

DiscAnn Program

4. – 5. October 2021

Location: [Festsaal ALTE AULA](#), Münzgasse 30, Universität Tübingen

Website: <https://discannworkshop.github.io>

Organized by the SFB 833, Project A4

Agenda

3. Oktober

From 19:00 Warm up: [Neckarmüller](#)

4. October

9:15 Welcome

9:30 - 10:45 Invited Talk: [Contrast in discourse and in argumentation](#)

Manfred Stede

10:45 - 11:15 Coffee Break

11:15 - 12:15 [Is there less annotator agreement when the discourse relation is underspecified?](#)

Jet Hoek, Merel C.J. Scholman Ted J.M. Sanders

12:15 - 14:00 Lunch Break

14:00 - 15:00 [German parenthetical discourse markers between perception and cognition - A case in point for an explorative approach to corpus data](#)

Regina Ziehe

15:00 - 15:15 Coffee Break

15:15 - 16:15 [Developing an Annotation System for Communicative Functions for a Cross-Layer ASR System](#)

Barbara Schuppler and Anneliese Kelterer

16:15 - 17:15 [Contextual Choice between Synonymous Pairs of Metaphorical and Literal Expressions: An Empirical Study and Novel Dataset to tackle or to address the Question](#)

Prisca Piccirilli and Sabine Schulte im Walde

17:15 - 18:30 Walk through Tübingen

19:30 Workshop Dinner: [Restaurant Mauganeschtle](#)

5. October

9:00 - 10:00	Invited Talk: Attending to more discourse relational phenomena <i>Bonnie Webber (per Zoom)</i>
10:00 - 11:00	Combined discourse representations: Coherence relations and questions under Discussion <i>Arndt Riester, Amalia Canes Nápoles, Jet Hoek</i>
11:00 - 11:30	Coffee Break
11:30 - 12:30	Advancing Neural Question Generation for Formal Pragmatics: Learning when to generate and when to copy <i>Kordula De Kuthy, Madeeswaran Kannan, Haemant Santhi Ponnusamy and Detmar Meurers</i>
12:30 - 13:15	General Discussion
13:15	Lunch and Farewell

Talks and Abstracts

Contrast in discourse and in argumentation

Manfred Stede, Universität Potsdam

Comparing common inventories of discourse relations (PDTB, RST, SDRT) reveals that the field of „Contrast“ is divided in somewhat different ways that do not invite a straightforward mapping. Correspondingly, evidence from the Potsdam Commentary Corpus (PCC) shows significant mismatches between the annotations of „Contrast“ relations in PDTB and RST. This sense of disorder leads me to re-study the meaning of German „contrastive“ connectives and to propose an inventory of functions they can fulfil in discourse. We annotated the PCC connectives with these functions, which helps in categorizing the reasons for the PDTB-RST mismatches. Finally, we examine the role of Contrast (relations, connective functions) specifically in argumentative text that compares and weights different stances toward the topic in question.

Attending to more discourse relational phenomena

Bonnie Webber, University of Edinburgh

Over the years, our focus on a relatively few challenging problems in annotating and recognizing discourse relations has meant that other, possibly equally challenging problems with discourse relations have been ignored.

In particular, although the pandemic has sent many of us into a blue funk and messed with our ability to concentrate, perhaps sharing some of these other discourse relational phenomena might give us something fresh to think about, as we emerge into what will be a very different research world for all of us.

Among the phenomena I want to lay out here are: focus particles and the sense of discourse relations, phrases that seem part of both arguments to a discourse relation, other modifiers of discourse connectives and their effect on the interpretation of discourse relations, and various ways to think about multiple discourse relations.

Is there less annotator agreement when the discourse relation is underspecified?

Jet Hoek, Centre for Language Studies Radboud University Nijmegen

Merel C.J. Scholman Language Science and Technology Saarland University

Ted J.M. Sanders Utrecht Institute of Linguistics OTS Utrecht University

When annotating coherence relations, interannotator agreement tends to be lower on implicit relations than on relations that are explicitly marked by means of a connective or a cue phrase. This paper explores one possible explanation for this: the additional inferencing involved in interpreting implicit relations compared to explicit relations. If this is the main source of disagreements, agreement should be highly related to the specificity of the connective. Using the CCR framework, we annotated relations from TED talks that were marked by a very specific marker, marked by a highly ambiguous connective, or not marked by means of a connective at all. We indeed reached higher inter-annotator agreement on explicit than on implicit relations. However, agreement on underspecified relations was not necessarily in between, which is what would be expected if agreement on implicit relations mainly suffers because annotators have less specific instructions for inferring the relation.

German parenthetical discourse markers between perception and cognition - A case in point for an explorative approach to corpus data

Regina Zieleke, University of Tübingen

This paper discusses the potential of corpus data for the derivation of linguistic research questions. The case in point relates to parenthetical discourse markers involving the perception verb 'to see' (English you see, French tu vois) in French-/English-German parallel corpora. We argue that preliminary results taking into account 'meta-data' (i.e. type of discourse) can point towards a hypothesis on when German equivalents for you see and tu vois are encoded by perception verbs ((wie) du siehst (also)) vis-à-vis cognition verbs (weiß / verstehst du).

Developing an Annotation System for Communicative Functions for a Cross-Layer ASR System

Barbara Schuppler SPSC Laboratory Graz University of Technology, Austria

Anneliese Kelterer Department of Linguistics, University of Graz, Austria

The investigation of conversational speech requires the close collaboration of linguists and speech technologists to develop new modeling techniques that allow the incorporation of various knowledge sources. This paper presents a progress report on the ongoing interdisciplinary project "Cross-layer language models for conversational speech" with a focus on the development of an annotation system for communicative functions. We discuss the requirements of such a system for the application in ASR as well as for the use in phonetic studies of talk-in-interaction, and illustrate emerging issues with the example of turn management.

Contextual Choice between Synonymous Pairs of Metaphorical and Literal Expressions: An Empirical Study and Novel Dataset to tackle or to address the question

Prisca Piccirilli and Sabine Schulte im Walde, Institute for Natural Language Processing, University of Stuttgart

Research on metaphorical language detection and interpretation has produced a large number of resources mainly focusing on metaphoric vs. literal uses of specific expressions, and on metaphor paraphrases. As to our knowledge, however, no existing NLP resource provides a basis for understanding the choice between a synonymous pair of a literal and a metaphorical expression. E.g., why would one favor the use of grasp a term over understand a term in a given context, and does the preceding context prime for one or the other usage? We address these questions and provide an empirical study and a novel resource: Based on 50 pairs of English synonymous literal/metaphorical verb-object and subject-verb expressions in discourse, we asked participants in crowdsourcing experiments to (1) rate the degree of metaphoricity of a discourse, and (2) choose the expression that fits best. Our resource contains a total of 1,000 discourses and is ready to be exploited for computational research on discourse conditions for metaphorical vs. literal expression choices.

Combined discourse representations: Coherence relations and questions under discussion

Arndt Riester, Universität Bielefeld

Amalia Canes Nápoles, Universität zu Köln

Jet Hoek, Centre for Language Studies Radboud University Nijmegen

We analyze a text according to three different discourse theories; CCR, RST and QUD trees. We discuss differences with respect to segmentation and show how coherence relations can be mapped onto a discourse representation based on questions under discussion.

Advancing Neural Question Generation for Formal Pragmatics: Learning when to generate and when to copy

Kordula De Kuthy, Madeeswaran Kannan, Haemant Santhi Ponnusamy and Detmar Meurers, University of Tübingen

Question generation is an interesting challenge for current neural network architectures given that it combines aspects of language meaning and forms in complex ways. Recent work also highlighted the role that questions and question generation can play conceptually in formal pragmatics for linking the information structure of sentences to the discourse structure of texts in so-called Question-under-Discussion (QuD) approaches. In this talk, we show that the sequence to sequence architecture employed in the previous work fails to capture a key property of the task: the required question-answer congruence ensures that the lexical material needed for the question is explicitly given by the answer generated from. Extending the architecture with a pointer component helps overcome

this shortcoming. In addition, we explore the viability of form-based and more fine-grained encodings such as character or subword representations for question generation.

Furthermore, we enrich the models with part-of-speech and semantic role information to improve question phrase generation. The resulting approaches quantitatively advances the state of the art in terms of BLEU scores and question well-formedness, and we qualitatively discuss key linguistic characteristics of the generated question.