The Utility of Discourse Parsing Features for Predicting Argumentation Structure

Freya Hewett, Roshan Rane, Nina Harlacher, Manfred Stede
University of Potsdam
{hewett,rane,harlacher,stede}@uni-potsdam.de

Outline

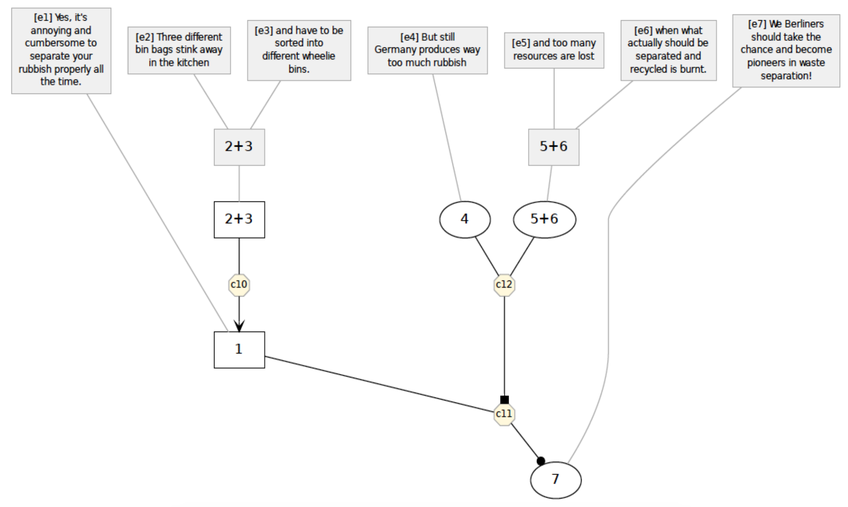
- Aim: empirically analyse the relationship between discourse structure and argumentation structure (follow-up of Peldszus & Stede 2016)

- Automatically parsed the Microtexts: Rhetorical Structure Theory (RST) and Penn Discourse Treebank (PDTB)
- Compared the parsed RST trees to the gold RST annotations
- Compared the two discourse parses to the argumentation annotation

- Reproduced the state-of-the-art Evidence Graph model for Microtexts from Afantenos et al. (2018)
- Experimented with new discourse features and features from previous experiments for predicting argumentation

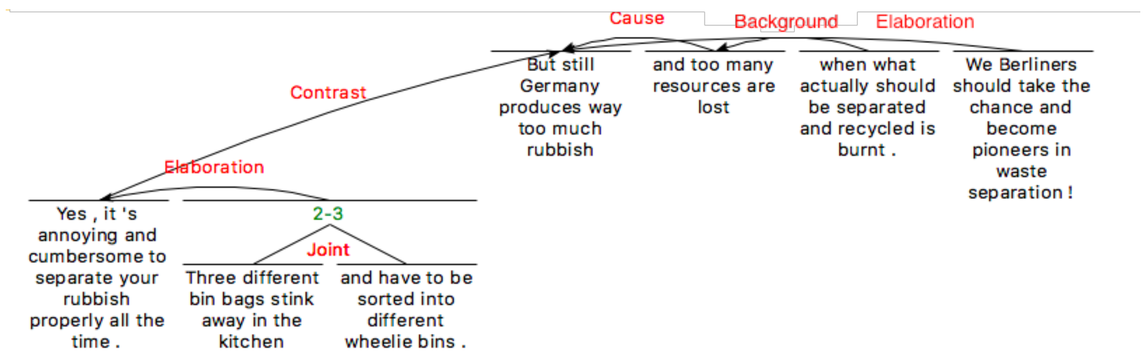
Microtexts corpus – ARG annotation

- Four levels: function (fu), role (ro), central claim (cc), attachment (at)



RST

- Represents the global coherence of the text
- Parser from Feng & Hirst (2014) with no additional training



PDTB

- Examines discourse relations by identifying explicit and implicit discourse connectives
- Annotates two arguments of the connective as well as its 'sense'
- Parser from Lin et al. (2014) with no additional training

Three different bin bags stink away in the kitchen and have to be sorted into different wheelie bins *but* still Germany produces way too much rubbish and too many resources are lost when what actually should be separated and recycled is burnt = *Comparison*
too much rubbish and too many resources are lost *when* what actually should be separated and recycled is burnt = *Temporal*

Dependency conversion

Converted the tree structures to dependencies to examine common edges in both annotations

RST vs. ARG

	reb	join	sup	und	link	exa	NONE
elaboration	22	23	88	6	4	3	115
same-unit	2	0	1	0	0	0	8
joint	2	13	1	1	10	0	32
contrast	7	2	3	28	0	0	19
temporal	0	0	0	0	0	0	1
evaluation	3	0	3	0	0	0	7
summary	0	0	0	0	0	0	1
explanation	1	1	8	1	0	1	7
cause	0	3	8	1	1	0	3
topic-comment	0	0	0	0	0	0	1
background	0	9	14	0	0	0	7
attribution	0	4	0	0	0	0	3
condition	0	14	0	0	0	0	0
enablement	0	1	1	0	0	0	0
manner-means	0	1	1	0	0	0	0
comparison	0	0	2	0	0	0	0
NONE	65	10	114	23	6	4	0

PDTB vs. ARG

	join	und	reb	sup	link	exa	NONE
Temporal.Synchrony	6	1	0	10	0	0	1
Expansion.Conjunction	3	0	2	0	2	0	24
Comparison.Contrast	0	21	5	0	0	0	18
Expansion.Alternative	0	0	1	0	0	0	0
Contingency.Cause	0	0	0	9	0	0	7
Expansion.Instantiation	0	0	0	0	0	1	0
Contingency.Condition	5	0	0	2	0	0	2
Temporal.Asynchronous	0	1	0	1	0	0	0
Comparison.Concession	1	0	2	0	0	0	1
NONE	61	25	71	189	10	2	0

When edges do match, the relations seem mostly to correspond. However, many edges have no corresponding edge in the respective annotation.

Results

	Full relation set			
features	**cc**	**ro**	**fu**	**at**
Default features	0.722 (+/- 0.068)	0.467 (+/- 0.054)	0.224 (+/- 0.015)	0.673 (+/- 0.034)
Default, RST	0.729 (+/- 0.068)	0.600 (+/- 0.049)	0.278 (+/- 0.034)	0.680 (+/- 0.033)
Default, RST, RST+	0.732 (+/- 0.068)	0.582 (+/- 0.049)	0.305 (+/- 0.048)	0.685 (+/- 0.026)
Default, PDTB	0.771 (+/- 0.073)	0.720 (+/- 0.048)	0.420 (+/- 0.056)	0.691 (+/- 0.030)
Default, RST, RST+, PDTB	0.759 (+/- 0.078)	0.721 (+/- 0.045)	0.417 (+/- 0.050)	0.703 (+/- 0.031)
Default, Afantenos et al. (2018)	**0.854 (+/- 0.057)**	**0.737 (+/- 0.052)**	0.444 (+/- 0.044)	0.720 (+/- 0.023)
All	0.852 (+/- 0.054)	0.728 (+/- 0.056)	**0.461 (+/- 0.044)**	**0.732 (+/- 0.027)**

Macro-averaged F1 score

RST parser evaluation

- Compared parser output to the gold annotations
- Converted the more fine-grained relations to the smaller set used by the parser

	Span	Nuclearity	Relation
Precision	0.372	0.288	0.125
Recall	0.309	0.243	0.106
F1	0.338	0.264	0.115

References:
Stergos Afantenos, Andreas Peldszus, and Manfred Stede. 2018. Comparing decoding mechanisms for parsing argumentative structures. Argument and Computation, 9(3):177–192.
Vanessa Wei Feng and Graeme Hirst. 2014. A linear time bottom-up discourse parser with constraints and post-editing. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 511–521.
Ziheng Lin, Hwee Tou Nh, and Min-Yen Kan. 2014. A pdtb-styled end-to-end discourse parser. Natural Language Engineering , 20:151–184.
Michael O'Donnell. 2000. Rsttool 2.4 a markup tool for rhetorical structure theory. In Proceedings of the International Natural Language Generation Conference (INLG'2000) , pages 253–256.
Andreas Peldszus and Manfred Stede. 2016. Rhetorical structure and argumentation structure in monologue text. In Proceedings of the Third Workshop on Argument Mining (ArgMining2016) , pages 103–112.

Features

For classifier	Feature description	Tag
fu, ro	Absolute and relative number of all grand/children of segment	RST+
fu, ro	Absolute and relative number of all grand/parents of segment	RST+
fu, ro	Relative count of grandchildren which occur before & after the segment	RST+
fu, ro	Relative count of grandparents which occur before & after the segment	RST+
at	Relative and absolute distance to the parent and the direction	RST+
at	Whether the segment is involved in an multi-nuclear relation	RST+
cc, fu, ro, at	Whether the segment has any PDTB connections with its neighbouring segments	PDTB
cc, fu, ro	Count of incoming & outgoing PDTB connectives (0,1 or 2)	PDTB
cc, fu, ro, at	Level one and two of the PDTB semantic relation	PDTB
cc, fu, ro, at	Raw text of PDTB connective	PDTB

Main takeaways

- Few PDTB relations identified by parser
- Evaluation of RST parser reveals many differences between predictions and gold annotations (but manual analysis confirms that many parses are plausible)

- PDTB features yielded an improved accuracy, RST+ features less so
- Overall, our new features (PDTB & RST+) helped to achieve a better F1 score for the function and attachment classifier
- The results for the simple relation set were similar, with the exception of the function classifier, which achieved an F1 score of 0.750 with the full feature set, an improvement of 0.011