Argumentation Mining on the Web from Information Seeking Perspective

Ivan Habernal[†‡]**, Judith Eckle-Kohler**[†‡]**, Iryna Gurevych**[†‡]

[†]Ubiquitous Knowledge Processing Lab (UKP-TUDA)
Department of Computer Science, Technische Universität Darmstadt
[‡]Ubiquitous Knowledge Processing Lab (UKP-DIPF)
German Institute for Educational Research
`www.ukp.tu-darmstadt.de`

Abstract

In this paper, we argue that an annotation scheme for argumentation mining is a function of the task requirements and the corpus properties. There is no one-size-fits-all argumentation theory to be applied to realistic data on the Web. In two annotation studies, we experiment with 80 German newspaper editorials from the Web and about one thousand English documents from forums, comments, and blogs. Our example topics are taken from the educational domain.

To formalize the problem of annotating arguments, in the first case, we apply a Claim-Premise scheme, and in the second case, we modify Toulmin's scheme. We find that the choice of the argument components to be annotated strongly depends on the register, the length of the document, and inherently on the literary devices and structures used for expressing argumentation. We hope that these findings will facilitate the creation of reliably annotated argumentation corpora for a wide range of tasks and corpus types and will help to bridge the gap between argumentation theories and actual application needs.

1 Introduction

Argumentation mining apparently represents an emerging field in Natural Language Processing (NLP) with publications appearing at mainstream conferences, such as ACL (Cabrio and Villata, 2012; Feng and Hirst, 2011; Madnani et al., 2012) or COLING (Stab and Gurevych, 2014; Levy et al., 2014; Wachsmuth et al., 2014a). In particular, there is an increasing need for tools capable of understanding argumentation on the large scale, because in the current information overload, humans cannot feasibly process such massive amounts of data in order to reveal argumentation. Unfortunately, even current Web technologies (such as search engines or opinion mining services) are not suitable for such a task. This drives the research field to the next challenge – argumentation mining on the Web. The abundance of freely available (yet unstructured, textual) data and possible applications of such tools makes this task very appealing.

Our research into argumentation mining is motivated by the *information seeking perspective*. The key sources are discussions (debates) about controversies (contentions) targeted at a particular topic which is of the user's interest. The scope is not limited to a particular media type as the source types can range from the on-line newspapers' editorials to user-generated discourse in social media, such as blogs and forum posts, covering different aspects of the issues. Understanding positions and argumentation in on-line debates helps users to form their opinions on controversial issues and also fosters personal and group decision making (Freeley and Steinberg, 2008, p. 9). The main task would be to identify and extract the core argumentation (its formal aspects will be discussed later) and present this new knowledge to users. By utilizing argumentation mining methods, users can be provided with the most relevant information (arguments) regarding the controversy under investigation.

Although argumentation mining on the Web has already been partly outlined in the literature (Schneider et al., 2012; Sergeant, 2013), the requirements and use-case scenarios differ substantially. Various tasks are being solved, most of them depending on the domain, e.g., product reviews or political contentions. As a result, different interpretations of arguments and argumentation have been developed in NLP, and therefore, most of the existing researches are not directly adaptable.
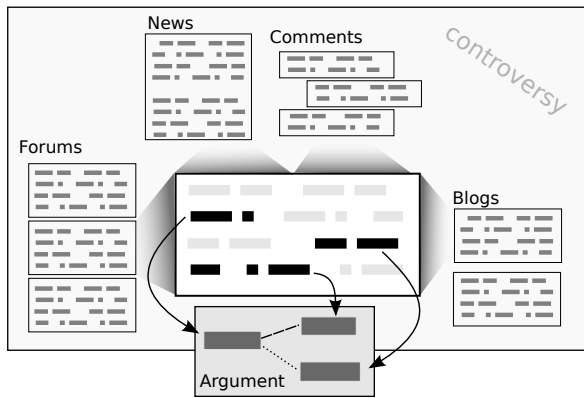
Figure 1: Schematic overview of argumentation mining on the Web

Morover, not all of the related research works are tightly connected to argumentation theories (de Moor et al., 2004; Villalba and Saint-Dizier, 2012; Cabrio et al., 2013b; Llewellyn et al., 2014). However, we feel that it is vital to ground NLP research in argumentation mining in existing work on argumentation.

In this article, we will particularly focus on bridging the gap between argumentation theories and actual application needs that has not been targeted in the relevant literature. We will support our findings by comprehensively surveying existing works and presenting results from two extensive annotation studies.

Our main findings and suggestions can be summarized as follows: First, the use-case of any research in argumentation mining must be clearly stated (i.e., in terms of expected outcomes). Second, properties of the data under investigation must be taken into account, given the variety of genres and registers (Biber and Conrad, 2009). Third, an appropriate argumentation model must be chosen according to the requirements. Therefore, we claim that it is not possible to formulate a single argumentation mining perspective that would be applicable to the Web data in general.

2 Relation to Argumentation Theories

Research on argumentation is widely interdisciplinary, as it spreads across philosophy and rhetoric (Aristotle and Kennedy (translator), 1991; Perelman and Olbrechts-Tyteca, 1991; Walton et al., 2008), informal and formal logic (Dung, 1995; Henkemans, 2000; Stoianovici, 2009; Schneider et al., 2013; Hunter, 2013), educational research (Weinberger and Fischer, 2006;

Noroozi et al., 2013), pragmatics (Xu and Wu, 2014), psychology (Larson et al., 2004), and many others. Given so many different perspectives on investigating argumentation, there is a plethora of possible interpretations of argumentation mining. Thus, finding a common understanding of this evolving field is a fundamental challenge.

For NLP research, this overwhelming amount of related works brings many theoretical and practical issues. In particular, there is no one-size-fits-all argumentation theory. Even argumentation researchers disagree on any widely-accepted ultimate concept. For example, Luque (2011) criticizes the major existing approaches in order to establish a new theory which is later again severely criticized by other in-field researches (Andone, 2012; Xie, 2012). Given this diversity of perspectives, NLP research cannot simply adopt one particular approach without investigating its theoretical background as well as its suitability for the particular task.

2.1 What we do not tackle

Given the breath of argumentation mining just outlined, we would also like to discuss aspects that do not fit into our approach to argumentation mining, namely macro argumentation and evaluation using formal frameworks.

First, we treat argumentation as a product (micro argumentation or monological models), not as a process (macro argumentation or dialogical models). While dialogical models highlight the process of argumentation in a dialogue structure, monological models emphasize the structure of the argument itself (Bentahar et al., 2010, p. 215). Therefore, we examine the relationships between the different components of a given argument, not a relationship that can exist between arguments.[1] Exploring how argumentation evolves between parties in time remains out of our scope.

Second, we do not tackle any logical reasoning, defeasibility of reasoning, or evaluating argumentation with formal frameworks in general. Although this is an established field in informal logic (Prakken, 2010; Hunter, 2013; Hunter, 2014), such an approach might not be suitable directly for Web data as it assumes that argumentation is logical (such a strong assumption cannot be guar-

---

[1]For further discussion see, e.g., (Blair, 2004; Johnson, 2000; Reed and Walton, 2003) or Micheli (2011) who summarizes the distinction between the process (at a pragmatic level) and the product (at a more textual level).

anteed). Furthermore, acceptability of arguments also touches the fundamental problem of the target audience of the argument, as different groups have different perceptions. Crosswhite et al. (2004) point out that "one of the key premises from which the study of rhetoric proceeds is that influencing real audiences is not simply a matter of presenting a set of rational, deductive arguments."

2.2 Common terminology

Let us set up a common terminology. *Claim* is "the conclusion we seek to establish by our arguments" (Freeley and Steinberg, 2008, p. 153) or "the assertion put forward publicly for general acceptance" (Toulmin et al., 1984, p. 29). *Premises* are "connected series of sentences, statements, or propositions that are intended to give reasons of some kind for the claim" (Freeley and Steinberg, 2008, p. 3).

3 Related Work

3.1 Opinion mining perspective

In existing works on argumentation mining of the Web data, the connection is often made to opinion mining (Liu, 2012). From the users' point of view, opinion mining applications reveal *what people think about something.* The key question which brings argumentation on the scene is *why do they think so?* – in other words, explaining the reasons behind opinions.

Villalba and Saint-Dizier (2012) approach aspect-based sentiment of product reviews by classifying discourse relations conveying arguments (such as justification, reformulation, illustration, and others). They build upon Rhetorical Structure Theory (RST) (Mann and Thompson, 1987) and argue that rhetorical elements related to explanation behave as argument supports.

For modeling argumentation in social media, Schneider et al. (2012) suggest using Dung's framework (Dung, 1995) with Walton schemes (Walton et al., 2008), but do not provide evidence for such a decision. They admit that "It is far from clear how an argument [...] can be transformed into a formal argumentation scheme so that it can be reasoned in an argumentation framework" (Schneider et al., 2012, p. 22).

Schneider and Wyner (2012) focus on the product reviews domain and develops a number of argumentation schemes (inspired by (Walton et al., 2008)) based on manual inspection of their cor-

pus. Appropriateness of such an approach remains questionable. On the one hand, Walton's argumentation schemes are claimed to be general and domain independent. On the other hand, evidence from the field shows that schemes might not be the best means for analyzing user-generated argumentation. In examining real-world political argumentation from (Walton, 2005), Walton (2012) found out that 37.1% of the arguments collected did not fit any of the fourteen schemes they chose so they created new schemes ad-hoc. Cabrio et al. (2013a) select five argumentation schemes from Walton and map these patterns to discourse relation categories in the Penn Discourse TreeBank (PDTB) (Prasad et al., 2008), but later they define two new schemes that they discovered in PDTB. These findings confirm that the schemes lack coverage for dealing with real argumentation in natural language texts.

3.2 Previous works on annotation

Table 1 summarizes the previous research on annotating argumentation. Not only it covers related work from the NLP community but also studies from general discourse analysis (Newman and Marshall, 1991; Walton, 2012) and road-maps or position papers (Schneider and Wyner, 2012; Peldszus and Stede, 2013a; Sergeant, 2013). The heterogeneity of used argumentation models and the domains under investigation demonstrates the breath of the argumentation mining field. We identified the following research gaps.

- Most studies dealing with Web data use some kind of proprietary model without relation to any argumentation theory (Bal and Saint-Dizier, 2010; Rosenthal and McKeown, 2012; Conrad et al., 2012; Schneider and Wyner, 2012; Villalba and Saint-Dizier, 2012; Florou et al., 2013; Sergeant, 2013; Wachsmuth et al., 2014b; Llewellyn et al., 2014).

- Inter-annotation agreement (IAA) that reflects reliability of the annotated data is either not reported (Feng and Hirst, 2011; Mochales and Moens, 2011; Walton, 2012; Florou et al., 2013; Villalba and Saint-Dizier, 2012), or is not based on a chance-corrected measure (Llewellyn et al., 2014).

This motivates our research into annotating Web data relying on a model based on a theoretical
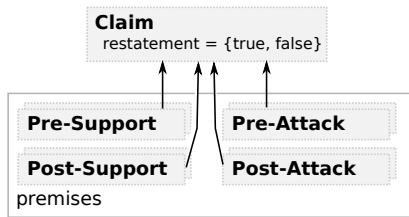
Figure 2: Claim-Premise scheme. Note that the relations (arrows) are only illustrative; they are implicitly encoded in the roles of the particular argument components.

background in argumentation and reporting IAA that would confirm suitability of the model and reliability of the annotated data.

4 Annotating argumentation in Web data

Up until now, we have used the terms argumentation and argument in their common meaning without any particular formal definition. We will now elaborate on annotation schemes and discuss their suitability and reliability for the Web data.

4.1 Annotation Schemes

Because of the lack of a single general-purpose argumentation model (cf. discussion in §1), we present here two different schemes.[2] Both are built upon foundations in argumentation theories, but they differ in their granularity, expression power, and other properties.

4.1.1 Claim-Premises scheme

The Claim-Premises scheme is widely used in previous work on argumentation mining, e.g., (Palau and Moens, 2009; Florou et al., 2013; Peldszus and Stede, 2013b). It defines an argument as consisting of a (possibly empty) set of premises and a single claim; premises either support or attack the claim (Besnard and Hunter, 2008). We adopted this general scheme for the purpose of annotating arguments in long Web documents (Kluge, 2014). According to this adopted version of the scheme, claims, restatements and premises are subsumed under the term argument component; a restatement of a claim is also considered as claim and is part of the same argument. The scheme is depicted in Figure 2.

Premises either support or attack a claim, i.e., there is a support or attack relation between each

premise and a claim. The simplest way to represent the support and attack relations is to attach labels to adjacent argument components, which indicate their argumentative role. The span of argument components is left unspecified, allowing for argument components spanning a clause or one to several sentences. Using the six labels *claim, restatement, pre-claim support, post-claim support, pre-claim attack and post-claim attack*, a linear sequence of non-nested arguments can be represented.

While graph structures where nodes stand for argument components, and edges for support or attack relations are a more general way to represent arguments (equivalent to, i.e., (Dung, 1995) or (Freeman, 1991)), it is unclear which additional benefits such a more fine-grained annotation of arguments brings for the annotation of Web documents. In a pre-study performed by Kluge (2014), the possibility to annotate nested arguments turned out to be a drawback, rather than an advantage, because the inter-annotator agreement dropped considerably.

Suitability of the scheme The main advantage of the Claim-Premises scheme is its simplicity. Therefore, it is particularly suited for annotating arguments in long Web documents, such as news articles, editorials or blog posts. Kluge (2014) found that most documents of these text types consist of three major parts: an introductory part, summarizing the document content in one or two paragraphs, the main part, presenting a linear sequence of arguments, and an optional concluding part summarizing the main arguments.

The Claim-Premise scheme can be used to provide an overview of the claims and their supporting or attacking premises presented in a long Web document. From an information seeking perspective, arguments could be clustered by similar claims or similar premises, and then ranked in the context of a specific information need by a user. In a similar way, this scheme could be used for automatic summarization.

However, the Claim-Premises scheme does not allow to distinguish between different kinds of premises supporting the claim. Hence, fine-grained distinctions of premises into specific factual evidence versus any kind of common ground can not be captured.

---

[2]An exhaustive overview of various argumentation models, their taxonomy, and properties can be found in (Bentahar et al., 2010).

Source	Arg. Model	Domain	Size	IAA
Newman and Marshall (1991)	Toulmin	legal domain (People vs. Carney, U.S. Supreme Court)	qualitative	N/A
Bal and Saint-Dizier (2010)	proprietary	socio-political newspaper editorials	56 documents	Cohen's κ (0.80)
Feng and Hirst (2011)	Walton (top 5 schemes)	legal domain (Aracuraria corpus, 61% subset annotated with Walton scheme)	$\approx$ 400 arguments	not reported claimed to be small
Georgila et al. (2011)	proprietary	general discussions (negotiations between florists)	21 dialogues	Krippendorf's α (0.37-0.56)
Mochales and Moens (2011)	Claim-Premise based on Freeman	legal domain (Aracuraria corpus, European Human Rights Council)	641 documents w/ 641 arguments (Aracuraria) 67 documents w/ 257 arguments (EHRC)	not reported
Walton (2012)	Walton (14 schemes)	political argumentation	256 arguments	not reported
Rosenthal and McKeown (2012)	opinionated claim, sentence level	blogposts, Wikipedia discussions	4000 sentences	Cohen's κ (0.50-0.57)
Conrad et al. (2012)	proprietary (spans of arguing subjectivity)	editorials and blogpost about Obama Care	84 documents	Cohen's κ (0.68) on 10 documents
Schneider and Wyner (2012)	proprietary, argumentation schemes	camera reviews	N/A (proposal/position paper)	N/A
Schneider et al. (2012)	Dung + Walton	unspecified social media	N/A (proposal/position paper)	N/A
Villalba and Saint-Dizier (2012)	proprietary, RST	hotel reviews, hi-fi products, political campaign	50 documents	not reported
Peldszus and Stede (2013a)	Freeman + RST	Potsdam Commentary Corpus	N/A (proposal/position paper)	N/A
Florou et al. (2013)	none	public policy making	69 argumentative segments / 322 non-argumentative segments	not reported
Peldszus and Stede (2013b)	based on Freeman	not reported, artificial documents created for the study	23 short documents	Fleiss' κ multiple results
Sergeant (2013)	N/A	Car Review Corpus (CRC)	N/A (proposal/position paper)	N/A
Wachsmuth et al. (2014b)	none	hotel reviews	2100 reviews	Fleiss' κ (0.67)
Llewellyn et al. (2014)	proprietary, no argumentation theory	Riot Twitter Corpus	7729 tweets	only percentage agreement reported
Stab and Gurevych (2014)	Claim-Premise based on Freeman	student essays	90 documents	Krippendorf's α_U (0.72) Krippendorf's α (0.81)

Table 1: Previous works on annotating argumentation. IAA = Inter-annotation agreement; N/A = not applicable.
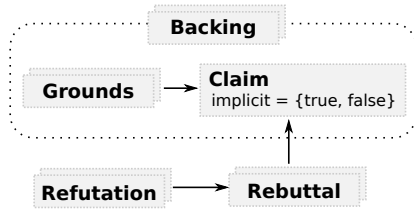
Figure 3: Extended Toulmin's scheme. Note that the relations (arrows) are only illustrative; they are implicitly encoded in the roles of the particular argument components.

4.1.2 Toulmin's scheme

The Toulmin's model (Toulmin, 1958) is a conceptual model of argumentation, in which different components play distinct roles. In the original form, it consists of six components: *claim*, *data (grounds)*, *warrant*, *backing*, *qualifier*, and *rebuttal*.

The roles of *claim* and *grounds* correspond to the definitions introduced earlier (*claim* and *premises*, respectively). The role of *warrant* is to justify a logical inference from *grounds* to *claim*. To assure the trustworthiness of the *warrant*, *backing* provides further set of information. *Qualifier* limits the degree of certainty under which the argument should be accepted and *rebuttal* presents a situation in which the *claim* might be defeated. For examples of arguments based on Toulmin's original model see, e.g., (Freeley and Steinberg, 2008, Chap. 8).

Based on our experiments during annotation pre-studies, we propose an extension of the Toulmin's model by means of (1) omitting the *qualifier* for stating modality, as people usually do not state the degree of cogency, (2) omitting the *warrant* as reasoning for justifying the move from grounds to claims is not usually explained, (3) extending the role of *backing* so it provides additional set of information to back-up the argument as a whole but is not directly bound to the *claim* as the *grounds* are, and (4) adding *refutation* which attacks the *rebuttal* (attacking the attack). The scheme is depicted in Figure 3.

Suitability of the scheme As pointed out by Bentahar et al. (2010), many argumentation systems make no distinction between their premises, despite the fact that in arguments expressed in natural language we can typically observe premises playing different roles. Toulmins' scheme allows such a distinction using the set of different com-

ponents (roles). "By identifying these roles, we can present the arguments in a more readily understandable fashion, and also identify the various ways in which the argument may be accepted or attacked" (Bentahar et al., 2010, p. 216).

Toulmin's model, as a general framework for modeling static monological argumentation (Bentahar et al., 2010), has been used in works on annotating argumentative discourse (Newman and Marshall, 1991; Chambliss, 1995; Simosi, 2003; Weinberger and Fischer, 2006). However, its complexity and the fact that the description of the components is informal and sometimes ambiguous, poses challenges for an application of the model on real-world data, especially user-generated discourse on the Web. Moreover, some of the components are usually left implicit in argumentation, such as the warrant or even the claim (Newman and Marshall, 1991).

5   Preliminary results of annotation studies

In order to examine the proposed approaches, we conducted two extensive independent annotation studies. The central controversial topics were related to education. One distinguishing feature of educational topics is their breadth, as they attract researchers, practitioners, parents, or policymakers. Since the detailed studies are being published elsewhere, we summarize only the main results and outcomes in this paper.

In the first study, we used the Claim-Premises scheme for annotating a dataset of web documents consisting of 80 documents from six current topics related to the German educational system (e.g., mainstreaming, staying down at school), which is described in (Kluge, 2014). The dataset contains (newspaper) articles, blog posts, and interviews. It was created by Vovk (2013) who manually selected documents obtained from a focused crawler and the top 100 search engine hits (per topic).

In the second study, the annotation was split into two stages. In the first stage, we annotated 990 English comments to articles and forums posts with their argumentativeness (persuasiveness). The source sites were identified using a standard search engine and the content was extracted manually; we chose the documents randomly without any pre-filtering. In the second stage, we applied the extended Toulmin's scheme on 294 argumentative English comments to arti-

cles and forums posts and 57 English newspaper editorials and blog posts. The topics cover, e.g., mainstreaming,[3] single-sex schools, or home-schooling, among others.

Measuring inter-annotator agreement For any real large-scale annotation attempt, measuring inter-annotator agreement (IAA) is crucial in order to estimate the reliability of annotations and the feasibility of the task itself. Both annotation approaches share one common sub-task: labeling spans of tokens with their corresponding argumentation concept, the boundaries of the spans are not known beforehand. Therefore, the most appropriate measure here is the unitized Krippendorf's α_U as the annotators identify and label the units in the same text (Krippendorff, 2013). Other measures, such as Cohen's κ or Fleiss' π, expect the units (boundaries of the argument component) to be known beforehand, which is not the case here.

5.1 Outcomes of annotating with Claim-Premises scheme

During an annotation study of 6 weeks, three annotators (one inexperienced annotator and two experts) annotated 80 documents belonging to six topics. On average, each annotator needed 23 hours to annotate the 3863 sentences. The annotators marked 5126 argument components (53% premises, 47% claims) and 2349 arguments, which is 2.2 argument components per argument. On average, 74% of the tokens in the dataset are covered by an argument component indicates that the documents are in fact highly argumentative. An average claim spans 1.1 sentences, whereas an average premise spans 2.2 sentences.

While the IAA scores appeared to be non-substantial, ranging from $\alpha_U{=}34.6$ (distinguishing all 6 annotation classes and non-argumentative) to $\alpha_U{=}42.4$ (distinguishing between premises, claims and non-argumentative), they are in line with previous results: Peldszus and Stede (2013b) report $\alpha_U{=}42.5$ for their sentence-level annotation study.

By analysing typical patterns of argument components used in arguments, Kluge (2014) found that almost three quarters of arguments (72.4%) consist of one claim and one premise. In 59.5% of these arguments, the support follows the claim,

---

[3]Discussion about benefits or disadvantages of including children with special needs into regular classes.

Argument Component	Comments, Forums	Blogs	Articles
Claim	0.57	0.17	0.23
Grounds	0.64	0.32	0.11
Backing	0.41	-0.16	0.28
Rebuttal	0.33	-0.02	0.00
Refutation	0.06	0.35	0.00

Table 2: IAA scores (Krippendorf's α_U) from annotations using the Toulmin's scheme.

whereas only in 11.6% of the arguments, the support precedes the claim. The corresponding patterns consisting of attack and claim are significantly less frequent: only 3.4% of the arguments consist of a claim and an attack.

Annotated examples can be found in §A.1.

5.2 Outcomes of annotating with Toulmin's scheme

In the first stage, three independent annotators labeled 524 out of 990 documents as argumentative/persuasive on the given topic. Total size of this dataset was 130,085 tokens (mean 131, std. dev. 139) and 6,371 sentences (mean 6.44, std. dev. 6.53). Agreement on the first sub-set of this dataset of 300 documents was 0.51 (Fleiss' π, three annotators per document), the second sub-set (690 documents) was then annotated by two annotators with agreement 0.59 (Cohen's κ). This stage took in total about 17 hours per annotator.

In the second phase that took about 33 hours per annotator, a collection of comments and forum posts (294 documents) was randomly chosen from the previously labeled argumentative documents from the previous stage together with 49 blog posts and 8 newspaper articles. The total size of this dataset was 345 documents, containing 87,286 tokens (mean 253.00, std. dev. 262.90) and 3,996 sentences (mean 11.58, std. dev. 11.72). Three independent annotators annotated the whole dataset in multiple phases. After each phase, they discussed discrepancies, resolved issues and updated the annotation guidelines. The inter-annotator agreement was measured on the last phase containing 93 comments and forum posts, 8 blogs, and 6 articles. During the annotations, 2 articles and 4 forum posts/comments were also discarded as non-argumentative.

Agreement (Krippendorf's α_U) varies significantly given different argumentation components

and registers, as shown in Table 2. Given these results, we formulate the following conclusions.

This scheme seems to fit well short documents (forum posts and comments) as they tend to bring up one central *claim* with a support (*grounds*). Its suitability for longer documents (blogposts and editorials) is doubtful. We examined the annotation errors and found that in well-structured documents, the annotators were able to identify the concepts reliably. However, if the discussion of the controversy is complex (many sub-aspects are discussed) or follows a dialogical manner, application the Toulmin's scheme is all but straightforward.

Furthermore, the distinction between *grounds* and *backing* also allows to capture different kinds of evidence. Authors purposely use *grounds* to explicitly support their *claim*, while *backing* mostly serves as an additional information (i.e., author's personal experience, referring to studies, etc.) and the argument can be still acceptable without it. However, boundaries between these two components are still fuzzy and caused many disagreements.

We show few annotation examples (as agreed by all annotators after the study) in §A.2.

6 Observations

In this section, we would like to summarize some important findings from our annotation studies.

6.1 Data heterogeneity

Variety or registers There exist many on-line registers that carry argumentation to topics under investigation, such as newspaper reports (i.e., events), editorials (opinions), interviews (single party, multiple parties), blogposts,[4] comments to articles and blogs (threaded allowing explicit discussion, linear with implicit discussion by quoting and referencing), discussion forums, Twitter, etc.

Short versus long documents Different document lengths affect the style of argumentation. Short documents (i.e., Tweets in the extreme case) have to focus on the core of the argument. By contrast, long documents, such as blog posts or editorials, may elaborate various aspects of the topic and usually employ many literary devices, such as

narratives, quotations from sources, or direct and indirect speech.

Well structured newspaper articles versus poorly structured user-generated content Producing a well-understandable argument is actually a human skill that can be acquired by learning; many textbooks are available on that topic, e.g., (Sinnott-Armstrong and Fogelin, 2009; Weston, 2008; Schiappa and Nordin, 2013). Thus, it is very likely that, for example, trained journalists in editorials and lay people in social media will produce very different argumentation, in terms of structure, language, etc.

6.2 Properties of argumentation in user-generated discourse

Non-argumentative texts Distinguishing argumentative from non-argumentative discourse is a necessary step that has to be undertaken before annotating argument components. While in newspaper editorials some parts (such as paragraphs) may be ignored during argument annotation (Kluge, 2014), in comments and forum posts we had to perform an additional step to filter documents that do not convey any argumentation or persuasion (cf. §5.2 or Example 4 in §A.2).

Implicit argumentation components in Toulmin's model As already reported by Newman and Marshall (1991), some argument components are not explicitly expressed. This is mostly the case of *warrant* in the original Toulmin's model; we also discarded this component from our extension. However, even the *claim* is often not stated explicitly, as seen in example 3 (§A.2). The claim reflects the author's stance and can be understood (inferred) by readers, but is left implicit.

Other rhetorical dimensions of argument All the models for argumentation discussed so far focus solely on the *logos* part of the argument. However, rhetorical power of argumentation also involves other dimensions, namely *pathos*, *ethos*, and *kairos* (Aristotle and Kennedy (translator), 1991; Schiappa and Nordin, 2013). These have never been tackled in computational approaches to modeling argumentation. Furthermore, figurative langauge, fallacies, or narratives (see example 3 in §A.2) are prevalent in argumentation on the Web.

---

[4]In contrast to traditional publisher, bloggers do not have to comply with strict guidelines or the use of formal language (Santos et al., 2012).

6.3 Recommendations

Based on the experience from the annotation studies, we would like to conclude with the following recommendations: (1) selection of argumentation model should be based on the data at hand and the desired application; our experiments show that Toulmin's model is more expressive than the Claim-Premise model but is not suitable for long documents, (2) annotating argumentation is time-demanding and error-prone endeavor; annotators thus have to be provided with detailed and elaborated annotation guidelines and be extensively trained (our experiments with crowdsourcing were not successful).

7 Follow-up use cases

Understanding argumentation in user-generated content can foster future research in many areas. Here we present two concrete applications.

7.1 Understanding argumentative discourse in education

Computer-supported argumentation has been a very active research field, as shown by Scheuer et al. (2010) in their recent survey of various models and argumentation formalisms from the educational perspective. Many studies on computer-supported collaboration and argumentation (Noroozi et al., 2013; Weinberger and Fischer, 2006; Stegmann et al., 2007) can directly benefit from NLP techniques for automatic argument detection, classification, and summarization. Instead of relying on scripts (Dillenbourg and Hong, 2008; Scheuer et al., 2010; Fischer et al., 2013) or explicit argument diagramming (Scheuer et al., 2014), collaborative platforms can further provide scholars with a summary of the whole argumentation to the topic, reveal the main argumentative patterns, provide the weaknesses of other's arguments, as well as identify shortcomings that need to be improved in the argumentative knowledge construction. Automatic analysis of micro-arguments can also help to overcome the existing trade-off between freedom (free-text option) and guidance (scripts) (Dillenbourg and Hong, 2008).

7.2 Automatic summarization of argumentative discourse

When summarizing argumentative discourse, knowledge of the underlying structure of the argument is a valuable source. Previous work in this

area includes, e.g., opinion-based summarization of blogposts (a pilot task in TAC 2008[5]). Carenini and Cheung (2008) compared extractive and abstractive summaries in controversial documents and found out that a high degree of controversiality improved performance of their system. Similarly, presenting argumentation in a condensed form (the large concepts of the argument are compressed or summarized) may improve argument comprehension. This approach would mainly utilize tools for document compression (Qian and Liu, 2013).

8 Conclusions

In this article, we formulated our view on argumentation mining on the Web and identified various use-case scenarios and expected outcomes. We thoroughly reviewed related work with focus on Web data and annotation approaches. We proposed two different annotation schemes based on their theoretical counterparts in argumentation research and evaluated their suitability and reliability for Web data in two extensive independent annotation studies. Finally, we outlined challenges and gaps in current argumentation mining on the Web.

Acknowledgments

A Annotated examples

A.1 News articles using Claim-Premises scheme

Example 1

[*claim:* „Die Umstellung zu G8 war schwierig“, sagt Diana.] [*support:* In den Sommerferien nach dem Sitzenbleiben holte sie das nach, was ihr die G8er voraus hatten: Lateinvokabeln, Stochastik, Grammatik. „Den Vorteil, durch das Wiederholen den Stoff noch mal zu machen, hatte ich nicht.“]

[*claim:* "The change [to G8] was difficult," says Diana.] [*support:* (Since) After staying down, she had to catch up with the G8 students during her summer holiday, studying Latin vocabulary,

stochastics, and grammar. "I did not have the advantage of reviewing previous material."]

Example 2

[*claim:* Lehrer wird man, weil das ein sicherer Beruf ist.] [*support:* So denken noch immer viele junge Leute, die sich für eine Pädagogenlaufbahn entscheiden. Gut acht von zehn Erstsemestern, die 2009 mit einem Lehramtsstudium anfingen, war dieser Aspekt ihres künftigen Berufs wichtig oder sogar sehr wichtig. Keine andere Studentengruppe, die die Hochschul-Informations-System GmbH HIS befragte, legt so viel Wert auf Sicherheit.]

[*claim:* People become teachers because it is a safe job.] [*support:* This is what more and more young people who decide to become a teacher think. Well over eight of 10 freshman students who started to study to become teachers in 2009 considered this an important or very important aspect. No other group of students interviewed by the HIS set that much value on safeness.]

Example 3

[*claim:* Für die Unis sind Doktoranden günstige Arbeitskräfte.] [*support:* Eine Bekannte hatte mit ihrem Doktorvater zu kämpfen, der versuchte, sie noch am Institut zu halten, als ihre Arbeit längst fertig war. Er hatte immer neue Ausreden, weshalb er noch keine Note geben konnte. Als sie dann auch ohne Note einen guten Job bekam, auerhalb der Uni, spielte sich eine Art Rosenkrieg zwischen den beiden ab. Bis heute verlangt er von ihr noch Nacharbeiten an der Dissertation. Sie schuftet jetzt spätabends und am Wochenende für ihren Ex-Prof, der natürlich immer nur an ihrem Fortkommen interessiert war.]

[*claim:* At university, graduate students are cheap employees.] [*support:* An acquaintance struggled with her Ph.D. supervisor, who tried to keep her in his group at any rate, even though she had already completed her thesis. He pled more and more excuses for not yet grading her work. When she finally found a good job outside university even without a final grade a martial strife arose. Still today, he asks her to rework her dissertation. Now, she is drudging for her ex-supervisor, who always only wanted the best for her, late in the evening or on the weekend.]

A.2 Forum posts using extended Toulmin's scheme

Example 1

[*backing:* I'm a regular education teacher. I have students mainstreamed into my class every year.] [*grounds:* My opinion is that it needs to be done far more judiciously than it is done now- if six exceptional children are put in my class, that is the equivalent of putting an entire special ed classroom into my regular class.] [*grounds:* I personally feel like these kids are shortchanged- some of them are good kids who need an adult close by and able to give more focused attention. In a class of 30+, this isn't going to happen consistently.] [*grounds:* And some of the ones who come to me have legally imposed modifications, some of which have little or no bearing on what I teach, so I am not allowed to handle my class in a way I think it should be done. That impairs my efficiency as an educator.] [*grounds:* Also, some have so many modifications that for all intents and purposes they are merely taking a special ed class whose physical location just happens to be in a regular classroom.] [*claim:* From my point of view, mainstreaming is not a terrible idea, but it is lamentable in its execution, and because of that, damaging in its results.]

Comments Quite a good argument with an explicit claim, few grounds and some backing.

Example 2

tara_mommy:
I agree with you too, which is why I said:
[*rebuttal:* There are obviously cases where this isn't going to work. Extreme behavioral trouble, kids that just aren't able to keep up with what they're learning in average classes, etc.] [*claim:* But on the whole, I like mainstreaming.]

Comments Only claim and rebuttal; no supporting grounds.

Example 3

I think as parents of the child you have to be certain and confident that your child is ready to mainstream. If not, it can backfire on the child. [*backing:* My child was in "preschool handicapped" from age 2-5. We tried to mainstream him in kindergarten, but he had a hard time adjusting. So the school got him a one on one para and it helped a bit. 2 grades later, he still has

a one on one aide but doing EXCELLENT.]
Our goal is for him to not have a one on one by middle school. We took him off meds and we have a strong behavior plan, he sees therapists, and it is hourly teaching and redirecting with him. Truth be told College may not be in his future, but we will do everything in our power to try to get him there.

Comments The claim is implicit, the author is slightly against mainstreaming. Mainly storytelling, which is not considered as grounds but as backing. The typos (using 'l' instead of 'I') are kept uncorrected.

Example 4

My lo has mild autism, he has only just been diagnosed, he is delayed in some areas (but not others), he goes to ms school, and has some one to one (this should increase now, I hope). There is one TA and a full time TA who supports another child with autism. It's a smallish school.
He isn't disruptive (well he sometimes doesn't do as asked and can be a little akward), he has never been aggressive in anyway, he is very happy.
I am worried about his future (high school)after reading this.
Sarah x

Comments Not an argumentative/persuasive text.

References

Corina Andone. 2012. Bermejo-Luque, Lilian. Giving Reasons. A Linguistic-Pragmatic Approach to Argumentation Theory. *Argumentation*, 26(2):291–296.

Aristotle and George Kennedy (translator). 1991. *On Rhetoric: A Theory of Civil Discourse*. Oxford University Press.

Bal Krishna Bal and Patrick Saint-Dizier. 2010. Towards Building Annotated Resources for Analyzing Opinions and Argumentation in News Editorials. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 1152–1158. European Language Resources Association (ELRA).

Jamal Bentahar, Bernard Moulin, and Micheline Bélanger. 2010. A taxonomy of argumentation models used for knowledge representation. *Artificial Intelligence Review*, 33:211–259.

Philippe Besnard and Anthony Hunter. 2008. *Elements of argumentation*, volume 47. MIT press Cambridge.

Douglas Biber and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press.

J. Anthony Blair. 2004. Argument and its uses. *Informal Logic*, 24:137151.

Elena Cabrio and Serena Villata. 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.

Elena Cabrio, Sara Tonelli, and Serena Villata. 2013a. From Discourse Analysis to Argumentation Schemes and Back: Relations and Differences. In João Leite, Tran Cao Son, Paolo Torroni, Leon Torre, and Stefan Woltran, editors, *Proceedings of 14th International Workshop on Computational Logic in Multi-Agent Systems*, volume 8143 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg.

Elena Cabrio, Serena Villata, and Fabien Gandon. 2013b. A support framework for argumentative discussions management in the web. In Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, *The Semantic Web: Semantics and Big Data*, volume 7882 of *Lecture Notes in Computer Science*, pages 412–426. Springer Berlin Heidelberg.

Giuseppe Carenini and Jackie Chi Kit Cheung. 2008. Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controversiality. In *Proceedings of the Fifth International Natural Language Generation Conference*, INLG '08, pages 33–41, Stroudsburg, PA, USA. Association for Computational Linguistics.

Marilyn J. Chambliss. 1995. Text cues and strategies successful readers use to construct the gist of lengthy written arguments. *Reading Research Quarterly*, 30(4):778–807.

Alexander Conrad, Janyce Wiebe, and Rebecca Hwa. 2012. Recognizing arguing subjectivity and argument tags. In Roser Morante and Caroline Sporleder, editors, *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, pages 80–88, Jeju Island, Korea. Association for Computational Linguistics.

Jim Crosswhite, John Fox, Chris Reed, Theodore Scaltsas, and Simone Stumpf. 2004. Computational models of rhetorical argument. In Chris Reed and Timothy J. Norman, editors, *Argumentation Machines*, volume 9 of *Argumentation Library*, pages 175–209. Springer Netherlands.

Aldo de Moor, Lilia Efimova, and Aldo De Moor. 2004. An Argumentation Analysis of Weblog Conversations. In *Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004)*, volume 197, pages 1–16.

Pierre Dillenbourg and Fabrice Hong. 2008. The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning*, 3(1):5–23.

Phan Minh Dung. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321 – 357.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 987–996, Stroudsburg, PA, USA. Association for Computational Linguistics.

Frank Fischer, Ingo Kollar, Karsten Stegmann, and Christof Wecker. 2013. Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, 48(1):56–66.

Eirini Florou, Stasinos Konstantopoulos, Antonis Koukourikos, and Pythagoras Karampiperis. 2013. Argument extraction for supporting public policy formulation. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 49–54, Sofia, Bulgaria. ACL.

Austin J. Freeley and David L. Steinberg. 2008. *Argumentation and Debate*. Cengage Learning, Stamford, CT, USA, 12th edition.

James B Freeman. 1991. *Dialectics and the macrostructure of arguments: A theory of argument structure*, volume 10 of *Trends in Linguistics*. De Gruyter.

Kallirroi Georgila, Ron Artstein, Angela Nazarian, Michael Rushforth, David Traum, and Katia Sycara. 2011. An annotation scheme for cross-cultural argumentation and persuasion dialogues. In *Proceedings of the SIGDIAL 2011 Conference: the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 272–278, Portland, Oregon. Association for Computational Linguistics.

A. Francisca Snoeck Henkemans. 2000. State-of-the-art: The structure of argumentation. *Argumentation*, 14(4):447–473.

Anthony Hunter. 2013. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, January.

Anthony Hunter. 2014. Probabilistic qualification of attack in abstract argumentation. *International Journal of Approximate Reasoning*, 55(2):607–638, January.

Ralph H Johnson. 2000. *Manifest rationality: A pragmatic theory of argument*. Routledge.

Roland Kluge. 2014. Automatic Analysis of Arguments about Controversial Educational Topics in Web Documents, Master Thesis, Ubiquitious Knowledge Processing Lab, TU Darmstadt.

Klaus Krippendorff. 2013. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications, 3rd edition.

Meredith Larson, M. Annae Britt, and Aaron Larson. 2004. Disfluencies in comprehending argumentative texts. *Reading Psychology*, 25:205–224.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. Context dependent claim detection. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, August. To appear.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Clare Llewellyn, Claire Grover, Jon Oberlander, and Ewan Klein. 2014. Re-using an Argument Corpus to Aid in the Curation of Social Media Collections. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 462–468.

Lilian Bermejo Luque. 2011. *Giving Reasons: A Linguistic-Pragmatic Approach to Argumentation Theory*, volume 20 of *Argumentation Library*. Springer Netherlands.

Nitin Madnani, Michael Heilman, Joel Tetreault, and Martin Chodorow. 2012. Identifying high-level organizational elements in argumentative discourse. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 20–28, Stroudsburg, PA, USA. Association for Computational Linguistics.

William C. Mann and Sandra A. Thompson. 1987. Rhetorical structure theory: A theory of text organization. Technical report, Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA.

Raphaël Micheli. 2011. Arguing Without Trying to Persuade? Elements for a Non-Persuasive Definition of Argumentation. *Argumentation*, 26(1):115–126, September.

Raquel Mochales and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22, April.

S. Newman and C. Marshall. 1991. Pushing Toulmin Too Far: Learning From an Argument Representation Scheme. Technical report, Xerox Palo Alto Research Center 3333 Coyote Hill Road, Palo Alto, CA 94034.

Omid Noroozi, Armin Weinberger, Harm J.a. Biemans, Martin Mulder, and Mohammad Chizari. 2013. Facilitating argumentative knowledge construction through a transactive discussion script in CSCL. *Computers & Education*, 61:59–76, February.

Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107, New York, NY, USA. ACM.

Andreas Peldszus and Manfred Stede. 2013a. From Argument Diagrams to Argumentation Mining in Texts:. *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1):1–31, January.

Andreas Peldszus and Manfred Stede. 2013b. Ranking the annotators : An agreement study on argumentation structure. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperabilty with Discourse*, pages 196–204. Association for Computational Linguistics.

Chaim Perelman and Lucie Olbrechts-Tyteca. 1991. *The New Rhetoric*. University of Notre Dame Press.

Henry Prakken. 2010. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1(2):93–124, June.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse TreeBank 2.0. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 1–4. European Language Resources Association (ELRA).

Xian Qian and Yang Liu. 2013. Fast joint compression and summarization via graph cuts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1492–1502, Seattle, Washington, USA, October. Association for Computational Linguistics.

Chris Reed and Douglas Walton. 2003. Argumentation schemes in argument-as-process and argument-as-product. In *Proceedings of the conference celebrating informal Logic*, volume 25.

Sara Rosenthal and Kathleen McKeown. 2012. Detecting Opinionated Claims in Online Discussions. In *2012 IEEE Sixth International Conference on Semantic Computing*, pages 30–37. IEEE, September.

Rodrygo LT Santos, Craig Macdonald, Richard MC McCreadie, Iadh Ounis, Ian Soboroff, et al. 2012. Information retrieval on the blogosphere. *Foundations and Trends in Information Retrieval*, 6(1):1–125.

Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M. McLaren. 2010. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1):43–102.

Oliver Scheuer, BruceM. McLaren, Armin Weinberger, and Sabine Niebuhr. 2014. Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science*, 42(2):127–157.

Edward Schiappa and John P. Nordin. 2013. *Argumentation: Keeping Faith with Reason*. Pearson UK, 1st edition.

Jodi Schneider and Adam Wyner. 2012. Identifying Consumers' Arguments in Text. In Diana Maynard, Marieke van Erp, and Brian Davis, editors, *Semantic Web and Information Extraction SWAIE 2012*, pages 31–42.

Jodi Schneider, B Davis, and Adam Wyner. 2012. Dimensions of argumentation in social media. In *Lecture Notes in Computer Science*, volume 7603, pages 21–25. Springer Berlin Heidelberg.

Jodi Schneider, Tudor Groza, and Alexandre Passant. 2013. A review of argumentation for the social semantic web. *Semantic Web*, 4(2):159–218.

Alan Sergeant. 2013. Automatic Argumentation Extraction. In *ESWC 2013*, pages 656–660. Springer-Verlag Berlin Heidelberg.

Maria Simosi. 2003. Using Toulmin's framework for the analysis of everyday argumentation: Some methodological considerations. *Argumentation*, 17:185–202.

Walter Sinnott-Armstrong and Robert J. Fogelin. 2009. *Understanding Arguments: An Introduction to Informal Logic*. Cengage Learning, 8 edition.

Christian Stab and Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*, August. To appear.

Karsten Stegmann, Armin Weinberger, and Frank Fischer. 2007. Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2(4):421–447.

Dragan Stoianovici. 2009. Formal Logic vs. Philosophical Argument. *Argumentation*, 24(1):125–133, January.

Stephen Toulmin, Richard Rieke, and Allan Janik. 1984. *An Introduction to Reasoning*. Macmillan, 2nd edition.

Stephen E. Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.

Maria Paz Garcia Villalba and Patrick Saint-Dizier. 2012. Some Facets of Argument Mining for Opinion Analysis. In Bart Verheij, Stefan Szeider, and Stefan Woltran, editors, *Proceedings of Fourth International Conference on Computational Models of Argument, COMMA 2012*.

Artem Vovk. 2013. Discovery and Analysis of Public Opinions on Controversial Topics in the Educational Domain, Master Thesis, Ubiquitious Knowledge Processing Lab, TU Darmstadt.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, and Gregor Engels. 2014a. Modeling Review Argumentation for Robust Sentiment Analysis. In *Proceedings of the 25th International Conference on Computational Linguistics COLING 2014*, page To appear, Dublin, Ireland.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, Gregor Engels, and Tsvetomira Palakarska. 2014b. A Review Corpus for Argumentation Analysis. In Alexander Gelbukh, editor, *15th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 14)*, pages 115–127. Springer.

Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

Douglas Walton. 2005. *Fundamentals of Critical Argumentation*. Critical Reasoning and Argumentation. Cambridge University Press, 1 edition.

Douglas Walton. 2012. Using Argumentation Schemes for Argument Extraction: A Bottom-Up Method. *International Journal of Cognitive Informatics and Natural Intelligence*, 6(3):33–61.

Armin Weinberger and Frank Fischer. 2006. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1):71–95, January.

Anthony Weston. 2008. *A Rulebook for Arguments*. Hackett Pub Co., 4 edition.

Yun Xie. 2012. Review of Giving Reasons. *Informal Logic*, 32(4).

Cihua Xu and Yicheng Wu. 2014. Metaphors in the perspective of argumentation. *Journal of Pragmatics*, 62:68–76, February.