Encoding Discourse Structure: Comparison of RST and QUD

Sara Shahmohammadi[1*], Hannah Seemann[2*], Manfred Stede[1], Tatjana Scheffler[2]
[1] Department of Linguistics, University of Potsdam, Germany
[2] Department for German Language and Literature, Ruhr-University Bochum, Germany
shahmohammadi@uni-potsdam.de hannah.seemann@rub.de
stede@uni-potsdam.de tatjana.scheffler@rub.de

Abstract

We present a quantitative and qualitative comparison of the discourse trees defined by the Rhetorical Structure Theory and Questions under Discussion models. Based on an empirical analysis of parallel annotations for 28 texts (blog posts and podcast transcripts), we conclude that both discourse frameworks capture similar structural information. The qualitative analysis shows that while complex discourse units often match between analyses, QUD structures do not indicate the centrality of segments.

1 Introduction

Rhetorical structure Theory (RST) (Mann and Thompson, 1988) and the Question under Discussion (QUD) model (e.g., Ginzburg, 1996; Roberts, 2012; Onea, 2019) are two accounts of discourse structure that stem from different research fields and aim to explain different phenomena (speaker intentions and rhetoric versus information structure). However, they share a fundamental formal assumption: that discourse structure is to be represented as a tree that is constructed by recursively combining adjacent "elementary units" of the discourse. For QUD, "discourse" originally meant primarily dialog, while RST was designed for monologue text. Nonetheless, researchers have occasionally explored ways to apply one theory also to the mode of the other.

In this paper, we systematically compare these two approaches to discourse structure, based on empirical observations in a novel multi-media corpus. While some researchers have previously noted the intuitive similarities of RST based trees and QUD based trees on a theoretical level, this work presents the first study where both frameworks are systematically applied to a corpus of both spoken and written data, and compared in a quantitative and qualitative manner.

We first present part of a novel corpus of German blog posts (monologue) and podcast transcripts (dialog). There is a loose 1:1 correspondence between the two, in that the blogs are descriptions of what is being discussed in the podcasts. To our knowledge, this is one of the first corpora that are annotated in parallel with RST and QUD structures. Our aim is to compare the annotated material so that insights into the descriptive and explanatory power of the two approaches can be gained from an empirical perspective of studying authentic data. We make the annotated corpus available to facilitate follow-up research.

To enable the quantitative comparison, we automatically map manually-annotated RST trees to Riester (2019)-style QUD trees. We make the conversion tool available as a web application (and will later release the code). Based on the common format, we perform a quantitative analysis of the similarity of RST and QUD discourse trees, for which we propose an evaluation measure. In addition, our thorough qualitative comparisons show that QUD trees do not indicate the centrality of segments, and often fail to cover relations such as concession and contrast. On the other hand, the topic progression and speaker change within a dialog is captured in QUD analyses but may be missing from RST.

2 Related Work

In annotating both media present in our corpus, we apply both discourse models outside of what they have been primarily designed for. While QUD has been applied to monologue texts before (Riester et al., 2018; Westera et al., 2020), it is most centrally applied to (short) dialogs. On the other hand, RST aims to capture the intentional structure constructed by the writer and is thus designed for monologue, but it has been occasionally applied to dialog, as well (e.g., Stent, 2000).

The QUD framework has been developed to capture aspects of the information structure of sen-

[*] Equal contribution

Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023), pages 89–98

tences and certain specific pragmatic phenomena (Ginzburg, 1996; Roberts, 2012). Only recently has it been used for annotating larger texts, and compared to models of the coherence structure of discourse.

Hunter and Abrusán (2017) compare QUD structures to those proposed by Segmented Discourse Representation Theory (SDRT). They argue that although there likely is no QUD corresponding to every discourse relation, "QUDs correspond to complex discourse units in a discourse graph" (p. 41), that is, topics that lead to grouping discourse units together.

Onea (2019) similarly starts by comparing QUD and SDRT and argues that formally analyzing the erotetic (i.e., question) structure of a discourse can be useful to understand its meaning and its relation structure, for example its SDRT representation. He develops a method for mapping (parts of) question graphs to SDRT representations, and takes a close look at the *Result* relation as a case study. He argues that QUD theories (in particular, models based on potential questions) have repercussions for the larger discourse structure of a text (as represented for example in SDRT).

Riester et al. (2018) offer the first detailed guidelines for segmenting discourse and annotating QUD trees in authentic text, discussing individual texts from English, French, and German and from three different genres. One claim is that the same guidelines apply to monologue (newspaper articles) and dialog (interviews) alike. In later work, Riester et al. (2021) compare QUD, RST, and the CCR approach to discourse structure for one text. Regarding segmentation rules in QUD and RST, they point out that QUD calls for smaller segments than RST when information-structural factors suggest a discourse contribution, e.g., for contrastive foci. Conversely, adjunct clauses can sometimes be separate segments in RST (e.g., in *Circumstance* relations) but not in QUD. As for the relations, the authors show how RST relations can be integrated into a QUD tree notation, and discuss some typical mappings.

Riester (2019) presents a proposal to include both sub- and coordinating relations in a tree that combines QUD and SDRT, based on the approach by Klein and von Stutterheim (1987). For example, the temporal progression of a discourse can be represented by questions asking about each point in time: *What happened at* $t_{1:n}$*?*

Finally, we mention that an early (but somewhat inconclusive) debate in the RST community on the interplay of "intentional" and "informational" coherence relations (e.g., (Moore and Pollack, 1992)) foreshadowed the kind of duality that RST and QUD embody.

3 Data and Method

We carry out an empirical comparison of QUD- and RST-based annotations of the same texts in two media. The idea is based on the assumption that while the QUD and RST frameworks cannot be directly mapped onto one another, both aim to capture the overall coherence of a discourse in a tree-like fashion. Thus, previous work such as (Hunter and Abrusán, 2017; Riester et al., 2021) has proposed to study the correspondences in these discourse trees by looking at the relation between rhetorical relations and question-based structures. Our study is the first, to our knowledge, which carries out a parallel annotation of both spoken and written texts in the two frameworks.

Our data and annotation process is described here.

3.1 Data

The corpus contains texts from two media: podcast transcripts and their corresponding blog posts, both in German. Furthermore, the corpus contains different domains: business podcasts that are produced by companies like DELL or Deutsche Telekom and science podcasts that cover topics from various fields of science and politics. For this analysis, we use 14 blog posts and chunks of 14 podcast transcripts. The blog posts are composed of 26 EDUs on average. The contiguous discourse chunks we annotated from the transcripts consist of 17 EDUs on average. Table 1 shows the size of the resulting sub-corpus.

medium	# episodes	# EDUs	# tokens
blog posts	14	364	4,204
transcripts	14	502	4,980
total	28	866	9,184

Table 1: Corpus overview, EDU and token count.

3.2 Annotation

The texts have been manually annotated in both frameworks, RST and QUD. To simplify the com-

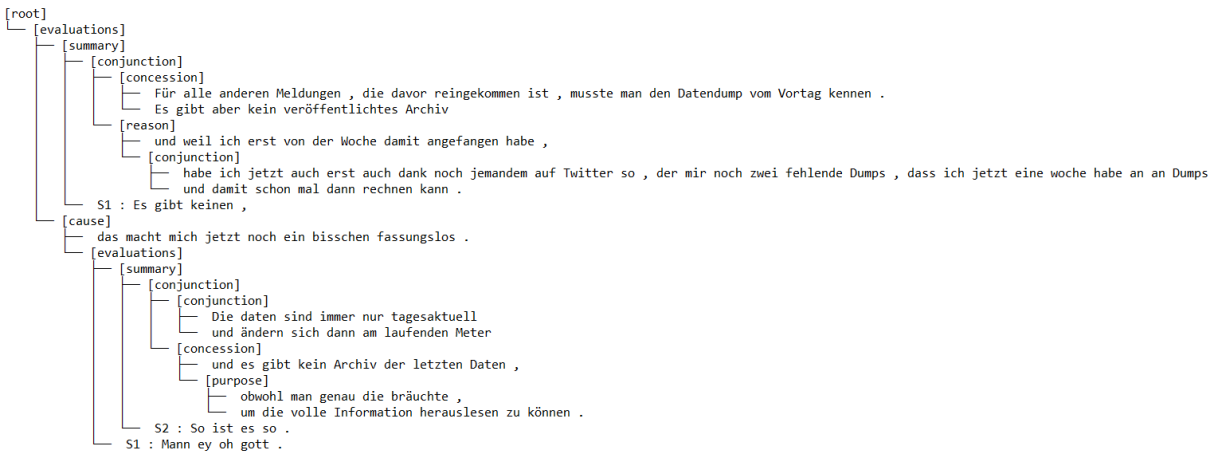


Figure 1: A QUD tree converted from an RST tree (UKW024-p3).

parison between the annotations, we use the RST segmentation according to the Potsdam Commentary Corpus guidelines (Stede, 2016) for both frameworks. For now, we assume EDUs as viable segments for QUD annotation, even though there are differences compared to the usual QUD segmentation, as discussed by Riester et al. (2021). Other than the segmentation, the QUD annotation follows the guidelines defined by Riester et al. (2018). The RST annotation mainly follows the guidelines proposed by Stede et al. (2017), with a few changes.[1]

The annotations were conducted by one person for each model, and revised by a second annotator (a co-author of this paper).[2] The annotated files can be found in the project's GitHub repository[3]. For reasons of space, not all examples referred to in our analysis are shown in the text, but they can be found under the file name given, e.g. CRE210_Transcript_p3. An example of the same file annotated with both discourse models is presented in Figure 2. The English translation of this example is given in (4).[4]

4 Converting RST to QUD Trees

Both RST structures and QUD trees encode discourse structure formally as trees that span over the entire discourse. In RST, intermediate nodes are discourse relations that group (typically) two segments, (typically) a nucleus and satellite. In QUD, intermediate nodes are explicit or implicit questions which guide the discourse; children are (partial) answers to these questions. Disregarding node labels for intermediate nodes, trees with the same yield can be mapped onto each other by comparing just the branching structure.

To evaluate the similarity of RST and QUD trees, we converted the RST trees to a format similar to the QUD trees that can be quantitatively compared to the QUD annotation. Figure 1 shows the converted version of the RST tree in Figure 2.

To convert an RST tree to the QUD format, we take a discourse relation in an RST tree as an intermediate node (implicit question) in a QUD structure. The satellite and nucleus of the relation are daughters of this intermediate node at the same level of nesting.[5] The details of the conversion will be made available in the project repository on GitHub.

5 Quantitative Analysis of RST and QUD Correspondence

We automatically evaluated the similarity of the RST and QUD discourse structures quantitatively

[1] To account for particularities of speech, we added a 'completion' relation that is used in podcast conversations, if a speaker says something that is not complete, e.g. does not have a verb, and then completes it later. In addition, we extended the 'restatement' relation to allow being used as a forward-looking relation. This way, it also covers cases of a speaker's self-correction. It is noteworthy to know that there is no "question" relation in these guidelines. Such a relation exists in some RST guidelines, for instance in the annotation guide proposed by Carlson and Marcu (2001).

[2] For future work, we will add a second annotation in each framework and inter-annotator-agreement.

[3] https://github.com/mohamadi-sara20/rst-qud-comparison

[4] All examples are our own translations of the original German data.

[5] It is also possible to convert an RST tree into a QUD tree where the satellite is nested one more level compared to the nucleus. That is, to consider the satellite a subtopic of the nucleus. This way, information on nuclearity will not be lost in the conversion. However, we found this less similar to a QUD structure, so it was not used for the final analysis.

```
[Q0 What is the way things are?]
└─ [Q1 Welche Folge hat das?]
   ├─ Für alle anderen Meldungen, die davor reingekommen ist, musste man den Datendump vom Vortag kennen .
   └─ [Q2 Ist das möglich?]
      ├─ Es gibt aber kein veröffentlichtes Archiv
      └─ [Q3 Gibt es andere Möglichkeiten als ein öffentliches Archiv?]
         ├─ und weil ich erst von der Woche damit angefangen habe ,
         ├─ habe ich jetzt auch erst auch dank noch jemandem auf Twitter so , der mir noch zwei fehlende Dumps , dass ich jetzt eine woche habe an an Dumps
         ├─ und damit schon mal dann rechnen kann .
         └─ [Q4 Wie bewertet S1 diese Situation?]
            ├─ S1: Es gibt keinen ,
            ├─ das macht mich jetzt noch ein bisschen fassungslos .
            └─ [Q5 Wie genau ist die Situation?]
               ├─ Die daten sind immer nur tagesaktuell
               ├─ und ändern sich dann am laufenden Meter
               ├─ und es gibt kein Archiv der letzten Daten ,
               └─ [Q6 Wofür ist ein Archiv notwendig?]
                  ├─ obwohl man genau die bräuchte ,
                  ├─ um die volle Information herauslesen zu können .
                  └─ [Q7 Bestätigt S2 diese Zusammenfassung?]
                     ├─ S2: So ist es so .
                     └─ [Q8 Wie bewertet S1 die Situation?]
                        └─ S1: Mann ey oh gott .
```

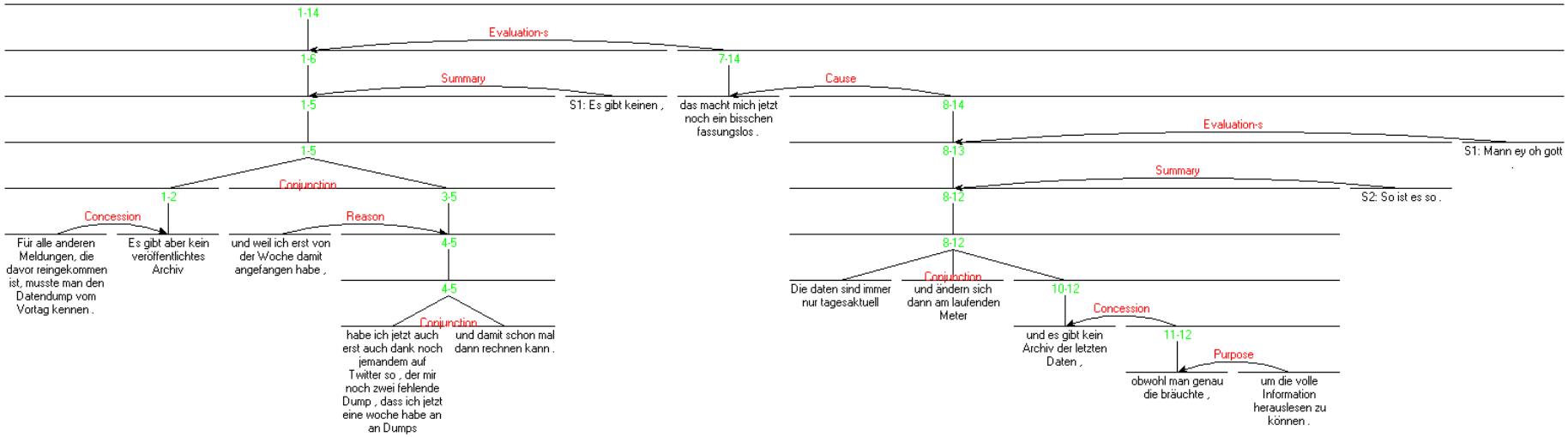


Figure 2: Representations of (4) in QUD (top/left) and RST (bottom/right) (UKW024-p3).

across the whole corpus. To do this, we computed a variant of the PARSEVAL measure known from evaluating (syntactic) constituency parse trees. This measure compares unlabelled trees A and B: First, for all nodes N_A in tree A, one determines whether tree B contains a node with the same yield (= concatenation of all the text dominated by a node) as N_A. In the case of comparing an automatic parse tree with a gold standard, this would reflect precision. Second, for all nodes N_B in B, the same is repeated ($\sim$ recall). We finally compute the harmonic mean of both directions to determine the similarity of RST and QUD structures.

We use unlabelled parseval scores because intermediate nodes are labelled with relations in the case of RST trees, and with questions in QUD trees.[6] Note that the standard parseval score includes leaf nodes in its computation (in the case of syntactic trees, POS tags), and that this practice typically leads to much higher scores than excluding leaves. In our computation, we also include leaves, not only because it is commonly done, but also because in our tree structures, certain error cases can only be reflected when including leaf nodes. This happens in particular, when an explicit text segment is used as an explicit question under discussion (i.e., internal node) in a QUD tree. For this reason, we also redefine the "yield" of a node to be all explicit text dominated by that node, both when it is represented in a leaf node and when it is in an intermediate node.

medium	Q→R	R→Q	f-score
blogs	0.87	0.68	0.76
transcripts	0.85	0.63	0.73
total	0.86	0.65	0.74

Table 2: (Micro-averaged) parseval scores comparing RST and QUD discourse trees.

The results of the quantitative comparison are shown in Table 2. It can be seen that there is a large amount of overlap in the tree structures between QUD and RST frameworks, with an average similarity (parseval) score of 0.74. In addition, we see that the blog posts show higher similarity across frameworks than the podcast transcripts. This is the case even though the blog posts annotated here are on average longer than the transcript snippets (in a longer discourse, there are more possibilities for mismatches in discourse structure annotations). However, we can observe that the similarity of discourse structure between QUD and RST trees is quite high, comparable to inter-annotator agreement within the same framework.

[6]We cannot compute labelled parseval scores because the RST node labels are (a fixed set of) coherence relations and the QUD node labels are (totally free) natural language questions, some of which actually are part of the discourse and thus represented as leaf nodes in the RST trees. To compare QUD trees automatically in general, one would need to define an evaluation method that can rate the equivalence of natural language questions, a task we leave for future work.

6 Qualitative Comparison of RST and QUD Trees

To further evaluate the correspondences between the analyses, we take a closer look at our annotations. First, we inspected the five pairs of trees that received the lowest matching scores in the quantitative comparison; here we note that a frequent source of mismatch is RST's tendency to build a complex discourse unit in cases where QUD attaches the material locally (see below). Then we turned also to the other pairs, trying to generalize sources of misalignment. Section 6.1 compares the way complex discourse units are constructed in both models.The overall structure of texts annotated with both models seems to often be similar, yet there are instances where the structures are quite different. Sections 6.2 and 6.3 compare how translatable different rhetorical relations are to QUD trees and whether RST trees are able to represent typical characteristics of dialog, like speaker changes.

6.1 Comparison of Complex Discourse Units

As discussed in the previous section, both RST trees and QUD trees seem to similarly cluster EDUs into groups, which is particularly beneficial in higher-level units. It is possible to decompose an RST tree into prominent sub-trees. EDUs in the same sub-tree or cluster are closer to each other than they are to other EDUs. In a QUD tree, EDUs grouped together are put under the same parent question and hence address the same topic or rather, answer sub-questions of an overarching question under discussion. One example can be seen in the RST tree in Figure 2: EDUs 1–6 make up one cluster, while EDUs 7–14 are grouped together.

The tendency for QUD and RST analyses to group discourse units similarly is noticeable in most of the trees of the corpus, but there are also exceptions. Figure 2 shows an example where QUD

and RST trees seem to capture different aspects of the functions of discourse unit 6. As evident in the figure, RST subtree 1-5 discusses the problem of the unavailability of public data archives. Unit 6 repeats the same idea, without the details. According to the guidelines, one possible relation holding between 6 and 1-5 is the Summary relation. On the other hand, unit 7 starts with the pronoun 'Das' ('that'), which expresses an evaluation of the current condition of the archives. Hence, the best attachment point for it is unit 6.

However, if the evaluation is attached to 6, the summary relation cannot be chosen to relate it to the prior discourse. On the other hand, if the evaluation is attached to 1-6 instead of only 6, it would mean ignoring the intention behind this repetition. This is where the difference between the two trees arises: The RST tree groups 6 with the previous discourse, and therefore loses the ideal attachment point for the evaluation. The QUD tree, on the other hand, groups 6 with the discourse following it, and hence fails to fully observe the function of this repetition.

This pattern is not limited to discourse management in dialogs but is also seen in monologues. In example (1), part (1-a) discusses the fact that there is a great difference between theory and practice in IT security, and brings some evidence to support this claim. Part (1-b) repeats this idea with fewer details and different wording, and (1-c) evaluates the situation.

(1) a. And what we actually see in crime out there is that the so-called cybercrime is working on a level way lower than, for example, academic research. In academic research, we invent amazing new procedures of cryptography, helping us against quantum computers. What I think, I think this research is important. But what really happens to us is that companies are being hacked because some recipients click on e-mail attachments.

b. Meaning, there is a huge difference between the technically complex and artistic attacks in academic research and street crime.

c. And it makes sense if you compare that to a purse thief.

(DELL001_Transcript)

The main reason for adding (1-b) is to bring attention back to the 'difference between theory and practice in IT security' in order to later evaluate it. If the evaluation is attached to (1-b), the summary relation cannot be chosen. If the evaluation is attached to Summary(a, b) instead of just (1-b), it would mean ignoring the intention behind this repetition. This is, again, the source of the difference between the two trees: The RST tree groups (1-b) with the previous discourse, and hence loses the ideal attachment point for the evaluation, while the QUD tree groups it with the discourse following it, neglecting the function of the repetition.

Another example of the two representations capturing different aspects of a conversation is shown in (2). In this example, the second sentence by speaker 2 (2-c) has two functions: It is an evaluation of the speaker's knowledge of the current topic but at the same time part of a turn-taking-mechanism, offering speaker 1 to evaluate on the net research.

(2) Context: The first human settlements, beginnings of agriculture, and domestication of animals.

a. Speaker 1: I just looked it up, and it is not really well. . . It seems to be disputed when the dog really joined us, the process seems to have been over a very long time and gradually.

b. Speaker 2: Yes, maybe it fluctuates, too, that is, how cultures reacted to it.

c. Well, with your net research, you probably know more about it than I do.

d. S1: Well, I'm not reading all of it now. There is a lot to say here about *canis lupus*, but sometime around 15,000 and 100,000 years. It really is vaguely defined.

(CRE210_Transcript_p5)

Double functions like these are not rare in our data, many utterances have a discourse-managing function on top of the propositional content. Thus, they may be interpreted differently by different annotators, leading to a different annotation of the same text independent of systematic differences between the two discourse models.

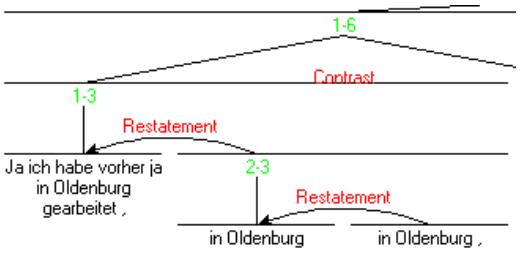
6.2 Comparison Between Relations

Restatement. Figure 4 shows an example of a 'restatement' relation in both QUD and RST trees.

```
[Q6 Which products contain fructose?]
├── Apples, juice, dairy products, cereal, and some sweets have one thing in common:
├── they contain fructose.
└── [Q7 What is the problem with fructose?]
    ├── What health-conscious and sweet teeth often don't know:
    ├── Excessive consumption of fructose can lead to stomach ache and diarrhea.
    └── [Q8 How much fructose do products contain?]
        ├── Some drinks, for example, contain up to 40g of fructose per liter.
        ├── [Q9 What amount of fructose is problematic?]
        │   └── But consuming more than 35g per meal is considered problematic.
        └── Even a cup of yogurt with low sugar content can contain up to 15g of fructose.
```

Figure 3: Concession represented in QUD tree, translated to English (VBZ011_Blog).

Instances of 'restatement' are all at the same level in a QUD tree (since they all answer the same question), while an RST tree does not represent this as a parallel structure; it forms a tree that is right-branching. Our RST annotation guidelines allow the restatement of adjacent units to be successfully modeled.[7]




```
[Q1 Wie ist der Werdegang der Sprecherin?]
├── Ja ich habe vorher ja in Oldenburg gearbeitet ,
├── in Oldenburg
└── in Oldenburg ,
```

Figure 4: Restatement represented in RST tree (top) and QUD tree (bottom) (FG029-p2).

Although restating an idea does not introduce a new topic, there are definitely intentions behind it; sometimes the speaker needs time to think and so they buy time by repeating themselves, sometimes interlocutors want to make a topic more memorable and do so by restating it. It is also possible to restate a previous topic, which is not the current topic, in order to make it salient in current discourse. A QUD structure would not be concerned with these functions of a 'restatement' relation. In contrast, the fine-grained representation of discourse relations provided in RST or SDRT would distinguish such cases of 'restatement'.

[7]Our guidelines respect the adjacency constraint, so it is not possible to mark non-adjacent restatements. That is, if somewhere the fifth EDU restates the content of the first EDU, we cannot model that with a restatement relation according to our guidelines, as it would need a non-adjacent edge.

Background. A background relation in RST introduces background information in order to enable the reader to understand a more central claim. What is annotated with a 'background' relation in an RST tree has been modeled in different fashions in the QUD annotation. Sometimes it is represented in the QUD tree with a nested structure, with a series of nested questions and answers leading to the more central claim. But in different cases what is a 'background' satellite in the RST tree is annotated in the QUD tree to share the same immediate parent of the central claim – and therefore without a nested structure.

This is probably due to the QUD model being concerned with different aspects of discourse than the RST model. A QUD representation does not aim at presenting what is the most important or central claim of discourse but rather the series of sub-questions that are talked about.

Concession and Contrast. Since there is no question type that can be answered with a 'but'-statement (Scheffler, 2013), it may be expected that a QUD representation of a discourse is unable to model contrastive relations. Q9 in the QUD tree shown in Figure 3 is annotated with the 'concession' relation in the corresponding RST tree.

In the QUD tree, the concession is represented by an additional sub-question – as is also suggested by Riester et al. (2021). This way, concessions can, in principle, be modeled in a QUD representation. However, the QUD representation cannot explicitly show which part of the concession is the expectation or the violation of the expectation – the information conveyed by the nucleus and the satellite in an RST representation is lost.

A similar problem arises when dealing with 'contrast' relations. Even though contrasts can be modeled by a parallel structure in a QUD tree, see example (3) Q4.1 and Q4.2, the contrastive meaning

is not explicitly represented due to the list character of this representation. Thus, it is not impossible to represent a tree that has contrastive relations in a QUD structure in general, but the resulting representation is not explicitly contrastive.

(3) A1: And then there seems to have been some shift in the climate,
{Q2: What did this shift in the climate cause?}
A2': causing Africa to dry in its center,
A2": that is changing more and more into a savannah landscape.
{Q3: What were the resulting changes in humanity?}
A3: and then there seems to have been a split in the evolution
{Q4: What was the result of this split?}
{Q4.1: What did the first half do?}
A4.1: one part further tried to eat vegetarian food
{Q4.2: What did the second half do?}
A4.2: and the others started to hunt.

(CRE210_Transcript_p2)

6.3 Comparison of Speaker Changes

While the QUD model is made to deal with explicit questions and speaker changes in dialog, RST is not. Still, in some examples, both representations deal with the speaker change in a similar way: A different speaker than the person before takes the turn, their utterance has its own subtree giving more information, elaboration, or repeating what has been said, then both models go back to a higher discourse level (see figure 5).

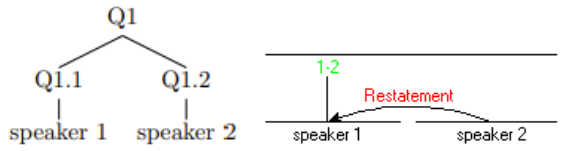


Figure 5: Representations of the same speaker change in QUD (left) and RST tree (right).

But in most cases of speaker changes, the utterances by each speaker are separate sub-trees connected on a higher level in the RST tree, while in the QUD, the second speaker's utterance does not form its own sub-tree. Since RST and QUD deal differently with topical progression, examples like (4) are represented in different ways.

(4) Context: Daily reports of COVID case numbers and the German government's archive infrastructure.
a. Speaker 2: For all the other reports that came in before, you have to know all the data dumps from the day before. But there is no public archive and because I only started this week, I now have only, thanks to someone on Twitter who gave me two missing dumps, so that I now have one week of dumps and do calculations with it.
b. Speaker 1: There is none, that really stuns me. The data is always only on a daily basis and changing constantly and there is no archive of the previous data, even though you would need exactly this to be able to extract all the information.
c. S2: That's how it is.
d. S1: Man that sucks.

(UKW024_Transcript_p2)

In the RST representation, an utterance containing a summary of the previous topic and a transition to the next topic will always be parts of two different sub-trees. On the other hand, in the QUD representation, the transition to the next topic will be a child of the previous topic, as long as both topics are closely connected to each other - see figure 2 for both representations of (4). This means that in the RST representation, the utterance by speaker 1 is split up between two different sub-trees while in the QUD representation, it is not. Unsurprisingly, example (4) has one of the lowest similarities (0.68) in our quantitative comparison between discourse annotations.

Another problem occurs if a speaker change is accompanied by an explicit question. While this is what QUD is meant to model and has no problem dealing with, an explicit question that has no other function than introducing a new topic cannot be appropriately represented in an RST tree without introducing an additional discourse relation.

7 Discussion

In this paper, we have carried out a systematic comparison of the discourse structures induced by RST and QUD frameworks for the same texts from two media, blog posts and podcast transcripts. Our annotations of the 28 texts show that both frameworks

can be successfully applied to monologue texts as well as dialogs.

We compare the branching structure of the resulting RST and QUD trees by first providing a method that converts an RST tree into an equivalent QUD representation. We compare these representations to the manually annotated QUD trees and find an overall similarity of 0.74 – similar to or surpassing the agreement scores between human annotators for discourse structure annotation tasks. This shows that the two frameworks cover some of the same information for our corpus. The similarity between analyses is higher for the blog posts, indicating that the topic structure (QUD) and coherence/intentional structure (RST) are more closely related for monologue texts than for dialog (where two speakers have to agree on how the overall discourse progresses).

Finally, we provide a detailed qualitative comparison of the way complex discourse units are mapped across frameworks, about how certain discourse relations can be represented in QUD trees, and the effects of speaker changes in dialog. We discuss that while the overall structure of QUD and RST trees often matches approximately, QUD trees do not indicate the centrality of discourse segments and cannot represent certain types of relations easily, such as concession and contrast. In contrast, the topic progression within a discourse is captured in QUD analyses but may be missing from RST.

Limitations

We have carried out all analyses according to our best abilities. Nevertheless, it should be noted that RST structures and QUD structures were annotated by distinct researchers. While all annotations have been double-checked by at least one other expert for plausibility, in many cases there are alternative analyses of the texts which may also be applicable (as is usually the case for discourse structure). Since we do not have direct access to the discourse creators and their goals, this limitation is unavoidable in corpus studies.

Ethics Statement

The data reported on in this paper was collected within the research project named below. For all texts, explicit consent was obtained from the creators to use the data for annotation and scientific analysis. The data was automatically and manually preprocessed and reformatted, and manually annotated for discourse structure (among other levels). All annotations were carried out by researchers during their regular work time, either PhD students (authors of this paper) or student research assistants with regular work contracts. The data will be long-term archived and made available to other researchers according to the laws that apply. We see no other ethical issues with our data or research practices.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 317633480 – SFB 1287. We want to thank Elisa Lübbers for helping with RST annotation.

References

Lynn Carlson and Daniel Marcu. 2001. Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, 54(2001):56.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. *Logic, language and computation*, 1:221–237.

Julie Hunter and Márta Abrusán. 2017. Rhetorical structure and QUDs. In *New Frontiers in Artificial Intelligence*, pages 41–57. Springer International Publishing.

Wolfgang Klein and Christiane von Stutterheim. 1987. Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 109:163–183.

William Mann and Sandra Thompson. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *TEXT*, 8:243–281.

Johanna Moore and Martha Pollack. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, 18(4):537–544.

Edgar Onea. 2019. Underneath rhetorical relations. The case of result. In Malte Zimmermann, Klaus von Heusinger, and Edgar Onea, editors, *Questions in Discourse: Volume 2: Pragmatics*, pages 194–250. Brill, Leiden, The Netherlands.

Arndt Riester. 2019. Constructing QUD trees. In Malte Zimmermann, Klaus von Heusinger, and Edgar Onea, editors, *Questions in Discourse: Volume 2: Pragmatics*, pages 164–193. Brill, Leiden, The Netherlands.

Arndt Riester, Lisa Brunetti, and Kordula De Kuthy. 2018. Annotation guidelines for questions under discussion and information structure. In *Information structure in lesser-described languages. Studies in prosody and syntax*, pages 403–443. John Benjamins.

Arndt Riester, Amalia Canes Nápoles, and Jet Hoek. 2021. Combined discourse representations: Coherence relations and Questions under Discussion. In *Proceedings of the First Workshop on Integrating Perspectives on Discourse Annotation*, pages 26–30, Tübingen, Germany. Association for Computational Linguistics.

Craige Roberts. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Tatjana Scheffler. 2013. *Two-dimensional semantics: Clausal adjuncts and complements*, volume 549 of *Linguistische Arbeiten*. Walter de Gruyter.

Manfred Stede, editor. 2016. *Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0*. Number 8 in Potsdam cognitive science series. Universitätsverlag Potsdam, Potsdam.

Manfred Stede, Maite Taboada, and Debopam Das. 2017. Annotation guidelines for rhetorical structure. *Manuscript. University of Potsdam and Simon Fraser University*.

Amanda Stent. 2000. Rhetorical structure in dialog. In *INLG'2000 proceedings of the first international conference on natural language generation*, pages 247–252.

Matthijs Westera, Laia Mayol, and Hannah Rohde. 2020. TED-Q: TED talks and the questions they evoke. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1118–1127, Marseille, France. European Language Resources Association.