Aligning Discourse and Argumentation Structures using Subtrees and Redescription Mining

Laurine Huber, Yannick Toussaint, Charlotte Roze, Mathilde Dargnat, Chloé Braud

Terminology

Central Claims (CC) = standpoint, conclusion, nucleus in RST

Premises:

- Support = evidence, justification
- Attack
  - rebut = counter-argumentation, objection (denying the validity of a claim)
  - undercut = counter-consideration (denying relevance of a premise for a claim ; Peldszus and Stede [2013] denying the relevance of a statement in support another)

background

- arg-microtexts-multilayer corpus
  - a collection of human-subjected short texts, originally in German and professionally translated to English
  - 3 annotation schemes (RST, SDRT and ARG) use an identical segmentation
- Peldszus and Stede [2016] alignment analysis: 1-to-1 mapping have mismatches because of granularity differences in annotations
  - RST 28 relations 'fine-grained' - vs - ARG 5 relations 'coarse'
- investigate similarities between RST and Argumentative structures via subtrees based on data mining techniques (graph and redescription mining)

methodology 3-step

1. Transformation of sequentially-labeled trees from RST and ARG into two distinct unlabeled structure trees
    - texts and sequentiality are not taken into account
    - root nodes : CC
    - unlabeled nodes : DUs and AUs
    - edges : relations
    - Output = encoded RST and ARG initial trees
2. Subtrees extraction & binary data-tables (views) creation
    - Subtrees extraction : subgraph mining gSpan with a min support of 2 (the subtree graphs appear in at least 2 graphs)
    - 2 respective views : binary, rows = texts and columns = features (subtrees)
    - Output = binary views/tables of one RST and one ARG
3. Corresponding alignment of RST-ARG via redescription mining
    - ReReMi algorithm: compute Jaccard for all pairs of queries, rank and keep best *n* pairs, and apply operations (addition, deletion, edition) on queries til no more improvement on Jaccard. Heuristic and generic. Query max len = 4

1(ARG)-to-many(RST) mapping

- granularity in labeling relations
  - ARG 5 relations vs RST 28 relations
- granularity in structure
  - linked, convergent, serial, divergent, single

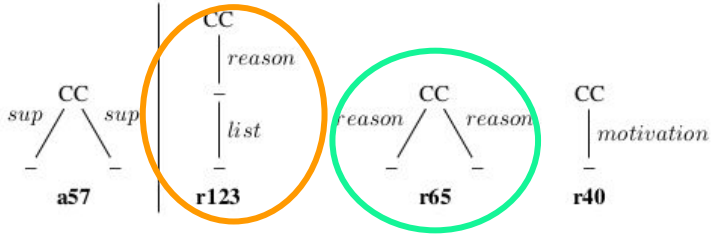

Figure 2: Subtrees corresponding to features of $Rd1$.

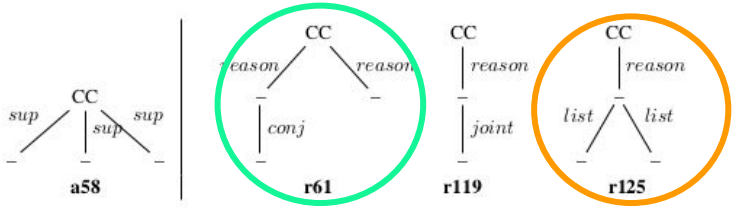

Figure 3: Subtrees corresponding to features of $Rd2$.

- parametrise ReReMi to find corresponding redescription for every atomic query?
  - Rd1 -> Rd2 : specialization
  - non-parallel redescription subtree in specialization (r119)

many(ARG)-to-1(RST) mapping

- Deeper elements also helps to clarify the redescriptions and substructures in subtrees
- Informative, despite of the low Jaccard value

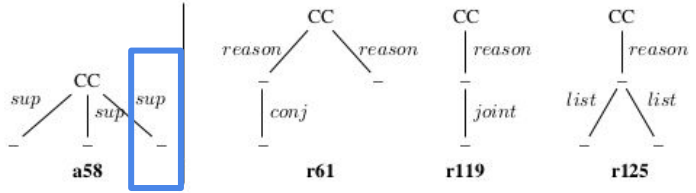

Figure 3: Subtrees corresponding to features of $Rd2$.

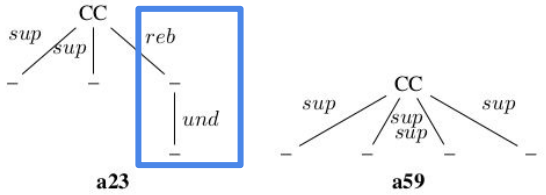

Figure 4: Subtrees corresponding to ARG features of $Rd3$, for `r125` feature see Figure 4

es of $Rd2$.

	q1	q2	J(q1,q2)	# texts
$Rd1$	`a57`	`r40 ∨ r65 ∨ r123`	0.691	54
$Rd2$	`a58`	`r61 ∨ r119 ∨ r125`	0.351	13
$Rd3$	`a23 ∨ a59`	`r125`	0.3	8

Table 1: Examples 3 redescriptions. `aX` and `rX` correspond resp. to ARG and RST subtrees.

conclusion

- Automatic process
- Reparametrize ReReMi
  - associate a conjunction of RST subtrees to each ARG subtree
  - get a higher number of redescriptions and possibly longer queries
- Include links to text segments to enable a fairer alignment between ARG and RST structures
- extend to other formalisms (SDRT)

reference

- Andreas Peldszus and Manfred Stede. From argument diagrams to argumentation mining in texts: A survey. International Journal of Cognitive Informatics and Natural Intelligence, 7(1),pages 1–31, 2013.
- Andreas Peldszus and Manfred Stede. Rhetorical structure and argumentation structure in monologue text. In 3rd Workshop on Argumentation Mining, pages 103–112, 2016a.