Aligning Discourse and Argumentation Structures using Subtrees and Redescription Mining

Laurine Huber\textsuperscript{1}, Yannick Toussaint\textsuperscript{1}, Charlotte Roze\textsuperscript{1}, Mathilde Dargnat\textsuperscript{2} and Chloé Braud\textsuperscript{1}

\textsuperscript{1}LORIA, Université de Lorraine, France
\textsuperscript{2}ATILF, Université de Lorraine, France

ArgMining
August 1, 2019
Discourse structure

- Semantic and pragmatic relations between text segments (reason, cause, concession ...)
- Rhetorical Structure Theory [Mann and Thompson, 1988] (RST)
- Distinction between nucleus and satellite

(1) People are getting older on average, (2) but they are not sicker because of it.
Argumentation Structure

- Argumentation relations between text segments (support, attack, ...)
- Macro-structure of argumentation [Freeman, 2011]
- Distinction between premisse and conclusion

1. One should not reintroduce capital punishment
2. Since no one can claim the right to rule over the life of another human being
So what?

**Goal:** Understand the similarities between discourse and argumentation structures.

- Building bridges between theories
- Improve Argument Mining systems
Corpus

- ArgMicroTexts corpus [Peldszus and Stede, 2015] *
- 112 short argumentative texts
- 18 controversial questions

"Should Germany introduce the death penalty?"

1: The death penalty is a legal means that as such is not practicable in Germany.
2: For one thing, inviolable human dignity is anchored in our constitution,
3: and furthermore no one may have the right to adjudicate upon the death of another human being.
4: Even if many people think that a murderer has already decided on the life or death of another person,
5: this is precisely the crime that we should not repay with the same.

* available online
Corpus

ArgMicroTexts corpus [Peldszus and Stede, 2015] *

112 short argumentative texts
18 controversial questions

"Should Germany introduce the death penalty?"

1: The death penalty is a legal means that as such is not practicable in Germany.
2: For one thing, inviolable human dignity is anchored in our constitution,
3: and furthermore no one may have the right to adjudicate upon the death of another human being.
4: Even if many people think that a murderer has already decided on the life or death of another person,
5: this is precisely the crime that we should not repay with the same.

* available online
Corpus

- ArgMicroTexts corpus [Peldszus and Stede, 2015] *
- 112 short argumentative texts
- 18 controversial questions

"Should Germany introduce the death penalty?"
1: The death penalty is a legal means that as such is not practicable in Germany.
2: For one thing, inviolable human dignity is anchored in our constitution,
3: and furthermore no one may have the right to adjudicate upon the death of another human being.
4: Even if many people think that a murderer has already decided on the life or death of another person,
5: this is precisely the crime that we should not repay with the same.

* available online
Corpus

- ArgMicroTexts corpus [Peldszus and Stede, 2015] *
- 112 short argumentative texts
- 18 controversial questions

"Should Germany introduce the death penalty?"
1: The death penalty is a legal means that as such is not practicable in Germany.
2: For one thing, inviolable human dignity is anchored in our constitution,
3: and furthermore no one may have the right to adjudicate upon the death of
another human being.
4: Even if many people think that a murderer has already decided on the life or death
of another person,
5: this is precisely the crime that we should not repay with the same.

* available online
Macro-structure of argumentation [Peldszus and Stede, 2016]

RST

(SDRT [Lascarides and Asher, 2007])

(a) ARG annotation  
(b) RST annotation
Overview of the approach

**Goal**: can we align ARG and RST at the subtree level?

1. Representing ARG and RST structures as trees

2. Building two descriptions of each text
   - ARG and RST descriptions
   - A description is a set of subtrees

3. Aligning set of subtrees that describe almost the same set of texts
Representing ARG and RST structures as trees

**Goal:** Unify and anonymise the structures.

- Transform ARG and RST structures into labeled trees
- Keep only structure, no text
Representing ARG and RST structures as trees

Goal: Unify and anonymise the structures.

- Transform ARG and RST structures into labeled trees
- Keep only structure, no text
Representing ARG and RST structures as trees: ARG

ARG annotation

- **Root**: central claim
- **Parent**: conclusion
- **Child**: premisse
Representing ARG and RST structures as trees: \textit{RST}

RST annotation

- **Root**: most central nucleus
- **Parent**: nucleus
- **Child**: satellite

The death penalty is a legal means that as such is not practicable in Germany.

- For one thing, inviolable human dignity is anchored in our constitution,
- and furthermore no one may have the right to adjudicate upon the death of another human being.
- Even if many people think that a murderer has already decided on the life or death of another person,
- this is precisely the crime that we should not repay with the same.

\begin{itemize}
\item CC
\item reason
\item conj
\item conces
\item RST tree derivation
\end{itemize}
Building two descriptions of the corpus

**Goal:** Produce 2 descriptions of each texts in term of subtrees

1. Extract all subtrees of ARG
2. Extract all subtrees of RST

Frequent subgraph mining: gSpan [Yan and Han, 2002]
Building two descriptions of the corpus

**Goal:** Produce 2 descriptions of each texts in term of subtrees

1. Extract all subtrees of ARG
2. Extract all subtrees of RST

Frequent subgraph mining: gSpan [Yan and Han, 2002]
Building two descriptions of the corpus: subtrees extraction

CC

sup

reb

sup

und

L. Huber (LORIA)
Building two descriptions of the corpus: subtrees extraction
Building two descriptions of the corpus: subtrees extraction

▶ \( f \) is the frequency of occurrence of subtrees in the corpus
Building two descriptions of the corpus: subtrees extraction

- keep subtrees with $f \geq 2$
Goal: Find an ARG description and a RST description that characterize almost the same set of objects

- Two different descriptions of each text
  - \( ARG = \{a_0, a_1, ..., a_{98}\} \)
  - \( RST = \{r_0, r_1, ..., r_{311}\} \)

- A set of objects: a set of texts from the corpus

- A text \( t_i \) is described by
  - a subset of \( ARG \)
  - a subset of \( RST \)
Redescription mining

**Goal:** Find an ARG description and a RST description that characterize almost the same set of objects

- Two different descriptions of the each text
  - \( \text{ARG} = \{a_0, a_1, ..., a_{98}\} \)
  - \( \text{RST} = \{r_0, r_1, ..., r_{311}\} \)

- A set of objects: a set of texts from the corpus
- A text \( t_i \) is described by
  - a subset of \( \text{ARG} \)
  - a subset of \( \text{RST} \)
Redescription mining

\[ Rd_1 : a_{57} \leftrightarrow \emptyset \]
Redescription mining

\[ Rd1 : a57 \leftrightarrow r123 \]
Redescription mining

\[ Rd1 : a57 \leftrightarrow r123 \lor r65 \]
Redescription mining

\[
Rd_{1} : a_{57} \leftrightarrow r_{123} \lor r_{65} \lor r_{40}
\]
A redescription is pair of queries
- $q_{Arg}$ a logical formulae over the $Arg$ subtrees
- $q_{Rst}$ a logical formulae over the $Rst$ subtrees

$q_{Arg}$ and $q_{Rst}$ should describe **almost** the same set of texts

"Almost": given a similarity threshold calculated with Jaccard index

$$\text{Jacc}(q_{Arg}, q_{Rst}) = \frac{\text{supp}(q_{Arg} \land q_{Rst})}{\text{supp}(q_{Arg} \lor q_{Rst})}$$
Experiment setup

- Algorithm: ReRemi
- Conjunctions and disjunctions allowed
- Length of the query limited to 4
- Output: 35 redescriptions
### Results

<table>
<thead>
<tr>
<th>id</th>
<th>q1</th>
<th>q2</th>
<th>J(q1,q2)</th>
<th># texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rd1</td>
<td>a57</td>
<td>r123 ∨ r65 ∨ r40</td>
<td>0.691</td>
<td>54</td>
</tr>
<tr>
<td>Rd2</td>
<td>a58</td>
<td>r61 ∨ r119 ∨ r125</td>
<td>0.351</td>
<td>13</td>
</tr>
<tr>
<td>Rd3</td>
<td>a23 ∨ a59</td>
<td>r125</td>
<td>0.3</td>
<td>8</td>
</tr>
</tbody>
</table>

3 over 35 obtained redescriptions

aX and rX correspond to ARG and RST subtrees respectively.
Results

Rd1 : a57 \leftrightarrow r123 \lor r65 \lor r40

RST is more fine grained than ARG
Well captured information

(a) ARG annotation

(b) RST annotation
Anonymization lead to wrong captured patterns

(a) **ARG annotation**

(b) **RST annotation**

L. Huber (LORIA)
Results

\[ Rd2 : a_{58} \leftrightarrow r_{61} \lor r_{119} \lor 125 \]

Rd2 is a specialization of Rd1
Results

$Rd3 : a_{23} \lor a_{59} \iff r_{125}$

$2 \neq \text{ARG representations of the one RST subtree}$
Conclusion

- Turn a linguistic problem into a Data Mining problem
- Systematic, generic and automatic comparison
- Understand the links between $\neq$ theories
Future work

- Take segments into account
- Play with parameters of ReReMi
- Propose an exhaustive analysis of the redescriptions
- Investigate other Data Mining formalisms 
  (e.g. FCA, association rules)
- Extend to other formalisms 
  (e.g. SDRT)
Thank you!

laurine.huber
yannick.toussaint
charlotte.roze
chloebraud
mathilde.dargnat

}@loria.fr

}@atalf.fr
Acknowledgments

This work was supported partly by the french PIA project "Lorraine Université d’Excellence“, reference ANR-15-IDEX-04-LUE, and the PEPS blanc from CNRS (INS2I).
References I


References II

An annotated corpus of argumentative microtexts.

Rhetorical structure and argumentation structure in monologue text.
Yan, X. and Han, J. (2002).
gSpan: graph-based substructure pattern mining.