Introduction
000

Corpus
00

Methodology
0000000000000000

Results
000000

Conclusion
000

Acknowledgments and references
0000

Aligning Discourse and Argumentation Structures using Subtrees and Redescription Mining

Laurine Huber[1], Yannick Toussaint[1], Charlotte Roze[1],
Mathilde Dargnat[2] and Chloé Braud[1]

[1]LORIA, Université de Lorraine, France
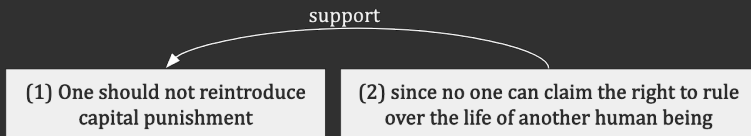[2]ATILF, Université de Lorraine, France

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
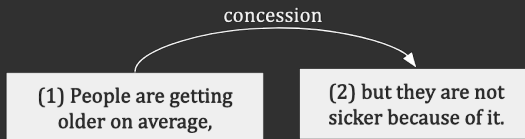
Argumentation Structure

▶ Argumentation relations between text segments (*support, attack, ...*)

▶ Macro-structure of argumentation [Freeman, 2011]

▶ Distinction between premisse and conclusion

support

| (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

Introduction
000

Corpus
●○

Methodology
000000000000000

Results
000000

Conclusion
000

Acknowledgments and references
0000

Corpus

- ▶ ArgMicroTexts corpus [Peldszus and Stede, 2015] *
- ▶ 112 short argumentative texts
- ▶ 18 controversial questions
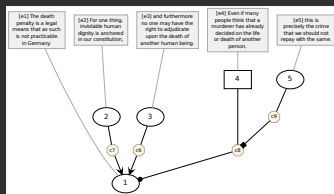
"Should Germany introduce the death penalty?"
1: The death penalty is a legal means that as such is not practicable in Germany.
2: For one thing, inviolable human dignity is anchored in our constitution,
3: and furthermore no one may have the right to adjudicate upon the death of another human being.
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,
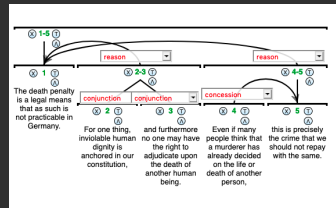**5:** this is precisely the crime that we should not repay with the same.

* available online

Introduction
000

Corpus
●○

Methodology
○○○○○○○○○○○○○○○

Results
○○○○○○

Conclusion
○○○

Acknowledgments and references
○○○○

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

Introduction
ooo

Corpus
oo●

Methodology
oooooooooooooooo

Results
oooooo

Conclusion
ooo

Acknowledgments and references
oooo

Corpus

▶ Macro-structure of argumentation [Peldszus and Stede, 2016]
▶ RST
▶ (SDRT [Lascarides and Asher, 2007])

(a) ARG annotation

(b) RST annotation

Introduction
000

Corpus
00

Methodology
●○○○○○○○○○○○○○○○

Results
000000

Conclusion
000

Acknowledgments and references
0000

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
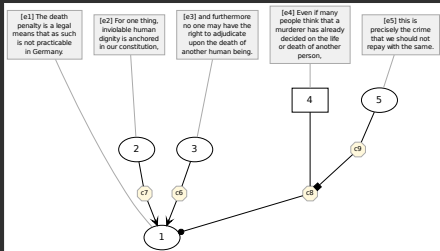
ARG tree derivation

▶ **Root**: central claim
▶ **Parent**: conclusion
▶ **Child**: premisse

Introduction
000

Corpus
00

Methodology
0000●000000000000

Results
000000

Conclusion
000

Acknowledgments and references
0000

Representing ARG and RST structures as trees : *RST*
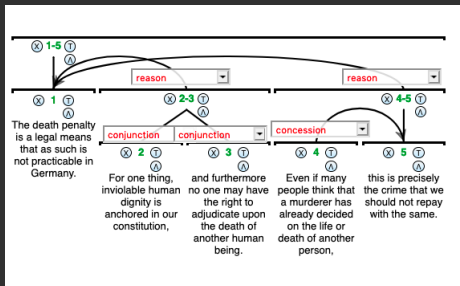
RST annotation
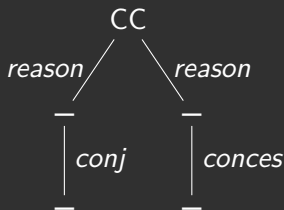
RST tree derivation

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

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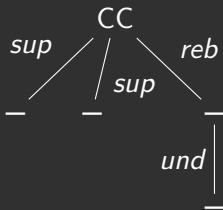
Goal: Produce 2 descriptions of each texts in term of subtrees

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

Frequent subgraph mining: gSpan [Yan and Han, 2002]

Introduction
000

Corpus
00

Methodology
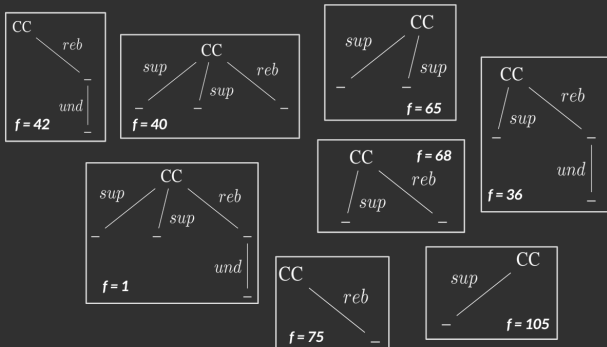00000●0000000000

Results
000000

Conclusion
000

Acknowledgments and references
0000

Building two descriptions of the corpus: subtrees extraction

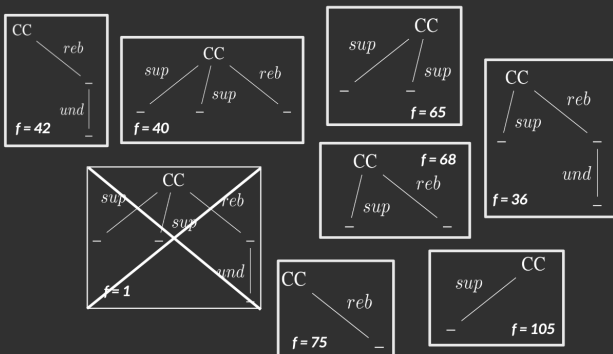# Building two descriptions of the corpus: subtrees extraction
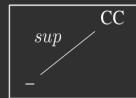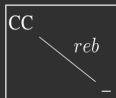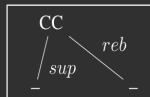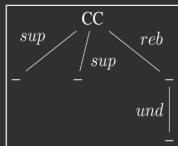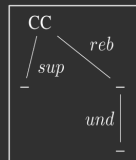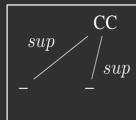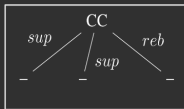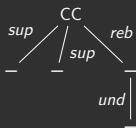
Building two descriptions of the corpus: subtrees extraction

▶ f is the frequency of occurrence of subtrees in the corpus

Building two descriptions of the corpus: subtrees extraction

▶ keep subtrees with $f \geq 2$

Redescription mining

Goal: Find an ARG description and a RST description that characterize almost the same set of objects
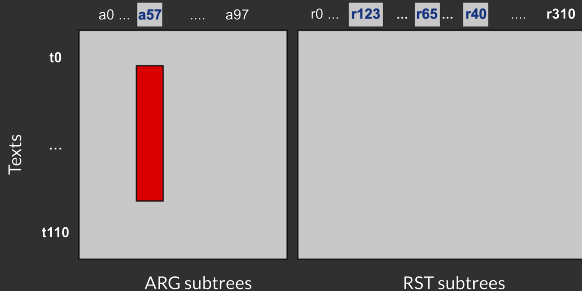
▶ Two different descriptions of the each text

    ▶ $ARG = \{a0, a1, ..., a98\}$

    ▶ $RST = \{r0, r1, ..., r311\}$

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

▶ A text t_i is described by
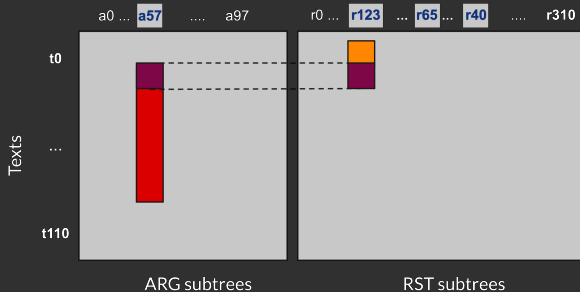
    ▶ a subset of ARG
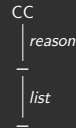
    ▶ a subset of RST

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
  ▶ $ARG = \{a0, a1, ..., a98\}$
  ▶ $RST = \{r0, r1, ..., r311\}$
▶ 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

Introduction
000

Corpus
00

Methodology
○○○○○○○○○○●○○○○○

Results
000000

Conclusion
000

Acknowledgments and references
0000

Redescription mining

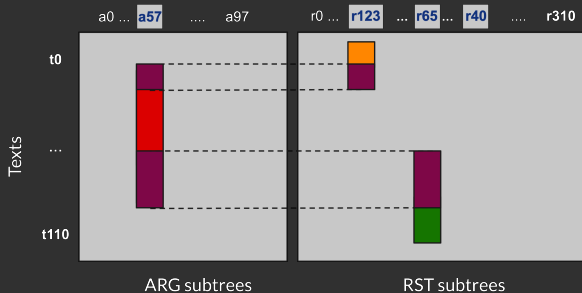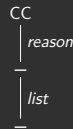$$Rd1 : a57 \leftrightarrow \varnothing$$
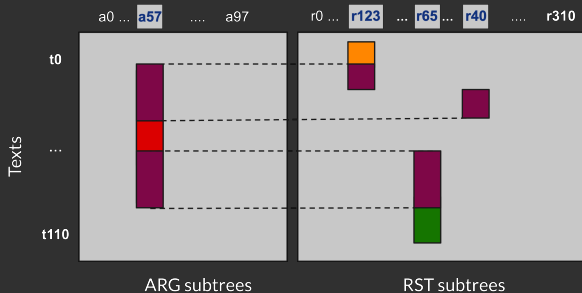
Redescription mining
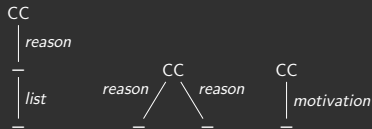

$$Rd1 : a57 \longleftrightarrow r123$$

Redescription mining

$$Rd1 : a57 \longleftrightarrow r123 \lor r65$$

Introduction
ooo

Corpus
oo

Methodology
ooooooooooooooo●oo

Results
oooooo

Conclusion
ooo

Acknowledgments and references
oooo

Redescription mining

$$Rd1 : a57 \longleftrightarrow r123 \lor r65 \lor r40$$

Redescription mining

- ▶ A redescription is pair of queries
    - ▶ *qArg* a logical formulae over the *Arg* subtrees
    - ▶ *qRst* a logical formulae over the *Rst* subtrees
- ▶ *qArg* and *qRst* should describe **almost** the same set of texts
- ▶ *"Almost"*: given a similarity threshold calculated with Jaccard index

$$Jacc(qArg, qRst) = \frac{supp(qArg \land qRst)}{supp(qArg \lor qRst)}$$

Experiment setup

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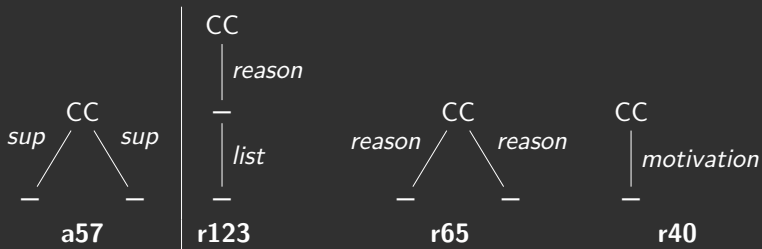- ▶ Length of the query limited to 4
- ▶ Output: 35 redescriptions

Results

id	q1	q2	J(q1,q2)	# texts
*Rd*1	a57	r123 ∨ r65 ∨ r40	0.691	54
*Rd*2	a58	r61 ∨ r119 ∨ r125	0.351	13
*Rd*3	a23 ∨ a59	r125	0.3	8

3 over 35 obtained redescriptions
aX and rX correspond to *ARG* and *RST* subtrees respectively.
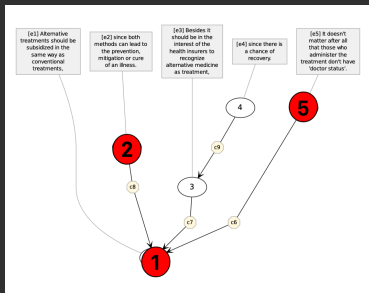
Results

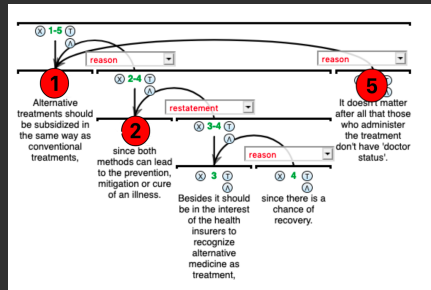$$Rd1 : a57 \longleftrightarrow r123 \vee r65 \vee r40$$

RST is more fine grained than ARG
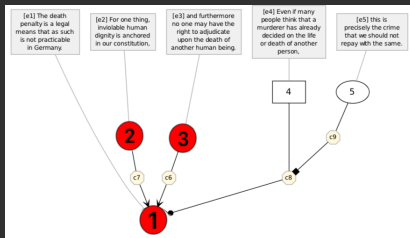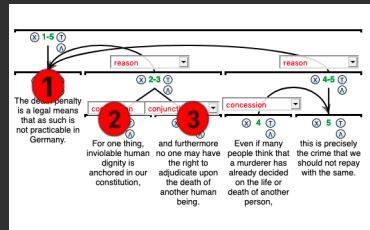
Well captured information

(a) ARG annotation

(b) *RST* annotation

Anonymization lead to wrong captured patterns
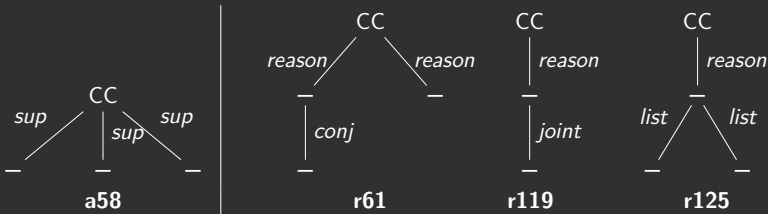

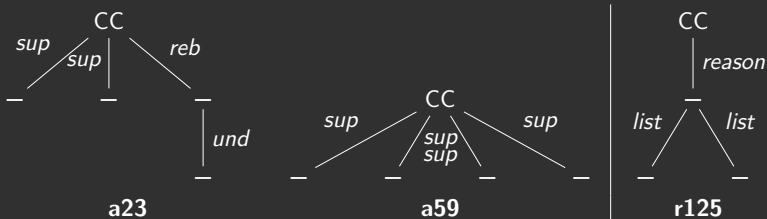(a) *ARG* annotation

(b) *RST* annotation

Results

$$Rd2 : a58 \longleftrightarrow r61 \vee r119 \vee 125$$

Rd2 is a specialization of Rd1

Results

$$Rd3 : a23 \lor a59 \longleftrightarrow r125$$

a23 **a59** **r125**

$2 \neq$ **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 ≠ 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)

Introduction
000

Corpus
00

Methodology
000000000000000

Results
000000

Conclusion
00●

Acknowledgments and references
0000

Thank you!

laurine.huber
yannick.toussaint
charlotte.roze   @ loria . fr
chloe.braud
mathilde.dargnat   @ atilf . fr

Acknowledgments

References I

📄 Freeman, J. B. (2011).
Argumentation Structure: Representation and Theory.
Springer, Dordrecht.

📄 Lascarides, A. and Asher, N. (2007).
Segmented Discourse Representation Theory: Dynamic
Semantics With Discourse Structure.
In Bunt, H. and Muskens, R., editors, *Computing Meaning*,
volume 3. Springer Netherlands, Dordrecht.

📄 Mann, W. and Thompson, S. (1988).
Rhetorical structure theory: Towards a functional theory of
text organization.
TEXT, 8:243–281.

References II

Peldszus, A. and Stede, M. (2015).
An annotated corpus of argumentative microtexts.
In *Proceedings of the First European Conference on Argumentation: Argumentation and Reasoned Action*, volume 2, pages 801–816.

Peldszus, A. and Stede, M. (2016).
Rhetorical structure and argumentation structure in monologue text.
In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 103–112, Berlin, Germany. Association for Computational Linguistics.

References III

📄 Yan, X. and Han, J. (2002).
gSpan: graph-based substructure pattern mining.
In *2002 IEEE International Conference on Data Mining, 2002.
Proceedings.*, pages 721–724, Maebashi City, Japan. IEEE.