

Discourse structure in the Potsdam Commentary Corpus: Human annotation, human disagreement, and automatic parsing

IPI PAN, Warsaw, April 2025

Manfred Stede
Applied Computational Linguistics
Universität Potsdam

Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

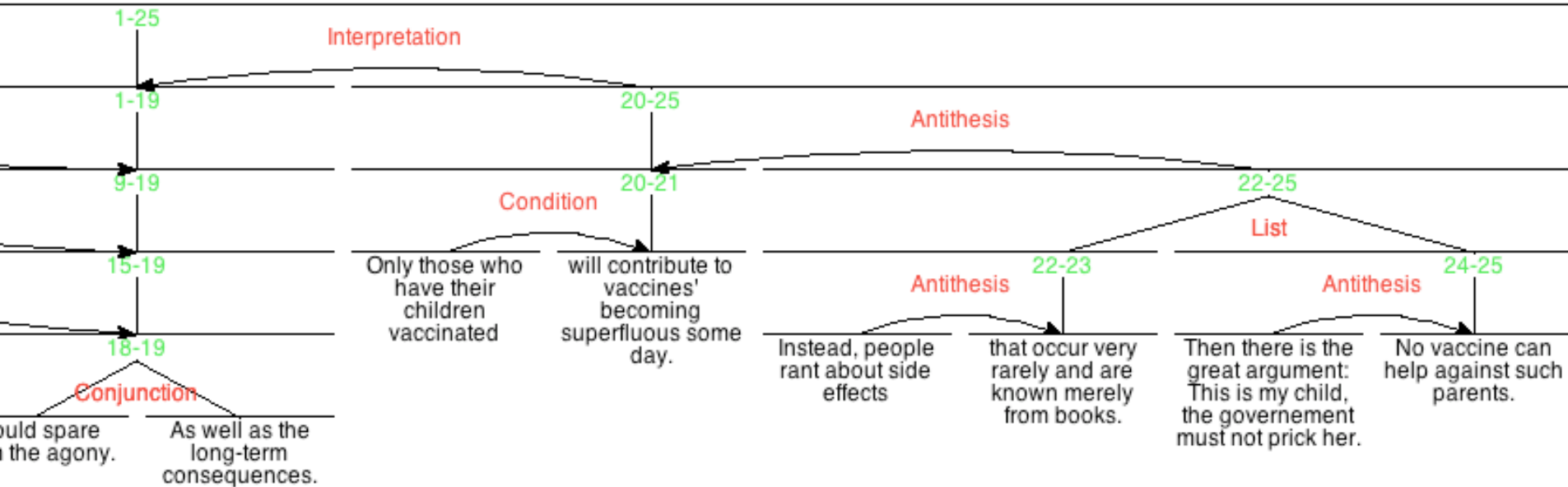
Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

Rhetorical Structure Theory

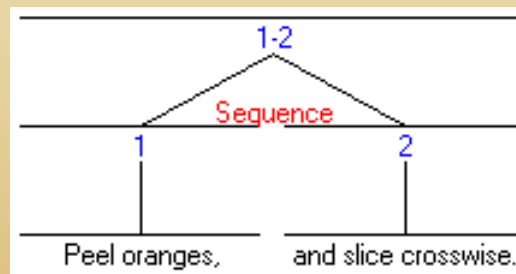
- Text is being segmented into **elementary discourse units** (EDUs)
- Adjacent units are being connected by a **coherence relation**
- Result is a **tree structure** spanning the whole text
- Almost all relations consist of a **nucleus** and a **satellite** (supporting, optional)
- Relation set is divided in
 - **subject matter** („semantic“)
 - **presentational** („pragmatic“)
 - **textual**
relations

Rhetorical Structure Theory

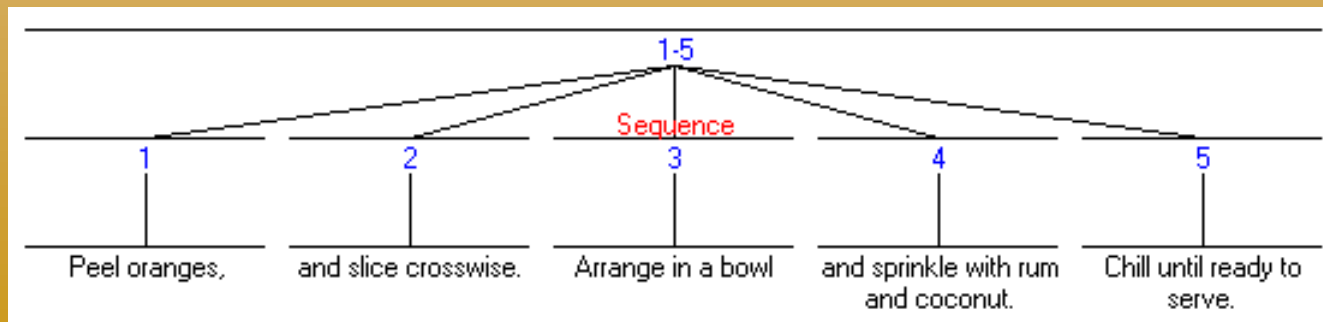


Multinuclear (coordinate) relations

- At the sub-sentential level
 - *Peel oranges, and slice crosswise.*

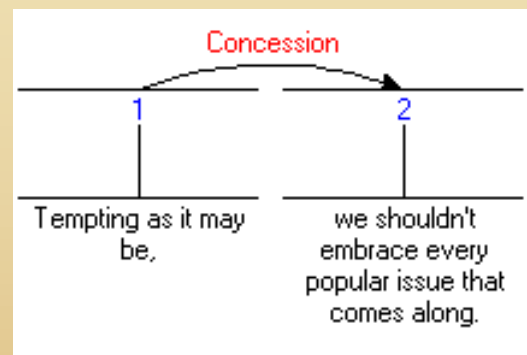


- But also across sentences
 - *Peel oranges, and slice crosswise. Arrange in a bowl and sprinkle with rum and coconut. Chill until ready to serve.*

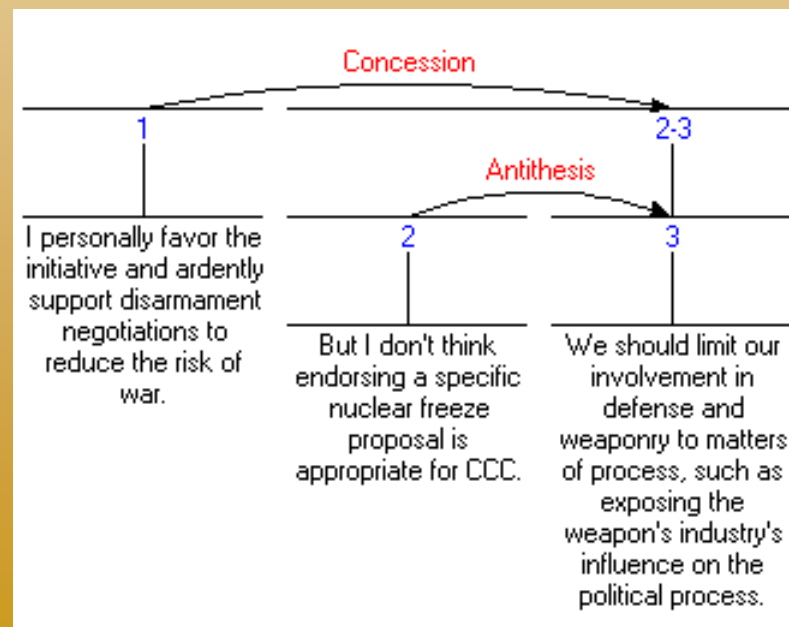


Mononuclear (subordinate) relations

- Sub-sentential **Concession** relation



- **Concession** across sentences
 - Nucleus (spans 2-3) made up of two spans in an **Antithesis** relation



Example: Evidence

EVIDENCE

- **N:** A subjective statement/thesis/claim, which R might not accept or might not regard as sufficiently important or positive.
- **S:** A statement that R is likely to accept; usually an “objective” description of a fact.
- **N/S:** Understanding S makes it easier for R to accept N, or to share the particular viewpoint of W.
- **Effect:** R’s belief in N is increased.
- **Typical connectives:** Causal connectives.
- **Example:** *Debate about having two subjects 'Religion' and 'LER' at school*
[And now even our state government seems determined to remove this apparent equality between the two subjects.]_N [Stolpe, Reiche and others do say Yes to the possible compromise offered by the Karlsruhe court, but they also decree: There cannot be any voluntary subject area LER/Religion.]_S
(maz-6159, PCC)

RST relations

- Subject-matter / semantic

- Circumstance
- Condition
- Elaboration
- Otherwise
- Unless
- Interpretation
- Means
- Cause / Result
- Purpose
- Solutionhood

- Textual

- Preparation
- Restatement
- Summary

- Presentational / pragmatic

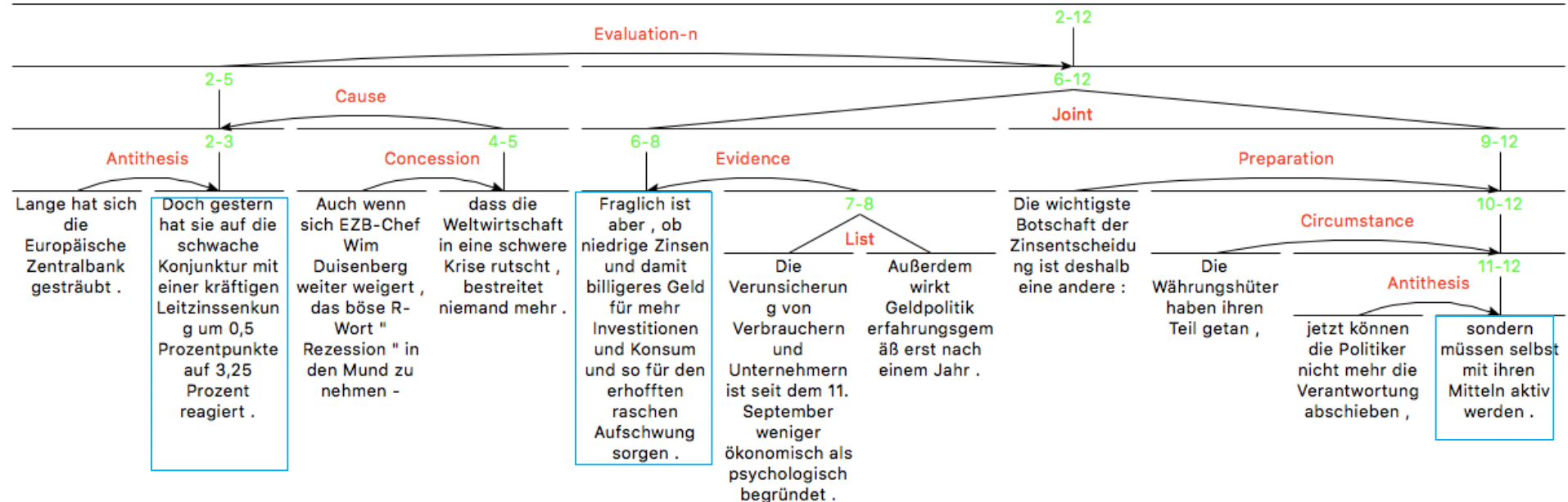
- Background
- Antithesis
- Concession
- Evidence
- Reason
- Justify
- Evaluation
- Motivation
- Enablement

- Multinuclear

- Contrast
- Sequence
- List
- Conjunction
- Joint

Strong nuclearity principle

When a relation holds between two larger segments, it in particular holds between their most central nuclei.



RST: Goal

Text has a **plausible** analysis/representation

$\langle = \rangle$

Text is **coherent**

Some RST corpora

- RST Discourse Treebank (Eng)
 - 385 texts: WSJ
 - annotation manual (78 relations)
- GUM (Eng)
 - continuously growing corpus, multiple genres, multiple annotation layers
- SFU Discourse Relations Reference Corpus (Eng)
 - 65 texts: WSJ, movie and book reviews
- Potsdam Commentary Corpus (Ger)
 - 176 texts: German newspaper editorials
 - annotation manual (31 relations)
- Parallel Discourse Annotation Corpus (Eng, Ger)
 - 123 short texts: argumentation, RST, SDRT

Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

Potsdam Commentary Corpus (PCC)

- 176 commentaries (32K words) from *Märkische Allgemeine Zeitung* (a regional daily)
- Growing number of *pro&contra* editorials from *Tagesspiegel* (a relatively-regional daily)
- Five layers of annotation
- Why a regional newspaper?
 - Language features fairly straightforward constructions and not too much conceptual complexity - presumably easier to parse
- Why commentary?
 - Interesting *rhetorical* structure
 - Perspective: automatic analysis of argumentation

Impfpflicht gegen Kinderkrankheiten?

[1] Kein Kind weiß heute noch, was Pocken sind. [2] So ein Glück. [3] Als die Pockenimpfung 1854 eingeführt wurde, [4] glaubten manche Menschen, [5] dass sich ihr Kopf in einen Kuhkopf verwandelt, [6] wenn sie sich impfen lassen. [7] Denn der Impfstoff wurde damals aus der Haut von Rindern hergestellt. [8] Heute ist diese furchtbare Krankheit ausgerottet. [9] Dank einer entschlossenen, weltweiten Impfkampagne. [10] Aber es gibt noch: Masern, Kinderlähmung, Diphtherie, Mumps, Röteln, Hepatitis B, Tuberkulose, Keuchhusten. [11] Daran sterben, vor allem in den Entwicklungsländern, jährlich immer noch Millionen Kinder. [12] In Deutschland werden diese Krankheiten von vielen Eltern offenbar nicht ernst genommen. [13] Weil sie sie gar nicht mehr kennen! [14] Denn mit Impfstoffen wurde erreicht, [15] dass diese Infektionen nur noch sporadisch auftreten. [16] Doch wer aus eigenem Erleben weiß, [17] wie schrecklich Kinder leiden, [18] wenn sie ,nur‘ Masern oder Keuchhusten haben, [19] sollte ihnen dies ersparen. [20] Und auch die gesundheitlichen Folgewirkungen. [21] Nur wer impfen lässt, hilft mit, dass Impfungen eines Tages überflüssig werden. [22] Stattdessen wird über Nebenwirkungen von Impfstoffen schwadroniert, [23] die höchst selten auftreten und die man erst Recht nur aus Büchern kennt. [24] Dann gibt es noch das schöne Argument: Das ist mein Kind, das darf der Staat nicht pieken. [25] Gegen solche Eltern hilft auch keine Impfung.

Mandatory vaccination against children's diseases?

[1] Today, children don't know anymore what pox are. [2] What a joy. [3] When pox vaccination was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] if they got themselves vaccinated. [7] For the vaccine was made from cattle's skin at the times. [8] Nowadays this dreadful disease is exterminated. [9] Thanks to a determined, world-wide vaccination campaign. [10] But there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] Because they don't know them anymore! [14] For it has been achieved with vaccines [15] that these infections hit only rarely today. [16] But those who have experienced [17] how terribly children suffer [18] when they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] As well as the long-term consequences. [21] **Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day.** [22] Instead, people rant about side effects [23] that occur very rarely and are known merely from books. [24] Then there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Layers of Annotation

- **Syntax**

- TIGER scheme, *Annotate* tool

- **Coreference**

- PoCoS scheme, *MMAX* tool

- **Information Structure**

- SFB632IS scheme, *Exmaralda* tool

- **Shallow discourse structure**

- PDTB scheme, *ConnAnno* tool

- **Rhetorical Structure (RST)**

- Mann/Thompson scheme, *RSTTool* (<http://www.wagsoft.com/RSTTool/>)

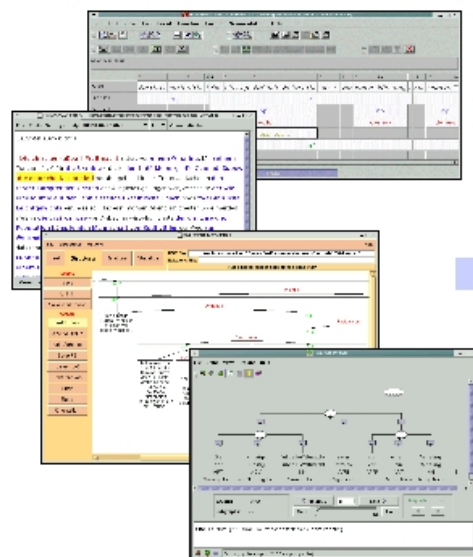
Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

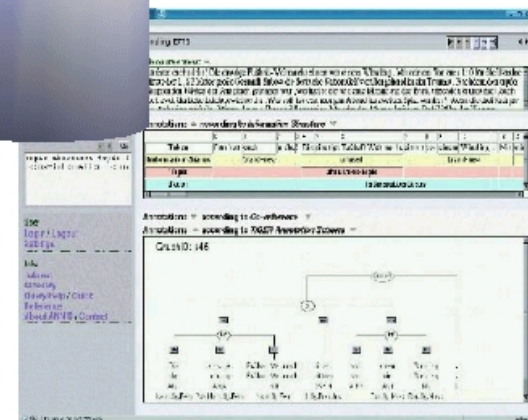
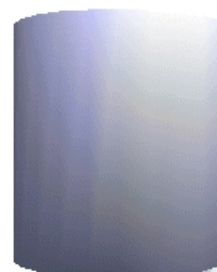
Data Annotation
*Exmaralda, MMAX,
 RST Tool, annotate*

PAULA: XML-based
 Standoff Interchange
 Format

ANNIS
 Linguistic Database:
 Visualization
 and Retrieval

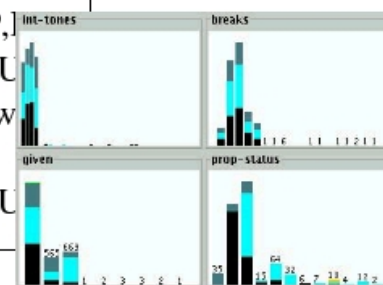


`<markList`
`type="tok"`
`xml:base=`
`"text.xml">`
`...`
`</markList>`



Statistical Evaluation
WEKA

@ATTRIBUTE
 const {NP,
 @ATTRIBUTE
 given {new
 acc, giv}
 @ATTRIBUTE



not logged in [Login](#)

Help/Examples
Query Result

Base text
Token Annotations

1 / 12
Displaying Results 1 - 10 of 120
Result for: cat="NP" & tok & #1 > * #2 & primmark_type="connect ...

Path: pcc2.1 > maz-10205 (tokens 78 - 103)
left context: 5 right context: 5

Schröder	hat	Angst	,	dass	die	öffentliche	Meinung	wegen	der	Kriegsteilnahme	kippt	und	sich	gegen	ihn
Schröder	haben	Angst	,	dass	der	öffentlich	Meinung	wegen	der	Kriegsteilnahme	kippen	und	sich	gegen	ihn
Nom.Sg.Masc	3.Sg.Pres.Ind	Acc.Sg.Fem	--	--	Nom.Sg.Fem	Pos.Nom.Sg.Fem	Nom.Sg.Fem	--	*.Sg.Fem	*.Sg.Fem	3.Sg.Pres.Ind	--	3.Acc.Sg	--	3.Acc.
NE	VAFIN	NN	\$,	KOUS	ART	ADJA	NN	APPR	ART	NN	VVFIN	KON	PRF	APPR	PPER

☐ information structure (grid)
☐ discourse referents (grid)
☒ sentential syntax (tree)

☐ coreference (discourse)
☐ rhetorical structure (rst)

Querying with AQL

- Example: two English sentences with coreferential subjects, where the second one is a pronoun

```
cat="S" & node & cat="S" & pos="PRP" &  
#1 >[func="sb"] #2 &  
#3 >[func="sb"] #4 &  
#4 ->coref #2 &  
meta::language="en"
```

Annis Query Language: Operators

	Name	Illustration	Options
.	direct precedence	A B	
.*	indirect precedence	A x y z B	.n,m
>	direct dominance	A B	>secedge >[func="OA"]
>*	indirect dominance	A ... B	>n,m
->LABEL	Labeled pointing relation	LABEL ↓ ↓ B A	->coref[type =apposition]
->LABEL*	Labeled pointing path	LABEL LABEL ↓ ↓ B x y z A	
=	identical coverage	AAA BBB	
o	overlap	AAA BBB	_ol_ _or_
i	inclusion	AAA B	

	Name	Illustration
l	left aligned	AAA BB
r	right aligned	AA BBB
>@l	left-most child	A / \ B x y
>@r	right-most child	A / \ x y B
\$	Common parent node	x / \ A B
\$*	Common ancestor node	x ... / \ A B
#x:arity=n	Arity	x / \ 1 ... n
#x:tokenarity=n	Tokenarity	x ... / \ 1 ... n
#x:root	Root	<u>x</u>

Query Builder

Search Form

AnnisQL:

pos="VVFIN" & node &
cat="S" & node & #3 > #1 &
#3 > [tiger:func="SB"] #2 &
#1, #2 & #4, #1 & #3

Match Count:

Valid Query

More Corpora

<input type="checkbox"/> Name	Texts	Token
<input type="checkbox"/> ONTONOTES_v1.6_4	100	53875
<input type="checkbox"/> ONTONOTES_v1.6_small	4	6450
<input type="checkbox"/> falko_docDay	1	252
<input type="checkbox"/> pcc-11	11	1939
<input checked="" type="checkbox"/> pcc176	176	33222
<input type="checkbox"/> pcc3	3	573
<input type="checkbox"/> pcc3_mmax2exmaralda	3	573

Simple Search

Query Builder

Statistics

Show Result

Create Node

Edge Add Clear X

Field	op	Value
cat	=	S

X

> [tiger:func="OA"]

X

>

X

> [tiger:func="SB"]

Edge Add Clear X

Field	op	Value
pos	=	VVFIN

X

.

X

.

Edge Add Clear X

Field	op	Value
pos	=	VVFIN

X

.

X

.

Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

Annotation Guidelines: Process

1. **Read** the text completely
2. Divide the text into **EDUs**
3. Informally mark **topic boundaries** between sections of the text
4. Informally mark EDUs that seem specifically **important** for the text
5. Consider all pairs of **adjacent EDUs**:
clear relations? - add them to the analysis
6. In a bottom-up way, **connect segments** to one another, considering the topic boundaries and the nuclearity status

Annotation Guidelines: EDUs

- Which types of subordinate clauses are EDUs
 - *It is completely clear that the mayor needs to resign.*
 - *The chairperson should stay in office, while the mayor needs to resign.*

Annotation Guidelines: EDUs

- Which types of subordinate clauses are EDUs
 - *It is completely clear that the mayor needs to resign.*
 - *The chairperson should stay in office, while the mayor needs to resign.*

Annotation Guidelines: Relations

BACKGROUND

- **N/S:** Understanding S makes it easier for R to understand the content of N; without the background information in S, it would be difficult to comprehend N. In a text, S mostly but not always precedes N. A Background S at the beginning of the text often serves to introduce the topic of the text.
- **Effect:** R's ability to comprehend N is increased.
- **Typical connectives:** Rarely signalled by connectives.
- **Example:** [Until 1984, Burkina Faso was called Obervolta.]_S [According to an EMNID poll, many Europeans today believe that they are two different countries.]_N
- **Remark:** This relation only rarely holds between two EDUs; mostly it will connect larger segments. Many editorials start with a Background S, which serves as starting point for W's subsequent assessment.

Annotation Guidelines: Hints

- Genuine **structural ambiguity**
 - e.g., One „claim“ followed by two „arguments“
- Genre-specific heuristic: prefer **pragmatic** relations
- Prefer more **“informative”** relations
(e.g., causality beats elaboration)
- Fine-grained **differences** between **similar** relations
 - Contrast, Antithesis, Concession
 - causal relations

Text

Structurer

Relations

Statistics

RST File:

Relations File:

Modes:

Segment

Edit

Segment:

Sentences

Paragraphs

The city of Barcelona passed a law on Friday to curb tourism as visitors have begun to overwhelm the city and anger local residents.

Last year the city's 1.6 million residents were heavily outnumbered by an estimated 32 million visitors, about half of them day-trippers.

The new law comes after more than 25 years of relentless promotion of the city as a tourist destination, and coincides with a planned "occupation" on Saturday of La Rambla, a street that has come to symbolise what many view as the excessive and unsustainable number of tourists.

The occupation has been organised by the Barcelona Urban Neighbourhood Association, along with more than 40 residents and community associations.

Text

Structurer

Relations

Statistics

RST File:

Relations File:

Modes

Link

Unlink

Collapse/Expand

Actions

Add Span

Add MultiNuc

Add Schema

Save PS

Save PDX

Print Canvas

Undo

Redo

Orientation

The city of
Barcelona
passed a law
on Friday to
curb tourism

as visitors have
begun to
overwhelm the
city and anger
local residents.

Last year the
city's 1.6
million
residents were
heavily
outnumbered
by an
estimated 32
million visitors,
about half of
them day-
trippers.

The new law
comes after
more than 25
years of
relentless
promotion of
the city as a
tourist
destination,

and coincides
with a planned
"occupation"
on Saturday of
La Rambla, a
street that has
come to
symbolise what
many view as
the excessive
and
unsustainable
number of
tourists. The
occupation has
been organised
by the
Barcelona
Urban
Neighbourhood
Association,
along with
more than 40
residents and
community
associations.

Text

Structurer

Relations

Statistics

RST File:

/Users/stede/Desktop/test.rs3

Relations File:

/Applications/RSTTool311/Relation-Sets/ClassicMT.rel

Modes

Link

Unlink

Collapse/Expand

Actions

Add Span

Add MultiNuc

Add Schema

Save PS

Save PDX

Print Canvas

Undo

Redo

Orientation

The city of Barcelona passed a law on Friday to curb tourism

2-3

Evidence

as visitors have begun to overwhelm the city and anger local residents.

Last year the city's 1.6 million residents were heavily outnumbered by an estimated 32 million visitors, about half of them day-trippers.

The new law comes after more than 25 years of relentless promotion of the city as a tourist destination,

5-6

Background

and coincides with a planned "occupation" on Saturday of La Rambla, a street that has come to symbolise what many view as the excessive and unsustainable number of tourists.

The occupation has been organised by the Barcelona Urban Neighbourhood Association, along with more than 40 residents and community associations.

Text

Structurer

Relations

Statistics

RST File:

/Users/stede/Desktop/test.rs3

Relations File:

/Applications/RSTTool311/Relation-Sets/ClassicMT.rel

Modes

Link

Unlink

Collapse/Expand

Actions

Add Span

Add MultiNuc

Add Schema

Save PS

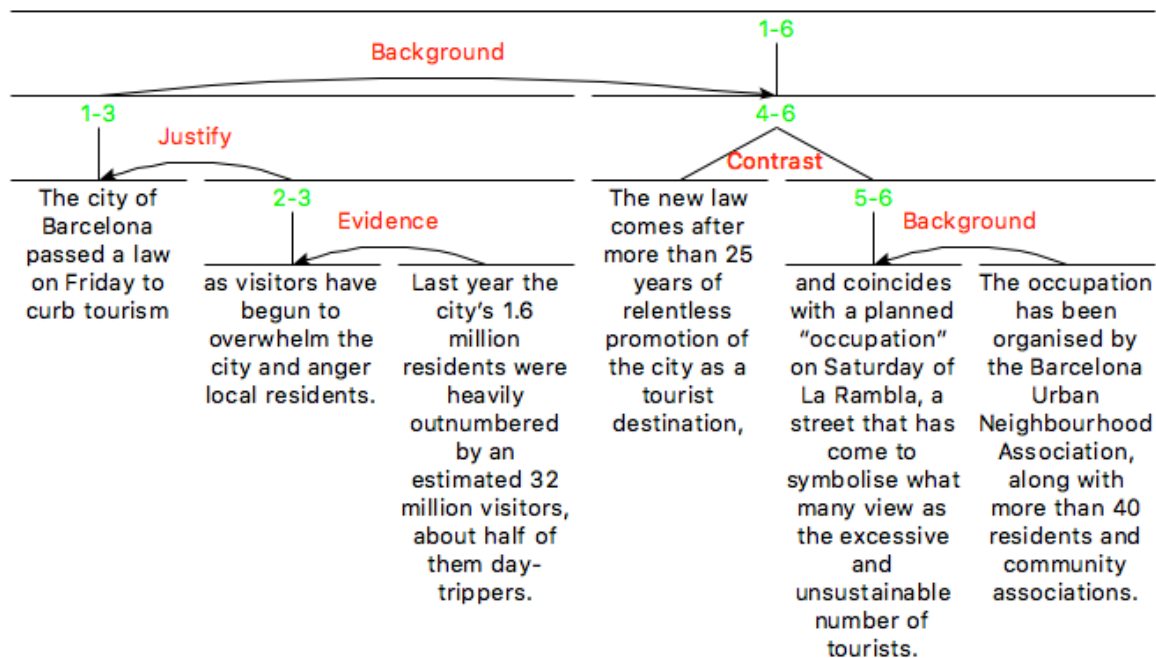
Save PDX

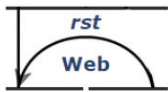
Print Canvas

Undo

Redo

Orientation





rstWeb - Browser Annotation of Rhetorical Structure Theory

rstWeb is an open source, browser based annotation tool for discourse analyses in [Rhetorical Structure Theory](#) and its enhanced version, [eRST](#). It is meant to support collaborative, online annotation projects using just a web browser, without the need to install software for annotators.

New: Version 4 is now out, supporting enhanced Rhetorical Structure Theore ([eRST](#))! For more information, see [our paper](#).

Do RST and eRST: Online — Collaborative — Browser Based!

For a server installation, rstWeb needs **Python > 2.6.X** for its backend (versions at least up to **Python 3.11** have been extensively tested), which accesses a SQLite database. The client runs in JavaScript and only needs a browser. User management and configurations use Michael Foord's *logintools*, and the front end uses the excellent [jsPlumb](#), [jQuery](#) and [Font Awesome](#), all of which are packaged with rstWeb and require no installation. To take screenshots of analyses automatically, it is optionally possible to install [Selenium](#) and [PhantomJS](#), which automate saving scaled screenshots that fit your analysis size at high quality. This project's source code is available from [Github](#) - feel free to fork, contribute, get in touch and let me know what you're working on.



Quantitative annotator agreement

Measuring parsing accuracy (P = Precision, R = Recall).

	Spans		Nuclearity		Relations	
Constituents	Human	System	Human	System	Human	System
1-1	*	*	S	S	Attribution	Attribution
2-2		*		S		Elaboration
3-3		*		N		Span
4-4	*	*	N	N	Contrast	Contrast
5-5	*		N		Span	
6-6	*		S		Elaboration	
7-7	*	*	N	N	Same-Unit	Joint
2-3	*	*	N	N	Contrast	Contrast
5-6	*	*	N	N	Same-Unit	Joint
2-4	*	*	S	N	Contrast	Span
5-7	*	*	N	N	Contrast	Contrast
1-4		*		N		Contrast
2-7	*		N		Span	
	R = 7/10, P = 7/10		R = 6/10, P = 6/10		R = 4/10, P = 4/10	

Quantitative annotator agreement

- Measured on 18 PCC texts (10%)
- Two well-trained annotators
- EDU segmentation was given

- Span: 0.85
- Nuclearity: 0.65
- Relations: 0.44

Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

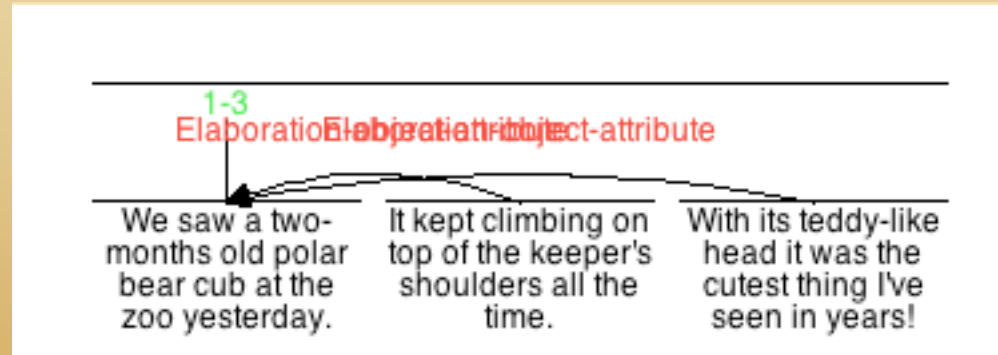
RST annotation

- Annotators don't agree very much with each other.
- >> not a problem.
- Annotators don't agree very much with themselves.
- >> a problem!
- RST tree is supposed to capture the **subjective interpretation** by the analyst - but it should do so as **transparently** and **unambiguously** as possible.

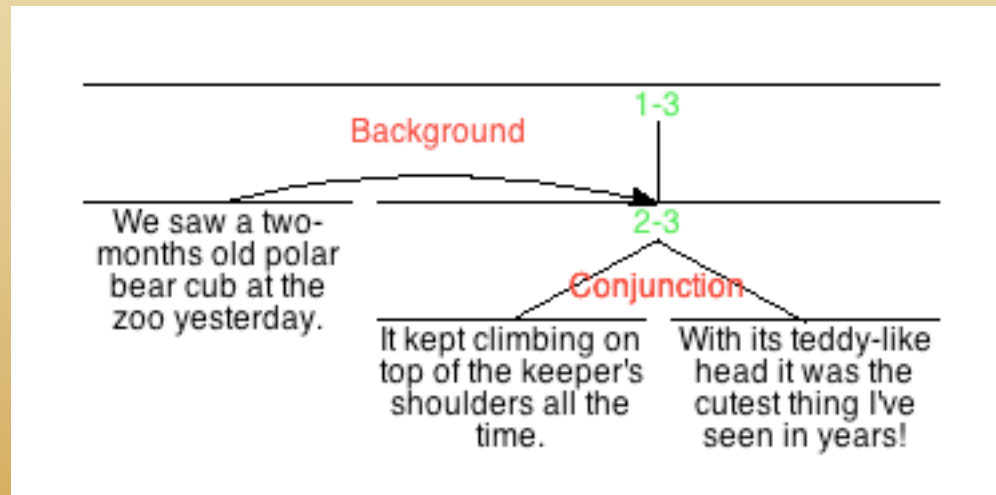
RST annotation: ambiguities

We saw a two-months old polar bear cub at the zoo yesterday. It kept climbing on top of the keeper's shoulders all the time. With its teddy-like head it was the cutest thing I've seen in years!

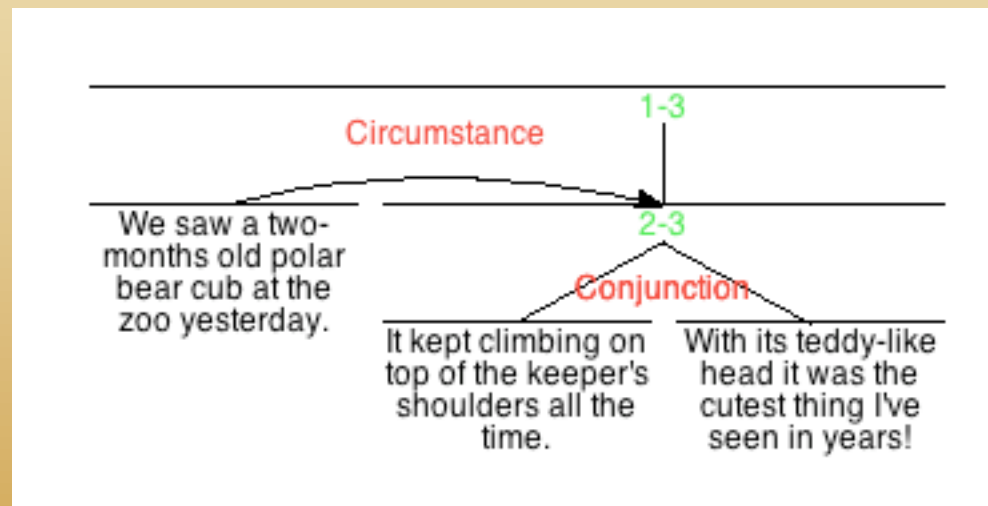
RST annotation: ambiguities



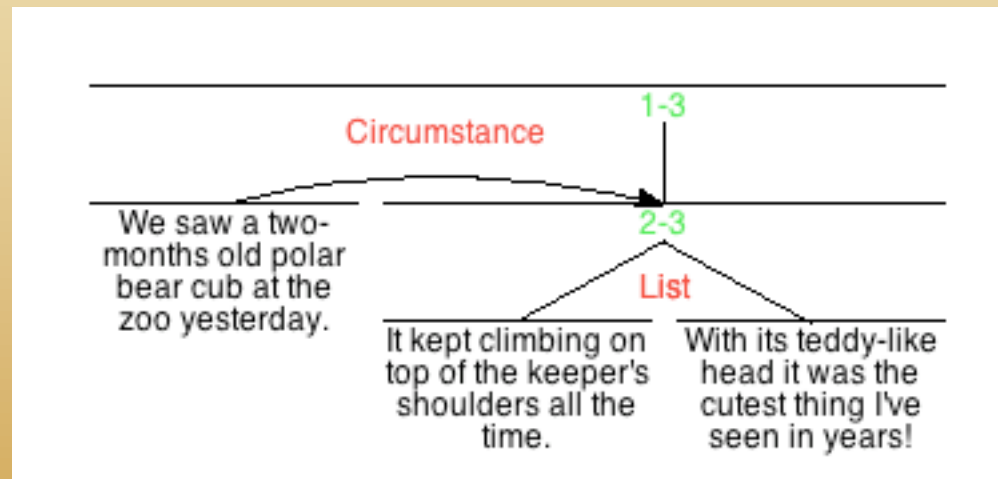
RST annotation: ambiguities



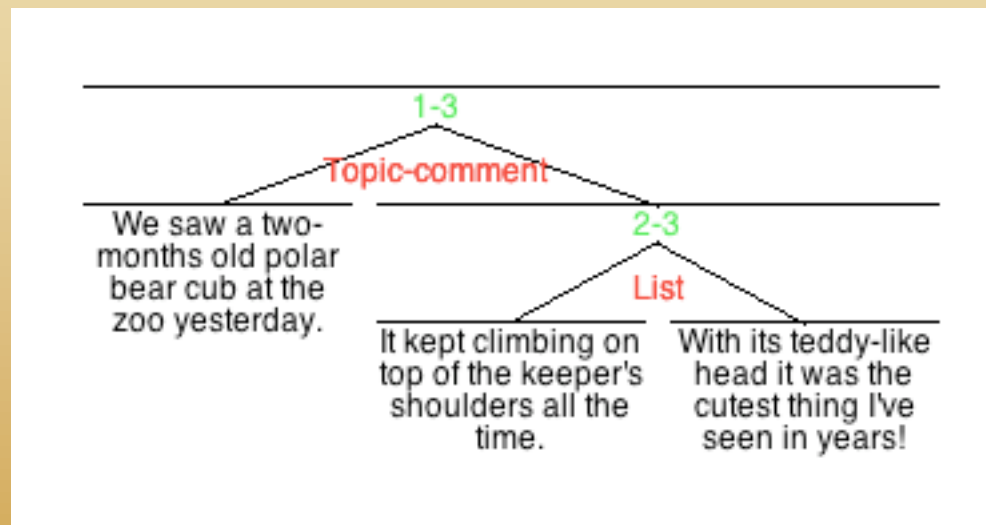
RST annotation: ambiguities



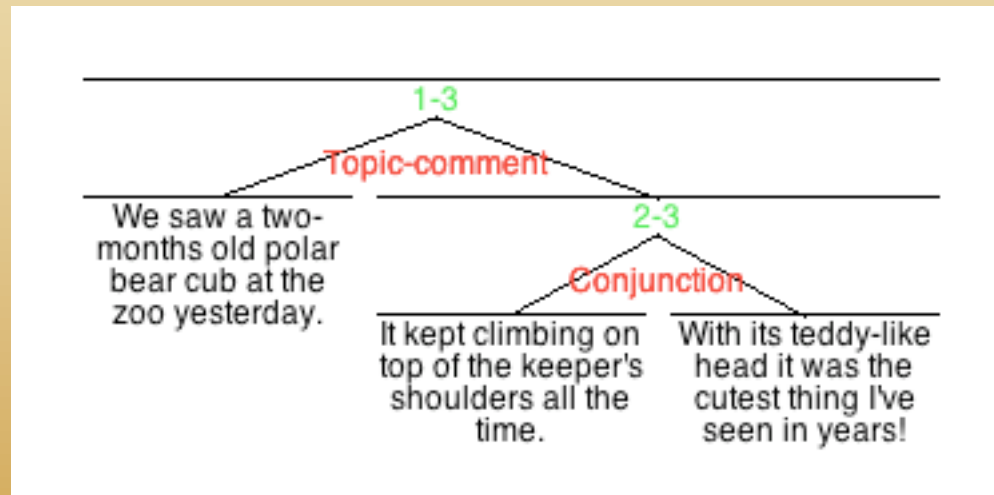
RST annotation: ambiguities



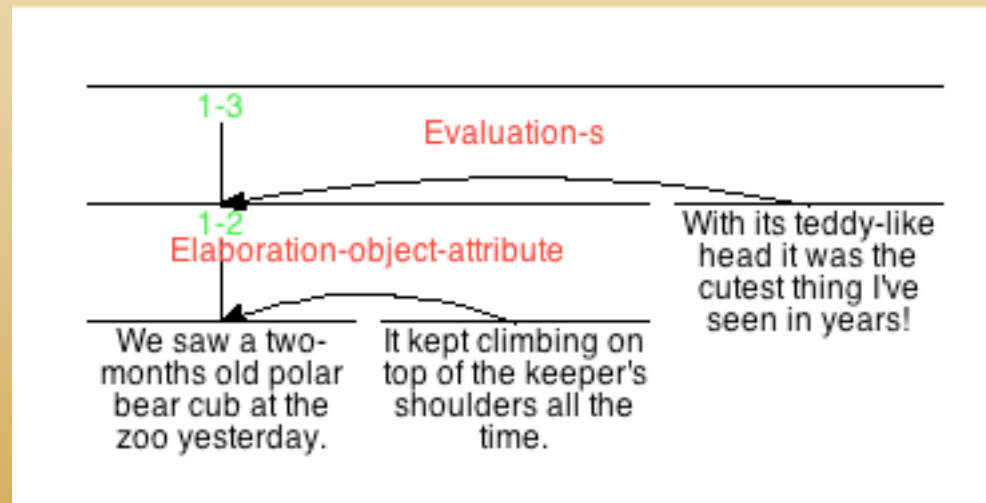
RST annotation: ambiguities



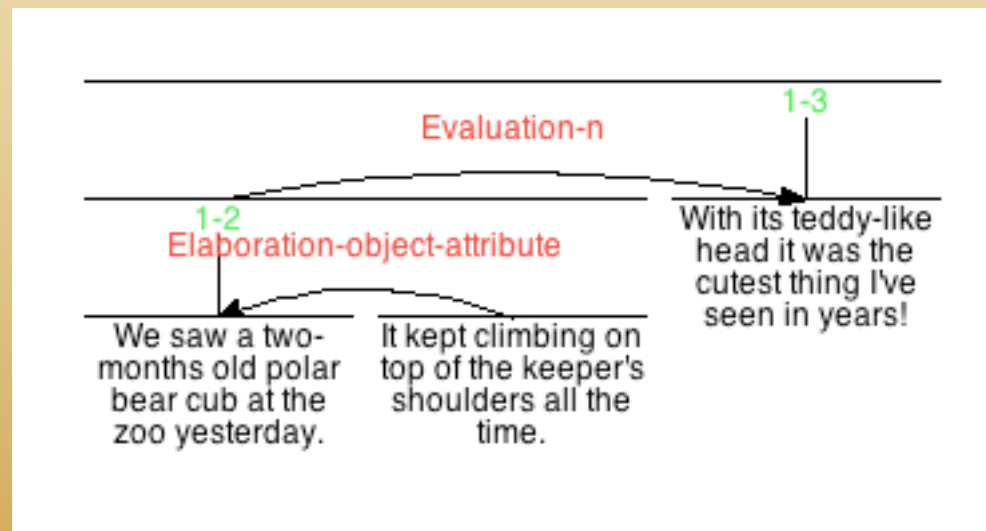
RST annotation: ambiguities



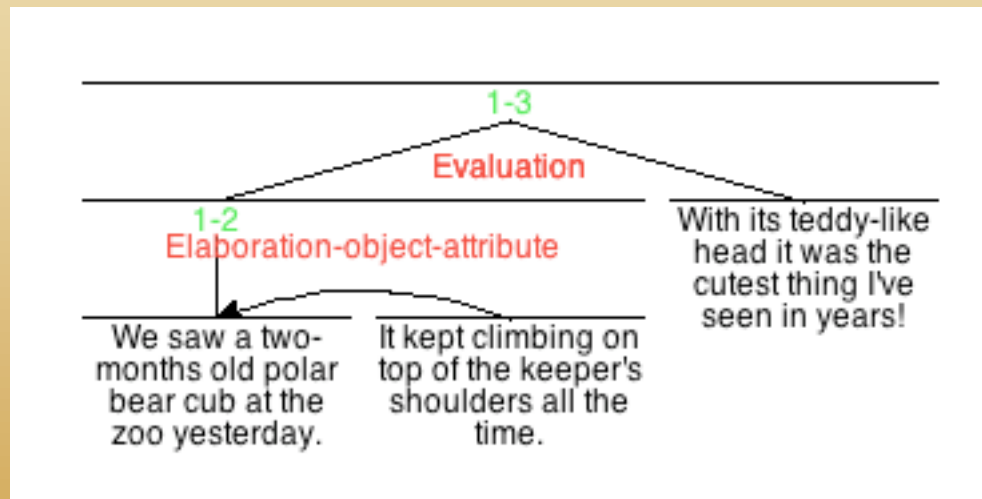
RST annotation: ambiguities



RST annotation: ambiguities



RST annotation: ambiguities



Additional data: German

- **Blog posts**

- Various publishers
- Announcements or summaries of accompanying podcasts
- 78 RST trees / 1,309 relations / 1,387 EDUs

- **Simplified German news (APA)**

- 25 double-annotated texts from Austrian News Agency
- (Simplification to levels B1 and A2)
- 61 RST trees / 852 relations / 938 EDUs

- **Total (with PCC):**

315 RST trees / 5,096 relations / 5,436 EDUs

Additional data: English

- RST Discourse Treebank (Carlson/Marcu 2003)
 - 385 Wall Street Journal texts
 - 50 texts have two annotations
 - 26 of these have identical segmentation
- UNSC-RST corpus (Zaczynska/Stede 2023)
 - 84 Texts from UN Security Council speeches

Mapping out the disagreements

RST Tace: Enumerate the decision points Simplification: we assume identical EDU segmentation

Stems 1		Stems 2	ID	Category	CS-A	Relation-A	Num-A	C1-A	CN-A	CS-A	A1-A	A2-A	Texts	Texts	CS-B	Relation-B	Num-B	C1-B	CN-B	CS-B	A1-B	A2-B	Texts	Texts	Matching	N	R	O	A	Agreement	Disagreement		
10	BS-EE		11	Interchangeable relations	11	elaboration	---	11.0	11.0	S	10.0	10.0	Weniger wäre schädlicher als der Eindruck, die Hauptstadt könne den Hals nicht voll kriegen.	Zum Beispiel, indem sie sich in den Wettbewerb um eine deutsche Kandidatur für die Olympischen Spiele 2016 dringt.	11	e-elaboration	---	11.0	11.0	S	10.0	10.0		Weniger wäre schädlicher als der Eindruck, die Hauptstadt könne den Hals nicht voll kriegen.	Zum Beispiel, indem sie sich in den Wettbewerb um eine deutsche Kandidatur für die Olympischen Spiele 2016 dringt.	Completely identical	CS	✓	✓	✓	NCA	✓R	
11	F-EE		12	Interchangeable relations	12-14	elaboration	---	12.0	14.0	S	10.0	11.0	Weniger wäre schädlicher als der Eindruck, die Hauptstadt könne den Hals nicht voll kriegen. Zum Beispiel, indem sie sich in den Wettbewerb um eine deutsche Kandidatur für die Olympischen Spiele 2016 dringt.	Gegen Hamburg, das Leipzig in der nationalen Ausschreibung knapp unterlag, gegen die anderen Bewerber, die das Risiko bereits einmal auf sich genommen haben. Das Risiko nämlich, für einen sehr geringen Erfolg Geld auszugeben.	12-13	e-elaboration	---	12.0	14.0	S	10.0	11.0		Weniger wäre schädlicher als der Eindruck, die Hauptstadt könne den Hals nicht voll kriegen. Zum Beispiel, indem sie sich in den Wettbewerb um eine deutsche Kandidatur für die Olympischen Spiele 2016 dringt.	Gegen Hamburg, das Leipzig in der nationalen Ausschreibung knapp unterlag, gegen die anderen Bewerber, die das Risiko bereits einmal auf sich genommen haben. Das Risiko nämlich, für einen sehr geringen Erfolg Geld auszugeben.	C1-C2 and A1-A2	✓	✓	✓	✓	NCA	✓R	
15	SoU-Cau		16	Interchangeable relations	18	result	---	18.0	18.0	S	16.0	17.0	Schon eine Olympia-Bewerbung als solche kostet so einige Millionen - die wieder an anderer Stelle gespart werden müssten.	Wer wäre wohl bereit zu zahlen: die Lehrer, die Eltern, die OpenHouse?	18	elaboration	---	18.0	18.0	S	16.0	17.0		Schon eine Olympia-Bewerbung als solche kostet so einige Millionen - die wieder an anderer Stelle gespart werden müssten.	Wer wäre wohl bereit zu zahlen: die Lehrer, die Eltern, die OpenHouse?	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
20	SoU-Cau		7	Interchangeable relations	8	reason	---	7.0	8.0	S	6.0	6.0	Es ist offensichtlich, dass dieser Bundeswehr-Einsatz bei vielen Grünen an die Wurzeln ihrer Identität rührt.	Kommt der Kosovo-Krieg noch mit dem Kampf für die Menschenrechte in Europa begründete werden, geht es nun um die militärische Solidarität mit den USA in einem weit entfernten Dritte-Welt-Land.	7-8	elaboration	---	7.0	8.0	S	6.0	6.0		Es ist offensichtlich, dass dieser Bundeswehr-Einsatz bei vielen Grünen an die Wurzeln ihrer Identität rührt.	Kommt der Kosovo-Krieg noch mit dem Kampf für die Menschenrechte in Europa begründete werden, geht es nun um die militärische Solidarität mit den USA in einem weit entfernten Dritte-Welt-Land.	C1-C2 and A1-A2	✓	✓	✓	✓	NCA	✓R	
36	BS-Corj		13	Interchangeable relations	1011	conjunction	---	10.0	10.0	N	11.0	11.0	dass es wie der Krieg in Afghanistan stoppen	nach eine deutsche Beteiligung daran verhindern können.	1011	rel	---	10.0	10.0	N	11.0	11.0		dass es wie der Krieg in Afghanistan stoppen	nach eine deutsche Beteiligung daran verhindern können.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
43	SoU-BE	Item auch Interp sein	8	Interchangeable relations	9-16	background	---	9.0	16.0	S	1.0	8.0	Die Zahl der Arbeitslosen ist im Vergleich zur Vorwoche leicht gesunken. Derzeit sind 606.987 Personen beim AMS arbeitslos gemeldet oder in Schulung, das sind um 3.500 Personen weniger als noch eine Woche. 326.396 Personen sind auf Arbeitslehre, 69.991 nehmen an Schulungsmaßnahmen teil, geht aus aktuellen Zahlen des Arbeitsmarkts heraus. Im Anstieg wurde bei den Anmeldungen zur Kurzarbeit verzeichnet, diese kletterten von 114.961 auf 136.995.	"Trotz der für diese Zeit des Jahres üblichen saisonalen Effekte verzeichnen wir derzeit sinkende Arbeitslosenquoten. Bemerkenswert ist, dass die Arbeitslosigkeit mit derzeit 606.987 arbeitslosen gemeldeten Personen inklusive Schulungsmaßnahmen und Teilnehmern an Arbeitslehre 326.396 Personen unter dem Niveau der Vergleichswerte 2019 liegt", erklärte Arbeitsminister Martin Kocher (ÖVP) am Dienstag bei der Aussendung. Die rückläufigen Zahlen zeigen, dass die Auswirkungen der Pandemie auf den Arbeitsmarkt nur mehr in geringen Bereichen wie der Langzeitarbeitslosigkeit prägen, heißt es in der Aussendung. Positiv sei auch, dass sich die Jugendarbeitslosigkeit nach nationaler Definition auf niedrigerem Niveau als vor der Krise bewegt. Die Quote lag derzeit bei 6,9 Prozent, 2019 waren es noch 7,8 Prozent.	9-16	elaboration	---	9.0	16.0	S	1.0	8.0		Die Zahl der Arbeitslosen ist im Vergleich zur Vorwoche leicht gesunken. Derzeit sind 606.987 Personen beim AMS arbeitslos gemeldet oder in Schulung, das sind um 3.500 Personen weniger als noch eine Woche. 326.396 Personen sind auf Arbeitslehre, 69.991 nehmen an Schulungsmaßnahmen teil, geht aus aktuellen Zahlen des Arbeitsmarkts heraus. Im Anstieg wurde bei den Anmeldungen zur Kurzarbeit verzeichnet, diese kletterten von 114.961 auf 136.995.	"Trotz der für diese Zeit des Jahres üblichen saisonalen Effekte verzeichnen wir derzeit sinkende Arbeitslosenquoten. Bemerkenswert ist, dass die Arbeitslosigkeit mit derzeit 606.987 arbeitslosen gemeldeten Personen inklusive Schulungsmaßnahmen und Teilnehmern an Arbeitslehre 326.396 Personen unter dem Niveau der Vergleichswerte 2019 liegt", erklärte Arbeitsminister Martin Kocher (ÖVP) am Dienstag bei der Aussendung. Die rückläufigen Zahlen zeigen, dass die Auswirkungen der Pandemie auf den Arbeitsmarkt nur mehr in geringen Bereichen wie der Langzeitarbeitslosigkeit prägen, heißt es in der Aussendung. Positiv sei auch, dass sich die Jugendarbeitslosigkeit nach nationaler Definition auf niedrigerem Niveau als vor der Krise bewegt. Die Quote lag derzeit bei 6,9 Prozent, 2019 waren es noch 7,8 Prozent.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
58	F		8	Interchangeable relations	8	elaboration	---	8.0	8.0	S	5.0	7.0	In Neuen beispielsweise bleibt inzwischen jeder sechste der jährlich entfallenden 4000 Stühle an der Straße zurück, weil er nicht mit Laub von Straßenbäumen, sondern mit Garten- oder Hausabfällen gefüllt ist.	Tendenz steigend.	8	evaluation-e	---	8.0	8.0	S	5.0	7.0		In Neuen beispielsweise bleibt inzwischen jeder sechste der jährlich entfallenden 4000 Stühle an der Straße zurück, weil er nicht mit Laub von Straßenbäumen, sondern mit Garten- oder Hausabfällen gefüllt ist.	Tendenz steigend.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
63	BS-EE		14	Interchangeable relations	14	e-elaboration	---	14.0	14.0	S	13.0	13.0	Inmitten rund 28 000 Mark fallen sich damit pro Jahr einengen.	Geld, das an anderen Stellen dringend benötigt wird.	14	elaboration	---	14.0	14.0	S	13.0	13.0		Inmitten rund 28 000 Mark fallen sich damit pro Jahr einengen.	Geld, das an anderen Stellen dringend benötigt wird.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
64	SoU-Cau		13	Interchangeable relations	15	summary	---	15.0	17.0	S	1.0	14.0	Reise vom Sonntagstraktor, keine Mühlsteinen oder die alle Barsteppes gehören nicht in den Leutskau. Das wird jedes Kind aber nicht jeder hat sich daran. Als Jahre wieder findet sich in vielen Städten und so manche Kuratist. In Neuen beispielsweise bleibt inzwischen jeder sechste der jährlich entfallenden 4000 Stühle an der Straße zurück, weil er nicht mit Laub von Straßenbäumen, sondern mit Garten- oder Hausabfällen gefüllt ist. Tendenz steigend, immer mehr verschwendungsgeliebte Gegenstände verkaufen. Beim Müll Regel kooperieren statt ihr verschwendungsgeliebt zu entsorgen. Die sind nicht als teuer zu einem kommen. Denn angesichts der zunehmenden Müllkosten und neuer Kassen ist es nur noch eine Frage der Zeit, bis die Betreiber der für den Bürger finanziellen Lastenbegründung entstehen. Inmitten rund 28 000 Mark fallen sich damit pro Jahr einengen. Geld, das an anderen Stellen dringend benötigt wird.	Entsorgung nur noch gegen Vorkasse, das könnte schon bald Realität werden. Denn seit Jahren appliziert die Stadt an die Einsicht ihrer Bürger, und jedes Jahr aufs Neue zeigt sich, dass diese Äpfel (bisher) schief sind.	15	result	---	15.0	17.0	S	1.0	14.0		Reise vom Sonntagstraktor, keine Mühlsteinen oder die alle Barsteppes gehören nicht in den Leutskau. Das wird jedes Kind aber nicht jeder hat sich daran. Als Jahre wieder findet sich in vielen Städten und so manche Kuratist. In Neuen beispielsweise bleibt inzwischen jeder sechste der jährlich entfallenden 4000 Stühle an der Straße zurück, weil er nicht mit Laub von Straßenbäumen, sondern mit Garten- oder Hausabfällen gefüllt ist. Tendenz steigend, immer mehr verschwendungsgeliebte Gegenstände verkaufen. Beim Müll Regel kooperieren statt ihr verschwendungsgeliebt zu entsorgen. Die sind nicht als teuer zu einem kommen. Denn angesichts der zunehmenden Müllkosten und neuer Kassen ist es nur noch eine Frage der Zeit, bis die Betreiber der für den Bürger finanziellen Lastenbegründung entstehen. Inmitten rund 28 000 Mark fallen sich damit pro Jahr einengen. Geld, das an anderen Stellen dringend benötigt wird.	Entsorgung nur noch gegen Vorkasse, das könnte schon bald Realität werden. Denn seit Jahren appliziert die Stadt an die Einsicht ihrer Bürger, und jedes Jahr aufs Neue zeigt sich, dass diese Äpfel (bisher) schief sind.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
72	SoU	weniger Korrektur	2	Interchangeable relations	2	circumstance	---	2.0	2.0	S	3.0	3.0	Denn wenn sich jetzt unentwirrt von Schläm an in Wiesenburg ein Förderverein für die Grund- und Gesamtschule gründet.	zeigt das, welchen Stellenwert einer guten schulischen Ausbildung inzwischen eingebracht wird.	2	condition	---	2.0	2.0	S	3.0	3.0		Denn wenn sich jetzt unentwirrt von Schläm an in Wiesenburg ein Förderverein für die Grund- und Gesamtschule gründet.	zeigt das, welchen Stellenwert einer guten schulischen Ausbildung inzwischen eingebracht wird.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
73	SoU-Cau		3	Interchangeable relations	4	evidence	---	4.0	4.0	S	5.0	5.0	Wie anderswo bereits seit längeren erfolgreich praktiziert, wird die Schule so stärker in den Mittelpunkt der Gesellschaft gerückt.	Wie anderswo bereits seit längeren erfolgreich praktiziert.	4	circumstance	---	4.0	4.0	S	5.0	5.0		Wie anderswo bereits seit längeren erfolgreich praktiziert, wird die Schule so stärker in den Mittelpunkt der Gesellschaft gerückt.	Wie anderswo bereits seit längeren erfolgreich praktiziert.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
76	SoU-Cau	Item auch Resultate oder	6	Interchangeable relations	7	reason-n	---	6.0	7.0	S	4.0	4.0	Wie anderswo bereits seit längeren erfolgreich praktiziert, wird die Schule so stärker in den Mittelpunkt der Gesellschaft gerückt.	Wie ist kein abgeschlossenes Gebilde mehr, sondern Teil der Kommunen.	7	result	---	6.0	7.0	S	4.0	4.0		Wie anderswo bereits seit längeren erfolgreich praktiziert, wird die Schule so stärker in den Mittelpunkt der Gesellschaft gerückt.	Wie ist kein abgeschlossenes Gebilde mehr, sondern Teil der Kommunen.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
80	SoU		10	Interchangeable relations	9	restatement	---	9.0	9.0	S	8.0	8.0	An die Stelle von "die oben müssten mal ..." ist Eigeninitiative gegeben.	Selbst etwas anpacken und nicht auf Almosen warten, scheint nun die Devise zu sein.	9	interpretation	---	9.0	9.0	S	8.0	8.0		An die Stelle von "die oben müssten mal ..." ist Eigeninitiative gegeben.	Selbst etwas anpacken und nicht auf Almosen warten, scheint nun die Devise zu sein.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
81	F		11	Interchangeable relations	12-13	elaboration	---	12.0	13.0	S	11.0	11.0	Bleibt zu hoffen, dass sich nicht nur den Wiesenburger Förderverein möglichst viele Menschen aus diversen Gruppen der Gesellschaft anschließen.	Beispielsweise können Handwerker und Betriebe der Region so häufiger Kontakte zu künftigen Lehrlingen knüpfen oder im Gespräch mit Lehrern ihre Anforderungen an die oft kritisierte Ausbildung in den Schulen verdeutlichen.	12-13	reason	---	12.0	13.0	S	11.0	11.0		Bleibt zu hoffen, dass sich nicht nur den Wiesenburger Förderverein möglichst viele Menschen aus diversen Gruppen der Gesellschaft anschließen.	Beispielsweise können Handwerker und Betriebe der Region so häufiger Kontakte zu künftigen Lehrlingen knüpfen oder im Gespräch mit Lehrern ihre Anforderungen an die oft kritisierte Ausbildung in den Schulen verdeutlichen.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
94	F	Nu evaluiert nicht den 5	10	Interchangeable relations	9	cause	---	9.0	9.0	S	10.0	10.0	Die Zeiten, die die Literatur Aufzählungen- und Bekanntheit mitteilen müsste, sind zum Glück vorbei.	Deshalb ist es auch gut, dass das Umweltministerium in diesem Jahr in engerem Sinne auf fachliche Leistungen setzt.	9	evaluation-e	---	9.0	9.0	S	10.0	10.0		Die Zeiten, die die Literatur Aufzählungen- und Bekanntheit mitteilen müsste, sind zum Glück vorbei.	Deshalb ist es auch gut, dass das Umweltministerium in diesem Jahr in engerem Sinne auf fachliche Leistungen setzt.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
98	F-EE		12	Interchangeable relations	12	e-elaboration	---	12.0	12.0	S	11.0	11.0	Buane Pauben bekommt den Preis für ein erfindendes Buchsch.	das dem Leser die rätselhafte Welt der Pflanzen näherbringt.	12	elaboration	---	12.0	12.0	S	11.0	11.0		Buane Pauben bekommt den Preis für ein erfindendes Buchsch.	das dem Leser die rätselhafte Welt der Pflanzen näherbringt.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
110	BS-EE	Item Korrektur, aber bei 16	10	Interchangeable relations	1112	contrast	---	11.0	11.0	N	12.0	12.0	Haute tie sie von "94 bis '99 müßig geschäft", 14 Millionen Defizit der Altkasse abzubauen.	Hoffen nun immer neue Millionenbörcher, beispielsweise im Sozialist, die kaum noch zu stoppen sind.	1112	sequence	---	11.0	11.0	N	12.0	12.0		Haute tie sie von "94 bis '99 müßig geschäft", 14 Millionen Defizit der Altkasse abzubauen.	Hoffen nun immer neue Millionenbörcher, beispielsweise im Sozialist, die kaum noch zu stoppen sind.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
140	BS-OS	die positiven Details der 9	9	Interchangeable relations	10	evidence	---	10.0	11.0	S	9.0	9.0	Das es Menschen gibt in dieser Stadt, die etwas bewirken können, hat die Umsetzung der jetzt ausgeführten Urban-Programme gezeigt.	Das es Menschen gibt in dieser Stadt, die etwas bewirken können, hat die Umsetzung der jetzt ausgeführten Urban-Programme gezeigt.	10	elaboration	---	10.0	11.0	S	9.0	9.0		Das es Menschen gibt in dieser Stadt, die etwas bewirken können, hat die Umsetzung der jetzt ausgeführten Urban-Programme gezeigt.	Das es Menschen gibt in dieser Stadt, die etwas bewirken können, hat die Umsetzung der jetzt ausgeführten Urban-Programme gezeigt.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
141	SoU-Cau	Nurlesen / Referenzfrage	10	Interchangeable relations	11	reason-n	---	11.0	11.0	S	10.0	10.0	Innervall der vergangenen sechs Jahre hat die Bahnhofsvorstadt ihr Gesicht fast völlig verändert.	Aus dem früheren Nachbarnkennzeichen ist ein familienfreundlicher Kiez geworden.	11	elaboration	---	11.0	11.0	S	10.0	10.0		Innervall der vergangenen sechs Jahre hat die Bahnhofsvorstadt ihr Gesicht fast völlig verändert.	Aus dem früheren Nachbarnkennzeichen ist ein familienfreundlicher Kiez geworden.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
198	BS-EE		12	Interchangeable relations	11	elaboration	---	11.0	11.0	S	10.0	10.0	Als das beginnt nach einer Übergangszeit mit Informationsaktivitäten die facto Kontrolle der Impfindigkeit.	Als die jure ab Februar gelten soll.	11	e-elaboration	---	11.0	11.0	S	10.0	10.0		Als das beginnt nach einer Übergangszeit mit Informationsaktivitäten die facto Kontrolle der Impfindigkeit.	Als die jure ab Februar gelten soll.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
199	F		6	Interchangeable relations	10	condition	---	10.0	10.0	S	9.0	9.0	und es schlicht einem Nichtnutzer sogar noch Aggression entgegen.	weist er freundlich drauf hin, dass es unnötiglich müßig...	10	elaboration	---	10.0	10.0	S	9.0	9.0		und es schlicht einem Nichtnutzer sogar noch Aggression entgegen.	weist er freundlich drauf hin, dass es unnötiglich müßig...	Completely identical	CS	✓	✓	✓	✓	NCA	✓R
202	SoU-Corj	BS passt nicht, denn bei 16	10	Interchangeable relations	12	conjunction	---	1.0	1.0	N	2.0	2.0	Amerikaner würden sich bei dieser Frage wahrscheinlich auf den Boden werfen	und lauthals lachen.	12	sequence	---	1.0	1.0	N	2.0	2.0		Amerikaner würden sich bei dieser Frage wahrscheinlich auf den Boden werfen	und lauthals lachen.	Completely identical	CS	✓	✓	✓	✓	NCA	✓R

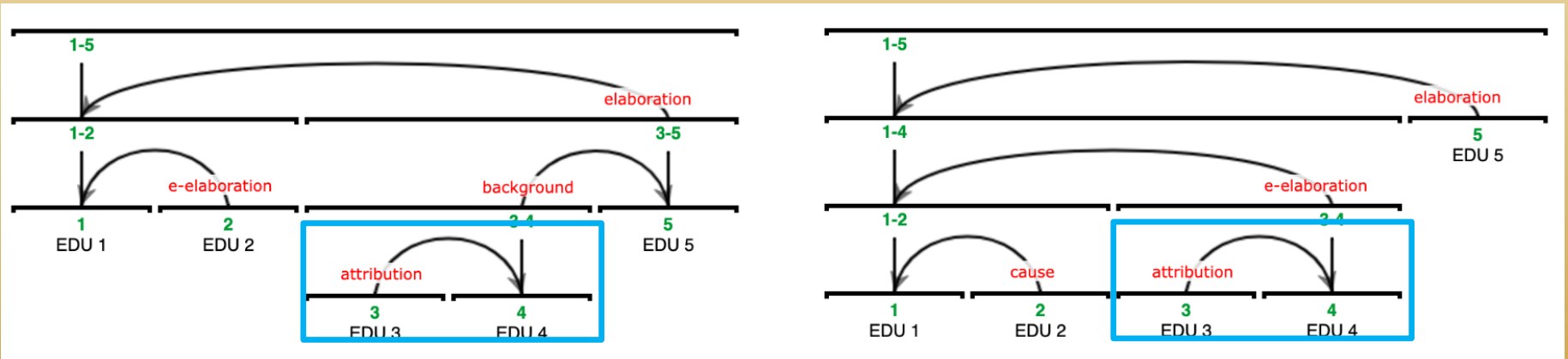
S. Wan, T. Kutschbach, A. Lüdeling, M. Stede. RST-Tace: A tool for automatic comparison and evaluation of RST trees. Proc. of the Workshop on Discourse Relation Parsing and Treebanking, 2019

Step 1: Formal classes

- Perfect match
- Relation mismatch
 - same spans, but different relation
- Scope mismatch
 - same overall span
 - same relation, different split point
 - different relations, different split points
 - different overall spans, but:
 - identical relation, identical split point
 - (...)
- Left/right priority mismatch
 - The same EDU is attached left vs. right
- No match

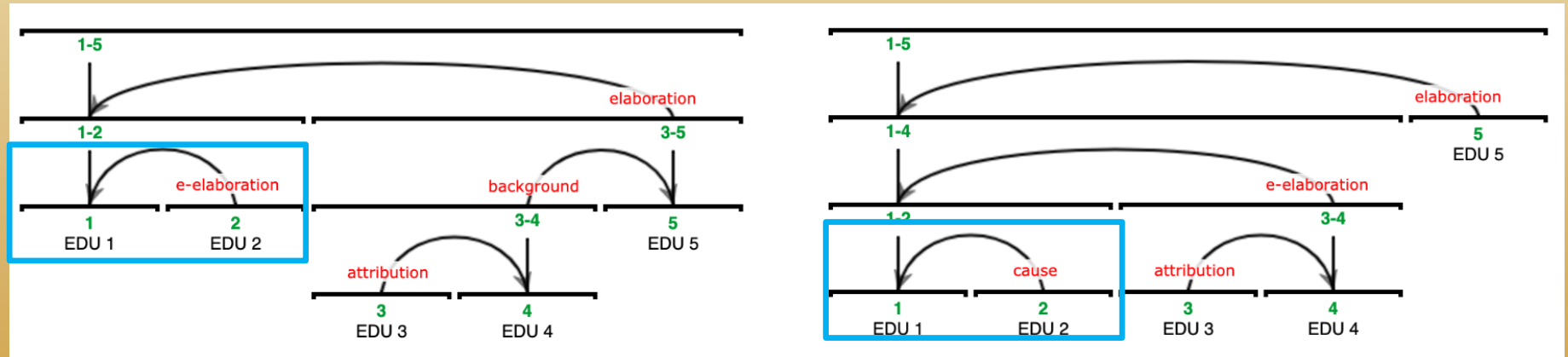
Step 1: Formal classes

Perfect match



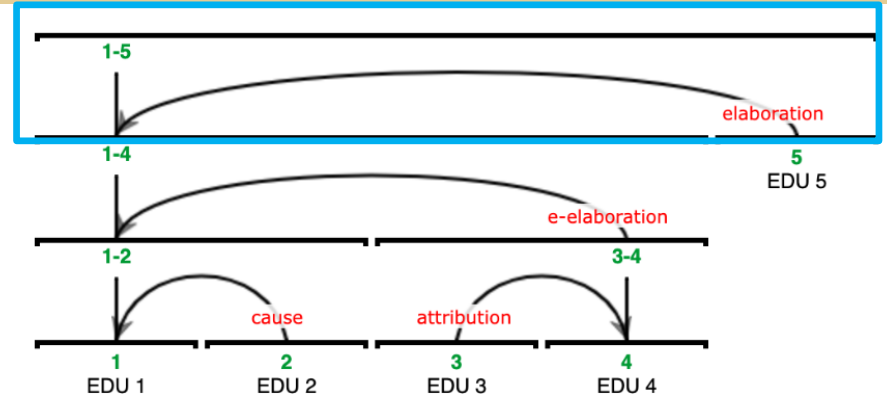
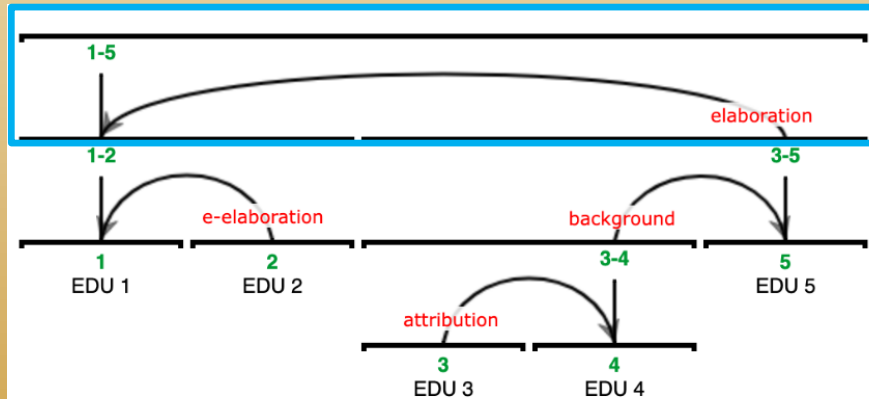
Step 1: Formal classes

Relation mismatch



Step 1: Formal classes

Scope mismatch: same rel., same span, different split

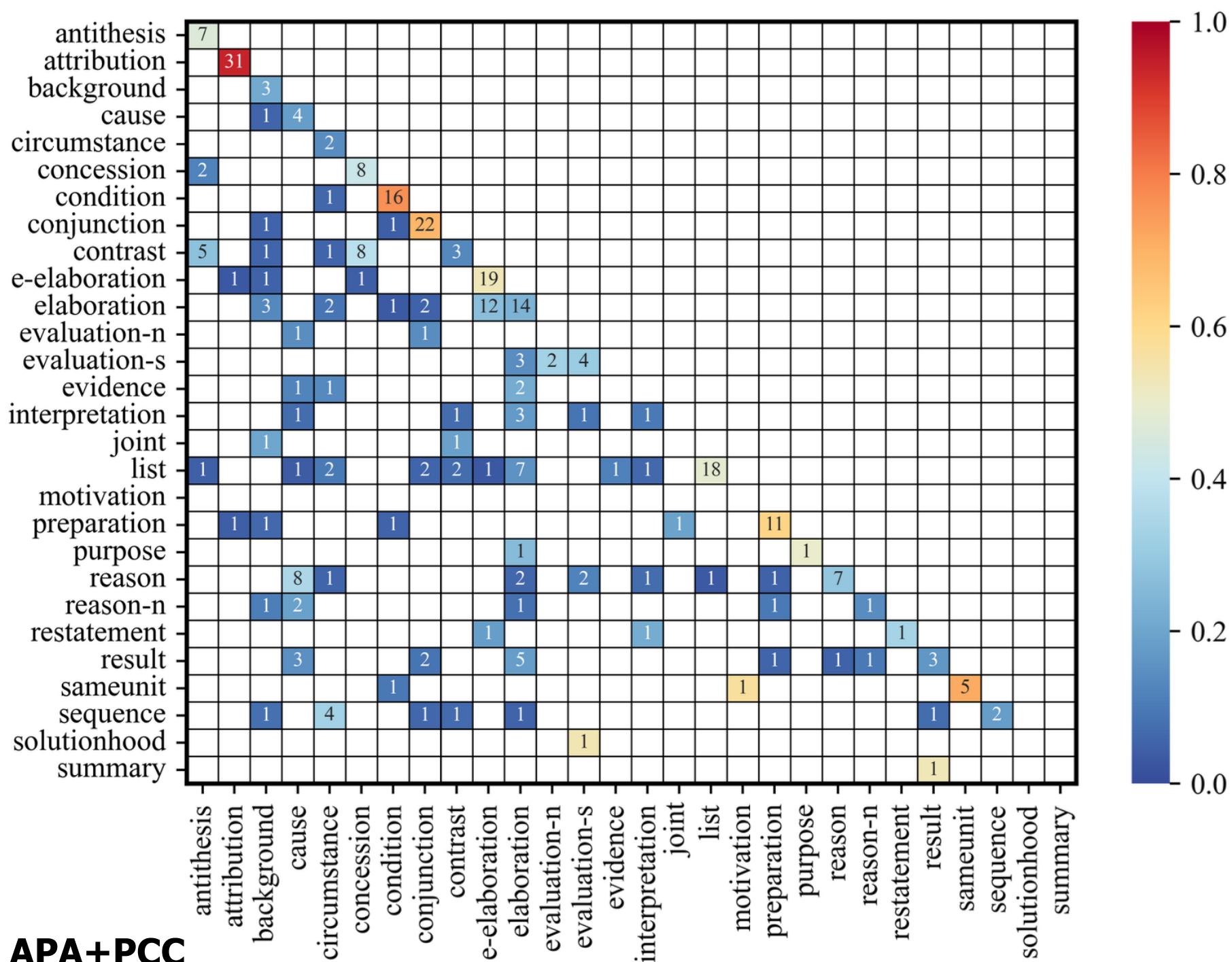


Step 1: Formal classes

Subcorpus	APA+PCC				UNSC				RST-DT			
Size	46 texts		640 EDUs		84 texts		1346 EDUs		26 texts		768 EDUs	
Agreement	N	R	C	A	N	R	C	A	N	R	C	A
	.50	.33	.46	.42	.60	.38	.55	.51	.56	.37	.53	.49
Tace output bin	<i>n</i>		Span length		<i>n</i>		Span length		<i>n</i>		Span length	
Perfect match	183 (26%)		3.1		410 (29%)		4.4		397 (49%)		5.9	
Relation mismatch	135 (20%)		3.7		288 (20%)		4.2		165 (20%)		4.9	
Scope mismatch	152 (22%)		7.1		301 (21%)		5.9		125 (15%)		11.3	
Left/right mismatch	25 (4%)		3.2		49 (3%)		3.2		8 (1%)		4.4	
No match	197 (28%)		7.4		369 (26%)		7.6		115 (14%)		13.6	

Step 2: Status of mismatches

- An „omniscient“ third annotator rates the disagreements
 - **Disagree:** one annotation is not good
 - **Both:** a comprehensive analysis would like to use both labelings
 - **Vague:** the situation can be seen this way or that way
 - **Either/or:** both annotations are plausible, but they are mutually exclusive
- So far, we analysed the **relation** and **scope** mismatches in APA+PCC (= 287 instances) and in RST-DT (= 290 instances)



Step 3: Reasons for disagreement

- Formal structural alternatives
 - E.g., a „claim“ followed by two separate „evidences“
- Relation definition overlap
 - Elaboration, Background, ... are broadly applicable
 - Contrast, Antithesis, Concession are difficult to distinguish
- Epistemic status of propositions
 - subjective / objective statement => Reason or Cause
- Presupposed knowledge / bias
 - non-/identity of referents of NPs
 - Did event A cause event B?

Example: Relation def. overlap in the original RST proposal

■ Antithesis

- *Constraint on Nucleus:* Writer has positive regard for nucleus
- *Constraints on Nucleus+Satellite:* N and S are in contrast; because of the incompatibility **one cannot have positive regard for both situations**; comprehending S and the incompatibility increases reader's positive regard for N
- *Intention of Writer:* Reader's positive regard for N is increased

■ Concession

- *Constraint on Nucleus:* Writer has positive regard for nucleus
- *Constraint on Satellite:* **Writer is not claiming that S does not hold**
- *Constraints on Nucleus+Satellite:* Writer acknowledges a potential or apparent incompatibility between N and S; recognizing the compatibility between N and S increases Reader's positive regard for N
- *Intention of Writer:* Reader's positive regard for N is increased

Step 3: Reasons for disagreement

- Assignment of „importance“ to spans
 - decides on nuclearity
 - => influences relation choice
 - => influences span range: strong nucl. principle or not?
- Global text structure
 - Topic switch (a „global“ move) versus local coherence relation
 - Genre-specific „zones“: opening – core – rhetorical ending
- Scope of adverbial connectives (and other exprs)
 - (...) *This* lead to (...)
 - (...), *Nonetheless*, it was decided to (...)

Overview

- Rhetorical Structure Theory (RST)
- Potsdam Commentary Corpus (PCC)
- PCC in the ANNIS Linguistic Database
- Human Annotation
 - Annotation Guidelines
 - Problems with RST
 - Dis/agreement analysis
- Automatic RST Parsing with Re-trained DPLP

Additional data: German

- Blog posts
 - Various publishers
 - Announcements or summaries of accompanying podcasts
 - 78 RST trees / 1,309 relations / 1,387 EDUs
- Simplified German
 - 25 texts from Austrian News Agency
 - (Simplification to levels B1 and A2)
 - 61 RST trees / 852 relations / 938 EDUs
- Total (with PCC):
315 RST trees / 5,096 relations / 5,436 EDUs

DPLP parser (Ji/Eisenstein 2014)

- Feature-based shift-reduce parser
 - position, dependency syntax, Brown clusters
- Advantages
 - well-documented
 - lightweight
 - easily adaptable to other languages
- Disadvantage
 - some outdated libraries needed replacement
- => **Publicly-available Docker image**

Results

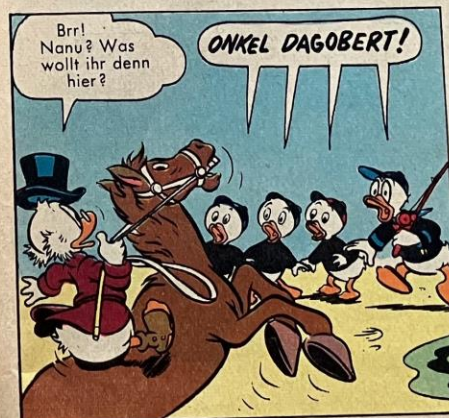
	S	N	R
PCC	0.77	0.52	0.28
PCC*	0.77	0.55	0.35
Blogs	0.81	0.61	0.40
APA	0.81	0.56	0.32
APA+Blogs+PCC*	0.78	0.56	0.36
Blogs+PCC* \rightarrow PCC*	0.77	0.54	0.34
Blogs+PCC* \rightarrow Blogs	0.82	0.64	0.43
PCC* \rightarrow APA	0.77	0.48	0.24
APA \rightarrow PCC*	0.75	0.47	0.24
APA \rightarrow Blogs	0.78	0.53	0.31
Blogs \rightarrow APA	0.76	0.45	0.21
PCC* \rightarrow Blogs	0.80	0.59	0.39
Blogs \rightarrow PCC*	0.76	0.50	0.28
Human agreement	0.85	0.65	0.44

	antithesis	attribution	background cause	circumstance	concession	condition	conjunction	contrast	e-elaboration	elaboration	evaluation-S	evidence	interpretation	list	means	preparation	purpose	reason	reason-N	restatement	result	sameunit	sequence	solutionhood	solutionhood-	summary	
antithesis	5	0	1	0	0	0	1	0	1	2	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
attribution	1	12	0	0	0	0	2	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
background	0	0	3	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
cause	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
circumstance	0	0	0	0	4	0	3	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
concession	1	0	0	0	1	19	3	2	0	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
condition	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
conjunction	1	0	0	0	0	0	0	51	0	0	1	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0
contrast	0	0	0	0	0	4	0	5	0	1	1	1	0	1	6	0	0	1	5	1	0	0	0	0	0	0	0
e-elaboration	0	0	0	0	0	0	0	1	0	23	1	0	0	0	3	0	1	0	0	0	0	0	1	0	0	0	0
elaboration	1	1	1	0	0	1	0	2	1	1	5	0	1	1	1	0	0	0	2	0	0	0	0	0	0	0	0
evaluation-S	0	0	1	0	0	2	0	1	1	0	1	0	0	1	2	0	1	0	2	0	0	0	0	0	0	0	0
evidence	0	0	0	1	0	0	0	0	2	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0
interpretation	0	0	1	0	0	1	0	1	6	1	1	1	0	2	2	0	1	0	3	0	0	0	0	0	0	0	0
list	0	3	2	0	0	2	1	8	3	4	3	1	1	1	25	0	1	1	3	0	0	0	4	0	0	0	0
means	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
preparation	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
purpose	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
reason	0	0	0	0	0	1	0	1	1	5	0	1	0	4	7	0	0	0	13	0	0	0	0	0	0	0	0
reason-N	0	0	0	0	0	0	0	0	0	0	1	1	0	2	3	0	0	0	3	0	0	0	0	0	0	0	0
restatement	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
result	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0
sameunit	0	0	0	0	0	2	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	7	0	0	0	0
sequence	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
solutionhood	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
solutionhood-N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
summary	0	0	0	0	0	0	0	1	0	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0

	additive	attribution	causal	commentary	condition	context	contrastive	elaboration	means	purpose	sameunit	sequence	summary-restatement
additive	71	2	8	6	1	3	3	9	0	1	3	0	0
attribution	3	12	0	0	3	1	0	0	0	1	0	0	0
causal	4	0	26	11	0	2	5	11	0	0	0	0	0
commentary	4	0	6	8	0	3	3	8	0	0	0	0	0
condition	0	0	0	0	12	1	0	0	0	0	0	0	0
context	2	0	1	0	2	15	1	3	0	0	0	0	0
contrastive	12	0	15	0	2	2	28	6	0	0	0	0	0
elaboration	8	0	8	1	0	2	1	26	0	0	1	0	0
means	0	0	0	0	0	0	1	0	0	0	0	0	0
purpose	0	0	1	0	1	0	0	1	0	4	0	0	0
sameunit	2	0	2	0	0	1	2	3	0	0	7	0	0
sequence	5	0	0	0	0	0	0	0	0	0	0	0	0
summary-restatement	2	0	0	1	0	0	0	1	0	0	0	0	0

Summary

- RST: a model of text coherence
- PCC: a multi-layer corpus
- ANNIS: a database for visualizing and querying multilayer corpora
- Ambiguity and vagueness in RST annotation
- Baseline parsing results for PCC/APA/Blog corpus
- Ongoing work
 - Integrate „legitimate disagreement“ into the overall workflow of annotation, analysis, parsing, evaluation (cf. **perspectivism**)
 - and ...



Thank you!

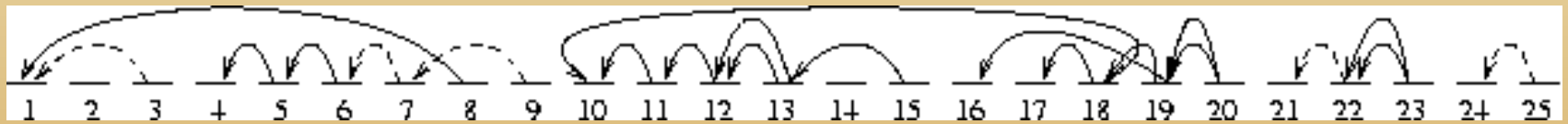
Mandatory vaccination against children's diseases?

[1] Today, children don't know anymore what pox are. [2] What a joy. [3] When pox vaccination was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] if they got themselves vaccinated. [7] For the vaccine was made from cattle's skin at the times. [8] Nowadays this dreadful disease is exterminated. [9] Thanks to a determined, world-wide vaccination campaign. [10] But there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] Because they don't know them anymore! [14] For it has been achieved with vaccines [15] that these infections hit only rarely today. [16] But those who have experienced [17] how terribly children suffer [18] when they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] As well as the long-term consequences. [21] Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day. [22] Instead, people rant about side effects [23] that occur very rarely and are known merely from books. [24] Then there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Mandatory vaccination against children's diseases?

[1] Today, children don't know anymore what **pox** are. [2] What a joy. [3] When **pox vaccination** was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] if they got themselves vaccinated. [7] For the vaccine was made from cattle's skin at the times. [8] Nowadays **this dreadful disease** is exterminated. [9] Thanks to a determined, world-wide vaccination campaign. [10] But there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] Because they don't know them anymore! [14] For it has been achieved with vaccines [15] that these infections hit only rarely today. [16] But those who have experienced [17] how terribly children suffer [18] when they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] As well as the long-term consequences. [21] Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day. [22] Instead, people rant about side effects [23] that occur very rarely and are known merely from books. [24] Then there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Referential Structure



Mandatory vaccination against children's diseases?

[1] Today, children don't know anymore what pox are. [2] What a joy. [3] When pox vaccination was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] if they got themselves vaccinated. [7] For the vaccine was made from cattle's skin at the times. [8] Nowadays this dreadful disease is exterminated. [9] Thanks to a determined, world-wide vaccination campaign. [10] But there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] Because they don't know them anymore! [14] For it has been achieved with vaccines [15] that these infections hit only rarely today. [16] But those who have experienced [17] how terribly children suffer [18] when they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] As well as the long-term consequences. [21] Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day. [22] Instead, people rant about side effects [23] that occur very rarely and are known merely from books. [24] Then there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Thematic Structure

Pox and pox vaccination									Other diseases and vaccination						Measles and pertussis					Pto Vaccination				
○									○											○				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

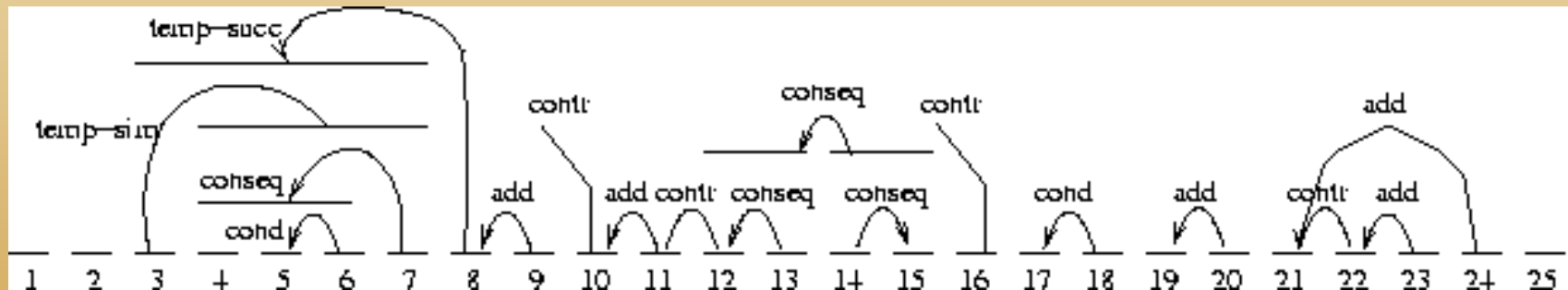
Mandatory vaccination against children's diseases?

[1] Today, children don't know anymore what pox are. [2] What a joy. [3] **When** pox vaccination was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] **if** they got themselves vaccinated. [7] **For** the vaccine was made from cattle's skin at the times. [8] Nowadays this dreadful disease is exterminated. [9] **Thanks to** a determined, world-wide vaccination campaign. [10] **But** there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] **Because** they don't know them anymore! [14] **For** it has been achieved with vaccines [15] that these infections hit only rarely today. [16] **But** those who have experienced [17] how terribly children suffer [18] **when** they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] **As well as** the long-term consequences. [21] Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day. [22] **Instead**, people rant about side effects [23] that occur very rarely and are known merely from books. [24] **Then** there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Conjunctive Relations

- temporal
 - simultaneous, succession
- consequential
 - manner, consequence, condition, purpose, concession
- comparative
 - similarity, contrast, reformulation
- additive
 - addition, alternation
- Relations can be *directed* but not *weighted* - there is no nuclearity

Conjunctive Relations



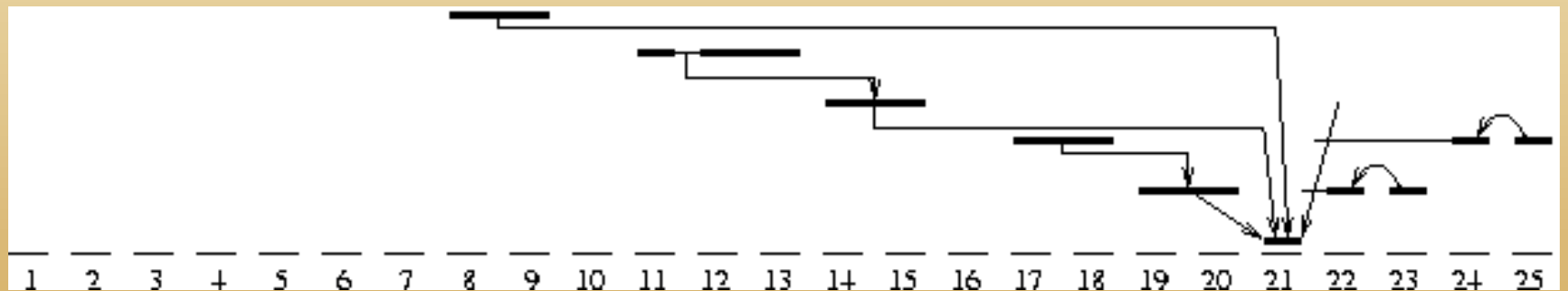
Intentional structure

- Illocutions (inspired by Schmitt 00, Searle 76)
 - *Reportivum*: writer describes a state of affairs
 - *Identifikativum*: writer characterizes own state of mind, health, etc.
 - *Estimativum*: writer presents proposition as probably true
 - *Evaluativum*: writer presents a personal opinion
 - *Appellativum*: writer orders or suggests an action
- *Support Relations* (subset of RST)
 - Ease-understanding (*Background*)
 - Encourage-acting (*Motivation*)
 - Ease-acting (*Enablement*)
 - Encourage-believing (*Evidence*)
 - Encourage appreciating (*Antithesis, Concession*)
- Compare „types of argument“ (e.g., Eggs 00):
 - deontic
 - epistemic
 - ethic/aesthetic

Mandatory vaccination against children's diseases?

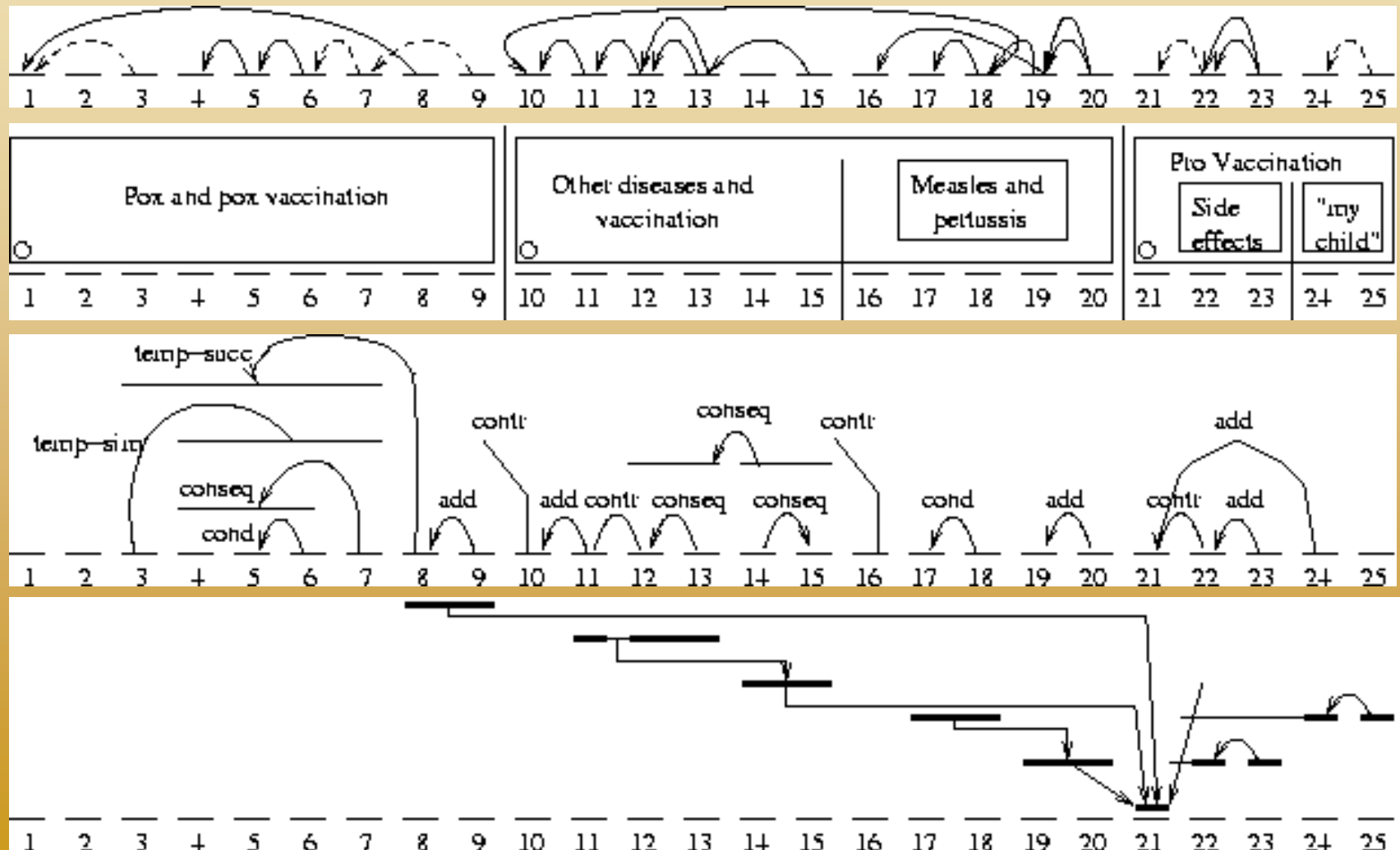
[1] Today, children don't know anymore what pox are. [2] What a joy. [3] When pox vaccination was introduced in 1854, [4] quite a few people believed [5] that their head would turn into a cow's head [6] if they got themselves vaccinated. [7] For the vaccine was made from cattle's skin at the times. [8] Nowadays this dreadful disease is exterminated. [9] Thanks to a determined, world-wide vaccination campaign. [10] But there still are other diseases: Measles, polio, diphtheria, mumps, rubella, hepatitis B, tuberculosis, pertussis. [11] Millions of children die of these, especially in less developed countries. [12] In Germany, many parents apparently don't take these diseases seriously. [13] Because they don't know them anymore! [14] For it has been achieved with vaccines [15] that these infections hit only rarely today. [16] But those who have experienced [17] how terribly children suffer [18] when they come down with ,just' measles or pertussis, [19] should spare them the agony. [20] As well as the long-term consequences. [21] Only those who have their children vaccinated will contribute to vaccines' becoming superfluous some day. [22] Instead, people rant about side effects [23] that occur very rarely and are known merely from books. [24] Then there is the great argument: This is my child, the government must not prick her. [25] No vaccine can help against such parents.

Argument structure



(inspired by Freeman 1993)

Text understanding: Relating levels of analysis



Conclusion (1):

No single complete tree no more

- Two partial hierarchies: conjunctive relations, intentional (argument) structure
- One „probably“ complete hierarchy: thematic structure
- One segment can be related (on different levels) to two other units
- Intentional (argument) hierarchy can relate non-adjacent segments
(in RST: „strong nuclearity principle“, Marcu 00)
- No nuclearity, except for *support* relations on level of intentions/argument

Conclusion (2):

No „semantic“ relations no more

- Recall Grosz/Sidner 86: There is no point in trying to enumerate types of semantic relationships
- *Cause/Result* seems intuitively straightforward, but relations like *Solutionhood* open the door to unbounded relations sets
- Cf. the *Until* relation (Rösner/Stede 92):
Add oil until the level reaches the upper mark.
- Cf. the *Report-Shower-And-Mention-Fruit* relation (Dale 93):
John took a shower. He really likes Granny Smith apples.

Conclusion (3): Benefits of MLA

- Assumptions underlying the high-level decisions (e.g., argument structure) can be made explicit: interpretation becomes more transparent
- Annotations on one level can be changed without technically affecting the others
- New levels can be introduced at leisure (e.g., modality)
- Corpus annotation is expensive - we need to be effective!
 - Resulting corpus should be as *useful* as possible
 - Annotation work can be distributed across time, space, annotators
 - Can use the „ideal“ tool for each level

Outlook:

Some research issues fo MLA

- How does sentential information structure relate to aspects of discourse structure?
- To what extent can argument structure be predicted from „lower“ levels?
- Prospects for semi-automatic annotation

Some related work

- Grosz/Sidner 86: linguistic structure, attention structure, intention structure
- Bateman 01: conjunctive relations versus „deep“ relations
- *Penn Discourse Treebank*
(Webber, Joshi et al.)
- Multi-dimension annotation discussion in the dialogue community, e.g.: *DAMSL*; Bunt 07