

Rhetorical structure and argumentation structure in monologue text

Andreas Peldszus

Applied Computational Linguistics
UFS Cognitive Science
University of Potsdam
peldszus@uni-potsdam.de

Manfred Stede

Applied Computational Linguistics
UFS Cognitive Science
University of Potsdam
stede@uni-potsdam.de

Abstract

On the basis of a new corpus of short “microtexts” with parallel manual annotations, we study the mapping from discourse structure (in terms of Rhetorical Structure Theory, RST) to argumentation structure. We first perform a qualitative analysis and discuss our findings on correspondence patterns. Then we report on experiments with deriving argumentation structure from the (gold) RST trees, where we compare a tree transformation model, an aligner based on subgraph matching, and a more complex “evidence graph” model.

1 Introduction

Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) was designed to represent the structure of a text in terms of coherence relations holding between adjacent text spans, where the same set of relations is being used to join the “elementary discourse units” (EDUs) and, recursively, the larger spans. The result is a tree structure that spans the text completely; there are no “gaps” in the analysis, and there are no crossing edges. The relations are being defined largely in terms of speaker intentions, so that the analysis is meant to capture the “plan” the author devised to influence his or her audience. The developers of RST had not explicitly targeted one particular text type or discourse mode (instructive, argumentative, descriptive, narrative, expository), but when we assume that the text is argumentative, the very nature of the RST approach suggests that it might in fact capture the underlying argumentation quite well.

Systems for automatic RST parsing have been built since the early 00s, with recent approaches including (Ji and Eisenstein, 2014) and (Joty et

al., 2015). Hence, a potentially useful architecture for argumentation mining could involve an RST parser as an early step that accomplishes a good share of the overall task. How feasible this is has so far not been determined, though.

On the theoretical side, different opinions have been voiced in the literature on the role of RST trees for argumentation analysis; we summarize the situation below in Section 2. All these opinions were based on the experiences that their authors had made with manually applying RST and with analyzing argumentation, but they were not based on systematic empirical evidence. In contrast, in this paper we use a new resource that we recently released (Stede et al., 2016), which offers annotations of both RST and argumentation structure analyses on a corpus of 112 short texts. Our previous paper presented a first rough analysis of the correlations between RST and argumentation. The present paper builds on those preliminary results and makes two contributions:

- We provide a qualitative analysis that examines the commonalities and differences between the two levels of representation in the corpus, and seeks explanations for them.
- We report on experiments in automatically mapping RST trees to argumentation structures, for now on the basis of the manually-annotated “gold” RST trees.

Following the discussion of related work in Section 2, Section 3 gives a brief introduction to the corpus and the annotation schemes that are used for argumentation and for RST. Then, Section 4 presents our qualitative (comparative) analysis, and Section 5 the results of our experiments on automatic analysis. Finally, Section 6 relates these two endeavours and draws conclusions.

2 Related work

In this section, we summarize the positions that have so far been taken in the literature on the status of RST analyses for argumentation.

The view that performing an RST analysis essentially subsumes the task of determining argumentation structure was advanced by Azar (1999), who argued that RST’s nucleus-satellite distinction is crucial for distinguishing the two roles in a dialectical argumentative relationship, and that, in particular, five RST relations should be regarded as providing argumentative support for different types of claims: Motivation for calls for action; Antithesis and Concession for increasing positive regard toward a stance; Evidence for forming a belief; Justify for readiness to accept a statement. Azar illustrated that idea with a few short sample texts that he analyzed in terms of RST trees using these relations. In one of these examples, however, Azar made the move of combining two non-adjacent text segments into a single node in the RST tree (representing the central claim), which is in conflict with a basic principle of RST. This indicates that Azar borrowed certain aspects from RST but ignored others. In our earlier work (Peldszus and Stede, 2013), we posited that the underlying phenomenon of non-adjacency creates a problem for RST-argumentation mapping in general, i.e., it is not limited to discontinuous claims: Both Support and Attack moves can be directed to material that occurs in non-adjacent segments.

A small portion of RST’s ideas was incorporated into the annotation of argumentation performed by Kirschner et al. (2015) on student essays. The authors used standard argumentative Support and Attack relations, and to these added the coherence relations Sequence and Detail for capturing specific argumentative moves; the relation definitions are inspired by those used in RST.

Green (2010) proposed a “hybrid” tree representation called ArgRST, which combines RST’s nuclearity principle and some of its relation definitions with additional annotations capturing aspects of argumentation: The analyst can add implicit statements to the tree (enthymemes in the argumentation), and in parallel to RST relations, the links between segments can also be labeled with relations from the scheme of Toulmin (1958) and with those proposed by Walton et al. (2008). Also, the representation allows for noncontiguous premises and conclusions. More recently, Green

(2015) argued that the hybrid representation does not readily carry over to a different text genre (biomedical research articles), and she concluded that RST and argumentation structure operate on two levels that are subject to different motivations and constraints, and thus should be kept distinct.

We also subscribe to the view that (at least for many text genres) distinguishing rhetorical structure and argumentation structure is important for capturing the different aspects of a text’s coherence on the one hand, and its pragmatic function on the other. Also, we wish to emphasize the conflict between segment *adjacency* (a central feature of RST’s account of coherence) and *non-adjacency* (a pervasive phenomenon in argumentative function of portions of text). Still, it remains to be seen to what extent an RST analysis can in principle *support* an argumentation analysis, e.g. in a pipeline architecture; shedding light on this question is our goal for this paper.

3 The corpus

Below we provide a very brief description of the data and annotations that we provided in (Stede et al., 2016); for more details, see that paper. Notice that the layers of annotations had been produced independently by different people, thus inviting a posthoc comparison, which we will perform in the next sections. For reasons of space, we do not give further details on RST here; the interested reader should consult (Mann and Thompson, 1988) or (Taboada and Mann, 2006).

3.1 Data

The argumentative microtext corpus (Peldszus and Stede, 2016) is a freely available collection of 112 short texts that were collected from human subjects, originally in German. Subjects received a prompt on an issue of public debate, usually in the form of a yes/no question (e.g., Should shopping malls be open on Sundays?), and they were asked to provide their answer to the question along with arguments in support. They were encouraged to also mention potential objections. The target length suggested to the subjects was five sentences. After the texts were collected, they were professionally translated to English, so that the corpus is now available in two languages. An example of an English text is:

Health insurance companies should naturally cover alternative medical treat-

ments. Not all practices and approaches that are lumped together under this term may have been proven in clinical trials, yet its precisely their positive effect when accompanying conventional western medical therapies that's been demonstrated as beneficial. Besides, many general practitioners offer such counselling and treatments in parallel anyway - and who would want to question their broad expertise?

In (Stede et al., 2016), two new annotation layers are introduced for the corpus: Discourse structure in terms of RST, and in terms of Segmented Discourse Representation Theory (Asher and Lascarides, 2003). Importantly, these two as well as the argumentation annotation use an identical segmentation into elementary discourse units (EDUs).

3.2 Argumentation structure representation

The annotation of argumentation structure follows the scheme outlined in (Peldszus and Stede, 2013), which in turn is based on the work of Freeman (1991). It posits that the argumentative text has a central claim (henceforth: CC), which the author can back up with statements that are in a Support relation to it; this is a transitive relation, leading to “serial support” in Freeman’s terms. A statement can also have multiple Supports; these can be independent (each Support works on its own) or linked (only the combination of two statements provides the Support). Also, the scheme distinguishes between “standard” and “example” support, whose function originates from providing an illustration, or anecdotal evidence.

When the text mentions a potential objection, this segment is labeled as bearing the role of “opponent’s voice”; this goes back to Freeman’s insight that any argumentation, even if monological, is inherently dialectical. The segment will be in an Attack relation to another one (which represents the proponent’s voice), and the scheme distinguishes between Rebut (denying the validity of a claim) and Undercut (denying the relevance of a premise for a claim). When the author proceeds to refute the attack, the attacking segment itself is subject to a Rebut or Undercut relation.

The building blocks of such an analysis are Argumentative Discourse Units, which often are larger than EDUs: Multiple discourse segments

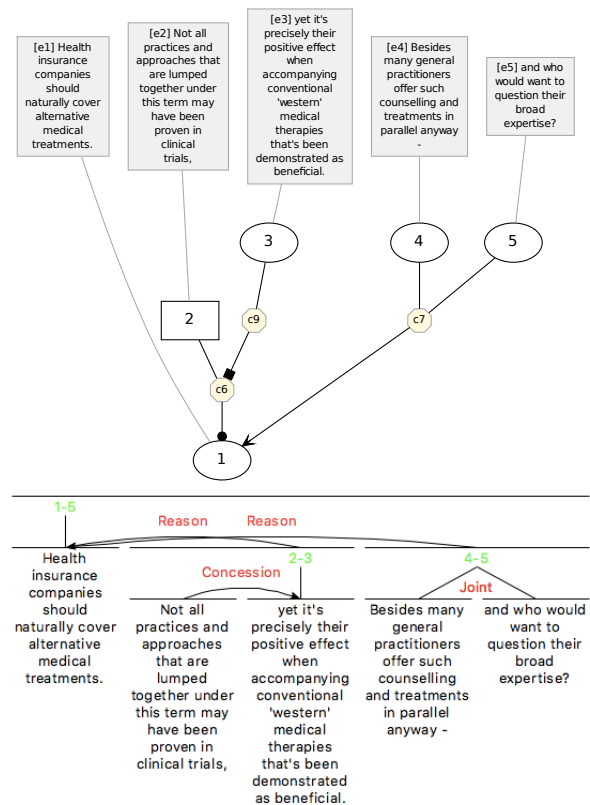


Figure 1: Example ARG and RST structure

play a common argumentative role. In such cases, the EDUs are linked together by a meta-relation called Join. The argumentation and RST analyses of the sample text are shown in Figure 1.

4 Matching RST and argumentation: Qualitative analysis

When introducing the corpus (Stede et al., 2016), we provided figures on how edges in the RST tree map to edges in the argumentation graph (which can be calculated straightforwardly, because both representations build on the same segmentation). We found that, ignoring the labels, 60% of the edges are common to both structures. As can be expected, argumentative Support mostly corresponds to Reason or Justification; however, 39% of the Supports do *not* have a corresponding RST edge. Furthermore, 72% of all Rebuts and 33% of Undercuts do not have a corresponding RST edge. Thus, the correspondences between the layers are certainly not trivial. In order to understand the mismatches, we undertook a qualitative analysis that focuses on the three central notions of the argumentation: the central claim and its mapping to RST nuclearity, the Support relations, and the configurations of Attack/Counterattack.

4.1 CC and Nuclearity

Recall that Azar (1999) already pointed out the importance of RST's notion of 'nucleus' for representing argumentation. To operationalize the analogy, it is important to make use of the "strong nuclearity principle" (Marcu, 2000), according to which the most important segment(s) of a text can be found by following the RST tree from its root down the nucleus links to the leaf nodes. If there are only mononuclear relations along the way, there is a single most important segment (henceforth: RSTnuc); otherwise, there are multiple ones. A natural first question therefore is whether the RSTnuc segment corresponds to the CC in argumentation. We found that for 95 texts, i.e., the vast majority of the 112 texts (85%), this is the case. Considering the goal of RST analysis, which is to capture the main intention of the writer, this is the expected default case.

But what happens in the 17 mismatches? In five cases, RSTnuc and ARGcc are indeed disjoint. Four of these are due to the thesis being stated early in the text and once again (as a paraphrase) later on. It is thus left to the annotator to decide which formulation s/he considers more apt to play the central role of the text – and these decisions happen to have led to different results in the four texts. In the fifth, the thesis is not explicitly stated; here, too, there are two plausible options for choosing the most important segment of the text.

In 12 texts, RSTnuc and ARGcc overlap, which can be due to two reasons. (i) In five texts, ARGcc consists of two EDUs, with the RSTnuc being one of them. This is due to an RST relation that is not argumentatively relevant (mostly Condition). (ii) Seven texts show the reverse situation: A multinuclear RST relation induces >1 RSTnuc. In three of these cases, this seems due to an unclear text; the author's position remains somewhat ambiguous, and the RST annotator considered different statements as equally important. The ARG annotation, on the other hand, was committed to making a decision on the CC (as stated in the guidelines). In the remaining four cases, we find minor differences in interpretation, where the RST decision might well be influenced by surface features, in particular the presence of coordinating conjunctions, which suggest a parallel structure for a coherence-oriented analysis. ARG analysis, on the other hand, encourages the annotator to ab-

stract from linguistic realization and to consider the underlying pragmatic relationships.

4.2 Support

Of the 261¹ ARG-Supports, 132 have a corresponding edge in the associated RST tree, with a label that is clearly compatible with Support: Reason, Justify, Evidence, Motivation, or Cause. And of the 112 texts, 26 have only such canonical SUPPORTs (and three texts do not have Supports at all). Together these are 23% of the texts, so that 77% contain non-canonical Support. This calls for closer investigation, and we found two groups:

(i) 12 Support relations have a corresponding RST edge that is labeled with an "unexpected" relation: Elaboration, Background, Result, Interpretation, Antithesis, Concession, or a multinuclear relation. These are instances of the dichotomy between accounting for the local coherence versus the underlying argumentation; in fact, this corresponds to a discussion that originated shortly after the introduction of RST and pointed out the potential conflict between an "informational" versus an "intentional" analysis (Moore and Pollack, 1992).

(ii) 117 Support relations do not have a corresponding edge in the RST tree. The reasons can be subclassified as follows, with the observed frequency given in parentheses. (These attributes can combine, so the numbers add to more than 117.)

- The RST segment participates in a multinuclear relation (List, Conjunction, Joint), or in the pseudo-relation Same-Unit. Hence it can be reached directly by following the respective edges. (70)
- Relation disagreement: The RST annotator did not see a Support-like relation, but used something else (most often Background or Elaboration). (21)
- Transitivity mismatch: ARG and RST annotations do not agree on serial versus joint support, i.e., whether a segment supports a claim directly or only indirectly. (16)
- Grain size: In a segment, RST uses a non-argumentative relation such as Condition, so that the nuclearity assignment does not match that of the segmentation in the ARG-Join relation. (9)

¹This number diverges by 25 from that given by Stede et al. (2016), because for technical reasons they excluded from their statistics 10 texts that have discontinuous segments.

- Consequence of the different nuclearity structures we mentioned in the previous subsection. (5)
- Different or same reason: RST and ARG annotators differed in whether two segments constitute the same Reason/Support, or separate ones. (4)

4.3 Attack

Finally, we study what RST constellations correspond to attack configurations in the ARG tree. For the time being, we do not distinguish Rebut from Undercut.² We discuss the cases in increasing order of complexity and give the number of texts where the instance occurs (which is almost identical to the number of instances).

1. Text does not have any attacks in ARG. (16)
2. A single attack node in ARG, or a joined pair; these are leaf nodes. This is the situation where an attack is not being countered – the author considers his other Supports to implicitly outweigh the attack. (24) – Variant: The attack is not a leaf but supported by another opponent-voice node. (7) – We treat these together, and of the 31, 24 have a “canonical” RST counterpart: The attacking segment is also a leaf node, and its is connected via one of the RST relations Antithesis, Contrast, Concession. The remaining 7 have a “non-canonical” RST counterpart: The opponent voice is not reflected in the RST tree, or a local attachment of an attacking subordinate segment leads to a non-canonical relation.
3. Similar to (2), but instead of one there are two separate attack nodes in ARG. In all of these cases, the RST tree combines the two attacks in a Conjunction relation. (7)
4. The attack is being countered: An opponent-voice-segment has both an outgoing and an incoming attack. For illustration, consider Figure 1 above (the “incoming” attack of node 2 there is an undercut). In general, there are three structural subclasses. (i): Both attack and counterattack are individual segments (36), as is the case in Fig. 1. The structures can be straightforwardly compared to their RST correspondents as follows:
 - Canonical-a: The counterattack corresponds to a backward Conces-

²Intuitively, we see no evidence that this distinction is reflected in RST, but it needs to be determined quantitatively.

sion/Antithesis, and the whole is the satellite of a canonical support relation (Reason, Justify, ...). (22) This is shown in Fig. 1.

- Canonical-b: Likewise, but the whole participates first in some multinuclear relation (List, Joint), which in turn is the satellite of a canonical support. (6)
 - Non-canonical: RST annotator did not see argumentative function as most important for capturing local coherence. (8)
- (ii) Slightly more complex: The counterattack has >1 segment. (16)
- Canonical: The counterattack subtree gets some RST analysis, and the overall construction is as described in the previous category. (13)
 - Noncanonical: reason as in (i). (3)
- (iii) More complex: The attack has >1 segment. (8)
- Canonical: overall construction is as described above. (6)
 - Noncanonical: Support corresponds to Interpretation/Elaboration. (2)

4.4 Summary

We found a large proportion of CC, Support and Attack configurations to correspond to “canonical” configurations in RST trees – i.e., subtrees that intuitively reflect the argumentative functions (under the definitions of the RST relations). While so far we looked at the correspondence only in the direction ARG→RST, this result still suggests that an automatic mapping from RST to ARG tree can be feasible; this will be the topic of the next section. Furthermore, a central purpose of the manual analysis was to determine the reasons for mismatches, which can inform theoretical considerations on the relationship between RST and argumentation. For reasons of space, we cannot go into detail, but our central observation is that RST analysis is subject to a tension between accounting for the *local* coherence or for the *global* one using underlying intentions, i.e., the argumentation. As we noted earlier, this has been discussed in the RST community early on — but it has never been resolved. The issue is likely to be much more pronounced in longer texts than in the microtexts we are studying here. In principle, the specific RST annotation guidelines could ask annotators to clearly prefer

one or the other perspective; this would shift the original goal of the theory, but probably would do better justice to the data.

Considering the option of annotating an “argumentation-oriented” RST tree, the question arises to what extent it can be theoretically adequate. Of central importance is the correspondence between RSTnuc and ARGcc; we found that for all the mismatches in the corpus, it is possible to construct a plausible alternative RST tree such that the two are identical or at least overlapping (when the granularities of the analyses don’t match exactly). Another issue is the presence of crossing edges, which occur in seven ARG graphs in the corpus. Since this is likely to occur more often in longer texts, it remains a fundamental issue; we will return to it at the end.

5 Deriving argumentation structure from rhetorical structure automatically

In order to automatically map between RST and argumentation, it is very helpful to have both layers in the same technical format. To that end, our joint work with colleagues in Toulouse supplied a common dependency structure representation (Stede et al., 2016). In the following, we use that version of the corpus. For illustration, see Figure 2 for the dependency conversion of the example text.

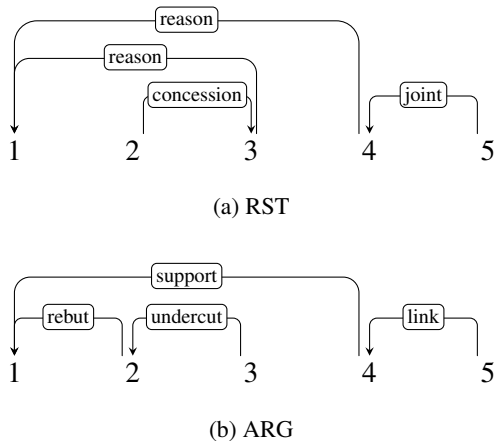


Figure 2: Dependency conversion example

5.1 Models

We have implemented three different models: A simple heuristic tree-transformation serves as a baseline, against which we compare two data-driven models. All models and their parameters are described in the following subsections.

In our study, we follow the experimental setup of (Peldszus and Stede, 2015). We use the same train-test splits, resulting from 10 iterations of 5-fold cross validation, and adopt their evaluation procedure, where the correctness of predicted structures is assessed in four subtasks:

- **attachment (at)**: Given a pair of EDUs, are they connected? [yes, no]
- **central claim (cc)**: Given an EDU, is it the central claim of the text? [yes, no]
- **role (ro)**: Given an EDU, is it in the [proponent]’s or the [opponent]’s voice?
- **function (ro)**: Given an EDU, what is its argumentative function? Here, we use the fine-grained relation set available in the data. [support, example, rebut, undercut, link, join]

Note that the argumentative role of each segment is not explicitly coded in the structures we predict below, but is inferred from the chain of supporting (role preserving) and attacking (role switching) relations from the central claim (by definition in proponent’s voice) to the segment of interest.

5.1.1 Heuristic baseline

The baseline model (**BL**) produces an argumentation structure that is isomorphic to the RST tree. RST relations are mapped to argumentative functions, based on the most frequently aligning class as reported in (Stede et al., 2016) – see Figure 3. For the two relations marked with an asterisk, no direct edge alignments could be found, and thus we assigned them to the class of the non-argumentative *join*-relation. The argumentative *example* and *link*-relations were not frequent enough to be captured in this mapping.

We expect this baseline to be not an easy one to beat. It will predict the central claim correctly already for 85% of the texts, due to the correspondence described in Section 4.1. Also, as we saw above, 60% of the unlabelled edges should be mappable. Finally, the argumentative role is covered quite well, too: The chain of supporting and attacking relations determining the role is likely to be correct on an EDU basis, if the relation mapping is correct, and even if attachment is wrongly predicted.

5.1.2 Naive aligner

Our naive aligner model (**A**) learns the probability of subgraphs in the RST structure mapping to

support: background, cause, evidence, justify, list, motivation, reason, restatement, result
rebut: antithesis, contrast, unless
undercut: concession
join: circumstance, condition, conjunction, disjunction, e-elaboration, elaboration, evaluation-s, evaluation-n, interpretation*, joint, means, preparation, purpose, sameunit, solutionhood*

Figure 3: Mapping of RST relations to ARG relations, used in the heuristic baseline.

subgraphs of the argumentative structure.

For training, this model applies a subgraph alignment algorithm yielding connected components with n nodes occurring in the undirected, unlabelled version of both the RST and the argumentative structures. It extracts the directed, labelled subgraphs for these common components for both structures and learns the probability of mapping one to the other over the whole training corpus.

For prediction, all possible subgraphs of size n in the input RST tree are extracted. If one maps to an argumentation subgraph according to the mapping learned on the training corpus, the corresponding argumentation subgraph is added to an intermediary multi-graph. After all candidate subgraphs have been collected, all equal edges are combined and their individual probabilities accumulated. Finally, a tree structure is decoded from the intermediary graph using the minimum spanning tree (MST) algorithm (Chu and Liu, 1965; Edmonds, 1967).

The model can be instantiated with different subgraph sizes n . Choosing $n = 2$ only learns a direct mapping between RST and ARG edges. Choosing larger n can reveal larger structural patterns, including edges that cannot be directly aligned. Most importantly, the model can be trained with more than one subgraph size n : for example, model **A-234** simultaneously extracts subgraphs of the size $n = [2, 3, 4]$, so that the edge probabilities of differently large subgraphs add up.

The collected edges of all candidate subgraphs do not necessarily connect all possible nodes. In this case, no spanning tree can be derived. We thus initialize the intermediary multi-graph as a total graph with low-scored default edges of the type *unknown*. These should only be selected by the MST algorithm when there is no other evidence for connecting to unconnected subgraphs. The number of predicted unknown edges thus serves as an indicator of the coverage of the learnt model. In

base features incl. 2-node subgraph features:
- absolute and relative position of the segment in the text
- binary feature whether it is the first or the last segment
- binary feature whether it has incoming/outgoing edges
- number of incoming/outgoing edges
- binary feature for each type of incoming/outgoing edge
3-node subgraph features:
- all relation chains of length 2 involving this segment
4-node subgraph features:
- all relation chains of length 3 involving this segment

Figure 4: Segment feature sets

- direction of the potential link (forward or backward)
- distance between the segments
- whether there is a RST relation between the segments
- type of the RST relation between the segments or None

Figure 5: Segment-pair features

evaluation, unknown edges are interpreted as the majority relation type, i.e. as support.

Finally, we added an optional root-constraint (**+r**) to the model: It forbids outgoing edges from the node corresponding to the RST central nucleus, and therefore effectively enforces the ARG structure to have the same root as the RST tree.

5.1.3 Evidence graph model

We implemented a variant of the evidence graph model (**EG**) of (Peldszus and Stede, 2015). In this model, four base classifiers are trained for the four levels of the task (cc, ro, fu and at). For each possible edge, the predictions of these base classifiers are combined into one single edge score. Again, MST decoding is used to select the globally optimal tree structure.

The combined edge score reflects the probability of attachment, the probability of not being the central claim (similar to the root constraint in the alignment model), the probability of a role switch between the connected nodes and the probability of the corresponding edge type. Jointly predicting these different levels has been shown to be superior over the single prediction of the base classifiers.

Our model differs from the original one in two respects: First, our model is trained on the new version of the corpus, featuring a finer segmentation into EDUs, and it considers the full relation set (in contrast to the reduced relation set of just Support and Attack). Second and more importantly, our base classifiers are trained exclusively on a new feature set reflecting aspects of the input RST tree, and do not use any linguistic features.

The segment features are shown in Figure 4. We distinguish three feature groups: base features including edges (EG-2), base features plus 3-node subgraph features (EG-23), and the latter plus 4-node subgraph-features (EG-234). Base classifiers for the cc, ro, and fu-level are trained on segment features. The at-level base classifier is trained on segment features for the source and the target node, as well as on relational features, shown in Figure 5.

As in the original model, the base classifiers perform an inner cross-validation on the training data to optimize the hyperparameters of the log-linear SGD classifier (Pedregosa et al., 2011). We do not optimize the weighting of the base classifiers for score combination here, because we had shown in the original experiments that an equal weighting yields competitive results (Peldszus and Stede, 2015).

5.2 Results on gold RST trees

Scores are reported as averages over the 50 train-test-splits, with macro-averaged F1 as the metric. For significance testing, we apply the Wilcoxon signed-rank test on the macro-averaged F1 scores and assume a significance level of $\alpha = 0.01$. The evaluation results are shown in Table 1.

All alignment models including at least subgraphs of size $n=3$ (A-23*) improve over the baseline (BL) in predicting the relation type (fu) and the attachment (at). Considering larger subgraphs helps even more, and it decreases the rate of unknown edges.³ On the role level, the baseline is unbeaten. For central claim identification, the alignment model performs poorly. Adding the root constraint yields exactly the baseline prediction for the central claim, but also improves the results on all other levels, with the cost of an increased rate of unknown edges. The clear improvement over the baseline for the relation type (fu) indicates that the probability estimates of the alignment models capture the relations better than the heuristic mapping to the most frequently aligning class in the baseline. Furthermore, extraction of larger subgraphs gradually increases the result on both the fu and the at level, showing us that there are subgraph regularities to be learnt which are not captured when assuming isomorphic trees.

³Note that when testing the A-234 model on training data, only very few unknown edges are predicted (less than 1%), which indicates that more data might help to fully cover all of them.

model	cc	ro	fu	at	unknown
BL	.861	.896	.338	.649	
A-2	.578	.599	.314	.650	10.6%
A-23	.787	.744	.398	.707	7.5%
A-234	.797	.755	.416	.719	7.0%
A-2345	.794	.762	.424	.721	6.8%
A-2+r	.861	.681	.385	.682	13.9%
A-23+r	.861	.783	.420	.716	11.3%
A-234+r	.861	.794	.434	.723	10.8%
A-2345+r	.861	.800	.443	.725	10.7%
EG-bc-2	.899	.768	.526	.747	
EG-bc-23	.907	.845	.525	.749	
EG-bc-234	.906	.847	.526	.750	
EG-2	.918	.843	.522	.744	
EG-23	.919	.869	.526	.755	
EG-234	.918	.868	.530	.754	

Table 1: Evaluation scores of all models on the gold RST input trees reported as macro-avg. F1

For the evidence graph model, we will first investigate the performance of the base classifiers (EG-bc-*), before we discuss the results of the decoder. The difference between the three feature sets is most important here. Comparing the classifier that only uses the basic feature set (EG-bc-2) against the one with extra features for 3-node subgraphs (EG-bc-23), we find the greatest improvement on the argumentative role level with an extra +7.7 points macro F1 score. Central claim identification also profits with a minor gain of +0.8 points. Interestingly, the local models for function and attachment are not effected by the richer feature sets. Extending the features even to 4-node subgraphs (EG-bc-234), does not further improve the results on any level.

The evidence graph decoding models (EG-*) combine the predictions of the base classifiers to a global optimal structure. The model using the base classifiers with the smallest feature set (EG-2) already outperforms the best alignment model on all levels significantly and beats the baseline on all levels but argumentative role. We attribute this improvement to three aspects of the model: First, the learning procedure of the base classifiers is superior to that of the alignment model. Second, the base classifiers not only learn regularities between RST and ARG but also positional properties of the target structures. Finally, the joint prediction of the different levels in the evidence graph model helps to compensate weaknesses of the local models by enforcing constraints in the combination of the individual predictions: Com-

paring the base classifier’s predictions (EG-bc-2) with the decoded predictions (EG-2), we observe a boost of +7.5 points macro F1 on the role level and a small boost of +1.9 points for central claim through joint prediction.

Adding features for larger subgraphs further improves the results: EG-23 beats EG-2 on all levels, but the improvement is significant only for role and attachment. EG-234, though, differs from EG-23 only marginally and on no level significantly. Note, that the gain from joint prediction is less strong with better base classifiers, but still valuable with +2.4 points on the role level and +1.2 points for central claim.

In conclusion, the baseline model remained unbeaten on the level of argumentative role. This was already expected, as the sequence of contrastive relations in the RST tree is very likely to map to a correct sequence of proponent and opponent role assignments. On all other levels, the best results for mapping gold RST trees to fine-grained argumentation structures are achieved by the EG-23(4) model.

6 Summary and Outlook

We presented the first empirical study on the relationship between discourse structure (here in terms of Rhetorical Structure Theory) and argumentation structure. In the qualitative analysis, we found a large proportion of “canonical” correspondences between RST subtrees and the central notions of argumentation, with the remaining mismatches being due to an inherent ambiguity of RST analysis (informational versus intentional) and to more technical aspects of granularity (multinuclear relations). By using annotation guidelines that “drive” the annotator toward capturing underlying argumentation, the correspondence could be considerably higher. There remain problems with non-adjacency in the ARG structure, however. These are likely to increase when texts are larger than our microtexts.

For mapping the gold RST trees to ARG structure, we compared three mapping mechanisms: A heuristic baseline, transforming RST trees to isomorphic trees with corresponding argumentative relations; a simple aligner, extracting matching subgraph pairs from the corpus and applying them to unseen structures; and one fairly elaborate evidence graph model, which trains four classifiers and combines their predictions for decoding

globally optimal structures. The latter achieved promising results (with the exception of *prima facie* low numbers for argumentative function, but recall we are using a much larger tag set than all the related work). This confirms the conclusion from the qualitative study, and it invites the next step, which is to use our mapping procedure on the predictions of state-of-the-art RST parsers.

References

- Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.
- M. Azar. 1999. Argumentative text as rhetorical structure: An application of Rhetorical Structure Theory. *Argumentation*, 13:97–114.
- Y. J. Chu and T. H. Liu. 1965. On the shortest arborescence of a directed graph. *Science Sinica*, 14:1396–1400.
- Jack Edmonds. 1967. Optimum Branchings. *Journal of Research of the National Bureau of Standards*, 71B:233–240.
- James B. Freeman. 1991. *Dialectics and the Macrostructure of Argument*. Foris, Berlin.
- Nancy Green. 2010. Representation of argumentation in text with Rhetorical Structure Theory. *Argumentation*, 24:181–196.
- Nancy Green. 2015. Annotating evidence-based argumentation in biomedical text. In *Proceedings of the IEEE Workshop on Biomedical and Health Informatics*.
- Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, June. Association for Computational Linguistics.
- Shafiq Joty, Giuseppe Carenini, and Raymond Ng. 2015. Codra: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3):385–435.
- Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2015. Linking the thoughts: Analysis of argumentation structures in scientific publications. In *Proceedings of the 2015 NAACL-HLT Conference*. Association for Computational Linguistics, June.
- William Mann and Sandra Thompson. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.
- Daniel Marcu. 2000. *The theory and practice of discourse parsing and summarization*. MIT Press, Cambridge/MA.

- Johanna Moore and Martha Pollack. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to automatic argument mining: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31.
- Andreas Peldszus and Manfred Stede. 2015. Joint prediction in mst-style discourse parsing for argumentation mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 938–948, Lisbon, Portugal, September. Association for Computational Linguistics.
- Andreas Peldszus and Manfred Stede. 2016. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015 / Vol. 2*, pages 801–816, London. College Publications.
- 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 International Conference on Language Resources and Evaluation (LREC)*, Portoroz.
- Maite Taboada and William Mann. 2006. Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, 8(4):423–459.
- Stephen Toulmin. 1958. *The Uses of Argument*. Cambridge University Press, Cambridge.
- D. Walton, C. Reed, and F. Macagno. 2008. *Argumentation schemes*. Cambridge University Press, Cambridge.