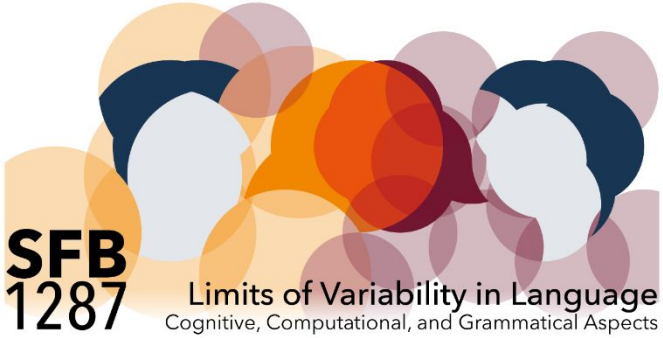




# Disagreements in analyses of rhetorical text structure: A new dataset and first analyses

*Freya Hewett & Manfred Stede, University of Potsdam*

31.07.2025

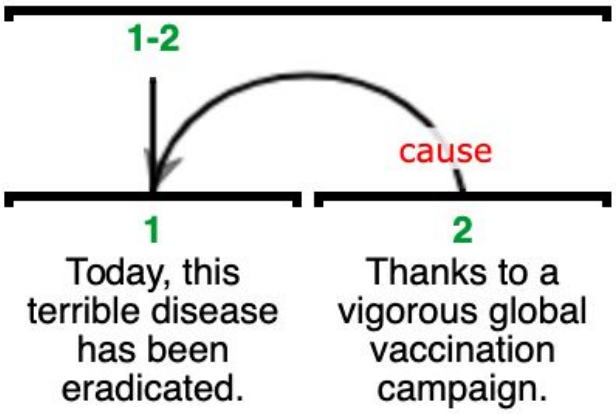
# Contributions

- new **dataset** of English and German texts with two parallel analyses
  - *total of 156 texts*
- statistics on **conflicting annotation** decisions
  - *2919 text segments categorised automatically*
- a **typology of reasons** for these conflicting annotations
  - *480 instances of disagreement analysed manually*

# Dataset (I)

Rhetorical Structure Theory (RST; Mann & Thompson, 1987)

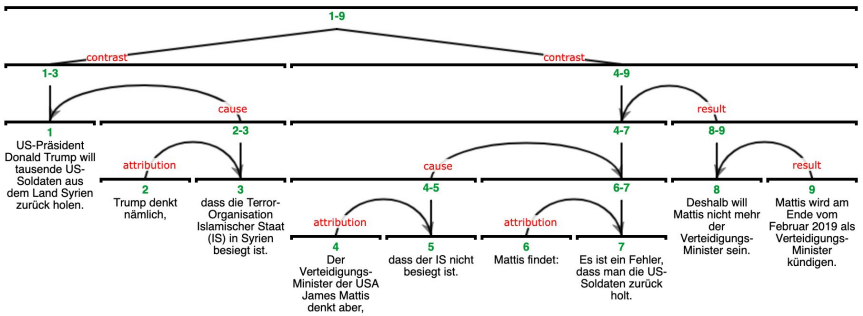
1. Text is segmented into “minimal discourse units”
2. Coherence relations between the units are determined
3. Within each relation, a nucleus and satellite is determined



# Dataset (II)

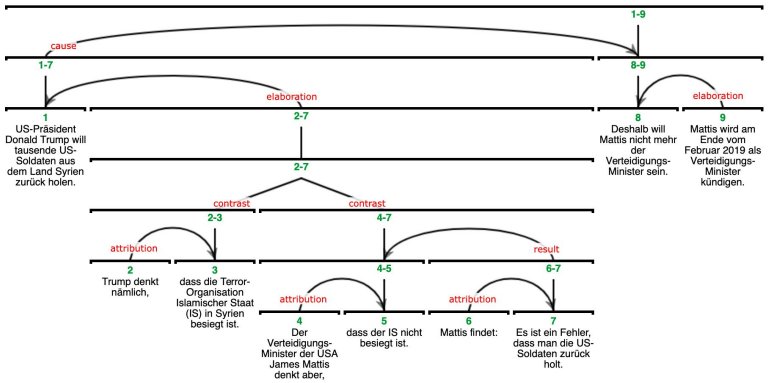
Rhetorical Structure Theory (RST; Mann & Thompson, 1987)

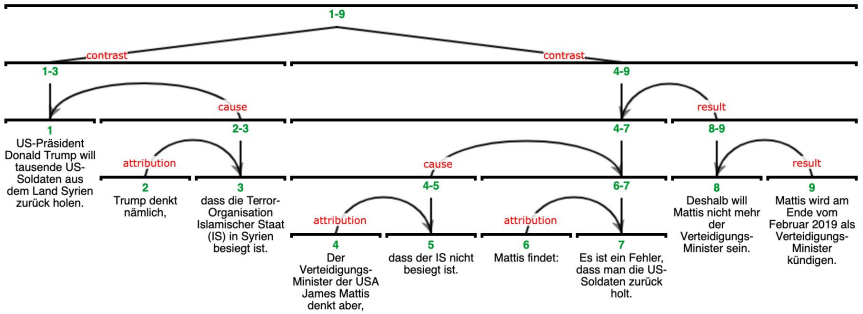
1. Text is segmented into minimal discourse units
2. Coherence relations between the units are determined
3. Within each relation, a nucleus and satellite is determined
4. Units and spans of units are annotated iteratively until a tree is formed for the whole text



# Dataset (III)

- RST annotations consist of multiple components: *relations, nuclearity, span (attachment point, constituents)*…
- They represent the annotator's interpretation of the text
- … lots of room for disagreement



# Dataset (IV)

- Our dataset consists of 156 English and German texts, where each text has two parallel RST analyses

- English:
  - **UNSC-RST** (Zaczynska & Stede, 2024): transcripts of speeches from the UN Security Council
  - **RST-DT** (Carlson et al., 2003): articles from the Wall Street Journal

- German:
  - **APA-RST** (Hewett, 2023): simplified newspaper articles
  - **PCC*** (Shahmohammadi & Stede, 2024): commentaries from local newspapers

# Conflicting annotations (I)

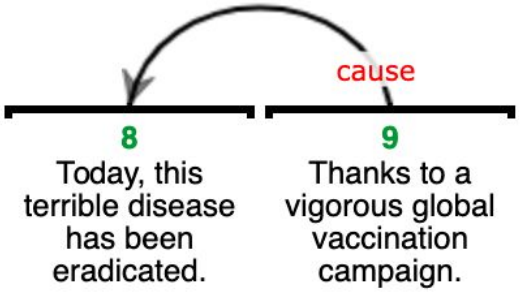
RSTTace (Wan et al., 2019):

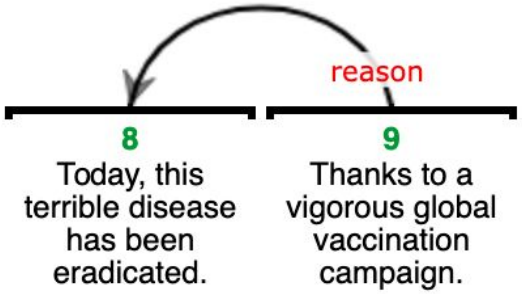
- tool used to calculate inter-annotator agreement for RST annotations
- provides a table describing the “matching” spans

# Conflicting annotations (I)

RSTTace (Wan et al., 2019):

- tool used to calculate inter-annotator agreement for RST annotations
- provides a table describing the “matching” spans




| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | | 6 | condition | ← | 6 | 6 | S | 5 | 5 | N | | 6 | condition | ← | 6 | 6 | S | 4 | 5 | N | Completely identical CS | ✓ | ✓ | ✓ | ✗ | NRC | |
| 7 | | 7 | cause | ← | 7 | 7 | S | 3 | 6 | N | | 7 | reason | ← | 7 | 7 | S | 3 | 6 | N | Completely identical CS | ✓ | ✗ | ✓ | ✓ | NCA | ≠R |
| 8 | | 9 | reason | ← | 9 | 9 | S | 8 | 8 | N | | 9 | cause | ← | 9 | 9 | S | 8 | 8 | N | Completely identical CS | ✓ | ✗ | ✓ | ✓ | NCA | ≠R |
| 9 | | 11-13 | result | ← | 11 | 13 | S | 10 | 10 | N | | 11-13 | elaboration | ← | 11 | 13 | S | 10 | 10 | N | Completely identical CS | ✓ | ✗ | ✓ | ✓ | NCA | ≠R |
| 10 | | 12 | elaboration | ← | 12 | 12 | S | 11 | 11 | N | | 12 | elaboration | ← | 12 | 12 | S | 11 | 11 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ | NRCA | |
| 11 | | 14-20 | reason | → | 14 | 20 | S | 21 | 29 | N | | 8-13\|14-20 | list | ↔ | 8 | 13 | N | 14 | 20 | N | Partially identical CS | ✗ | ✗ | ✗ | ✗ | | N/N-N/S, ≠R |

# Conflicting annotations (I)

RSTTace (Wan et al., 2019):

- tool used to calculate inter-annotator agreement for RST annotations
- provides a table describing the “matching” spans

| ID | CS-A | Relation-A | Nuc-A | C1-A | C2-A | CN-A | A1-A | A2-A | AN-A | CS-B | Relation-B | Nuc-B | C1-B | C2-B | CN-B | A1-B | A2-B | AN-B | Matching | N | R | C | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | preparation | → | 1 | 2 | S | 3 | 31 | N | 1 | preparation | → | 1 | 2 | S | 3 | 31 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 2 | 3-31 | evidence | → | 1 | 31 | S | 32 | 37 | N | 3-31 | evidence | → | 1 | 31 | S | 32 | 39 | N | Completely identical CS | ✓ | ✓ | ✓ | × |
| 3 | 32-37\|38-41 | joint | ↔ | 1 | 37 | N | 38 | 41 | N | | | | | | | | | | No matching | × | × | × | × |
| 4 | | | | | | | | | | 40-41 | evaluation-s | ← | 40 | 41 | S | 1 | 39 | N | No matching | × | × | × | × |
| 5 | 2 | reason | ← | 2 | 2 | S | 1 | 1 | N | 2 | reason | ← | 2 | 2 | S | 1 | 1 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 6 | 3 | background | → | 3 | 3 | S | 4 | 4 | N | 3 | background | → | 3 | 3 | S | 4 | 4 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 7 | 3-7 | evidence | → | 3 | 7 | S | 8 | 8 | N | 3-7 | evidence | → | 3 | 7 | S | 8 | 8 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 8 | 8\|9 | contrast | ↔ | 3 | 8 | N | 9 | 9 | N | 8\|9 | contrast | ↔ | 3 | 8 | N | 9 | 9 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 9 | 4\|5 | contrast | ↔ | 3 | 4 | N | 5 | 7 | N | 4\|5 | contrast | ↔ | 3 | 4 | N | 5 | 7 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 10 | 3-9\|10-24\|25 | list | ↔ | 3 | 9 | N | 10 | 31 | N | 3-9\|10\|13\|25 | list | ↔ | 3 | 9 | N | 10 | 31 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 11 | 6 | evidence | ← | 6 | 7 | S | 5 | 5 | N | 6 | evidence | ← | 6 | 7 | S | 5 | 5 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 12 | 7 | circumstance | ← | 7 | 7 | S | 6 | 6 | N | 7 | condition | ← | 7 | 7 | S | 6 | 6 | N | Completely identical CS | ✓ | × | ✓ | ✓ |
| 13 | 10-24\|25 | list | ↔ | 10 | 24 | N | 25 | 31 | N | 10\|13\|25 | list | ↔ | 10 | 12 | N | 13 | 31 | N | Partially identical CS | ✓ | ✓ | × | × |
| 14 | 10\|13 | contrast | ↔ | 10 | 12 | N | 13 | 24 | N | 13\|25 | list | ↔ | 13 | 24 | N | 25 | 31 | N | Partially identical CS | ✓ | × | × | × |
| 15 | 11 | cause | → | 11 | 11 | S | 12 | 12 | N | 11\|12 | conjunction | ↔ | 11 | 11 | N | 12 | 12 | N | C1=C2 and A1=A2 | × | × | × | ✓ |
| 16 | 12 | elaboration | ← | 11 | 12 | S | 10 | 10 | N | 11-12 | elaboration | ← | 11 | 12 | S | 10 | 10 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 17 | 14-24 | evidence | ← | 14 | 24 | S | 13 | 13 | N | 14-24 | evidence | ← | 14 | 24 | S | 13 | 13 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 18 | 14\|16-17\|19-20 | list | ↔ | 14 | 15 | N | 16 | 24 | N | 14\|16-17\|19-20\|22-24 | list | ↔ | 14 | 15 | N | 16 | 24 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 19 | 15 | elaboration | ← | 15 | 15 | S | 14 | 14 | N | 15 | elaboration | ← | 15 | 15 | S | 14 | 14 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 20 | 16-17\|19-20 | list | ↔ | 16 | 18 | N | 19 | 24 | N | 16-17\|19-20\|22-24 | list | ↔ | 16 | 18 | N | 19 | 24 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 21 | 17 | concession | ← | 17 | 17 | S | 16 | 16 | N | 17 | concession | ← | 17 | 17 | S | 16 | 16 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 22 | 18 | evaluation-s | ← | 18 | 18 | S | 16 | 17 | N | 18 | evaluation-s | ← | 18 | 18 | S | 16 | 17 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 23 | 20 | purpose | ← | 20 | 20 | S | 19 | 19 | N | 19\|20 | sequence | ↔ | 19 | 19 | N | 20 | 20 | N | C1=A2 and A1=C2 | × | × | × | × |
| 24 | 21 | concession | → | 21 | 21 | S | 22 | 24 | N | 21 | concession | → | 21 | 21 | S | 22 | 24 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 25 | 22-24 | e-elaboration | ← | 21 | 24 | S | 19 | 20 | N | 19-20\|22-24 | list | ↔ | 19 | 20 | N | 21 | 24 | N | C1=A2 and A1=C2 | × | × | × | × |
| 26 | 23-24 | evidence | ← | 23 | 24 | S | 22 | 22 | N | 23 | evidence | ← | 23 | 24 | S | 22 | 22 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 27 | 23\|24 | contrast | ↔ | 23 | 23 | N | 24 | 24 | N | 24 | elaboration | ← | 24 | 24 | S | 23 | 23 | N | C1=A2 and A1=C2 | × | × | × | × |
| 28 | 26-31 | evidence | ← | 26 | 31 | S | 25 | 25 | N | 26-31 | evidence | ← | 26 | 31 | S | 25 | 25 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 29 | 26\|27\|28\|30\|31 | list | ↔ | 26 | 26 | N | 27 | 31 | N | 26\|27\|28-30\|31 | list | ↔ | 26 | 26 | N | 27 | 31 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 30 | 27\|28\|30\|31 | list | ↔ | 27 | 27 | N | 28 | 31 | N | 27\|28-30\|31 | list | ↔ | 27 | 27 | N | 28 | 31 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 31 | 28\|30\|31 | list | ↔ | 28 | 28 | N | 29 | 31 | N | 28\|29 | list | ↔ | 28 | 28 | N | 29 | 30 | N | Partially identical CS | ✓ | ✓ | ✓ | × |
| 32 | 29 | interpretation | → | 29 | 29 | S | 30 | 30 | N | 30 | elaboration | ← | 30 | 30 | S | 29 | 29 | N | C1=A2 and A1=C2 | × | × | × | × |
| 33 | 30\|31 | list | ↔ | 29 | 30 | N | 31 | 31 | N | 28-30\|31 | list | ↔ | 28 | 30 | N | 31 | 31 | N | Partially identical CS | ✓ | ✓ | × | ✓ |
| 34 | 32 | concession | → | 32 | 32 | S | 33 | 35 | N | 32 | concession | → | 32 | 32 | S | 33 | 35 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 35 | 33-35 | antithesis | → | 32 | 35 | S | 36 | 36 | N | 33-35 | antithesis | → | 32 | 35 | S | 36 | 39 | N | Completely identical CS | ✓ | ✓ | ✓ | × |
| 36 | 36 | reason | → | 32 | 36 | S | 37 | 37 | N | 36 | reason | → | 36 | 36 | S | 37 | 37 | N | Completely identical CS | ✓ | ✓ | × | ✓ |
| 37 | 34 | circumstance | ← | 34 | 34 | S | 33 | 33 | N | 34 | circumstance | ← | 34 | 34 | S | 33 | 33 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 38 | 35 | evidence | ← | 35 | 35 | S | 33 | 34 | N | 35 | evidence | ← | 35 | 35 | S | 33 | 34 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 39 | 38 | circumstance | → | 38 | 38 | S | 39 | 39 | N | 38 | cause | → | 38 | 38 | S | 39 | 39 | N | Completely identical CS | ✓ | × | ✓ | ✓ |
| 40 | 39\|40-41 | joint | ↔ | 38 | 39 | N | 40 | 41 | N | 39 | elaboration | ← | 38 | 39 | S | 36 | 37 | N | Partially identical CS | × | × | × | × |
| 41 | 40\|41 | list | ↔ | 40 | 40 | N | 41 | 41 | N | 40\|41 | list | ↔ | 40 | 40 | N | 41 | 41 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |

| Results | Nuclearity | Relation | Constituent | Attachment point | RST trees |
|---|---|---|---|---|---|
| F-Measure: | (33 of 40) → 0.804 | (30 of 41) → 0.731 | (29 of 41) → 0.707 | (29 of 41) → 0.707 | 0.737 |
| Inter-Annotator Agreement: | 0.71 | 0.698 | 0.57 | 0.59 | 0.567 |

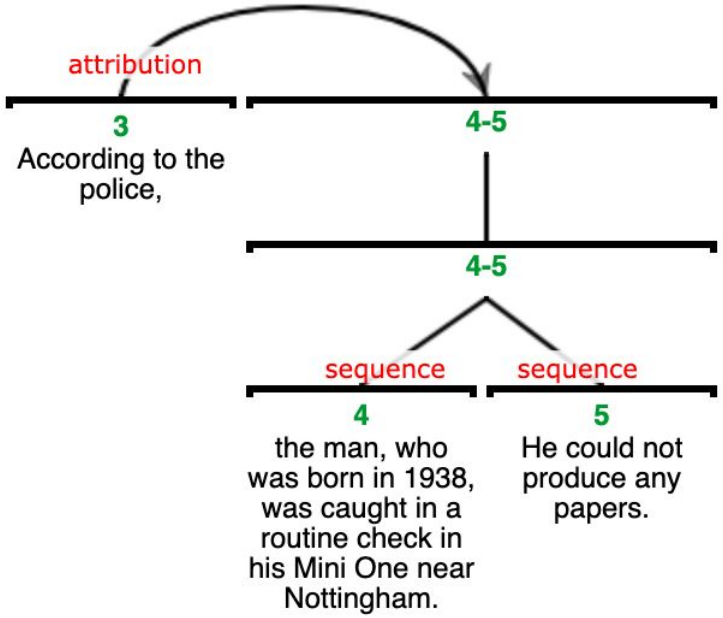
# Conflicting annotations (II)

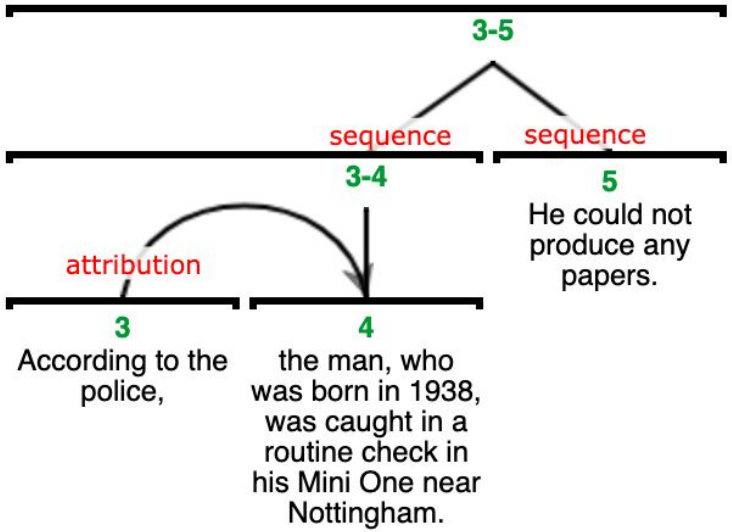
We use the output from Tace to calculate five main categories:

| **Category** | **Percentage in all subcorpora** |
|---|---|
| Perfect match | 35% |
| Relation mismatch | 20% |
| Scope mismatch | 19% |
| Left/right mismatch | 3% |
| No match | 23% |

# Conflicting annotations (III)

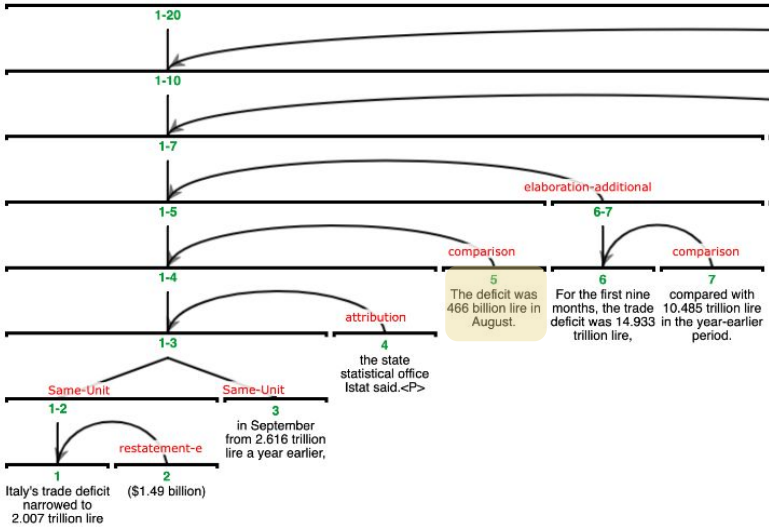
Scope mismatch:

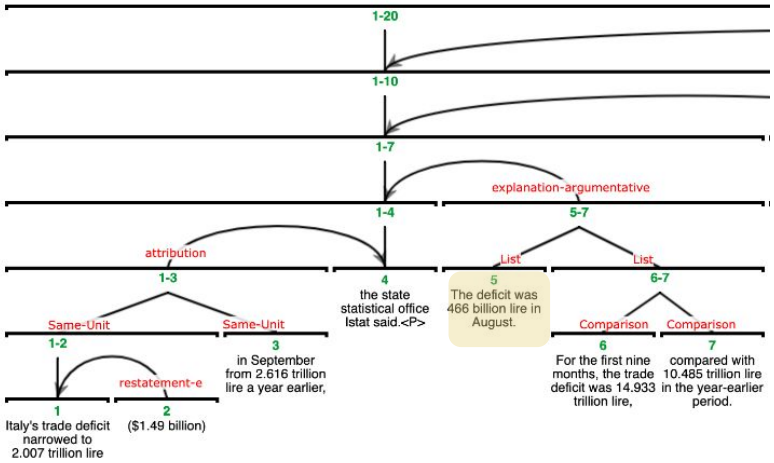


# Conflicting annotations (IV)

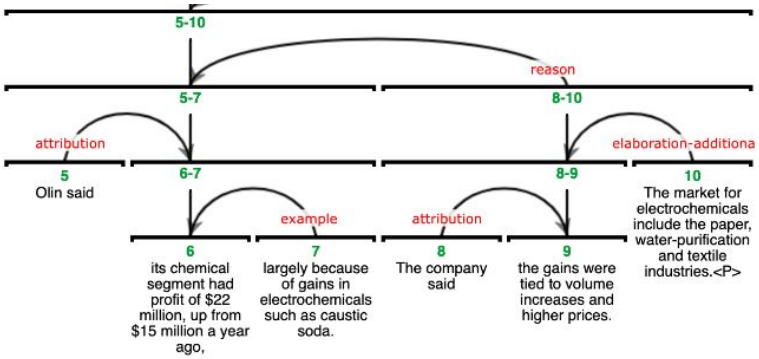
Left/right mismatch:

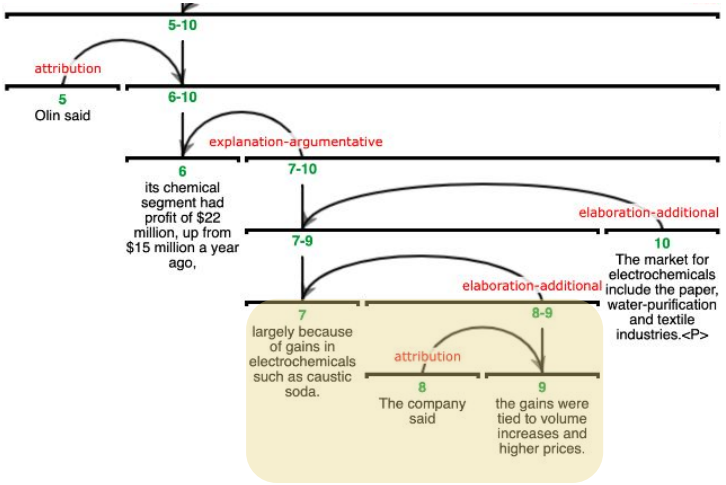


# Conflicting annotations (V)

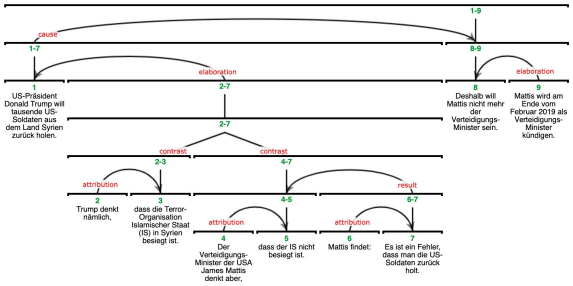
No match:

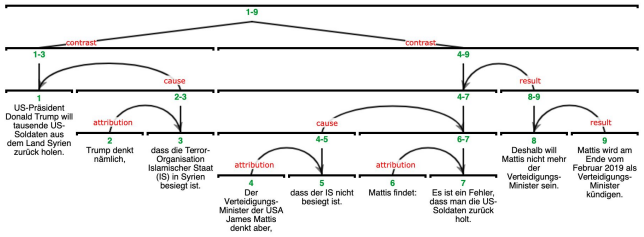


# Analysis steps

1. Two parallel annotations




2. Output from Tace + our categories

| ID | CS-A | Relation-A | Nuc-A | C1-A | C2-A | CN-A | A1-A | A2-A | AN-A | CS-B | Relation-B | Nuc-B | C1-B | C2-B | CN-B | A1-B | A2-B | AN-B | Matching | N | R | C | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | preparation | → | 1 | 2 | S | 3 | 31 | N | 1 | preparation | → | 1 | 2 | S | 3 | 31 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 2 | 3-31 | evidence | → | 1 | 31 | S | 32 | 37 | N | 3-31 | evidence | → | 1 | 31 | S | 32 | 39 | N | Completely identical CS | ✓ | ✓ | ✓ | × |
| 3 | 32-37/38-41 | joint | ↔ | 1 | 37 | N | 38 | 41 | N | | | | | | | | | | No matching | × | × | × | × |
| 4 | | | | | | | | | | 40-41 | evaluation-s | ← | 40 | 41 | S | 1 | 39 | N | No matching | × | × | × | × |
| 5 | 2 | reason | ← | 2 | 2 | S | 1 | 1 | N | 2 | reason | ← | 2 | 2 | S | 1 | 1 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 6 | 3 | background | → | 3 | 3 | S | 4 | 4 | N | 3 | background | → | 3 | 3 | S | 4 | 4 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 7 | 3-7 | evidence | → | 3 | 7 | S | 8 | 8 | N | 3-7 | evidence | → | 3 | 7 | S | 8 | 8 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 8 | 8/9 | contrast | ↔ | 3 | 8 | N | 9 | 9 | N | 8/9 | contrast | ↔ | 3 | 8 | N | 9 | 9 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 9 | 4/5 | contrast | ↔ | 3 | 4 | N | 5 | 7 | N | 4/5 | contrast | ↔ | 3 | 4 | N | 5 | 7 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 10 | 3-9/10-24/25 | list | ↔ | 3 | 9 | N | 10 | 31 | N | 3-9/10/31/325 | list | ↔ | 3 | 9 | N | 10 | 31 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 11 | 6 | evidence | ← | 6 | 7 | S | 5 | 5 | N | 6 | evidence | ← | 6 | 7 | S | 5 | 5 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 12 | 7 | circumstance | ← | 7 | 7 | S | 6 | 6 | N | 7 | condition | ← | 7 | 7 | S | 6 | 6 | N | Completely identical CS | ✓ | × | ✓ | ✓ |
| 13 | 10-24/25 | list | ↔ | 10 | 24 | N | 25 | 31 | N | 10/1325 | list | ↔ | 10 | 12 | N | 13 | 31 | N | Partially identical CS | ✓ | ✓ | × | × |
| 14 | 10/13 | contrast | ↔ | 10 | 12 | N | 13 | 24 | N | 13/25 | list | ↔ | 13 | 24 | N | 25 | 31 | N | Partially identical CS | ✓ | × | × | × |
| 15 | 11 | cause | → | 11 | 11 | S | 12 | 12 | N | 11/12 | conjunction | ↔ | 11 | 11 | N | 12 | 12 | N | C1=C2 and A1=A2 | × | × | × | ✓ |
| 16 | 12 | elaboration | ← | 11 | 12 | S | 10 | 10 | N | 11-12 | elaboration | ← | 11 | 12 | S | 10 | 10 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 17 | 14-24 | evidence | ← | 14 | 24 | S | 13 | 13 | N | 14-24 | evidence | ← | 14 | 24 | S | 13 | 13 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 18 | 14/16-17/19-20 | list | ↔ | 14 | 15 | N | 16 | 24 | N | 14/16-17/19-20/22-24 | list | ↔ | 14 | 15 | N | 16 | 24 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 19 | 15 | elaboration | ← | 15 | 15 | S | 14 | 14 | N | 15 | elaboration | ← | 15 | 15 | S | 14 | 14 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |
| 20 | 16-17/19-20 | list | ↔ | 16 | 18 | N | 19 | 24 | N | 16-17/19-20/22-24 | list | ↔ | 16 | 18 | N | 19 | 24 | N | C1=C2 and A1=A2 | ✓ | ✓ | ✓ | ✓ |
| 21 | 17 | concession | ← | 17 | 17 | S | 16 | 16 | N | 17 | concession | ← | 17 | 17 | S | 16 | 16 | N | Completely identical CS | ✓ | ✓ | ✓ | ✓ |

3. Manual analysis of annotation decisions

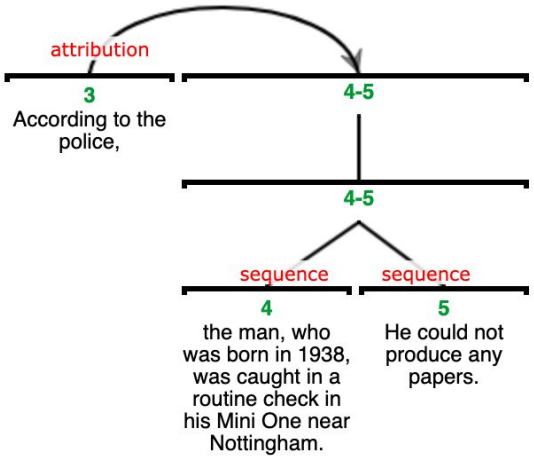
Four categories for the status of a mismatch:
Disagree, Both, Vague, Either/Or

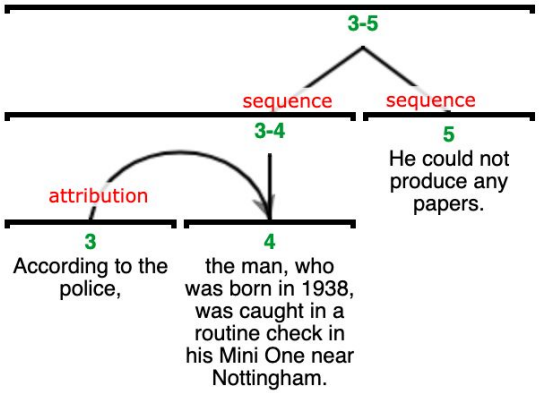
# Reasons for disagreement (I)

- Formal structural alternatives
- Relation definition overlap
- Epistemic status of propositions
- Presupposed knowledge / subjective bias
- Assignment of ‘importance’
- Text structure
- Scope of adverbial connectives etc.

# Reasons for disagreement (II)

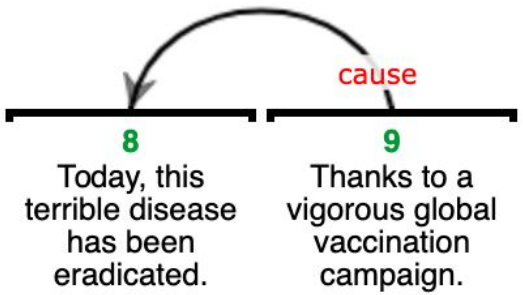
- Formal structural alternatives
- Relation definition overlap
- Epistemic status of propositions
- Presupposed knowledge / subjective bias
- Assignment of ‘importance’
- Text structure
- **Scope of adverbial connectives etc.**

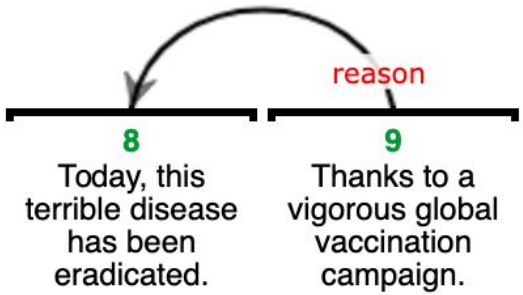


# Reasons for disagreement (III)

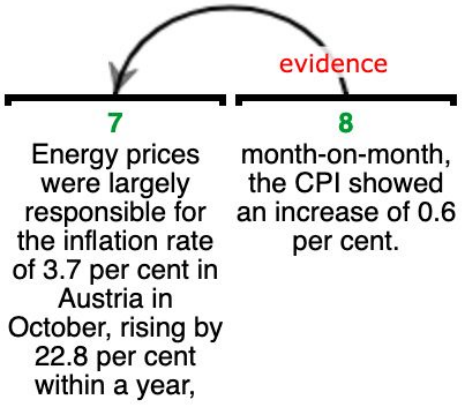
- Formal structural alternatives
- Relation definition overlap
- **Epistemic status of propositions**
- Presupposed knowledge / subjective bias
- Assignment of ‘importance’
- Text structure
- Scope of adverbial connectives etc.

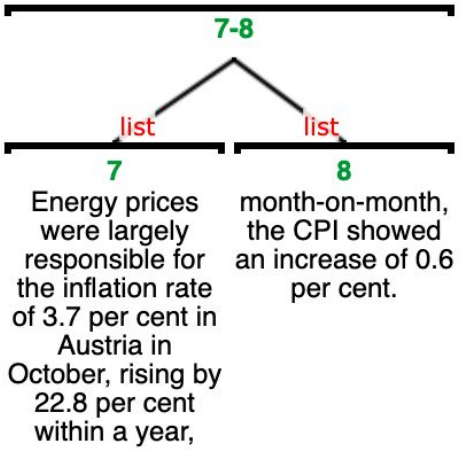


# Reasons for disagreement (IV)

- Formal structural alternatives
- **Relation definition overlap**
- Epistemic status of propositions
- **Presupposed knowledge / subjective bias**
- Assignment of ‘importance’
- Text structure
- Scope of adverbial connectives etc.



# Reasons for disagreement (V)

- Ambiguities of language (e.g. scope of attribution)
- Ambiguities of annotation guidelines (e.g. relation definitions)
- Differences between annotators (e.g. subjective bias)

# Reasons for disagreement (V)

- Ambiguities of language (e.g. scope of attribution)
- Ambiguities of annotation guidelines (e.g. relation definitions)
- Differences between annotators (e.g. subjective bias)

Implications:

- Incorporating uncertainty in discourse parsing
- Improving annotation guidelines
- Insights into coherence

# Thank you!

Parallel annotations for 130 texts plus scripts to convert Tace output:
`https://github.com/discourse-lab/RSTmulti/`

**References**


Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In van Kuppevelt, J., & Smith, R. (Ed.), *Current Directions in Discourse and Dialogue*. Kluwer, Dordrecht.

Hewett, F. (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.

Mann, W. C., & Thompson, S. A. (1987). Rhetorical Structure Theory: A Theory of Text Organization. In L. Polanyi (Ed.), *The Structure of Discourse*. Ablex Publishing Corporation.

Shahmohammadi, S., & Stede, M. (2024). Discourse parsing for German with new RST corpora. In *Proceedings of the 20th Conference on Natural Language Processing (KONVENS 2024)*, pages 65–74, Vienna, Austria. Association for Computational Linguistics.

Wan, S., Kutschbach, T., Lüdeling, A., & Stede, M. (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*, 88–96.

Zaczynska, K., & Stede, M. (2024). Rhetorical strategies in the UN security council: Rhetorical Structure Theory and conflicts. In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 15–28, Kyoto, Japan. Association for Computational Linguistics.

# Relations

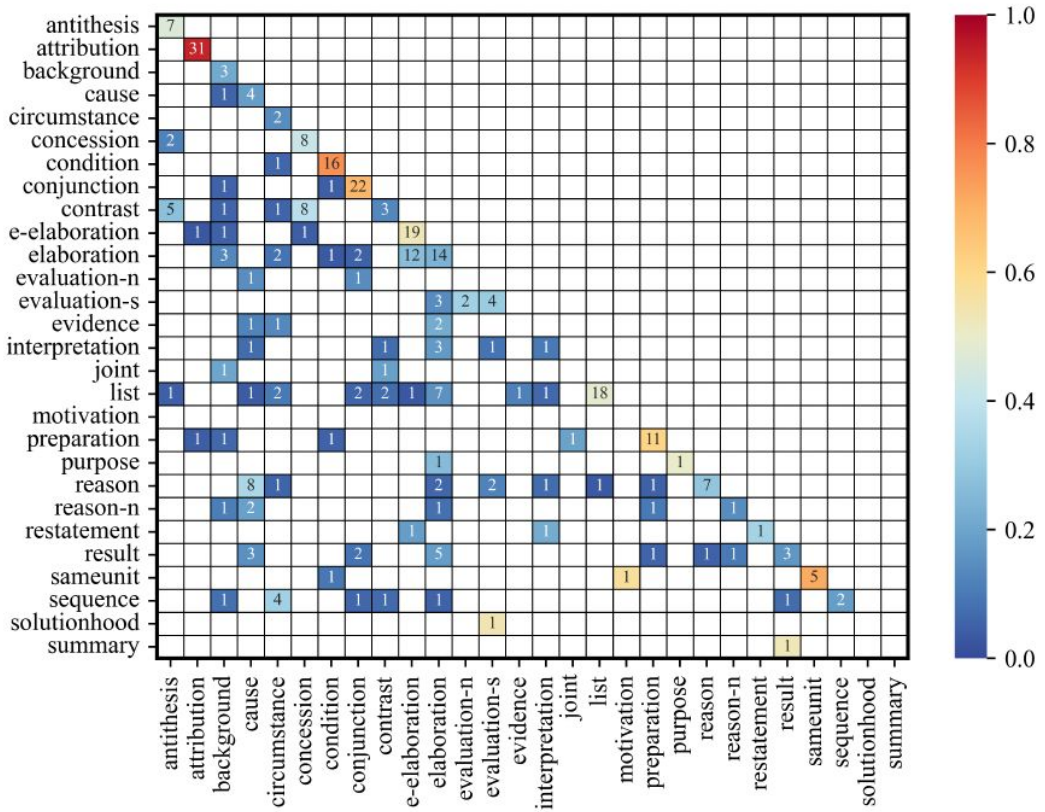



Figure 4: Relations in the categories 'Perfect match' or 'Relation mismatch' in the double annotated subsets of the German-language subcorpora (APA+PCC).