

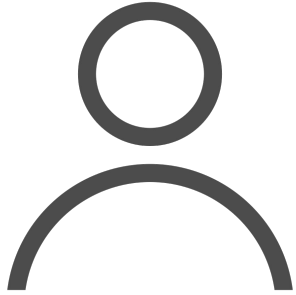
Theory-based Argument Quality for Advanced Argument Retrieval

Opportunities and Challenges

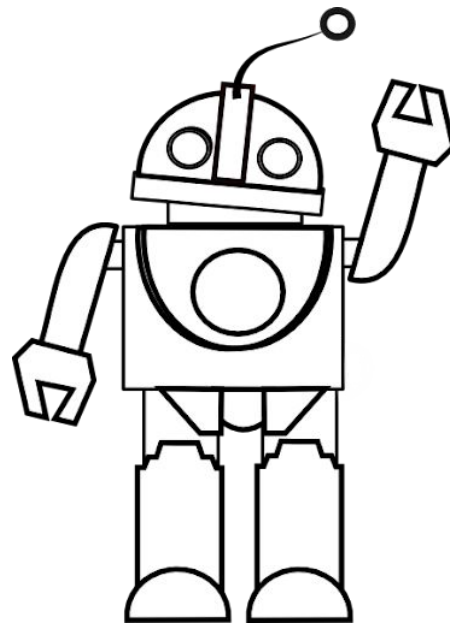
Anne Lauscher @Touché 2021



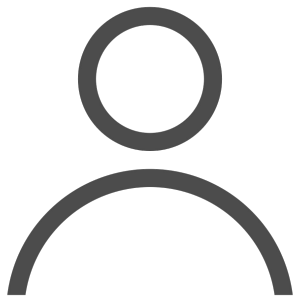
Argument Retrieval



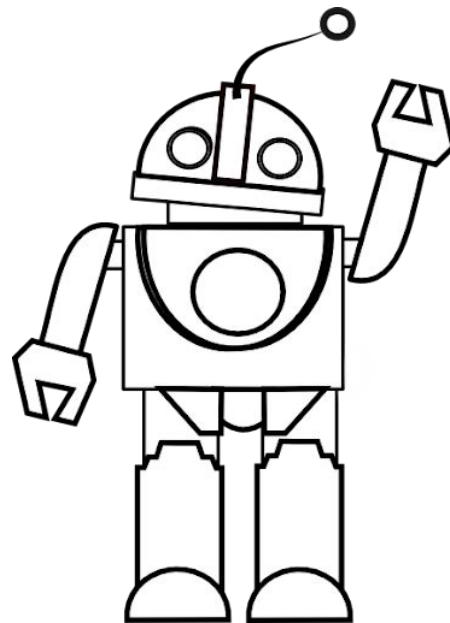
Argument Retrieval



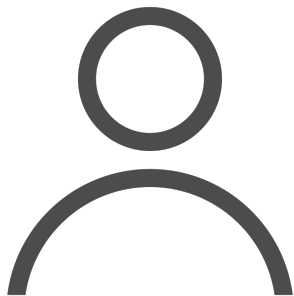
Argument Retrieval



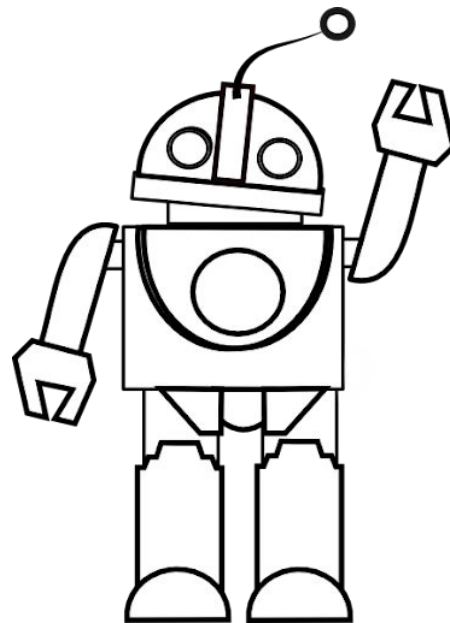
Why should we allow
gay marriage?
Show me arguments!



Argument Retrieval



Why should we allow
gay marriage?
Show me *good* arguments!

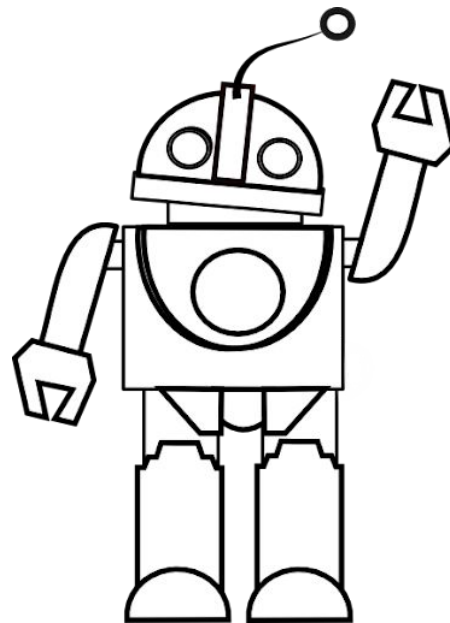


Argument Retrieval

Argument retrieval should be guided by the **quality** of arguments!



Why should we allow
gay marriage?
Show me **good** arguments!



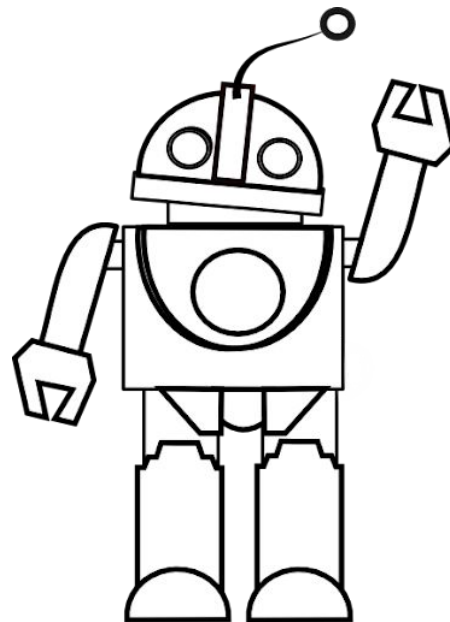
Touché — Evaluation

“Our human assessors will label the retrieved documents manually, both for their general topical relevance and for the rhetorical quality, i.e., “well-writtenness” of the document:

(1) whether a document contains arguments (...) and whether the text has a good style of speech (...),

(2) whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,

(3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”



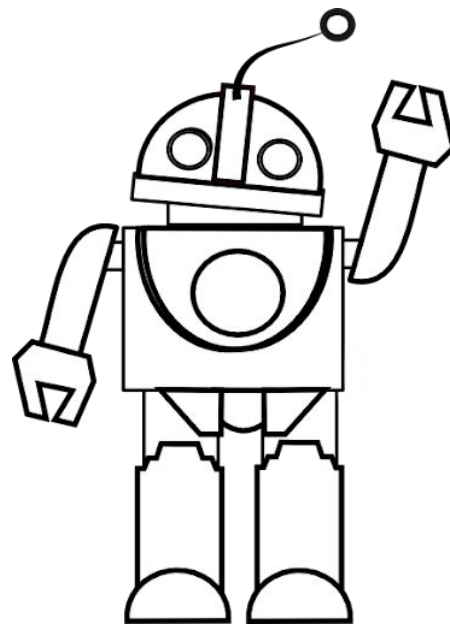
Touché — Evaluation

“Our human assessors will label the retrieved documents manually, both for their general **topical relevance** and for the **rhetorical quality**, i.e., “well-writtenness” of the document:

(1) whether a document contains arguments (...) and whether the text has a good style of speech (...),

(2) whether the text has a proper sentence structure and is easy to read and follow, whether it can be well understood,

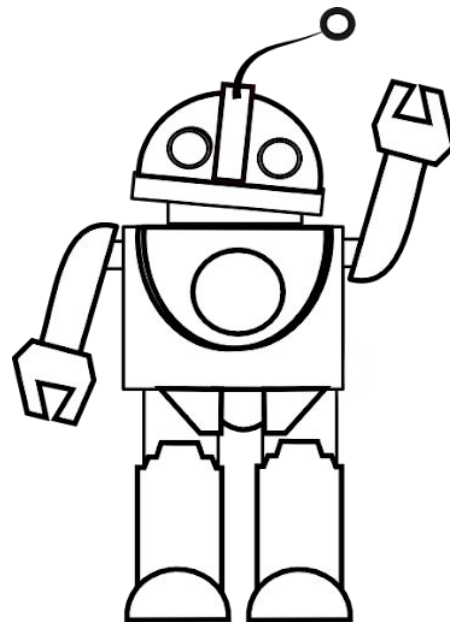
(3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”



Touché — Evaluation

“Our human assessors will label the retrieved documents manually, both for their general **topical relevance** and for the **rhetorical quality**, i.e., “well-writtenness” of the document:

- (1) **whether a document contains arguments** (...) and whether the text has a **good style** of speech (...),
- (2) whether the text has a **proper sentence structure** and is **easy to read** and follow, whether it can be well understood,
- (3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”



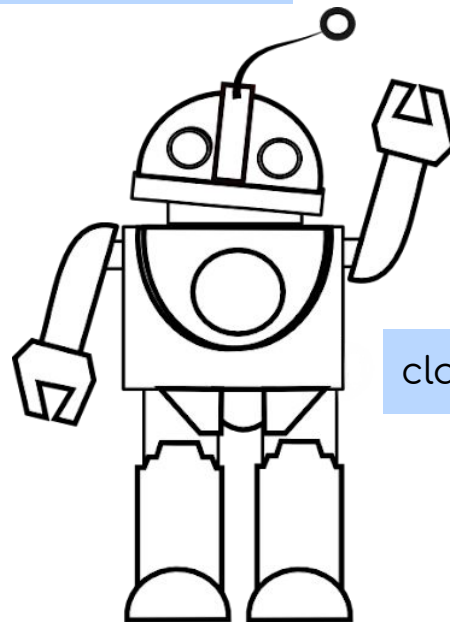
Touché — Evaluation

“Our human assessors will label the retrieved documents manually, both for their general **topical relevance** and for the **rhetorical quality**, i.e., “well-writtenness” of the document:

- (1) **whether a document contains arguments** (...) and whether the text has a **good style** of speech (...),
- (2) whether the text has a **proper sentence structure** and is **easy to read** and follow, whether it can be well understood,
- (3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”

appropriateness?

argumentativeness?



arrangement?

clarity?

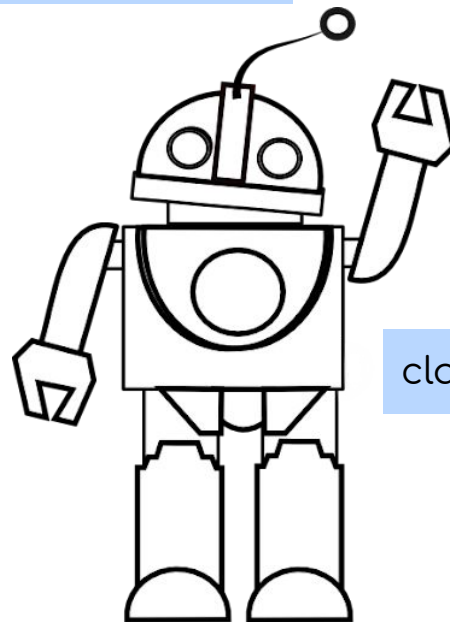
Touché — Evaluation

“Our human assessors will label the retrieved documents manually, both for their general **topical relevance** and for the **rhetorical quality**, i.e., “well-writtenness” of the document:

- (1) **whether a document contains arguments** (...) and whether the text has a **good style** of speech (...),
- (2) whether the text has a **proper sentence structure** and is **easy to read** and follow, whether it can be well understood,
- (3) whether it includes profanity, has typos, and makes use of other detrimental style choices.”

appropriateness?

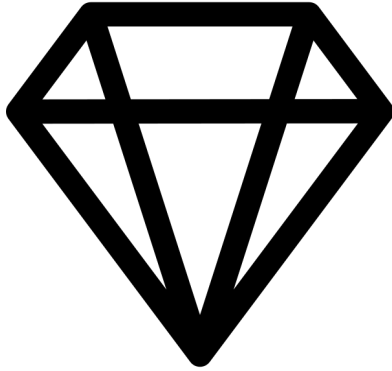
argumentativeness?



arrangement?

clarity?

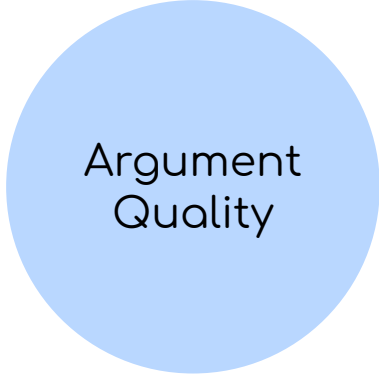
Theory-based Argument Quality



What is this? And what can we gain?

Theory-based Argument Quality

(Wachsmuth et al., 2017a)



Argument
Quality

A large, light blue circle is centered on the right side of the slide. Inside the circle, the words "Argument" and "Quality" are stacked vertically in a black, sans-serif font.

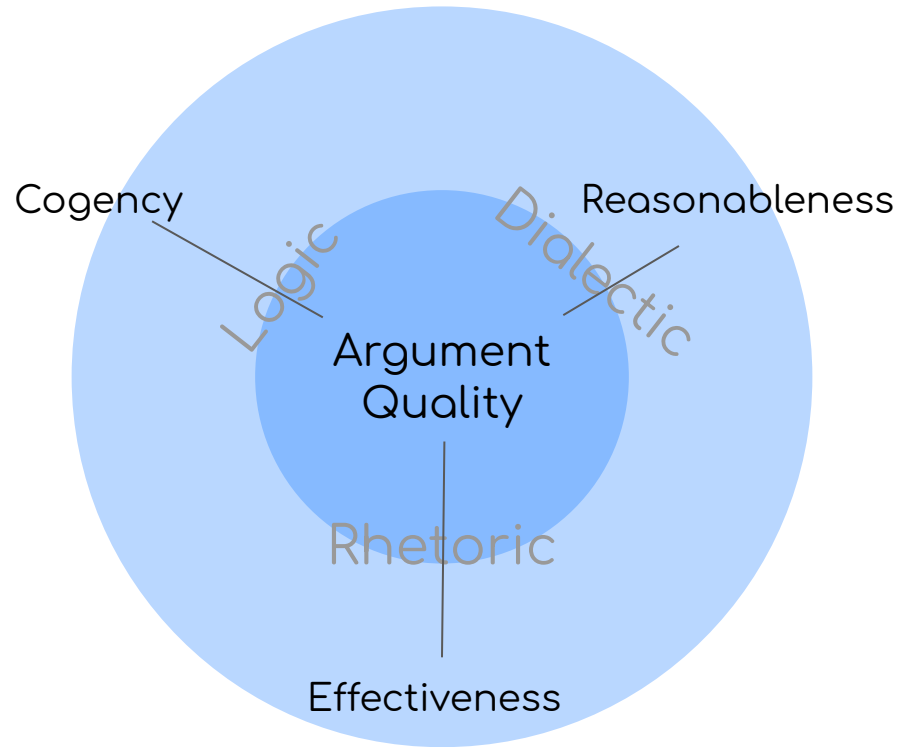
Theory-based Argument Quality

(Wachsmuth et al., 2017a)



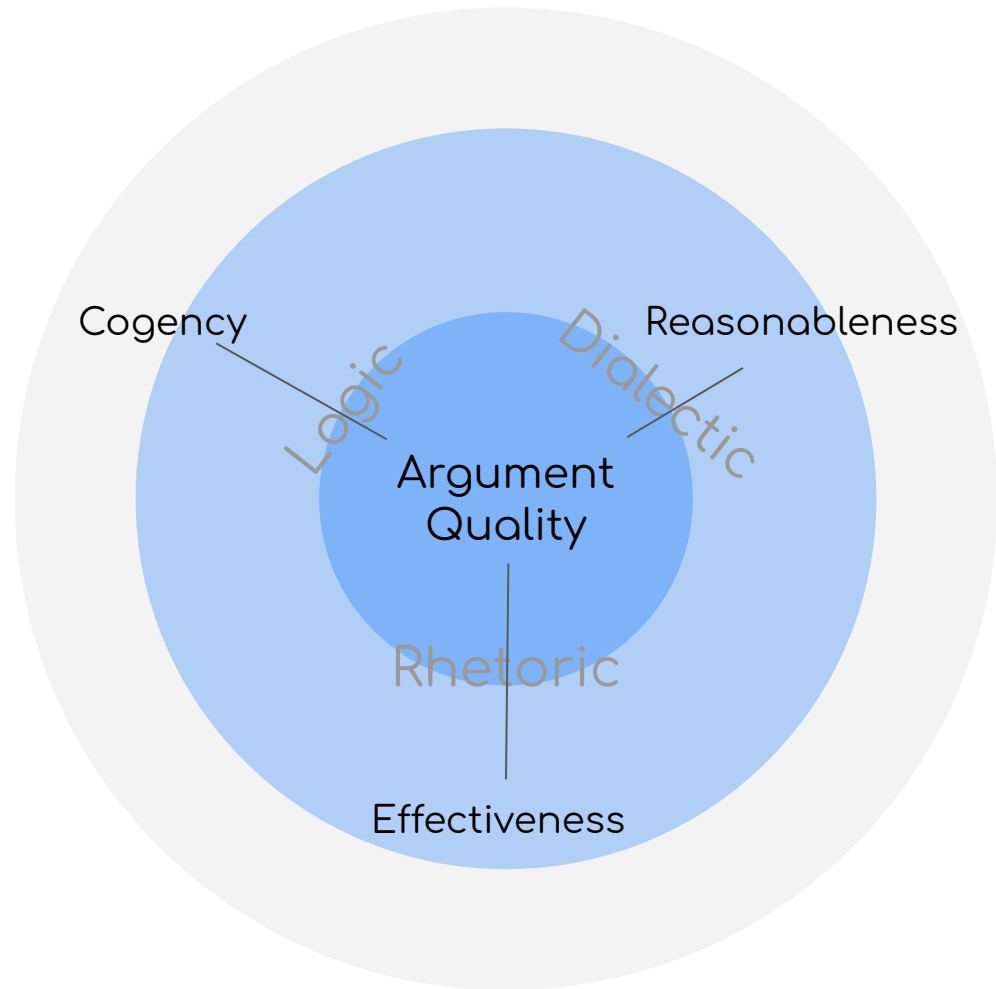
Theory-based Argument Quality

(Wachsmuth et al., 2017a)



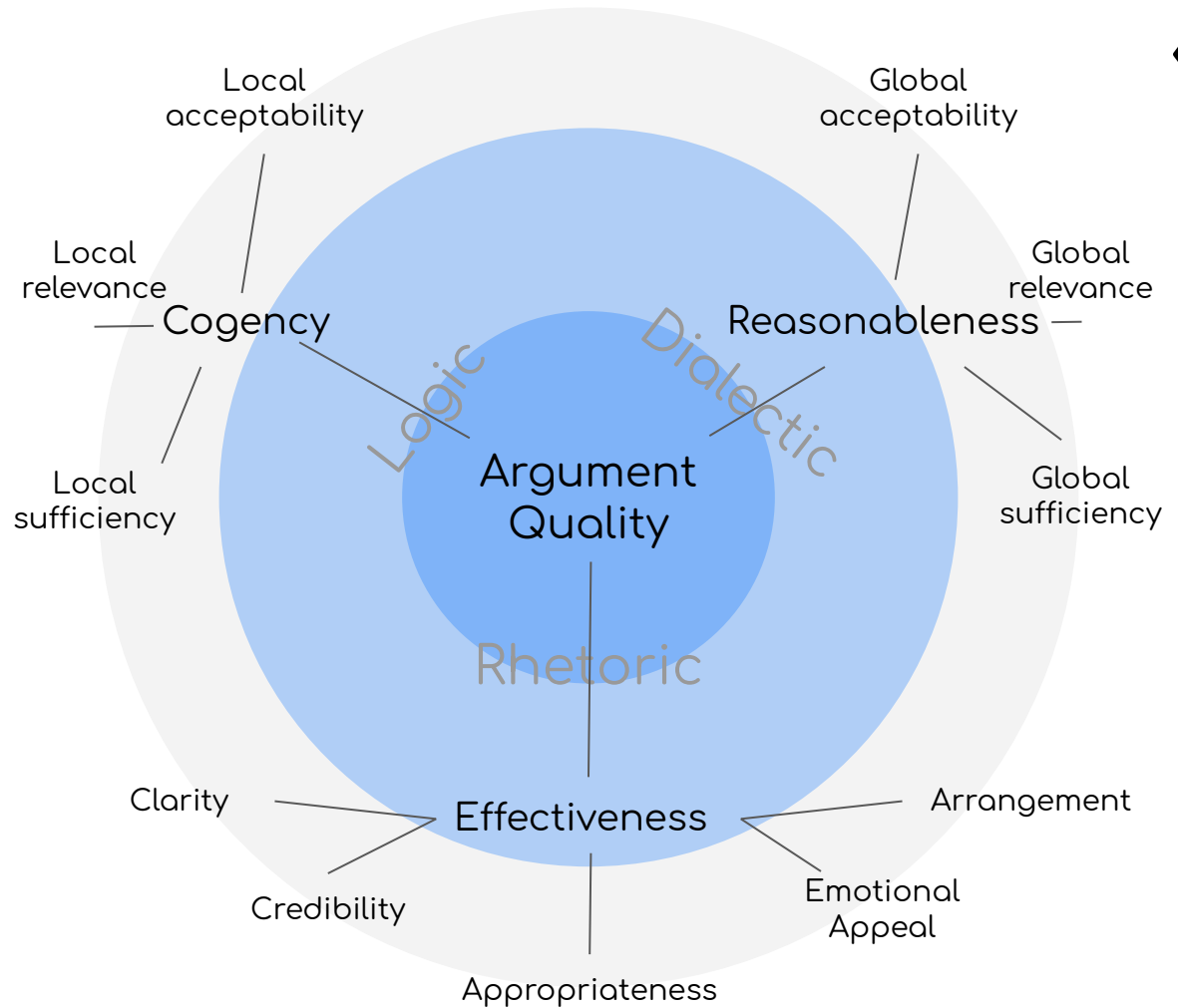
Theory-based Argument Quality

(Wachsmuth et al., 2017a)



Theory-based Argument Quality

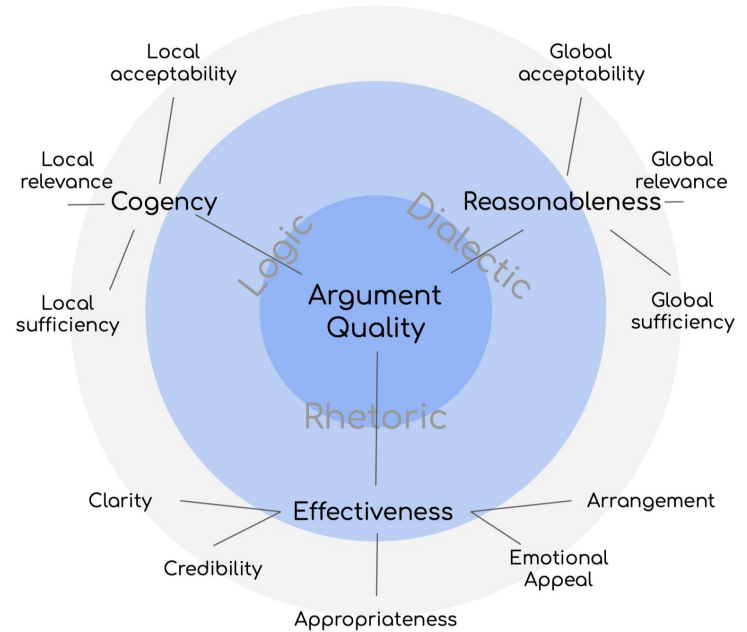
(Wachsmuth et al., 2017a)





Theory-based Argument Quality

Should we allow gay marriage?

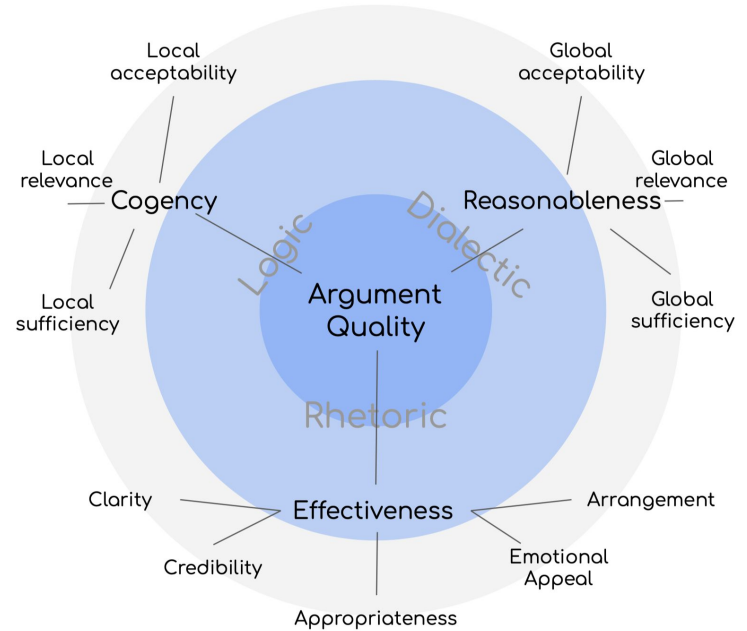




Theory-based Argument Quality

Should we allow gay marriage?

No, *\$%§"\$%&&!!!!!

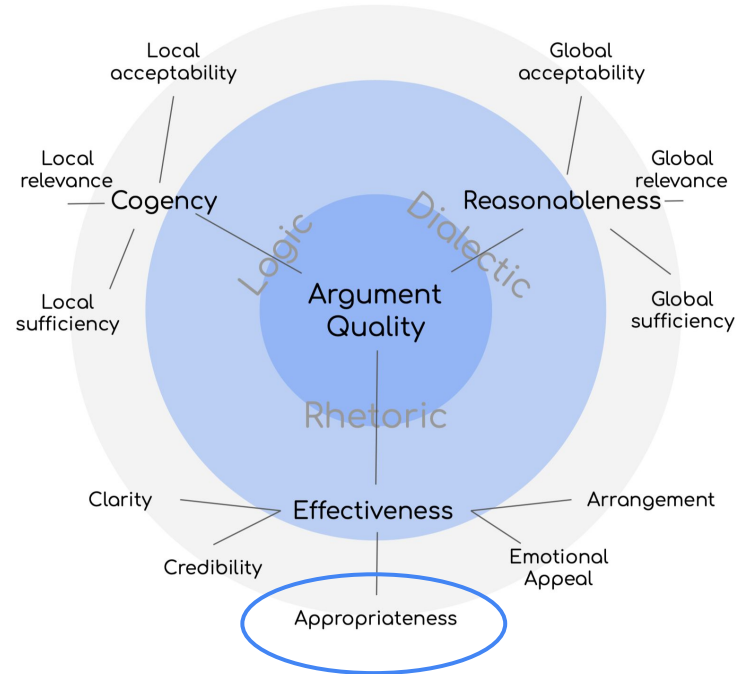




Theory-based Argument Quality

Should we allow gay marriage?

No, *\$%§"\$%&&!!!!!





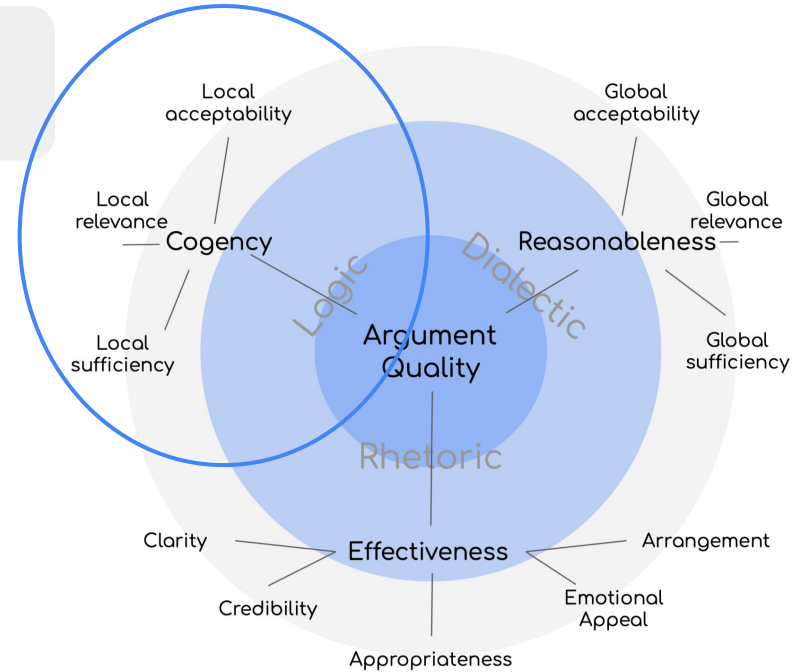
Theory-based Argument Quality

Should we allow gay marriage?

Yes, <claim>, because

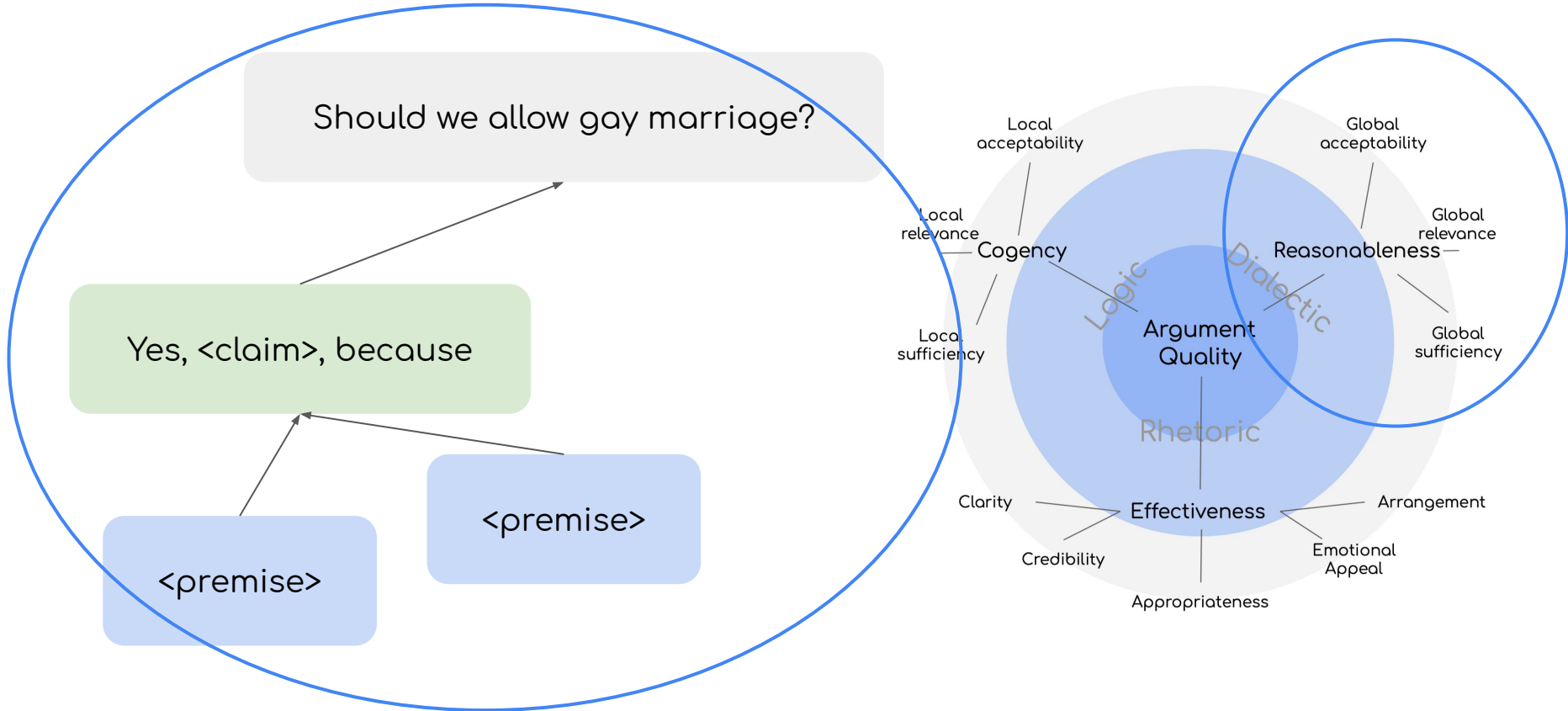
<premise>

<premise>

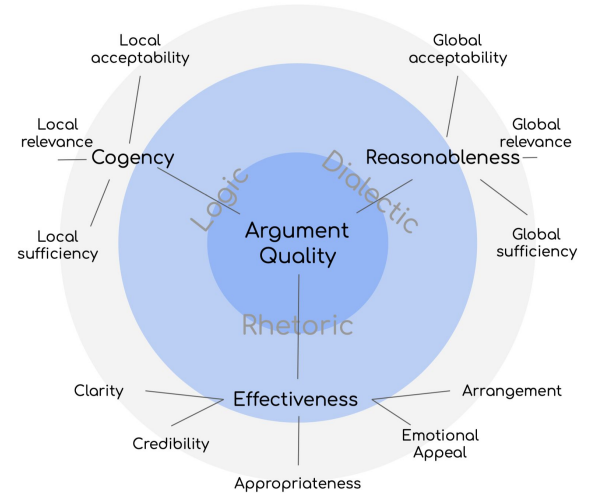




Theory-based Argument Quality



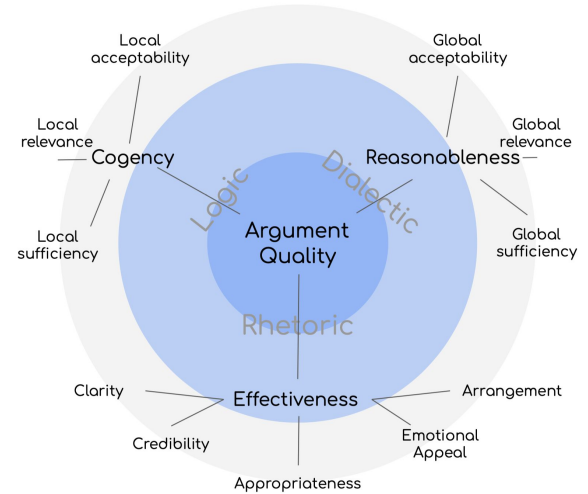
Opportunities for Argument Retrieval



Opportunities for Argument Retrieval



Holistic assessment of “good” arguments based on a theoretically-grounded notion



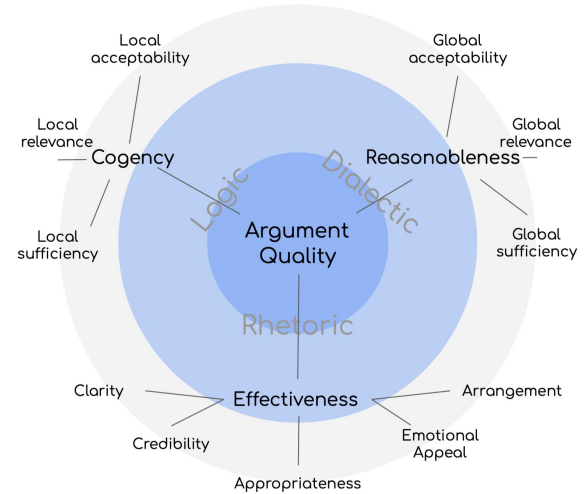
Opportunities for Argument Retrieval



Holistic assessment of “good” arguments based on a theoretically-grounded notion



More interpretable

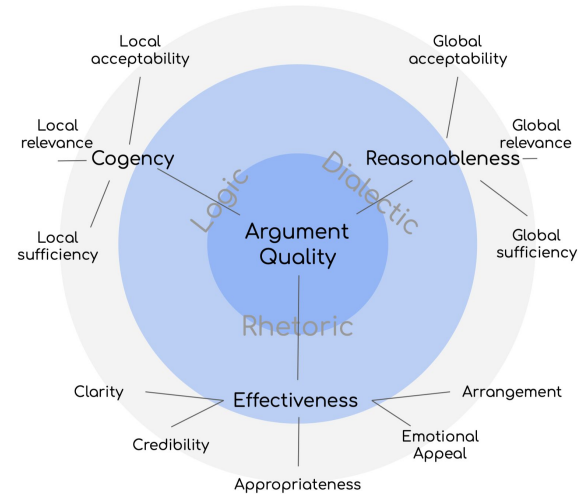


Opportunities for Argument Retrieval

◆ Holistic assessment of “good” arguments based on a theoretically-grounded notion

◆ More interpretable

◆ Potential for more focused and targeted, e.g., user-specific, retrieval



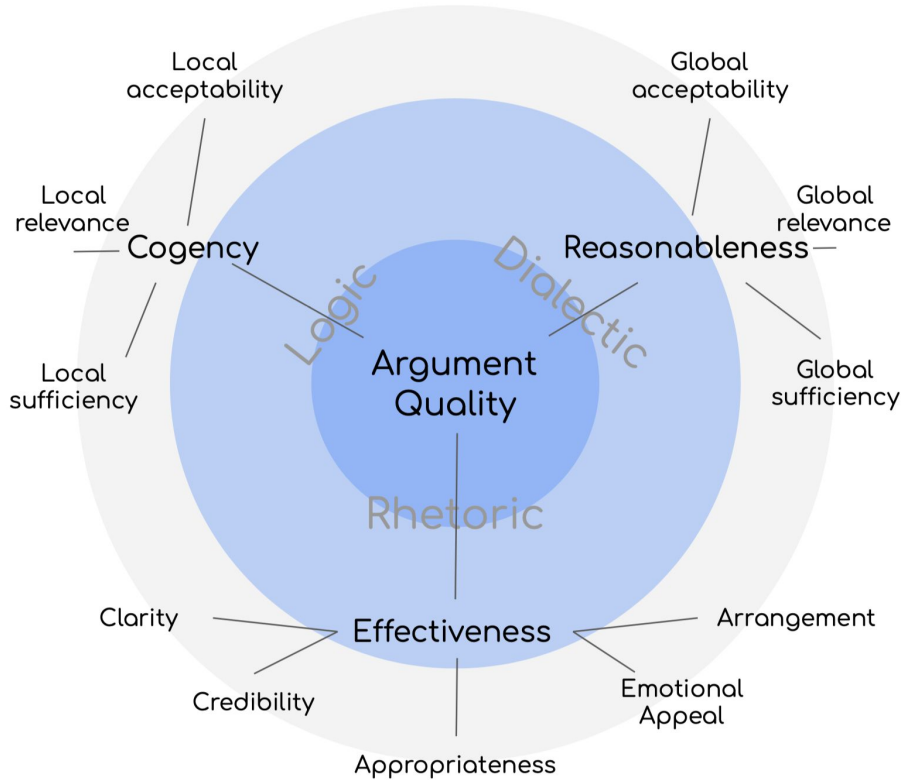


Where do we stand?



Taxonomy of theory-based AQ

(Wachsmuth et al., 2017a,b)

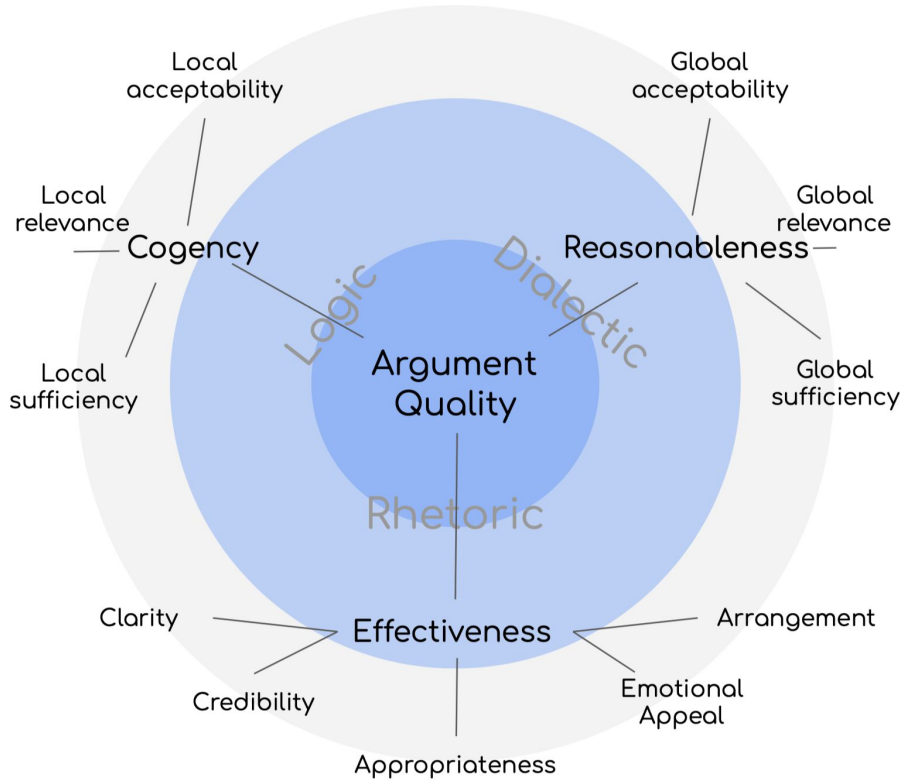


- assessment is challenging
- crowdsourcing is possible
- guidelines and task need to be simplified



Taxonomy of theory-based AQ

(Wachsmuth et al., 2017a,b)



- assessment is challenging
- crowdsourcing is possible
- guidelines and task need to be simplified

But, until last year,

- No large corpus
- No computational model



Grammarly Argument Quality Corpus (GAQCorpus)

First multi-domain corpus and largest English corpus annotated with theory-based Argument Quality scores

Lauscher, A., Ng, L., Napoles, C., & Tetreault, J. (2020, December). Rhetoric, Logic, and Dialectic: Advancing Theory-based Argument Quality Assessment in Natural Language Processing. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 4563-4574).

Ng, L., Lauscher, A., Tetreault, J., & Napoles, C. (2020, December). Creating a Domain-diverse Corpus for Theory-based Argument Quality Assessment. In *Proceedings of the 7th Workshop on Argument Mining* (pp. 117-126).



GAQCorpus

Simplifications

- Reduction of the taxonomy
 - Keep overall Argument Quality and three higher-level dimensions
 - Translate lower-level aspects to guiding questions
- Instruction Modifications
- 5-point scale



GAQCorpus

Annotation Process

- Guideline development with four expert annotators
 - Fluent or native in English
 - Advanced degree in Linguistics
- Pilot studies with crowd and experts
- Web Interface



GAQCorpus

Validation of our Simplifications

- 200 randomly sampled arguments and [gold annotations](#) from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, [TvsP](#))



GAQCorpus

Validation of our Simplifications

- 200 randomly sampled arguments and [gold annotations](#) from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, [TvsP](#))
- Crowd-sourced annotations with our simplified guidelines ([Ours](#))



GAQCorpus

Validation of our Simplifications

- 200 randomly sampled arguments and **gold annotations** from Dagstuhl-ArgQuality-Corpus-V2 (Wachsmuth et al., 2017a)
- Crowd-sourced annotations from (Wachsmuth et al., 2017b, **TvsP**)
- Crowd-sourced annotations with our simplified guidelines (**Ours**)

	Cogency	Effectiveness	Reasonableness	Overall
Ours	0.46	0.48	0.48	0.55
TvsP	0.27	0.38	0.13	0.43

Krippendorff's
alpha between
expert and
crowd
annotations



GAQCorpus

Data

- Debate forums (Debates)
 - Convince Me
 - Change My View
- Community Q&A forums (CQA)
 - Yahoo Answers: Law & Ethics
- Review forums (Reviews)
 - Yelp restaurant reviews





GAQCorpus

Results

- Total arguments: 5,285
- Three domains
- Portions:
 - Crowd annotations only (10 votes)
 - Expert annotations (1-3 votes)
 - Overlapping portions
with expert and crowd annotations



GAQCorpus Results

- Total arguments: 5,285
- Three domains
- Portions:
 - Crowd annotations only (10 votes)
 - Expert annotations (1-3 votes)
 - Overlapping portions with expert and crowd annotations

	Total	Train	Dev	Test
CQA	2,085	1,109	476	500
Debates	2,100	1,093	469	538
Reviews	1,100	700	300	100
All	5,285	2,902	1,245	1,138



Experiments

Models

Argument Length

SVR with lexical features

SVR with semantic features

Feature-rich SVR (Wachsmuth et al., 2016)

Single Task Learning w. BERT (BERT ST)



Experiments

Models

Argument Length

SVR with lexical features

SVR with semantic features

Feature-rich SVR (Wachsmuth et al., 2016)

Single Task Learning w. BERT (BERT ST)



Experiments Results

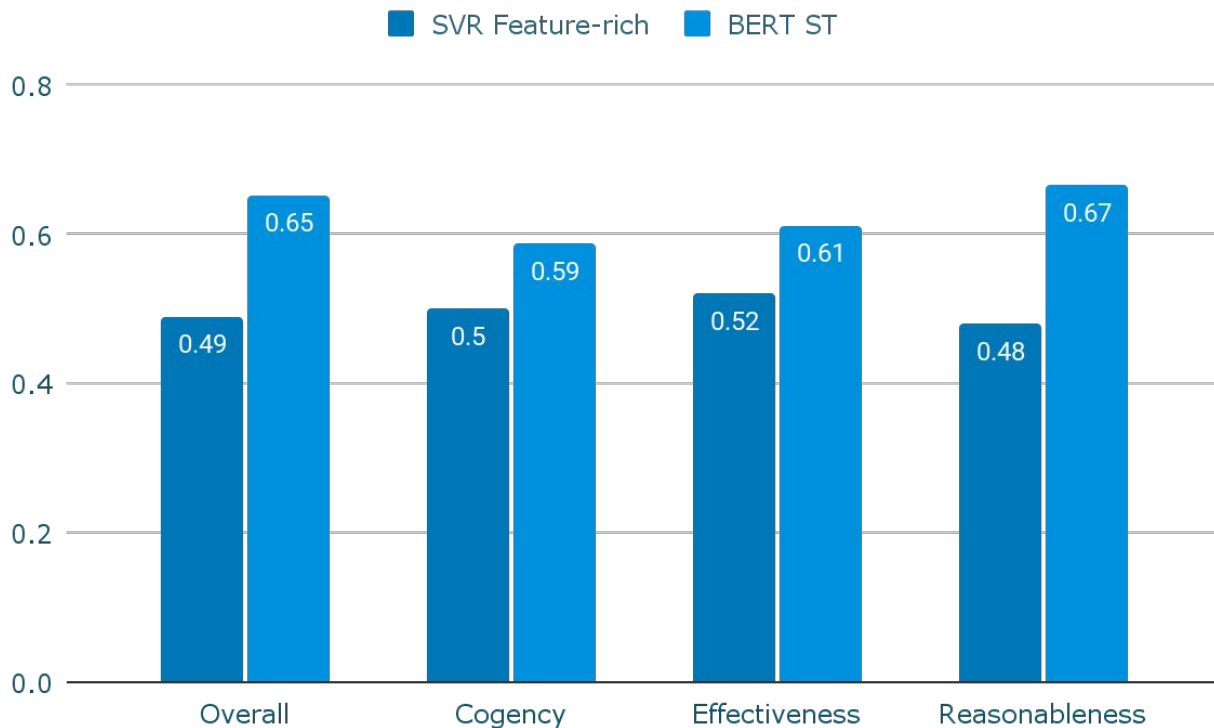
Pearson correlations of our model predictions with the annotation scores for the **Community Q&A** domain on our newly created GAQcorpus.





Experiments Results

Pearson correlations of our model predictions with the annotation scores for the Community Q&A domain on our newly created GAQcorpus.





Experiments

Models

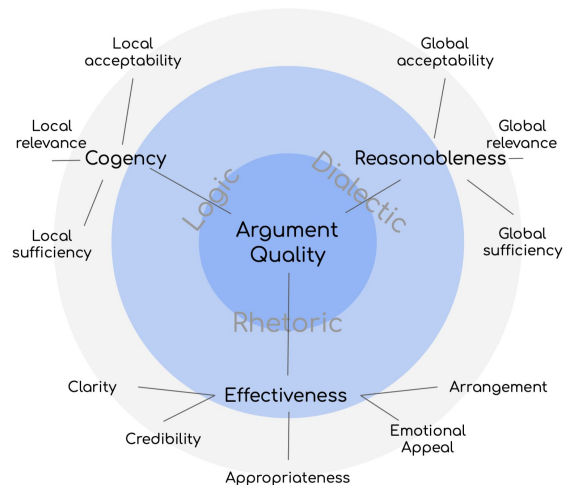
Argument Length

SVR with lexical features

SVR with semantic features

Feature-rich SVR

Single Task Learning w. BERT (BERT ST)





Experiments Models

Argument Length

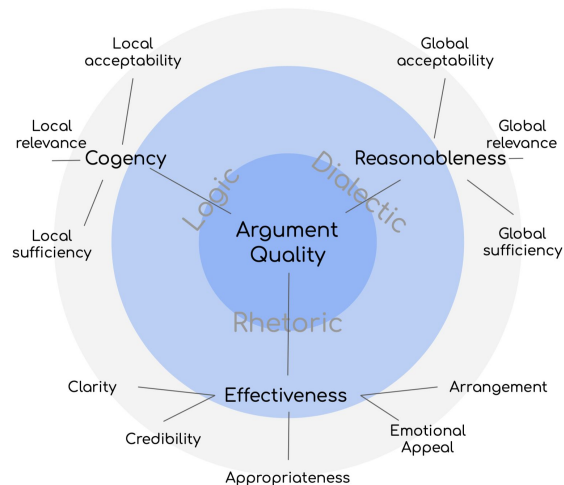
SVR with lexical features

SVR with semantic features

Feature-rich SVR

Single Task Learning w. BERT (BERT ST)

Flat MT Learning (BERT MT Flat)





Experiments

Models

Argument Length

SVR with lexical features

SVR with semantic features

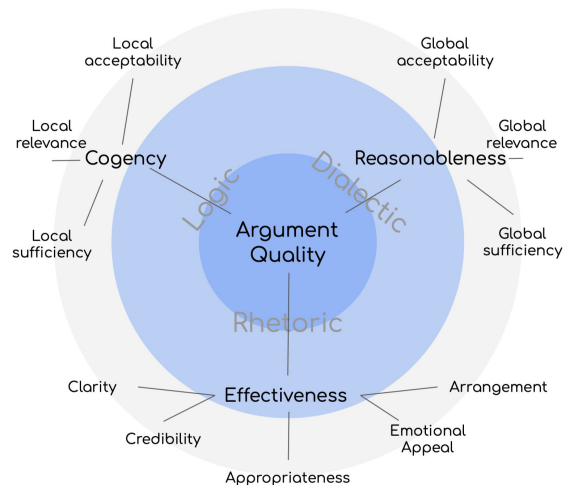
Feature-rich SVR

Single Task Learning w. BERT (BERT ST)

Flat MT Learning (BERT MT Flat)

Hierarchical MT Learning (BERT MT Hier)

Concatenation of the hidden document representation with the predictions for the lower-level dimensions for predicting overall AQ

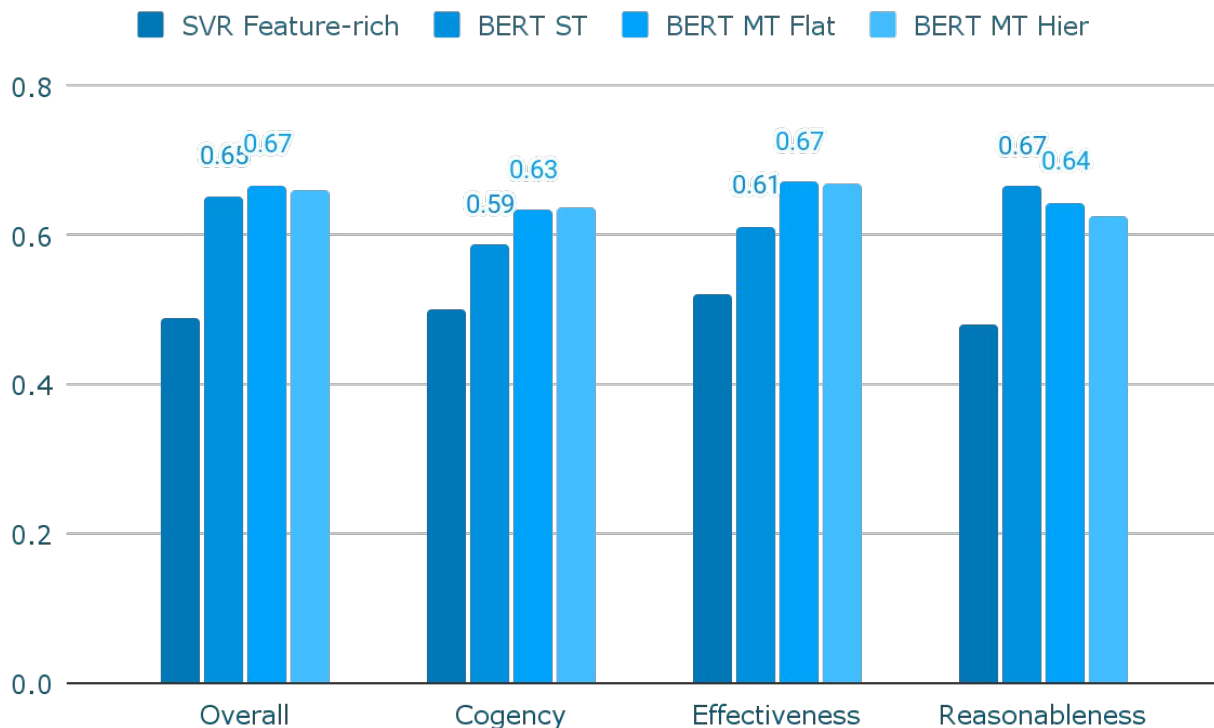


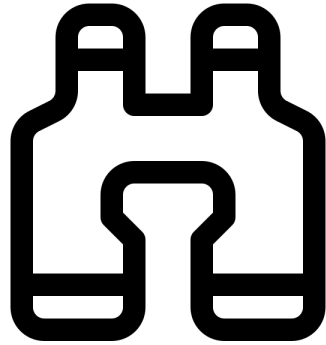


Experiments Results

Pearson correlations of our model predictions with the annotation scores for the Community Q&A domain on our newly created GAQcorpus.

Overall, the multi-task models outperform the single task model in 9 out of 12 cases.

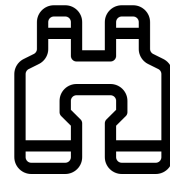




What challenges are we facing?

Challenge 1: Resources

A first larger scale corpus is available





Challenge 1: Resources

A first larger scale corpus is available

- only covers the four higher-level dimensions (overall AQ, cogency, effectiveness, reasonableness)
- only covers three domains
- only covers English



Challenge 1: Resources

A first larger scale corpus is available

- only covers the four higher-level dimensions (overall AQ, cogency, effectiveness, reasonableness)
- only covers three domains
- only covers English

We need to consider more domains and languages.
And what about an even finer-grained assessment?

Challenge 2: Advanced Knowledge

Intrinsically evaluating the quality makes sense,
but there is much more to it ...





Challenge 2: Advanced Knowledge

Intrinsically evaluating the quality makes sense, but there is much more to it ...

- Knowledge about the cultural background
- Knowledge about the audience
- Knowledge about the speaker
- Commonsense knowledge & world knowledge

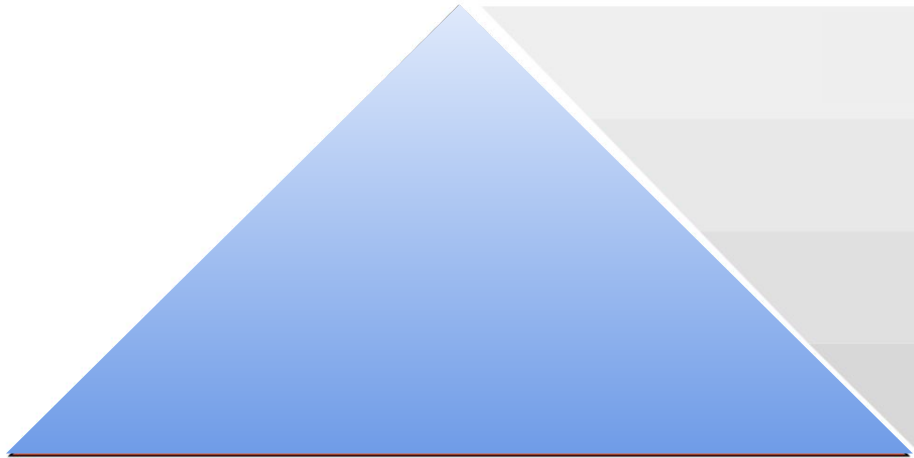


The Role of Knowledge in Argumentation

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

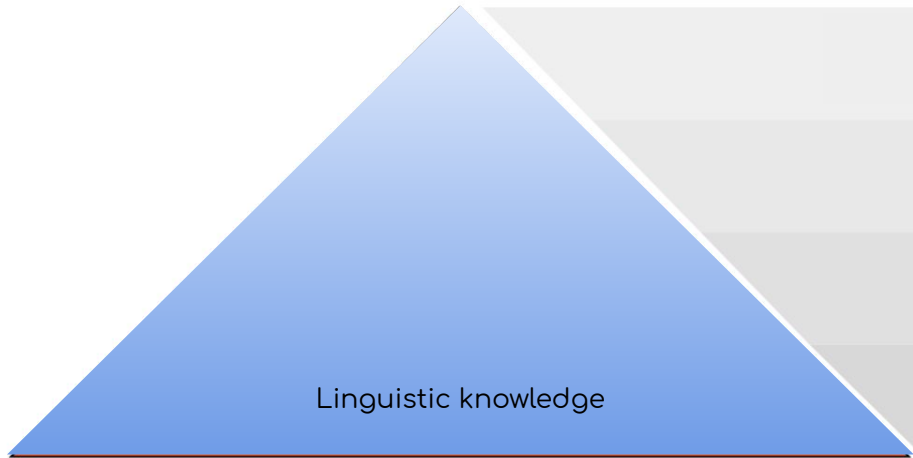


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

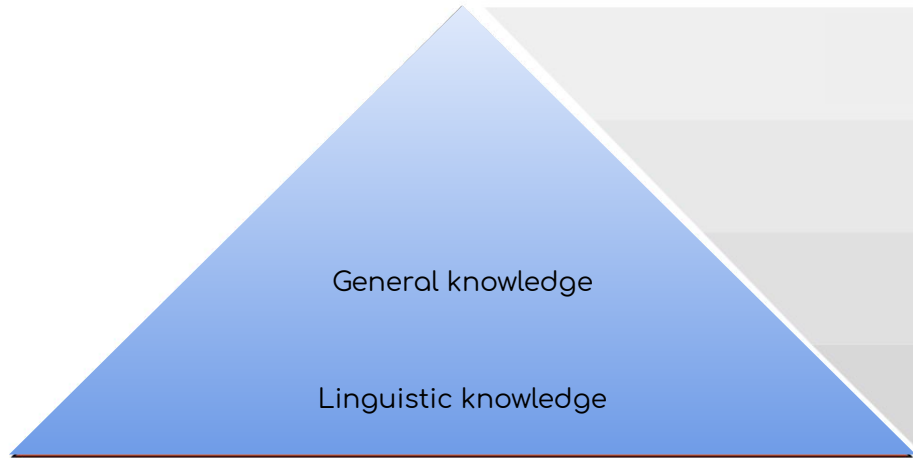


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

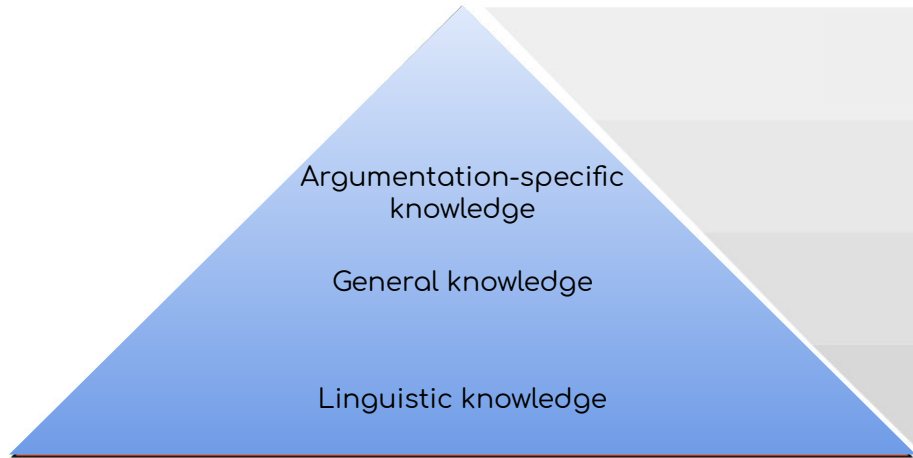


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

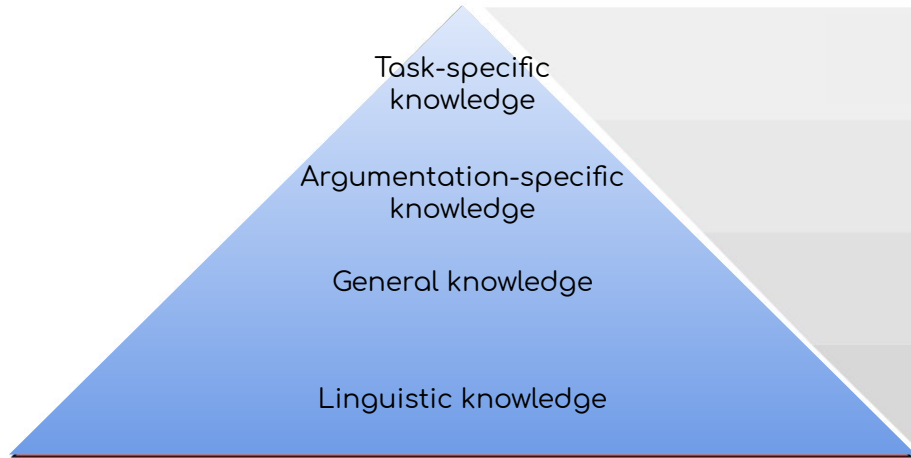


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

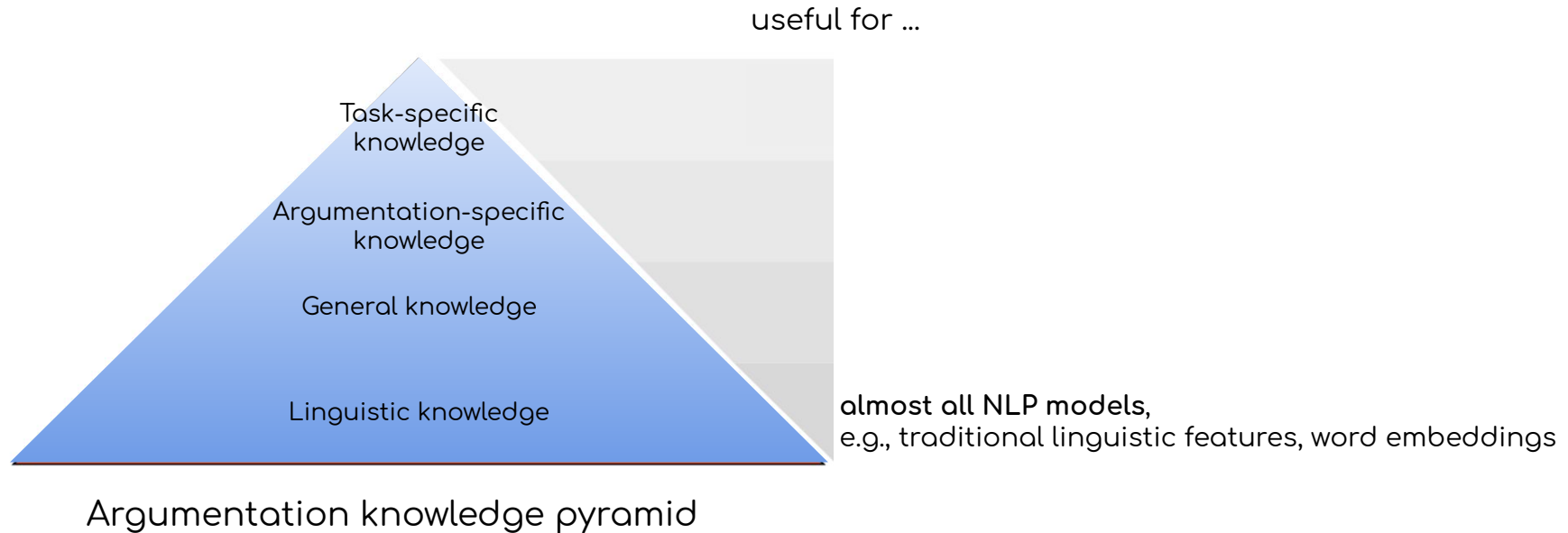


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



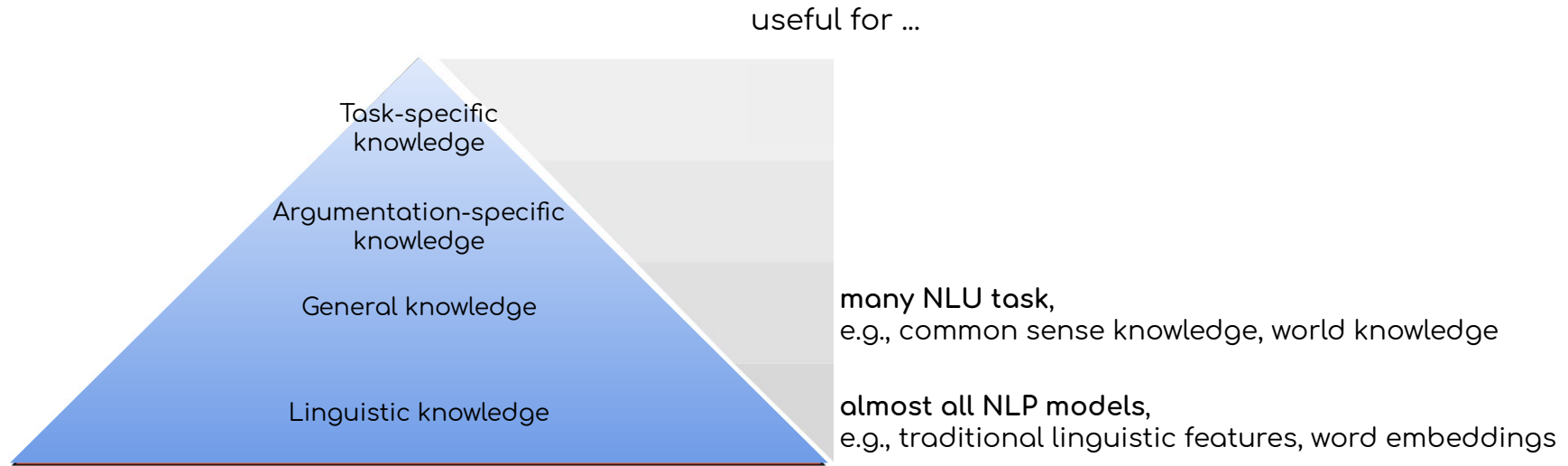
The Role of Knowledge in Argumentation



Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

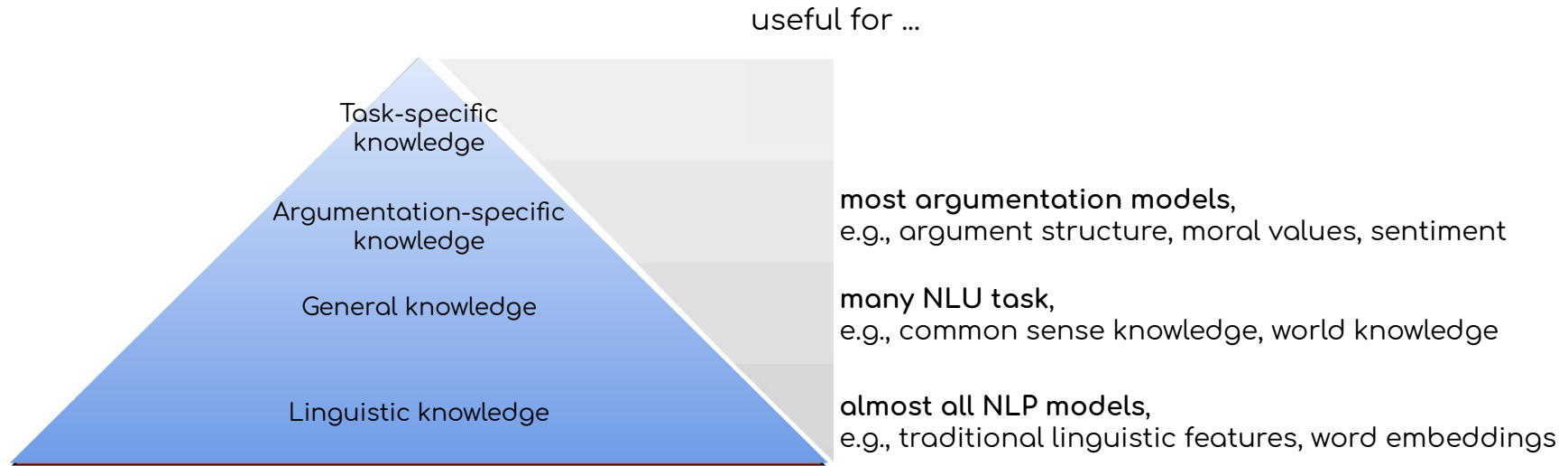


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

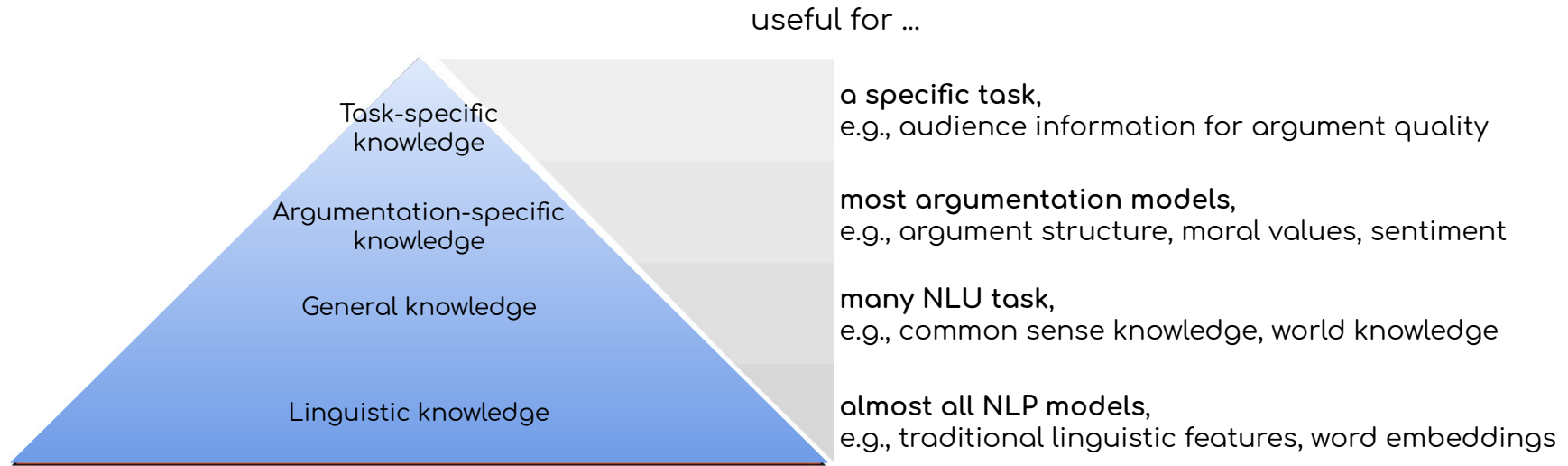


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation

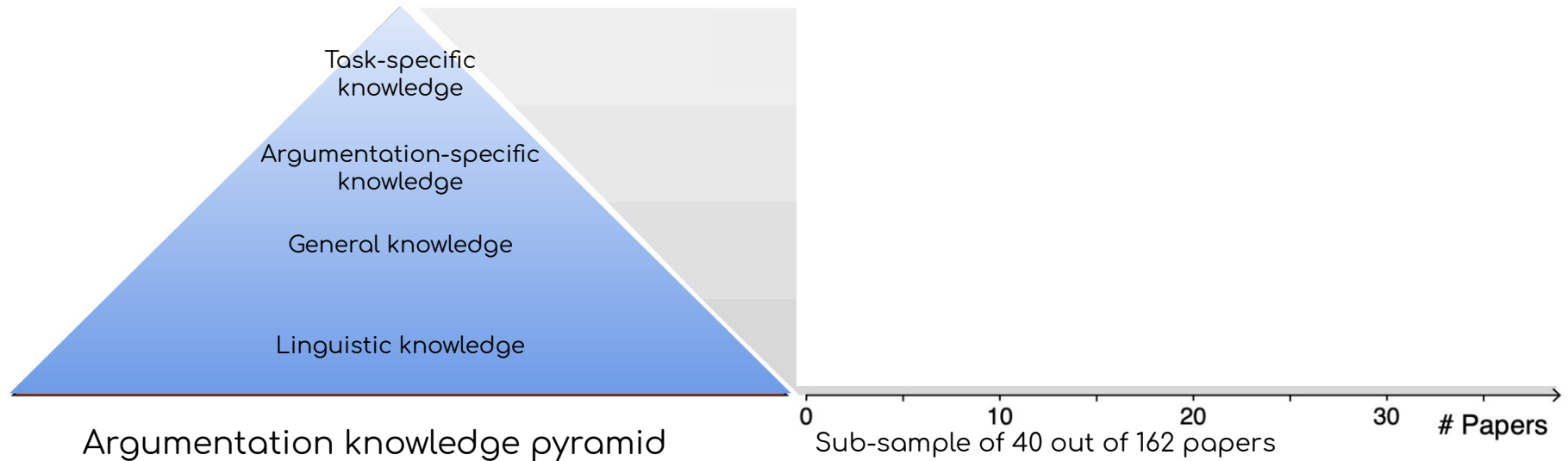


Argumentation knowledge pyramid

Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



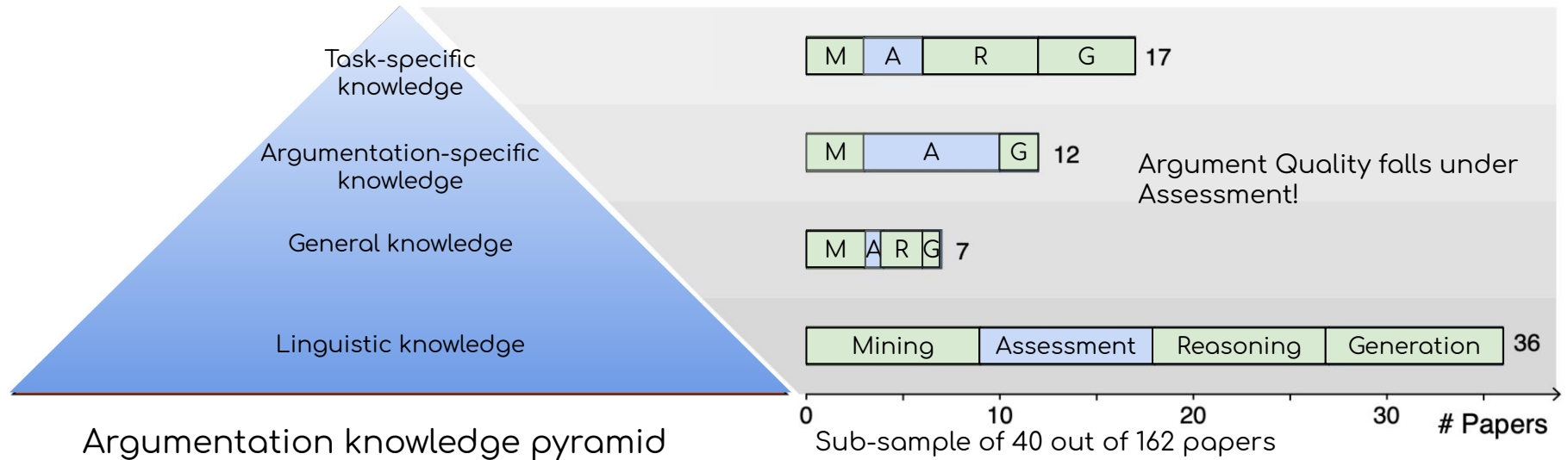
The Role of Knowledge in Argumentation



Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



The Role of Knowledge in Argumentation



Lauscher, A., Wachsmuth, H., Gurevych, I., & Glavaš, G. (2021). Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation. *arXiv preprint arXiv:2107.00281*.



Challenge 2: Advanced Knowledge

Intrinsically evaluating the Quality makes sense, but there is much more to it ...

- Knowledge about the cultural background
- Knowledge about the audience
- Knowledge about the speaker
- Commonsense knowledge & world knowledge

Where to obtain the information needed?
How to model it?

Challenge 3: Ethics



If argument retrieval influences people's opinions ...

... and argument retrieval is guided by argument quality

... and we also have evidence that argumentative corpora are biased

(Spliethover and Wachsmuth, 2020)



Challenge 3: Ethics

If argument retrieval influences people's opinions ...

... and argument retrieval is guided by argument quality

... and we also have evidence that argumentative corpora are biased

(Spliethover and Wachsmuth, 2020)

How can we ensure that the models' assessments are not unfairly biased?

Theory-based AQ for Advanced Argument Retrieval



Our Opportunities



Our Position



Our Challenges

Theory-based AQ for Advanced Argument Retrieval



Our Opportunities

- Holistic perspective
- Theoretically -grounded
- Improved interpretability
- More focused/ targeted retrieval



Our Position



Our Challenges

Theory-based AQ for Advanced Argument Retrieval



Our Opportunities

- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval



Our Position

- Data and computational models for three domains
- Intrinsic assessment



Our Challenges

Theory-based AQ for Advanced Argument Retrieval



Our Opportunities

- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval



Our Position

- Data and computational models for three domains
- Intrinsic assessment



Our Challenges

- Resources for more domains and languages, finer-grained assessment?
- Advanced knowledge?
- Ethical aspects, bias?

Theory-based AQ for Advanced Argument Retrieval



Our Opportunities

- Holistic perspective
- Theoretically-grounded
- Improved interpretability
- More focused/targeted retrieval



Our Position

- Data and computational models for three domains
- Intrinsic assessment



Our Challenges

- Resources for more domains and languages, finer-grained assessment?
- Advanced knowledge?
- Ethical aspects, bias?

Thank you very much for your interest!

References

Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., ... & Stein, B. (2017a, April). Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 176-187).

Wachsmuth, H., Naderi, N., Habernal, I., Hou, Y., Hirst, G., Gurevych, I., & Stein, B. (2017b, July). Argumentation quality assessment: Theory vs. practice. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 250-255).

Wachsmuth, H., Al Khatib, K., & Stein, B. (2016, December). Using argument mining to assess the argumentation quality of essays. In *Proceedings of COLING 2016, the 26th international conference on Computational Linguistics: Technical papers* (pp. 1680-1691).

Spliethöver, M., & Wachsmuth, H. (2020, December). Argument from Old Man's View: Assessing Social Bias in Argumentation. In *Proceedings of the 7th Workshop on Argument Mining* (pp. 76-87).