheridan Feucht, Rachel Avram, Alexander Wey, Muskaan Garg, Kate Spitalnic, Carsten Eickhoff, Ellie Pavlick, Bjorn Sandstede and Steven A. Sloman

17:40–17:55 *Revisiting Shallow Discourse Parsing in the PDTB-3: Handling Intra-sentential Implicits*

Zheng Zhao and Bonnie Webber

17:55–18:05 *Improving Multi-Party Dialogue Discourse Parsing via Domain Integration*

Zhengyuan Liu and Nancy Chen

18:05–18:15 *discopy: A Neural System for Shallow Discourse Parsing*

René Knaebel

Nov. 11, 2021

09:00–10:00 Invited Talk

Inter annotator agreement in discourse annotation – the role of domain knowledge and individual differences

Vera Demberg

10:25–11:15 Discourse and Multilinguality

10:25–10:35 *Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces*

Ekaterina Lapshinova-Koltunski, Heike Przybyl and Yuri Bizzoni

10:35–10:50 *Capturing document context inside sentence-level neural machine translation models with self-training*

Elman Mansimov, Gábor Melis and Lei Yu

10:50–11:05 *DMRST: A Joint Framework for Document-Level Multilingual RST Discourse Segmentation and Parsing*

Zhengyuan Liu, Ke Shi and Nancy Chen

11:05–11:15 *Visualizing Cross-Lingual Discourse Relations in Multilingual TED Corpora*

Zae Myung Kim, Vassilina Nikoulina, Dongyeop Kang, Didier Schwab and Laurent Besacier

11:15–12:00 Discussion and Closing of Main CODI Workshop

12:00–13:30 Lunch Break

Nov. 11, 2021 (continued)

13:30–15:30 DISRPT 2021 Shared Task Session 1

13:30–13:45 *The DISRPT 2021 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification*
Amir Zeldes, Yang Janet Liu, Mikel Iruskieta, Philippe Muller, Chloé Braud and Sonia Badene

13:50–14:20 *A Transformer Based Approach towards Identification of Discourse Unit Segments and Connectives*
Sahil Bakshi and Dipti Misra Sharma

14:20–14:50 *Multi-lingual Discourse Segmentation and Connective Identification: MELODI at DISRPT2021*
Morteza Ezzabady, Philippe Muller, and Chloé Braud

14:50–15:20 *Delexicalised Multilingual Discourse Segmentation for DISRPT 2021 and Tense, Mood, Voice and Modality Tagging for 11 Languages*
Tillmann D’onicke

15:30–16:00 Coffee Break

16:00–17:30 DISRPT 2021 Shared Task Session 2

16:00–16:30 *A Unified Approach to Discourse Relation Classification in nine Languages*
Hanna Varachkina and Franziska Pannach

16:30–17:00 *DisCoDisCo at the DISRPT2021 Shared Task: A System for Discourse Segmentation, Classification, and Connective Detection*
Luke Gessler, Shabnam Behzad, Yang Janet Liu, Siyao Peng, Yilun Zhu, Amir Zeldes

17:00–17:15 Closing Remarks

