Towards
Assessing Argumentation Annotation —
A First Step

Anna Lindahl, Lars Borin & Jacobo Rouces

University of Gothenburg

August 1, 2019

Introduction

- Annotation of Swedish news editorials with Walton's argumentation schemes.
- Initial effort to evaluate the suitability and usefulness of these schemes for argumentation mining.

Argument schemes

- Walton's argumentation schemes are made up by a set of premises and a conclusion, and a label for the scheme.

ARGUMENT FROM CONSEQUENCES:
Premise: If A is brought about, then good (bad) consequences will (may plausibly) occur.
Conclusion: A should (not) be brought about.

Data set

- 30 editorials from Swedish newspapers (1973).
- Total about 19,000 words, on average 640 words/editorial.
- Originally compiled by Hedquist[1], also annotated with emotive language.

---

[1]Rolf Hedquist. 1978. Emotivt språk: En studie idagstidningars ledare [Emotive language: A studyin newspaper editorials]. Umeå University, Dept. of Nordic Languages, Umeå.

Annotation

- Two annotators with linguistic training.
- Instructed to use Walton et al.'s book on Argumentation schemes[2], no further instructions.
- An argument consists of a conclusion and one or more premises, plus a scheme.
  - Any span of text can be a conclusion or premise.
  - No pre-annotated structures.

---

[2]Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. Argumentation Schemes. Cambridge University Press, Cambridge

Example of an annotated argument

Premise: 'A shift of power will result in us not risking any socialistic experiment during the elected term and instead we can further build on the foundations of the welfare society.'
Conclusion: 'Voters should vote for the opposition'
Scheme: ARGUMENT FROM CONSEQUENCES

Results

- Annotator 1 annotated more arguments than annotator 2.
- Annotator 2 annotated more premises per argument on average.

	Annotator 1	Annotator 2
No. of arguments	345	195
Avg. no. of premises per arg.	1.26	2.03
Total no. of units	782	591

Results cont.

- The same schemes among the most used for both annotators, except the top used scheme.

A1	Count	A2	Count
Evidence to a Hypothesis	105	Correlation to Cause	42
Consequences	90	Sign	22
Sign	47	Consequences	20
Cause to Effect	30	Cause to Effect	18
Falsification of a Hypothesis	30	Popular Practice	17

Inter-annotator agreement (IA)

- IA is measured according to below:

$$IA = 2 * m/(a_1 + a_2) \tag{1}$$

- where m is the number of matches, $a1$ and $a2$ is the number of annotated conclusions, premises or schemes for respective annotator.
- Two conclusions or premises are considered as matching if their string overlap is above a threshold, α, of 0.9 or 0.5.
- m is also used for comparing matching schemes, but then no overlap is used.

Conclusions

- More matches and higher IA for lower overlap ratio.

Conclusions	α	
	0.9	0.5
m	71	92
IA	0.26	0.34

Premises

- Given a matching conclusion, premises were compared in two ways:
  - At least one premise matches.
  - All premises match.
- Of the previous 71 matching conclusions, 20 have at least one premise matching.

	α	
At least one matching premise	0.9	0.5
m	20	33
IA, within matching conclusions	0.56	0.71
IA, within all arguments	0.07	0.12
All premises match		
m	6	9
IA, within matching conclusions	0.17	0.20
IA, within all arguments	0.02	0.03

Premises

- Premises without conclusions(α=0.9)
  - 74 arguments where at least one premise matches.
  - 14 arguments where all premises match.
- The same premise can be used for different conclusions, and a conclusion can have different premises.

Different premises, same conclusion

Premise A1: 'It is already showing in the form of increasing oil and gas prices.'
Premise A2: 'We are not especially used to saving anything in this country.'
Conclusion A1 & A2 : 'But now the energy crisis is not far away'
Scheme A1: ARGUMENT FROM SIGN
Scheme A2: ARGUMENT FROM CAUSE TO EFFECT

Same premise, different conclusions

Premise A1 & A2 : 'A shift of power will result in us not risking any socialistic experiment during the elected term and instead we can further build on the foundations of the welfare society.'

Conclusion A1: 'Voters should vote for the opposition'

Conclusion A2: 'Do not vote away collaboration!'

Scheme A1: ARGUMENT FROM CONSEQUENCES

Scheme A2: CAUSAL SLIPPERY SLOPE ARGUMENT

Schemes

- Given both a matching conclusion and all premises, 2 schemes matched. (for $\alpha=0.9$)
- Comparing only matching conclusions results in higher IA (9 matches).
- Comparing only premises has 3 scheme matches.

	α	
Scheme matches, given conclusion	0.9	0.5
m	9	10
IA, within matching conclusion	0.25	0.22
IA, within all arguments	0.02	0.02

Groups of schemes

- Groups suggested in Walton et al.'s book as a classification system for schemes.
- The groups resulted in 3 matches with both conclusion, premises and scheme.
- Comparing only conclusions increased IA from 0.25 to 0.48 (17 instead of 9 matches).
- Comparing only premises gave 4 matches.

	α	
Matching schemes	0.9	0.5
m	3	7
IA, within matching	0.08	0.15
IA, within all arguments	0.01	0.03

Co-occurring schemes

- Argument from consequences and Argument from popular practice co-occur much more than the other schemes. (12 times.)

ARGUMENT FROM CONSEQUENCES:
Premise: If A is brought about, then good (bad) consequences will (may plausibly) occur.
Conclusion: A should (not) be brought about.

ARGUMENT FROM POPULAR PRACTICE:
Premise: If a large majority (everyone, nearly everyone, etc.) does A, or acts as though A is the right (or an acceptable) thing to do, then A is a prudent course of action.
Premise: A large majority acts as though A is the right thing to do.
Conclusion: A is a prudent course of action.

Conclusions & Future work

- The annotators differ a lot, this could be because of
  - ▶ The instructions.
  - ▶ The structure of the task.
  - ▶ The schemes themselves.
- Groups improved the results.

- Future work:
  - ▶ Same schemes, new instructions.
  - ▶ Groups of schemes, new instructions.
  - ▶ Possibly change the annotation task.
  - ▶ New argumentation model/scheme.

Thank you for listening!