

# Towards Unification of Discourse Annotation Frameworks

Yingxue Fu

School of Computer Science

University of St Andrews

KY16 9SX, UK

yf30@st-andrews.ac.uk

## Abstract

Discourse information is difficult to represent and annotate. Among the major frameworks for annotating discourse information, RST, PDTB and SDRT are widely discussed and used, each having its own theoretical foundation and focus. Corpora annotated under different frameworks vary considerably. To make better use of the existing discourse corpora and achieve the possible synergy of different frameworks, it is worthwhile to investigate the systematic relations between different frameworks and devise methods of unifying the frameworks. Although the issue of framework unification has been a topic of discussion for a long time, there is currently no comprehensive approach which considers unifying both discourse structure and discourse relations and evaluates the unified framework intrinsically and extrinsically. We plan to use automatic means for the unification task and evaluate the result with structural complexity and downstream tasks. We will also explore the application of the unified framework in multi-task learning and graphical models.

## 1 Introduction

A text is not a simple collection of isolated sentences. These sentences generally appear in a certain order and are connected with each other through logical or semantic means to form a coherent whole. In recent years, modelling beyond the sentence level is attracting more attention, and different natural language processing (NLP) tasks use discourse-aware models to obtain better performance, such as sentiment analysis (Bhatia et al., 2015), automatic essay scoring (Nadeem et al., 2019), machine translation (Sim Smith, 2017), text summarization (Xu et al., 2020) and so on.

As discourse information typically involves the interaction of different levels of linguistic phenomena, including syntax, semantics, pragmatics and information structure, it is difficult to represent and annotate. Different discourse theories and dis-

course annotation frameworks have been proposed. Accordingly, discourse corpora annotated under different frameworks show considerable variation, and a corpus can be hardly used together with another corpus for natural language processing (NLP) tasks or discourse analysis in linguistics. Discourse parsing is a task of uncovering the underlying structure of text organization, and deep-learning based approaches are used in recent years. However, discourse annotation takes the whole document as the basic unit and is a laborious task. To boost the performance of neural models, we typically need a large amount of data.

Due to the above issues, the unification of discourse annotation frameworks has been a topic of discussion for a long time. Researchers have proposed varied methods to unify discourse relations and debated over whether trees are a good representation of discourse (Egg and Redeker, 2010; Lee et al., 2008; Wolf and Gibson, 2005). However, existing research either focuses on mapping or unifying discourse relations of different frameworks (Bunt and Prasad, 2016; Benamara and Taboada, 2015; Sanders et al., 2018; Demberg et al., 2019), or on finding a common discourse structure (Yi et al., 2021), without giving sufficient attention to the issue of relation mapping. There is still no comprehensive approach that considers unifying both discourse structure and discourse relations.

Another approach to tackling the task is to use multi-task learning so that information from a discourse corpus annotated under one framework can be used to solve a task in another framework, thus achieving synergy between different frameworks. However, existing studies adopting this method (Liu et al., 2016; Braud et al., 2016) do not show significant performance gain by incorporating a part of discourse information from a corpus annotated under a different framework. How to leverage discourse information from different frameworks

remains a challenge.

Discourse information may be used in downstream tasks. [Huang and Kurohashi \(2021\)](#) and [Xu et al. \(2020\)](#) use both coreference relations and discourse relations for text summarization with graph neural networks (GNNs). The ablation study by [Huang and Kurohashi \(2021\)](#) shows that using coreference relations only brings little performance improvement but incorporating discourse relations achieves the highest performance gain. While different kinds of discourse information can be used, how to encode different types of discourse information to improve discourse-awareness of neural models is a topic that merits further investigation.

The above challenges motivate our research on unifying different discourse annotation frameworks. We will focus on the following research questions:

**RQ1:** Which structure can be used to represent discourse in the unified framework?

**RQ2:** What properties of different frameworks should be kept and what properties should be ignored in the unification?

**RQ3:** How can entity-based models and lexical-based models be incorporated into the unified framework?

**RQ4:** How can the unified framework be evaluated?

The first three questions are closely related to each other. Automatic means will be used, although we do not preclude semi-automatic means, as exemplified by [Yi et al. \(2021\)](#). We will start with the methods suggested by existing research and focus on the challenges of incorporating different kinds of discourse information in multi-task learning and graphical models.

The unified framework can be used for the following purposes:

1. A corpus annotated under one framework can be used jointly with another corpus annotated under a different framework to augment data, for developing discourse parsing models or for discourse analysis. We can train a discourse parser on a corpus annotated under one framework and compare its performance with the case when it is trained on augmented data, similar to [Yi et al. \(2021\)](#).
2. Each framework has its own theoretical foundation and focus. A unified framework may have the potential of combining the strengths of different frameworks. Experiments can

be done with multi-task learning so that discourse parsing tasks of different frameworks can be solved jointly. We can also investigate how to enable GNNs to better capture different kinds of discourse information.

3. A unified framework may provide a common ground for exploring the relations of different frameworks and validating annotation consistency of a corpus. We can perform comparative corpus analysis and obtain new understanding of how information expressed in one framework is conveyed in another framework, thus validating corpus annotation consistency and finding some clues for solving problems in a framework with signals from another framework, similar to [Poláková et al. \(2017\)](#) and [Bourgonje and Zolotarev \(2019\)](#).

## 2 Related Work

### 2.1 An Overview of Discourse Theories

A number of discourse theories have been proposed. The theory by [Grosz and Sidner \(1986\)](#) is one of those earlier few whose linguistic claims about discourse are also computationally significant ([Mann and Thompson, 1987](#)). With this theory, it is believed that discourse structure is composed of three separated but interrelated components: linguistic structure, intentional structure and attentional structure. The linguistic structure focuses on cue phrases and discourse segmentation. The intentional structure mainly deals with why a discourse is performed (discourse purpose) and how a segment contributes to the overall discourse purpose (discourse segment purpose). The attentional structure is not related to the discourse participants, and it records the objects, properties and relations that are salient at each point in discourse. These three aspects capture discourse phenomena in a systematic way, and other discourse theories may be related to this theory in some way. For instance, the Centering Theory ([Grosz et al., 1995](#)) and the entity-grid model ([Barzilay and Lapata, 2008](#)) focus on the attentional structure, and the Rhetorical Structure Theory (RST) ([Mann and Thompson, 1988](#)) focuses on the intentional structure.

The theory proposed by [Halliday and Hasan \(1976\)](#) studies how various lexical means are used to achieve cohesion, these lexical means including reference, substitution, ellipsis, lexical cohesion and conjunction. Cohesion realized through the

first four lexical means is in essence anaphoric dependency and conjunction is the only source of discourse relation under this theory (Webber, 2006).

The other discourse theories can be divided into two broad types: relation-based discourse theories and entity-based discourse theories (Jurafsky and Martin, 2018). The former studies how coherence is achieved with discourse relations and the latter focuses on local coherence achieved through shift of focus, which abstracts a text into a set of entity transition sequences (Barzilay and Lapata, 2008).

RST is one of the most influential relation-based discourse theories. The RST Discourse Treebank (RST-DT) (Carlson et al., 2001) is annotated based on this theory. In the RST framework, discourse can be represented by a tree structure whose leaves are Elementary Discourse Units (EDUs), typically clauses, and whose non-terminals are adjacent spans linked by discourse relations. The discourse relations can be symmetric or asymmetric, the former being characterized by equally important spans connected in parallel, and the latter typically having a nucleus and a satellite, which are assigned based on their importance in conveying the intended effects. An RST tree is built recursively by connecting the adjacent discourse units, forming a hierarchical structure covering the whole text. An example of RST discourse trees can be seen in Figure 1.

Another influential framework is the Penn Discourse Treebank (PDTB) framework, which is represented by the Penn Discourse Treebank (Prasad et al., 2008, 2018). Unlike the RST framework, the PDTB framework does not aim at achieving complete annotation of the text but focuses on local discourse relations anchored by structural connectives or discourse adverbials. When there are no explicit connectives, the annotators will read adjacent sentences and decide if a connective can be inserted to express the relation. The annotation is not committed to a specific structure at the higher level. PDTB 3.0 adopts a three-layer sense hierarchy, including four general categories called classes at the highest level, the middle layer being more specific divisions, which are called types, and the lowest layer containing directionality of the arguments, called subtypes. An example of the PDTB-style annotation is shown as follows (Prasad et al., 2019):

*The Soviet insisted that aircraft be brought into the talks,*(implicit=but){arg2-as-denier} **then ar-**

**gued for exempting some 4,000 Russian planes because they are solely defensive.**

The first argument is shown in italics and the second argument is shown in bold font for distinction. As the discourse relation is implicit, the annotator adds a connective that is considered to be suitable for the context.

The Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003) is based on the Discourse Representation Theory (Kamp and Reyle, 1993), with discourse relations added, and discourse structure is represented with directed acyclic graphs (DAGs). Elementary discourse units may be combined recursively to form a complex discourse unit (CDU), which can be linked with another EDU or CDU (Asher et al., 2017). The set of discourse relations developed in this framework overlap partly with those in the RST framework but some are motivated from pragmatic and semantic considerations. In Asher and Lascarides (2003), a precise dynamic semantic interpretation of the rhetorical relations is defined. An example of discourse representation in the SDRT framework is shown in Figure 2, which illustrates that the SDRT framework provides full annotation, similar to the RST framework, and it assumes a hierarchical structure of text organization. The vertical arrow-headed lines represent subordinate relations, and the horizontal lines represent coordinate relations. The textual units in solid-line boxes are EDUs and  $\pi'$  and  $\pi''$  represent CDUs. The relations are shown in bold.

## 2.2 Research on Relations between Different Frameworks

The correlation between different frameworks has been a topic of interest for a long time. Some studies explore how different frameworks are related, either in discourse structures or in relation sets. Some studies take a step further and try to map the relation sets of different frameworks.

### 2.2.1 Comparison/unification of discourse structures of different frameworks

Stede et al. (2016) investigate the relations between RST, SDRT and argumentation structure. For the purpose of comparing the three layers of annotation, the EDU segmentation in RST and SDRT is harmonized, and an “argumentatively empty” JOIN relation is introduced to address the issue that the basic unit of the argumentation structure is coarser than the other two layers. The annotations are con-

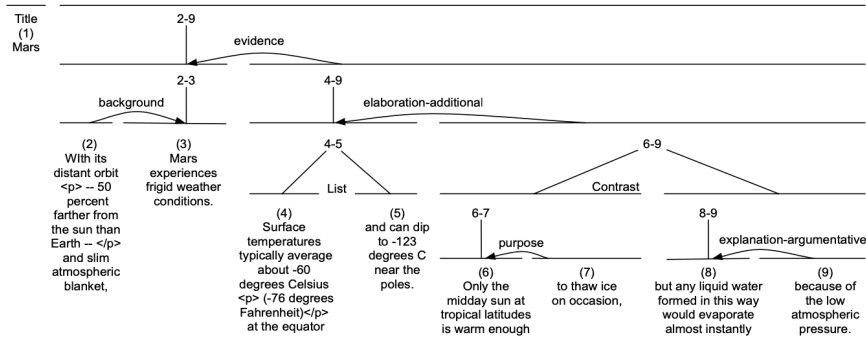


Figure 1: An RST discourse tree, originally from [Marcu \(2000a\)](#).

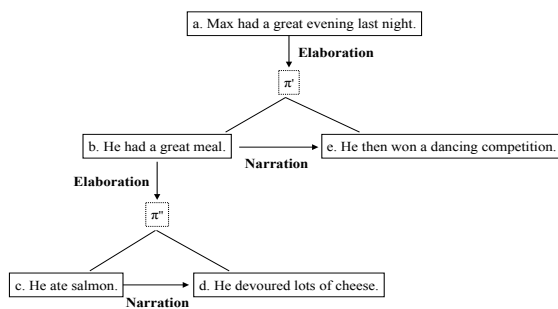


Figure 2: SDRT representation of the text *a. Max had a great evening last night. b. He had a great meal. c. He ate salmon. d. He devoured lots of cheese. e. He then won a dancing competition.* The example is taken from [Asher and Lascarides \(2003\)](#).

verted to a common dependency graph format for calculating correlations. To transform RST trees to the dependency structure, the method introduced by [Li et al. \(2014\)](#) is used. The RST trees are binarized and the left-most EDU is treated as the head. In the transformation of the SDRT graphs to the dependency structure, the CDUs are simplified by a *head replacement strategy*. The authors compare the dependency graphs in terms of common edges and common connected components. The relations of the argumentation structure are compared with those of RST and SDRT, respectively, through a co-occurrence matrix. Their research shows the systematic relations between the argumentation structure and the two discourse annotation frameworks. The purpose is to investigate if discourse parsing can contribute to automatic argumentation analysis. The authors exclude the PDTB framework because it does not provide full discourse annotation.

[Yi et al. \(2021\)](#) try to unify two Chinese discourse corpora annotated under the PDTB framework and the RST framework, respectively, with

a corpus annotated under the dependency framework. They use semi-automatic means to transform the corpora to the discourse dependency structure which is presented in [Li et al. \(2014\)](#). Their work shows that the major difficulty is the transformation from the PDTB framework to the discourse dependency structure, which requires re-segmenting texts and complementing some relations to construct complete dependency trees. They use the same method as [Stede et al. \(2016\)](#) to transform the RST trees to the dependency structure. Details about relation mapping across the frameworks are not given.

## 2.2.2 Comparison/unification of discourse relations of different frameworks

The methods of mapping discourse relations of different frameworks presented by [Scheffler and Stede \(2016\)](#), [Demberg et al. \(2019\)](#) and [Bourgonje and Zolotareno \(2019\)](#) are empirically grounded. The main approach is to make use of the same texts annotated under different frameworks.

[Scheffler and Stede \(2016\)](#) focus on mapping between explicit PDTB discourse connectives and RST rhetorical relations. The Potsdam Commentary Corpus ([Stede and Neumann, 2014](#)), which contains annotations under both frameworks, is used. It is found that the majority of the PDTB connectives in the corpus match exactly one RST relation and mismatches are caused by different segment definitions and focuses, i.e., PDTB focuses on local/lexicalized relations and RST focuses on global structural relations.

As the Potsdam Commentary Corpus only contains explicit relations under the PDTB framework, [Bourgonje and Zolotareno \(2019\)](#) try to induce implicit relations from the corresponding RST annotation. Since RST trees are hierarchical and the PDTB annotation is shallow, RST relations that connect complex spans are discarded.



Moreover, because the arguments of explicit and implicit relations under the PDTB framework are determined based on different criteria, only RST relations that are signalled explicitly are considered in the experiment. It is shown that differences in segmentation and partially overlapping relations pose challenges for the task.

Demberg et al. (2019) propose a method of mapping RST and PDTB relations. Since the number of PDTB relations is much smaller than that of RST relations for the same text, the PDTB relations are used as the starting point for the mapping. They aim for mapping as many relations as possible while making sure that the relations connect the same segments. Six cases are identified: direct mapping, which is the easiest case; when PDTB arguments are non-adjacent, the Strong Compositionality hypothesis (Marcu, 2000b) (i.e., if a relation holds between two textual spans, that relation also holds between the most important units of the constituent spans) is used to check if there is a match when the complex span of an RST relation is traced along the nucleus path to its nucleus EDU; in the case of multi-nuclear relations, it is checked if a PDTB argument can be traced to the nucleus of the RST relation along the nucleus path; the mismatch caused by different segmentation granularity is considered innately unalignable and discarded; centrally embedded EDUs in RST-DT are treated as a whole and compared with an argument of the PDTB relation; and the PDTB ENTREL relation is included to test its correlation with some RST relations that tend to be associated with cohesion.

Other studies are more theoretical. Hovy (1990) is the first to attempt to unify discourse relations proposed by researchers from different areas and suggests adopting a hierarchy of relations, with the top level being more general (from the functional perspective: ideational, interpersonal and textual) and putting no restrictions on adding fine-grained relations, as long as they can be subsumed under existing taxonomy. The number of researchers who propose a specific relation is taken as a vote of confidence of the relation in the taxonomy. The study serves as a starting point for research in this direction. There are a few other proposals for unifying discourse relations of different frameworks to facilitate cross-framework discourse analysis, including: introducing a hierarchy of discourse relations, similar to Hovy (1990), where the top level is general and fixed, and the lowest level is

more specific and allows variations based on genre and language (Benamara and Taboada, 2015), finding some dimensions based on cognitive evidence where the relations can be compared with each other and re-grouped (Sanders et al., 2018), and formulating a set of core relations that are shared by existing frameworks but are open and extensible in use, with the outcome being ISO-DR-Core (Bunt and Prasad, 2016). When the PDTB sense hierarchy is mapped to the ISO-DR-Core, it is found that the directionality of relations cannot be captured by the existing ISO-DR-Core relations and it remains a question whether to extend the ISO-DR-Core relations or to redefine the PDTB relations so that the directionality of arguments can be captured (Prasad et al., 2018).

### 3 Research Plan

RST-DT is annotated on texts from the Penn Treebank (Marcus et al., 1993) that have also been annotated in PDTB. The texts are formally written Wall Street Journal articles. The English corpora annotated under the SDRT framework, i.e., the STAC corpus (Asher et al., 2016) and the Molweni corpus (Li et al., 2020), are created for analyzing multi-party dialogues, making it difficult to be used together with the other two corpora. Therefore, in addition to RST-DT and PDTB 3.0, we will use the ANNODIS corpus (Péry-Woodley et al., 2009), which consists of formally written French texts. We will first translate the texts into English with an MT system and then manually check the translated texts to reduce errors.

In the following, the research questions and the approach in our plan will be discussed. These questions are closely related to each other and the research on one question is likely to influence how the other questions should be addressed. They are presented separately just for easier description.

**RQ1:** Which structure can be used to represent discourse in the unified framework?

Although there is a lack of consensus on how to represent discourse structure, in a number of studies, the dependency structure is taken as a common structure that the other structures can be converted to (Muller et al., 2012; Hirao et al., 2013; Venant et al., 2013; Li et al., 2014; Yoshida et al., 2014; Stede et al., 2016; Morey et al., 2018; Yi et al., 2021). This choice is mainly inspired by research in the field of syntax, where the dependency grammar is better studied and its computational and

representational properties are well-understood<sup>1</sup>. The research by Venant et al. (2013) provides a common language for comparing discourse structures of different formalisms, which is used in the transformation procedure presented by Stede et al. (2016). Another possibility is the constrained directed acyclic graph introduced by Danlos (2004). While Venant et al. (2013) focus on the expressivity of different structures, the constrained DAG is motivated from the perspective of strong generative capacity (Danlos, 2008). Although neither of the studies deals with the PDTB framework, since they are both semantically driven, we believe it is possible to deal with the PDTB framework using either of the two structures. We will start with the investigation of the two structures.

Another issue is how to maintain one-to-one correspondence during the transformation of the original structure and the unified structure back and forth. As indicated by Stede et al. (2016), the transformation from the RST or SDRT structures into dependency structures always produces the same structure, but going back to the initial RST or SDRT structures is ambiguous. Morey et al. (2018) introduces head-ordered dependency trees in syntactic parsing (Fernández-González and Martins, 2015) to reduce the ambiguity. We may start with a similar method.

As is clear from Section 2, using the dependency structure as a common ground for studying the relations between different frameworks is not new in existing literature, but comparing the RST, PDTB and SDRT frameworks with this method has not yet been done. This approach will be our starting point, and the suitability of the dependency structure in representing discourse will be investigated empirically. The SciDTB corpus (Yang and Li, 2018), which is annotated under the dependency framework, will be used for this purpose.

**RQ2:**<sup>2</sup> What properties of different frameworks should be kept and what properties should be ignored in the unification?

We present a non-exhaustive list of properties, which we consider to have considerable influence on the unified discourse structure.

1. Nuclearity: Marcu (1996) uses the nuclearity principle as the foundation for a formal treatment of compositionality in RST, which

means that two adjacent spans can be joined into a larger span by a rhetorical relation if and only if the relation holds between the most salient units of those spans. This assumption is criticized by Stede (2008). The remedy provided by Stede (2008) is to separate different levels of discourse information, which is in line with the suggestions in Knott et al. (2000) and Moore and Pollack (1992). Our strategy is to keep this property in the initial stage of experimentation. The existing methods for transforming RST trees to dependency structure (Hirao et al., 2013; Li et al., 2014) rely heavily on the nuclearity principle and we will use these methods in the transformation and see what kinds of problems this procedure will cause, particularly with respect to the PDTB framework, which does not enforce a hierarchical structure for complete coverage of the text.

2. Sentence-boundedness: The RST framework does not enforce well-formed discourse subtrees for each sentence. However, it is found that 95% of the discourse parse trees in RST-DT have well-formed sub-trees at the sentence level (Soricut and Marcu, 2003). For the PDTB framework, there is no restriction on how far an argument can be from its corresponding connective: it can be in the same sentence as the connective, in the sentence immediately preceding that of the connective, or in some non-adjacent sentence (Prasad et al., 2006). Moreover, the arguments are determined based on the *Minimality Principle*, which means that clauses and/or sentences that are minimally required for the interpretation of the relation should be included in the argument, and other spans that are relevant but not necessary can be annotated as supplementary information, which is labeled depending on which argument it is supplementary to (Prasad et al., 2008). The SDRT framework developed in Asher and Lascarides (2003) does not specify the basic discourse unit, but in the annotation of the ANNODIS corpus, EDU segmentation follows similar principles as RST-DT. The formation of CDU and the attachment of relations are where SDRT differs significantly from RST. A segment can be attached to another segment from the same sentence, the same paragraph or a larger context,

<sup>1</sup>In communication with Bonnie Webber, January, 2022.

<sup>2</sup>In communication with Bonnie Webber, January, 2022. We thank her for pointing out this aspect.

and by one or possibly more relations. A CDU can be of any size and can have segments that are far apart in the text, and relations may be annotated within the CDU<sup>3</sup>.

The differences in the criteria on location and extent for basic discourse unit identification and relation labeling of the RST framework and the PDTB framework may be partly attributed to different annotation procedures. In RST, EDU segmentation is performed first and EDU linking and relation labelling are performed later. The balance between consistency and granularity is the major concern behind the strategy for EDU segmentation (Carlson et al., 2001). In contrast, in PDTB, the connectives are identified first, and their arguments are determined afterwards. Semantic relatedness is given greater weight and the location and extent of the arguments can be determined more flexibly. On the whole, neither SDRT nor PDTB shows any tendency of sentence-boundedness. We will investigate to what extent the tendency of sentence-boundedness complicates the unification and what the consequences are if entity-based models and lexical-based models are incorporated.

3. Multi-sense annotation: As shown above, SDRT and PDTB allow multi-sense annotation while RST only allows one relation to be labeled. The single-sense constraint actually gives rise to ambiguity because of the multi-faceted nature of local coherence (Stede, 2008). For the unification task, we assume that multi-sense annotation is useful. However, we agree with the view mentioned in Stede (2008) that incrementally adding more relations as phenomena are being recognized is not a promising direction. There are two possible approaches: one is to separate different dimensions of discourse information (Stede, 2008) and the other is to represent different kinds of discourse information simultaneously, similar to the approach adopted in Knott et al. (2000). While multi-level annotation may reveal the interaction between discourse and other linguistic phe-

nomena, it is less helpful for developing a discourse parser and requires more efforts in annotation. The second approach may be conducive to computationally cheaper discourse processing when proper constraints are introduced.

**RQ3:** How can entity-based models and lexical-based models be incorporated into the unified framework?

The PDTB framework believes that lexical-based discourse relations are associated with anaphoric dependency, which is anchored by discourse adverbials (Webber et al., 2003) and annotated as a type of explicit relations. As for entity-based relations, PDTB uses the ENTREL label to annotate this type of relations when neither explicit nor implicit relations can be identified and only entity-based coherence relations are present. In the RST framework, the ELABORATION relation is actually a relation between entities. However, it is encoded in the same way as the other relations between propositions, which bedevils the framework (Knott et al., 2000). Further empirical studies may be needed to identify how different frameworks represent these different kinds of discourse information. The main challenge is to use a relatively simple structure to represent different types of discourse information while keeping the complexity relatively low.

**RQ4:** How can the unified framework be evaluated?

We will use intrinsic evaluation to assess the complexity of the discourse structure.

Extrinsic evaluation will be used to assess the effectiveness of the unified framework. The downstream tasks in the extrinsic evaluation include text summarization and document discrimination, which are two typical tasks for evaluating discourse models. The document discrimination task asks a score of coherence to be assigned to a document. The originally written document is considered to be the most coherent, and with more permutations, the document becomes less coherent. For comparison with previous studies, we will use the CNN and Dailymail dataset (Hermann et al., 2015) for the text summarization task, and use the method and dataset<sup>4</sup> in Shen et al. (2021) to control the degree of coherence for the document discrimination task.

<sup>3</sup>See section 3 of the ANNODIS annotation manual, available through <http://w3.erss.univ-tlse2.fr/textes/publications/CarnetsGrammaire/carnGram21.pdf>

<sup>4</sup>[https://github.com/AiliAili/Coherence\\_Modelling](https://github.com/AiliAili/Coherence_Modelling)

Previous studies that use multi-task learning and GNNs to encode different types of discourse information will be re-investigated to test the effectiveness of the unified framework.

As we may have to ignore some properties, we will examine what might be lost with the unified framework.

## 4 Conclusion

We propose to unify the RST, PDTB and SDRT frameworks, which may enable discourse corpora annotated under different frameworks to be used jointly and achieve the potential synergy of different frameworks. The major challenges include determining which structure to use in the unified framework, choosing what properties to keep and what to ignore, and incorporating entity-based models and lexical-based models into the unified framework. We will start with existing research and try to find a computationally less expensive way for the task. Extensive experiments will be conducted to investigate how effective the unified framework is and how it can be used. An empirical evaluation of what might be lost through the unification will be performed.

## 5 Acknowledgements

We thank Bonnie Webber for valuable feedback that greatly shaped the work. We are grateful to the anonymous reviewers for detailed and insightful comments that improved the work considerably, and Mark-Jan Nederhof for proof-reading the manuscript. The author is funded by University of St Andrews-China Scholarship Council joint scholarship (NO.202008300012).

## 6 Ethical Considerations and Limitations

The corpora are used in compliance with the licence requirements:

The ANNODIS corpus is available under Creative Commons By-NC-SA 3.0.

RST-DT is distributed on Linguistic Data Consortium:

Carlson, Lynn, Daniel Marcu, and Mary Ellen Okurowski. RST Discourse Treebank LDC2002T07. Web Download. Philadelphia: Linguistic Data Consortium, 2002.

PDTB 3.0 is also distributed on Linguistic Data Consortium:

Prasad, Rashmi, et al. Penn Discourse Treebank Version 3.0 LDC2019T05. Web Download.

Philadelphia: Linguistic Data Consortium, 2019.

**Bender Rule** English is the language studied in this work.

## References

- Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. [Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).
- Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.
- Nicholas Asher, Philippe Muller, Myriam Bras, Lydia Mai Ho-Dac, Farah Benamara, Stergos Afantenos, and Laure Vieu. 2017. [ANNODIS and related projects: case studies on the annotation of discourse structure](#). In *Handbook of Linguistic Annotation*, pages 1241–1264. Springer.
- Regina Barzilay and Mirella Lapata. 2008. [Modeling local coherence: An entity-based approach](#). *Computational Linguistics*, 34(1):1–34.
- Farah Benamara and Maite Taboada. 2015. [Mapping different rhetorical relation annotations: A proposal](#). In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*, pages 147–152, Denver, Colorado. Association for Computational Linguistics.
- Parminder Bhatia, Yangfeng Ji, and Jacob Eisenstein. 2015. [Better document-level sentiment analysis from RST discourse parsing](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2212–2218, Lisbon, Portugal. Association for Computational Linguistics.
- Peter Bourgonje and Olha Zolotareno. 2019. [Toward cross-theory discourse relation annotation](#). In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 7–11, Minneapolis, MN. Association for Computational Linguistics.
- Chloé Braud, Barbara Plank, and Anders Søgaard. 2016. [Multi-view and multi-task training of RST discourse parsers](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1903–1913, Osaka, Japan. The COLING 2016 Organizing Committee.
- Harry Bunt and Rashmi Prasad. 2016. [ISO DR-Core \(ISO 24617-8\): Core concepts for the annotation of discourse relations](#). In *Proceedings 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)*, pages 45–54.



- Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2001. [Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory](#). In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*.
- Laurence Danlos. 2004. [Discourse dependency structures as constrained DAGs](#). In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, pages 127–135, Cambridge, Massachusetts, USA. Association for Computational Linguistics.
- Laurence Danlos. 2008. Strong generative capacity of rst, sdr and discourse dependency dags. In *Constraints in discourse*. Citeseer.
- Vera Demberg, Merel CJ Scholman, and Fatemeh Torabi Asr. 2019. [How compatible are our discourse annotation frameworks? insights from mapping rst-dt and pdtb annotations](#). *Dialogue & Discourse*, 10(1):87–135.
- Markus Egg and Gisela Redeker. 2010. [How complex is discourse structure?](#) In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Daniel Fernández-González and André F. T. Martins. 2015. [Parsing as reduction](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1523–1533, Beijing, China. Association for Computational Linguistics.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. [Centering: A framework for modeling the local coherence of discourse](#). *Computational Linguistics*, 21(2):203–225.
- Barbara J. Grosz and Candace L. Sidner. 1986. [Attention, intentions, and the structure of discourse](#). *Computational Linguistics*, 12(3):175–204.
- Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. 1976. *Cohesion in English (1st ed.)*. Routledge.
- Karl Moritz Hermann, Tomáš Kočický, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. [Teaching machines to read and comprehend](#). In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15*, page 1693–1701, Cambridge, MA, USA. MIT Press.
- Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. 2013. [Single-document summarization as a tree knapsack problem](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1515–1520, Seattle, Washington, USA. Association for Computational Linguistics.
- Eduard H. Hovy. 1990. [Parsimonious and profligate approaches to the question of discourse structure relations](#). In *Proceedings of the Fifth International Workshop on Natural Language Generation*, Linden Hall Conference Center, Dawson, Pennsylvania. Association for Computational Linguistics.
- Yin Jou Huang and Sadao Kurohashi. 2021. [Extractive summarization considering discourse and coreference relations based on heterogeneous graph](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3046–3052, Online. Association for Computational Linguistics.
- Daniel Jurafsky and James H Martin. 2018. Speech and language processing (draft). *preparation [cited 2022 Jan 3]*, Available from: <https://web.stanford.edu/~jurafsky/slp3>.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Publishers.
- Alistair Knott, Jon Oberlander, Michael O'Donnell, and Chris Mellish. 2000. [Beyond elaboration: The interaction of relations and focus in coherent text](#). In *Text Representation: Linguistic and Psycholinguistic Aspects, chapter 7*, pages 181–196. John Benjamins.
- Alan Lee, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2008. Departures from tree structures in discourse: Shared arguments in the penn discourse treebank. In *Proceedings of the constraints in discourse iii workshop*, pages 61–68.
- Jiaqi Li, Ming Liu, Min-Yen Kan, Zihao Zheng, Zekun Wang, Wenqiang Lei, Ting Liu, and Bing Qin. 2020. [Molweni: A challenge multiparty dialogues-based machine reading comprehension dataset with discourse structure](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2642–2652, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Sujian Li, Liang Wang, Ziqiang Cao, and Wenjie Li. 2014. [Text-level discourse dependency parsing](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 25–35, Baltimore, Maryland. Association for Computational Linguistics.
- Yang Liu, Sujian Li, Xiaodong Zhang, and Zhifang Sui. 2016. [Implicit discourse relation classification via multi-task neural networks](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- William C Mann and Sandra A Thompson. 1987. *Rhetorical structure theory: A theory of text organization*. University of Southern California, Information Sciences Institute Los Angeles.

- William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.
- Daniel Marcu. 1996. [Building up rhetorical structure trees](#). In *Proceedings of the National Conference on Artificial Intelligence*, pages 1069–1074.
- Daniel Marcu. 2000a. [The rhetorical parsing of unrestricted texts: a surface-based approach](#). *Computational Linguistics*, 26(3):395–448.
- Daniel Marcu. 2000b. *The theory and practice of discourse parsing and summarization*. MIT press.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. [Building a large annotated corpus of English: The Penn Treebank](#). *Computational Linguistics*, 19(2):313–330.
- Johanna D. Moore and Martha E. Pollack. 1992. [A problem for RST: The need for multi-level discourse analysis](#). *Computational Linguistics*, 18(4):537–544.
- Mathieu Morey, Philippe Muller, and Nicholas Asher. 2018. [A dependency perspective on RST discourse parsing and evaluation](#). *Computational Linguistics*, 44(2):197–235.
- Philippe Muller, Stergos Afantenos, Pascal Denis, and Nicholas Asher. 2012. [Constrained decoding for text-level discourse parsing](#). In *Proceedings of COLING 2012*, pages 1883–1900, Mumbai, India. The COLING 2012 Organizing Committee.
- Farah Nadeem, Huy Nguyen, Yang Liu, and Mari Ostendorf. 2019. [Automated essay scoring with discourse-aware neural models](#). In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 484–493, Florence, Italy. Association for Computational Linguistics.
- Marie-Paule Péry-Woodley, Nicholas Asher, Patrice Enjalbert, Farah Benamara, Myriam Bras, Cécile Fabre, Stéphane Ferrari, Lydia-Mai Ho-Dac, Anne Le Draoulec, Yann Mathet, Philippe Muller, Laurent Prévot, Josette Rebeyrolle, Ludovic Tanguy, Marianne Vergez-Couret, Laure Vieu, and Antoine Widlöcher. 2009. [ANNODIS: une approche outillée de l’annotation de structures discursives](#). In *Actes de la 16ème conférence sur le Traitement Automatique des Langues Naturelles. Articles courts*, pages 41–46, Senlis, France. ATALA.
- Lucie Poláková, Jirí Mírovský, and Pavlína Synková. 2017. Signalling implicit relations: A PDTB-RST comparison. *Dialogue & Discourse*, 8(2):225–248.
- R. Prasad, Eleni Miltsakaki, Nikhil Dinesh, Alan Lee, Aravind K. Joshi, Livio Robaldo, and Bonnie Lynn Webber. 2006. [The Penn Discourse Treebank 2.0 annotation manual](#).
- 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 Sixth International Conference on Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Rashmi Prasad, Bonnie Webber, and Alan Lee. 2018. [Discourse annotation in the PDTB: The next generation](#). In *Proceedings 14th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 87–97, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Rashmi Prasad, Bonnie Webber, Alan Lee, and Aravind Joshi. 2019. [Penn Discourse Treebank Version 3.0 LDC2019T05](#).
- Ted JM Sanders, Vera Demberg, Jet Hoek, Merel CJ Scholman, Fatemeh Torabi Asr, Sandrine Zufferey, and Jacqueline Evers-Vermeul. 2018. [Unifying dimensions in coherence relations: How various annotation frameworks are related](#). *Corpus Linguistics and Linguistic Theory*.
- Tatjana Scheffler and Manfred Stede. 2016. [Mapping PDTB-style connective annotation to RST-style discourse annotation](#). In *Proceedings of the 13th Conference on Natural Language Processing*, pages 242–247.
- Aili Shen, Meladel Mistica, Bahar Salehi, Hang Li, Timothy Baldwin, and Jianzhong Qi. 2021. [Evaluating document coherence modeling](#). *Transactions of the Association for Computational Linguistics*, 9:621–640.
- Karin Sim Smith. 2017. [On integrating discourse in machine translation](#). In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 110–121, Copenhagen, Denmark. Association for Computational Linguistics.
- Radu Soricut and Daniel Marcu. 2003. [Sentence level discourse parsing using syntactic and lexical information](#). In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 228–235.
- Manfred Stede. 2008. Disentangling nuclearity. ‘Subordination’ versus ‘Coordination’ in Sentence and Text: A cross-linguistic perspective, 98:33.
- Manfred Stede, Stergos Afantenos, Andreas Peldszus, Nicholas Asher, and Jérémy Perret. 2016. [Parallel discourse annotations on a corpus of short texts](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 1051–1058, Portorož, Slovenia. European Language Resources Association (ELRA).
- Manfred Stede and Arne Neumann. 2014. [Potsdam commentary corpus 2.0: Annotation for discourse research](#). In *Proceedings of the Ninth International*

*Conference on Language Resources and Evaluation (LREC'14)*, pages 925–929, Reykjavik, Iceland. European Language Resources Association (ELRA).

Antoine Venant, Nicholas Asher, Philippe Muller, Pascal Denis, and Stergos Afantenos. 2013. [Expressivity and comparison of models of discourse structure](#). In *Proceedings of the SIGDIAL 2013 Conference*, pages 2–11, Metz, France. Association for Computational Linguistics.

Bonnie Webber. 2006. [Accounting for discourse relations: constituency and dependency](#). *Intelligent linguistic architectures*, pages 339–360.

Bonnie Webber, Matthew Stone, Aravind Joshi, and Alistair Knott. 2003. [Anaphora and discourse structure](#). *Computational Linguistics*, 29(4):545–587.

Florian Wolf and Edward Gibson. 2005. [Representing discourse coherence: A corpus-based study](#). *Computational Linguistics*, 31(2):249–287.

Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. [Discourse-aware neural extractive text summarization](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5021–5031, Online. Association for Computational Linguistics.

An Yang and Sujian Li. 2018. [SciDTB: Discourse dependency TreeBank for scientific abstracts](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 444–449, Melbourne, Australia. Association for Computational Linguistics.

Cheng Yi, Li Sujian, and Li Yueyuan. 2021. [Unifying discourse resources with dependency framework](#). In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1058–1065, Huhhot, China. Chinese Information Processing Society of China.

Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata. 2014. [Dependency-based discourse parser for single-document summarization](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1834–1839, Doha, Qatar. Association for Computational Linguistics.