A News Editorial Corpus for Mining Argumentation Strategies

Khalid Al-Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, Benno Stein
Faculty of Media, Bauhaus-Universität Weimar, Germany
`<firstname>.<lastname>@uni-weimar.de`

Abstract

Many argumentative texts, and news editorials in particular, follow a specific strategy to persuade their readers of some opinion or attitude. This includes decisions such as when to tell an anecdote or where to support an assumption with statistics, which is reflected by the composition of different types of argumentative discourse units in a text. While several argument mining corpora have recently been published, they do not allow the study of argumentation strategies due to incomplete or coarse-grained unit annotations. This paper presents a novel corpus with 300 editorials from three diverse news portals that provides the basis for mining argumentation strategies. Each unit in all editorials has been assigned one of six types by three annotators with a high Fleiss' κ agreement of 0.56. We investigate various challenges of the annotation process and we conduct a first corpus analysis. Our results reveal different strategies across the news portals, exemplifying the benefit of studying editorials—a so far underresourced text genre in argument mining.

1 Introduction

News editorials define a genre of written argumentative discourse whose main goal is persuasiveness. In a news editorial, an author states and defends a thesis that conveys his or her stance on a controversial topic usually related to the public interest. Editorials do not only aim to persuade readers of some opinion, but they often also propagate particular ideologies or recommend certain attitudes to the community, e.g., a specific action towards an upcoming event (van Dijk, 1992). To achieve persuasion, a news editorial follows a particular *argumentation strategy* that the author expects to be most suitable for the target audience, i.e., the author composes a series of claims, assumptions, and different types of evidence while using argumentative language and structure (van Dijk, 1995). This does not only cover the resort to quantitative features of text (e.g., related to lexical style, cohesion, or rhetorical structure), but it also refers to the roles, positions, flows, and relations of specific argumentative discourse units.

The study of news editorials is beneficial for several tasks, such as the creation of persuasive writing strategies for writing assistance systems and qualitative media content analysis. At the same time, the rapid expansion of online news portals increases the need for algorithms that can analyze an editorial's discourse *automatically*. The needed analyses include argumentation mining and evidence detection, both of which are studied in computational argumentation, an emergent area of computational linguistics. While several text corpora for such analyses have recently been published for different domains and genres, a respective resource with news editorials is missing to this day. Moreover, existing corpora stick to coarse-grained and/or incomplete annotations of the units of an argumentative discourse (see Section 2 for details), which renders the mining of an author's argumentation strategy impossible.

In this paper, we present a novel corpus with 300 news editorials evenly selected from three diverse online news portals: *Al Jazeera*, *Fox News*, and *The Guardian*. The aim of the corpus is to study (1) the mining and classification of fine-grained types of argumentative discourse units and (2) the analysis of argumentation strategies pursued in editorials to achieve persuasion. To this end, each editorial contains manual type annotations of all units that capture the role that a unit plays in the argumentative discourse,

Title. *An education was my path to financial security. Then I got my student loan bill.*

Editorial. *I have a very distinct memory from my first day of college: My family's minivan slowly pulling into my dormitory's parking lot, through a crowd of first-year students flanked by helicopter parents and, in retrospect, probably hungover orientation week advisers. I remember thinking "Hurry up! I'm ready to start my real life."*

I had no idea what I was really rushing towards.

As the only daughter of Nigerian immigrants with a tenuous-at-best toehold on the middle class, college was billed as the only path to financial security. *"No one can ever take away your education," my father would say repeatedly.* While that may be true, *two degrees later someone could take away my access to decent housing because of my shit credit, thanks to the nearly $60,000 in student loans I've essentially defaulted on since graduating from the University of Chicago and Northwestern University.*

It seems a college education is part of the American dream that's easy to buy (or borrow) into, but hard to pay off.

With tuition soaring, and the middle class shrinking along with their incomes, many students and their families are left holding incredibly expensive bags. *In 2013, 69% of graduating seniors at public and private nonprofit colleges took out student loans to pay for college*, and *"about one-fifth of new graduates' debt was in private loans," according to the Project on Student Debt.* Even public schools - long considered a more affordable option - are less accessible: *public colleges increasingly rely on tuition dollars as state funding continues to fall (25% and 23%, respectively, in 2012, compared to 17% and 23% in 2003).* *The country's cumulative student loan debt ($1.1tn) has surpassed car loans ($875bn) and credit card debt ($659bn).* Though college graduates make more than their peers who only graduated from high school, for many, monthly student loans leach into that extra $17,500 in salary.

Yet the party line that college education is the middle class' only hope for upward mobility persists - *it will even be the message of President Obama's last stop on his "SOTU Spoiler" tour in Knoxville, Tennessee.*

"In today's economy," Dan Pfeiffer, the president's senior advisor, wrote on Medium, "access to a college education is the surest ticket to the middle class -- and the President's proposals will help more young people punch that ticket."

As someone who punched that ticket twice, I'm still waiting for my express bus to the middle class. *The modest income I make as an entrepreneur with a day job is whittled away each month thanks to loan payments (plus interest) to various financial intuitions that feel more like bounty hunters than supporters of middle-class aspirants.*

With that $60,000 in student loans hanging over me, I'm still waiting to start the "real" life I'd always imagined for myself. It's just that now I want one with its possibilities a little less hampered by student debt.

Types of units

- Anecdote
- Assumption
- Common ground
- Other
- Statistics
- Testimony

Figure 1: Example news editorial from the presented corpus. Each argumentative discourse unit of all 300 editorials in the corpus has been manually assigned one of six types, four of which are shown here.

such as *assumption* or *statistics*. The corpus consists of $14,313$ units of six different types, each annotated by three professional annotators from the crowdsourcing platform *upwork.com*. Figure 1 shows the type annotations of one editorial in the corpus. The editorial has been taken from The Guardian.

Based on the results of the annotation process, we analyze the agreement between annotators in order to scrutinize the major cases of disagreement as well as to designate the complex issues that humans face in classifying types of argumentative discourse units in editorials. Considering the number and complexity of the types, the obtained inter-annotator agreement of 0.56 in terms of Fleiss' κ can be seen as high. In a first brief statistical analysis of the corpus, we investigate differences in the type distribution between the three portals, which indicate divergent argumentation strategies.

To conclude, the contribution of this paper is three-fold: (1) We propose the first annotation scheme for mining argumentation strategies that captures the type of each unit of an argumentative text. Unlike previous schemes, it relies on a complete annotation of the text on a fine-grained level. (2) We introduce a novel corpus with 300 news editorials which are manually annotated according to the scheme. Despite the resort to a so far under-resourced text type, the corpus contains a considerably larger number of unit annotations than comparable existing corpora. In addition, we provide a detailed analysis of the inter-annotator agreement and of the reliability of the resulting annotations. The corpus is freely available in order to foster research on computational argumentation.[1] (3) We present insightful findings that indicate the potential of our corpus for analyzing the argumentation strategies of news editorials. As far as we know, our corpus is the first which allows studying such strategies in monological opinionated text.

2 Related Work

Many recent publications in the area of computational argumentation are concerned with the construction of annotated resources, which is as a fundamental step towards building automatic systems that analyze the argumentative structure of texts. Unlike most previous corpora, the annotations of our corpus provide a classification of argumentative units based on their content. In the most simple case, other works distinguish only argumentative discourse units from other text, as we do in (Al-Khatib et al., 2016). Some

---

[1] Webis-Editorials-16 corpus, http://www.uni-weimar.de/en/media/chairs/webis/corpora/

corpora contain unit annotations based on the role units take in arguments, e.g., premise or conclusion (Stab and Gurevych, 2014; Habernal and Gurevych, 2015). Such annotations represent the general argumentative structure, but they do not encode the means an author uses to persuade the readers.

Previous corpora that contain content-based unit annotations are not suitable for analyzing argumentation strategies due to their annotation method or to the genre of the contained texts. Similar to our corpus, the CE corpus of Wikipedia articles (Aharoni et al., 2014; Rinott et al., 2015) also considers types of evidence (anecdotes, statistics, and testimony). However, evidence is annotated only where it relates to a set of given topics, so no complete annotation of the article's discourse is provided. Park and Cardie (2014) use annotations to distinguish verifiable from non-verifiable evidence, which is not related to argumentation strategies. The corpus of Habernal and Gurevych (2016) includes logos and pathos annotations. While this would be appropriate for studying argumentation strategies, the corpus consists of short forum discussions, which follow no strategic plan.

Several corpora serve to study relations between the argumentative discourse units in a text (Boltužić and Šnajder, 2014; Ghosh et al., 2014; Peldszus and Stede, 2015; Stab and Gurevych, 2014), e.g., whether a unit supports or attacks another unit or the thesis of the text. Some patterns of such relations would certainly be relevant for an analysis of argumentation strategies, like the rebuttal of a common ground that seems to counter the author's stance. However, similar to the low number of attack relations in persuasive essays (Stab and Gurevych, 2014), we found few insightful patterns of this kind in editorials.

While we focus on the argumentation strategy of an entire text, the 60 schemes of Walton et al. (2008) can be seen as representing the strategy of single arguments, although at a much finer granularity. Annotating a complete argumentative text according to all these schemes is very difficult. The only available resource with scheme annotations we are aware of, the Araucaria corpus (Reed and Rowe, 2004), contains annotations of single arguments but not of whole texts.

The construction of the corpus introduced in this paper has already been sketched in (Kiesel et al., 2015). Due to the pure form of argumentation found in news editorials, we discuss suitability of the corpus as a resource for shared tasks on argument mining there. Most existing work on news editorials in computational linguistics studies sentiment and opinions (Yu and Hatzivassiloglou, 2003; Wilson and Wiebe, 2003; Bal, 2009). A first conceptual study of the relation between the opinions in a news editorial and its argumentative structure is described in (Bal and Saint-Dizier, 2009). Besides, to our knowledge, the only work in this regard is the very recent work of Chow (2016) on Chinese editorials. Unfortunately, the annotation of this corpus is restricted to the argumentativeness of paragraphs as a whole, which makes the corpus unsuitable for analyzing argumentation strategies.

3 Model

We now introduce the model (in terms of annotation scheme) that we propose for analyzing the argumentation strategy of a news editorial. The model is focused on the sequential structure of argumentative discourse, which benefits a reliable statistical recognition of strategy patterns. To this end, we separate an editorial into argumentative discourse units of six different types where each type represents a particular role in the discourse. While our model is in line with related work on evidence types (Rinott et al., 2015), we assign a type to each unit in order to capture an editorial's overall argumentation strategy.

In particular, we see argumentative discourse units as the smallest elements of the argumentative discourse of an editorial. They represent the propositions stated by the editorial's author to discuss, directly or indirectly, his or her thesis. In general, propositions affirm or deny that certain entities have certain attributes. An entity may be an object (e.g., *milk*), a being (e.g., *Obama* or *we*), or an abstract concept (e.g., *learning to cooperate*). Technically, we define a unit based on the notion of propositions as follows:

> **Argumentative Discourse Unit:** An argumentative discourse unit is the minimum text span that completely covers one or more propositions. It always includes a subject (or a placeholder, such as "which") and a verb, and it needs to include an object if grammatically required. It spans at most one sentence.

The units of a news editorial play different roles in the editorial's argumentative discourse. E.g., some

represent knowledge or beliefs of the author or other people, and some serve as evidence in favor or against the truth of other units. We assume each unit to refer to exactly one of six types:

1. **Common Ground:** The unit states common knowledge, a self-evident fact, an accepted truth, or similar. It refers to general issues, not to specific events. Even if not known in advance, it will be accepted without proof or further support by all or nearly all possible readers.

    Example: *"History warns us what happens when empires refuse to teach known values that strengthen societies and help protect them from enemies intent on their destruction."*

2. **Assumption:** The unit states an assumption, conclusion, judgment, or opinion of the author, a general observation, possibly false fact, or similar. To make readers accept it, it is, or it would need to be supported by other units.

    Example: *"For too long young people have relied on adults who have done too little to stop the violation of the rights of the children for whom they were responsible."*

3. **Testimony:** The unit gives evidence by stating or quoting that a proposition was made by some expert, authority, witness, group, organization, or similar.

    Example: *"According to The Yazidi Fraternal Organization (YFO), thousands of young Yazidi women and children are being used by ISIL as sex slaves."*

4. **Statistics:** The unit gives evidence by stating or quoting the results or conclusions of quantitative research, studies, empirical data analyses, or similar. A reference may but needs not always be given.

    Example: *"Of the total of 779 men and boys that have been detained at Guantanamo Bay since 2002, only nine have been convicted of any crime."*

5. **Anecdote:** The unit gives evidence by stating personal experience of the author, an anecdote, a concrete example, an instance, a specific event, or similar.

    Example: *"In 1973, it deployed 18,000 troops with 300 tanks to save Damascus during the 'October War'."*

6. **Other:** The unit does not or hardly adds to the argumentative discourse or it does not match any of the above classes.

    Example: *"Happy New Year!"*

Our hypothesis is that these six types suffice to capture the main structural characteristics of argumentation strategies in news editorials. At the same time, they define an annotation scheme for a fine-grained mining of argumentative discourse units. This scheme represents the basis for the corpus we constructed.

4 Corpus Construction

This section describes the construction and annotation process of the news editorial corpus based on the proposed model (Section 3). The purpose of the corpus is to study different argumentation strategies in news editorials in terms of the argumentative discourse units they use.

4.1 Data Acquisition and Preparation

Before the annotation, the editorials are selected from three diverse news portals and decomposed into clause-like segments in order to ease the annotation process to achieve scale.

Selection of Argumentative News Editorials The corpus consists of editorials from *aljazeera.com*, *foxnews.com*, and *theguardian.com*. This selection of news portals cover diverse cultures and styles. They are internationally well-known and have separate editorial sections. We randomly selected 100 editorials from each portal that (1) are published within the same short time interval (December 2014 and January 2015) to facilitate a topical overlap, (2) sparked at least a small discussion (had at least 5 comments), and (3) contain at least 250 words (to filter out texts that just pose a question instead of arguing).

Type	Editorials	Total	Mean	Std. dev.	Median	Min	Max
Tokens	**All editorials**	**287364**	**957.88**	**257.28**	**932**	**298**	**1894**
	Al Jazeera	106430	1064.30	236.05	1033	440	1671
	Fox News	86415	864.15	226.36	855	298	1613
	Guardian	94519	945.19	267.13	906	481	1894
Sentences	**All editorials**	**11754**	**39.18**	**13.00**	**37**	**12**	**114**
	Al Jazeera	3962	39.62	10.55	38	16	75
	Fox News	3912	39.12	13.45	39	12	104
	Guardian	3880	38.80	14.65	36	18	114
Paragraphs	**All editorials**	**4664**	**15.55**	**6.48**	**15**	**2**	**45**
	Al Jazeera	1896	18.96	5.15	19	7	33
	Fox News	1689	16.89	6.71	16	2	45
	Guardian	1079	10.79	4.29	10	5	31
Segments	**All editorials**	**35665**	**118.88**	**38.21**	**116**	**28**	**309**
	Al Jazeera	11521	115.21	31.68	113	32	218
	Fox News	11315	113.15	35.4	112	28	231
	Guardian	12829	128.29	44.58	122	59	309

Table 1: Distribution of tokens, sentences, paragraphs, and segments in the corpus before annotation.

Pre-Segmentation of Argumentative Discourse Units   To allow for an annotation at larger scale, we automatically segmented the editorials before the annotation but then allowed annotators to merge adjacent segments to discard incorrect unit boundaries. In this setup, the annotators do not have to choose the exact unit boundaries, which simplifies the annotation process while making the evaluation of the annotator-agreement more intuitive. A similar manual approach was used by Park and Cardie (2014).

In detail, the applied segmentation algorithm, which we will make publicly available, starts a new segment at the beginning or end of every clause not preceded by a relative pronoun. Clauses were identified using a state-of-the-art dependency parser (Manning et al., 2014) and the clause tags from the Penn Treebank Guidelines (Bies et al., 1995). The heuristic behind the segmentation was chosen based on a careful analysis of news editorials as well as of the persuasive essays corpus from (Stab and Gurevych, 2014), since essays resemble editorials in the way they compose argumentative discourse units. An evaluation of the segmentation algorithm on that corpus yielded very satisfying results: The algorithm segmented the 90 essays into 5132 segments. Only nine of these segments should have been split further, as they overlapped with several ground-truth units from the essay corpus. On the other hand, the segmentation was somewhat too fine-grained, namely, the 1552 ground-truth units were split into 3637 segments. In our setup, however, the annotators then perform the necessary segment merges. Table 1 shows statistics about the size of the corpus and its three sub-corpora after segmentation.

4.2 Annotation Process

Given the 300 selected news editorials, an annotation process was performed in order to identify all argumentative discourse units in each newspaper editorial, including an assignment of one of the six types from Section 3 to each unit. The main steps of this process are summarized in the following.

Task Definition   First, each editorial had to be read as whole in order to understand the main topic and to follow the stance of the editorial's author.

As the annotation task, one out of eight classes had to be chosen for each segment of each editorial (see pre-segmentation above): (1–6) Any of the six types of argumentative discourse units of our model from Section 3, (7) *no unit*, when the segment does not belong to a unit, and (8) *continued*, when the segment needs to be merged with subsequent segments in order to obtain a unit. In case (8), the class assigned to the last segment determines the class of the merged unit.

The annotation guidelines given to the annotators contained the type definitions from Section 3 and a few clear and controversial examples for each type. In addition, we pointed out that the correct classification of a segment may require looking at surrounding segments. Also, no distinction should be made between true and false propositions (e.g., a wrong testimony should still be classified as testimony).

	Common ground	Assumption	Anecdote	Testimony	Statistics	Other	No unit	Continued	Overall
Fleiss' κ	0.114	0.613	0.399	0.591	0.582	0.152	0.365	0.684	**0.560**

Table 2: Inter-annotator agreement in the main annotation process, quantified in terms of Fleiss' κ.

Pilot Annotation Study   A pilot study was conducted on nine editorials to evaluate the guidelines and to select the annotators. For this, three editorials were chosen from each portal. They are not part of the corpus, but were acquired and segmented in the same fashion.

We decided to conduct the annotation process via the professional crowdsourcing platform *up-work.com* in order both to increase scalability and to obtain independent annotators. The list of candidate annotators comprised ten freelancers. All of them were native English speakers, had at least a bachelor's degree, and had already knowledge about argumentation theories from their education.

Seven annotators completed all nine editorials, taking around 30 minutes per editorial on average. The Fleiss' κ agreement score for all seven annotators was a moderate 0.433 (J. Richard Landis, 1977). As we observed remarkable drops in the agreement caused by either of three specific annotators, we decided to exclude those annotators and keep the remaining four for the main annotation study.

An error analysis of the annotation of the four annotators revealed insightful hard cases. For instance, the annotators had difficulties to distinguish between *common ground* and *anecdote* for units discussing a specific event that is well-known universally. Also, there was notable disagreement between *common ground* and *assumption*. This was expected, though, since the distinction of these two types appears more subjective than for other type combinations. Nevertheless, the agreement between the four annotators for all types was substantial with $\kappa = 0.606$. Therefore, we decided not to modify our scheme, but only to clarify the type definitions and to add some additional examples that clarify these hard cases.

Main Annotation Process   The 300 corpus editorials were evenly distributed among the four annotators. Each annotator got 225 editorials to annotate, 75 from each news portal. Accordingly, each editorial was annotated by three annotators.

4.3 Annotation Results

We analyzed the results of the annotation process in order to examine (1) the reliability of the corpus and (2) the major disagreements in units and types between the annotators. Our main findings are as follows:

Inter-Annotator Agreement   In terms of Fleiss' κ, the overall agreement is 0.56. As broken down in Table 2, however, the types *common ground* and *other* have only a slight agreement, while the annotators achieved fair agreement for *no unit* and *anecdote* as well as moderate or substantial agreement for the remaining four types. Moreover, for 94.4% of all segments at least two of three annotators agree on one type, suggesting that a resort to majority agreement is very adequate. Considering that the annotators had to decide among eight different classes for every segment, such agreement seems high in overall terms. Therefore, we conclude that the annotations of the corpus can be seen as reliable.

Disagreement Analysis   To analyze the disagreement between the annotators, we created the confusion probability matrix (CPM, Cinková et al. (2012)) for all classes shown in Table 3. Each matrix cell shows the probability of choosing the column's class, given that another annotator chose row's class. Table 3 reveals the five class-pairs where annotators are most confused between:

1. Disagreement between *other* and *assumption* (0.324). An explanation may be that the annotators interpreted the intention of the author of a respective editorial differently in some segments.

   An example unit that led to confusion is *"I just don't get it"* after another unit *"the rave reviews for the first episode make me feel like a teetotaller at a lock-in"*. The first unit could be interpreted as an implicit assumption about the reviews in the second unit, say, that the review is corrupt or hard to understand. However, it could also simply be seen as an interjection not belonging to any argument.

2. Disagreement between *common ground* and *assumption* (0.562). Although we revised our guidelines to resolve the ambiguity of these types, their distinction still seems to be hard in practice.

	Common ground	Assumption	Testimony	Statistics	Anecdote	Other	No unit	Continued
Common ground	0.129	**0.562**	0.012	0.005	**0.163**	0.012	0.075	0.042
Assumption	0.035	0.701	0.017	0.010	0.075	0.014	0.066	0.083
Testimony	0.006	0.134	0.618	0.016	0.087	0.002	0.034	0.104
Statistics	0.006	0.195	0.042	0.603	0.074	0.002	0.037	0.040
Anecdote	0.037	0.277	0.041	0.013	0.451	0.016	0.059	0.104
Other	0.018	**0.324**	0.006	0.003	0.101	0.166	**0.310**	0.073
No unit	0.008	0.114	0.008	0.003	0.027	0.023	0.440	**0.377**
Continued	0.001	0.036	0.006	0.001	0.012	0.001	0.094	0.849

Table 3: Probability confusion matrix for all pairs of annotated types of argumentative discourse units.

For example, the unit *"To see a movie legally you must leave your house, queue up, ask someone for a ticket and then sit down in the company of others"* can be viewed as *common ground* if the annotator believes that most people agree with this statement, meaning there is no need for justification. In contrast, it is viewed as an *assumption* if people are assumed to disagree to some extent, e.g., because a DVD can be bought and watched legally at home.

3. Disagreement between *common ground* and *anecdote* (0.163). Confusion between these types occurred in cases where there was a distinct fact that the editorial's author uses to support his stance.

   For example, *Iraq's Sunnis were the leading force within the Iraqi army since its foundation on January 6, 1921.* This declaration was used to support the author's claim that the Sunnis respect their army and see it as a national institution of unrivaled prestige.

4. Disagreement between *other* and *no unit* (0.310). Without clear reason, these classes seem to have been used interchangeably sometimes.

5. Disagreement between *no unit* and *continued* (0.377). The main reason for such disagreement was that the annotators dealt with discourse markers and connectives inconsistently.

   E.g., in case of the subsequent segments (1) *"According to the administration"* and (2) *"the film by Nakoula Basseley Nakoulahad sparked spontaneous riots to defend Muhammad's honor"*, the first was partly seen as *no unit*, although our guideline specified to consider such segments as one unit.

Post-Processing of the Annotations For the final version of the corpus, the corpus segments were consolidated using the *majority vote* for each segment: If at least two workers agreed on the class of a segment, the segment was classified accordingly. Else, one of this paper's authors selected one of three suggested classes. Based on the disagreement analysis and a manual inspection of the annotations, we found a few general misclassifications that could be fixed semi-automatically. While overruling some decisions of the annotators, we thereby achieve a more consistent annotation, which is crucial for learning based on the corpus. In particular, we conducted the following post-processing steps:

- A considerable number of segments was annotated as *no unit*, although it should have been merged with the next segment. We reviewed several instances of this problem, such as conditional statements (e.g., of the form "if A then B") or relations that are not argumentative but temporal or spatial (e.g., of the form "when A then B"). Where necessary, we then merged the respective segments.

- According to our definitions, only non-rhetorical questions should be labeled as *no unit*. However, many rhetorical questions were also classified as *no unit*, even though they had, in our view, a clear argumentative function: most times implicitly conveying claims, recommendations, or similar. Following our definitions, we reclassified them as *assumption*.

- Second person voice segments were often classified as *no unit*, possibly due to the unintended interpretation that a unit requires an *explicit* subject. Nearly all of them are appeals, recommendations, or similar. As above, we thus reclassified them as *assumption*.

In addition to the corrections above, we excluded periods, commas, or similar punctuation at the end of segments and put them in separate *no unit* segments. This is important to prevent unit type classifiers from misleadingly learning to identify particular types based on these characters.

Type	Total	Mean	Std. dev.	Median	Min	Max	Percent
Common ground	241	0.80	1.53	0	0	13	1.7%
Assumption	9792	32.64	12.42	32	3	86	68.4%
Anecdote	2603	8.68	9.12	7	0	77	18.2%
Testimony	1089	3.63	5.42	2	0	44	7.6%
Statistics	421	1.40	2.76	0	0	19	2.9%
Other	167	0.56	1.64	0	0	24	1.2%
All units	14313	47.71	14.28	46	14	132	100%

Table 4: The distribution of types of argumentative discourse units in the created corpus.

4.4 The Corpus

Table 4 presents some statistics of the final corpus, obtained after post-processing. We observe that the most frequent type of argumentative discourse unit is *assumption* covering almost 68.4% of all units. The *anecdote* type represents about 18.2%, surpassing the *testimony* (7.6%), *statistics* (2.9%), and *common ground* (1.7%). *Other*, finally, only refers to a very low percentage of units (1.2%). On one hand, this supports the hypothesis that editorials are a rich source for argumentation. On the other hand, it serves as strong evidence that the six proposed types of units cover most units found in editorials.

5 Argumentation Strategies

Aside from the general use of our corpus for the development and evaluation of approaches to argumentation mining, the corpus serves to investigate how authors argue in news editorials in order to persuade the readers. We do not actually analyze such argumentation strategies here. Rather, we present some basic findings that indicate the potential of our corpus for analysis in this regard.

In particular, Table 5 shows detailed statistics about the types of argumentative discourse units in the corpus. Overall, we see that the length of news editorials is quite stable across the three news portals, with a mean between 48.76 (*The Guardian*) and 52.34 units (*Fox News*). Some very short (minimum 14 units) and very long editorials (maximum 132 units) exist, though.

Regarding the distribution of the types, some general tendencies as well as some insightful differences can be observed. Generally, more than two third of an editorial usually comprises assumptions. This is not surprising, as the type *assumption* covers both claims and any other propositions that may require justification. While *The Guardian* has the highest proportion of assumptions (71.7%), it represents the median for most other types. Fox News more strongly relies on *common ground*, with more than one unit of that type on average. Even more clearly, 8.7% of all units in Fox News editorials is *testimony* evidence, about twice as many on average as in The Guardian (4.55 vs. 2.53). In contrast, Al Jazeera seems to put more emphasis on *anecdote*. At least, it spreads anecdotes across more units (21.0% of all). Interestingly, all three portals behave very similar in their resort to *statistics* at the same time.

Altogether, these numbers suggest that our corpus is worthy of being analyzed regarding argumentation strategies. While we leave such analyses to future research, we expect that especially the sequence of types of argumentative discourse units in an editorial can be decisive. Similar findings have been reported for the impact of discourse functions on the quality of essays (Persing et al., 2010), the inference of intentions from rhetorical moves (Teufel, 2014), and the generality of sentiment flows across review domains (Wachsmuth et al., 2015). Possibly, an adequate granularity (editorial level vs. paragraph level) and abstraction (all types vs. majority in paragraphs vs. ...) may have to be found for editorials, though, because the high average number of units allows for significant variance in respective sequences.

6 Discussion and Conclusion

Although news editorials are considered as one of the purest forms of argumentative text, so far few works exist in computational linguistics that study them. In this paper, we have presented the development of an annotated corpus for the mining of an editorial's argumentative discourse and the analysis of its argumentation strategy. We expect that such an analysis will contribute to the computational assessment of the quality and persuasiveness of monological argumentation. In recent work, we have analyzed the argumentative structure of persuasive essays in order to assess their argumentation quality (Wachsmuth

Type	Editorials	Total	Mean	Std. dev.	Median	Min	Max	Percent
Common ground	**All editorials**	**241**	**0.80**	**1.53**	**0**	**0**	**13**	**1.7%**
	Al Jazeera	59	0.59	0.97	0	0	5	1.2%
	Fox News	104	1.04	2.04	0	0	13	2.0%
	Guardian	78	0.78	1.36	0	0	10	1.6%
Assumption	**All editorials**	**9792**	**32.64**	**12.42**	**32**	**3**	**86**	**68.4%**
	Al Jazeera	3294	32.94	10.79	33	3	65	66.9%
	Fox News	3002	30.02	13.16	30	5	86	57.4%
	Guardian	3496	34.96	12.68	32	6	73	71.7%
Anecdote	**All editorials**	**2603**	**8.68**	**9.12**	**7**	**0**	**77**	**18.2%**
	Al Jazeera	1036	10.36	10.19	8	0	71	21.0%
	Fox News	727	7.27	6.67	6	0	37	13.9%
	Guardian	840	8.40	9.82	6	0	77	17.2%
Testimony	**All editorials**	**1089**	**3.63**	**5.42**	**2**	**0**	**44**	**7.6%**
	Al Jazeera	381	3.81	4.61	3	0	22	7.7%
	Fox News	455	4.55	7.42	2	0	44	8.7%
	Guardian	253	2.53	3.09	2	0	16	5.2%
Statistics	**All editorials**	**421**	**1.40**	**2.76**	**0**	**0**	**19**	**2.9%**
	Al Jazeera	141	1.41	2.60	0	0	17	2.9%
	Fox News	143	1.43	3.23	0	0	19	2.7%
	Guardian	137	1.37	2.37	0	0	12	2.8%
Other	**All editorials**	**167**	**0.56**	**1.64**	**0**	**0**	**24**	**1.2%**
	Al Jazeera	12	0.12	0.41	0	0	2	0.2%
	Fox News	83	0.83	1.20	0	0	5	1.6%
	Guardian	72	0.72	2.49	0	0	24	1.5%
All units	**All editorials**	**14313**	**47.71**	**14.28**	**46**	**14**	**132**	**100.0%**
	Al Jazeera	4923	49.23	12.23	48	21	81	100.0%
	Fox News	5234	52.34	15.64	50	17	123	100.0%
	Guardian	4876	48.76	16.55	46	22	132	100.0%

Table 5: Distribution of types of argumentative discourse units in the complete corpus and in the subcorpus of each news portal. Percentages refer to the proportions of units in the respective (sub-) corpus.

et al., 2016). While we focused on the logos means of persuasion there, argumentation strategies actually bring together logos, pathos, and ethos (Aristotle, 2007).

Our corpus is based on a fine-grained model (in terms of an annotation scheme) that we propose for capturing the different types of argumentative discourse units found in news editorials and in similar argumentative texts. We have detailed the annotation process and we have presented empirical evidence for typical characteristics of editorials and their variance across news portals. While not being considered in the model, we point out that units sometimes may have different types, e.g., they may represent both testimonial and statistical evidence. To create a clear classification setting, we decided to assign exactly one type to each unit, though, and to give the annototators guidelines about what type to prefer in what context. Partly, they are already encoded in the presented type definitions. Aside from that, we observed rather low agreement for the infrequent type *common ground*. Still, the resulting annotations may be valuable for research questions related to argumentation quality or persuasiveness. For instance, some authors use specific terms such as "in fact" or "for sure" before assumptions to let them appear as common ground. Our corpus helps to detect such cases.

In general, there is room to extend our corpus in future work. Most evidently, a deeper analysis of the argumentative discourse of news editorials will need to consider the relations between units that make up arguments. Among others, the relations will also reveal what are the main claims of an editorial. Unlike related text genres such as persuasive essays, however, editorials usually have no clear hierarchical argumentative structure, partly due to a frequent resort to enthymemes. This makes a consistent annotation of relations very challenging. We thus decided to focus on the types of units in the given form of the corpus, thereby ending up with a reliable corpus of reasonable size that we will make freely available. Recently, the need for according benchmark corpora was discussed in the emerging community of computational argumentation in order to allow for shared tasks and similar (Gurevych et al., 2016).

References

Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics. In *Proceedings of the First Workshop on Argumentation Mining*, pages 64–68, Baltimore, Maryland. Association for Computational Linguistics.

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein. 2016. Cross-Domain Mining of Argumentative Text through Distant Supervision. In *15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 16)*, pages 1395–1404. Association for Computational Linguistics, June.

Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse (George A. Kennedy, Translator)*. Clarendon Aristotle series. Oxford University Press.

Bal Krishna Bal and Patrick Saint-Dizier. 2009. Towards and Analysis of Argumentation Structure and the Strength of Arguments in News Editorials. In *AISB Symposium on Persuasive Technologies*, pages 55–63.

Bal Krishna Bal. 2009. Towards an analysis of opinions in news editorials: How positive was the year? In *Proceedings of the Eighth International Conference on Computational Semantics*, IWCS-8 '09, pages 260–263, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ann Bies, Mark Ferguson, Karen Katz, Robert MacIntyre, Victoria Tredinnick, Grace Kim, Mary Ann Marcinkiewicz, and Britta Schasberger. 1995. Bracketing Guidelines for Treebank II Style Penn Treebank Project. Technical report, University of Pennsylvania.

Filip Boltužić and Jan Šnajder. 2014. Back up your Stance: Recognizing Arguments in Online Discussions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 49–58, Baltimore, Maryland, June. Association for Computational Linguistics.

Marisa Chow. 2016. Argument identification in chinese editorials. In *NAACL Student Research Workshop 2016*. Association for Computational Linguistics.

Silvie Cinková, Martin Holub, and Vincent Kríž. 2012. Managing Uncertainty in Semantic Tagging. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '12, pages 840–850, Stroudsburg, PA, USA. Association for Computational Linguistics.

Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, and Matthew Mitsui. 2014. Analyzing Argumentative Discourse Units in Online Interactions. In *Proceedings of the First Workshop on Argumentation Mining*, pages 39–48, Baltimore, Maryland, June. Association for Computational Linguistics.

Iryna Gurevych, Eduard H. Hovy, Noam Slonim, and Benno Stein. 2016. Debating Technologies (Dagstuhl Seminar 15512). *Dagstuhl Reports*, 5(12):18–46.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127–2137.

Ivan Habernal and Iryna Gurevych. 2016. Argumentation mining in user-generated web discourse. *Computational Linguistics*, page (in press). Submission received: 2 April 2015; revised version received: 20 April 2016; accepted for publication: 14 June 2016. Pre-print available at http://arxiv.org/abs/1601.02403.

Gary G. Koch J. Richard Landis. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174.

Johannes Kiesel, Khalid Al-Khatib, Matthias Hagen, and Benno Stein. 2015. A Shared Task on Argumentation Mining in Newspaper Editorials. In *Proc. of the 2nd Workshop on Argumentation Mining*, pages 35–38.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Joonsuk Park and Claire Cardie. 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.

Andreas Peldszus and Manfred Stede. 2015. Joint prediction in mst-style discourse parsing for argumentation mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 938–948, Lisbon, Portugal, September. Association for Computational Linguistics.

Isaac Persing, Alan Davis, and Vincent Ng. 2010. Modeling Organization in Student Essays. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 229–239.

Chris Reed and Glenn Rowe. 2004. Araucaria: Software for Argument Analysis, Diagramming and Representation. *International Journal on Artificial Intelligence Tools*, 13.

Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim. 2015. Show Me Your Evidence – An Automatic Method for Context Dependent Evidence Detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 440–450.

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*, pages 1501–1510.

Simone Teufel. 2014. Scientific Argumentation Detection as Limited-Domain Intention Recognition. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and Natural Language Processing*, pages 101–109.

Teun A. van Dijk. 1992. Racism and argumentation:"Race Rio" Rhetoric in Tabloib Editorials . In *In van Emeren et al. (Eds.), Argumentation illuminated*.

Teun A. van Dijk. 1995. Opinions and Ideologies in Editorials. In *Proceedings of the 4th International Symposium of Critical Discourse Analysis, Language, Social Life and Critical Thought*, Athens, June.

Henning Wachsmuth, Johannes Kiesel, and Benno Stein. 2015. Sentiment Flow – A General Model of Web Review Argumentation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 601–611, Lisbon, Portugal.

Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2016. Using Argument Mining to Assess the Argumentation Quality of Essays. In *Proceedings of the 26th International Conference on Computational Linguistics*.

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

Theresa Wilson and Janyce Wiebe. 2003. Annotating opinions in the world press. In *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, pages 129–136, Stroudsburg, PA, USA. Association for Computational Linguistics.