Applying Rhetorical Structure Theory to Student Essays for Providing Automated Writing Feedback

Shiyan Jiang, Kexin Yang, Chandrakumari Suvarna,
Pooja Casula, Mingtong Zhang, Carolyn Penstein Rosé
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA
{shiyanj,kexiny, csuvarna,
pcasula, mingtonz, cprose,}@andrew.cmu.edu

Abstract

We present a package of annotation resources, including annotation guideline, flowchart, and an Intelligent Tutoring System for training human annotators. These resources can be used to apply Rhetorical Structure Theory (RST) to essays written by students in K-12 schools. Furthermore, we highlight the great potential of using RST to provide automated feedback for improving writing quality across genres.

1 Introduction

Recent work in automated essay scoring focuses on local features of writing, often simply to predict grades, though sometimes to offer feedback (Burstein et al., 2003; Wilson et al., 2017). Our focus is specifically at the rhetorical structure level. Structural writing feedback is designed for helping writers to develop a clear structure in which sentences and paragraphs are well-organized (Huang et al., 2017). Researchers have made much progress in providing feedback for enhancing writing structure with the development of intelligent writing systems, such as Writing Mentor (Madnani et al., 2018) and Writing Pal (Roscoe and McNamara, 2013). However, structural writing feedback generated from existing systems is either locally situated in individual sentences or not specific enough for students to take actions. This paper presents how RST can be used to provide global structural feedback for improving writing quality and discusses future work about providing automated writing feedback with deep learning technology. Our contributions are 1) presenting RST annotation resources that can be used to annotate student essays and 2) highlighting the huge potential of using RST annotation for providing automated writing feedback in K-12 education.

2 Creating an Annotated RST Corpus of Student Writing

Though there is an existing data source annotated with RST (Carlson et al., 2002), for our effort we required a corpus of student writing that was annotated with RST. We obtained a student writing corpus through our partnership with TurnItIn.com. Here we describe the data we received, our effort to develop a coding manual for RST applied to this data for our purposes, and the resulting coded corpus.

2.1 Source data

Our data is drawn from a set of 137 student essays from Revision Assistant (Woods et al., 2017), which is an automated writing feedback system developed by TurnItIn.com. Of the 137 essays, 58 are from two genres (i.e., analysis and argumentative writing) and were the primary focus of our effort to design and develop resources to support our annotation effort, including a fine-grained annotation flowchart, guideline, and an intelligent tutoring system (ITS) for training human annotators. As a test of generality, we analyzed the remaining 79 essays, which were randomly sampled from four genres (i.e., analysis, argumentative, historical analysis, and informative writing).

2.2 Goal of annotation

The goal of annotation is to represent an essay in a rhetorical structure tree whose leaves are Elementary Discourse Units (EDUs) (Stede et al., 2017). In the tree, EDUs and spans of text are connected with rhetorical relations (explained in section 2.3). We assume a well-structured essay will have meaningful relations connecting the portions. When meaningful relations connecting EDUs or spans cannot be identified, the assumption is that a revision of structure is needed. The goal of

Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 163–168
Minneapolis, MN, June 6, 2019.

our envisioned automatically generated feedback is to point out these opportunities for improvement through restructuring to students.

More specifically, a span is formed by EDUs connected with rhetorical relations and usually includes multiple EDUs. For example, Figure 1 represents a tree that includes six EDUs (28-33) and four spans (span 29-31, 28-31, 32-33, and 28-33). In some cases, a single EDU is a span when there are no EDUs connecting with it.

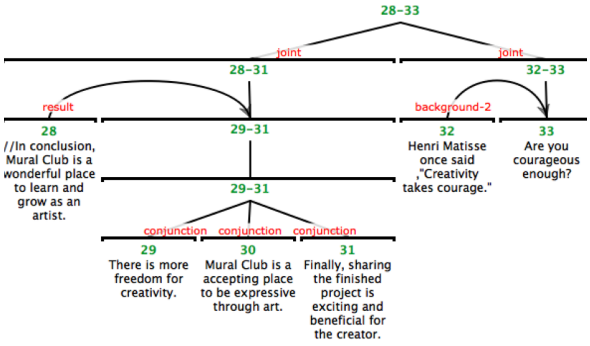

Figure 1: Example of RST annotation with rstWeb

Notice that the EDUs of text at the leaf nodes are mostly single sentences. We segment essays with sentences to represent essays with higher level structure and trigger structural feedback. We used individual sentences as EDUs to provide writing feedback between sentences and paragraphs. Namely, each EDU is a complete sentence, which could be indicated by a full stop, exclamation, or question mark. However, students might use several sentences in prompt sources or punctuation in the wrong way. In these two scenarios, the criteria of punctuation cannot be used to segment essays into EDUs. Instead, we treated continuous sentences in prompt sources as one EDU and segmented essays based on correct punctuation.

2.3 Adaptation of RST for our Data

Though our goal was to retain as much of the spirit of RST as possible, we adjusted its definitions and scope in order to tailor it for our data. We could not share the dataset due to privacy issues. Instead, we clearly demonstrate how to adapt RST for annotating student essays in this section. Annotators need to identify rhetorical relations between three kinds of units: EDUs, spans, and paragraphs. These relations can be divided into two categories: Nucleus-Satellite (N-S) relation and multi-nuclear relation. N-S relation represents the relation between units that are not equally important while multi-nuclear relation represents the relation between equally important units. We developed a guideline for annotators to understand the definition and examples of these relations.

Combine	**Eliminate**	**Change**
Conjunction	Condition	Background
Sequence	Unless	Justify
List	Purpose	Preparation
	Disjunction	Summary

Table 1: Adaptation of RST relations.

Mann and Thompson (1987) defined 23 rhetorical relations and the set of relations has been augmented with eight more relations. We combined, eliminated, and made minor changes (e.g., dividing one relation into multiple ones) to some relations for the purpose of providing meaningful writing feedback (see Table 1).

Specifically, we use Conjunction to represent the relations of Conjunction, Sequence, and List. These three relations all represent sentences being conjoined to serve a common purpose, we combined them because it is doubtful that distinguishing the nuance between these homogeneous relations will have much benefit for triggering writing feedback. In addition, we eliminated the relations of Condition, Unless, Purpose, and Disjunction as they rarely occurred between sentences.

Furthermore, based on the characteristics of student essays, we made minor changes to the relations of Background, Justify, Preparation, and Summary. We divided the relation of Background into two relations, namely the relations of Background-1 and Background-2. Background-1 refers to the relation that describes two closely connected units in which one of them includes pronouns (e.g., it or them) pointing at something mentioned in the other one. This makes it necessary to present one unit for readers to understand the other unit. Background-2 refers to the relation that describes two loosely related units in which one unit increases the reader's ability to understand the other one. We made this distinction because these two relations are very frequently seen in students' essays, yet they can potentially prompt for different writing feedback.

In terms of the relation of Justify, we used the common scenario of two units being inseparable (i.e., one unit is a question and the other unit is the

answer) to identify it. This differs from the relation of Solutionhood as it refers to a pair of answer and question, instead of problem and solution.

In addition, we extended the definition of the relation of Preparation. Our definition of Preparation includes the common scenario of one unit being the generic description and the other unit being the detailed description. For instance, one unit is: "I have three reasons to be vegetarian", and the other unit is: "First, it is healthier, second, it protects animals, and the third is that it saves the earth from global warming." This type of sentence pairs fit the definition of Preparation which describes that the reader's comprehending the Satellite increases the reader's readiness to accept the writer's right to present the Nucleus.

For the relation of Summary, we only looked at the paragraph level granularity. One unit being the summary of parts of a paragraph is not useful for providing feedback and not much different from the relation of Restatement, while one unit summarizing all other units in a paragraph could indicate a high-quality student writing. Therefore, we only considered cases where one unit summarizes a whole paragraph for providing feedback.

While these changes may seem arbitrary, we find it necessary to make these changes during our annotation process to reduce confusion, increase inter-rater reliability and identify relations that can reveal the structure of student essays and trigger meaningful writing feedback. Specifically, the first and second author independently annotated all essays. Any inconsistencies were discussed and resolved resulting in 100% agreement.

2.4 Annotation process

The structure of the coding manual is driven by the process we advocate to human annotators and we followed a top-down annotation strategy (Iruskieta et al., 2014). Overall, the annotation process is meant to consist of five steps:

First step: Segment an essay into EDUs. This step is explained in subsection 2.2.

Second step: Identify central claims in each paragraph. In this step, annotators should first read the whole essay and understand its line of argumentation. Then annotators should identify EDUs that are central claims in each paragraph. Identifying central claims is useful for deciding whether two units are equally important in the third step.

Third step: Identify rhetorical relations between EDUs. In this step, annotators can use rstWeb, a tool for RST annotation developed by Zeldes (2016), to decide the relations between adjacent EDUs in each paragraph from left to right. Specifically, annotators should first determine whether two adjacent EDUs are equally important. The more important EDU is a Nucleus while the other EDU is a Satellite. To identify whether two EDUs are equally important, annotators can use the flowchart in Figure 2. Then annotators should follow a flowchart (Jiang et al., 2019) to identify the relation. The order of relations in the flowchart is based on the ease they can be excluded. Namely, the easier it is to decide whether one relation applies or not, the earlier it appears in the flowchart. If no relation can be used to describe the relation, then the left EDU is the end of a span. A span is formed by EDUs connected with rhetorical relations, as described in subsection 2.2.

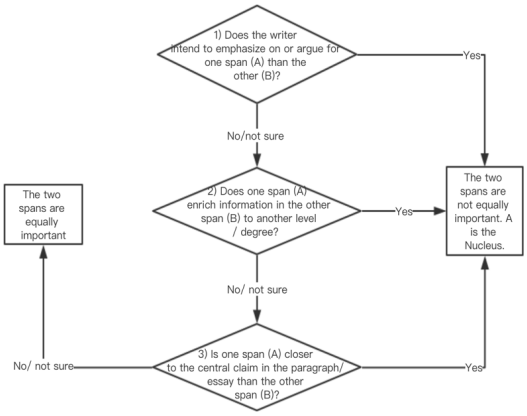

Figure 2: Flowchart for identifying importance

Fourth step: Identify rhetorical relations between spans. In this step, annotators should identify relations between spans within each paragraph from left to right. When identifying relations between two spans, annotators should use the same flowchart in the third step to determine relations between the Nucleus of two spans. If no relation exists between spans, annotators should use Joint to build the paragraph into a tree.

Fifth step: Identify rhetorical relations between paragraphs. Annotators should identify relations between paragraphs from left to right. Similar to the fourth step, annotators should determine the relation between the Nucleus of two paragraphs. If any of the two paragraphs contain the relation of Joint, it indicates that spans in the paragraph do

not have strong relations. In this case, the relation of Joint should be used to connect two paragraphs.

2.5 Practical RST Intelligent Tutor

Based on the flowchart and guideline that make up our coding manual, we developed an Intelligent Tutoring System (ITS) to help novice annotators learn RST annotation efficiently. We built the Practical RST Intelligent Tutor using Cognitive Tutor Authoring Tools (CTAT), an authoring tool for ITS (Aleven et al., 2009). This tutor (Figure 3, access upon request) is hosted on an open platform TutorShop that provides free access for research or public use. As shown in Figure 3, annotators are first presented with three RST relations including their definitions, examples of sentence pairs, conjunction phrases, and groups. Conjunction phrases refer to connection words or phrases that can naturally conjoin two spans. For example, "because" can be used to connect two spans that indicate the relation of Cause. Groups refer to the categories of relations: progressive, supplementary, conjunct, repeating, contrast or no relation. These categories represent a higher level of RST relations. Annotators are then guided to identify the relation of a given sentence pair, and are scaffolded with step by step procedures and hints to complete the task.

To develop the system, we conducted two rounds of cognitive task analysis (Crandall et al., 2006), respectively with five subjects who had no prior experience in RST and three subjects with experience in RST. After analyzing think-aloud data from the first round, we found that novice annotators regularly referred back to the definition of RST relations, compared given sentence pairs with provided examples, and inserted conjunction phrases between annotation units to see whether it made sense logically. Based on these findings, we developed an initial intelligent tutoring system. We further ran a pilot study involving target users with background or experience in RST. These users further provided feedback on both the interface and instructional design. We refined our tutor accordingly with additional features of arranging problems from easy to hard, adjusted granularity of step-loop, and used more harmonious visual design. This intelligent tutor also takes advantage of the Bayesian Knowledge Tracing algorithm (Baker et al., 2008) developed at Carnegie Mellon University to provide adaptive problem selection, which can assist learners to achieve mastery in four knowledge components (i.e. identifying groups, conjunction phrases, nuclearity, and relations) about identifying RST relations (Koedinger et al., 2012).

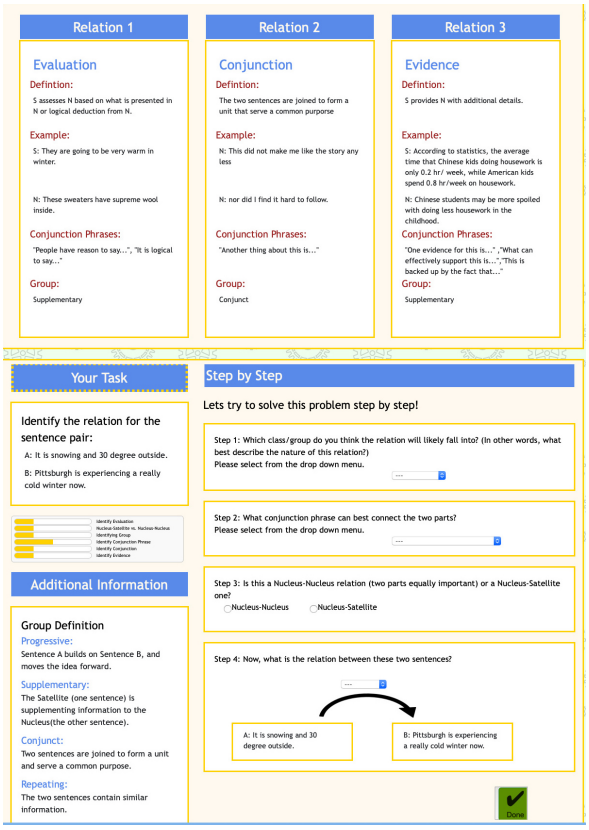

Figure 3: Interface of Practical RST Intelligent Tutor

3 From RST Analysis to Writing Feedback

Here we explain the potential of using RST for providing structural writing feedback across genres and for specific genres.

RST can be used to provide writing feedback for enhancing coherence across genres. Coherence refers to how sentences in an essay are connected and how an essay is organized. RST could be used to provide actionable writing feedback for increasing the level of coherence that traditional automated coherence scores were deemed insufficient to realize. Specifically, the relation of Joint indicates a low level of coherence. As an example, Figure 1 is an annotation of one paragraph from student writing. This paragraph includes two spans (i.e., span 28-31 and span 32-33) that are not connected clearly. In span 28-31, the writer listed three benefits of joining a club. In span 32-33, the

writer might intend to encourage people to join the club while the intention is not clear as there is no mention of joining the club. The RST tree had the potential of giving more concrete context for low-level coherence and in this way, students could identify where they can make revisions for clearer structure.

In terms of providing feedback in specific genres, the combination of relations can indicate high-quality writing. For example, presenting and analyzing evidence is an indication of high-quality argumentative writing (Gleason, 1999). Researchers have made much effort in predicting whether there is evidence in student writing and pointed out the need for future studies in examining how evidence was used to show the soundness and strength of arguments. RST can be used to meet the need with predicting the combination of relations, such as the combination of evidence and interpretation or the combination of evidence and evaluation.

Furthermore, RST is valuable to provide writing feedback in analysis writing. Making comparisons is a common structure of well-organized analysis writing. It's easy to identify sentences involving comparison locally. However, identifying the whole structure of making comparisons in an essay remains to be a challenging automation task. RST has the potential to address the challenge by illustrating a global comparative structure with the relation of Contrast, Antithesis, or Concession.

4 Conclusion

We take full advantage of RST in providing structural feedback for enhancing writing quality across genres. Currently, based on the work from Li et al. (2014), we are building an RST parser that can generate RST trees to represent student essays automatically with deep learning techniques. In the future, we plan to build the work from Fiacco et al. (2019) to generate RST trees more accurately and efficiently. Our long term goal is to embed these techniques in a writing tutor like Revision Assistant and conduct large-scale classroom studies to evaluate the effect of RST trees in writing instruction.

Acknowledgments

This work was funded in part by NSF grant #1443068 and a Schmidt foundation postdoc fellowship. We thank our collaborators at TurnItIn.com for providing student essays. We also thank Vincent Aleven and Jonathan Sewall for design suggestions and guidance in building the practical RST intelligent tutor.

References

Vincent Aleven, Bruce McLaren, Jonathan Sewall, and Kenneth R. Koedinger. 2009. Example-tracing tutors: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 19(2):105.

Ryan S. Baker, Albert T. Corbett, and Vincent Aleven. 2008. More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian Knowledge Tracing. In *International conference on intelligent tutoring systems*, pages 406–415.

Jill Burstein, Daniel Marcu, and Kevin Knight. 2003. Finding the write stuff: automatic identification of discourse structure in student essays. *IEEE Intelligent Systems*, 18:32–39.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2002. *RST Discourse Treebank*. Linguistic Data Consortium, University of Pennsylvania.

Beth Crandall, Gary Klein, and Robert R. Hoffman. 2006. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. A Bradford Book.

James Fiacco, Elena Cotos, and Carolyn Rose. 2019. Towards enabling feedback on rhetorical structure with neural sequence models. In *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, pages 301–319.

Mary M . Gleason. 1999. The role of evidence in argumentative writing. *Reading and Writing Quarterly*, 15(1):81–106.

Yi-Ching Huang, Jiunn-Chia Huang, Hao-Chuan Wang, and Jane Yung jen Hsu. 2017. Supporting ESL writing by prompting crowdsourced structural feedback. In *Proceedings of the Fifth AAAI Conference on Human Computation and Crowdsourcing*, pages 71–78.

Mikel Iruskieta, Arantza Daz de Ilarraza, and Mikel Lersundi. 2014. The annotation of the central unit in rhetorical structure trees: A key step in annotating rhetorical relations. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 466–475.

Shiyan Jiang, Kexin Yang, Chandrakumari Suvarna, and Carolyn Ros. 2019. Guideline and Flowchart for Rhetorical Structure Theory Annotation. Technical report, Carnegie Mellon University, School of Computer Science.

Kenneth R. Koedinger, Albert T. Corbett, and Charles Perfetti. 2012. The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36:757–798.

Jiwei Li, Rumeng Li, and Eduard Hovy. 2014. Recursive deep models for discourse parsing. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2061–2069.

Nitin Madnani, Jill Burstein, Norbert Elliot, Beata Beigman Klebanov, Diane Napolitano, Slava Andreyev, and Maxwell Schwartz. 2018. Writing mentor: Self-regulated writing feedback for struggling writers. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 113–117.

William C. Mann and Sandra A. Thompson. 1987. *Rhetorical Structure Theory: A Theory of Text Organization*. University of Southern California, Information Sciences Institute.

Rod Roscoe and Danielle McNamara. 2013. Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105(4):1010–1025.

Manfred Stede, Maite Taboada, and Debopam Das. 2017. Annotation guidelines for rhetorical structure. Version 1.0.

Joshua Wilson, Rod Roscoe, and Yusra Ahmed. 2017. Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34:16–36.

Bronwyn Woods, David Adamson, Shayne Miel, and Elijah Mayfield. 2017. Formative essay feedback using predictive scoring models. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2071–2080.

Amir Zeldes. 2016. rstWeb-a browser-based annotation interface for Rhetorical Structure Theory and discourse relations. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 1–5.