

# Developing an argument annotation scheme based on a semantic classification of arguments

Lea Kawaletz<sup>1</sup>, Heidrun Dorgeloh<sup>1</sup>, Stefan Conrad<sup>2</sup> and Zeljko Bekcic<sup>2</sup>

<sup>1</sup>Department of English and American Studies, <sup>2</sup>Department of Computer Science, HHU Düsseldorf, Germany

## Arguments

Arguments are functionally complex

- Argument components (e.g. Stab and Gurevych 2014)
  - Claim:** Controversial statement, provides the topic of the argument
  - Premise:** Supports or attacks the claim by providing evidence or expressing reasoning

polarity	epistemic	ethical	deontic
positive	x is true	x is good	do x
negative	x is false	x is bad	don't do x

## The overall goal

- Create an annotated inventory of arguments
  - Linguistic & computational experiments
  - Enrich discourse and genre theory & improve machine learning

## The annotation scheme

We have developed a repository of semantic templates

category	positive claim	negative claim
epistemic		
support	and this is true because	and this is false because
attack	but this is not true because	but this is not false because
ethical		
support	and this is good because	and this is bad because
attack	but this is bad because	but this is good because
deontic		
support	and this should be done because	and this should not be done because
attack	but don't do this because	but do do this because



- (1) a. [M]asking should be mandated and enforced.  
 b. and this should be done because [i]t's not just about your individual risk tolerance, but about keeping everyone safe.

## The problem

Argument identification is hard

- Argumentation is often not straightforward
  - Arguments are linguistically variable (e.g. Dorgeloh and Wanner 2010)
  - A lot of information is implicit (e.g. Moens 2018)

- Even trained annotators have a hard time identifying claims and premises
- Argument mining approaches typically rely on recurrent patterns

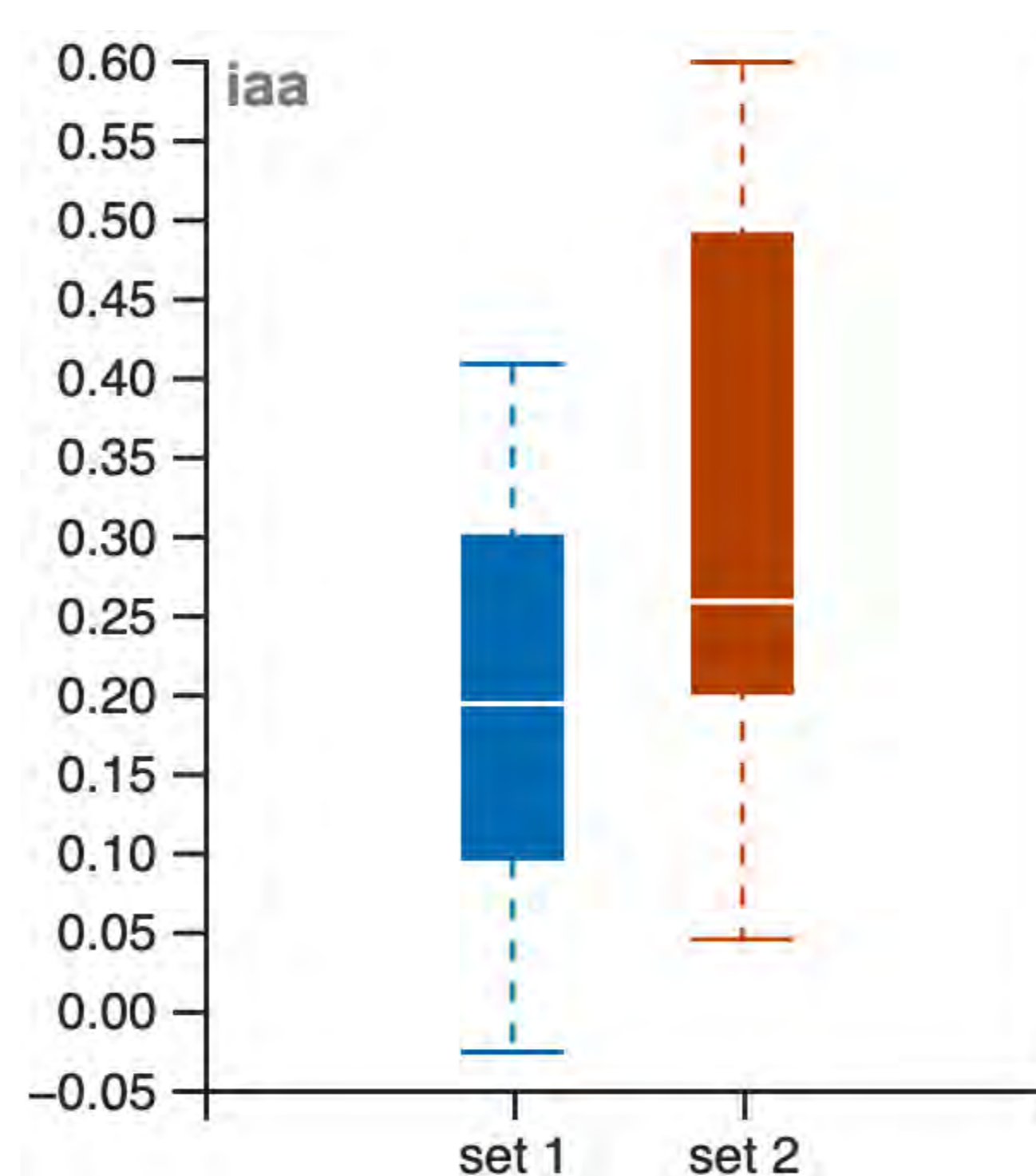
## The idea

- Utilize functional complexity of argumentation
  - Operationalize the annotation scheme by using semantic templates
  - Develop a topic-independent classification scheme

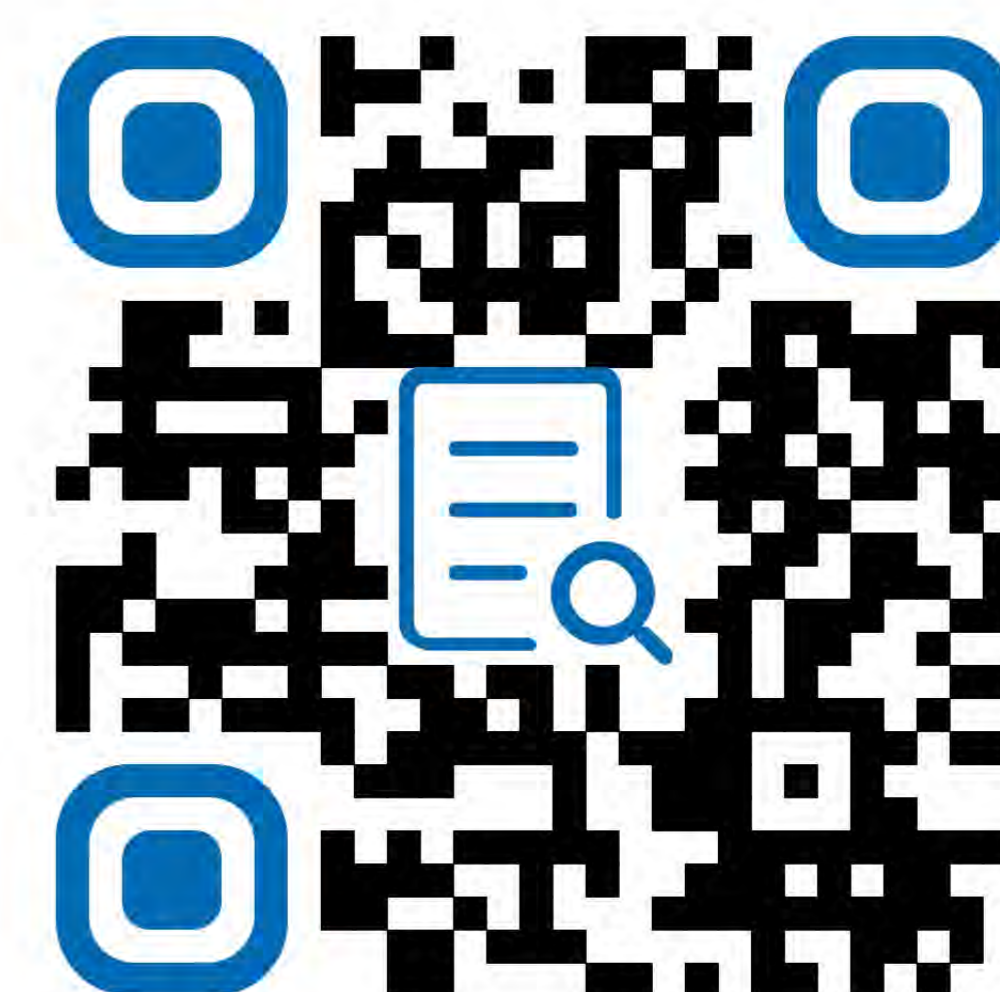
## Results

We find a clear trend toward increasing inter-annotator agreements

Annotators report an overall improvement of the process



- Facilitated argument identification
- Smoother discussions
- Difficult claim classification



## Inter-annotator agreements

The inter-annotator agreements of set 1 are highly insufficient

Classifying claims is not harder than identifying them

subset id	iaa	# of annotators	subset id	iaa	iaa	# of
				p/c/∅	p/ep/et/d/∅	annotators
1-1	10	0.2713	3			
1-1	11	0.4078	3			
1-1	12	0.2646	3			
1-2	13	0.1932	3			
1-2	14	-0.0268	3			
1-2	15	0.3851	3			
1-2	16	0.3002	3			
1-2	17	0.0123	3			
1-2	18	0.1705	3			
1-2	19	0.0941	3			
1-2	20	0.3853	3			
1-2	21	0.1891	3			
1-2	22	0.0681	3			
2-1	23	0.1811	0.181	4		
2-1	24	0.2809	0.2657	4		
2-2	25	0.0951	0.113	3		
2-2	26	0.2516	0.2798	3		
2-2	27	0.5906	0.5953	3		
2-2	28	0.2496	0.2391	3		
2-2	29	0.0446	0.0428	3		
2-2	30	0.2638	0.2623	3		
2-2	31	0.5531	0.5525	3		
2-2	32	0.2046	0.2027	3		
2-2	33	0.5982	0.5829	3		
2-2	34	0.4704	0.4753	3		

## Corpus compilation

- COVID-19-related news opinion texts from *The New York Times*
  - Set 1: 13 texts, 15,299 words
  - Set 2: 12 texts, 14,167 words
  - Total: 25 texts, 29,466 words
- Annotated for argument components (sets 1 & 2) and functions (set 2) using the INCEpTION tool (Klie et al. 2018)
- Monitoring iaa (Krippendorff's unitizing alpha, Krippendorff et al. 2016)

## Next steps

- Refine work flow further → Improve iaa, annotator recall & corpus
  - Only editorials
  - Text length 40-70 sentences
  - Pre-assessment of texts
  - Curation of annotations

## Selected references

- H. Dorgeloh and A. Wanner, editors. *Syntactic Variation and Genre*. Number 70 in Topics in English Linguistics. De Gruyter Mouton, Berlin/New York, 2010.
- J.-C. Klie, M. Bugert, B. Boulosa, R. E. de Castilho, and I. Gurevych. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. ACL, 2018.
- K. Krippendorff, Y. Mathet, S. Bouvry, and A. Widlöcher. On the reliability of unitizing textual continua: Further developments. *Quality & Quantity*, 50(6):2347–2364, 2016.
- M.-F. Moens. Argumentation mining: How can a machine acquire common sense and world knowledge? *Argument & Computation*, 9(1):1–14, 2018.
- C. Stab and I. Gurevych. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014*, pages 1501–1510, Dublin, 2014. Dublin City University and ACL.