Discourse-based Argument Segmentation and Annotation

Ekaterina Saveleva, Volha Petukhova, Marius Mosbach and Dietrich Klakow

The 17th Joint ACL-ISO Workshop on Interoperable Semantic Annotation

UNIVERSITÄT DES SAARLANDES

SIC Saarland Informatics Campus

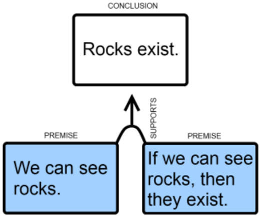
Argumentation Structure

claim or conlcusion

is an assertion that the argument aims to prove, an idea which is either true or false, put forward by somebody as true;

evidence or premise

is the proposition which give reasons or grounds for drawing the conclusion (Palau and Moens, 2009)



Discourse Structure

A **discourse relation** (rhetorical relation) is a description of how two segments of discourse are logically connected to one another.

Elementary discourse units (EDU) – a minimal building block of a discourse tree, e.g. prosodic units, turns of talk, clauses, sentences.

Discourse Annotation Schemes

A **discourse relation** (rhetorical relation) is a description of how two segments of discourse are logically connected to one another.

- Rhetorical Structure Theory
- Penn Discourse Tree Bank

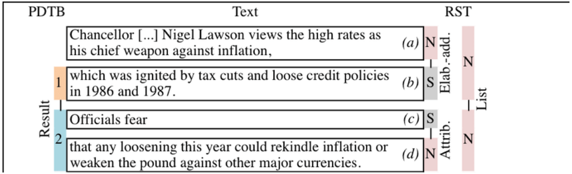


Figure 1: PDTB and RST-DT annotations for a paragraph of wsj 1172. 1 refers to Arg1 in PDTB; 2 refers to Arg2. N refers to Nucleus in RST; S refers to Satellite. (a-d) refer to RST-DT's EDUs (Demberg et al., 2019)

Discourse Relations

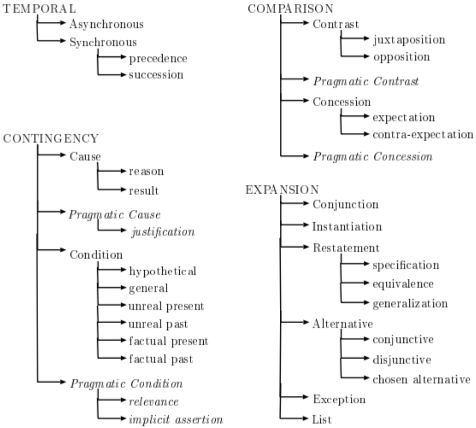


Figure 2: Hierarchy of sense tags in PDTB 2.0 (Prasad et al., 2008)

Corpora

Penn Discourse Tree Bank (Prasad et al., 2008)

- social media (WSJ)
- discourse structure, discourse semantics
- 40600 discourse unit pairs

Use of dispersants was approved when **a test on the third day showed some positive results**, officials said. (CONTINGENCY:Cause:reason)

Chairman Krebs says *the California pension fund is getting a bargain price that wouldn't have been offered to others.* In other words: **The real estate has a higher value than the pending deal suggests.** (EXPANSION:Restatement:equivalence)

Corpora

Dagstuhl15512
ArgQuality (Wachsmuth et al., 2017)

- debate portal arguments
- 15 argument quality dimensions
- 320 arguments on 15 topics

It would be better to have a lousy father than to not a have a father at all. If you were to grow up with out a father you would always have that empty void feeling about it. And if your father is lousy it helps you strive to be better than him. It could also cause you to start being responsible faster because the father isn't taking care of business. Like me and my brother, my father is lousy but I'm still happy i know who he is. Just because i cant imagine wondering about it.

- cogency: 3 (high)
- acceptability: 2 (medium)
- relevance: 1 (low)
- sufficiency: 3 (high)

Experimental Workflow

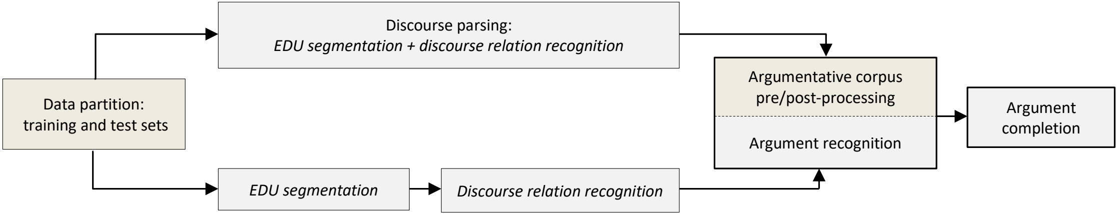


Figure 3: Experimental workflow of the argument structure recognition.

Discourse Parsing Experiments

Experimental setting	**F1 score** (%)
GS + EP (partial match)	46.80*
Auto + EP (partial match)	38.18*
GS + EP (exact match)	33.00*
Auto + EP (exact match)	20.64*
EDU span identification	22.61**
EDU span identification & relation recognition	21.20**

Table 1: Performance (F1 scores) of the PDTB parser developed by Lin et al. (2010) on various tasks. * evaluation performed on the section 23 of the PDTB 2.0 corpus; ** evaluation performed on the on full PDTB 2.0 corpus.

Discourse Parsing Experiments

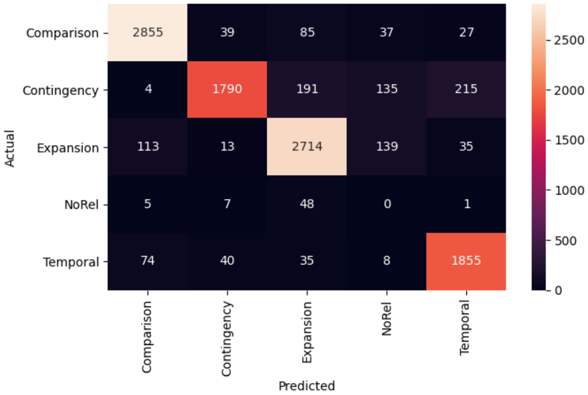


Figure 4: Confusion matrix for L1 relation classification with the PDTB parser.

Deep Learning Approach: Discourse Segmentation

- neural discourse segmentation model by Wang et al., (2018) based on BiLSTM-CRF framework.
- trained on the RST-DT dataset

Precision (%)	**Recall (%)**	**F1 (%)**
92.9	95.7	94.3

[Mrs Yeargin is lying.] [They found students in an advanced class a year earlier who said she gave them similar help.]

[If they had this much trouble with Chicago North Western,] [they are going to have an awful time with the rest.]

Deep Learning Approach: Discourse Segmentation

Statistics		F1 score (%)
total segments	123780	
exact matches	13420 (10.84%)	68.55
partial matches	56847 (45.92%)	

Table 2: Segmentation performance in terms of F1 score on PDTB 2.0 with an overview of the exact and partial matches; segmented with NeuralEDUSeg (Wang et al., 2018).

Discourse Relation Classification

[Mrs Yeargin is lying.] **(CONTINGENCY: Cause)** [They found students in an advanced class a year earlier who said she gave them similar help.]

[Mrs Yeargin is lying.] [If they had this much trouble with Chicago North Western,] **(CONTINGENCY:Condition)** [they are going to have an awful time with the rest.]

Deep Learning Approach: Discourse Relation Recognition

# classes	Statistics training set	test set	Accuracy (%)
2 classes [1]	62172	4655	88.86
5 classes [2]	19145	4655	66.37
10 classes [3]	12070	4471	53.64

Table 3: Relation recognition accuracy scores for different classification scenarios on PDTB 2.0 data; classified with XLNet model (Wang et al., 2018).

[1] Rel, NoRel

[2] Expansion, Comparison, Contingency, Temporal, NoRel

[3] Expansion.Conjunction, Expansion.Restatement, Temporal.Synchrony, Contingency.Cause, Comparison.Contrast, Comparison.Concession, Expansion.Instantiation, Temporal.Asynchronous, Contingency.Condition, NoRel

Discourse Segmentation: Argumentative Data

Match type	F1 score (%)
exact match	47.94
partial match	79.83

Table 4: Performance on EDU segmentation task with the NeuralEDUSeg model (Wang et al., 2018) on the Dagstuhl15512 ArgQuality corpus

Discourse Segmentation: Argumentative Data

[I believe] [it should not be done] [just to discipline a child.] → [I believe [it should not be done just to discipline a child.]

[Congress have no power] [to pass a legislation] [forcing religious institutions about marriage.] → [Congress have no power to pass a legislation forcing religious institutions about marriage.]

[it doesn't break the Separation between Church and State] [ruled by the Supreme Court.] → [it does n't break the Separation between Church and State ruled by the Supreme Court.]

[It would be hard for me to turn in the one] [I love.] → [It would be hard for me to turn in the one I love.]

Discourse Relation Recognition: Argumentative Data

# Classes	Accuracy (%)
5 classes	60.22
10 classes	50.48

Table 5: Performance for 5- and 10-class discourse relation classification on the Dagstuhl15512 ArgQuality corpus with the fine-tuned XLNet-large model.

Discourse Relation Recognition: Argumentative Data

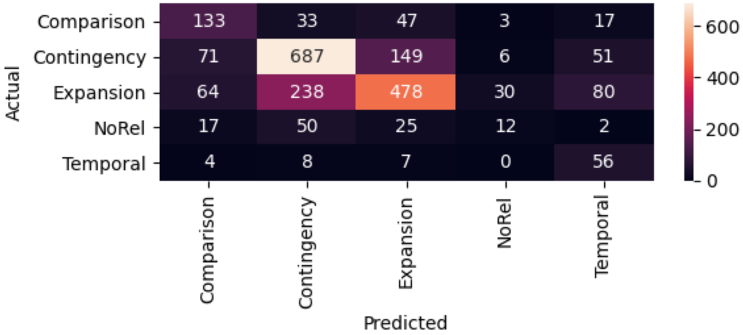


Figure 5: Confusion matrix for 5-class discourse relation classification task on Dagstuhl15512 ArgQuality.

Discourse Relation Recognition: Argumentative Data

So a lousy father is better than none. `Comparison.Concession` (that is of course assuming that he is not abusive in any way)

Physical education does absolutely nothing for the children 's health and/or lifestyle . `Expansion.Instantiation` Let me describe my PE experience. Throughout my public education career , PE has been mandatory for each year.

I think common good is better than personal pursuit `Comparison.Contrast` Yes personal pursuit is important.

it wouldn't be so easily for you to become fat `Contingency.Condition` (of course you would also need to keep a balanced diet)

Conclusions

- review of the discourse-based approaches to argumentative discourse analysis;
- discourse processing tools assessment on non-argumentative and argumentative texts;
- annotation of Dagstuhl15512 ArgQuality with discourse relations and its further release to the research community.

Future Work

- exploration of new discourse processing tools and experiments with mapping between different discourse frameworks, e.g. RST-RT to PDTB;
- reconstruction and analysis of discourse-based argumentation schemes

Thank you for your attention!