# Rhetorical Strategies in the UN Security Council: Rhetorical Structure Theory and Conflicts

**Karolina Zaczynska** and **Manfred Stede**
University of Potsdam, Applied Computational Linguistics
Potsdam, Germany
{lastname}@uni-potsdam.de

## Abstract

More and more corpora are being annotated with Rhetorical Structure Theory (RST) trees, often in a multi-layer scenario, as analyzing RST annotations in combination with other layers can lead to a deeper understanding of texts. To date, prior work on RST for the analysis of diplomatic language however, is scarce. We are interested in political speeches and investigate what rhetorical strategies diplomats use to communicate critique or deal with disputes. To this end, we present a new dataset with RST annotations of 82 diplomatic speeches aligned to existing Conflict annotations (UNSC-RST). We explore ways of using rhetorical trees to analyze an annotated multi-layer corpus, looking at both the relation distribution and the tree structure of speeches. In preliminary analyses we already see patterns that are characteristic for particular topics or countries.

## 1 Introduction

The United Nations Security Council (UNSC) meetings offer a unique longitudinal, cross-thematic resource on diplomatic interactions. Transcriptions of these meetings (Schönfeld et al., 2019) are a valuable corpus to study language use and communication style in an international relations context. In this paper, we study rhetorical style in diplomatic speech, by analyzing UNSC speeches from the perspective of Rhetorical Structure Theory (RST) (Mann and Thompson, 1988).

RST aims to capture the structure of a text by combining its elementary discourse units (EDUs) into one single, hierarchical tree structure. RST trees have proven to be useful in several downstream tasks, including characterizing genre distinctions (Sun et al., 2021; Liu and Zeldes, 2023), investigating text complexity (Hewett, 2023; Williams and Power, 2008) and fake news analysis (Rubin and Vashchilko, 2012; Popoola, 2017). However, little work has been done on RST in political and diplomatic context, with a notable exception presented by Zeldes (2017). We address this gap by presenting a new corpus of 82 UNSC speeches annotated with RST trees. The resulting corpus (henceforth referred to as UNSC-RST) overlaps with our earlier work (Zaczynska et al., 2024), in which we annotated verbal Conflicts in UNSC speeches. In this paper, we present a multi-layer corpus of both RST trees and linguistic markers of Conflicts. We demonstrate how combining the two layers can reveal strategies in verbalizing disputes in a diplomatic setting. The main contributions of this paper are:

First, we present a new corpus with RST annotations for 82 diplomatic speeches from the UNSC. We adopt the RST annotation guidelines from earlier work (Carlson and Marcu, 2001; Zeldes, 2017; Stede et al., 2017), but make amendments tailored to the characteristics of diplomatic language. We include and discuss inter-annotator agreement, and publish our annotation guidelines.

Second, we combine our obtained RST annotations with earlier annotations of Conflict over the same texts, and use insights from argumentation analysis (Stede, 2016), to demonstrate how conclusions can be drawn on strategies to express Conflict. We compare the rhetorical style used by different countries (the five permanent members of the UNSC, plus Ukraine) and in different topics (debates concerning the situation in Ukraine, and the Women, Peace and Security agenda), and show, for example, that Conflicts are not as often supported by causal or justification relations as one might expect.

Our work provides an empirical basis for Political Science and International Relations researchers who are interested in understanding rhetorical styles used by representatives of different countries and in different contexts. The dataset, guidelines and code are available at: `https://github.com/linatal/rhetorical_UNSC`



*Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 15–28
September 18–20, 2024. 

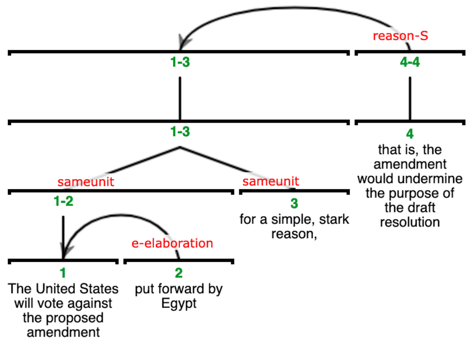



Figure 1: RST subtree from UNSC-RST (S/PV.7658, United States of America)

## 2 Background

This section first provides an overview of earlier work related to RST, and then describes the UNSC Conflicts Corpus that our work is based on.

### 2.1 RST Theory and Corpora

RST (Mann and Thompson, 1988) is a theory for analyzing the organization of texts and looks at discourse from an intention-driven perspective. It represents the structure of text in terms of coherence relations between text spans and captures the "plan" the author devised to influence their audience. Annotating texts with RST consists of two main steps: 1) segmenting the text into so-called Elementary Discourse Units (EDUs) and 2) organizing the EDUs into a single, hierarchical tree-structure. The result is a tree with hierarchically weighted EDUs, capturing the relative importance of each unit. Fig. 1[1] shows an RST tree with EDUs and discourse relations relations between EDUs. Most relations express a hierarchical relation between EDUs; they connect a less important EDU (called the *satellite*) to the more more important one (the *nucleus*). In Fig. 1, EDU 4 is supporting the decision described in EDUs 1-3 by providing a REASON for the decision. Some relations, however, join equally-weighted EDUs, such as SAME-UNIT, which in the example connects two EDUs (1 and 3) that are interrupted by an E- ELABORATION (2).

Existing RST corpora such as the RST Discourse Treebank (RST-DT) (Carlson et al., 2001, 2002), the RST layer of the Georgetown University Multi-layer corpus (GUM) (Zeldes, 2017) and the RST layer of the Potsdam Commentary Corpus (henceforth: PCC-RST) (Stede et al., 2017) each come with their own, slightly different versions of annotation guidelines. The guidelines of our UNSC-RST corpus are based on both the RST-DT and PCC-RST: For EDU segmentation, we use the RST-DT guidelines, and for relation annotation, we adopt (and slightly modify) the relation set from the PCC-RST (see Section 3.1 for more information on our relation set).

Our UNSC-RST corpus is an addition to the collection of RST-annotated texts, of which, to the best of our knowledge, only one covers texts from the political domain: The GUM corpus, since its v7.0.0 version, includes 15 speeches given in the UN General Assembly (16,720 tokens).[2] In comparison, the UNSC-RST corpus contains more speeches (82 vs. 15 in GUM) and more tokens (56,535 vs. 16,720 in GUM).

Obtaining RST trees automatically is the goal of RST parsing (Nguyen et al., 2021; Kobayashi et al., 2021; Liu and Zeldes, 2023), and RST trees have been used for downstream tasks such as text quality assessment (Skoufaki, 2020), summarization (Altmami and Menai, 2020), sentiment analysis (Kraus and Feuerriegel, 2019), and argument mining (Hewett et al., 2019).

### 2.2 The UNSC Conflicts Corpus

Our RST annotations are done over the same speeches as the Conflict annotations in the UNSC Conflicts corpus (UNSCon) (Zaczynska et al., 2024). There, Conflicts are defined as verbalized disagreements or critique towards someone present at the UNSC debate (and the term Conflict does not refer to a military or physical conflict). There are different sub-types of Conflict:

- *Direct Negative Evaluations (Direct_NegEval)* describe Conflicts where the speaker directly directs the critique to another country.
  Example: *This is a claim that takes Russia's distortion of international law to a new level.* (S/PV.7165, United Kingdom and Northern Ireland)[3]
- *Indirect Negative Evaluations (Indirect_NegEval)* describe Conflicts where some intermediate entity serving as a proxy is criticized instead of the other country directly. This can be done, for example, by criticizing

[1] All RST examples are taken from UNSC-RST. We provide an official debate ID, beginning with S/PV and the country of the speaker, for each example.

[2] https://github.com/amir-zeldes/gum/releases/tag/V7.0.0.

[3] Examples are taken from UNSC debates on the situation in Ukraine.

a group acting on behalf of another country, or by criticizing a resolution the other country is supporting.
Example: *It is clear where responsibility lies: with the senseless violence of armed separatists and with those who have supported, equipped and advised them.* (S/PV.7165, United Kingdom and Northern Ireland)

- *Challenging* statements accuse another country of not telling the truth (see example below).
- *Corrections* rectify the allegedly false statement.
Example: *To conclude, one of our colleagues said that Kyiv had extended a hand to Moscow and that we had refused to reciprocate. (Challenge)*
*But the problem is not with Moscow; it has to do with the fact that Kyiv should have been the one to extend a hand to its people and regions, [...]. (Correction)* (S/PV.7138, Russian Federation)

## 3 Annotations and Data

In the following, we describe our annotation guidelines, the annotation procedure, and corpus statistics.

### 3.1 RST Guidelines Expansion

The first step in RST annotation is EDU segmentation. EDUs are sentences or smaller units (mostly clauses). Since in the UNSCon the speeches are already segmented into EDUs for its Conflicts annotation, we directly use their segmentation and refer to Zaczynska et al. (2024) for details on segmentation. The second step in RST annotation consists of choosing discourse relations to link EDUs. The next section describes our modifications to the PCC-RST relations guidelines.

#### 3.1.1 Additional Relations

We use the discourse relation set of (Stede et al., 2017), and include four additional relations (all taken from RST-DT, except for TOPIC-COMMENT, which is from GUM): SAME-UNIT, ATTRIBUTION, TEXTUAL-ORGANIZATION, and TOPIC-COMMENT. Since the sentence structure in the UNSC speeches is relatively complex (see Zaczynska et al. (2024, Table 1)) we found many cases where the EDU was interrupted by one or more embedded discourse units. To connect interrupted EDUs we use the SAME-UNIT relation. We also include ATTRIBUTION, which serves to identify the speaker or source of a statement, because for the analysis of Conflicts it can be important to see whether speakers refer to other sources or to themselves (for example, when accusing someone of a false statement, like in *Challenge*-type Conflicts). We use TEXTUAL-ORGANIZATION to make links between different structural elements, such as between the title and the body of the text, or between a section heading and the following text. TOPIC-COMMENT is used for EDUs that do not contribute propositional content to the discourse, including back-channeling, incomplete utterances, and fillers.

#### 3.1.2 Merging Relations

In the guidelines by Stede et al. (2017), REASON and JUSTIFY both describe EDUs that aim to change the attitude of the reader. The difference is that for REASON, the claim is supported by a subjective assessment, while JUSTIFY describes a general basic attitude of the writer. Because this difference seems not relevant for our genre here, we decided to merge both relations and call them REASON.

#### 3.1.3 Rhetorical Questions

A particular challenge was the annotation of *rhetorical questions*, which appear quite frequently in the speeches. In RST-DT, they are labeled as RHETORICAL-QUESTION, which is a sub-type of TOPIC-COMMENT. However, ideally an RST relation should express the purpose of a unit in relation to another one, rather than characterizing a single unit in itself. Since rhetorical questions often have the purpose to emphasize for example a REASON for a claim, or the EVALUATION of a situation or statement, we decided to use these relations, instead of the general TOPIC-COMMENT relation. We only use TOPIC-COMMENT in cases where it is possible to remove the rhetorical question without losing essential information. For more details on the RST relations, we refer to the RST annotation guideline amendment provided in our repository.

### 3.2 RST-Annotation Procedure

We used the RSTWeb annotation tool for tree building (Zeldes, 2016). Five annotators were trained for over a month for the first round of RST annotations. Then we conducted parallel annotations for a subset of 32 speeches, with two annotations per speech, based on the guidelines from Stede et al. (2017). For statistical evaluation we use the tool

RST-Tace (Wan et al., 2019), which is based on a qualitative method for comparing RST trees as described in Iruskieta et al. (2015). We computed inter-annotator agreement and found an overall average kappa of 0.44. The kappa score for nuclearity (defining the relative importance of an EDU) is 0.43; for relations it is 0.31; for constituents (the central nucleus) it is 0.43, and for attachment points (the direction of the relation) it is 0.51.

A confusion matrix providing more information about disagreements is given in Appendix A. Note that for the gold annotation we added four relations to the list of relations (see section 3.1.1). Most of the mismatches in the annotations can be related to semantic similarity of the chosen relations. For example, a frequent source of disagreement was LIST vs. CONJUNCTION. Both are essentially enumerating EDUs of the same importance, one using typographical connectors like commas or semicolons, the latter using conjunctions like *and* and *or*. Another frequent disagreement was between E-ELABORATION and ELABORATION. This has also been reported by Hewett (2023). Both relations state that the topic of the discourse is being continued in a more specific way, but for E-ELABORATION, the additional information is only on a single entity.

After we obtained the preliminary annotations for IAA calculation, we proceeded to form the adjudicated gold annotations. Two annotators (one is an author of this paper) annotated the entire corpus of 82 speeches, and continuously discussed progress via chat and in weekly meetings, thus creating the gold annotations according to the updated guidelines.

For the final trees, we decided to make use of the given paragraph breaks within the speech transcriptions. This means we first annotated adjacent EDUs for all paragraphs individually and then completed the tree for the whole speech. This way, we speeded up the annotation process for longer speeches. Another advantage was that it enables us to compare sub-tree structures and discourse relation distributions, as well as to find local most-important EDUs within the paragraphs (see Section 4).

### 3.3 UNSC-RST Corpus Statistics

The UNSC-RST corpus includes 85 speeches and therefore 85 RST trees with 60.87 EDUs per tree on average and 11.32 tokens per EDU on average (56,535 tokens in total). It covers almost all of the speeches from the UNSCon.[4] The smallest tree has only seven EDUs (S/PV.7138_spch016, Jordan), whereas the largest one has 194 EDUs (S/PV.7165_spch019, Ukraine). There are six debates in total, covering two topics: Four debates (61 speeches) on the situation in Ukraine (from 2014), and two debates (24 speeches) on the "Women, Peace, and Security" agenda (both from 2016) dealing with gender aspects in security issues. The corpus includes 578 paragraphs, which are seven paragraphs on average per speech, with a maximum of 20 paragraphs per speech.

## 4 Methods

In this section, we describe the kinds of quantitative and qualitative analyses that we performed; the corresponding results will follow in the next section.

### 4.1 Distribution of Discourse Relations

Inspired by Popoola (2017); Hewett (2023) and others, we first look at the discourse relation distribution. We compare the frequency of RST relations and Conflict annotations per EDU on the leaf nodes (EDUs on the lowest level). In order to compare the distribution of relations between Conflicts, we look at the percentage of RST relations used per Conflict type. PCC-RST divides the set of RST relations into four groups according to their function: (1) *Pragmatic relations* serve to change the attitude of another person; (2) *semantic relations* describe states of affairs in the world; (3) *textual relations* organize the text and make its understanding easier; and (4) *multinuclear relations* enlist two or more EDUs of same importance in a relatively weak rhetorical relation. For our purposes here, we separately build the group of (5) *contrastive relations* that focus on differences or incompatibility of two propositions, often by weighting one as more important than the other. We have not assigned ATTRIBUTIONS to any group because they represent the purely formal action of marking reported speech, without additional rhetorical effect.

Since we are interested in how a Conflict is embedded in the text structure, we also compare the distribution of discourse relations within paragraphs. Thus we compare paragraphs with at least one Conflict annotation to those having no Conflict annotation.

[4]Two speeches were missing in the UNSC-RST at the time we conducted the experiments described in this paper.

We assume that diplomats use more pragmatic RST relations for Conflicts than for Non-Conflicts, because speakers can use pragmatic relations to motivate their criticism of another party, and to strengthen potential coalitions against the criticized position. They can also appeal to the criticized country to change their behavior or to take/refrain from a particular action. The results on relation distribution are in section 5.1.

### 4.2 Analyzing the Tree Structure: Nuclearity Mass Distribution

Besides relation distribution, we inspect the tree structure resulting from the RST annotation. The *central nucleus* (CN) is interpreted as the central statement of the text covered by the tree, and can be reached starting at the top of the tree by following only 'nucleus' edges towards the leaf nodes (Mann et al., 1992). Looking at the overall shape of the tree, we can distinguish between "deeper" RST trees that are centered around one core EDU to which there is a single distinctive longest path, and "flatter" trees that have several more or less equally weighted EDUs. Stede (2016) found that for short argumentative texts, deeper trees correlate with more strongly opinionated texts, in comparison to flat trees that can signal more descriptively-oriented text. Making use of the Conflict annotation for the analysis, we were interested in a potential difference between RST trees used for paragraphs with a high proportion of Conflicts versus Non-Conflicts. We look at two levels for the analysis:

**Topics** The UNSC Conflicts corpus includes two topics, each with a different potential for Conflict. The first topic encompasses debates from 2014 about the Ukraine crisis ("Ukraine"), dealing with military conflict in which there are opposing conflicting parties. The second topic encompasses the Women, Peace and Security ("WPS") agenda, dealing with norm debates. Generally, the Ukraine debates have a more confrontational nature, whereas the WPS debates are largely about reporting on the current situation. Therefore, we expect the Ukraine debates to be more argumentative than the WPS ones.

**Countries** We compare speeches given by the permanent members of the UNSC: China, France, Russian Federation, United Kingdom, and the Unites States of America. For the Ukraine agenda, we additionally include speeches given on behalf of Ukraine.

We evaluate two methods to analyze the tree structures described in (Stede, 2016), who used it for the depth of argumentation on a small-scale analysis, and adapt the methods on a larger scale for Conflicts in diplomatic speech. More precisely, we describe two methods for characterizing the depth of an RST tree, both based on the so-called Nuclearity Mass (NM) distribution (Stede, 2016). The first Nuclearity Mass (NM1) value considers solely the number of central nodes, whereas the second Nuclearity Mass (NM2) also takes into account the distance of each node from the root. Central nuclei (CNs) are those EDUs that have zero or one satellite relations on the path from the leaf EDU node to the root of the tree.[5]

(1) NM1 describes the proportion of CNs to all leaf nodes. For example, the set of leaf nodes in Fig. 1 consists of four EDUs with two CNs. The NM1 value for this tree is therefore 0.5 (2/4).

(2) NM2 additionally includes the length of the path from the leaf node up to the root ($l_i$). NM2 is the sum of $l_i$ of the CNs, divided by the sum of all $l_i$. In the example, the root node of the subtree comprises EDUs 1-3. The $l_i$ value for CNs is 13 (4+5+4); the $l_i$ value for the full subtree is 16 (4+5+4+3). Given the multinuclear relations in this tree (EDUs 1-3), the NM2 value is 0.81 (13/16).

## 5 Results and Discussion

### 5.1 Relation Distribution

In this section, we discuss the overlap of Conflict types and the frequency of RST relations when only considering leaf nodes (Fig. 2 and 3) and inside a paragraph (Fig. 4). Note that in Fig. 2 we did not include relations that indicate mere textual organization (such as SAME-UNIT) or that are too infrequent (less than 10 occurrences both for leaf nodes and paragraphs). We merged the causal relations REASON-N and REASON (to REASON) because they only differ in how they weight two EDUs, i.e. whether the cause is more important than the reason or the other way around. Similarly,

[5]Following Stede (2016), we allow one satellite relation for CN, since we often encounter pairs of EDUs where the satellite elaborates the nucleus but still is strongly connected to the content of the nucleus (i.e., not digressing).

we merge EVALUATION-N and EVALUATION-S (to EVALUATION).

**Attribution:** Looking at ATTRIBUTION relations in Fig. 3, we notice a high proportion of *Challenging* (18.29%) and *Correcting* (6.29%) Conflicts. The high frequency of this relation is to some degree expected since *Challenges* are questioning the truthfulness of statements by another party and therefore are also reporting on what someone has (allegedly) said. *Corrections* are correcting an allegedly false statement, potentially citing a source of information (recall that ATTRIBUTIONs mark reported speech).

**Pragmatic Relations:** In section 4.1, we speculated that diplomats use more pragmatic relations for Conflicts than *Non-Conflicts* because these discourse relations describe the argumentation of the speaker, like justifying a thesis that the author has proposed (EVIDENCE, REASON), or evaluating a state of affairs from the author's perspective (EVALUATION). In fact, EVALUATION is slightly more often used in *Direct_NegEval* (2.06%) than for *Non-Conflicts* (1.52%), and EVIDENCE appears more often in *Indirect_NegEval* (1.13%) than in *Non-Conflicts* (0.93%) (Fig. 3). Nevertheless, Conflicts in general are less often annotated with EVALUATION or other pragmatic relations relations than *Non-Conflicts* (3.6% pragmatic relations in Conflicts, 5.52% in *Non-Conflicts*) (Fig. 2).

When including the upper levels of the tree (Fig. 4), we see that paragraphs with *NegEval* Conflicts have only slightly more occurrences of RST relations expressing a justification with EVIDENCE and EVALUATION than paragraphs without Conflicts. Nevertheless, REASONS are found more often for *Non-Conflicts* than for Conflicts.

**Contrastive Relations:** Contrastive relations are generally more frequently used in Conflicts than in *Non-Conflicts* (Fig. 2) (4.02% versus 2.5%). Looking at the Conflict types in more detail (Fig. 3), we see that especially *Challenge* and *Correction* have a high proportion of ANTITHESIS and CONTRAST relations, which focus on the difference (CONTRAST) or incompatibility (ANTITHESIS) of two statements, and therefore the co-occurrence is to be expected. For *Direct_NegEval* we see a peak for CONCESSION, which compares two incompatible states of affairs while regarding the content of one (the nucleus) more important than the other.

**Multinuclear and Semantic Relations:** We observe a high peak for CONJUNCTIONS for Conflicts and *Non-Conflicts*, which marks an enumeration and expresses otherwise little extra meaning. Semantic relations describing, for example, local or temporal CIRCUMSTANCES, causal relations expressing RESULT or PURPOSE appear proportionally more often in Conflicts, especially in cases marked as *Direct_NegEval*.

Summarizing these results, we discuss possible first interpretations of the rhetorical strategies we can discern from the relation distribution analysis. For a more extensive discussion, we would need more qualitative analysis involving domain experts, to be able to generalize what the relation distribution could implicate for rhetorical strategies used in the UNSC.

Contrary to our hypothesis that Conflicts are more often justified than no Conflicts, in our corpus, pragmatic/justifying relations such as EVALUATION or REASON occur with similar frequencies in texts that do or do not contain Conflicts. On the other hand, we see some semantic relations, such as E-ELABORATION and PURPOSE, more often used with Conflicts than for *Non-Conflicts*. Looking into the speeches, we find that within a Conflict statement, often not only the actions of others are criticized, but especially the ascribed intention of the actors performing the action. These cases are annotated as PURPOSE, which could explain the generally high frequency of this relation.

Further, contrastive relations are more frequently used for Conflicts than for no Conflicts. We saw in a first qualitative study for the WPS debates that diplomats frequently place a positive statement in front of a direct critique that is then contrasted with the latter. Our annotators often used CONTRAST or CONCESSION to relate those two parts, which can indicate a rhetoric strategy to de-emphasize the verbalized critique. Again, these observations will need to be doublechecked with domain experts and tested on more data, but we include them here to exemplify what kind of analysis our corpus potentially enables.

### 5.2 Tree Structure Analysis

#### 5.2.1 Nuclearity Mass and Tree Size

At first, we computed the NM1 and NM2 values for complete RST trees, but after some consideration, looked at subtrees within paragraphs instead. The

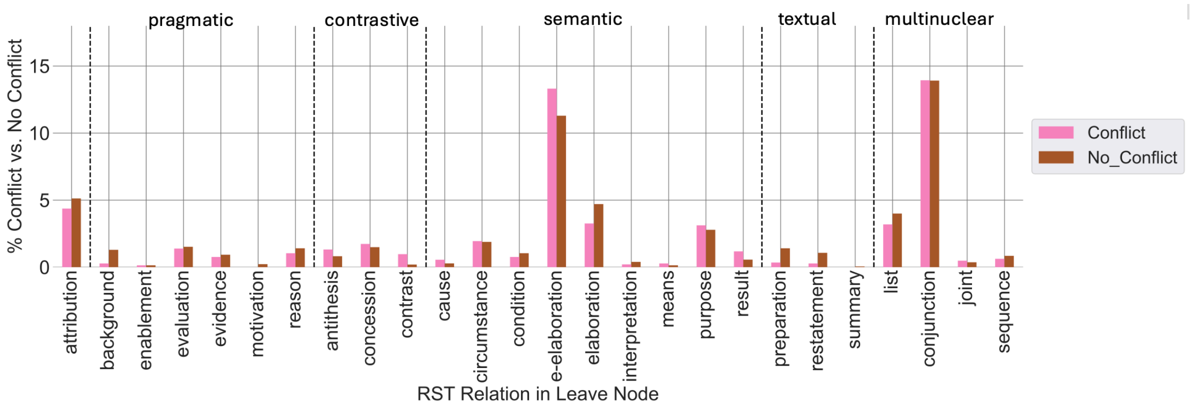



Figure 2: Normalized frequency of RST relations. The relations are grouped by their function.

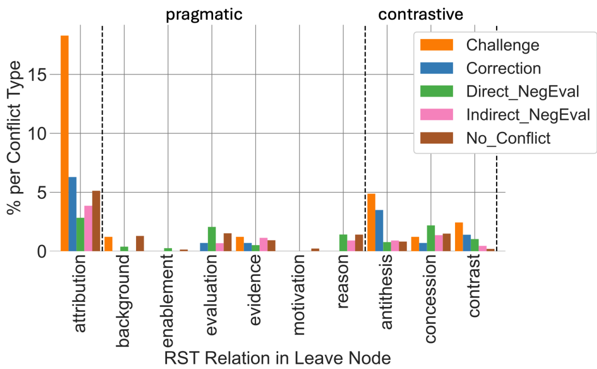



Figure 3: Normalized frequency of RST relations per Conflict type in leave nodes.

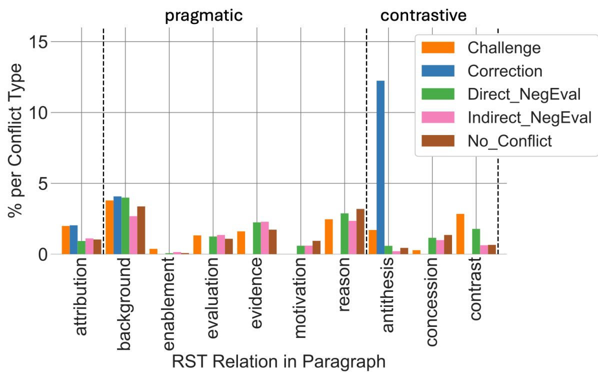



Figure 4: Normalized frequency of RST relations in paragraphs.

reason is that the NM value is sensitive to the size of the tree. In practice, annotators tend to establish a hierarchy between two EDUs, and choose multinuclear relations much less frequently (often for listings). Only multinuclear relations, which assign an equal weight to discourse units, lead to multiple CNs. As a consequence, we observe that the larger the tree, the smaller the NM value. Since the speeches in the UNSC Corpus have a large variety of tree length (see Section 3.3), this observation is especially important for our UNSC-RST.

To quantify this, the standard deviation for number of EDUs per speech/entire tree is 42.67, and for the number of CNs per speech it is six times lower (7.0). Looking at the same values for paragraphs, the standard deviation for EDUs per paragraph is 5.6, and for CNs it is 2.1, which is only 2.7 times lower. Since both NM measures are based on the ratio of leaf nodes to CNs, we decided to continue inspecting subtrees at the paragraph level in order to achieve better comparability of the trees.

#### 5.2.2 Results and Discussion Nuclearity Mass

For a paragraph to be labeled as *Conflict*, we define that at least one third of the EDUs in the paragraph should be marked with one of the Conflict types. Otherwise, the paragraphs are marked as *Non-Conflict*. Note that for the analysis of discourse relation distribution in paragraphs (Section 5.1), only one EDU had to contain a Conflict type to be marked as Conflict, since Conflict types are too sparsely distributed to establish a higher threshold.

**Topics:** Broadly comparing the values for both measures NM1 and NM2, we see that they show similar results, but the NM2 values are generally smaller than NM1 values. Looking at Figure 5 on the left, showing the distribution NM values using both measurements, we see that the values for NM1 are higher than for NM2, but both measurements show that the NM distribution is slightly lower for Ukraine than for WPS. The fact that the WPS debates have more discourse units of equal importance is in line with our expectations, as the

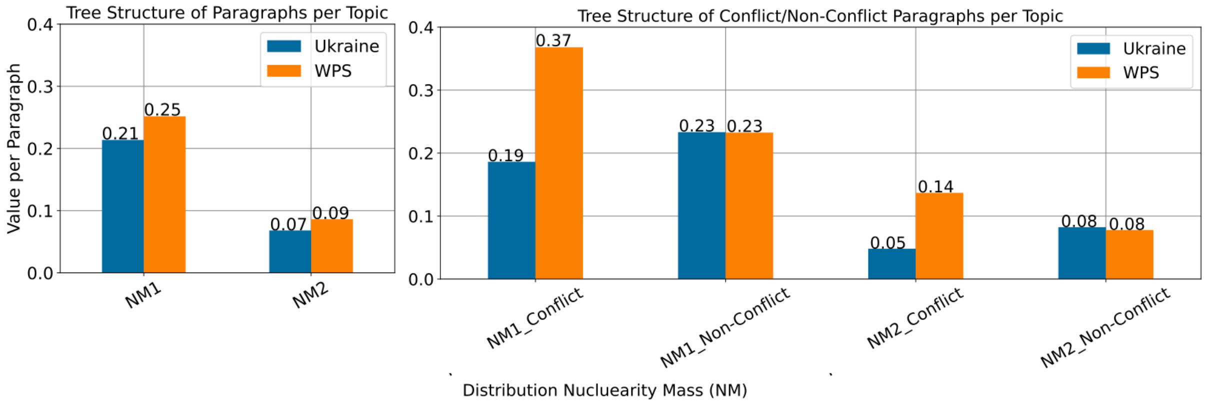



Figure 5: Mean Distribution of NM for Ukraine (194 Conflicts paragraphs, 271 Non-Conflicts) and WPS debates (16 Conflict paragraphs and 97 Non-Conflicts).

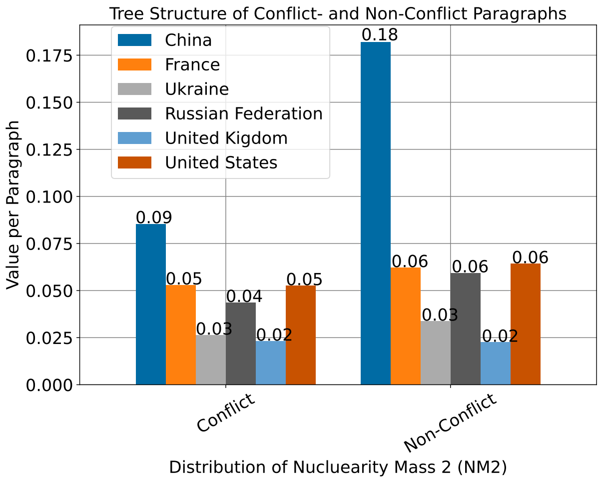



Figure 6: Mean Distribution of NM per Country, comparing Conflict versus Non-Conflict paragraphs.

WPS debates are often about summarizing what has been achieved in terms of gender and security issues and the situation in different countries.

**Topics and Conflicts:** Interestingly, comparing the topics with Conflicts versus *Non-Conflicts* paragraphs, we see that the difference between the topics is only in the Conflict, and that paragraphs with low proportion of Conflict types have similar NM Density values for both topics and both measures NM1 and NM2 (0.23 NM1 and 0.08 NM2 for both topics). One possible explanation would be that the Conflicts in WPS are rhetorically embedded and there is not one central message to which all the discourse units are leading (0.37 for NM1 and 0.14 NM2 for Conflict respectively). For Ukraine, on the other hand, it seems to be the opposite, with smaller values of 0.19 NM1 and 0.05 NM2 in Conflicts for Ukraine, and therefore having deeper tree structures towards one EDU. Whether this means that the Conflicts in Ukraine are formulated with more intensity must be assessed by political scientists, but it would be a possible conclusion of the tree structures that we find.

**Countries and Conflicts:** Since, as mentioned above, both NM measures show similar values, just on a different scale, we will only look at the NM2 value for the statistics by country (Figure 6, the bar charts for both NM1 and NM2 are in Appendix D). The countries we compare are Ukraine (37 Conflict paragraphs, 36 Non-Conflicts), Russian Federation (29, 40), USA (32, 30), China (4, 18), United Kingdom and Northern Ireland (17, 27), and France (16, 28).

We see that the speeches given by China show the highest distribution of NM2 for both Conflicts and Non-Conflicts, which is insofar interesting as the diplomatic style of the Chinese government until the late 2010s is in fact known as using cooperative rhetoric and avoiding controversy (Yuan, 2023). We also notice a comparably large distance between the average Conflict (0.09) and Non-Conflict (0.18) values in the evaluated Chinese speeches in comparison to other speeches. This might point to a greater style change when expressing critique for the Chinese speeches than for other countries, using more non-argumentative style for *Non-Conflicts* and more argumentative for Conflicts. Nevertheless, we are looking only at four Conflict paragraphs for China, and we would need a larger corpus for greater validity.

All countries have lower NM values than China, with the lowest for both Conflicts and *Non-Conflicts* for Ukraine and the United Kingdom. This indicates an argumentative style that is focusing on one or a few statements and being more argumentative, in contrast to China. Also in Contrast to China, for Conflicts, the distribution of NM is almost similar to that of *Non-Conflicts*. This may indicate that the countries are not changing their rhetorical style when expressing Conflict as much as might be expected. Also for Conflict and *Non-Conflict*, the highest value for both is that of China, followed by the Russian Federation, France and the United States, and finally by Ukraine and the United Kingdom with the lowest NM values.

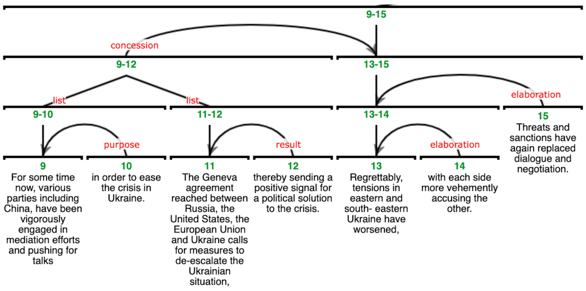



Figure 7: RST Paragraph with EDUs 13-15 being a Conflict (*Direct_NegEval*) with NM1 0.64 and NM2 0.1 (S/PV.7138, China).

## 6 Conclusion

We present a new corpus with RST annotations on 82 speeches given in the UNSC, aligned with Conflict annotations from the UNSC Conflict Corpus. We report an average inter-annotator agreement of 0,44. By jointly analyzing RST trees and Conflict annotations, we demonstrate how rhetorical analysis can help characterizing the verbalized disagreements or critiques as being more argumentative or having a more narrative style. Comparing paragraphs that contain Conflicts with those that do not, we see that the former on average have rhetorical structures that focus on a central statement, rather than having several statements of same importance. Comparing speeches of six countries in the Council, we only see a larger difference between Conflicts and *Non-Conflicts* for the Chinese speeches. When comparing values between countries, they maintain their rhetorical style, with China always having the flattest, and the United Kingdom the most centralized rhetorical structure.

We see the work presented here as one of the first to use RST to analyze the rhetorical style of diplomats. More generally, we contribute to exploring ways of using RST trees in the analysis of a multilayer corpus. In future work we want to expand not only the corpus with more topics and speeches, but also the set of analysis methods. For example, we will have a closer look at patterns of rhetorical relations, and whether some relations co-occur more often than others, which might yield more insights on rhetorical strategies used by diplomats. Based on our presented tree structure analysis, it would also be interesting to compare trees that contain an EDU marked as Conflict as their central nucleus (and thus highlight the criticism) with trees where the Conflicts are hidden in higher parts of the trees (which might serve to weaken it). Our analyses show promising results, and open up a new direction of research, combining Conflict annotations (which are less time-consuming to obtain than RST trees) with manually evaluated and corrected RST parser output, in order to investigate on larger scale in potential future work.

## Limitations

For the analysis, we work with speeches translated into English, which may introduce a bias in the analysis of rhetorical structures, as the annotators pay close attention to linguistic subtleties in order to extract the discourse relationship between text segments. When comparing the rhetorical styles of diplomatic speeches, we need to be aware that the style of individual diplomats can also bring about a change in the strategies we see. In order to analyze this, and rhetorical style in general on a larger scale, we would need more data. The relatively small corpus size is due to the time-consuming process of annotating the RST trees, which took over 5 months. To accelerate the process, we plan to evaluate the performance of RST parsers trained on the latest version of the GUM corpus, which includes political speeches.

## Acknowledgments

We thank our annotators Dietmar Benndorf (main annotator), Louis Cho, Elisa Lübbers, Hugo Meinhof, and Costanza Rasi for their work and valuable feedback on the annotation guidelines. Furthermore, we would like to thank the reviewers for their engaged and helpful comments. The research was supported by the Deutsche Forschungsgemeinschaft (DFG), project (448421482) "Trajectories of Conflict: The Dynamics of Argumentation in the UN Security Council“.

## References


Nouf Ibrahim Altmami and Mohamed El Bachir Menai. 2020. CAST: A Cross-Article Structure Theory for Multi-Article Summarization. *IEEE Access*, 8:100194–100211.

Lynn Carlson and Daniel Marcu. 2001. Discourse Tagging Reference Manual. Technical Report ISI-TR-545, University of Southern California Information Sciences Institute.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2001. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2002. RSTDiscourse Treebank LDC2002T07. Linguistic Data Consortium. Web Download.

Freya Hewett. 2023. APA-RST: A Text Simplification Corpus with RST Annotations. In *Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023)*, pages 173–179, Toronto, Canada. Association for Computational Linguistics.

Freya Hewett, Roshan Prakash Rane, Nina Harlacher, and Manfred Stede. 2019. The utility of discourse parsing features for predicting argumentation structure. In *Proceedings of the 6th Workshop on Argument Mining*, pages 98–103, Florence, Italy. Association for Computational Linguistics.

Mikel Iruskieta, Iria da Cunha, and Maite Taboada. 2015. A qualitative comparison method for rhetorical structures: identifying different discourse structures in multilingual corpora. *Language Resources and Evaluation*, 49(2):263–309.

Naoki Kobayashi, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. 2021. Improving Neural RST Parsing Model with Silver Agreement Subtrees. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1600–1612, Online. Association for Computational Linguistics.

Mathias Kraus and Stefan Feuerriegel. 2019. Sentiment analysis based on rhetorical structure theory:Learning deep neural networks from discourse trees. *Expert Systems with Applications*, 118:65–79.

Yang Janet Liu and Amir Zeldes. 2023. Why Can't Discourse Parsing Generalize? A Thorough Investigation of the Impact of Data Diversity. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3112–3130, Dubrovnik, Croatia. Association for Computational Linguistics.

William C. Mann, Christian M.I.M. Matthiessen, and Sandra A. Thompson. 1992. Rhetorical structure theory and text analysis. In William C. Mann and Sandra A. Thompson, editors, *Pragmatics & Beyond New Series*, volume 16, page 39. John Benjamins Publishing Company.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281. Publisher: De Gruyter Mouton Section: Text & Talk.

Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty, and Xiaoli Li. 2021. RST Parsing from Scratch. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1613–1625, Online. Association for Computational Linguistics.

Olu Popoola. 2017. Using Rhetorical Structure Theory for Detection of Fake Online Reviews. In *Proceedings of the 6th Workshop on Recent Advances in RST and Related Formalisms*, pages 58–63. Association for Computational Linguistics.

Victoria L. Rubin and Tatiana Vashchilko. 2012. Identification of Truth and Deception in Text: Application of Vector Space Model to Rhetorical Structure Theory. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 97–106, Avignon, France. Association for Computational Linguistics.

Mirco Schönfeld, Steffen Eckhard, Ronny Patz, and Hilde van Meegdenburg. 2019. The UN Security Council debates 1995-2017. *CoRR*, abs/1906.10969.

Sophia Skoufaki. 2020. Rhetorical Structure Theory and coherence break identification. *Text & Talk*, 40(1):99–124. Publisher: De Gruyter Mouton.

Manfred Stede. 2016. Towards assessing depth of argumentation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3308–3317, Osaka, Japan. The COLING 2016 Organizing Committee.

Manfred Stede, Maite Taboada, and Debopam Das. 2017. Annotation Guidelines for Rhetorical Structure. *Unpublished manuscript*.

Kun Sun, Rong Wang, and Wenxin Xiong. 2021. Investigating genre distinctions through discourse distance and discourse network. *Corpus Linguistics and Linguistic Theory*, 17(3):599–624. Publisher: De Gruyter Mouton.

Shujun Wan, Tino Kutschbach, Anke Lüdeling, and Manfred Stede. 2019. RST-Tace A tool for automatic comparison and evaluation of RST trees. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 88–96, Minneapolis, MN. Association for Computational Linguistics.

Sandra Williams and Richard Power. 2008. Deriving Rhetorical Complexity Data from the RST-DT Corpus. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Shaoyu Yuan. 2023. Tracing China's diplomatic transition to wolf warrior diplomacy and its implications. *Humanities and Social Sciences Communications*, 10(1):837.

Karolina Zaczynska, Peter Bourgonje, and Manfred Stede. 2024. How Diplomats Dispute: The UN Security Council Conflict Corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8173–8183, Torino, Italia. ELRA and ICCL.

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, San Diego, California. Association for Computational Linguistics.

Amir Zeldes. 2017. The GUM Corpus: Creating Multilayer Resources in the Classroom. *Language Resources and Evaluation*, 51(3):581–612.

# A Appendix: Inter-Annotator Agreement

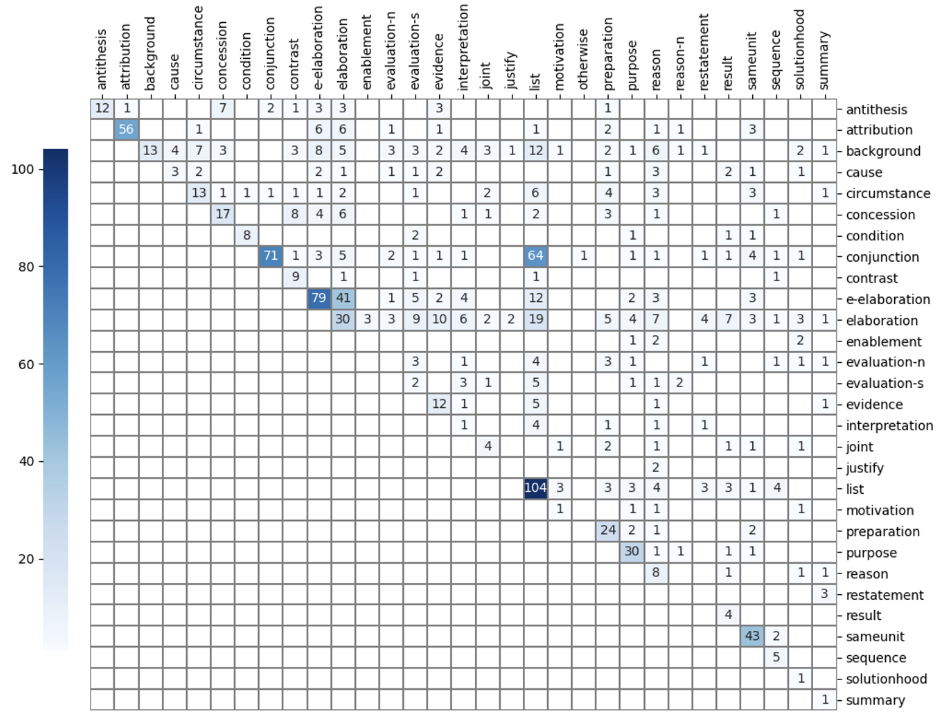
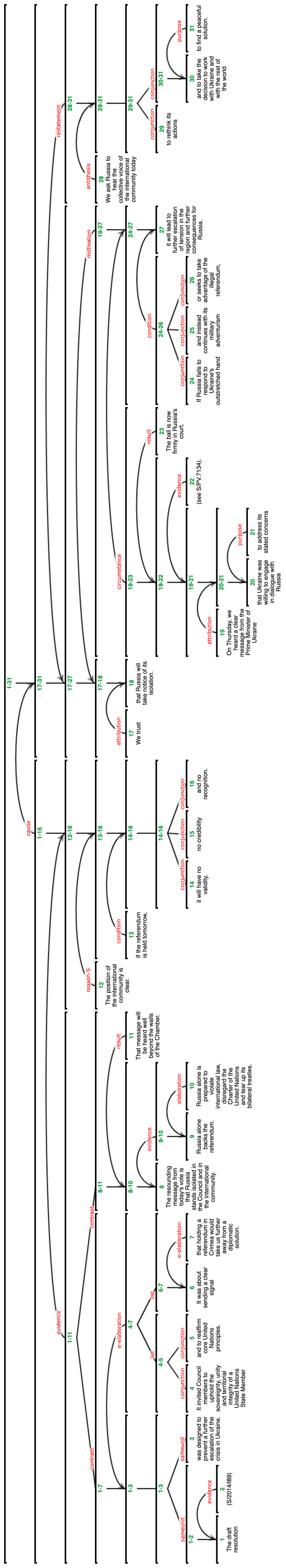



Figure 8: Confusion matrix with RST relations for two parallel annotations per RST-tree.

# B Appendix: Example RST Trees with different tree shapes

We included two example trees from the UNSC Conflicts corpus, where the first one has a clearly-identifiable central nucleus ("We trust that Russia will take notice of its isolation."). The second tree shows a tree with a higher distribution of NM with several EDUs having a multinuclear relations toward the top of the tree, and several points that are perceived as being equally important to the author of the text. For the upper tree in Figure 9, the average values per paragraph are 0.27 NM1 and 0.046 NM2; for the lower tree they are 0.64 and 0.15.



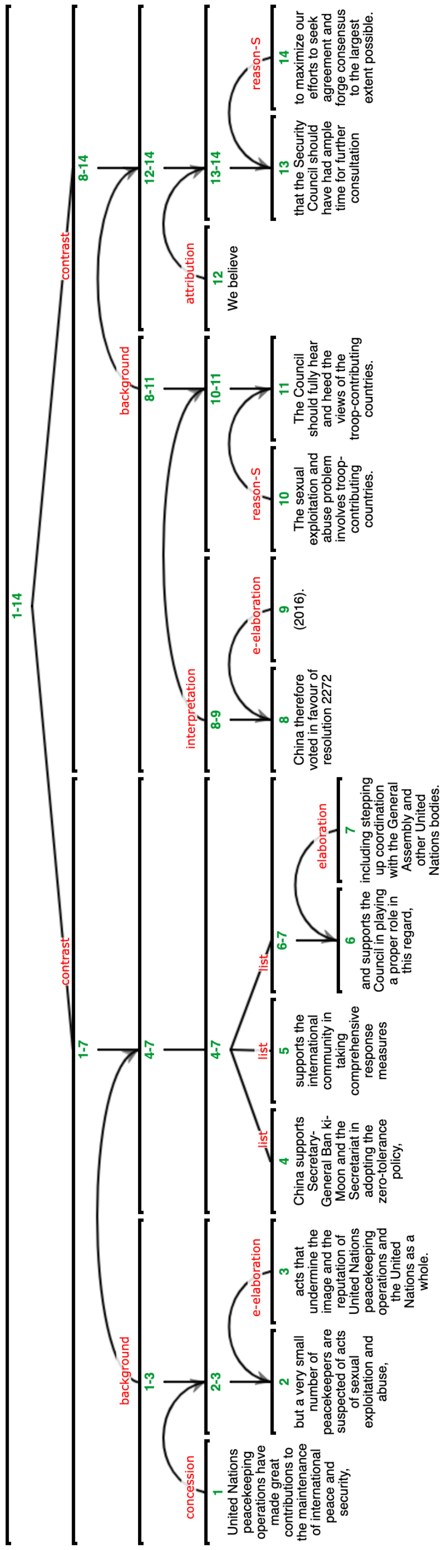



Figure 9: RST Example trees (S/PV.7138_spch006 by United Kingdom, and S/PV.7643_spch008 by China below) with different nuclearity mass distribution.

## C Statistics for RST Relation Distribution Bar Charts

| | Challenge | Correction | Direct NegEval | Indirect NegEval | Non-Conflict |
|---|---|---|---|---|---|
| paragraph #EDUs | 1,054 | 49 | 12,864 | 3,314 | 12,299 |
| leaf nodes #EDUs | 82 | 143 | 776 | 441 | 3,550 |

Table 1: Number of EDUs per Conflict Type

## D Nuclearity Mass per Country for both Measures NM1 and NM2

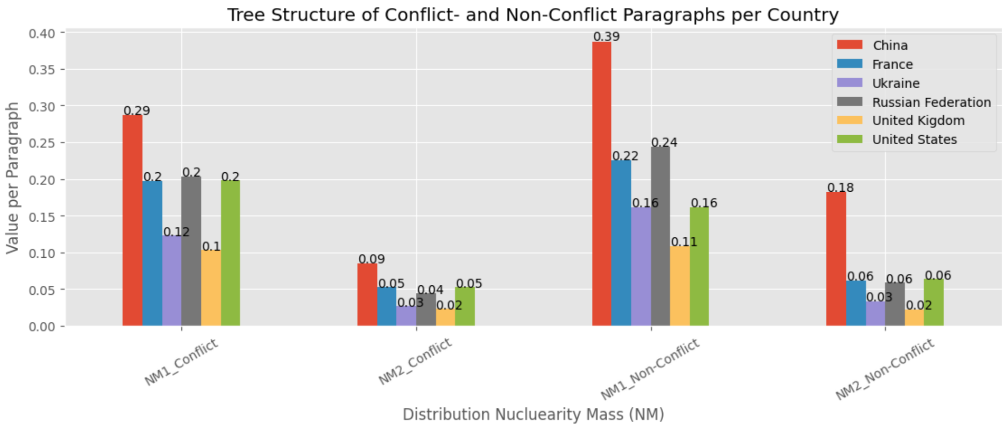



Figure 10: Mean Distribution of NM per Country, comparing Conflict versus Non-Conflict paragraphs.