

# Argumentation Technology for Artificial Intelligence

## Part 2: Argument Mining and Assessment

Philipp Cimiano, Khalid Al-Khatib / Benno Stein, **Henning Wachsmuth**

September 21, 2020



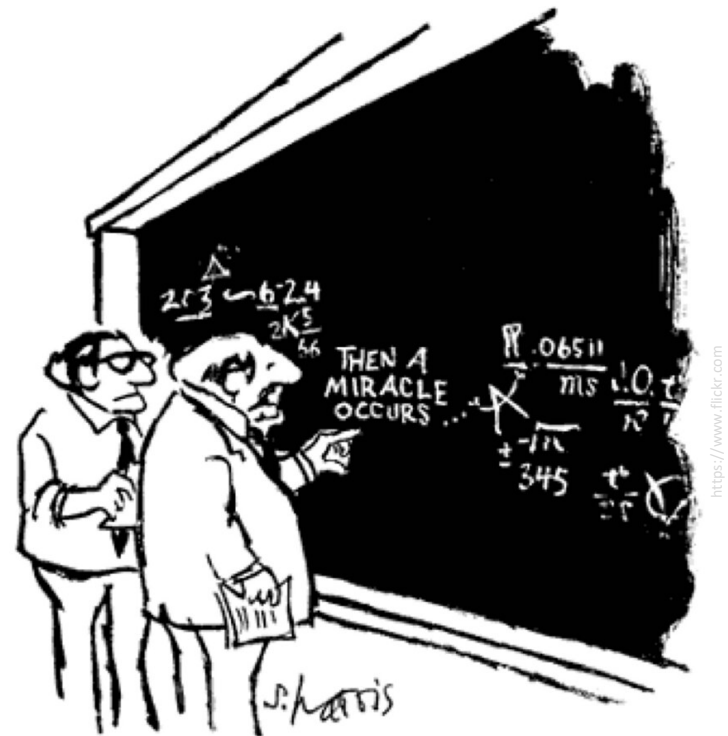
Bauhaus-Universität  
Weimar



# Argumentation in natural language

*” If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people. ”*

- **Real-world arguments**
  - Mostly not logically valid
  - Leave much implicit
  - May be hidden in longer texts
  - May be split over multiple texts
  - May depend on the context
- **Can we use formal approaches?**
  - Yes, but we need to *mine* arguments and *assess* their properties before
  - Natural language processing needed



“I think you should be more explicit here in step two.”

# Argument mining and assessment: Outline

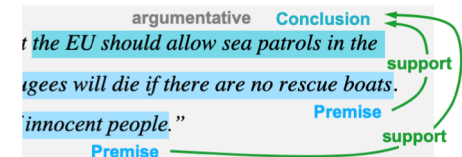
## 1. Stance classification

- How to determine whether an argument is pro or con?



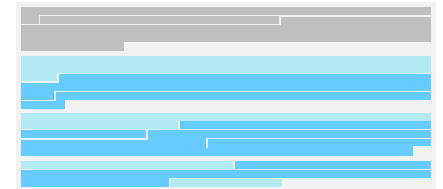
## 2. Argument mining

- How to find argument units and relations in text?



## 3. Argument mining for writing support

- How to leverage the output of argument mining?



## 4. Argumentation quality assessment

- How to judge whether an argument is good or bad?



## 5. Fallacy detection in online discussions

- How to identify argumentative flaws and their triggers?

"Again, how old are you?"

# 1. Stance Classification

# Stance classification: Introduction

- **Stance**

- Overall position of a person towards some target, such as an issue or statement
- To have/take a stance on a target means to be *pro* or *con* towards it (Somasundaran and Wiebe, 2010)



<https://pixabay.com>

- **Stance classification**

- Determination of the stance encoded in a text span towards a target
- Usually, the target is the (given) issue at discussion

**Con** towards "banning rescue boats"

*"If you wanna hear my view, I think that **the EU should allow sea patrols in the Mediterranean Sea.** **Many innocent refugees will die if there are no rescue boats.** **Nothing justifies to endanger the life of innocent people.**"*

**Pro** towards claim above

**Pro** towards claim above

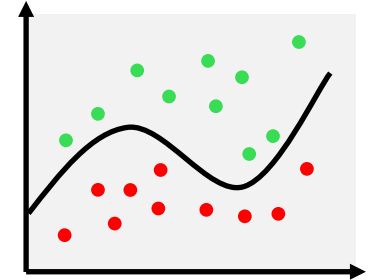
- **Notice**

- Stance classification usually comes *after* argument mining  
Here discussed first, because of its conceptual simplicity

# Background: Supervised text classification

- **Text classification**

- **Task.** Given a text, assign one class from a set of classes  
Stance classification is a text classification problem
- Usually done with supervised machine learning



- **Feature-based classification**

- Map text to feature vector, map feature vector to class label  
Features engineered manually or semi-automatically
- **Models.** Support vector machines, random forest, ...

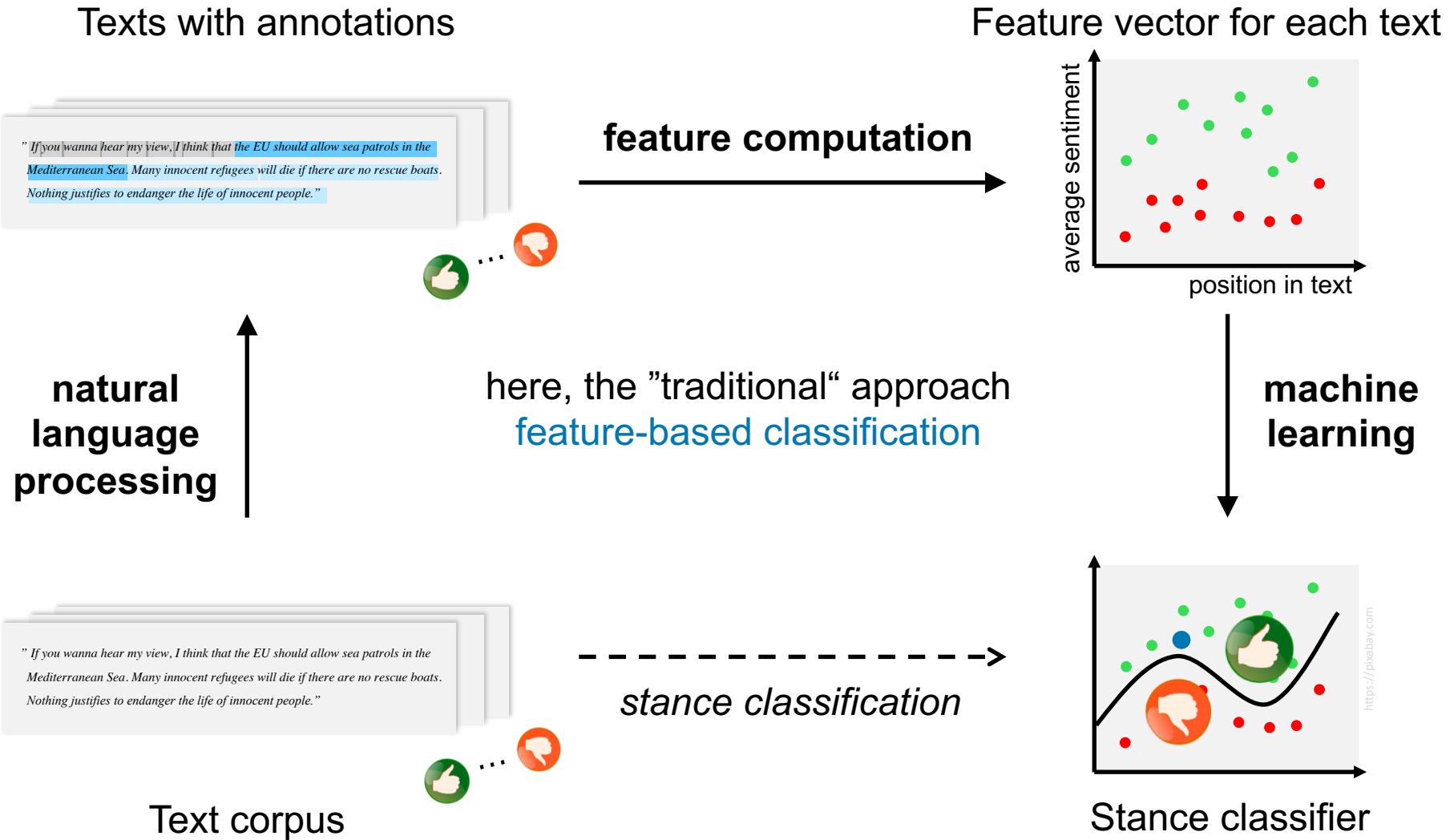
- **Neural classification** (usually works better, given enough data)

- Features (weights in neural networks) learned automatically
- **Models.** Convolutional neural networks, bi-directional LSTMs, ...

- **Sequence labeling** (applicable when a sequence of texts is classified)

- Like other techniques, but considering previous classifications
- **Models.** Conditional random fields, recurrent neural networks, ...

# How to develop a stance classification algorithm



# Background: Evaluation measures

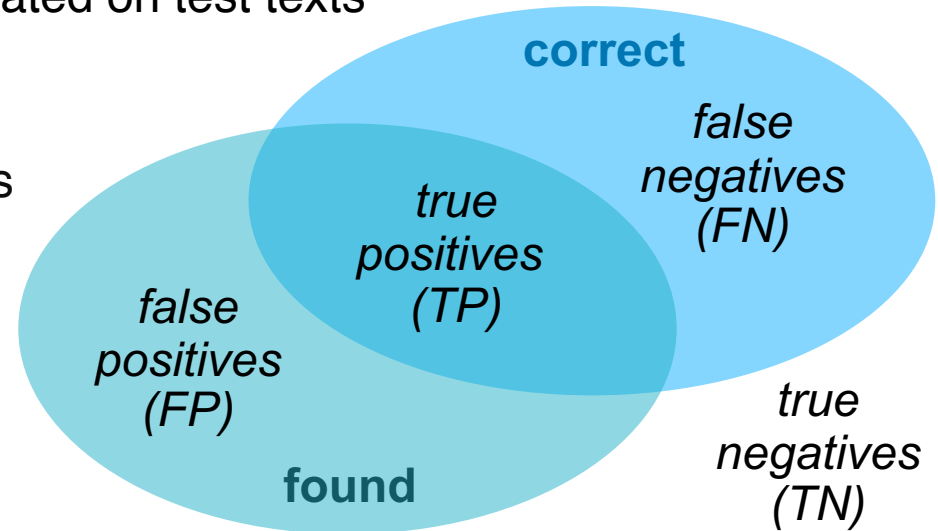
- **NLP is data-driven**

- Methods developed on training texts, output not always correct
- Effectiveness of methods evaluated on test texts

- **Effectiveness measures**

- **Accuracy.** Used if both positives and negatives important

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$



- **Precision, recall, and F<sub>1</sub>-score.** Used if positives in the focus

$$\text{Precision (P)} = \frac{TP}{TP + FP} \quad \text{Recall (R)} = \frac{TP}{TP + FN} \quad \text{F}_1\text{-score} = \frac{2 \cdot P \cdot R}{P + R}$$

- **Mean absolute/squared error.** Often used for numerical scores



# Modeling stance

- **Candidate features of the text** (Somasundaran and Wiebe, 2010, Hasan and Ng, 2013)

- **Bag-of-words.** Distribution of words or word n-grams
- **Core vocab.** Terms from subjectivity lexica
- **Discourse.** Connectives + relations between units
- **Sentiment.** Aspect or topic-directed polarity

... and many more...

→ accuracy up to 0.70

- **Candidate features of the context**

- Exploit author knowledge in dialog → up to 0.74 (Ranade et al., 2013)
- Exploit opposing views in dialog → up to 0.75 (Hasan and Ng, 2013)
- Connections between topics of claim and target (Bar-Haim et al., 2017)

→ 0.84 for most confident 10%, 0.65 overall (3 classes)

Alice: The EU should allow **sea patrols** in the Mediterranean Sea, to save the innocent refugees.

↓  
**stance tend to be opposite**

Bob: So naïve... having rescue boats makes even more people die trying.

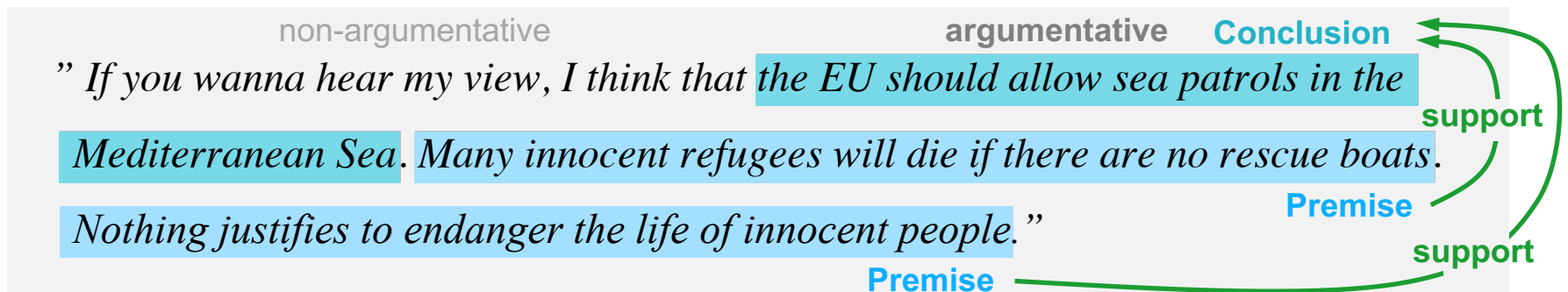
↘  
**stance tend to be the same**

Alice: Preventing **asylum seekers** from trying is another issue. For now, we need to support them.

## 2. Argument Mining

# Argument mining: Introduction

- **Argument mining** (aka argumentation mining)
  - Automatic identification of arguments in natural language text
- **Three main steps** (variations found in literature)
  - Segmenting a text into argument units and other parts
  - Classifying the type/role of each unit
  - Identifying and classifying relations between units



- **Why argument mining?**
  - Real-world arguments often "hidden" in longer text, possibly fragmented
  - Mining is the basis for any argument analysis and any application
    - Exception: Arguments, and their structure, already given in the source data

# Unit segmentation

- **Argument units** (aka argumentative discourse units)
  - Text segments with an argumentative function  
Usually, the premises and conclusions of arguments
- **Unit segmentation**
  - **Task.** Given a text, segment it into argument units and other parts
  - **Method.** Usually, token-level sequence labeling

non-argumentative

argumentative

*” If you wanna hear my view, I think that the EU should allow sea patrols in the Mediterranean Sea. Many innocent refugees will die if there are no rescue boats. Nothing justifies to endanger the life of innocent people.”*

- **State of the art** (Ajjour et al., 2017)
  - Rather reliable on narrow genres ( $F_1$  0.72–0.82), unsolved across them
  - Unit granularity differs: Anything between clauses and paragraphs

# Unit classification

- **Unit classes**

- Claim and evidence types  
(Rinott et al., 2015; Al-Khatib et al., 2017)

assumption statistics other  
anecdote common ground testimony

- Roles within argumentation  
(Stab and Gurevych, 2014; Habernal and Gurevych 2015)

conclusion premise thesis none

- **Unit classification**

- **Task.** Given an argument unit, assign one class from a set of classes
- **Method.** Usually, supervised text classification

Conclusion

” If you wanna hear my view, I think that *the EU should allow sea patrols in the Mediterranean Sea.* *Many innocent refugees will die if there are no rescue boats.*

Premise

*Nothing justifies to endanger the life of innocent people.*”

Premise

- **State of the art**

- Reliable on ”explicit” argumentation, such as essays ( $F_1$  0.87) (Stab, 2017)
- Still rather reliable on news editorials ( $F_1$  0.77) (Al-Khatib et al., 2017)  
Minority classes may be problematic, though

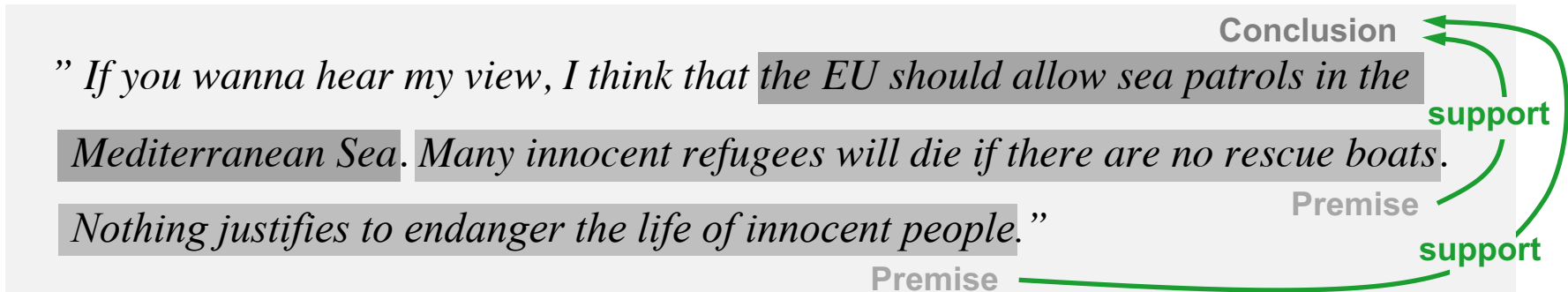
# Relation identification and classification

- **Argumentative relations**

- Within arguments (premises to conclusion) or between them (arg to arg)
- **Types.** Support or attack, partly more fine-grained

- **Relation identification and classification**

- **Task.** Given two units/arguments, what relation do they have, if any?
- **Method.** Various, e.g., with minimum spanning trees (Peldszus and Stede, 2015)



- **State of the art**

- Semi-reliable on narrow genres, such as essays ( $F_1$  0.73) (Stab, 2017)  
Identification works better than classification
- Relations hard to agree on for "hidden" arguments, such as in editorials

# 3. Argument mining for writing support

(Wachsmuth et al., 2016)

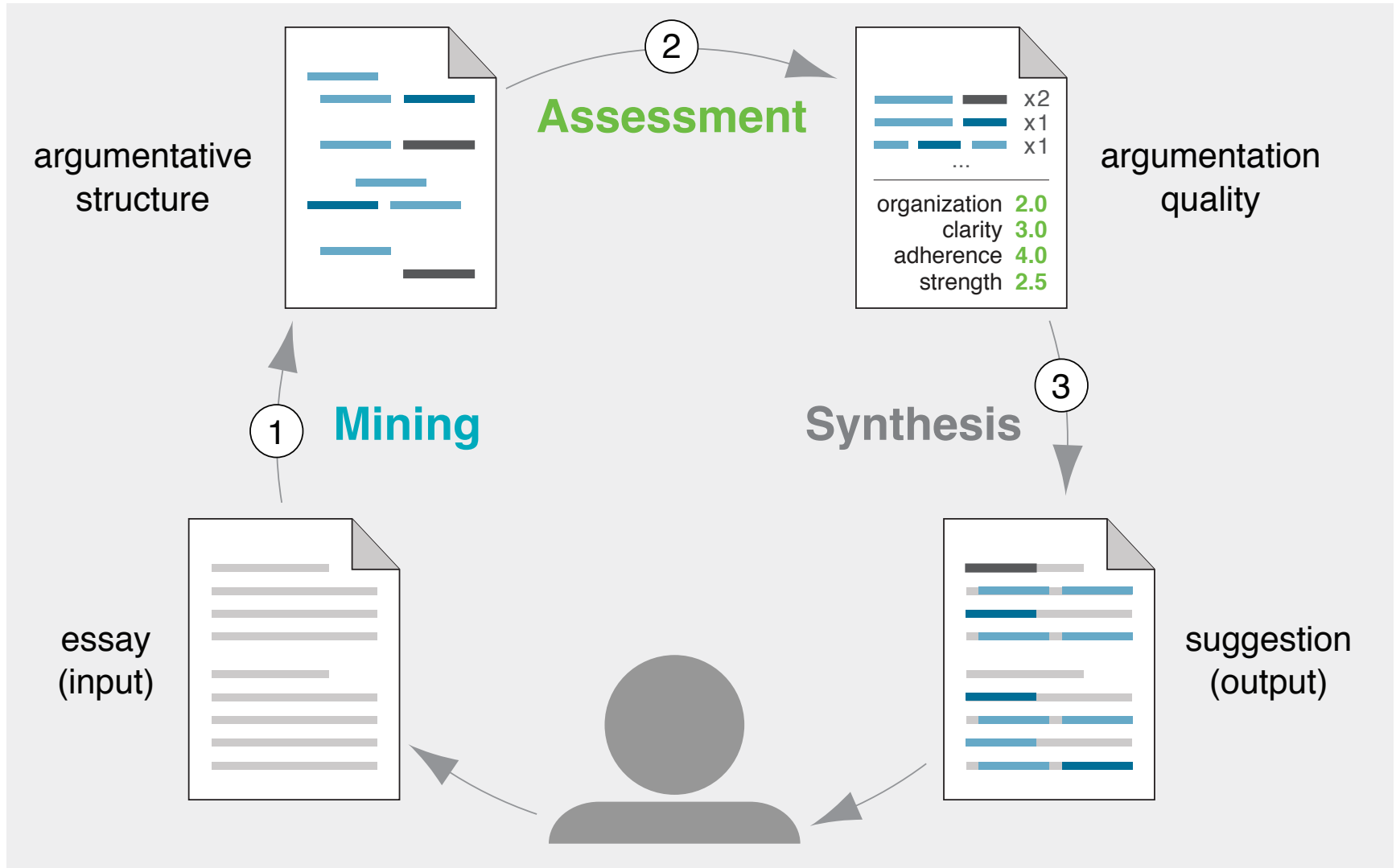


Khalid  
Al-Khatib



Benno  
Stein

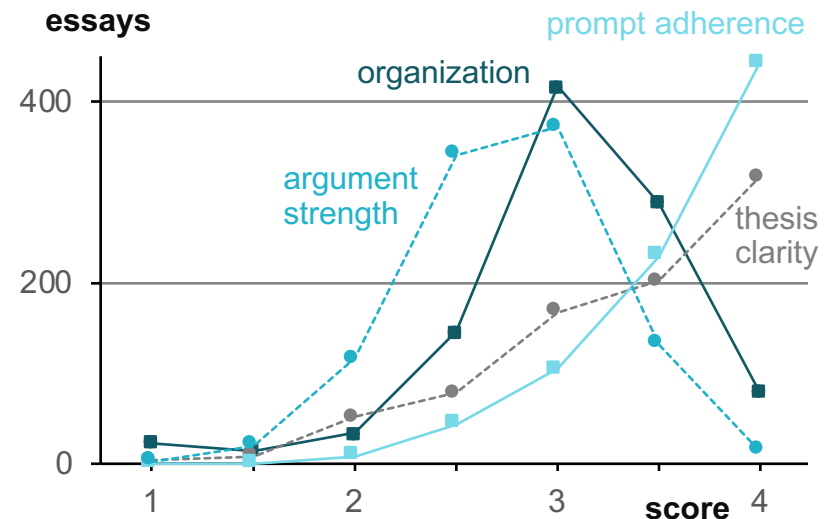
# Argument mining for writing support: Introduction





# Essay scoring based on argument mining

- **Research question**
  - Does argument mining help to score the quality of persuasive essays?
- **Quality dimensions** (Persing et al., 2010; Persing and Ng, 2013–2015)
  - **Organization**. How well is the essay arranged?
  - **Thesis clarity**. How easy to understand is the thesis?
  - **Prompt adherence**. How close does the essay stay to the issue?
  - **Argument strength**. How strong is the argument made for the thesis?
- **Data**
  - 800–1003 essays with scores in [1,4] for each dimension
- **Approach**
  1. Mine argument structure
  2. Analyze patterns in the structure
  3. Assess quality based on patterns



# Essay scoring approach: Mining and analysis

- **Mining**

- **Task.** Classify each sentence as thesis, conclusion, premise, or none
- **Data.** AAE corpus (Stab and Gurevych, 2014a)
- **Approach.** Support vector machine (SVM), several standard features

Approach	Accuracy	F <sub>1</sub>
Majority baseline	52.5	36.1
Stab and Gurevych (2014b)	<b>77.3</b>	72.6
<b>Our SVM classifier</b>	74.5	<b>74.5</b>

- **Analysis**

- **Task.** Compute most frequent unit role flows
- **Data.** All paragraphs of all 6085 essays in ICLE corpus (Granger et al., 2009)

Unit role flows	Average	First	Last
Conclusion, Premises	<b>25.1%</b>	–	13.1%
Conclusion, Premises, Conclusion	17.0%	–	<b>27.2%</b>
None, thesis	3.4%	<b>25.9%</b>	–
Premises, Conclusion	2.9%	–	2.7%

# Argument mining on example essay

- **Prompt**

"Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?"

**Organization** 3.0  
**Thesis clarity** 2.0  
**Prompt adherence** 4.0  
**Argument strength** 2.0

- **Essay**

None

"If we take a look back in time we are in a position to see man dreaming, philosophizing and using his imagination of whatever comes his way. We see man transcending his ego I a way and thus becoming a God - like figure. And by putting down these sacred words, what is taking shape in my mind is the fact that using his imagination Man is no longer this organic and material substance like his contemporary counterpart who is putting his trump card on science, technology and industrialization but Man is a way transcends himself through his imagination.

**Conclusion**

**Introduction**

For instance, if we take into account the Renaissance or Romantic periods of mankind and close our eyes we could see Shakespeare applying his imagination in the fancy world of his comedies: elf and nymphs circling the stage making it a dream that will lost forever in our minds. We could even hear their high-pitched weird chuckle piercing with a gentle touch our ears, but "open those eyes that must eclipse the day" and you'll see the high-tech wiping out every trace of the human elevated spirit that have dominated over the previous centuries. What we see now is "deux aux machina" or the fake "God from the machine" who with the touch of a button could unleash Armageddon.

**Premise**

**Body**

For poets and literate people of yore it was a common idea to transcend reality or to go beyond it by using their imagination not by using reason as we the homosapiens of our time do. For example, if we indulge in entertaining the idea of the film "The matrix" it has a lot to do with the period of Romanticism. But the difference is that a poet from that time could transcend reality, become one with Nature, and cruise wherever he wants using his imagination. Whereas now in the 21st century and in "The matrix" in particular the scientific type of Man thinks that at last he has succeeded in making travelling without boundaries via the virtual reality of his PC.

**Body**

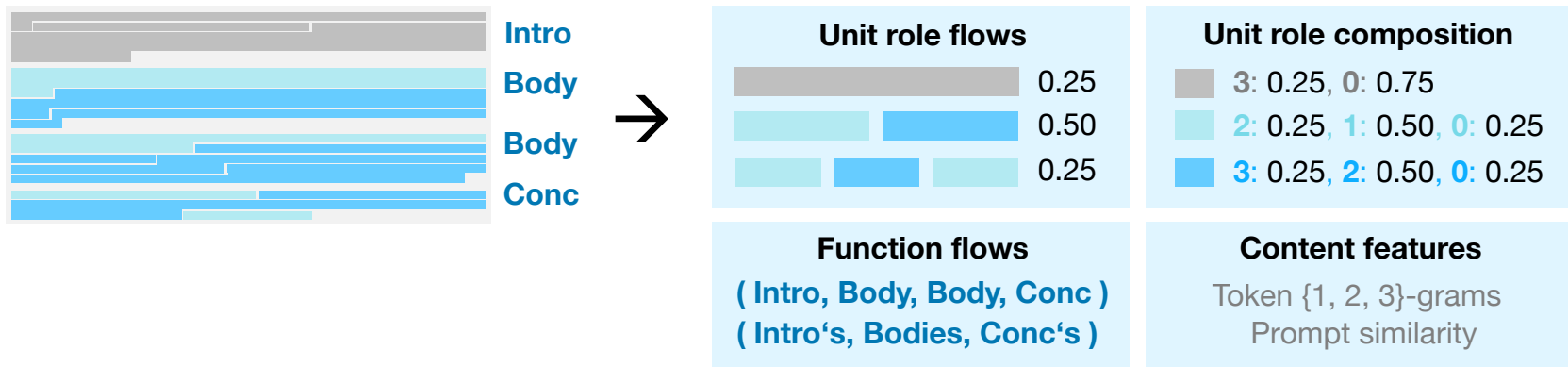
As a logical conclusion to my essay I would like to put only one thing. 'Wouldn't it be better if imagination makes the world go round'. If I was to answer this question, the answer would be positive, but given the aquisitive or consumer society conditions we live in let's make a match between imagination and science. It would be somewhat more realistic."

**Conclusion**

# Essay scoring approach: Assessment

- **Assessment**

- **Approach.** SVM regression, argument-specific and standard features



- **Evaluation.** Mean squared error for each quality dimension

Approach	Organization	Clarity	Adherence	Strength
Average baseline	0.349	0.469	0.291	0.266
Persing et al. (2010–2015)	0.175	<b>0.369</b>	<b>0.197</b>	0.244
<b>Our SVM regressor</b>	<b>0.164</b>	0.425	0.216	<b>0.226</b>
— Unit role flows	0.234	0.461	0.247	0.242
— Unit role composition	<b>0.194</b>	0.457	0.239	0.239
— Function flows	0.220	0.478	0.255	0.251
— Content features	0.336	<b>0.425</b>	<b>0.231</b>	<b>0.236</b>

## 4. Argumentation quality assessment

# Argumentation quality assessment: Introduction

- **Argumentation quality**

- Natural language arguments rarely logically valid
- Need to quantify how *strong* an argument or argumentation is.

*” Is a strong argument an effective argument which gains the adherence of the audience, or is it a valid argument, which ought to gain it? “*

(Perelman and Olbrechts-Tyteca, 1969)

- **Argumentation quality assessment**

- Absolute or relative judgment of specific quality dimensions
- Identification of flaws and fallacies

- **Critical for any application**

- **Argument search.** What argument to rank highest?
- **Writing support.** What argumentative flaws does a text have?
- **Decision making.** Which arguments outweigh others?



# What to assess and how to assess it

- **What to assess**

- Several, partly very subjective quality dimensions
- Different granularity levels

cogent?

effective?

reasonable?

Conclusion

” If you wanna hear my view, I think that *the EU should allow sea patrols in the Mediterranean Sea.* *Many innocent refugees will die if there are no rescue boats.*

Premise

*Nothing justifies to endanger the life of innocent people.”*

acceptability: 4 / 5

Premise

acceptable?

clear?

relevant?

more acceptable than

”It’s the main job of the EU to save people’s lives, no matter whether they belong here.“

- **How to assess**

- Absolute or relative?
- How *should* we vs. how *do* we argue?
- Based on manual assessments or ”objective“ properties?
- Include model of audience?

# Three aspects of argumentation quality

$$\frac{A \quad A \rightarrow B}{B}$$

Logic

*"A dialectical discussion derives its reasonableness from a dual criterion: problem validity and intersubjective validity."*

van Eemeren (2015)

Dialectic

$$\frac{A \quad A \rightarrow B}{B} \quad \frac{B \rightarrow C}{C}$$



<https://de.wikipedia.org>

*"An argument is cogent if its premises are relevant to its conclusion, individually acceptable, and together sufficient to draw the conclusion."*

Blair (2012)

Argumentation quality

Rhetoric

*"In making a speech, one must study three points: the means of producing persuasion, the style or language to be used, and the proper arrangement of the various parts."*

Aristotle (2007)

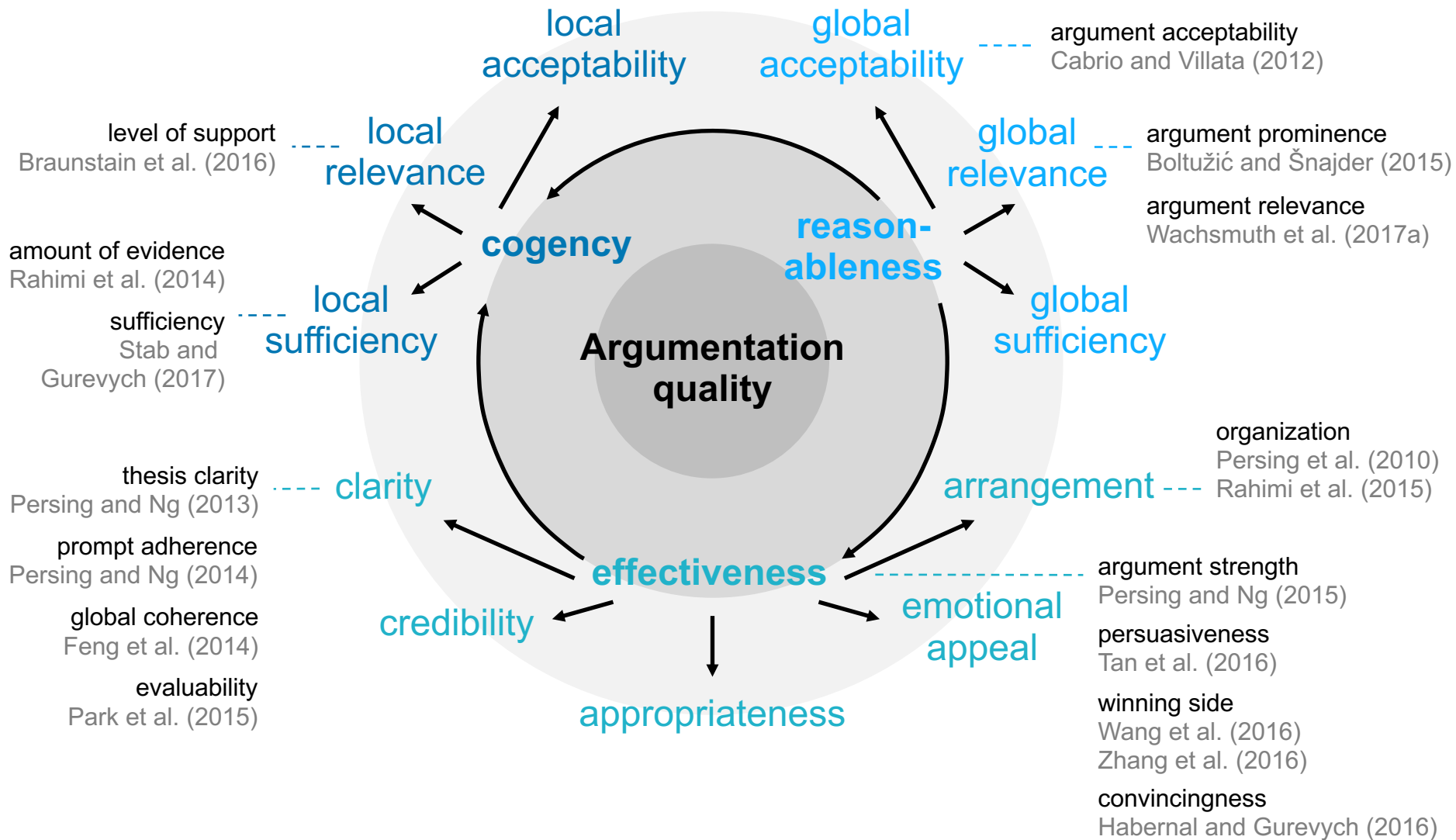
$$\frac{A \quad A \rightarrow B}{B}$$



<https://commons.wikimedia.org>



# A taxonomy of argumentation quality (Wachsmuth et al., 2017b)



# The role of participants in argumentation

## Author (or speaker)

- Argumentation is connected to the person who argues
- The same argument is perceived differently depending the author

*” The EU should allow rescue boats. Many innocent refugees will die if there are no rescue boats. “*



<https://pixabay.com>



<https://commons.wikimedia.org>

## Reader (or audience)

- Argumentation often targets a particular audience
- Different arguments and ways of arguing work for different persons

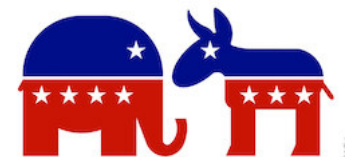
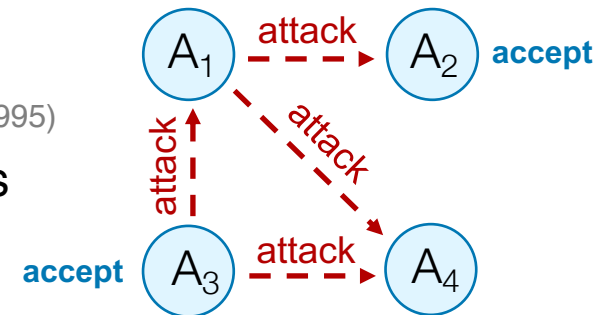
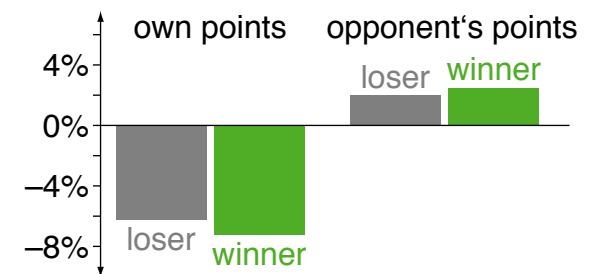
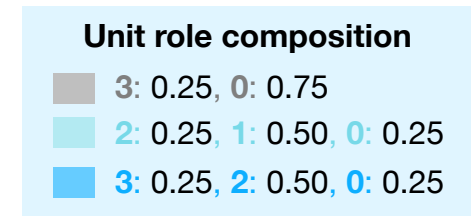
*” According to a recent UN study, the number of rescue boats had no effect on the number of refugees who try. “*



<https://pixabay.com>

# Selected quality assessment approaches

- **Absolute assessment** (Wachsmuth et al., 2016)
  - Regression of four essay quality dimensions
  - Features based on argument mining
- **Relative assessment** (Zhang et al., 2016)
  - Classification of the winner of a debate
  - Modeling own and attack of opponent's points
- **"Objective" assessment** (Cabrio and Villata, 2012)
  - Graph analysis to determine acceptability (Dung, 1995)
  - Textual entailment to obtain attacking arguments
- **Audience-specific assessment** (El Baff et al., 2020)
  - Analysis of editorial effectiveness for audience
  - Model of audience's ideology and personality



<https://flickr.com>

# 5. Fallacy detection in online discussions

(Habernal et al., 2018)



Ivan  
Habernal



Iryna  
Gurevych



Benno  
Stein



# Fallacy detection: Introduction

- **What is a fallacy?** (Tindale, 2007)
  - An argument with some (often hidden) flaw in its reasoning, i.e., it has a failed or deceptive scheme
- **Example types of fallacies**
  - **Ad-hominem.** Attacking the opponent instead of his or her arguments
  - **Red herring.** Reasoning based on an unrelated issue
  - **Appeal to ignorance.** Taking lack of evidence as proof for the opposite

*My girlfriend **won't** give me a gift for my birthday. I have received no indication to the contrary from her.*

*My flight tomorrow **won't** be delayed. I have received no indication to the contrary from the airline.*

(credit to Mario Treiber for this example)

- **Fallacies are hard to detect**
  - Structure identical to other arguments
  - Understanding and context knowledge needed

# A study of ad-hominem arguments on the web

- **Ad-hominem arguments**

- Attacking the opponent instead of his or her arguments
- 20% of online news comments uncivil (Coe et al., 2014)

That's an ad hominem fallacy Calvin!!



"YOU'RE FACE IS AN AD HOMINEM!!!"



- **Research questions**

- Can we identify ad-hominem automatically?
- What are triggers of ad-hominem?

- **Data**

- 2M posts from Reddit ChangeMyView
  - 3866 posts (0.2%) contain an ad-hominem argument
- Ad-hominem is deleted by moderators, but we obtained all comments from them



- **Reddit ChangeMyView (CMV)**

- An opinion poster (OP) states a view
- Others argue for the opposite
- OP gives  $\Delta$  to convincing posts

Deltas(s) from OP **CMV: Trump has done nothing of substance since being elected to office.**

This is kind of a counter to the other post made recently about Trump being a great president.

He pointed out things like the economy which was growing

# Ad-hominem on CMV

"Reading comprehension is your friend"

"Ever have discussions with narcissistic idiots on the internet? They are so tiring"

"You still refuse to acknowledge that you used a strawman argument against me"

"little buddy"

"Thank you so much for all your pretentious explanations"

"To say that people intrinsically understand portion size is idiotic."

"You started with a fallacy and then deflected."

"Please dont waste peoples time pretending to know what you're talking about"

"boy"

"Did you even read this?"

"Read what I posted before acting like a pompous ass"

"Do you even know what you're saying?"

"Again, how old are you?"

"You're making the claims, it's your job to prove it. Don't you know how debating works?"

"You're obviously just Nobody with enough brains to operate a computer could possibly believe something this stupid"

"Your second paragraph is fairly idiotic"

"You have no capability to understand why"

"Wow. Someone sounds like a bit of an anti-semite"

"you dumb fuck"

"Your just an asshole"

"Possible lie any harder?"

"How can you explain that? You can't because it will hurt your feelings to face reality"

"You're just a dishonest troll"

"You are just a liar."

"Can you also use Google?"

"Willful ignorance is not something I can combat"

"You're trash at debating."

"You're too dishonest to actually quote the verse because you know it's bullshit"

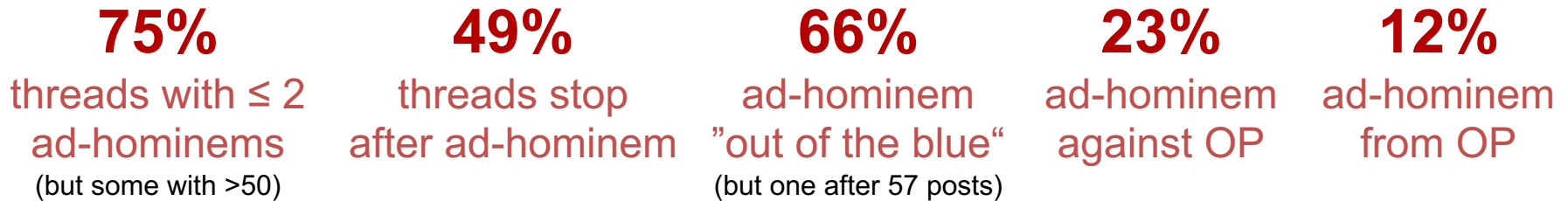
"You're using troll tactics"

"sir"

"If you can't grasp the concept, I can't help you"

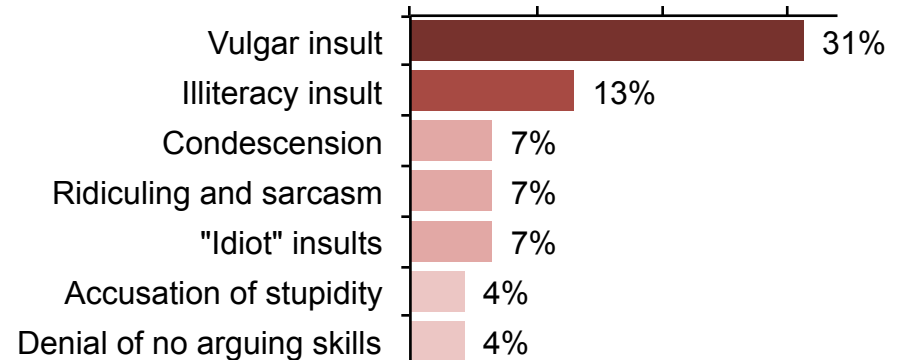
# Identification of ad-hominem

- **Distribution of ad-hominem**



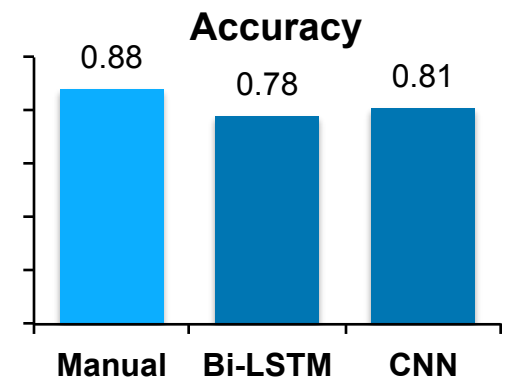
- **Types of ad-hominem**

- Ad-hominem annotated in 400 arguments by 7 crowdworkers
- 15 types derived manually from their annotations



- **Identification of ad-hominem**

- **Manual.** 100 balanced arguments (50 ad-hominem) classified by 6 workers
- **Automatic.** 7242 balanced arguments classified by 2 neural classifiers (Bi-LSTM & CNN)

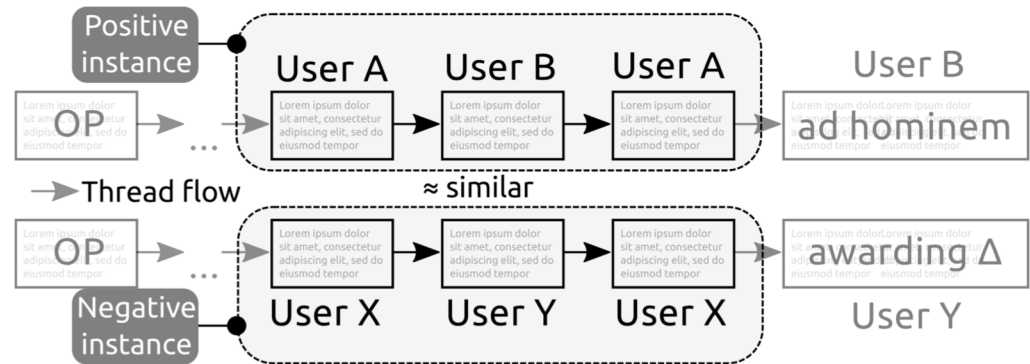




# Analysis of triggers of ad-hominem

- **Prediction of ad-hominem**

- Attentive LSTM trained on 2852 argument 3-tuples
- Accuracy 0.72
- Manual attention analysis



(OOV\_comment\_begin) If only you would n't rely on [ fallacious ] ( http : OOV ) [ arguments ] ( http : OOV ) to make your point. So no , I do n't realize how stupid and naive I am. All I 've realized is that you are n't actually prepared to have an actual discussion .

(OOV\_comment\_begin) What god do you believe in ? And it 's not a fallacy when it 's very comparable to the most popular gods .

(OOV means out-of-vocabulary)

- **Terms with much attention**

- Mostly topic-independent rhetorical devices
- A few loaded keywords, such as "racist"
- Partly meta about argumentation

**vulgar intensifiers**  
"... the fuck..."

**direct imperatives**  
"You should..."

**missing evidence**  
"unsupported claims!"

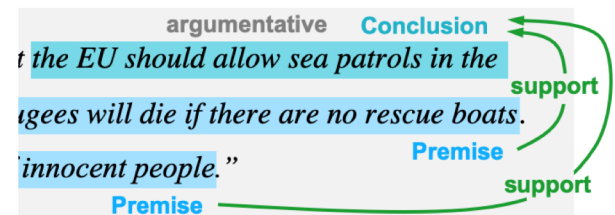
**bad argumentation**  
"You're grasping at straws" ...

# Conclusion

# Conclusion

- **Argument mining and assessment**

- Finding arguments in natural language text
- Classifying stance and other properties
- Assessing quality dimensions and flaws



- **State of the art**

- Most tasks now tackled with neural approaches
- In narrow domains, reasonable effectiveness achievable
- Robust ”off-the-shelf“ algorithms rare so far



- **Role within argumentation technology**

- Builds on argumentation theory and computational linguistics
- Needed to process natural language arguments
- Converts arguments to (semi-) structured information



# References

- **Ajjour et al. (2017)**. Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. Unit Segmentation of Argumentative Texts. In Proceedings of the Fourth Workshop on Argument Mining, pages 118–128, 2017.
- **Al-Khatib et al. (2017)**. Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. Patterns of Argumentation Strategies across Topics. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1362–1368, 2017.
- **Aristotle (2007)**. Aristotle (George A. Kennedy, Translator). On Rhetoric: A Theory of Civic Discourse. Clarendon Aristotle series. Oxford University Press, 2007.
- **Bar-Haim et al. (2017)**. Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. Stance Classification of Context-Dependent Claims. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, pages 251–261, 2017.
- **Blair (2012)**. J. Anthony Blair. Groundwork in the Theory of Argumentation. Springer Netherlands, 2012.
- **Boltužic and Snajder (2015)**. Filip Boltužic and Jan Snajder. Identifying Prominent Arguments in Online Debates using Semantic Textual Similarity. In Proceedings of the 2nd Workshop on Argumentation Mining, pages 110–115, 2015.
- **Braunstain et al. (2016)**. Liora Braunstain, Oren Kurland, David Carmel, Idan Szpektor, and Anna Shtok. Supporting Human Answers for Advice-seeking Questions in CQA Sites. In Proceedings of the 38th European Conference on IR Research, pages 129–141, 2016.
- **Cabrio and Villata (2012)**. Elena Cabrio and Serena Villata. Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 208–212, 2012.

# References

- **Collins and Duffy (2001)**. Michael Collins and Nigel Duffy. 2001. Convolution kernels for natural language. In *Advances in Neural Information Processing Systems 14*, pages 625–632.
- **Dung (1995)**: Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–357, 1995.
- **El Baff et al. (2020)**. Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. Analyzing the Persuasive Effect of Style in News Editorial Argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3154–3160, 2020.
- **Feng and Hirst (2011)**. Vanessa Wei Feng and Graeme Hirst. Classifying Arguments by Scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 987–996, 2011.
- **Feng et al. (2014)**. Vanessa Wei Feng, Ziheng Lin, and Graeme Hirst. The Impact of Deep Hierarchical Discourse Structures in the Evaluation of Text Coherence. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 940–949. Dublin City University and Association for Computational Linguistics, 2014.
- **Freeman (2011)**. *Argument Structure: Representation and Theory*. Springer, 2011.
- **Granger et al. (2009)**. Sylviane Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. *International Corpus of Learner English (version 2)*, 2009.
- **Habernal and Gurevych (2015)**. Exploiting Debate Portals for Semi-supervised Argumentation Mining in User-generated Web Discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127– 2137, 2015.

# References

- **Habernal and Gurevych (2016a)**. Ivan Habernal and Iryna Gurevych. 2016. Which Argument is More Convincing? Analyzing and Predicting Convincingness of Web Arguments using Bidirectional LSTM. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1589–1599.
- **Habernal et al. (2018a)**. Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. Before Name-calling: Dynamics and Triggers of Ad Hominem Fallacies in Web Argumentation. In Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 386–396, 2018.
- **Hasan and Ng (2013)**. Kazi Saidul Hasan and Vincent Ng. Stance Classification of Ideological Debates: Data, Models, Features, and Constraints. In Proceedings of the Sixth International Joint Conference on Natural Language Processing, pages 1348--1356, 2013.
- **Mooney and Bunescu (2006)**. Raymond J. Mooney and Razvan C. Bunescu. 2006. Subsequence kernels for relation extraction. In Advances in Neural Information Processing Systems 18, pages 171–178. MIT Press.
- **Park et al. (2015)**. Joonsuk Park, Cheryl Blake, and Claire Cardie. Toward Machine-assisted Participation in eRulemaking: An Argumentation Model of Evaluability. In Proceedings of the 15th International Conference on Artificial Intelligence and Law, pages 206–210, 2015.
- **Peldszus and Stede (2015)**. Andreas Peldszus and Manfred Stede. Joint Prediction in MST-style Discourse Parsing for Argumentation Mining. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 938–948, 2015.
- **Peldszus and Stede (2016)**. Andreas Peldszus and Manfred Stede. 2016. An annotated corpus of argumentative microtexts. In Argumentation and Reasoned Action: 1st European Conference on Argumentation.
- **Perelman and Olbrecht-Tyteca (1969)**. Chaïm Perelman and Lucie Olbrechts-Tyteca. 1969. The New Rhetoric: A Treatise on Argumentation (John Wilkinson and Purcell Weaver, translator). University of Notre Dame Press.

# References

- **Persing and Ng (2013):** Isaac Persing and Vincent Ng. Modeling Thesis Clarity in Student Essays. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 260–269, 2013.
- **Persing and Ng (2014):** Isaac Persing and Vincent Ng. Modeling Prompt Adherence in Student Essays. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 1534–1543, 2014.
- **Persing and Ng (2015):** Isaac Persing and V. Ng. Modeling Argument Strength in Student Essays. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 543–552, 2015.
- **Persing et al. (2010).** Isaac Persing, Alan Davis, and Vincent Ng. Modeling organization in student essays. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 229–239, 2010.
- **Rahimi et al. (2014).** Zahra Rahimi, Diane J. Litman, Richard Correnti, Lindsay Clare Matsumura, Elaine Wang, and Zahid Kisa. Automatic Scoring of an Analytical Response-to-Text Assessment. In Proceedings of the 12th International Conference on Intelligent Tutoring Systems, pages 601–610, 2014.
- **Rahimi et al. (2015).** Zahra Rahimi, Diane Litman, Elaine Wang, and Richard Correnti. Incorporating Coherence of Topics as a Criterion in Automatic Response-to-Text Assessment of the Organization of Writing. In Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 20–30, 2015.
- **Ranade et al. (2013):** Sarvesh Ranade, Rajeev Sangal, and Radhika Mamidi. Stance Classification in Online Debates by Recognizing Users’ Intentions. In: Proc. of the SIGDIAL 2013, 61–69, 2013.
- **Rinott et al. (2015).** Ruty Rinott, Lena Dankin, Carlos Alzate Perez, M. Mitesh Khapra, Ehud Aharoni, and Noam Slonim. Show Me Your Evidence — An Automatic Method for Context Dependent Evidence Detection. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 440–450, 2015.

# References

- **Somasundaran and Wiebe (2010)**: Swapna Somasundaran and Janyce Wiebe. Recognizing Stances in Ideological On-Line Debates. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 116–124, 2010.
- **Stab and Gurevych (2014a)**. Christian Stab and Iryna Gurevych. Annotating Argument Components and Relations in Persuasive Essays. In Proceedings of the 25th Conference on Computational Linguistics, pages 1501–1510, 2014.
- **Stab and Gurevych (2014b)**. Christian Stab and Iryna Gurevych. Identifying Argumentative Discourse Structures in Persuasive Essays. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pages 46–56, 2014.
- **Stab and Gurevych (2016)**. Christian Stab and Iryna Gurevych. 2016. Recognizing the absence of opposing arguments in persuasive essays. In Proceedings of the Third Workshop on Argument Mining (ArgMining2016), pages 113–118.
- **Stab (2017)**. Christian Stab. Argumentative Writing Support by means of Natural Language Processing, Chapter 5. PhD thesis, TU Darmstadt, 2017. <http://tuprints.ulb.tu-darmstadt.de/6006/1/PhD-Thesis-ChristianStab.pdf>
- **Stab and Gurevych (2017)**. Christian Stab and Iryna Gurevych. Recognizing Insufficiently Supported Arguments in Argumentative Essays. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pages 980–990, 2017.
- **Tan et al. (2016)**. Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In Proceedings of the 25th International Conference on World Wide Web, pages 613–624, 2016.
- **van Eemeren (2015)**. Frans H. van Eemeren. Reasonableness and Effectiveness in Argumentative Discourse: Fifty Contributions to the Development of Pragma-Dialectics. Argumentation Library. Springer International Publishing, 2015.



# References

- **Wachsmuth et al. (2016).** Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. Using Argument Mining to Assess the Argumentation Quality of Essays. In: Proceedings of the 26th International Conference on Computational Linguistics, pages 1680–1692, 2016.
- **Wachsmuth et al. (2017a).** Henning Wachsmuth, Benno Stein, and Yamen Ajjour. "PageRank" for Argument Relevance. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pages 1116–1126, 2017.
- **Wachsmuth et al. (2017b).** Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. Computational Argumentation Quality Assessment in Natural Language. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pages 176–187, 2017.
- **Walton et al. (2008).** Douglas Walton, Christopher Reed, and Fabrizio Macagno. Argumentation Schemes. Cambridge University Press, 2008.
- **Wang et al. (2017).** Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin. Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes. In: Transactions of the Association for Computational Linguistics 5, pages 219--232, 2017.
- **Zhang et al. (2016).** Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. Conversational Flow in Oxford-style Debates. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 136–141, 2016.