A Corpus of eRulemaking User Comments for Measuring Evaluability of Arguments

Joonsuk Park and Claire Cardie
Department of Computer Science, Williams College, Massachusetts, USA
Department of Computer Science, Cornell University, New York, USA
jpark@cs.williams.edu, cardie@cs.cornell.edu

Abstract

eRulemaking is a means for government agencies to directly reach citizens to solicit their opinions and experiences regarding newly proposed rules. The effort, however, is partly hampered by citizens' comments that lack reasoning and evidence, which are largely ignored since government agencies are unable to evaluate the validity and strength. We present *Cornell eRulemaking Corpus – CDCP*, an argument mining corpus annotated with argumentative structure information capturing the evaluability of arguments. The corpus consists of 731 user comments on Consumer Debt Collection Practices (CDCP) rule by the Consumer Financial Protection Bureau (CFPB); the resulting dataset contains 4931 elementary unit and 1221 support relation annotations. It is a resource for building argument mining systems that can not only extract arguments from unstructured text, but also identify what additional information is necessary for readers to understand and evaluate a given argument. Immediate applications include providing real-time feedback to commenters, specifying which types of support for which propositions can be added to construct better-formed arguments.

Keywords: argument mining, e-government, e-rulemaking, text analytics

1 Introduction

The U.S. federal agencies amend rules in a highly transparent manner, inviting public participation as they are finalized. This is legally ensured in part by the requirement that agencies publish descriptions and rationale behind newly proposed rules and solicit feedback from the public (Park et al., 2012; Farina and Newhart, 2013). However, the public participation tends to be dominated by large corporations and interest groups (CSFFR, 2009); *eRulemaking* is an ongoing effort to promote citizens' participation in federal policymaking by using the latest information technology to directly reach citizens and incorporate their feedback (Lubbers et al., 2012).

Government agencies consider reasoning and validity of supporting evidence, rather than a mere number of citizens supporting an argument, to determine how the rules should be adjusted to meet the needs of those who are directly affected. Thus, useful feedback consists of clear reasoning and objective evidence supporting factual claims (Park et al., 2015). However, many comments are not written this way, thwarting the government agencies' effort to communicate with citizens.

Consider the following comments from `www.regulationroom.org`, an eRulemaking website:

> (1) \$400 is enough compensation,$_A$ as it can cover a one-way fare across the US.$_B$ I checked in a passenger on a \$98.00 fare from east coast to Las Vegas the other day.$_C$
>
> (2) All airfare costs should include the passenger's right to check at least one standard piece of baggage.$_A$ All fees should be fully disclosed at the time of airfare purchase, regardless of nature (i.e. optional or mandatory).$_B$ Any changes in fees should be identified by air carriers at least 6 months prior to taking effect.$_C$

Comment 1 consists of propositions in support relations that collectively form a single argument: Proposition 1.C is an anecdotal evidence supporting Proposition 1.B, which in turn is a reason explaining why Proposition 1.A is true. Readers are able to make sense of the argument and evaluate its validity and strength, because each proposition is accompanied with a support of an appropriate type. (Figure 1 shows a sample annotation capturing the above discussion; see Section 3 for more details on the types of support and when they are appropriate.) In contrast, the propositions in Comment 2 are in no support relation with one another. In fact, each proposition functions as the conclusion of its own argument, where each argument contains no support for its conclusion. This renders it difficult for readers to understand the arguments, let alone evaluate them. Thus, Comment 1 is much more desirable for readers, whether it be government agencies or fellow citizens.

The aforementioned difference between the arguments

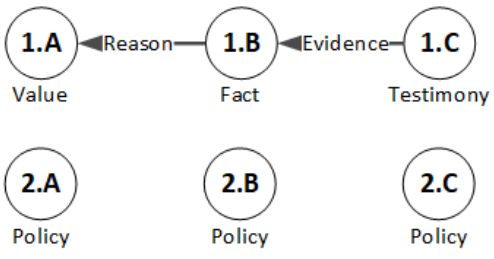

Figure 1: Annotated Example Comments

made in Comments 1 and 2 is captured by the notion of *evaluability* of argument proposed by Park et al. (2015)—are the propositions comprising a given argument adequately supported so as for readers to understand and evaluate the validity or strength of the whole argument?

We present *Cornell eRulemaking Corpus – CDCP*, an argument mining corpus annotated with argumentative structure information capturing the evaluability of arguments. We annotated 731 user comments on Consumer Debt Collection Practices (CDCP) rule by the Consumer Financial Protection Bureau (CFPB) posted on `www.regulationroom.org`; the resulting dataset contains 4931 elementary unit and 1221 support relation annotations. It will be a valuable resource for building argument mining systems that can not only extract arguments from unstructured text, but also identify ways in which a given argument can be improved with respect to its evaluability. Immediate applications include automatically ranking arguments based on their evaluability for a (crude) identification of read-worthy comments and providing real-time feedback to writers, specifying which types of support for which propositions can be added to construct better-formed arguments.

The remainder of this paper is organized as follows: We discuss related work (Section 2), provide an overview of the annotation scheme (Section 3), present an annotation study (Section 4) and describe the resulting dataset (Section 5).

2 Related Work

This paper presents a corpus for the purpose of mining and evaluating arguments in eRulemaking user comments. It is closely related to two areas of research: argument mining and argument quality assessment.

2.1 Argument Mining

Argument mining is a developing field of computational linguistics that aims at identifying argumentative structures in unstructured text. Extracting claims together with their respective premises allows us to go beyond opinion mining by considering the reasoning and rationale behind people's opinions. (Peldszus and Stede, 2013; Lippi and Torroni, 2016) Argument mining systems build on theoretical models of argument, which define argumentative components and their relations in a variety of ways. Famous models include the Toulmin Model (Toulmin, 1958) and argument schemes (Walton et al., 2008). The Toulmin Model is a general model of practical argumentation that can be instantiated in many forms. The three major components of the model are *claim*, *warrant*, and *data*, where *warrant* explains how *data* supports the *claim*. One criticism, which in turn make it challenging to build an argument mining system base on this model, is that the model leaves room for multiple interpretations. For example, according to Eemeren et al. (1987), *warrant* is indistinguishable from *data*. On the other hand, argument schemes capture specific patterns of argument that are in use; each argument scheme specifies specific premises for the given conclusion, as well as critical questions that can be used to examine the strength of the given argument (Walton, 1996; Blair, 2001). Having many specific premises, a subset of which may not be present in the text, makes it difficult for manual annotation and automatic classification. The sheer number of argument schemes also causes additional challenges in gathering enough examples for each scheme. In this work, we adopt a model uniquely designed to capture the evaluability of arguments, which is general enough to model diverse argumentative structures that appear in practical argumentation (Park et al., 2015).

Argument mining systems also differ in the domain, resulting in datasets consisting of newspaper articles (Reed et al., 2008), legal documents (Mochales and Moens, 2011), student essays (Stab and Gurevych, 2014), and eRulemaking user comments (Park and Cardie, 2014; Konat et al., 2016), to name a few. While ours is not the first eRulemaking dataset, the task is different; Park and Cardie (2014) targets elementary unit classification only, and Konat et al. (2016) focuses on identifying divisive issues between commenters by analyzing conflict relations found across multiple comments in a thread. In contrast, we examine support structures within a comment; our dataset contains both elementary unit and support relation annotation without cross-comment conflict annotation. Also, the user comments comprising our dataset are different from those in the aforementioned datasets.

2.2 Argument Quality Assessment

Measuring the quality of argument has long been a subject of discussion and research, leading to a variety of dimensions of quality (Toulmin, 1958; Perelman et al., 1969; van Eemeren and Grootendorst, 2004; Johnson and Blair, 2006; Wachsmuth et al., 2017). More recently, argument mining research is conducted with specific measures of quality depending on the domain and purpose, such as persuasiveness (Tan et al., 2016), strength (Persing and Ng, 2015), acceptability (Cabrio and Villata, 2012), and convincingness (Habernal and Gurevych, 2016). The measure of quality we are interested in is evaluability (Park et al., 2015). By examining arguments' evaluability, we aim to identify ways to improve them so that they can be better understood and evaluated. For example, we answer questions like, "Which propositions need additional reasons or evidence supporting them?" This is the type of constructive feedback that can help commenter improve their arguments, unlike quality measures that results in a single numeric score without specifying how an argument can be improved.

3 Annotation Scheme

The annotators annotated the elementary units and support relations defined in the argumentation model proposed by Park et al. (2015). In this section, we provide a brief overview of the model; please refer to the original paper for more details.

The goal of the model is to capture whether an argument consists of explicitly stated premises that allow readers to understand and evaluate the given argument. The model defines five types of elementary units that are prevalent in online comments, along with two types of support relations between the units.

3.1 Elementary Units

Proposition of Non-Experiential Fact (FACT) : This refers to an objective proposition "expressing or dealing with facts or conditions as perceived without distortion by personal feelings, prejudices, or interpretations."[1] By definition a FACT proposition has a truth value that can be verified with objective evidence. We restrict the notion of verifiability to pieces of evidence that may be available at the time the claim is made; predictions about future are considered unverifiable. Here are examples from the dataset:

- Recently, courts have held that debt collectors can escape 1692i's venue provisions entirely by pursuing debt collection through arbitration instead.
- banks can simply add this provision to their Loan Sale Agreements.
- That process usually takes as much a 2 years or more.

Proposition of Experiential Fact (TESTIMONY) : This refers to an objective proposition about the author's personal state or experience. One major characteristic of this type of objective propositions, as opposed to the non-experiential counterparts classified as FACT, is that it is often practically impossible to provide objective evidence in online commenting setting, in the form of URL or citation. That is, evidence for TESTIMONY is not publicly available in most cases. For example:

- Informing them that we wanted all debt collection to be written was also ignored.
- A neighbor who has since moved away has had her debts turned over to collection agencies.
- We receive repeated calls trying to get contact information, even though we request to be taken off their list.

Proposition of Value (VALUE) : This refers to a proposition containing value judgments without making specific claims about what should be done (If so, then it is a POLICY proposition.). Because of the subjectivity of value judgments, a VALUE proposition cannot be proved directly with objective evidence; however, providing a reason as support is feasible and appropriate. For example:

- That would be a starting point that can be expanded on as the system is fine tuned.
- Admittedly, their system is much more complex and dives much deeper than would be required for the debt industry.
- However, the double penalty against the consumer is certainly unfair.

Proposition of Policy (POLICY) : This refers to a proposition proposing a specific course of action to be taken. It typically contains modal verbs like "should" and "ought to." Just like VALUE, a POLICY proposition cannot be directly proved with objective evidence, and a proper type of support is a logical reason from which the proposition can be inferred. For example:

- They should not be allowed to contact anyone (other than the debtor him/herself) more than once.
- I say there ought to be sanctions, monetary sanctions, against these credit reporting agencies for making these mistakes and their cavalier attitude.
- Set up a system where the consumer is on equal footing with the debt collectors.

Reference to a Resource (REFERENCE) : This refers to a reference to a source of objective evidence. In online comments, a REFERENCE is typically a citation or a URL of a published work from a renowned source.For example:

- http://files.consumerfinance.gov/f/201309_cfpb_agency-brief_12-cv-04057.pdf
- http://www.myfico.com/CreditEducation/ImproveYourScore.aspx
- www.optoutprescreen.com

3.2 Support Relations

Reason : An elementary unit X is a *reason* for a proposition Y (of type POLICY, VALUE, FACT, or TESTIMONY) if X provides rationale for Y. For example:

- **Y:** I urge the CFPB to include in a rule language interpreting 1692i as requiring debt collectors to proceed in court, not through largely-unregulated arbitral forums. **X:** As the NAF studies reflect, arbitration has not proven a satisfactory alternative.

Evidence : An elementary unit X is *evidence* for a proposition Y (of type POLICY, VALUE, FACT, or TESTIMONY) if it proves whether proposition Y is true or not. The possible types of evidence are limited to TESTIMONY or REFERENCE based on previous studies on what constitutes justified grounds (Toulmin and Janik, 1979; Hitchcock, 2005). For example:

- **Y:** At least in Illinois there is a Caller ID spoofing law. **X:** http://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=1355ChapterID=24

[1] http://www.merriam-webster.com/

3.3 Evaluability

An argument is *evaluable* if all propositions comprising the given argument is supported by an explicit premise of an appropriate type, as summarized in Table 1. The underlying assumption is that readers are able to understand the gist of an argument—and at least roughly evaluate its strength—as long as one premise of an appropriate type is explicitly stated for each proposition.[2]

Once elementary units and support relations comprising an argument are identified, the evaluability of the given argument can be determined. This is done by comparing the appropriate types of support and the types of support present in the argument, if any. In the process, additional support that is necessary to make the given argument evaluable (e.g."A reason for proposition X needs to be provided.") can also be identified.

Proposition Type	POLICY	VALUE	FACT	TESTIMONY
Reason	✓	✓	✓	✓*
Evidence			✓	✓*

Table 1: Appropriate Support Types for Propositions
* Support can be provided, but it is not required.

4 Annotation Study

We annotated user comments on the Consumer Debt Collection Practices (CDCP) rule. The discussion regarding CDCP rule was hosted on `www.regulationroom.org` with a partnership with the CFPB. The goal was for the CFPB to hear about the first-hand experiences and concerns regarding debt collection practices. According to a voluntary user survey that asked the commenters to self-identify themselves, about 64% of the comments came from consumers, 22% from debt collectors, and the remainder from others, such as consumer advocates and counsellor organizations (Farina et al., 2017).

Each user comment was annotated by two annotators, who independently determined the types of elementary units and support relations among them using the GATE annotation tool (Cunningham et al., 2011). A third annotator manually resolved the conflicts to produce the final dataset.

An elementary unit is either a sentence or a clause; a sentence is split into smaller units if there are multiple independent clauses or an independent clause with a subordinate clause of interest, such as a because-clause. Non-argumentative portions of comments, such as greetings and names, were removed as elementary unit boundaries are determined in this way.

Inter-annotator agreement between 2 annotators is measured with Krippendorf's α (Krippendorff, 1980) with respect to elementary unit type (α=64.8%) and support relations (α=44.1%); IDs of supported elementary units are treated as labels for the supporting elementary units.[3]

The disagreements in elementary unit type annotation mostly occurred between VALUE vs TESTIMONY and VALUE vs FACT. The former is the case when a testimony spans multiple propositions and a few of them are subjective opinions about the experience. The latter often happens with an elementary unit that contains both subjective and objective expressions, e.g. "Unfortunate, but yes they are allowed to deny due process and get away with it." In this case, annotators had to determine the commenter's main intention—is it to express the emotion or state the fact? Depending on the answer, the given elementary unit was either marked as VALUE or FACT. (Allowing a more granular boundaries for elementary units can solve this type of disagreement; however, an undesirable effect of this is that automatic segmentation becomes more challenging.)

5 Dataset

The resulting dataset, *Cornell eRulemaking Corpus – CDCP*, consists of 731 comments, 4931 elementary units, and 1221 support relations as summarized in Table 2. About 45% of the elementary units are VALUE type, and most support relations are reasons. Table 3 describes annotated information in this dataset.

Figure 2 shows the types of supported elementary units and those of supporting elementary units. The percentage of supported elementary units decreases as the elementary unit's objectivity goes from the least objective (POLICY) to the most objective (REFERENCE). One reason is that it is easier to provide a reason as to why one thinks or feels something (POLICY and VALUE) than to justify factual propositions (FACT and TESTIMONY). Interestingly, even though both POLICY and VALUE are subjective, there is a notable difference in the support pattern; 51% of POLICY propositions are supported, whereas only 28% of VALUE propositions are supported. This means that when commenters propose a specific course of actions to be taken, they are more likely to provide support for it. This is because POLICY propositions are often the central claims of the comments, thus other propositions naturally support them. Also, unlike VALUE, a simple expression of one's thoughts and feelings, POLICY, a proposal to act in a certain way, is associated with persuasion, which benefits from explicitly stated reasoning.

A significant portion, roughly 75%, of support relation annotations are between adjacent elementary units. While commenters certainly tend to provide reasons immediately after the proposition to be supported, it is also easier for annotators to identify support relations in proximity. Thus, support relations in the wild may be not as skewed toward

[2] Please refer to Park et al. (2015) for formal definitions.

[3] Krippendorf's α is suitable for our purpose as it is compatible with various types of labeling, along with the ability to handle missing annotations.

POLICY	VALUE	FACT	TESTIMONY	REFERENCE	Elementary Units	Reason	Evidence	Support Relations
815	2182	785	1117	32	4931	1174	46	1220

Table 2: Number of Elementary Units and Support Relations in the Dataset (731 comments)

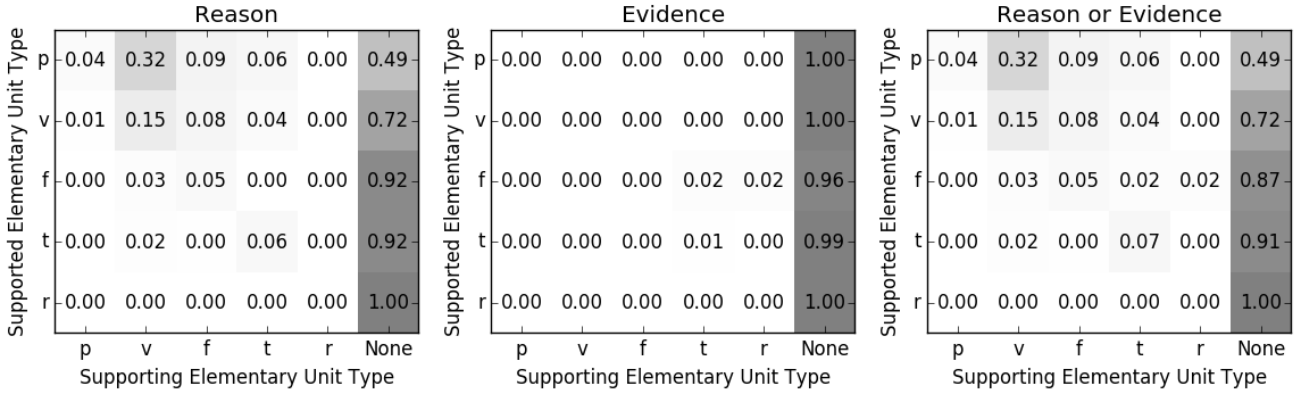

Figure 2: Types of Elementary Units in Support Relations (%)

Field	Description
ID	ID of the elementary unit
Text	Text of the elementary unit
Type	POLICY, VALUE, FACT, TESTIMONY or REFERENCE
Reasons	List of elementary unit IDs serving as reasons
Evidence	List of elementary unit IDs serving as evidence

Table 3: Annotated Information: Each comment annotation consists of a list of elementary units in the given comment with fields described in this table.

those between adjacent elementary units.

6 Conclusion

We have presented *Cornell eRulemaking Corpus – CDCP*, an argument mining corpus annotated with argumentative structure information capturing the evaluability of arguments. The corpus consists of 731 user comments on Consumer Debt Collection Practices (CDCP) rule by by the Consumer Financial Protection Bureau (CFPB) posted on `www.regulationroom.org`; the resulting dataset consists of 4931 elementary unit and 1221 support relation annotations. It will be a valuable resource for building argument mining systems that can not only extract arguments from unstructured text, but also identify which additional information is necessary for readers to understand and evaluate a given argument.

Future work includes: (1) construction of a larger corpus using the same or similar annotation scheme and (2) making use of the resources to train argument mining systems (Niculae et al., 2017) and subsequent applications, such as a commenting interface that provides real-time feedback to help commenters construct evaluable arguments. Domain adaptation is also desirable, since building an argument mining dataset for individual domains incurs a significant cost.

7 Bibliographical References

Blair, J. A. (2001). Walton's argumentation schemes for presumptive reasoning: A critique and development. *Argumentation*, 15(4):365–379.

Cabrio, E. and Villata, S. (2012). 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, Jeju Island, Korea, July. Association for Computational Linguistics.

CSFFR. (2009). Achieving the potential:the future of federal e-rulemaking. Technical report, Committee on the Status & Future of Federal e-Rulemaking/American Bar Association, Washington, DC.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., Heitz, T., Greenwood, M. A., Saggion, H., Petrak, J., Li, Y., and Peters, W. (2011). *Text Processing with GATE (Version 6)*.

Farina, C. R. and Newhart, M. J. (2013). Rulemaking 2.0: Understanding and getting better public participation.

Farina, C. R., Blake, C. L., Newhart, M. J., and Nam, C. (2017). Digital support for enhanced democratic participation in us rulemaking. In C. Prins, et al., editors, *Digital Democracy in a Globalized World*, chapter 10. Edward Elgar Publishing.

Habernal, I. and Gurevych, I. (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 (ACL 2016)*, volume Volume 1: Long Papers, page (to appear). Association for Computational Linguistics, August.

Hitchcock, D. (2005). Good reasoning on the toulmin model. *Argumentation*, 19(3):373–391.

Johnson, R. and Blair, J. (2006). *Logical Self-defense*. Key titles in rhetoric, argumentation, and debate series. International Debate Education Association.

Konat, B., Lawrence, J., Park, J., Budzynska, K., and Reed, C. (2016). A corpus of argument networks: Using graph properties to analyse divisive issues. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Sage commtext series. Sage Publications.

Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Technol.*, 16(2):10:1–10:25, March.

Lubbers, J., of Administrative Law, A. B. A. S., Practice, R., Government, A. B. A., and Division, P. S. L. (2012). *A Guide to Federal Agency Rulemaking*. ABA Section of Administrative Law and Regulatory Practice and Government and Public Sector Lawyers Division.

Mochales, R. and Moens, M.-F. (2011). Argumentation mining. *Artif. Intell. Law*, 19(1):1–22, March.

Niculae, V., Park, J., and Cardie, C. (2017). Argument mining with structured svms and rnns. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 985–995. Association for Computational Linguistics.

Park, J. and Cardie, C. (2014). Identifying appropriate support for propositions in online user comments. In *Proceedings of the First Workshop on Argumentation Mining*, pages 29–38, Baltimore, Maryland, June. Association for Computational Linguistics.

Park, J., Klingel, S., Cardie, C., Newhart, M., Farina, C., and Vallbé, J.-J. (2012). Facilitative moderation for online participation in erulemaking. In *Proceedings of the 13th Annual International Conference on Digital Government Research*, pages 173–182. ACM.

Park, J., Blake, C., and Cardie, C. (2015). Toward machine-assisted participation in erulemaking: An argumentation model of evaluability. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, ICAIL '15, pages 206–210, New York, NY, USA. ACM.

Peldszus, A. and Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *Int. J. Cogn. Inform. Nat. Intell.*, 7(1):1–31, January.

Perelman, C., Olbrechts-tyteca, L., Wilkinson, J., and Weaver, P. (1969). *The New Rhetoric*. University of Notre Dame P X.

Persing, I. and Ng, V. (2015). 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 (Volume 1: Long Papers)*, pages 543–552.

Reed, C., Palau, R. M., Rowe, G., and Moens, M.-F. (2008). Language resources for studying argument. In *LREC*. European Language Resources Association.

Stab, C. and Gurevych, I. (2014). Annotating argument components and relations in persuasive essays. In Junichi Tsujii et al., editors, *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, pages 1501–1510, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., and Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 613–624, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Toulmin, S.E., R. R. and Janik, A. (1979). *An Introduction to Reasoning*. Macmillan Publishing Company.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.

van Eemeren, F. and Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. A Systematic Theory of Argumentation: The Pragma-dialectical Approach. Cambridge University Press.

van Eemeren, F., Grootendorst, R., and Kruiger, T. (1987). *Handbook of argumentation theory ; a critical survey of classical backgrounds and modern studies*. PDA Series. Foris Publications.

Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., Hirst, G., and Stein, B. (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, April. Association for Computational Linguistics.

Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

Walton, D. (1996). *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates.