

Introduction to Discourse Relation Parsing and Treebanking (DISRPT): 7th Workshop on Rhetorical Structure Theory and Related Formalisms *

Amir Zeldes

Georgetown University
az364@georgetown.edu

Debopam Das

University of Potsdam
ddas@sfu.ca

Erick Galani Maziero

Federal University of Lavras
erick.maziero@ufla.br

Juliano Desiderato Antonio

Universidade Estadual de Maringá
jdantonio@uem.br

Mikel Iruskieta

University of the Basque Country
mikel.iruskieta@ehu.eus

Abstract

This overview summarizes the main contributions of the accepted papers at the 2019 workshop on Discourse Relation Parsing and Treebanking (DISRPT 2019). Co-located with NAACL 2019 in Minneapolis, the workshop's aim was to bring together researchers working on corpus-based and computational approaches to discourse relations. In addition to an invited talk, eighteen papers outlined below were presented, four of which were submitted as part of a shared task on elementary discourse unit segmentation and connective detection.

1 Introduction

Study of coherence relations in frameworks such as RST (Mann and Thompson, 1988), SDRT (Asher and Lascarides, 2003) and PDTB (Mitsakaki et al., 2004), has experienced a revival in the last few years, in English and many other languages (Matthiessen and Teruya, 2015; da Cunha, 2016; Iruskieta et al., 2016; Zeldes, 2016, 2017). Multiple sites are now actively engaged in the development of discourse parsers (Feng and Hirst, 2014; Joty et al., 2015; Surdeanu et al., 2015; Xue et al., 2016; Braud et al., 2017), as a goal in itself, but also for applications such as sentiment analysis, argumentation mining, summarization, question answering, or machine translation evaluation (Benamara et al. 2017; Gerani et al. 2019; Durrett et al. 2016; Peldszus and Stede 2016; Scarton et al. 2016 among many others). At the same time, evaluation of results in discourse parsing has proven complicated (see Morey et al. 2017), and progress in integrating results across discourse treebanking frameworks has been slow.

*Website at <https://sites.google.com/view/dsrpt2019> in conjunction with the Annual Conference of the NAACL 2019 in Minneapolis, MN.

DISRPT 2019 follows a series of biennial events on discourse relation studies, which were initially focused especially on RST, first in Brazil (2007, 2009, 2011, 2013) as part of Brazilian NLP conferences, and then in Spain in 2015 and in 2017, as part of the Spanish NLP conference¹ and INLG 2017.² The 2019 workshop aims to broaden the scope of discussion to include participants and program committee members from different discourse theories (especially, but not limited to, RST, SDRT and PDTB). We encouraged the submission of papers with a computational orientation, resource papers and work on discourse parsing, as well as papers that advance the field with novel theoretical contributions and promote cross-framework fertilization. A major theme and a related shared task on discourse unit identification across formalisms aimed to promote convergence of resources and a joint evaluation of discourse parsing approaches.

Fourteen theoretical and applied papers plus four papers for the shared task were accepted for the DISRPT 2019 workshop. A summary of these papers is provided below.

2 Workshop papers

In the first paper of the proceedings, Shi, Yung and Demberg (Shi et al., 2019) consider implicit discourse relation classification as one of the most challenging and important tasks in discourse parsing, due to the lack of connectives as strong linguistic cues. A principle bottleneck to further improvement is the shortage of training data (ca. $\approx 18k$ instances in the Penn Discourse Treebank (PDTB)). Shi et al. (2019) proposed to acquire additional data by exploiting connectives in transla-

¹<https://sites.google.com/site/workshoprst2015/>.

²<https://sites.google.com/site/workshoprst2017/>.

tion: human translators mark discourse relations which are implicit in the source language explicitly in the translation. Using back-translations of such explicited connectives improves discourse relation parsing performance. This paper addresses the open question of whether the choice of the translation language matters, and whether multiple translations into different languages can be effectively used to improve the quality of the additional data.

Scheffler, Aktaş, Das and Stede (Scheffler et al., 2019) introduce their pilot study applying PDTB-style annotation to Twitter conversations. They present their corpus of 185 Twitter threads and their relational annotation, including an inter-annotator agreement study. They discuss their observations as to how Twitter discourses differ from written news text with respect to discourse connectives and relations. They confirm their hypothesis that discourse relations in written social media conversations are expressed differently than in (news) text. They also find that connective arguments in Twitter often do not appear as full syntactic clauses, and that a few general connectives expressing EXPANSION and CONTINGENCY relations make up the majority of the explicit relations in their data.

Jiang, Yang, Suvarna, Cassula, Zhang and Rose (Jiang et al., 2019) present a package of annotation resources that can be used to apply RST to essays written by students. Furthermore, they highlight the great potential of using RST to provide automated feedback for improving writing quality across genres.

Ferracane, Page, Li and Erk (Ferracane et al., 2019) analyze how well news-trained segmenters perform segmentation in a small-scale medical corpus in English. While they find the expected drop in performance, the nature of the segmentation errors suggests that some problems can be addressed earlier in the pipeline, while others would require expanding the corpus to a trainable size to learn the nuances of the medical domain.

Das (2019) investigates the relationship between the notion of nuclearity as proposed in Rhetorical Structure Theory (RST) and the signalling of coherence relations, examining how mononuclear relations (e.g., ANTITHESIS, CONDITION) and multinuclear relations (e.g., CONTRAST, LIST) are indicated by relational signals, more particularly by discourse markers (e.g., ‘be-

cause’, ‘however’, ‘if’, ‘therefore’). He conducts a corpus study, examining the distribution of either type of relations in the RST Discourse Treebank (Carlson et al., 2002) and the distribution of discourse markers for those relations in the RST Signalling Corpus (Das and Taboada, 2018). The results show that discourse markers are used more often to signal multinuclear relations than mononuclear relations. The findings also suggest a complex relationship between the relation types and syntactic categories of discourse markers (subordinating and coordinating conjunctions).

Potter (2019) discusses the relational status of ATTRIBUTION in RST, which has been a matter of ongoing debate. Although several researchers have weighed in on the topic, and although numerous studies have relied upon attributional structures for their analyses, nothing approaching consensus has emerged. Potter’s paper identifies three basic issues which, he argues, must be resolved to determine the relational status of attributions. These are identified as the Discourse Units Issue, the Nuclearity Issue, and the Relation Identification Issue. These three issues are analyzed from the perspective of classical RST. A finding of this analysis is that the nuclearity and the relational identification of attribution structures are shown to depend on the writer’s intended effect, such that attributional relations cannot be considered as a single relation, but rather as attributional instances of other RST relations.

Bourgonje and Zolotareno (2019) attempt to automatically induce PDTB-style relations from RST trees. They work with a German corpus of news commentary articles, annotated for RST trees and explicit PDTB-style relations, and focus on inducing the implicit relations in an automated way. Preliminary results look promising as a high-precision (but low-recall) way of finding implicit relations where there is no shallow structure annotated at all, but mapping proves more difficult in cases where EDUs and relation arguments overlap, yet do not seem to signal the same relation.

Alkorta, Gojenola and Iruskieta (Alkorta et al., 2019) present the first results on the annotation of the Basque Opinion Corpus using RST, based on the assumption that discourse information is crucial for a better understanding of the text structure. It is also necessary to describe which part of an opinionated text is more relevant to decide how a text span can change the polarity (strengthen or

weaken) of other span by means of coherence relations. Their evaluation results and analysis show the main avenues to improve on a future annotation process. They have also extracted the subjectivity of several rhetorical relations and the results show the effect of sentiment words in relations and the influence of each relation in the semantic orientation value.

Wang, Gyawali, Bruno, Molloy, Evanini and Zechner (Wang et al., 2019) present a paper which aims to model the discourse structure of spontaneous spoken responses within the context of an assessment of English speaking proficiency for non-native speakers. Rhetorical Structure Theory (RST) has been commonly used in the analysis of discourse organization of written texts; however, limited research has been conducted to date on RST annotation and parsing of spoken language, in particular, non-native spontaneous speech. Due to the fact that the measurement of discourse coherence is typically a key metric in human scoring rubrics for assessments of spoken language, they conducted research to obtain RST annotations on non-native spoken responses from a standardized assessment of academic English proficiency. Subsequently, automatic parsers were trained on these annotations to process non-native spontaneous speech. Finally, a set of features were extracted from automatically generated RST trees to evaluate the discourse structure of non-native spontaneous speech, which were then employed to further improve the validity of an automated speech scoring system.

Gessler, Liu and Zeldes (Gessler et al., 2019) present a new system for open-ended discourse relation signal annotation in the framework of Rhetorical Structure Theory (RST), implemented on top of an online tool for RST annotation. The authors discuss existing projects annotating textual signals of discourse relations, which have so far not allowed simultaneously structuring and annotating words signaling hierarchical discourse trees, and demonstrate the design and applications of their interface by extending existing RST annotations in the freely available GUM corpus (Zeldes, 2017).

The paper by Liu (2019) points out that recent research on discourse relations has found that such relations are cued not only by discourse markers (DMs) but also by other textual signals, and that signaling information can be genre-specific. How-

ever, while several corpora exist with discourse relation signaling information such as the Penn Discourse Treebank (PDTB, Prasad et al. 2008 and the Rhetorical Structure Theory Signalling Corpus (RST-SC, Das and Taboada 2017), they all annotate a single text type, specifically the Wall Street Journal (WSJ) section of the Penn Treebank (PTB, Marcus et al. 1993), which is limited to the news domain. Liu's paper adapts signal identification and a signal anchoring scheme (Liu and Zeldes, 2019) to three more genres beyond news, and examines the distribution of signaling devices across relations and text types, providing a taxonomy of indicative signals found in her dataset.

For Iruskieta and Braud (2019), development of discourse parsers to annotate the relational discourse structure of a text is crucial for many downstream tasks. However, most existing studies focus on English, assuming quite a large dataset. Discourse data have been annotated for Basque, but training a system on these data is challenging since the corpus is very small. In their paper, Iruskieta and Braud create the first parser based on RST for Basque and investigate the use of data in another language to improve the performance of a Basque discourse parser. More precisely, they build a monolingual system using the small set of data available and investigate the use of multilingual word embeddings to train a system for Basque using data annotated for another language.

Wang, Kutschbach, Lüdeling and Stede (Wang et al., 2019) present RST-Tace, a tool for automatic comparison and evaluation of RST trees. RST-Tace serves as an implementation of Iruskieta's comparison method (Iruskieta et al., 2015), which allows trees to be compared and evaluated without the influence of decisions at lower levels in a tree in terms of four factors: constituent, attachment point, nuclearity and relation. RST-Tace can be used regardless of the language or the size of rhetorical trees. This tool aims to measure the agreement between two annotators. The result is reflected by F-measure and inter-annotator agreement. Both the comparison table and the result of the evaluation can be obtained automatically.

Shelmanov, Pisarevskaya, Chistova, Toldova, Kobozeva and Smirnov (Shelmanov et al., 2019) present results of the first experimental evaluation of machine learning models trained on Ru-RSTreebank (the first Russian corpus annotated within the RST framework). Various lexical,

quantitative, morphological, and semantic features were used. In rhetorical relation classification, an ensemble CatBoost model with selected features and a linear SVM model provide the best score (macro F1 = 54.67 ± 0.38). The authors discovered that most of the important features for rhetorical relation classification are related to discourse connectives derived from the lexicon of connectives for Russian and from other sources.

3 Shared task

As mentioned above, four papers addressed the shared task activity proposed for the workshop. More detailed information about the DISRPT 2019 shared task, along with quantitative results and system analyses, is provided in a separate report (Zeldes et al., 2019) accompanying these proceedings.

Yu, Zhu, Liu, Liu, Peng, Gong and Zeldes (Yu et al., 2019) present GumDrop, Georgetown University’s entry at the DISRPT 2019 Shared Task on automatic discourse unit segmentation and connective detection. The authors’ approach relies on model stacking, creating a heterogeneous ensemble of classifiers, which feed into a meta-learner for each final task: discourse unit segmentation and connective detection. The system encompasses three trainable component stacks: one for sentence splitting, one for discourse unit segmentation and one for connective detection. The flexibility of each ensemble allows the system to generalize well to datasets of different sizes and with varying levels of homogeneity.

Bourgonje and Schäfer (2019) describe a series of experiments applied to data sets from different languages and genres annotated for coherence relations according to different theoretical frameworks. Specifically, they investigate the feasibility of a unified (theory-neutral) approach to discourse segmentation. The authors apply a Random Forest and an LSTM based approach for all datasets and improve over a simple baseline assuming sentence or clause-like segmentation. Performance however varies considerably depending on language, and more importantly genre, with F-scores ranging from 0.73 to 0.944.

For Iruskieta, Bengoetxea, Salazar and Diaz de Ilarraza (Iruskieta et al., 2019), Elementary Discourse Units (EDUs) are quite similar across different theories. Segmentation is the very first stage on the way of rhetorical annotation. Still, each

annotation project adopted several decisions with consequences not only for the annotation of the relational discourse structure but also at the segmentation stage. In this shared task, the authors have employed pre-trained word embeddings, neural networks (BiLSTM+CRF) to perform the segmentation. They report F1 results for 6 languages: Basque (0.853), English (0.919), French (0.907), German (0.913), Portuguese (0.926) and Spanish (0.868 and 0.769) (for results on more datasets, see the report in Zeldes et al. 2019). Finally, they also pursued an error analysis based on clause typology for Basque and Spanish, in order to understand the performance of the segmenter.

According to Muller, Braud and Morey (Muller et al., 2019), segmentation is the first step in building practical discourse parsers, and is often neglected in discourse parsing studies. The goal is to identify the minimal spans of text to be linked by discourse relations, or to isolate explicit marking of discourse relations. Existing systems on English report F1 scores as high as 0.95, but they generally assume gold sentence boundaries and are restricted to English newswire texts annotated within the RST framework. Their paper presents a generic approach and a system, ToNy, a discourse segmenter developed for the DISRPT shared task where multiple discourse representation schemes, languages and domains are represented. In their experiments, the authors found that a straightforward sequence prediction architecture with pretrained contextual embeddings is sufficient to reach performance levels comparable to existing systems, when separately trained on each corpus. They report performance between 0.81 and 0.96 in F1 score. They also observed that discourse segmentation models only display a moderate generalization capability, even within the same language and discourse representation scheme.

References

- Jon Alkorta, Koldo Gojenola, and Mikel Iruskieta. 2019. Towards discourse annotation and sentiment analysis of the Basque Opinion Corpus. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.
- Farah Benamara, Maite Taboada, and Yannick Mathieu. 2017. Evaluative language beyond bags of

- words: Linguistic insights and computational applications. *Computational Linguistics*, 43(1):201–264.
- Peter Bourgonje and Robin Schäfer. 2019. Multilingual and cross-genre discourse unit segmentation. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Peter Bourgonje and Olha Zolotareno. 2019. Toward cross-theory discourse relation annotation. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Chloé Braud, Maximin Coavoux, and Anders Søgaard. 2017. Cross-lingual RST Discourse Parsing. In *Proceedings of EACL 2017*, pages 292–304, Valencia, Spain.
- L Carlson, D Marcu, and ME Okurowski. 2002. RST Discourse Treebank. Linguistic Data Consortium.
- Iria da Cunha. 2016. towards discourse parsing in Spanish. In *TextLink–Structuring Discourse in Multilingual Europe Second Action Conference Károli Gáspár University of the Reformed Church in Hungary Budapest, 11–14 April, 2016*, page 40.
- Debopam Das. 2019. Nuclearity in RST and signals of coherence relations. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Debopam Das and Maite Taboada. 2017. Signalling of coherence relations in discourse, beyond discourse markers. *Discourse Processes*, pages 1–29.
- Debopam Das and Maite Taboada. 2018. RST Signalling Corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*, 52(1):149–184.
- Greg Durrett, Taylor Berg-Kirkpatrick, and Dan Klein. 2016. Learning-based single-document summarization with compression and anaphoricity constraints. In *Proceedings of ACL 2016*, pages 1998–2008, Berlin.
- Vanessa Wei Feng and Graeme Hirst. 2014. A linear-time bottom-up discourse parser with constraints and post-editing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 511–521.
- Elisa Ferracane, Titan Page, Junyi Jessy Li, and Katrin Erk. 2019. From news to medical: cross-domain discourse segmentation. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Shima Gerani, Giuseppe Carenini, and Raymond Ng. 2019. Modeling content and structure for abstractive review summarization. *Computer Speech & Language*, (53):302–331.
- Luke Gessler, Yang Liu, and Amir Zeldes. 2019. A discourse signal annotation system for RST trees. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Mikel Iruskieta, Kepa Bengoetxea, Aitziber Atutxa, and Arantza Diaz de Ilarraza. 2019. Multilingual segmentation based on neural networks and pre-trained word embeddings. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Mikel Iruskieta and Chlo Braud. 2019. EusDisParser: improving an under-resourced discourse parser with cross-lingual data. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Mikel Iruskieta, Iria Da Cunha, and Maite Taboada. 2015. A qualitative comparison method for rhetorical structures: identifying different discourse structures in multilingual corpora. *Language resources and evaluation*, 49(2):263–309.
- Mikel Iruskieta, Gorka Labaka, and Juliano Desiderato Antonio. 2016. Detecting the central units in two different genres and languages: a preliminary study of Brazilian Portuguese and Basque texts. *Procesamiento del Lenguaje Natural*, (56):65–72.
- Shiyan Jiang, Kexin Yang, Chandrakumari Suvarna, Pooja Casula, Mingtong Zhang, and Carolyn Rose. 2019. Applying Rhetorical Structure Theory to student essays for providing automated writing feedback. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. 2015. Codra: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3):385–435.
- Yang Liu. 2019. Beyond the wall street journal: Anchoring and comparing discourse signals across genres. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Yang Liu and Amir Zeldes. 2019. Discourse relations and signaling information: Anchoring discourse signals in RST-DT. In *Proceedings of the Society for Computation in Linguistics*, volume 2, pages 314–317.
- William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank.
- Christian MIM Matthiessen and Kazuhiro Teruya. 2015. Grammatical realizations of rhetorical relations in different registers. *Word*, 61(3):232–281.

- Eleni Miltsakaki, Rashmi Prasad, Aravind K. Joshi, and Bonnie L. Webber. 2004. The Penn Discourse Treebank. In *Proceedings of LREC 2004*.
- Mathieu Morey, Philippe Muller, and Nicholas Asher. 2017. How much progress have we made on RST discourse parsing? A replication study of recent results on the RST-DT. In *Proceedings of EMNLP 2017*, pages 1319–1324, Copenhagen, Denmark.
- Philippe Muller, Chloé Braud, and Mathieu Morey. 2019. ToNy: Contextual embeddings for accurate multilingual discourse segmentation of full documents. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Andreas Peldszus and Manfred Stede. 2016. Rhetorical structure and argumentation structure in monologue text. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 103–112.
- Andrew Potter. 2019. The rhetorical structure of attribution. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, pages 2961–2968, Marrakesh, Morocco.
- Carolina Scarton, Daniel Beck, Kashif Shah, Karin Sim Smith, and Lucia Specia. 2016. Word embeddings and discourse information for quality estimation. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, volume 2, pages 831–837.
- Tatjana Scheffler, Berfin Akta, Debopam Das, and Manfred Stede. 2019. Annotating shallow discourse relations in Twitter conversations. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Artem Shelmanov, Dina Pisarevskaya, Elena Chistova, Svetlana Toldova, Maria Kobozeva, and Ivan Smirnov. 2019. Towards the data-driven system for rhetorical parsing of Russian texts. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Wei Shi, Frances Yung, and Vera Demberg. 2019. Acquiring annotated data with cross-lingual explicitation for implicit discourse relation classification. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Mihai Surdeanu, Tom Hicks, and Marco Antonio Valenzuela-Escarcega. 2015. Two practical Rhetorical Structure Theory parsers. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Demonstrations*, pages 1–5.
- Xinhao Wang, Binod Gyawali, James V. Bruno, Hillary R. Molloy, Keelan Evanini, and Klaus Zechner. 2019. Using Rhetorical Structure Theory to assess discourse coherence for non-native spontaneous speech. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Atapol Rutherford, Bonnie Webber, Chuan Wang, and Hongmin Wang. 2016. CoNLL 2016 shared task on multilingual shallow discourse parsing. *Proceedings of the CoNLL-16 shared task*, pages 1–19.
- Yue Yu, Yilun Zhu, Yang Liu, Yan Liu, Siyao Peng, Mackenzie Gong, and Amir Zeldes. 2019. GumDrop at the DISRPT2019 shared task: A model stacking approach to discourse unit segmentation and connective detection. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.
- Amir Zeldes. 2016. rstWeb - a browser-based annotation interface for Rhetorical Structure Theory and discourse relations. In *Proceedings of NAACL-HLT 2016 System Demonstrations*, pages 1–5, San Diego, CA.
- Amir Zeldes. 2017. [The GUM corpus: Creating multilayer resources in the classroom](#). *Language Resources and Evaluation*, 51(3):581–612.
- Amir Zeldes, Debopam Das, Erick Galani Maziero, Juliano Desiderato Antonio, and Mikel Iruskieta. 2019. The DISRPT 2019 Shared Task on elementary discourse unit segmentation and connective detection. In *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*, Minneapolis, MN.