

Mining, Assessing, and Improving Arguments in NLP and the Social Sciences

Gabriella Lapesa*, Eva Maria Vecchi†, Serena Villata‡, Henning Wachsmuth**

*GESIS Leibniz Institute for the Social Sciences and Heinrich-Heine University Düsseldorf,

†University of Stuttgart, Institute for Natural Language Processing

‡Université Côte d’Azur Inria CNRS I3S

**Leibniz University Hannover, Institute of Artificial Intelligence

*gabriella.lapesa@gesis.org, †eva-maria.vecchi@ims.uni-stuttgart.de,

‡villata@i3S.unice.fr, **h.wachsmuth@ai.uni-hannover.de

Abstract

Computational argumentation is an interdisciplinary research field, connecting Natural Language Processing (NLP) to other disciplines such as the social sciences. The focus of recent research has concentrated on *argument quality assessment*: what makes an argument good or bad? We present a tutorial with a strong interdisciplinary and interactive nature structured along three main coordinates: (1) the notions of argument quality (AQ) across disciplines (how do we recognize good and bad arguments?), with a particular focus on the interface between Argument Mining (AM) and Deliberation Theory; (2) the modeling of subjectivity (who argues to whom; what are their beliefs?); and (3) the generation of improved arguments (what makes an argument better?). The tutorial will also touch upon a series of topics that are particularly relevant for the LREC-COLING audience (the issue of resource quality for the assessment of AQ; the interdisciplinary application of AM and AQ in a text-as-data approach to political science), in line with the developments in NLP (LLMs for AQ assessment), and relevant for the societal applications of AQ assessment (bias and debiasing). We will involve the participants in two annotation studies on the assessment and the improvement of quality. The full materials of this tutorial can be found at <https://sites.google.com/view/argmintutorial-2024/home-page>.

Keywords: argument mining, quality assessment, annotation, data quality

1. Introduction

Computational argumentation is a field encompassing varying tasks on the automated analysis and synthesis of natural language arguments. Until recently, research in Natural Language Processing (NLP) mostly dealt with *Argument Mining* (AM), that is, the identification of argumentative claims that convey a stance towards some controversial issue, along with evidence given as reasons for the claims. AM has been studied for various genres (Mochales and Moens, 2011; Habernal and Gurevych, 2017; Dusmanu et al., 2017a) and argument models (Toulmin, 1958; Walton et al., 2008; Freeman, 2011).

Whether we conceptualize the function of argumentation as “reason giving” or “persuasion” (refer to Lawrence and Reed (2019) for a discussion of this dichotomy) the question of what makes an argument good (or better than another argument) has been at the core of research in argument mining (Wachsmuth et al., 2017; Lauscher et al., 2020; Marro et al., 2022). A first edition of this tutorial has been taught by the same authors of this tutorial at EACL 2023 (Lapesa et al., 2023). In the following, we present the main tutorial coordinates, shared with the previous edition (Section 1.1). This LREC-COLING 2024 edition, beyond the obvious update of the literature, will feature new topics that we de-

vised to fit the conference audience, to account for the fast pace of research in NLP, particularly in the context of large language models, and to broaden the interdisciplinary scope of the tutorial (Section 1.2).

1.1. Tutorial coordinates

In this tutorial, we start from the body of research on AM. Unlike earlier NLP tutorials on argumentation (Budzynska and Reed, 2019; Bar-Haim et al., 2021), however, our focus is a task that recently got into the center of attention: *argument quality assessment*, that is, to rate or to compare how good arguments are with respect to one or more defined quality dimensions.

The NLP Perspective: Assessing Argument Quality Let us start with the concrete example of argument quality annotations in Figure 1, taken from Lauscher et al. (2020). The topic is “freedom of speech”, and the stance is “against” (i.e., the government has the right to censorship). Quality is assessed here in four dimensions: *cogency* (is the conclusion adequately supported with acceptable, relevant, and sufficient premises?), *effectiveness* (how persuasive is the argument?), *reasonableness* (is the argument good in the context of the debate in which it is framed?), and *overall quality*.

The example illustrates the challenges which

<i>Title: Should 'blogging' be a capital crime? Iran is considering it...</i>				
<i>Stance: A government has the right to censor speech (...)</i>				
<i>Text: My government doesn't give me freedom of speech, so I have to argue for this side. Freedom of speech is bad because ... um ... then Our Leader's beliefs could be challenged. No one wants that. I mean, if everyone would just say and believe what Our Leader says to, we wouldn't need those firing squads altogether! Everyone wins.</i>				
	Cogency	Effectiveness	Reasonableness	Overall
Annotator 1	4	1	1	2
Annotator 2	4	5	3	4
Annotator 3	2	2	2	2

Figure 1: Argument quality assessment from Lauscher et al. (2020): Example argument, annotated for four dimensions by three annotators, with partial agreement.

we take as coordinates of this tutorial. The first challenge is the identification and definition of appropriate *dimensions* for quality assessment: for example, in this case, the effectiveness label conflates several aspects. The second challenge in quality assessment is *subjectivity*. In our example, the three annotators (linguistics experts) clearly disagree in their assessment. Lauscher et al. (2020) report that a crucial factor of disagreement of Annotators 1 and 2 was their perception of the ironic tone behind the text. Interestingly, for both of them, the text has a medium-high degree of cogency (so it is logically pretty “healthy”). A further challenge would be to improve the quality of this argument: How would we make this argument more effective? Do we need more irony, less irony, or a stronger statement of the stance?

To inform participants about argument quality, the tutorial will systematically review existing research on argument quality based on the literature (Wachsmuth et al., 2017), outlining the subjectiveness of quality dimensions as a key problem. In an interactive annotation session, participants will explore and discuss the assessment of quality on real-life arguments. They will be encouraged to take a critical standpoint to the annotation guidelines, learning in a concrete scenario how difficult it is to establish a trade-off between expressivity of the annotation schema and feasibility of the task.

The Social Science Perspective: Assessing Deliberative Quality To demonstrate the impact of argument quality in practice, the tutorial will bridge research in NLP with the social sciences, looking at deliberative democracy in particular. Deliberative democracy is an approach to democratic processes which does not focus on the output of decision-making, but on the discourse exchange that precedes it (Bächtiger and Parkinson, 2019). Crucially, deliberative theory scholars have been asking the same question as computational argumentation: What makes a contribution to a discussion good? This has led to the development of a *discourse qual-*

ity index to assess the quality of a discourse contribution (Steenbergen et al., 2003; Gerber et al., 2016).

Modeling Subjectivity Next, we will deal with subjectivity, modeling the parties involved in debates along with their values and beliefs. The connections of argument quality and deliberative quality highlight the subjective nature of argumentation, one of the three main coordinates of this tutorial. Subjectivity has been the trigger of an “affective turn” in both deliberative theory and computational argumentation. In the former, this has implied a switch from a purely rational perspective on deliberation to one which incorporates emotions, personal narratives, humor (Hoggett and Thompson, 2002; Black, 2020; Esau, 2018; Esau and Friess, 2022). In the latter, the affective turn has brought personal argumentation at center stage, highlighting the role played by human values (Kiesel et al., 2022), moral discourse (Alshomary et al., 2022), and narratives (Falk and Lapesa, 2022). In the tutorial, we aim to encourage participants to reflect on the two-fold role that subjectivity plays in quality assessment: subjective factors in quality assessment (e.g., interpretation of humor, as in the example above), and subjective factors in the production of an argument (e.g., all the “personal argumentation” ingredients listed before).

Improving Arguments The subjectivity topic will lead to another interactive session where the goal is to improve the quality of arguments. Limitations will be discussed as well as first research on quality-related argument generation (Gurcke et al., 2021; Skitalinskaya et al., 2023), before the tutorial concludes with an outlook on future perspectives.

1.2. Further topics

Data Quality What are the requirements for a high-quality resource to model AQ assessment? Is annotator disagreement necessarily a cue to bad quality? What is the role of human baselines in AQ assessment? Which sample should the annotated data be representative of? Which challenges are posed by crowdsourcing as an annotation method? We will wrap up every session with a dedicated slot for reflection on available resources and desiderata.

Bias and Debiasing Tightly related to the notion of data quality is the one of bias in AM datasets (Spliethöver and Wachsmuth, 2020) and debiasing methods for AM (Holtermann et al., 2022)

LLMs and AQ Assessment The fast developments and performance boosts offered by LLMs represent an incredible opportunity. What are the challenges and the potential risks of LLMs for AQ assessment?

Text-as-Data Approaches to political science AM and text-as-data approaches to political sci-

ence research find a natural overlap in the tasks of claim, stance, evidence detection. Moving a step forward, what is the relation between AQ and widely investigated phenomena in political science, such as electoral success or polarization?

2. Target Audience

The tutorial targets both participants who are new to the field of computational argumentation and those who need a comprehensive overview of techniques and applications. As the tutorial is interdisciplinary by design, it is also of interest to participants from a social sciences background who hope to integrate their knowledge within NLP. Finally, we expect the tutorial to attract attention from people interested in NLP techniques that currently impact the social and political world, in general. Basic knowledge of linguistics and computational linguistics is required.

3. Outline

Part I (60 min.) Mining Arguments

- Overview of computational argumentation
- Argument mining: Humans vs. computers
- Achieved results and open challenges
- Data quality: resources overview & reflection on desiderata

Part II (60 min.) The NLP Perspective: Assessing Argument Quality

- What makes an argument “good”?
- Logical, rhetorical, and dialectical dimensions of argument quality
- Subjectiveness as the key challenge for annotation and modeling
- Discussion of the notions of argument quality: Are they sufficient? Are they all necessary?
- Data quality: resources overview & reflection on desiderata

Part III (60 min.) Interactive Session 1

- Annotation: Assessment of sample arguments
- Consolidation: To what extent participants agree? Where not, and why?
- Discussion: What are alternative strategies to subjective quality annotation?

Part IV (60 min.) The Social Sciences Perspective

- Direct democracy, deliberative theories, and e-deliberation
- Deliberative quality: Features and annotation
- Integration of deliberative features in computational architectures
- Application: Argument quality for social good
- Application: Argument Mining in political science text-as-data research.

- Data quality: resources overview & reflection on desiderata

Part V (60 min.) Modeling Subjectivity

- Authors, audiences, and third parties
- Human values, moral foundations, narratives
- Issues with subjectivity: exploiting annotators’ disagreements
- Bias and debiasing
- Data quality: resources overview & reflection on desiderata

Part VI (60 min.) Interactive Session 2

- Annotation: Rewriting of sample arguments
- Consolidation: What was improved and how?
- Discussion: What can be improved, what not?

Part VII (60 min.) Conclusion: open challenges and lessons learned

- Generation Methods to improve argument quality
- Challenges: multilinguality, multimodality
- LLMs for AQ assessment
- Conclusions and next steps for the field

4. Diversity Considerations

We believe that exposing the students to the deliberative perspective of argumentation will be fruitful and enriching, as it might not be known to the typical *CL audience. It is our goal that participants leave our tutorial having learned the value of taking multiple disciplinary perspectives into account, even in a rather technical (logic- and NLP-oriented) subject such as computational argumentation. Besides, our focus on subjectivity and personal argumentation as positive features (and not bugs) brings individuals and their differences at center stage, contributing to inclusivity in the field.

5. Reading List

Survey Papers (Cabrio and Villata, 2018; Lawrence and Reed, 2019; Vecchi et al., 2021; Lauscher et al., 2022; Wang et al., 2023)

Mining Arguments (Habernal and Gurevych, 2017; Daxenberger et al., 2017; Dusmanu et al., 2017b; Schaefer and Stede, 2020)

Assessing Argument Quality (Wachsmuth et al., 2017; Lauscher et al., 2020; Marro et al., 2022; Ziegenbein et al., 2023)

Assessing Deliberative Quality (Steenbergen et al., 2003; Gerber et al., 2016)

Improving Arguments (Hua and Wang, 2018; Gurcke et al., 2021; Syed et al., 2023; Skitalinskaya and Wachsmuth, 2023; Skitalinskaya et al., 2023)

Challenges (Durmus et al., 2019; Toledo-Ronen et al., 2020; Spliethöver and Wachsmuth, 2020)

6. Presenters

Gabriella Lapesa is a team lead for Data Science Methods in the Department for Computational Social Sciences at the Leibniz Institute for Social Sciences (GESIS Köln) and a junior professor of Responsible Data Science and Machine Learning at the Heinrich-Heine University of Düsseldorf. She also leads the research group E-DELIB (*Powering-up E-DELIBeration: towards AI-supported moderation*) at the University of Stuttgart. Her research targets the intersection between NLP and the Social Sciences, with a general focus on the development of NLP methods to support social science research and real-world applications (i.e., moderation in deliberative discussions). She co-chaired the 9th Argument Mining workshop (2022) and co-taught a course and a tutorial on interdisciplinary Argument Mining, respectively ESSLLI 2022 (with E.M. Vecchi) and EACL 2023 (with the other authors of this proposal).

Eva Maria Vecchi holds a Ph.D. degree in cognitive and neurosciences. She is a postdoctoral researcher at the Institute for Natural Language Processing at IMS Stuttgart, working on the E-DELIB project. Her focus is on the interdisciplinary effort between NLP techniques for argument mining (AM) and theories in the social sciences with the goal of a more collaborative, productive, and ethical endeavor for e-Deliberation. She has taught courses and tutorials on AM and other topics, e.g., ESSLLI 2022 (with G. Lapesa) and EACL 2023 (with the authors of this proposal). Her current research aims at a better understanding of the role bias has in computational argumentation and e-Deliberation, particularly the impact it has on the models, implementation, and social aspects of computational argumentation.

Serena Villata is a research director in computer science at CNRS, and she pursues her research at the I3S laboratory in Sophia Antipolis (France). Her research area is computational argumentation, with a focus on legal and medical texts, political debates and social network harmful content (abusive language, disinformation). Her work conjugates argument-based reasoning frameworks with natural language arguments extracted from text. She is the author of over 150 scientific publications on the topic. She holds a Chair of the Interdisciplinary Institute for AI 3IA Côte d’Azur on “Artificial Argumentation for Humans”. Serena has co-chaired the 7th Workshop on Argument Mining at COLING 2020. She has also given tutorials on Argument Mining at ESSLLI 2017¹ and IJCAI 2016².

¹<https://www.irit.fr/esslli2017/courses/39.html>

²https://ijcai-16.org/index.php/welcome/view/accepted_tutorials/

Henning Wachsmuth is the head of the Natural Language Processing Group at Leibniz University Hannover. He is an internationally leading researcher on computational argumentation with about 70 publications on the topic, many at major NLP and AI venues. Other interests include social bias mitigation, computational reframing, and explainable NLP. Henning has co-chaired the 6th Workshop on Argument Mining at ACL 2019, and has given tutorials on argumentation at ASIRF 2018 (Cole and Achilles, 2019), EuroCSS 2018,³ KI 2019 (Benzmüller and Stuckenschmidt, 2019), and KI 2020 (Schmid et al., 2020). He is an initiator of the CLEF shared task series Touché on argument retrieval (Bondarenko et al., 2022), and co-chaired SemEval tasks on argument reasoning comprehension (Habernal et al., 2018), propaganda technique detection (Da San Martino et al., 2020), and identifying human values in arguments (Kiesel et al., 2023).

7. Ethics statement

The breadth of computational argumentation research, from previous focus on mining to more recent interest in assessment and improvement, encompasses huge benefit to various fields, e.g., NLP and Computational Social Sciences; however, we acknowledge the responsibility of the research to remain sensitive to the ethical concerns that are both generally shared in these fields as well as unique to automated assessment and improvement of arguments. Privacy concerns arise regarding the mining and analysis of private or sensitive data, such as social media posts, emails, or personal correspondence, without informed consent or when the data is not properly anonymized.

Argument quality assessment may be used in sensitive applications, e.g., argumentative writing support, legal or ethical decision-making processes, or guidance on political opinion formation, in which factual errors, bias concerns, and unfair evaluations are particularly problematic, as they may easily lead to or perpetrate wrong or shifted beliefs. Implementing measures to assess and improve arguments, particularly when incorporating subjectivity and human values, may open the door to the manipulation of arguments, such as strategically crafting arguments to achieve desired outcomes. In the contexts of social sciences, political campaigns, and social media, this is of considerable concern as it can lead to the spread of misinformation and unethical persuasion tactics at both a local and global level.

³<http://symposium.computationalsocialscience.eu/2018/>

8. Bibliographical References

- Milad Alshomary, Roxanne El Baff, Timon Gurcke, and Henning Wachsmuth. 2022. [The moral debater: A study on the computational generation of morally framed arguments](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8782–8797, Dublin, Ireland. Association for Computational Linguistics.
- Andre Bächtiger and John Parkinson. 2019. *Towards a New Deliberative Quality*. Oxford University Press, Cambridge, MA, USA.
- Roy Bar-Haim, Liat Ein-Dor, Matan Orbach, Elad Venezian, and Noam Slonim. 2021. [Advances in debating technologies: Building AI that can debate humans](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Tutorial Abstracts*, pages 1–5, Online. Association for Computational Linguistics.
- Christoph Benz Müller and Heiner Stuckenschmidt, editors. 2019. *KI 2019: Advances in Artificial Intelligence - 42nd German Conference on AI, Kasel, Germany, September 23-26, 2019, Proceedings*, volume 11793 of *Lecture Notes in Computer Science*. Springer.
- Laura Black. 2020. [Framing democracy and conflict through storytelling in deliberative groups](#). *Regular Issue*, 9(1).
- Alexander Bondarenko, Maik Fröbe, Johannes Kiesel, Shahbaz Syed, Timon Gurcke, Meriem Beloucif, Alexander Panchenko, Chris Biemann, Benno Stein, Henning Wachsmuth, Martin Potthast, and Matthias Hagen. 2022. Overview of touché 2022: Argument retrieval. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 311–336, Cham. Springer International Publishing.
- Katarzyna Budzyska and Chris Reed. 2019. [Advances in argument mining](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 39–42, Florence, Italy. Association for Computational Linguistics.
- Elena Cabrio and S. Villata. 2018. Five years of argument mining: a data-driven analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 5427–5433.
- Amelia W. Cole and Linda Achilles. 2019. [Autumn school for information retrieval and foraging 2018](#). *SIGIR Forum*, 52(2):87–91.
- Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. [SemEval-2020 task 11: Detection of propaganda techniques in news articles](#). In *14th International Workshop on Semantic Evaluation (SemEval 2020)*, pages 1377–1414, Barcelona (online). Association for Computational Linguistics.
- Johannes Daxenberger, Steffen Eger, Ivan Habernal, Christian Stab, and Iryna Gurevych. 2017. [What is the essence of a claim? cross-domain claim identification](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2055–2066, Copenhagen, Denmark. Association for Computational Linguistics.
- Esin Durmus, Faisal Ladhak, and Claire Cardie. 2019. [The role of pragmatic and discourse context in determining argument impact](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5668–5678, Hong Kong, China. Association for Computational Linguistics.
- Mihai Dusmanu, Elena Cabrio, and Serena Villata. 2017a. [Argument mining on Twitter: Arguments, facts and sources](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2317–2322, Copenhagen, Denmark. Association for Computational Linguistics.
- Mihai Dusmanu, Elena Cabrio, and Serena Villata. 2017b. [Argument mining on twitter: Arguments, facts and sources](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2317–2322.
- Katharina Esau. 2018. [Capturing citizens' values: On the role of narratives and emotions in digital participation](#). *Analyse & Kritik*, 40(1):55–72.
- Katharina Esau and Daniel Friess. 2022. [What creates listening online? exploring reciprocity in online political discussions with relational content analysis](#). *Journal of Deliberative Democracy*, 18(1):1–16.
- Neele Falk and Gabriella Lapesa. 2022. [Reports of personal experiences and stories in argumentation: datasets and analysis](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

- Papers), pages 5530–5553, Dublin, Ireland. Association for Computational Linguistics.
- James B. Freeman. 2011. *Argument Structure: Representation and Theory*. Springer.
- Marlené Gerber, André Bächtiger, Susumu Shikano, Simon Reber, and Samuel Rohr. 2016. Deliberative abilities and influence in a transnational deliberative poll (europolis). *British Journal of Political Science*, 48(4):1093–1118.
- Timon Gurcke, Milad Alshomary, and Henning Wachsmuth. 2021. [Assessing the sufficiency of arguments through conclusion generation](#). In *Proceedings of the 8th Workshop on Argument Mining*, pages 67–77, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1).
- Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. [SemEval 2018 Task 12: The Argument Reasoning Comprehension Task](#). In *12th International Workshop on Semantic Evaluation (SemEval 2018)*. Association for Computational Linguistics.
- Paul Hoggett and Simon Thompson. 2002. [Toward a democracy of the emotions](#). *Constellations*, 9(1):106–126.
- Carolin Holtermann, Anne Lauscher, and Simone Ponzetto. 2022. [Fair and argumentative language modeling for computational argumentation](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7841–7861, Dublin, Ireland. Association for Computational Linguistics.
- Xinyu Hua and Lu Wang. 2018. [Neural argument generation augmented with externally retrieved evidence](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 219–230, Melbourne, Australia. Association for Computational Linguistics.
- Johannes Kiesel, Milad Alshomary, Nicolas Handke, Xiaoni Cai, Henning Wachsmuth, and Benno Stein. 2022. [Identifying the human values behind arguments](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4459–4471, Dublin, Ireland. Association for Computational Linguistics.
- Johannes Kiesel, Milad Alshomary, Nailia Mirzakhmedova, Maximilian Heinrich, Nicolas Handke, Henning Wachsmuth, and Benno Stein. 2023. [SemEval-2023 task 4: ValueEval: Identification of human values behind arguments](#). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2287–2303, Toronto, Canada. Association for Computational Linguistics.
- Gabriella Lapesa, Eva Maria Vecchi, Serena Villata, and Henning Wachsmuth. 2023. [Mining, assessing, and improving arguments in NLP and the social sciences](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 1–6, Dubrovnik, Croatia. Association for Computational Linguistics.
- Anne Lauscher, Lily Ng, Courtney Napoles, and Joel Tetreault. 2020. [Rhetoric, logic, and dialectic: Advancing theory-based argument quality assessment in natural language processing](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4563–4574, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Anne Lauscher, Henning Wachsmuth, Iryna Gurevych, and Goran Glavaš. 2022. [Scientia Potentia Est—On the Role of Knowledge in Computational Argumentation](#). *Transactions of the Association for Computational Linguistics*, 10:1392–1422.
- John Lawrence and Chris Reed. 2019. [Argument mining: A survey](#). *Computational Linguistics*, 45(4):765–818.
- Santiago Marro, Elena Cabrio, and Serena Villata. 2022. Graph embeddings for argumentation quality assessment. In *Findings of EMNLP*.
- Raquel Mochales and Marie-Francine Moens. 2011. [Argumentation mining](#). *Artificial Intelligence and Law*, 19(1):1–22.
- Robin Schaefer and Manfred Stede. 2020. [Annotation and detection of arguments in tweets](#). In *Proceedings of the 7th Workshop on Argument Mining*, pages 53–58, Online. Association for Computational Linguistics.
- Ute Schmid, Franziska Klügl, , and Diedrich Wolter, editors. 2020. *KI 2020: Advances in Artificial Intelligence: 43rd German Conference on AI, Bamberg, Germany, September 21–25, 2020, Proceedings*. Springer-Verlag, Berlin, Heidelberg.
- Gabriella Skitalinskaya, Maximilian Spliethöver, and Henning Wachsmuth. 2023. [Claim optimization in computational argumentation](#). In *Pro-*

- ceedings of the 16th International Natural Language Generation Conference*, pages 134–152, Prague, Czechia. Association for Computational Linguistics.
- Gabriella Skitalinskaya and Henning Wachsmuth. 2023. [To revise or not to revise: Learning to detect improvable claims for argumentative writing support](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15799–15816, Toronto, Canada. Association for Computational Linguistics.
- Maximilian Spliethöver and Henning Wachsmuth. 2020. [Argument from old man’s view: Assessing social bias in argumentation](#). In *Proceedings of the 7th Workshop on Argument Mining*, pages 76–87, Online. Association for Computational Linguistics.
- Marco R. Steenbergen, Andre Baechtiger, Markus Spörndli, and Jürg Steiner. 2003. Measuring political deliberation: A discourse quality index. *Comparative European Politics*, 1:21–48.
- Shahbaz Syed, Timon Ziegenbein, Philipp Heinisch, Henning Wachsmuth, and Martin Potthast. 2023. [Frame-oriented summarization of argumentative discussions](#). In *Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue*, pages 114–129, Prague, Czechia. Association for Computational Linguistics.
- Orith Toledo-Ronen, Matan Orbach, Yonatan Bilu, Artem Spector, and Noam Slonim. 2020. [Multi-lingual argument mining: Datasets and analysis](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 303–317, Online. Association for Computational Linguistics.
- Stephen E. Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.
- Eva Maria Vecchi, Neele Falk, Iman Jundi, and Gabriella Lapesa. 2021. Towards argument mining for social good: A survey. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1338–1352.
- Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. [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*, pages 176–187, Valencia, Spain. Association for Computational Linguistics.
- Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.
- Xiaoou Wang, Elena Cabrio, and Serena Villata. 2023. [Argument and counter-argument generation: A critical survey](#). In *Natural Language Processing and Information Systems*, pages 500–510, Cham. Springer Nature Switzerland.
- Timon Ziegenbein, Shahbaz Syed, Felix Lange, Martin Potthast, and Henning Wachsmuth. 2023. [Modeling appropriate language in argumentation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4344–4363, Toronto, Canada. Association for Computational Linguistics.